CURRENT CONCEPTS AND INNOVATIVE RESEARCH IN ENGINEERING

All Sciences Academy

CURRENT CONCEPTS AND INNOVATIVE RESEARCH IN ENGINEERING

Editor
Asst. Prof. Dr. Umut ÖZKAYA

Current Concepts and Innovative Research in Engineering
Editor: Asst. Prof. Dr. Umut ÖZKAYA

Design: All Sciences Academy Design
Published Date: December 2025
Publisher's Certification Number: 72273
ISBN: 978-625-8536-30-0

© All Sciences Academy

www.allsciencesacademy.com
allsciencesacademy@gmail.com

CONTENT

1. Chapter **10**

Thermal Behavior and Effects of C Rates on Performance in Electric Vehicle Batteries

Ahmet MAVİ, Oğuz ARSLAN

2. Chapter **25**

Multilayer Perceptron Approach of Artificial Neural Networks in Fruit Mass Estimation

Sultan GÜNGÖR, Bünyamin DEMİR

3. Chapter **40**

Analyzing Ship Agency Selection Criteria in Maritime Transport Using Fuzzy AHP

Ercan YÜKSEKYILDIZ

4. Chapter **54**

Optimization of the Tribological Performance of Different Coating Types on the Piston-Ring Pair Using the Taguchi Method

Idris CESUR

5. Chapter **70**

Usability of Radiant Curing System for the Production of Aerospace Composites

Mustafa TAŞYÜREK, Soner ŞEN

6. Chapter **87**

An Research On Soil Microstructure

Nazile URAL

7. Chapter **111**

High Entropy Composites

Necati KOÇAK

8. Chapter **133**

Lightweight High Entropy Alloy

Necati KOÇAK

9. Chapter **159**

Ramjet Engines, Thermodynamic Analysis and Performance Evaluation

Soner ŞEN, Mustafa TAŞYÜREK

10. Chapter **177**

Study of Specific Heat Capacities of $CuGaSe_2$ and $CuAlS_2$ Semiconductors Using Einstein- Debye Approximation

Zafer DOĞAN, Tural MEHMETOĞLU

11. Chapter **193**

The Impact of Conducting Polymer Derivatives and Their Composites on Supercapacitor Performance

Nursel KARAOĞLAN

12. Chapter **208**

FPGA-Based Real-Time Dynamic System Accelerators: Modeling, Numerical Design, And Timing

Sertaç YAMAN, Mehmet Latif LEVENT

13. Chapter **226**

A Fully Automated LIBS–Machine Learning Workflow for Forensic Document Authentication using Mutated-DBSCAN

Mohammed Abdulbasit AL-AMERİ, Bünyamin CİYLAN, Basim MAHMOOD

14. Chapter **240**

Optimization of Financial Processes in ERP Systems with Artificial Intelligence

Murat DURMUS, Fatma AKALIN, Ismail Hakki CEDIMOGLU

15. Chapter **253**

Trajectory Control of a Constant-Load Servo System Using Fuzzy Iterative Learning PID (Fuzzy IL-PID) Controller

Mehmet Latif LEVENT, Sertaç YAMAN

16. Chapter **266**

Energy Management and Protection in Shipboard Dc Microgrids: A Critical Review of RB, MPC, and DRL Approaches

Mehmet Şefik ÜNEY

17. Chapter **286**

A Low-Cost, Modular IoT-Based Smart Parking System: Design, Prototyping, and Validation

Hadj ZERROUKI, Salima AZZAZ-RAHMANI

18. Chapter **302**

Wearable Smart Health Technology: Smart Bra

Makbule FİDAN, Mihriban KALKANCI

19. Chapter **315**

Investigation of Static and Dynamic Capillary Water Absorption in Travertines Under Water and Saline Conditions

Mustafa Yavuz ÇELİK

20. Chapter **336**

Effects of Freeze-Thaw Cycles in Pure and Salty Water on the Physical and Mechanical Properties of İscehisar Andesites

Mustafa Yavuz ÇELİK

21. Chapter **356**

Machine Learning-Based Prediction of Success Factors in Turkish Crowdfunding Startups

Nabil Abdusalam Abdelmutaleb IBRAHIM, Ömer KARAL, Ali DEGİRMENCİ

22. Chapter **381**

Utilization of Bio-Derived Eggshell Powder in Brake Friction Composites: Tribological Property Evaluation

Banu SUGÖZÜ, İlker SUGÖZÜ

23. Chapter **399**

Deep Learning–Based Vision Perception and Video Analytics

Emin GÜNEY

24. Chapter **411**

Deep Learning Architectures for Context Modeling: A Comprehensive Analysis of Recurrence, Gating and Attention Mechanism-Based Approaches

Hilal ÇELİK, Ramazan KATIRCI

25. Chapter **429**

Environmental Pharmaceutical Pollution: Ecosystems, Ecotoxicological Impacts, and Management Strategies

Buse KOCAER BİBER, Fatma Olcay TOPAÇ

26. Chapter **454**

Evaluating the Thermal Comfort Conditions in Traditional Mud Buildings of South-Eastern Nigeria

Akubue Jideofor ANSELM

27. Chapter **469**

Investigation of drilling parameters of CNT and graphene doped hybrid carbon-aramid fibre reinforced composite plates

Serkan ERDEM

28. Chapter **496**
Large Language Models: Foundations, Architectures and Applications
Tülay YILDIRIM

29. Chapter **511**
Energy and Exergy Analysis of Coal Combustion
M. Özgün KORUKÇU

30. Chapter **529**
TCDC: Thermal Circuit Designer and Calculator
M. Özgün KORUKÇU

31. Chapter **548**
A Simple Deep Learning Application for Apple Disease Classification
Nadir İLTEN

32. Chapter **562**
Investigation of The Effect of Shot Peening Process on The Mechanical and Surface Properties of Aluminum Materials
Ahmet DEMİRER

33. Chapter **583**
Scientific Advances in Sustainable Carbon Sources and Adsorption Technologies: A Bibliometric Analysis
Hatice ERDEM

34. Chapter **605**
Noise Management for Sustainable Urban Living
Melike YALILI KILIC, Semanur OZTURK

35. Chapter **622**
Mechanical Performance and Cost Analysis of Pumice Blocks with Afsin–Elbistan Fly Ash Addition
Mustafa EKEN

36. Chapter **630**
Geophysical Investigations of the Seamounts in the eastern Mediterranean
Sercan KAYIN, M. Nuri DOLMAZ

37. Chapter **646**
Transient Heat Transfer Analysis in Fuel Cell Plates Using a Nodal Network Approach
Hüseyin KAHRAMAN

38. Chapter **662**

Energy Balance Analysis of Layered Structures in Proton Exchange Membrane Fuel Cells

Hüseyin KAHRAMAN

39. Chapter **677**

Homogeneous Charge Compression Ignition (HCCI): A Comprehensive Overview of Performance, Emissions, and Future Prospects

Adem TÜYLÜ

40. Chapter **692**

Effects of Air Excess Ratio on Emissions in SI and CI Engines

Adem TÜYLÜ

41. Chapter **705**

Comparison of the Performance of LiBr- H_2O and NH_3–H_2O Fluid Pairs in Absorption Cooling Systems

Gamze ŞAHİN, Ahmet COŞKUN

42. Chapter **718**

Adsorption of Drugs with Zeolite Structures

Amani MARABEH, İbrahim İNANÇ

43. Chapter **732**

Integrating Human Factors into Operational Excellence: A FUCOM-Based Multidimensional Consistency-Optimized Approach in Socio-Technical Systems

Burcu YILMAZ KAYA

44. Chapter **752**

Industrial Energy Management: ISO 50001, Energy Auditing, and a PDCA Based Implementation Approach

Eren UYAN

45. Chapter **773**

Digital Work Study: A Multi-Dimensional Framework and Research Agenda

Erkan Sami KOKTEN

46. Chapter **792**

Machine Learning–Based Housing Price Prediction in King County: An Interpretable XGBoost Framework

Ersin NAMLI, Mustafa GÜLER

47. Chapter **813**

Interpretable and Explainable Machine Learning in Healthcare Decision Support Systems

Ferda BOZKURT

48. Chapter **837**

Functional Bioplastics: Next-Generation Materials with Active, Smart, and Barrier Properties

H. Duygu BİLGEN

49. Chapter **850**

Fundamental Principles of Construction Project Management

Abdullah Emre KELEŞ, Hümeyra TOPÇUDAN

50. Chapter **865**

Numerical Analysis of Different Bullet Impact Tests on Aluminum Plates of Varying Thicknesses

Serkan ERDEM, Mehmet Erbil ÖZCAN

51. Chapter **882**

Nano-TiO2 as Tribological Modifiers in Metal Matrix Composites: A State-of-the Art Review

Ali KAFALI, Mehmet Tayyip ÖZDEMİR

52. Chapter **902**

Frequency-Adjustable Antenna Design for RF Energy Harvesting Applications

Fettah KINAL, Mahmut Ahmet GOZEL, Mesud KAHRİMAN, Ozlem COSKUN

53. Chapter **915**

Accelerated Pavement Testing on Concrete Pavement

Muhammet Vefa AKPINAR

54. Chapter **926**

Assessment of Indoor Particulate Matter Concentrations in Supermarkets and Fitness Centers

Özlem ÖZDEN ÜZMEZ, Sıray SAĞDIÇ

55. Chapter **948**

Understanding Landslide Hazards: Causes and Assessment Methods

Samet BERBER

56. Chapter **969**

Adsorption Study of the System of Waste Foundry Sand/Methylene Blue

Gültuğ Gamze KUTLU, Sedef DİKMEN, Semra MALKOÇ

57. Chapter **997**

Performance Degradation of Photovoltaic Panels Caused by Dust Accumulation: A Case Study in Karabük Region

Melike ADAL ÇELİK, Yunus Emre KARASU

58. Chapter **1007**

Use Of Interlayer In Asphalt Concrete

Murat KARACASU, Ümmügülsüm ÜNAL

59. Chapter **1024**

Vibrational Characteristics of Multi Layered Sandwich Plates Featuring Hybrid Auxetic Core and FG Face Layers

Kerim Gökhan AKTAŞ, Fatih PEHLİVAN

60. Chapter **1041**

Recycling Methods of Waste Lithium-Ion Batteries

Seher POLAT

61. Chapter **1067**

An Intelligent Vision-Based Occupational Health and Safety Monitoring System Using Deep Learning

Seher POLAT

62. Chapter **1088**

Innovative Solutions for Minimizing Municipal and Industrial Sludge: Towards Sustainable Wastewater Management

Mirac Nur CINER, Serdar AYDIN, H. Kurtulus OZCAN, Emine ELMASLAR ÖZBAŞ

Thermal Behavior and Effects of C Rates on Performance in Electric Vehicle Batteries

Ahmet MAVİ[1]
Oğuz ARSLAN[2]

[1]Lecturer. Bilecik Şeyh Edebali University, Vocational school,
ahmet.mavi@bilecik.edu.tr, Orcid Id: 0000-0002-2779-9575
[2]Prof. Dr. Bilecik Şeyh Edebali University, Faculty of Engineering, Department of Mechanical Engineering,
oguz.arslan @ bilecik.edu.tr, Orcid Id: 000-0001-8233-831X

ABSTRACT

Lithium-ion battery performance in electric vehicles was determined by the effects of temperature and the C-rate. Battery operating efficiency and safety are highest between 20–40°C. At <0°C, ion transport slows, and internal resistance and polarization was increased. At >45°C, the risk of SEI degradation, gas formation, and thermal runaway increases. Heat generation is due to Joule, entropic, and side reactions. As the C-rate increases, thermal load and temperature differences increase. A high C-rate at low temperatures can lead to lithium plating (risk of short circuits). Battery degradation has two primary components, calendar aging and cycle aging. Both are sensitive to temperature and the C-rate. Therefore, the dynamic management of current by the BMS using temperature feedback and the target of $\Delta T \leq 5°C$ within the pack are critical. Thermal management systems include air (simple/economical), liquid (high heat transfer), two-phase (smaller temperature difference with latent heat), and PCM-assisted TEG (Thermo electric generator) approaches. Fluid selection for two-phase applications is determined by safety, environmental profile, and battery operating temperature ranges. R1234yf is the primary recommendation for applications, R1234ze is a strong alternative, and flammability management was required for R152a.

Keywords:

INTRODUCTION

The global transition to electric vehicles (EVs) was marked a significant turning point in terms of energy efficiency and sustainable transportation goals. Lithium-ion batteries, the most critical component in this process, was widely used due to their advantages such as high energy and power density, long cycle life, and low self-discharge rate. However, the performance and safety of these batteries are highly dependent on parameters such as operating temperature and charge-discharge rate. In particular, exceeding certain limits for battery temperature can lead to negative consequences such as capacity loss, increased internal resistance, and electrode degradation (Mavi and Arslan 2024).

In battery systems, the C-rate is defined as the ratio of the charge or discharge current to the nominal capacity and was a parameter that directly affects the battery's thermal behavior. At low C-rates, the battery operates more stably, while at high C-rates, Joule heat generation due to internal resistance increases rapidly. This leads to uncontrolled battery temperature increases, thermal instability, and, in the long term, accelerates the aging process. Conversely, at low temperatures, the battery's effective capacity decreases due to slower ion transport, decreasing energy conversion

efficiency and increasing potential differences at the electrode-electrolyte interface. Therefore, the relationship between temperature and the C-rate is critical for both instantaneous performance and long-term durability (Mavi and Arslan 2025).

Studies in the literature indicate that lithium-ion batteries exhibit different behaviors depending on the temperature range and C-rate. For example, while high efficiency was achieved at nominal conditions around 25 °C, ion diffusion resistance increases significantly at temperatures below 0 °C, and at conditions above 45 °C, SEI layer thickening and gas formation was observed (Bandhauer et al., 2011). Similarly, at discharge rates of 2 °C and above, heat generation increases with the increase in internal resistance, leading to rapid fluctuations in cell temperature (Aiello et al., 2024). However, most existing studies have examined the effects of temperature and C-rate independently. The simultaneous interaction of these two parameters and their combined effect on the overall thermal stability of battery systems was not sufficiently detailed (Shi et al., 2025).

This section examines the factors affecting performance in electric vehicle batteries. The study explains the effects of temperature changes on battery efficiency, internal resistance, capacity loss, and aging behavior. Furthermore, the relationship between the C-ratio and these variables was evaluated from a thermal management perspective. This provides a fundamental assessment of the optimal temperature range and appropriate C-ratio for both battery design and vehicle-level energy management.

THERMAL BEHAVIOR OF LITHIUM-ION BATTERIES

The thermal behavior of lithium-ion batteries is critical for both energy conversion efficiency and safety. Various internal losses occur as the battery converts chemical energy into electrical energy during operation (Alipour et al., 2020). These losses increase significantly, especially at high current densities or extreme temperature conditions. Each type of loss within the battery contributes to the total heat generation, causing temperature to vary with time and location within the cell (Guo et al., 2023; Tran et al., 2022). Therefore, the thermal characteristics of the battery are a key factor in both performance analyses and thermal management system design.

Heat Generation in the Battery

Heat in lithium-ion batteries originates from three primary sources. Joule (ohmic) heat, entropic heat, and heat from side reactions, Joule heat

arises from the effect of current flowing through the internal resistance and increases rapidly, especially at high C-rates, causing the cell temperature to rise. Entropic heat arises from changes in electrode potentials during charging and discharging. In some cases, it releases heat while in others, it absorbs heat from the environment (Böhm et al., 2025). At high temperatures, chemical reactions such as SEI layer degradation, electrolyte decomposition, and gas formation generate irreversible heat. When these processes proceed uncontrolled, cell stability decreases, and a chain reaction called "thermal runaway" can occur (Jiaqiang, et al., 2024). Heat generation rates at different C-rates is given in Figure 1.

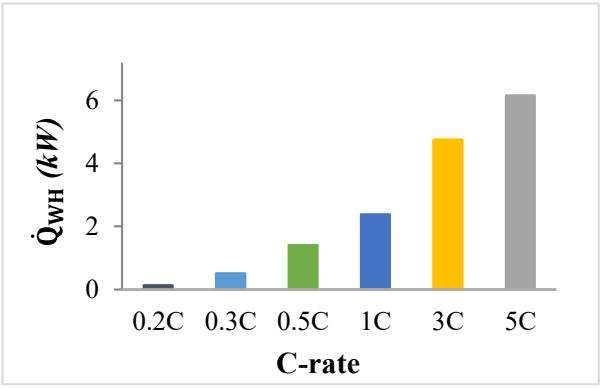


Figure. 6. Waste heat output at different C rates.

According to Figure. 6, the heat generated by the battery pack was 0.132, 0.410, 1.407, 2.458, 4.776, and 6.192 kW at 0.2C, 0.3C, 0.5C, 1C, 3C, and 5C, respectively [43].

Effect of Temperature on Battery Internal Resistance

Temperature has a direct and decisive effect on the battery's internal resistance. Cell internal resistance consists of three main components: electrolyte ionic resistance, electrode material resistance, and transition resistance at the electrode-electrolyte interface. At low temperatures, electrolyte viscosity increases, ion mobility decreases, and this leads to an increase in total internal resistance. This increased resistance leads to both capacity loss and a decrease in power output (Shen et al., 2022).

On the other hand, as temperature increases, electrolyte conductivity increases, internal resistance decreases, and short-term performance improvements was observed. However, as temperatures exceed 40°C, electrode materials begin to degrade and the SEI layer begins to chemically instability. This leads to long-term capacity loss and irreversible aging (Peled & Menkin, 2017). Therefore, maintaining the optimum temperature

range was crucial for both the battery's instantaneous performance and cycle life.

Thermal Operating Conditions of the Battery

Lithium-ion batteries operate most efficiently between 20 and 40°C; within this range, ion transport was balanced and reaction rates was controlled. Below 0°C, electrolyte fluidity decreases, ion diffusion slows, and cell potential differences increase, decreasing energy efficiency. Above 45°C, electrode surface degradation accelerates. SEI thickens, gas formation and internal pressure increase, internal resistance rises, and heating enters a vicious cycle. When this threshold is exceeded, thermal runaway can occur, which can lead to rapid temperature increases, electrolyte ignition, and the risk of fire (Mavi and Arslan 2025). Consequently, maintaining the battery temperature within the optimal range, balancing the thermal load due to the C-ratio, and maintaining thermal stability is essential. Therefore, temperature monitoring sensors and active cooling are mandatory in modern EV battery systems.

C-RATES AND THE EFFECT OF TEMPERATURE ON PERFORMANCE

The performance of lithium-ion batteries used in electric vehicles is closely related to the interaction between operating temperature and the C-rate. The C-rate indicates how quickly the battery was charged or discharged relative to its nominal capacity and directly affects both the energy efficiency and thermal behavior of the system. At higher C-rates, more current is drawn from the battery, resulting in increased thermal losses due to internal resistance, higher cell temperature, and decreased long-term electrochemical stability. In contrast, at lower C-rates, although the thermal load is reduced, the energy density and ability to meet the system's power requirements remain limited. Therefore, the optimal C-rate is decisived parameter for maintaining the balance between battery performance and thermal safety (Shi et al., 2025).

Effect of C-Rate on Battery Capacity

The battery's usable capacity is directly dependent on the applied C-rate. At low C-rates, lithium ions have sufficient time to diffuse through the electrode material, allowing the cell to discharge to full capacity. However, as the C-rate increases, ion transport rapidly becomes limiting; ions accumulate on the electrode surface, and the cycle is completed before the

active material was fully reacted. This leads to a decrease in practical capacity (Chen et al., 2022). For example, a cell providing 100% of its nominal capacity at 1C may only deliver 85-90% of its capacity when operated at 3C and 70-75% at 5C. This loss becomes more pronounced depending on temperature conditions. Because ion mobility is restricted at low temperatures, the capacity drop is much more severe at high C-rates (Luo et al., 2022). This is one of the main reasons for the range reduction in electric vehicles, especially in winter conditions.

Relationship between Temperature and C-Ratio

Temperature and C-ratio parameters are not independent of each other. On the contrary, they are two powerful variables that determine the battery's thermal and electrical responses. At high C-ratios, increasing Joule heat increases cell temperature. Increasing temperature, in turn, accelerates ion transport, improving short-term performance. However, this effect is temporary because excessive heat disrupts chemical equilibrium, leading to capacity loss and aging (Madani et al., 2025).

At low temperatures, the opposite interaction was observed. Even if the C-ratio is low, capacity loss and voltage drop occur due to the slowdown of ion diffusion. Therefore, optimizing the temperature-C-ratio balance is a fundamental task of battery management systems (BMS), especially in electric vehicles. Advanced BMS algorithms use real-time temperature measurements and current data to adjust the C-ratio and maintain the cell temperature within safe limits (Kurkin et al., 2025).

AGING (DEGRADATION) AND ITS EFFECTS ON LIFE

The performance of lithium-ion batteries decreases over time, limiting their cycle life. This decrease is due to the aging process, which results from both electrochemical and structural deterioration. The primary factors affecting battery life include temperature, C-rate, number of cycles, depth of charge/discharge (DoD), and storage conditions. Of these factors, temperature and C-rate, in particular, are the two most powerful parameters determining both short-term performance and long-term durability (Mavi and Arslan 2024).

The aging process was examined under two main headings. Calendar aging and cycle aging. Calendar aging refers to the capacity loss caused by chemical instabilities and SEI layer thickening over time, even when the battery is not in use. Cyclic aging, on the other hand, is the structural and electrochemical wear that occurs during the battery's charge-discharge

cycles. Both processes vary depending on temperature and C-rate (Mavi and Arslan 2025).

The Effect of High Temperature on Aging

High temperature is the primary factor accelerating aging in lithium-ion batteries. As temperature increases, side reactions at the electrode-electrolyte interface accelerate. The SEI layer thickens and its permeability decreases, restricting ion transport and increasing internal resistance. Electrolyte degradation can lead to gas formation and electrode swelling. With prolonged exposure to high temperatures, capacity loss, voltage shifts, and power density decline become more pronounced; for example, after 1000 cycles at 45°C, capacity retention has been reported to decrease to approximately 75%, while at 25°C, it remains at approximately 90%. Furthermore, processes such as metal ion dissolution at the cathode and SEI reformation at the anode accelerate, both increasing capacity loss and limiting safety (Feinauer et al., 2024).

The Interaction of Low Temperatures and High C-Rates

At low temperatures, the battery's internal resistance increases, ion transport slows, and polarization increases. This prevents the ions from distributing evenly across the electrodes when the battery is operated at high C-rates. Under high current densities, lithium plating (lithium plating) occurs, particularly on the anode surface. This phenomenon occurs when lithium ions accumulate as metallic lithium on the anode surface, leading to irreversible capacity loss (Tan et al., 2023).

Lithium plating not only reduces capacity but also poses a safety risk. The extension of metallic lithium dendrites over time can cause short circuits within the cell. Therefore, applying high C-rates at low temperatures significantly shortens battery life (Rahman and Alharbi, 2024). In general, the aging process is shaped by the combined effects of temperature, C-rate, and load conditions. Optimal temperature management and appropriate current profiles are effective in extending battery life.

THERMAL MANAGEMENT APPROACHES

Temperature control is essential for the safe, efficient, and long-lasting operation of lithium-ion batteries. Internal heat generated during operation creates temperature differences between cells, leading to capacity imbalances and premature aging. Therefore, thermal management systems (TMS) are a fundamental part of battery design and generally aim to maintain temperature within the 20-40°C range, reduce cell-to-cell temperature differences, and prevent the risk of thermal runaway (Mavi and Arslan 2024).

TMS processes cooling (heat removal) and heating (temperature increase in cold temperatures) within the battery. Depending on the application, passive (natural convection, PCM) or active (fan, pump, refrigerant, thermoelectric module) solutions are preferred (Mavi and Arslan 2025).

Air Cooled Systems

Air-cooled systems are the simplest and lowest-cost thermal management method. In these systems, a controlled airflow passes between the battery cells and heat is removed with the aid of a fan. Since air is used directly as the cooling medium, the system is easy to design and requires minimal maintenance.

However, the low thermal conductivity of air is a limiting factor in high-energy-density battery packs. Air cooling may be insufficient, particularly in hot ambient conditions or during high-C-rate discharges. This leads to increased temperature differences across the cell surfaces and localized overheating. Therefore, air-cooled systems are generally preferred in low- and medium-power electric vehicles, while they was used with additional support systems in high-power applications (Hwang et al., 2024).

Liquid Cooled Systems

Liquid-cooled systems are effective thermal management methods that utilize high-heat capacity coolants. In these systems, channels was placed between battery modules, and heat is transported by circulating water, glycol, or dielectric fluids. Liquid cooling has a much higher heat-carrying capacity than air, allowing for better control of temperature differences and ensuring uniform temperature distribution between cells (Akbarzadeh et al., 2021).

Cooling with two-phase fluid

Two-phase cooling utilizes the latent heat carried by the fluid during boiling and/or condensation, enabling it to maintain high heat loads at small temperature differences. Because it offers a higher heat transfer coefficient compared to single-phase cooling, it helps maintain cell temperatures within a narrow range. The primary design objective is to limit the temperature

difference between battery cells (ΔT, typically ≤3-5°C) and maintain the flow regime and operating pressure/stability within the appropriate range. This approach offers an effective solution for controlling the high temperatures that occur at high C-ratios (Mavi and Arslan 2024). The temperature distributions on the battery surface is given in figure 2.

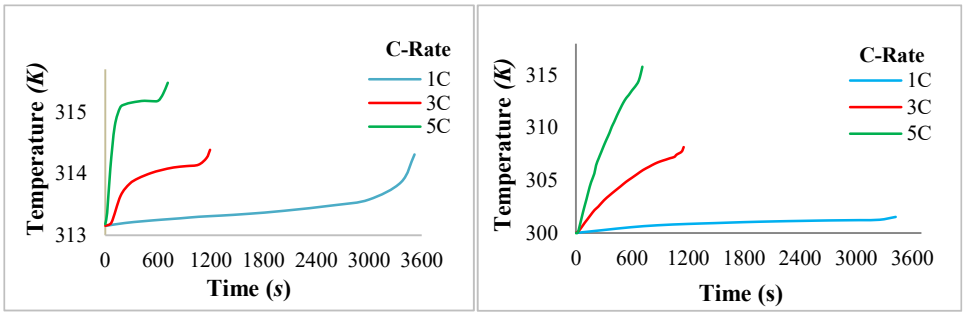


Figure 2. Cooled and uncooled battery surface temperatures (Mavi and Arslan 2024).

Figure 2 shows the time course of the battery surface temperature for discharge rates of 1C, 3C, and 5C under cooled and uncooled conditions. In the cooled condition, the two-phase cooler was operated with an inlet temperature of 40°C (313 K), resulting in a higher initial temperature. However, despite the heat flux, the temperature increase remained limited, and the temperature distribution was homogeneous (negligible at 1C, low at 3C, and a few kelvin temperature difference at 5C). In the uncooled condition, the temperature increased rapidly and progressively depending on the C-rate, with a significant early warming trend observed at 3C and especially at 5C. This comparison demonstrates that two-phase cooling effectively controls the surface temperature by controlling temperature fluctuations at high heat fluxes. The properties of the two-phase fluids used in the cooling process is given in Table 1.

Table 1. Thermophysical properties of the working fluids (Basu and Wilson, 1989; Bellair and Hood, 2019; Lemmon et al., 2010; Mavi and Arslan, 2025; Roy et al., 2020; Yu et al., 2010; Zhong et al., 2019).

Property	Unit	Refrigerant			
		R134a	**R1234yf**	**R1234ze**	**R152a**
Type	-	Wet	Dry	Dry	Wet
ODP	-	0	0	0	0
GWP	-	1200	4	6	124
ρ*	kg/m^3	4.258	4.859	4.775	2.759
c_p*	kJ/kg·K	0.85124	1.183	0.887	1.051
k*	W/m·K	0.0133	0.0134	0.0135	0.0141
$T_{critical}$	^{o}C	101	94.7	109	113.3
$P_{critical}$	kPa	4067	3382	3630	4496
Flammability	-	A1	A2L	A2L	A2
AIT	^{o}C	770	405	368	454

*At 25 ^{o}C and 1 atm, AIT: Auto ignition temperature.

The use of two-phase refrigerants was evaluated with R134a, R152a, R1234ze, and R1234yf on the basis of thermodynamic performance, environmental profile, and operational safety. The selection rests on three main criteria. First, achieving high convective heat transfer by exploiting latent heat. Second, minimizing environmental impact with ODP=0 and low GWP. Third, maintaining efficient operation over a wide temperature range. Since two-phase heat transfer is directly affected, the wet and dry vapor behaviors of the fluids were also considered, aiming for stable and efficient evaporation and condensation processes.

R134a was taken as the reference refrigerant because it remains widely used, including in Türkiye, and is classified A1 and non-flammable despite its high GWP. R152a offers strong heat transfer thanks to very low GWP and high latent heat, yet its A2 flammability makes safety measures critical. R1234ze, classified A2L and mildly flammable, stands out for sustainable applications with low GWP and balanced thermophysical properties, although its lower operating pressures can limit volumetric capacity. R1234yf was developed as the direct successor to R134a in mobile applications, providing low GWP and acceptable efficiency, and it has become widespread in the automotive sector. Overall, these four refrigerants can provide the performance balance required for sustainable and effective thermal management in electric vehicles while remaining aligned with international climate goals.

PCM AND TEG Supported Thermal Management

This section examines a hybrid approach that combines phase change materials (PCMs) with thermoelectric generators (TEGs). The objective is to passively limit fluctuations in battery temperature while recovering electrical energy from suitable temperature gradients. PCMs store a large amount of latent heat upon melting at a prescribed temperature, thereby providing passive thermal management that damps temperature excursions without additional energy input. PCM blocks placed around the cells absorb the excess heat released during charge and discharge. TEG modules are semiconductor devices that convert a temperature difference directly into electrical power. In battery cooling, TEGs can regulate the heat path and assist in removing excess heat while simultaneously enabling electricity generation from waste heat. This dual benefit can help supply auxiliary loads, particularly during high C-rate discharges when heat generation increases. Nonetheless, the electrical conversion efficiency of TEGs is typically low, so they are not sufficient on their own for large cooling demands.

Consequently, practical implementations adopt a hybrid architecture. The PCM suppresses temperature swings by clipping peak values, and the TEG harvests electricity from the resulting temperature differences. Under prolonged or successive high loads, the PCM may fully melt and lose its capacity to absorb additional heat, so PCM–TEG assemblies are commonly integrated with liquid or two-phase cooling loops (Mavi and Arslan, 2025). A comparative summary of thermal management systems is given in Table 2.

Table 2. Comparison of different thermal management systems.

Method	Advantages	Disadvantages	Area of Use
Air Cooling	Simple structure, low cost, easy maintenance	Low heat transfer, high temperature difference	Low-power EVs
Liquid Cooling	High heat transfer capacity, uniform temperature	High cost, risk of leakage	Mid- to high-power EVs
PCM and TEG-supported thermal management	PCM minimizes temperature fluctuations without wasting energy. TEG recovers electricity from waste heat; and provides stability at high loads.	PCM saturation and resolidification required at high loads for long periods, low conversion efficiency of TEG, integration/cost, and safety measures	Hybrid systems; high-C-rate transient loads, integrated EV packages with liquid or dual-phase circuits.

The selection of the thermal management method should be based on the system's energy density, environmental conditions, and cost objectives. Currently, the highest performance is achieved in hybrid structures that combine liquid cooling and PCM-TEG integrated systems.

These systems both maintain thermal stability and increase overall efficiency through energy recovery.

Consequently, the thermal management strategy in electric vehicle battery systems is a critical design parameter not only for temperature control but also for energy sustainability and safety. Appropriate thermal management is one of the most effective ways to extend battery life and increase system efficiency.

CONCLUSIONS AND RECOMMENDATIONS

Lithium-ion batteries operate most efficiently and safely between 20-40°C. As the C-ratio increases, the Joule temperature increases, and if the temperature is not controlled, performance and safety are negatively affected. At low temperatures, ion mobility decreases and internal resistance increases. Furthermore, a high C-ratio combined with high temperatures accelerates SEI degradation. A high C-ratio at low temperatures increases the risk of lithium plating. Therefore, the BMS should dynamically adjust the C-ratio using temperature feedback, target a pack-level $\Delta T \leq 5°C$, and implement climate-sensitive adaptive strategies (winter/summer). In practice, hybrid solutions with liquid cooling and PCM (with TEG support when necessary) can improve both thermal stability and efficiency. Furthermore, life models should consider the physical structure of the electrodes and the internal material degradation that change over time, along with temperature and C-ratio. Fluid selection in two-phase cooling systems should be based on criteria such as the fluids' safety classes, environmental profiles, operating pressures and temperatures, and target temperature ranges. R1234yf, with its low GWP and widespread use in the automotive ecosystem, is a practical primary candidate. R1234ze offers a strong alternative with its lower GWP and lower pressure levels (volume capacity/equipment size should be considered). Despite its good heat transfer potential, R152a requires additional safety and risk assessment due to its Class A2 flammability. Generally, fluid selection should be made by considering target temperature levels, package volume, and supply conditions.

REFERENCES

Aiello, L., Ruchti, P., Vitzthum, S., & Coren, F. (2024). Influence of pressure, temperature and discharge rate on the electrical performances of a commercial pouch li-ion battery. *Batteries*, *10*(3), 72.

Akbarzadeh, M., Kalogiannis, T., Jaguemont, J., Jin, L., Behi, H., Karimi, D., ... & Berecibar, M. (2021). A comparative study between air cooling and liquid cooling thermal management systems for a high-energy lithium-ion battery module. *Applied Thermal Engineering*, *198*, 117503.

Alipour, M., Ziebert, C., Conte, F. V., & Kizilel, R. (2020). A review on temperature-dependent electrochemical properties, aging, and performance of lithium-ion cells. *Batteries*, *6*(3), 35.

Bandhauer, T. M., Garimella, S., & Fuller, T. F. (2011). A critical review of thermal issues in lithium-ion batteries. *Journal of the electrochemical society*, *158*(3), R1.

Basu, R. S., & Wilson, D. P. (1989). Thermophysical properties of 1,1,1,2-tetrafluoroethane (R-134a). *International Journal of Thermophysics, 10*, 591-603.

Bellair, R. J., & Hood, L. (2019). Comprehensive evaluation of the flammability and ignitability of HFO-1234ze. *Process Safety and Environmental Protection, 132*, 273-284.

Böhm, K., Kondrakov, A., Markus, T., & Henriques, D. (2025). Understanding Heat Generation of LNMO Cathodes in Lithium-Ion Batteries via Entropy and Resistance. *Batteries*, *11*(10), 357.

Chen, Y., Key, J., O'regan, K., Song, T., Han, Y., & Kendrick, E. (2022). Revealing the rate-limiting electrode of lithium batteries at high rates and mass loadings. *Chemical Engineering Journal*, *450*, 138275.

Feinauer, M., Wohlfahrt-Mehrens, M., Hölzle, M., & Waldmann, T. (2024). Temperature-driven path dependence in Li-ion battery cyclic aging. *Journal of Power Sources*, *594*, 233948.

Guo, J., Guo, Q., Liu, J., & Wang, H. (2023). The polarization and heat generation characteristics of lithium-ion battery with electric–thermal coupled modeling. *Batteries,* 9(11), 529.

Hwang, F. S., Confrey, T., Reidy, C., Picovici, D., Callaghan, D., Culliton, D., & Nolan, C. (2024). Review of battery thermal management systems in electric vehicles. *Renewable and Sustainable Energy Reviews*, *192*, 114171.

Jiaqiang, E., Xiao, H., Tian, S., & Huang, Y. (2024). A comprehensive review on thermal runaway model of a lithium-ion battery: Mechanism, thermal, mechanical, propagation, gas venting and combustion. *Renewable Energy*, *229*, 120762.

Kurkin, A., Chivenkov, A., Aleshin, D., Trofimov, I., Shalukho, A., & Vilkov, D. (2025). Battery Management System for Electric Vehicles: Comprehensive Review of Circuitry Configuration and Algorithms. *World Electric Vehicle Journal*, *16*(8), 451.

Lemmon, E. W., Huber, M. L., & McLinden, M. O. (2010). *NIST reference fluid thermodynamic and transport properties-REFPROP (Version 9.0) [Computer software]*. National Institute of Standards and Technology. Gaithersburg, MD, USA.

Luo, H., Wang, Y., Feng, Y. H., Fan, X. Y., Han, X., & Wang, P. F. (2022). Lithium-ion batteries under low-temperature environment: challenges and prospects. *Materials*, *15*(22), 8166.

Madani, S. S., Shabeer, Y., Allard, F., Fowler, M., Ziebert, C., Wang, Z., ... & Khalilpour, K. (2025). A comprehensive review on lithium-ion battery lifetime prediction and aging mechanism analysis. *Batteries*, *11*(4), 127.

Mavi, A, and Arslan, O. (2024). Numerical investigation on the thermal management of Li-ion batteries for electric vehicles considering the cooling media with phase change for the auxiliary use. *Journal of Energy Storage*, *77*, 109964.

Mavi, A, and Arslan, O. (2025). A new TEG-supported thermal management system of the battery system for an electric vehicle. *Applied Thermal Engineering*, 128608.

Peled, E., & Menkin, S. (2017). SEI: past, present and future. *Journal of The Electrochemical Society*, *164*(7), A1703.

Rahman, T., & Alharbi, T. (2024). Exploring lithium-Ion battery degradation: A concise review of critical factors, impacts, data-driven degradation estimation techniques, and sustainable directions for energy storage systems. *Batteries*, *10*(7), 220.

Roy, J., Mahmud, H., Assadi, M., Iman, N., & Nikpey, H. (2020). Moving beyond gas: Can Bangladesh leapfrog and make the energy transition just by exploring the role of geothermal energy and gas infrastructure? *International Energy Journal, 20*(3A).

Shen, W., Wang, N., Zhang, J., Wang, F., & Zhang, G. (2022). Heat generation and degradation mechanism of lithium-ion batteries during high-temperature aging. *ACS omega*, *7*(49), 44733-44742.

Shi, D., Cui, Y., Shen, X., Gao, Z., Ma, X., Li, X., Fang, Y., Wang, S., & Fang, S. (2025). A review of the combined effects of environmental and operational factors on lithium-ion battery performance: Temperature, vibration, and charging/discharging cycles. *RSC Advances, 15*, 13272–13283.

Tran, M.-K., Mevawalla, A., Aziz, A., Panchal, S., Xie, Y., & Fowler, M. (2022). A review of lithium-ion battery thermal runaway modeling and diagnosis approaches. *Processes,* 10(6), 1192.

Tan, S., Shadike, Z., Cai, X., Lin, R., Kludze, A., Borodin, O., ... & Yang, X. Q. (2023). Review on low-temperature electrolytes for lithium-ion and lithium metal batteries. *Electrochemical Energy Reviews*, *6*(1), 35.

Yu, J., Xu, Z., & Tian, G. (2010). A thermodynamic analysis of a transcritical cycle with refrigerant mixture R32/R290 for a small heat pump water heater. *Energy and Buildings, 42*(12), 2431-2436.

Zhong, Q., Dong, X., Zhao, Y., Zhang, H., Wang, J., Guo, H., Shen, J., & Gong, M. (2019). Thermodynamic properties of (R1234yf + R290): Isochoric pρTx and specific heat capacity *c* measurements and an equation of state. *Journal of Chemical Thermodynamics, 129*, 36-43.

Multilayer Perceptron Approach of Artificial Neural Networks in Fruit Mass Estimation

Sultan GÜNGÖR[1]
Bünyamin DEMİR[2]

1- Graduate Student; Erciyes University, Graduate School of Natural and Applied Sciences. gungor.sultan@hotmail.com ORCID No: 0009-0000-7360-8533
2- Prof. Dr.; Mersin University, Faculty of Engineering Department of Mechanical Engineering. bd@mersin.edu.tr ORCID No: 0000-0002-6405-4724

ABSTRACT

Fruits should have meet specific quality attributes defined by both industrial standards and consumer expectations. Fruit quality is often determined by physical attributes—such as thickness, length, and width—which are directly related to fruit mass. The fast and efficient estimation of some physical characteristics of agricultural products such as mass and volume, improves and automates postharvest processes such as sorting, grading, and storage. Artificial Neural Network (ANN) structures are powerful tools which are used to predict the quality attributes of fruits and vegetables. Typical target applications of these ANN structures are classification, quality estimation from physical properties, fruit processing monitoring during storage. In this study, the fruit masses of Pink Lady apple cultivars were estimated by using Multilayer Perceptron approach of Artificial Neural Networks. The values for major axis length, minor axis length, perimeter, equivalent diameter, and area, measured by a hyperspectral camera, were used as inputs for the ANN structure. The Mean Absolute Percentage Error (MAPE), a key metric of model performance, was found to be 2.707%.

Keywords – Fruit, Fruit Mass, Artificial Neural Network, Estimator.

INTRODUCTION

Apple (Malus domestica L.), a temperate climate fruit, is the most widely cultivated pome fruit worldwide. The global annual production of apples is approximately 76.5 million tons, while Turkey contributes around 2.9 million tons, accounting for 3.8% of the global output. This places Turkey as the third-largest apple producer globally, following China and the United States (FAO, 2012).

In the domain of precision agriculture, identifying fruits based on their physical attributes is of significant importance. Image processing and computer vision are major scientific and technological disciplines that are widely employed for the detection of numerous agricultural products. The use of image processing systems in agriculture has markedly increased in recent years, as as it provides critical insights into production characteristics, reduces costs, support the maintenance of quality standards, and offer real-time actionable information. Image processing is an emerging technology that involves the acquisition, processing, and analysis of images to facilitate the non-destructive and objective evaluation of visual quality parameters in agricultural and food products. Techniques used in image analysis include image acquisition, preprocessing, and interpretation, which enable the quantification and classification of agricultural products based on image data (Mahendran et al., 2012).

There are numerous applications of image processing, including terrestrial and aerial mapping of natural resources, crop monitoring, precision agriculture, robotics, automated navigation, non-destructive inspection of product characteristics, quality control and classification in processing lines, and overall process automation (Cubero et al., 2011).

For the accurate design of classification systems, it is crucial to understand the relationships between fruit mass and other attributes such as length, width, thickness, volume, and projected area (Shahbazi & Rahmati, 2014). Agricultural and food products exhibit several unique features that distinguish them from engineering materials. These physical properties are key determinants of fruit quality, and identifying correlations among them facilitates quality control (Jannatizadeh et al., 2008). Among these, mass and volume are considered the most important characteristics for determining sizing systems (Torabi et al., 2013).

The physical properties of fruits and seeds can be determined using various measurement tools such as image analysis software, digital calipers, and micrometers (Frączek & Wróbel, 2006). However, numerical modeling of product shape based on traditional methods can be particularly labor-intensive for irregularly shaped products (Goni et al., 2007).

Numerous studies have been conducted on modeling fruit mass based on physical attributes. Some researchers have employed artificial neural networks (ANNs) as an effective tool for modeling and predicting certain physical properties of agricultural products. Ochoa-Martínez and Ayala-Aponte (2007) developed an ANN model to predict mass transfer kinetics during the osmotic dehydration of apples. Sabzi et al. (2013) used Adaptive Neuro-Fuzzy Inference System (ANFIS) and linear regression models to estimate fruit mass based on sphericity and projected area. Rad et al. (2015) predicted the final fruit mass of melon using ANN models. Schulze et al. (2015) applied three different models—Simple Linear Regression (SLR), Multiple Linear Regression (MLR), and ANN—to estimate mango fruit mass from geometric dimensions, comparing their predictive power and robustness through calibration and validation. Javadikia et al. (2017) utilized the ANFIS method for mass modeling of three orange cultivars based on fourteen physical properties. Demir et al. (2017), Kus et al. (2017), and Shabani et al. (2017) all reported that ANN models yielded considerably better results than traditional methods in predicting fruit characteristics.

This study was conducted to estimate apple fruit mass using physical properties obtained through hyperspectral imaging technology, which is widely adopted in technologically advanced countries and has recently gained traction in Turkey. A multilayer perceptron artificial neural network model was employed to perform the estimation.

MATERIALS AND METHODS

Materials

Hyperspectral Imaging System

A hyperspectral camera setup (PIKA-L, Resonon Inc., Bozeman, MT, USA) was employed to acquire the hyperspectral images of apple samples. The imaging system comprised a Pika hyperspectral imaging camera, a camera mounting stand, an illumination setup, and a software control system (Figure 1).

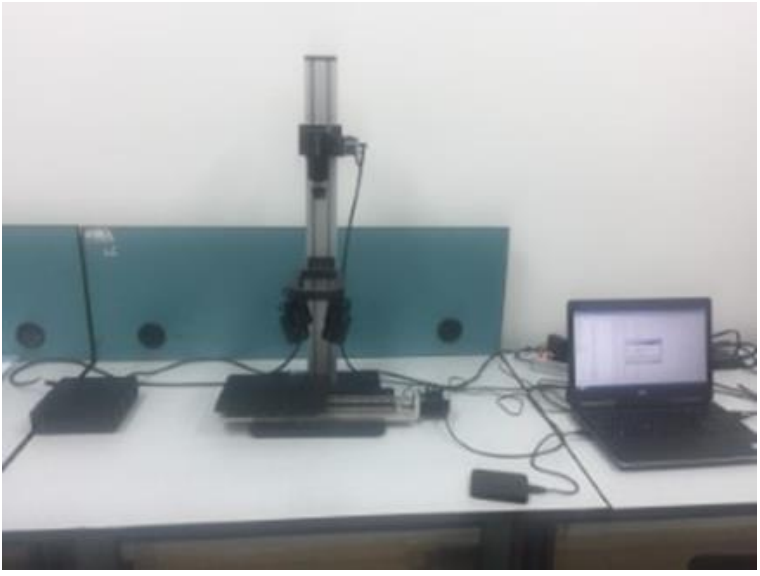


Figure 1: Hyperspectral imaging system

The camera captured reflectance data across 300 spectral bands within the 386–1028 nm range, with a spectral resolution of 2.1 nm. The objective lens used had a 17 mm focal length (maximum aperture F1.4), optimized for the near-infrared region. Hyperspectral images in the near-infrared and adjacent visible spectral ranges were acquired under artificial lighting conditions, providing a spatial resolution of approximately 50 pixels per mm^2. The camera was positioned approximately 50 cm above the platform to acquire the images (Figure 2).

To retain data in absolute reflectance, both the sensor response and illumination functions were disabled in reflectance mode, thereby eliminating any potential dark noise. Prior to image acquisition, the device underwent dark calibration by covering the objective lens. A piece of Teflon (K-Mac Plastics, MI, USA) was used for white calibration.

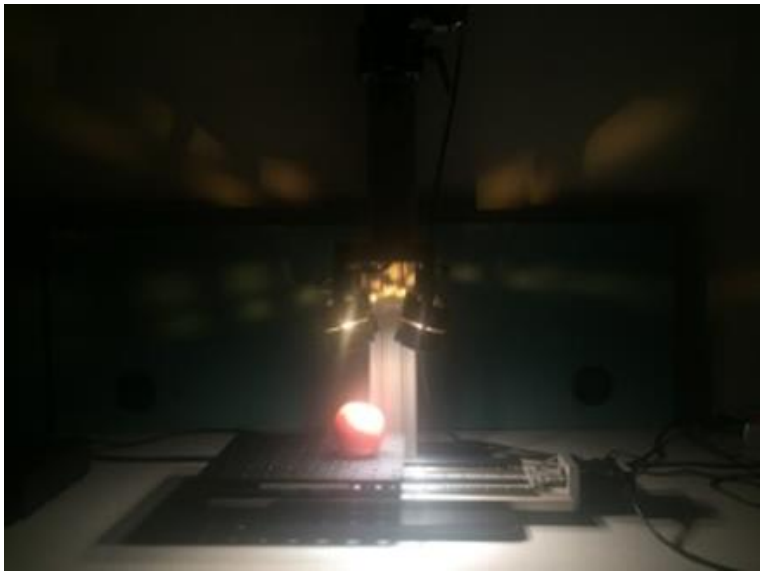


Figure 2: Hyperspectral imaging setup

Pink Lady Apple

The "Pink Lady" apple cultivars were used in this study. The apples were harvested from the orchard of a local enterprise located in Kayseri, Turkey.

Figure 3: Pink Lady apples

Methodology

Image Processing

Image segmentation is a crucial step in image analysis. If segmentation is not performed accurately, background information may be included in the region of interest (ROI), or some data loss may occur. Many

researchers segment the ROI manually, which is time-consuming, and often apply a single threshold value for a specific wavelength. In this study, an automated method was implemented to separate the apple region from the background. However, due to the non-homogeneous coloration of the apples and the presence of shadows in the background of certain images, this method was not successful for every sample.

To address this issue, standard deviations were calculated for all pixels, and thresholding was applied in this data plane. This approach yielded highly effective segmentation results.

The segmented binary image consisted of ones and zeros, where the zero values were ignored during processing. All calculations and image processing tasks were conducted using MATLAB R2016A (The MathWorks, Massachusetts, USA).

A total of 100 apples of the Pink Lady cultivar were used. Initially, their physical appearance was photographed, and features related to their physical dimensions were derived by converting pixel measurements into real-world dimensions. The measured features included Major Axis Length, Minor Axis Length, Perimeter, Equivalent Diameter, and Area.

Multilayer Perceptron (MLP) artificial neural network

Artificial neural network (ANN) methods operate using simple mathematical tools, making them faster than traditional super-resolution techniques. Their adaptive learning capability and ability to perform in real-time make them particularly attractive for various scientific applications. A multilayer perceptron (MLP) is a non-linear neural network consisting of multiple layers: an input layer, one or more hidden layers, and an output layer. The activation function of each neuron in the hidden layers is non-linear, typically sigmoid or tansigmoid. Each neuron is characterized by its activation function and is connected to all neurons in the subsequent layer. Each connection is defined by a weight factor, also referred to as synaptic weight (Mishra et al., 2013).

Numerous parameters must be considered when constructing an artificial neural network model. These parameters may vary depending on the input and output structure of the system. Given that the dataset consisted of only 100 samples—a relatively small size for training an ANN—additional measures were taken to enhance the reliability of the model.

Typically, ANNs comprise three main layers: the input layer, one or more hidden layers, and the output layer. In this study, the input layer included five features: Major Axis Length, Minor Axis Length, Perimeter, Equivalent Diameter, and Area. To eliminate the scale differences among these input variables, each input was min-max normalized to a 0–1 range, ensuring that the highest value equaled 1.

The hidden layers are responsible for establishing the relationship between the inputs and the output. At least one hidden layer is required; using multiple layers can help the model learn more complex patterns within

the training set. However, increasing the number of layers also extends the training time and may reduce the model's generalizability, as the network might memorize the training data rather than learning the underlying patterns. In other words, while the model may perform well on the training data, it may fail to accurately predict unseen data.

In this study, the number of hidden layers was limited to three due to the small dataset size (n = 100). Using fewer hidden layers would risk underfitting, whereas more layers could lead to overfitting. Overfitting would reduce the model's ability to generalize, particularly for the data not included in training.

Another key parameter in the hidden layers is the number of neurons per layer. In this study, this parameter was assigned dynamically to allow the algorithm to determine the most optimal configuration.

The output layer consisted of a single neuron representing the fruit's mass. Considering that ANNs tend to perform better with smaller magnitude values, mass values were converted from grams to kilograms prior to training.

The Multilayer Perceptron (MLP) architecture was selected as the predictive model due to its proven performance in similar studies involving mass prediction. The developed MLP artificial neural network model is illustrated in Figure 4.

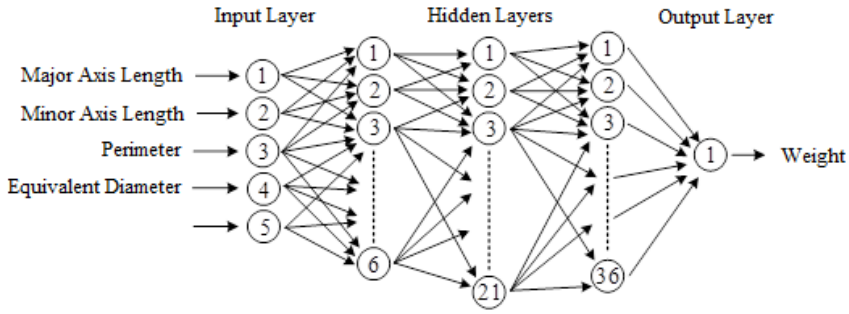


Figure 4: The developed MLP artificial neural network model

Ten out of the 100 samples were reserved for testing (10%), justified by K-fold cross-validation strategy, while the remaining 90 were used for training. To reduce the potential bias caused by selecting only 10 test samples, the K-Fold Cross-Validation method was applied for a more robust performance assessment.

In K-Fold cross-validation, the dataset is divided into K groups. In each iteration, one group is used for testing while the remaining K–1 groups are used for training. In the next iteration, a different group is selected for testing, and the rest are used for training without altering any model parameters. The previously used test group is then included in the training

set as a new group is selected for testing. This ensures that no test sample is used during training.

The key advantage of the K-Fold method is its ability to maximize data utilization in situations with limited sample sizes. It allows the model to be trained on nearly all available data while validating its performance on every subset. As a result, the model does not merely memorize the training set but learns patterns that generalize to unseen data. The only drawback of the method is that it requires K training runs instead of just one, thereby increasing computational cost—though only during training.

In this study, K was set to 10. In each iteration, 90 samples were used for training and 10 for testing. Thus, across 10 iterations, each of the 100 samples was used for testing exactly once (Figure 5).

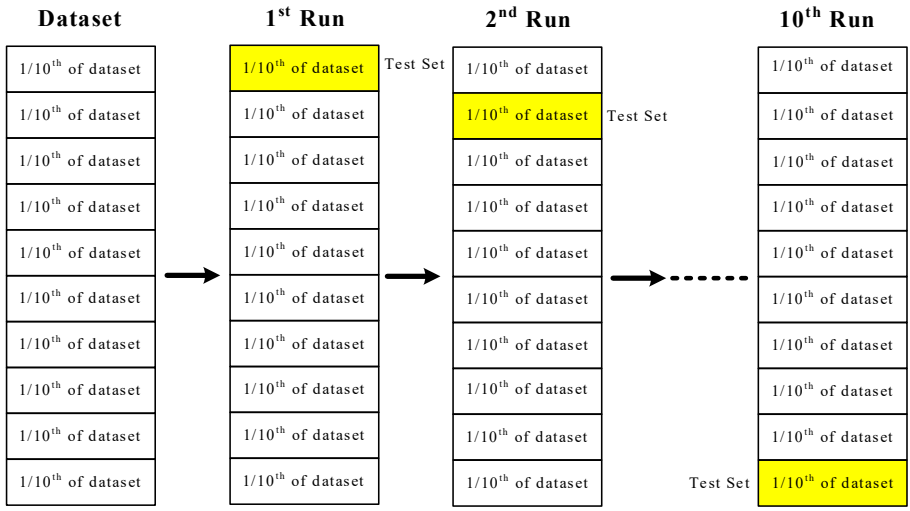


Figure 5: K-Fold validation workflow

Where n represents the total number of samples.

All parameters used in the MLP artificial neural network model are listed in Table 1:

Table 1: Parameters of the MLP artificial neural network model

Parameter Name	Selected Value
Activation Function	identity
Alpha Value	1.00E-06
Neurons in Hidden Layers	6, 21, 36
Learning Rate	Adaptive
Iterations	10,050
Random Seed	1,134,579
Solver	LBFGS

RESULTS AND DISCUSSION

The graph comparing the predicted masses produced by the Multilayer Perceptron (MLP) artificial neural network model, which was developed using the features Major Axis Length, Minor Axis Length, Perimeter, Equivalent Diameter, and Area, with the actual measured mass values is presented in Figure 6.

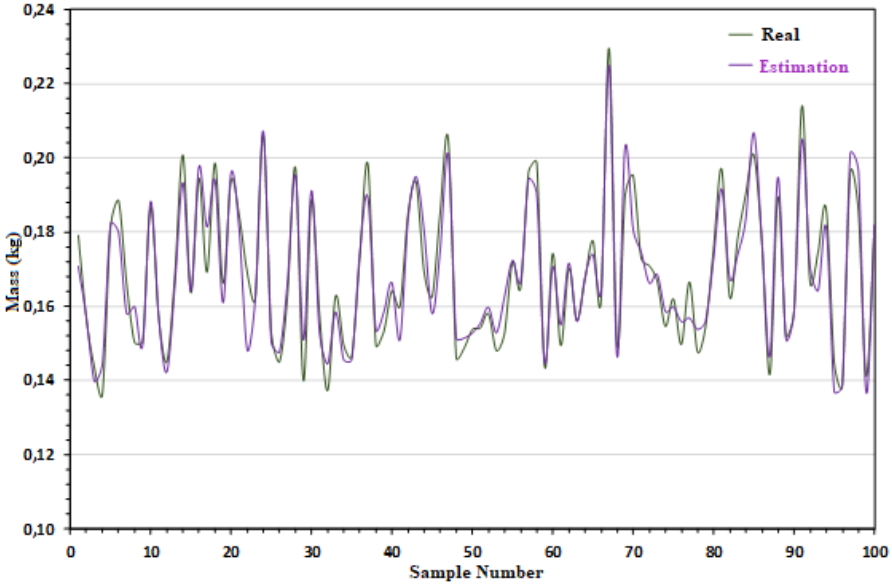


Figure 6: Comparison of actual and predicted fruit masses

Figure 6 illustrates the relationship between the actual and predicted fruit masses. The curve generated by the MLP model closely follows the trend of the actual data, indicating that the artificial neural network successfully modeled the data-generating process through a reliable predictive structure.

To validate the R^2 calculation between actual and predicted values, the analysis was also performed using Microsoft Excel, and the results were found to be consistent, as shown in Figure 7.

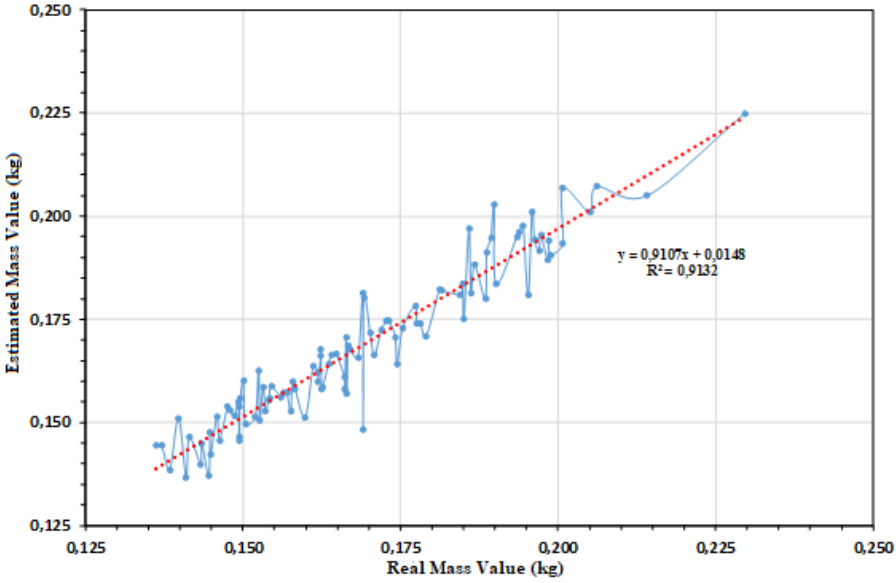


Figure 7: R^2 calculation graph for actual and predicted masses

The Absolute Percentage Error (APE) values obtained from the MLP artificial neural network model are presented in Table 2. According to the results, samples numbered 57 and 87 exhibited the highest absolute percentage error values. The mean coefficient of determination (R^2) across all samples was calculated as 0.9132, and the Mean Absolute Percentage Error (MAPE) was found to be 2.707.

Table 2: APE results of the MLP artificial neural network model

Sample No	Actual Mass (kg)	Estimated Mass (kg)	APE (%)	Sample No	Actual Mass (kg)	Estimated Mass (kg)	APE (%)
1	0.1362	0.1444	6.021	51	0.1664	0.1706	2.524
2	0.1371	0.1445	5.398	52	0.1665	0.1569	5.766
3	0.1384	0.1385	0.072	53	0.1667	0.1686	1.140
4	0.1398	0.1509	7.940	54	0.167	0.1678	0.479
5	0.141	0.1367	3.050	55	0.1683	0.1657	1.545
6	0.1415	0.1465	3.534	56	0.1691	0.1814	7.274
7	0.1433	0.1398	2.442	57	0.1691	0.1483	12.300
8	0.1435	0.1448	0.906	58	0.1692	0.1803	6.560
9	0.1446	0.137	5.256	59	0.1702	0.1717	0.881
10	0.1448	0.1476	1.934	60	0.1709	0.1663	2.692
11	0.1449	0.1423	1.794	61	0.172	0.1724	0.233
12	0.1459	0.1514	3.770	62	0.1728	0.1746	1.042
13	0.1463	0.1456	0.478	63	0.1732	0.1746	0.808
14	0.1476	0.1538	4.201	64	0.1742	0.1707	2.009
15	0.1479	0.1529	3.381	65	0.1745	0.1642	5.903
16	0.1487	0.1515	1.883	66	0.1754	0.1729	1.425
17	0.1493	0.155	3.818	67	0.1774	0.1783	0.507
18	0.1494	0.1465	1.941	68	0.1776	0.174	2.027
19	0.1495	0.1456	2.609	69	0.1782	0.174	2.357
20	0.1495	0.1539	2.943	70	0.1791	0.1708	4.634
21	0.1496	0.1559	4.211	71	0.1812	0.1822	0.552
22	0.1501	0.16	6.596	72	0.1815	0.182	0.275
23	0.1504	0.1496	0.532	73	0.1845	0.181	1.897
24	0.152	0.1513	0.461	74	0.1849	0.1836	0.703
25	0.1525	0.1626	6.623	75	0.1851	0.175	5.457
26	0.1527	0.1504	1.506	76	0.1859	0.1969	5.917
27	0.1532	0.1586	3.525	77	0.1862	0.1813	2.632
28	0.1536	0.1528	0.521	78	0.1868	0.1883	0.803
29	0.1541	0.1555	0.909	79	0.1885	0.18	4.509
30	0.1542	0.1559	1.102	80	0.1887	0.1912	1.325
31	0.1545	0.1587	2.718	81	0.1894	0.1948	2.851
32	0.1561	0.156	0.064	82	0.1899	0.2028	6.793
33	0.1565	0.1571	0.383	83	0.1902	0.1835	3.523
34	0.157	0.1571	0.064	84	0.1935	0.195	0.775
35	0.1576	0.1528	3.046	85	0.1938	0.196	1.135
36	0.1579	0.1598	1.203	86	0.1945	0.1977	1.645
37	0.1583	0.1582	0.063	87	0.1953	0.1808	7.424
38	0.1599	0.1511	5.503	88	0.1959	0.2011	2.654
39	0.1612	0.1637	1.551	89	0.1963	0.1943	1.019
40	0.1619	0.1621	0.124	90	0.197	0.1917	2.690
41	0.1619	0.1599	1.235	91	0.1973	0.1954	0.963
42	0.1624	0.1661	2.278	92	0.1984	0.1895	4.486
43	0.1624	0.1676	3.202	93	0.1986	0.194	2.316
44	0.1625	0.1581	2.708	94	0.1988	0.1906	4.125
45	0.1627	0.1585	2.581	95	0.2008	0.1934	3.685
46	0.1636	0.1641	0.306	96	0.2008	0.2068	2.988
47	0.1641	0.1664	1.402	97	0.2051	0.201	1.999
48	0.1648	0.1666	1.092	98	0.2062	0.2072	0.485
49	0.1661	0.161	3.070	99	0.2141	0.205	4.250
50	0.1662	0.1582	4.813	100	0.2296	0.2249	2.047

Measuring and quantitatively characterizing the mass of apple fruits is often challenging and time-consuming. However, postharvest fruit mass can be estimated indirectly using derived data.

Lorestani et al. (2012) demonstrated that in their study on limes, fruit volume-based modeling yielded a higher coefficient of determination (R^2) compared to mass-based models, suggesting volume is a more reliable predictor when modeling fruit characteristics.

Sabzi et al. (2013) emphasized that image processing has become one of the promising modern techniques in agricultural applications, such as mass and volume estimation, grading, surface damage detection, and the classification of different agricultural products. In their study, they extracted physical properties of various orange cultivars using image processing techniques and used ANFIS and SPSS models to estimate fruit mass. The results indicated that ANFIS was a suitable and effective tool for mass prediction based on physical features.

Soares et al. (2013) proposed a method for predicting banana bunch mass using artificial neural networks and concluded that the ANN model provided accurate estimations. Similarly, Rad et al. (2015) applied ANN models to predict the final fruit mass of melons, achieving high accuracy in mass estimation.

Javadikia et al. (2017) utilized image processing techniques to classify Bam, Khooni, and Thompson orange cultivars and estimated the fruit mass using the ANFIS model. In their approach, 70% of the data were used for training and 30% for testing. The best-performing models yielded determination coefficients (R^2) of 0.948, 0.99, and 0.98 for Bam, Khooni, and Thompson, respectively. The prediction accuracy for these models was within ±3.7 g, ±1.28 g, and ±3.2 g, respectively. These results confirmed the sufficiency and reliability of the ANFIS model for orange mass prediction.

Utai et al. (2019) presented seven different approaches to obtain fruit dimensions and estimate fruit mass using specific algorithms. They used two input parameters (area and thickness) or three parameters (length, width, and thickness), achieving a maximum success rate of 97% and a determination coefficient of 0.99. The study demonstrated the efficacy of algorithmic approaches in fruit mass estimation.

For irregularly shaped fruits, mass prediction based on image processing represents a constructive advancement in postharvest technologies. The success of image-based mass estimation depends largely on the accuracy of feature extraction during image analysis, the quality of input data fed into ANN models, the neural network structure, and the chosen learning algorithm. The robustness and sensitivity of ANN models are typically evaluated by examining and comparing the predicted outcomes with actual measured values.

CONCLUSION

In this study, hyperspectral images of Pink Lady apples were analyzed to extract several physical properties, which were then used as input features for estimating fruit mass. The results demonstrate that these features are sufficient for accurately predicting the mass of an individual apple. The findings suggest that the Multilayer Perceptron (MLP) artificial neural network is a valuable tool for simulating and optimizing individual fruit mass, showing strong potential for applications in automated postharvest handling and precision agriculture.

ACKNOWLEDGMENTS

This study was supported by the Mersin University Scientific Research Projects (BAP) Unit with project number 2018-3-AP4-3044

REFERENCES

Cubero, S., Aleixos, N., Moltó, E., Gómez-Sanchis, J., Blasco, J. (2011). Advances in machine vision applications for automatic inspection and quality evaluation of fruits and vegetables. Food and bioprocess technology, 4(4), 487-504.

Demir, B., Eski, I.,Ikbal, Kus, Z.A., Ercisli, S., (2017). Prediction of physical parameters of pumpkin seeds using neural network. Notulae Botanicae Horti Agrobotanici Cluj-Napoca, 45(1), 22-27.

Fao. (2012). Production crops. Online FAO statistics. http://www.fao. org/statistics. Accessed 20 Mar 2018.

Frączek, J., Wróbel, M. (2006). Methodic aspects of seed shape assessment. Inżynieria Rolnicza, 12(87), 155-163.

Goñi, S. M., Purlis, E., Salvadori, V. O. (2007). Three-dimensional reconstruction of irregular foodstuffs. Journal of Food Engineering, 82(4), 536-547.

Jannatizadeh, A., Naderi Boldaji, M., Fatahi, R., Ghasemi Varnamkhasti, M., & Tabatabaeefar, A. (2008). Some postharvest physical properties of Iranian apricot (Prunus armeniaca L.) fruit. International Agrophysics, 22(2), 125-131.

Javadikia, H., Sabzi, S., Rabbani, H. (2017). Machine vision based expert system to estimate orange mass of three varieties. International Journal of Agricultural and Biological Engineering, 10(2), 132-139.

Kus, Z. A., Demir, B., Eski, I., Gurbuz, F., Ercisli, S., (2017). Estimation of the colour properties of apples varieties using neural network. Erwerbs-Obstbau, 59(4), 291-299.

Lorestani, A. N., Jaliliantabar, F., Gholami, R. (2012). Mass and volume modelling of Persian lime (Citrus aurantifulia) with geometrical attributes. Journal of Agricultural Technology, 8(5), 1537-1543.

Mahendran, R., Jayashree, G. C., Alagusundaram, K. (2012). Application of computer vision technique on sorting and grading of fruits and vegetables. J Food Process Technol, 10, 2157-7110.

Mishra, S., Yadav, R. N., & Singh, R. P. (2013). A survey on applications of multi layer perceptron neural networks in DOA estimation for smart antennas. International Journal of Computer Applications, 83(17).

Ochoa-Martínez, C. I., Ayala-Aponte, A. A. (2007). Prediction of mass transfer kinetics during osmotic dehydration of apples using neural networks. LWT-Food Science and Technology, 40(4), 638-645.

Rad, M.R.N., Fanaei, H.R., Rad, M.R.P. (2015). Application of Artificial Neural Networks to predict the final fruit weight and random forest to select important variables in native population of melon (Cucumis melo L.). Scientia Horticulturae, 181, 108-112.

Sabzi, S., Javadikia, P., Rabani, H., Adelkhani, A. (2013). Mass modeling of Bam orange with ANFIS and SPSS methods for using in machine vision. Measurement, 46 (9), 3333-3341.

Schulze, K., Nagle, M., Spreer, W., Mahayothee, B., Müller, J. (2015). Development and assessment of different modeling approaches for size-mass estimation of mango fruits (Mangifera indica L., cv.'Nam Dokmai'). Computers and Electronics in Agriculture, 114, 269-276.

Shabani, A., Ghaffary, K. A., Sepaskhah, A. R., & Kamgar-Haghighi, A. A. (2017). Using the artificial neural network to estimate leaf area. Scientia Horticulturae, 216, 103-110.

Shahbazi, F., Rahmati, S. (2014). Mass modeling of persimmon fruit with some physical characteristics. Agricultural Engineering International: CIGR Journal, 16(1), 289-293.

Soares, J. D. R., Pasqual, M., Lacerda, W. S., Silva, S. O., Donato, S. L. R. (2013). Utilization of artificial neural networks in the prediction of the bunches' weight in banana plants. Scientia Horticulturae, 155, 24-29.

Torabi, A., Tabatabaekoloor, R., Hashemi, S. J. (2013). Volume modelling of three apple varieties based on physical parameters. Internat. J. Agricult. Food Sci. Technol, 4(5), 461-466.

Utai, K., Nagle, M., Hämmerle, S., Spreer, W., Mahayothee, B., Müller, J. (2019). Mass estimation of mango fruits (Mangifera indica L., cv.'Nam Dokmai') by linking image processing and artificial neural network. Engineering in Agriculture, Environment and Food, 12(1), 103-110.

Analyzing Ship Agency Selection Criteria in Maritime Transport Using Fuzzy AHP

Ercan YÜKSEKYILDIZ[1]

1- Asst. Prof. Dr; Samsun Üniversitesi Siyasal Bilgiler Fakültesi Denizcilik İşletmeleri Yönetimi Bölümü. ercan.yuksekyildizr@samsun.edu.tr ORCID No: 0000-0001-7199-8267

ABSTRACT

The criteria for choosing a ship agency in maritime logistics are identified, and their relative importance is evaluated in this study. Eight essential criteria—cost, service quality, dependability and reputation, regulatory compliance, communication, local knowledge and networks, financial strength, and technological infrastructure—were identified after a thorough analysis of the existing literature.Ten experts with at least eight years of experience in the maritime industry provided evaluations for the fuzzy analytical hierarchy process, which was used to address decision-making under uncertainty.The results show that the most important factors are reputation and dependability (20%), followed by service quality (17.8%) and regulatory compliance (13.9%). Cost was the fourth most important factor, with 12.6%. These results show that only looking at cost when choosing a ship agency is not enough; to get long-term results and a competitive edge, you need to put reliability, service quality, and following the rules first. The study offers valuable insights that can assist decision-makers in formulating effective strategic selection processes.

Keywords – Ship Agency, Fuzzy AHP, Agency Selection, Marine Transport, MCDM

INTRODUCTION

The maritime industry is a big part of trade around the world (UNCTAD, 2023). Approximately 80% of all international shipping is conducted by ships, making agents in this field very important. (Sevgili and Nas, 2017; Kokkinis and Pappis, 2006). These agents do a lot of different things, like making sure that ships run smoothly, that cargo is safe, and that the law is followed. Choosing the right agent will directly affect how competitive the business will be in the long run, how happy customers will be, how cost-effective it will be, and how well it will run (Thai, 2016; Panayides and Cullinane, 2002). But this selection process needs to look at a number of quantitative and qualitative factors at the same time (Seo and Bhattacharya, 2018).

Because of problems with global supply chains, it has been harder to make judgments in the maritime business during the past few years. The COVID-19 pandemic, geopolitical tensions, and fluctuations in energy prices have exacerbated these issues (Notteboom and Haralambides, 2020). In this scenario, tactics that solely look at cost haven't worked. Other things, such dependability, service quality, technology infrastructure, environmental sustainability, and following the rules, need to be taken into account as well (Yuen and Thai, 2015). Because of this, decision-makers are now utilizing

more advanced ways to deal with ambiguity and the necessity to look at more than one thing at a time.

Multi-Criteria Decision-Making (MCDM) techniques provide a systematic and analytical framework for addressing complex selection problems (Sener and Ozturk, 2015). The fuzzy analytical hierarchy process (FAHP) is one of these strategies that is very beneficial in areas where expert judgments are subjective and there is a lot of uncertainty. FAHP encompasses hierarchical Multi-Criteria Decision-Making (MCDM) techniques, which provide a systematic and analytical approach to resolving complex selection dilemmas (Sener and Ozturk, 2015). The fuzzy analytical hierarchy process (FAHP) is highly useful in situations when there is a lot of uncertainty and different experts have different perspectives. FAHP combines the hierarchical structure and pairwise comparison benefits of traditional AHP with the ability of fuzzy logic to show uncertainty (Kahraman and Ulukan, 2003). This integration enables the utilization of triangular fuzzy numbers to illustrate decision-makers' evaluations, hence enhancing the accuracy and reliability of criterion weight calculations (Chang, 1996).

Kahraman and Ulukan (2003) showed that fuzzy logic makes weight determination more reliable by turning language-based evaluations into a mathematical model. In the maritime sector, the flexibility of this approach in evaluating qualitative criteria—such as reliability, service quality, and crisis management—has made it a preferred evaluation method.

Over the past thirty years, the use of MCDM approaches to choose suppliers and agents has grown quickly. Ho and Dey (2010) examined 78 research published from 2000 to 2008, determining that AHP, TOPSIS, VIKOR, and fuzzy logic-based methodologies are the most prevalent techniques in supplier selection. While most of these studies have focused on criteria such as cost, quality, delivery time, and flexibility, maritime-specific factors—particularly reliability, regulatory compliance, and local port knowledge—have received comparatively limited attention in the literature.

Research in marine logistics shows that when choosing an agent, dependability and service quality are often more important than price. Thai (2016) conducted a study at the Port of Singapore and found that prompt service delivery was the most important factor affecting customer satisfaction. Yuen and Thai (2015) also said that how well ships can stay running and how well they can handle problems is closely related to the quality of port services. Panayides and Cullinane (2002) emphasized the significance of reliability and crisis management, whilst Seo and Bhattacharya (2018) highlighted the competitive advantage of adhering to international standards. Mehta and Mehta (1997) also said that how good the service is is very important for making customers happy.

This study employs the FAHP method to identify the criteria influencing agent selection in the maritime sector and determine their relative importance.

MATERIALS AND METHODS

To determine the weights of the criteria, the Fuzzy Analytic Hierarchy Process (FAHP) method was employed, and evaluations were obtained from ten experts with direct experience in the maritime sector. The participants were selected from ship agencies, port authorities, ship-owning companies, logistics firms, academia, and related support services, each with at least eight years of industry experience or academic expertise in the relevant field (Table 1). A fair evaluation of the criteria required a balanced combination of management, operations, academia, law, technology, and field experience. The FAHP analysis is more trustworthy because it comprises people who have varied jobs and points of view.

Table 1. Expert profiles and areas of responsibility

Expert Code	Position Title	Representation Area	Professional Experience	Expertise Description
U1	Ship Agency Operations Manager	Ship Agency	15 years	Port operations, agency service processes
U2	Port Authority Commercial Manager	Port Authority	22 years	Commercial relations, agency–port coordination
U3	Shipowner Company Operations Manager	Ship owning	11 years	Shipowner perspective, service quality assessment
U4	Maritime Transportation Specialist, Logistics Firm	Logistics	14 years	Supply chain, integration of maritime transportation
U5	Ship Agency Legal Counsel	Law	20 years	Regulatory compliance, contract management
U6	Academic (Faculty of Maritime Studies)	Academia	25 years	Theoretical modelling, decision-making methods
U7	Information Systems Specialist	Technology	8 years	Digital systems, electronic document management
U8	Agency Operations Officer	Ship Agency	8 years	Daily field operations, port procedures
U9	Agency Sales and Marketing Specialist	Ship Agency	10 years	Customer relations, service delivery
U10	Agency Customs and Documentation Officer	Ship Agency	14 years	Customs procedures, document flow, and regulatory follow-up

Fuzzy Analytic Hierarchy Process

Numerous Fuzzy Analytic Hierarchy Process (FAHP) approaches have been developed in the literature that integrate fuzzy set theory with hierarchical structure analysis. These methods provide a systematic framework for solving multi-criteria decision-making (MCDM) problems (Ertuğrul, 2007). The Extended FAHP method proposed by Chang (1996) was employed in this study. The main advantage of this method is that it is straightforward to calculate because it follows the usual AHP steps without adding any new ones. One of its main problems, though, is that it only works with triangular fuzzy numbers (Göksu and Güngör, 2008).

Let $X= \{x_1, x_2, ..., x_n\}$ be the set of objects and $U= \{u_1, u_2, ..., u_m\}$ be the set of goals, where each object is considered for achieving a particular goal. Thus, for each object, *m* extended analysis values are obtained, expressed as:

$$M_{g_i}^1, M_{g_i}^2, \dots, M_{g_i}^m, \; i= 1, 2, ..., n \qquad (1)$$

All $M_{g_i}^j$ ($j = 1, 2, ..., m$) values are triangular fuzzy numbers. The application steps of Chang's (1996) extended analysis method are as follows:

For the i-th object, the fuzzy synthetic extent value (*Si*) is calculated as:

$$S_i = \sum_{j=1}^{m} M_{g_i}^j \otimes \left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{g_i}^j\right]^{-1} \qquad v \qquad (2)$$

The degree of possibility that a fuzzy number $M_2= (l_2, m_2, u_2)$ is greater than another fuzzy number $M_1= (l_1, m_1, u_1)$is defined as:

$$V(M_2 \geq M_1) = \; sup_{y \geq x}\left[min\, \mu_{M_1}(x), \mu_{M_2}(x)\right] \qquad (3)$$

In comparison of degrees of probability,

$$V(M_2 \geq M_1) = hgt(M_1 \cap M_2) = \mu_{M_2}(d) \qquad (4)$$

$$\mu_{M_2}(d) = \begin{cases} 1 & if \;\; m_2 \geq m_1 \\ 0 & if \;\; l_1 \geq u_2 \\ \dfrac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)} & otherwise \end{cases} \qquad (5)$$

To compare M_1 and M_2, we need both the values of $V(M_1 \geq M_2)$ and $V(M_2 \geq M_1)$.

The degree possibility for a convex fuzzy number to be greater than k convex fuzzy numbers M_i ($i=1, 2, ..., k$) can be defined by:

$$V(M \geq M_1, M_2, \ldots, M_k)$$
$$= V[(M \geq M_1)\ and\ (M \geq M_2)\ and \ldots and\ (M \geq M_k)]$$
$$= \min V(M \geq M_i)\ and \quad i$$
$$= 1,2, \ldots, k \qquad (6)$$

For each S_i, the minimum degree of possibility over all $(S_i \geq S_k)$ is computed as:

$$d'(A_i) = \min V(S_i \geq S_k) \qquad (7)$$

For $k = 1, 2, \ldots, n\ and\ k \neq i$, then the weight vector is given by:

$$W' = \left(d'(A_1), d'(A_2), \ldots, d'^{(A_n)}\right)^T \qquad (8)$$

Via normalization, the normalized weight vectors are:

$$W = \left(d(A_1), d(A_2), \ldots, d(A_n)\right)^T \qquad (9)$$

This normalized vector (W) represents the relative weights of the criteria, where W is a nonfuzzy number.

RESULTS AND DISCUSSION

In the literature, the key criteria for ship agency selection are generally grouped under the categories of cost, service quality, reliability and reputation, regulatory compliance, communication, local knowledge and network, financial strength, and technological infrastructure (Mehta and Mehta, 1997; Lu, 2000; Kannan and Tan, 2002; Panayides and Cullinane, 2002; Kokkinis and Pappis, 2006; Özer and Çetin, 2012; Bayırhan and Nas, 2014; Sener and Ozturk, 2015; Sevgili and Nas, 2017; Seo and Bhattacharya, 2018; Özkan, 2020; Yorulmaz and Çelik, 2024)

To validate the eight main criteria sets identified from the literature and to incorporate any necessary additions, preliminary interviews were conducted with the participants. As a result, 22 sub-criteria associated with these main criteria were determined. The pairwise comparisons of the identified criteria and sub-criteria were evaluated using the fuzzy scale developed by Chang (1996) (Table 2).

Table 2. Scale used in FAHP according to Chang (1996)

Linguistic Importance	Fuzzy Scale	Reciprocal Scale
Equally important	(1, 1, 1)	(1/1, 1/1, 1/1)
Slightly more important	(1, 3, 5)	(1/5, 1/3, 1)
Strongly more important	(3, 5, 7)	(1/7, 1/5, 1/3)
Very strongly more important	(5, 7, 9)	(1/9, 1/7, 1/5)
Extremely more important	(7, 9, 9)	(1/9, 1/9, 1/7)

All pairwise comparison matrices obtained were aggregated using the geometric mean method, resulting in a consolidated fuzzy pairwise comparison matrix. The values obtained for the main criteria are presented in Table 3.

Table 3. Consolidated fuzzy pairwise comparison matrix

	K1	K2	K3	K4	K5	K6	K7	K8
K1	(1.0, 1.0, 1.0)	(0.28, 0.37, 0.52)	(0.21, 0.27, 0.38)	(0.50, 0.66, 1.00)	(1.55, 2.05, 2.85)	(2.25, 3.05, 4.05)	(2.85, 3.85, 5.05)	(2.05, 2.85, 4.05)
K2	(1.92, 2.70, 3.55)	(1.0, 1.0, 1.0)	(0.34, 0.50, 0.98)	(1.85, 2.55, 3.55)	(2.55, 3.55, 4.55)	(3.25, 4.25, 5.55)	(4.05, 5.05, 6.05)	(3.05, 4.05, 5.05)
K3	(2.65, 3.65, 4.85)	(1.02, 2.00, 2.95)	(1.0, 1.0, 1.0)	(2.55, 3.55, 4.55)	(3.55, 4.55, 5.55)	(4.55, 5.55, 6.55)	(5.05, 6.05, 7.05)	(4.05, 5.05, 6.05)
K4	(1.00, 1.50, 2.00)	(0.28, 0.40, 0.66)	(0.22, 0.28, 0.50)	(1.0, 1.0, 1.0)	(2.05, 3.05, 4.05)	(3.05, 4.05, 5.05)	(4.05, 5.05, 6.05)	(3.05, 4.05, 5.05)
K5	(0.35, 0.49, 0.65)	(0.22, 0.28, 0.50)	(0.18, 0.25, 0.33)	(0.25, 0.33, 0.50)	(1.0, 1.0, 1.0)	(2.05, 3.05, 4.05)	(3.05, 4.05, 5.05)	(2.05, 3.05, 4.05)
K6	(0.25, 0.33, 0.45)	(0.18, 0.25, 0.33)	(0.15, 0.20, 0.25)	(0.20, 0.25, 0.33)	(0.25, 0.33, 0.50)	(1.0, 1.0, 1.0)	(2.05, 3.05, 4.05)	(1.55, 2.05, 3.05)
K7	(0.20, 0.25, 0.33)	(0.16, 0.20, 0.25)	(0.14, 0.16, 0.20)	(0.16, 0.20, 0.25)	(0.20, 0.25, 0.33)	(0.25, 0.33, 0.50)	(1.0, 1.0, 1.0)	(1.55, 2.05, 3.05)
K8	(0.25, 0.33, 0.50)	(0.20, 0.25, 0.33)	(0.16, 0.20, 0.25)	(0.20, 0.25, 0.33)	(0.25, 0.33, 0.50)	(0.33, 0.50, 1.00)	(0.33, 0.50, 1.00)	(1.0, 1.0, 1.0)

Following the application steps of the FAHP method, calculated the fuzzy synthetic extent values (S_i), defuzzied weights (W'), and normalized final criterion weights (W) from the consolidated fuzzy pairwise comparison matrix (Table 4).

Table 4. Fuzzy synthetic extent values (S_i), defuzzied weights (W'), and normalized final criterion weights (W)

Criterion	(S_i) (l, m, u)	Defuzzied Weight (W')	Normalized Weight (W)
K1 Cost	(0.085, 0.125, 0.185)	0.1317	0.126
K2 Service Quality	(0.125, 0.180, 0.250)	0.1850	0.178
K3 Reliability & Reputation	(0.145, 0.205, 0.275)	0.2083	0.200
K4 Regulatory Compliance	(0.095, 0.140, 0.200)	0.1450	0.139
K5 Communication	(0.070, 0.100, 0.150)	0.1067	0.102
K6 Local Knowledge & Network	(0.060, 0.085, 0.125)	0.0900	0.086
K7 Financial Strength	(0.050, 0.075, 0.110)	0.0783	0.075
K8 Technological Infrastructure	(0.065, 0.090, 0.135)	0.0967	0.094

The same method was applied for the sub-criteria to compare pairwise and calculate local weights. The global weights were then obtained by multiplying the local weights of the sub-criteria by the weights of their corresponding main criteria (Table 5).

According to the FAHP results, the criterion Reliability and Reputation (K3) is the most important, with a weight of 20%. Service Quality (K2) comes in second with 17.8%, and Regulatory Compliance (K4) comes in third with 13.9%. This shows how important it is for the sector to fully follow all national and international rules in order to be reliable and sustainable. Cost (K1) ranks fourth with 12.6%.

Communication (K5) makes up 10.2% of the rest of the criterion, Technological Infrastructure (K8) makes up 9.4%, Local Knowledge and Network (K6) makes up 8.6%, and Financial Strength (K7) makes up 7.5%.

The findings of this study are broadly consistent with those reported in the literature. The identification of Reliability and Reputation as the most important criteria aligns with the emphasis on operational reliability, timely service, and crisis management found in studies such as Sevgili & Nas (2017), Thai (2016), Panayides & Cullinane (2002), and Yuen & Thai (2015). The preference for Service Quality over cost aligns with supplier selection literature, including the studies by Mehta and Mehta (1997), Kannan & Tan (2002), and Lu (2000).

Seo and Bhattacharya (2018) and Sevgili & Nas (2017) both say that full compliance with rules and standards is important for gaining a competitive edge and staying in business. This is in line with the idea that regulatory compliance should come in third place.. The Cost criterion ranking fourth reflects the general trend in local and international studies that price alone is not the sole determinant in selection decisions.

Table 5. Main and sub-criteria weights

Main Criterion	Sub-Criterion	Main Criterion Weight	Local Weight	Global Weight
K1 Cost	K1.1 Service fee	0.126	0.50	0.0630
	K1.2 Additional expenses		0.30	0.0378
	K1.3 Price transparency		0.20	0.0252
K2 Service Quality	K2.1 Operational speed	0.178	0.40	0.0712
	K2.2 Operational accuracy		0.35	0.0623
	K2.3 24/7 service capacity		0.25	0.0445
K3 Reliability & Reputation	K3.1 Timely service	0.200	0.45	0.0900
	K3.2 Crisis management		0.35	0.0700
	K3.3 Past performance		0.20	0.0400
K4 Regulatory Compliance	K4.1 Compliance with national regulations	0.139	0.40	0.0556
	K4.2 Compliance with international standards		0.35	0.04865
	K4.3 Full adherence to procedures		0.25	0.03475
K5 Communication	K5.1 Information flow	0.102	0.45	0.0459
	K5.2 Multilingual communication		0.30	0.0306
	K5.3 Proactive customer relations		0.25	0.0255
K6 Local Knowledge & Network	K6.1 Port procedure knowledge	0.086	0.50	0.0430
	K6.2 Local supplier relations		0.30	0.0258
	K6.3 Regional infrastructure knowledge		0.20	0.0172
K7 Financial Strength	K7.1 Payment guarantee	0.075	0.60	0.0450
	K7.2 Insurance capacity		0.40	0.0300
K8 Technological Infrastructure	K8.1 Digital tracking systems	0.094	0.55	0.0517
	K8.2 Electronic document management		0.45	0.0423

However, the relatively low weight assigned to Local Knowledge and Network in this study diverges from research in contexts such as irregular shipping routes and project cargo, as reported by Bayırhan and Nas (2014) and Özkan (2020). Similarly, the mid-to-lower ranking of Technological Infrastructure and Financial Strength contrasts with some recent studies that have assigned these criteria a higher priority.

CONCLUSION

This study identified the criteria considered by decision-makers in ship agency selection through a literature review and weighted them using the Fuzzy Analytic Hierarchy Process (FAHP) method. The results revealed that the Reliability and Reputation criterion (20%) holds the highest importance. This finding indicates that timely service delivery, crisis management, and past performance are even more critical than maritime sector costs.

Service quality (17.8%) and following the rules (13.9%) were the next most important things after reliability. To stay ahead of the competition, you need to be quick, precise, available all the time, and follow all national and international rules. The Cost criterion came in fourth with 12.6%. This shows that price is important, but it's not the only thing that matters when making a choice.

Other criteria, Communication (10.2%), Technological Infrastructure (9.4%), Local Knowledge and Network (8.6%), and Financial Strength (7.5%) were evaluated as supportive yet complementary factors. These results demonstrate that decision-makers value cost, quality, and supporting factors such as technology adoption, local connections, and financial assurance. In particular, technological infrastructure and local knowledge can provide strategic advantages in today's increasingly competitive global maritime industry.

Timely Service (K3.1) has the highest global weight at 9%, followed by Operational Speed (K2.1) at 7.12%, and Crisis Management (K3.2) at 7%. These results show that when choosing a ship agency, reliability and speed are more important than cost.

When making decisions, people should prioritize things like reliability, service quality, and following the rules over things that are only based on cost. Ship agencies should work on making their operations faster and more accurate, as well as their ability to handle emergencies, so they can better serve their customers. Furthermore, investments in compliance with international standards and digitalization will enhance their competitiveness. Policymakers

should push for the digital integration of port and customs procedures and make it easier for agencies to get to international certification processes.

The present study concentrates on general criteria for ship agency selection; subsequent research may integrate various transportation modes and agency profiles to enhance the literature.

A comparative analysis of the criteria prioritized by companies engaged in liner shipping versus tramp shipping may clarify sectoral distinctions.

Moreover, examining the reasons for the preference for separate agencies in different service sectors and assessing the criteria used by companies that transport different types of cargo (such as bulk cargo, containers, and project cargo) could significantly improve academic research and industry practices. Such extended studies would reveal how the importance of criteria changes depending on the context, enabling decision-makers to develop more targeted strategies. Moreover, the importance weights of the identified criteria for agency selection can be examined using other MCDM methods, such as TOPSIS and VIKOR, to conduct comparative analyses.

REFERENCES

Bayırhan, İ., & Nas, S. (2014). Düzensiz hatlarda hizmet veren gemi işletmelerinin gemi acentası seçim kriterleri: Merkezleri İzmir'de bulunan kuru ve dökme yük gemi işletmelerinde bir araştırma. Beykoz Akademi Dergisi, 2(2), 1–19.

Chan, F. T. S., & Kumar, N. (2007). Global supplier development considering risk factors using a fuzzy extended AHP-based approach. Omega, 35(4), 417–431.

Chang, D. Y. (1996). Applications of the extent analysis method on fuzzy AHP. European Journal of Operational Research, 95(3), 649–655. https://doi.org/10.1016/0377-2217(95)00300-2

Ertuğrul, İ. (2007). Fuzzy Analytic Hierarchy Process and its application to machine selection problem in a textile firm. Hacettepe University Journal of Economics and Administrative Sciences, 25(1), 171–192.

Göksu, A., & Güngör, İ. (2008). Fuzzy Analytic Hierarchy Process and its application of university preference ranking. Süleyman Demirel University Journal of Economics and Administrative Sciences, 13(3), 1–26.

Ho, W., & Dey, P. K. (2010). Multi-criteria decision-making approaches for supplier evaluation and selection: A literature review. European Journal of Operational Research, 202(1), 16–24.

Kahraman, C., & Ulukan, Z. (2003). Multi-criteria supplier selection using fuzzy AHP. Logistics Information Management, 16(6), 382–394.

Kannan, V. R., & Tan, K. C. (2002). Supplier selection and assessment: Their impact on business performance. Journal of Supply Chain Management, 38(3), 11–21.

Kaptanoğlu, D., & Özok, A. F. (2006). Akademik performans değerlendirmesi için bir bulanık model. İTÜ Dergisi/d Mühendislik, 5(1), 193–204.

Kokkinis, G., & Pappis, C. (2006). Freight forwarding in Greece: Services provided and choice criteria. EuroMed Journal of Business, 1(2), 64–81.

Lu, C. S. (2000). Logistics services in Taiwanese maritime firms. Transportation Research Part E: Logistics and Transportation Review, 36(2), 79–96.

Mehta, S. C., & Mehta, S. S. (1997). Important dimensions in ocean freight services: Shippers' perspective. Journal of Customer Service in Marketing & Management, 3(3), 85–100.

Notteboom, T. E., & Haralambides, H. E. (2020). Port management and governance in a post-COVID-19 era: Quo vadis? Maritime Economics & Logistics, 22(3), 329–352.

Özkan, M. (2020). Deniz tedarik zinciri yönetiminde gemi acentesinin ve deniz ulaştırma işletmesinin karşılıklı seçimi (Doktora tezi). Maltepe Üniversitesi, İstanbul, Türkiye.

Özer, T., & Çetin, İ. (2012). A study on the charter type choice of Turkish general cargo and dry bulk shipowners. The Asian Journal of Shipping and Logistics, 28(2), 203–226.

Panayides, P. M., & Cullinane, K. P. B. (2002). The vertical disintegration of ship management: Choice criteria for third-party selection and evaluation. Maritime Policy & Management, 29(1), 45–64.

Sevgili, C., & Nas, S. (2017). Taşıma işleri komisyoncularının gemi acentelerini tercih ölçütleri: İzmir Limanı uygulaması. International Journal of Management Economics and Business, 13(1), 155–165.

Sener, Z., & Ozturk, E. (2015). A QFD-based decision model for ship selection in maritime transportation. International Journal of Innovation, Management and Technology, 6(3), 202–205.

Seo, Y., & Bhattacharya, S. (2018). The ship management firm selection: The case of South Korea. Asian Journal of Shipping and Logistics, 34, 258–268. https://doi.org/10.1016/j.ajsl.2018.09.008

Thai, V. V. (2016). The impact of port service quality on customer satisfaction: The case of Singapore. International Journal of Maritime Economics, 18(4), 458–475.

United Nations Conference on Trade and Development. (2023). Review of maritime transport 2023. UNCTAD. https://unctad.org/system/files/official-document/rmt2023_en.pdf

Yuen, K. F., & Thai, V. V. (2015). Service quality and customer satisfaction in liner shipping. International Journal of Quality and Service Sciences, 7(2–3), 170–183.

Yorulmaz, M., & Çelik, Y. (2024). Gemi acenteliği ve seçim kriterleri. European Journal of Managerial Research (EUJMR), 8(15), 87–110.

Optimization of the Tribological Performance of Different Coating Types on the Piston-Ring Pair Using the Taguchi Method

Idris CESUR[1]

1- Prof. Dr. Idris CESUR; Sakarya University of Applied Sciences , Faculty of Technology, Department of Mechanical Engineering, icesur@subu.edu.tr ORCID No: 0000-0001-7487-5676.

ABSTRACT

One of the key factors that reduce energy efficiency in internal combustion engines is the friction and subsequent wear that occur between moving parts. In this context, the friction losses occurring on the contact surfaces between the ring and the cylinder play a decisive role in engine performance. This study experimentally evaluates the effects of two different coatings, based on chromium carbide and boron nitride, on the tribological performance of rings. In the experimental design, friction and wear behaviors were examined under different load and speed parameters, and the obtained results were analyzed using the Taguchi optimization method. The experimental findings revealed that at low speed conditions, higher wear rates were observed in both the ring and cylinder liner samples, and that with increasing load, there were significant increases in both friction force and wear rates. When comparing the coating types, it was found that the boron nitride-coated rings exhibited lower friction coefficients and wear values compared to the chromium carbide-coated samples. This study demonstrates that the appropriate selection of ring coating materials and operational parameters is crucial in reducing tribological losses in engine systems.

Keywords – Chromium Carbide, Boron Nitride, Piston-Ring Pair, Wear, Friction, Taguchi.

INTRODUCTION

In parallel with the rapidly increasing world population, a significant rise in the number of motor vehicles has been observed in recent years. This increase in vehicle numbers has considerably boosted fossil fuel consumption, thereby accelerating the depletion of petroleum reserves. These developments have made it imperative to use energy resources more efficiently and sustainably. Within this framework, research aimed at improving energy efficiency and enhancing system performance—particularly in the automotive sector—has gained growing importance. In the automotive industry, energy losses occurring in vehicles continue to be a fundamental issue in terms of both fuel economy and environmental sustainability (Cesur et al., 2024).

Internal combustion engines are energy conversion systems that continue to be widely used in vehicle technologies today. However, the thermal efficiency of these engines is largely limited by mechanical losses. In particular, the friction and wear occurring between the moving components inside the engine not only increase fuel consumption but also negatively affect the lifespan of engine parts (Holmberg and Erdemir, 2017). According to previous studies, up to 40% of the total mechanical losses within the engine originate from friction between the ring and the cylinder liner (Tian, 2002). Therefore, improving the tribological properties of ring surfaces is of critical importance for enhancing engine performance and reducing power losses.

Rings, working in conjunction with the piston, provide sealing in the combustion chamber and contribute to heat transfer. However, the continuous contact between these components under high temperatures, pressures, and shear stresses leads to significant wear (Taylor, 1998). To address this issue, recent years have seen a focus on various surface coating techniques and materials. In addition to traditional chromium and nickel coatings, advanced ceramic-based coatings are also being studied, particularly due to their high-temperature resistance, low friction coefficients, and superior wear resistance (Glaeser, 1992).

In this context, chromium carbide-based coatings are preferred in many tribological applications due to their high hardness and temperature resistance. Chromium carbide coatings, particularly when applied using thermal spray methods, can provide a stable surface structure even at high temperatures (Pawlowski, 2008). On the other hand, boron nitride-based coatings also exhibit remarkable performance, especially under dry sliding conditions, due to their low friction coefficient, good thermal conductivity, and chemical inertness (Kimura et al., 1999). The hexagonal form of boron nitride, which exhibits a graphite-like structure, is referred to as "white graphite" and is used as an additive to reduce friction in tribological systems due to its lubricating effect (Erdemir, 2005).

The literature highlights that the friction and wear between the piston ring and cylinder liner in internal combustion engines constitute a significant portion of the engine's total mechanical losses (Holmberg & Erdemir, 2017). In this context, coatings applied using surface engineering techniques play a crucial role in improving tribological performance. Ryk et al. reported a 25% reduction in friction in piston rings coated using the laser surface texturing (LST) method (Ryk et al., 2005). Tung and colleagues applied CrN and DLC coatings to stainless nitrided steel and chromium-plated steel piston rings and compared their tribological performance under different fuel types (Tung et al., 2002). Sheng and his team studied friction and wear properties in diesel engines by applying two different coatings on 1045 steel under solid lubrication, showing that ion-plated molybdenum achieved a lower friction coefficient and less wear (Sheng et al., 2023). The application of chromium carbide and NiCr matrix coatings using the HVOF method was found to increase wear resistance even at high temperatures (Li et al., 2019). Several studies have confirmed that hard coatings such as DLC, TiN, and CrN reduce friction losses in piston rings, thereby improving engine efficiency (Fan et al., 2022; Harvinder et al., 2024). It has been particularly noted that DLC coatings perform better under boundary lubrication conditions (Gustafsson et al., 2021). Boron nitride-based coatings, due to their lamellar structure, provide solid lubrication effects and reduce friction. Ni-W matrix boron nitride composite coatings have been reported to significantly enhance wear resistance both at high temperatures and under limited lubrication conditions (Aouadi et al. 2009). Furthermore, experiments on piston rings have shown

that CrN, TiAlN, and DLC coatings exhibit lower wear in linear oscillation tests compared to conventional rings (Szymanski et al., 2015). Studies evaluating the effects of multiple parameters using statistical optimization techniques like the Taguchi method have successfully analyzed the impact of optimum coating-load-temperature combinations on tribological performance (Yilmaz et al., 2020). It has been reported that chromium carbide and boron nitride coatings, when used separately or in hybrid form, result in a reduction in friction force and a significant increase in wear resistance. Pawlowski (2008) mentioned that chromium carbide–NiCr coatings offer high oxidation and wear resistance up to temperatures of 800 °C (Pawlowski, 2008). Similarly, a study by Zhang et al. (2012) demonstrated that boron nitride-coated surfaces exhibited a low friction coefficient of 0.15 under dry friction conditions and minimized surface damage (Zhang, 2012).

There are limited studies that directly compare chromium carbide and boron nitride-based coatings in piston ring-cylinder pairs or optimize the performance of these materials under varying load and speed parameters. This gap represents a significant research void, particularly considering the critical impact of coating material selection on performance in engine applications. In this context, it is of great importance to experimentally evaluate the tribological behavior of coating types and optimize them using multi-parameter statistical analysis methods.

In this study, the tribological performance of chromium carbide and boron nitride-based coatings applied to piston rings in internal combustion engines was examined under different load and speed conditions. Experimental data were analyzed using the Taguchi experimental design method, and the optimal coating type and process parameters were determined. In this context, the study is expected to make significant contributions to material selection and surface engineering in engine systems.

MATERIASL AND METHOD

Experimental Method

In this study, two different coating materials, chromium carbide and boron nitride, were applied to ring samples to evaluate the tribological interactions between the ring and cylinder liner in internal combustion engines. The friction and wear characteristics of the coatings were experimentally examined under different load and speed conditions using a wear testing device. The experimental design was optimized using the Taguchi method.

In the study, a wear testing device was used to determine the amount of wear and the friction coefficient. This device operates with a linear motion that simulates the contact conditions between the piston ring and the cylinder in engines (Figure 1).

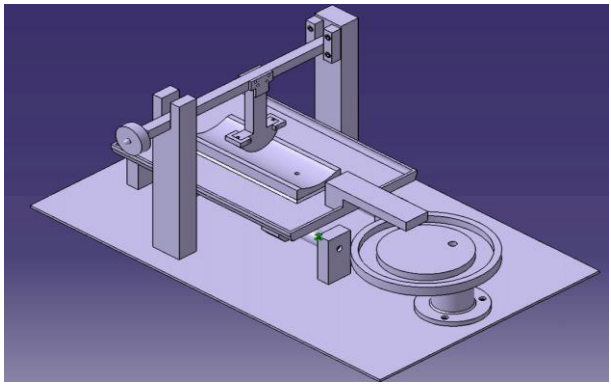


Figure 1: Wear Tester

The wear tests were conducted using a wear testing device where the ring and cylinder liner samples are in mutual contact. In the device, the ring and liner samples were scaled down to fit the test conditions. The rings used in the experiments were made of gray cast iron, a material commonly used in the market. The cylinder liner was made of low-carbon steel with a honed inner surface. The ring samples were divided into three groups: Uncoated (STD) piston rings, chromium carbide-coated piston rings, and boron nitride-coated piston rings.

In the experimental studies, the wear amount was determined using the mass loss method. For this purpose, an analytical balance with a measurement precision of 0.0001 g was used. Prior to the experiments, the cylinder liner and piston ring samples were cleaned with ethanol to remove any possible contaminants on the surface, and then dried to ensure complete evaporation of the ethanol. After drying, the initial mass measurements were taken, and these values were recorded as the reference. After each test, the samples were cleaned again with ethanol, dried, and their final mass was measured using the same method. The wear amounts were calculated by taking the difference between the mass values measured before and after the experiments.

To determine the friction coefficient, it is essential to accurately adjust the normal loads applied to the piston ring sample. These loads were applied by hanging weights of specific masses on the carrier arm, to which the ring sample was fixed, in a manner suitable for the experimental setup. This ensured controlled and repeatable loading conditions throughout the experiment.

The coating processes were carried out using the High-Velocity Oxygen Fuel (HVOF) spraying method. Chromium carbide and boron nitride coatings

were applied to the piston ring surface with an average thickness of 50 µm. The surface morphology and thickness of the coatings were characterized using a Scanning Electron Microscope (SEM). Hardness measurements were performed using a microhardness tester.

To evaluate the tribological behavior of the piston ring-cylinder pairs, a wear testing device based on the pin-on-line principle was used. During the tests, each experiment was conducted for a specific duration, and the friction force and temperature data generated during this time were continuously recorded. The wear amount was determined based on a 100-meter travel distance, with the mass of the samples measured before and after the experiment using a precision balance; the differences between these measurements provided the wear amount. In the experimental study, a three-level parameter arrangement was made, with applied loads set at 40 N, 60 N, and 80 N. The wear rotation speed was adjusted to 50 rpm, 75 rpm, and 100 rpm, and the effect of each parameter on the tribological performance was systematically examined.

In addition, to observe the effects of wear on the surface morphology, the surfaces of the piston ring and cylinder samples were examined using a Scanning Electron Microscope (SEM) after the experiments. The SEM images allowed for the microscopic evaluation of tribological interactions, such as coating integrity, microstructural deformations, surface cracks, and wear traces. These analyses provided a more detailed understanding of the contribution of the coating material to surface performance, and the experimental findings were supported by microstructural observations.

Taguchi Method

The data obtained from the friction coefficient and wear amount in the experimental studies were statistically evaluated to analyze the significance and reliability of the results. For this purpose, One-Way Analysis of Variance (ANOVA) was applied to determine the effects of different load, speed, and temperature parameters on tribological performance. Through the ANOVA analysis, the contribution of each control factor (load and rotational speed) to the system was identified, and it was determined which parameters had a statistically significant effect. These analyses allowed the results to be interpreted not only based on observations but also on quantitative data.

To determine the optimal experimental conditions, Signal-to-Noise (S/N) ratios were calculated for each experimental combination under the Taguchi experimental design framework. These ratios were used to identify the most suitable parameter levels for improving tribological performance. Based on the "smaller-the-better" criterion, S/N values were analyzed for both the friction coefficient and wear amount.

Thus, the parameter combinations that resulted in the lowest friction and wear values were statistically determined.

In the Taguchi method, the "smaller-the-better" criterion is used for performance metrics that are desired to be minimized, such as friction coefficient and wear. The S/N (Signal-to-Noise) ratio is calculated using the following formula:

$$S/N = -10 \cdot \log\left(\frac{1}{n} \cdot \sum_{i=1}^{n} Y_i^2\right)$$

Here:

- n: Number of Experimental Replications
- yi: The response value measured for each experiment (wear or friction coefficient)
- $\sum yi^2$: The sum of the squares of the measurement values

The L-9 standard orthogonal array of Taguchi was selected for the three different parameters and three different levels for each parameter (Table 1).

Table 1: Parameter Levels

	Level 1	Level 2	Level 3
A-Coating type	STD	Chromium carbide	Boron nitride
B-Speed, rpm	50	75	100
C-Load, N	40	60	80

EXPERIMENTAL RESULTS AND DISCUSSION

This study evaluated the tribological performance of the piston ring-cylinder pair, which is an important energy source in internal combustion engines, through the application of chromium carbide and boron nitride-based coatings. The experimental results, conducted under different load and speed parameters, are presented using the Taguchi method.

In Figure 2, the friction coefficient results obtained under different coating types, rotational speeds, and load values are presented. On the horizontal axis, there are three main parameter groups: coating type (STD, chromium carbide coating, and boron nitride coating), rotational speed (50, 75, 100 rpm), and load (40, 60, 80 N). On the vertical axis, the Signal-to-Noise (S/N) ratio of the friction coefficient is shown. This ratio is used to assess the stability of friction behavior in the experimental system. A lower S/N ratio corresponds to a lower friction coefficient and, therefore, better tribological performance.

When the graph is evaluated in general, it can be seen that the coating type has a significant effect on the friction coefficient. The highest friction coefficient was observed on the uncoated (STD) surface, while the chromium carbide coating reduced this value, and the boron nitride coating provided the

lowest friction coefficient. The reasons for this result include the increase in surface hardness and smoothness provided by the boron nitride coating, its reduction of oxidation, and the formation of a hard layer at the microstructural level. These properties minimize surface contact during friction, leading to lower energy losses and better wear resistance.

When the effect of rotational speed is examined, it is observed that the friction coefficient is relatively high at 50 rpm, while it decreases significantly at 75 and 100 rpm. The increase in rotational speed reduces the contact time between the surfaces, leading to a decrease in the friction coefficient. However, at very high speeds, the temperature increase may limit this reduction.

The effect of load value is more limited. When the load was increased from 40 N to 60 N, a slight increase in the friction coefficient was observed, and at 80 N, the value increased slightly. This indicates that although the increased load causes surface deformation, the coating layer is able to resist this deformation due to its high hardness and strength properties. In particular, the boron nitride coating maintained its structural integrity even under high loads, preventing an increase in the friction coefficient.

These results are largely consistent with studies in the literature. Kandeva et al. (2019) reported that boron coating reduced the friction coefficient by 30–40% under dry sliding conditions. Similarly, Babu et al. (2020) emphasized that chromium coating reduced friction by decreasing surface roughness but noted that its thermal resistance was lower compared to boron coating. Lee et al. (2018) found that increasing rotational speed strengthened the formation of the hydrodynamic lubrication film and resulted in a significant reduction in the friction coefficient. Additionally, in a study by Sahin et al. (2021), it was stated that the friction coefficient of boron-coated surfaces remained stable in the 40–80 N load range and that the coating layer was resistant to deformation.

In conclusion, the findings of this study demonstrate that boron nitride coating is the most effective method for reducing the friction coefficient, and it significantly improves tribological performance, especially under high-speed and moderate-load conditions. While uncoated surfaces exhibit high friction coefficients, leading to increased wear and energy losses, the boron nitride coating largely eliminates these negative effects.

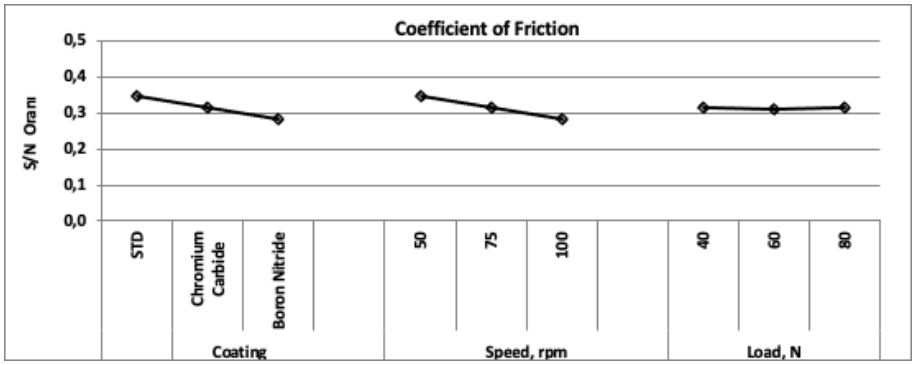


Figure 2: Variation of the S/N ratio in friction coefficient according to coating type, different rotational speeds, and load conditions.

In Figure 3, the statistical analysis of the cylinder wear experimental results is presented. The S/N ratio used here, as in the Taguchi method, is a measure of the stability and variability of the experimental results. A low S/N ratio for cylinder wear indicates a lower wear amount and, consequently, better surface durability.

When the graph is examined in general, it is observed that the coating type has the most significant effect on cylinder wear. The highest wear occurs on uncoated (STD) surfaces, while the wear rate decreases significantly with chromium carbide coating, and the lowest wear values are obtained with boron nitride coating. This result indicates that boron nitride coating significantly increases surface hardness and forms an effective barrier against oxidation and micro-pitting wear. Additionally, the boron nitride layer minimizes wear by providing resistance to adhesion and abrasive effects that may occur on the surface during friction. On the other hand, chromium carbide coating reduces surface roughness, resulting in some improvement, but it does not provide as high a level of protection as boron nitride coating.

When the figure is examined, it is observed that both the increase in rotational speed leads to a decrease in wear, and the increase in load results in an increase in cylinder wear. The increase in rotational speed reduces the contact time between the surfaces, leading to a reduction in wear. On the other hand, the change in wear with load can be explained by the fact that at lower loads, surface contact does not fully occur, causing microscopic wear to happen in a more irregular manner. As the load increases, the contact area and surface pressure also increase, which in turn enhances material deformation.

The results obtained in this study show a similar trend to many studies in the literature. Kandeva et al. (2019) reported that boron nitride coating significantly reduced surface wear due to its high hardness and chemical stability. Babu et al. (2020) stated that chromium coating increased wear resistance by reducing surface roughness, but exhibited lower performance under high temperature conditions compared to boron nitride coatings. Lee et

al. (2018) mentioned that increasing rotational speed weakened the lubrication film, leading to increased wear, while at lower speeds, the contact remained more stable. In a study by Sahin et al. (2021), it was reported that boron nitride-coated steel surfaces exhibited high resistance to deformation under different load conditions, and the increase in wear was limited.

In conclusion, the data obtained from this graph clearly demonstrate that boron nitride coating is the most effective method for reducing cylinder wear. While high wear rates are observed on uncoated surfaces, it is evident that boron nitride coating increases surface durability and provides resistance to temperature and load effects. Although an increase in rotational speed negatively impacts wear behavior, the surface protection provided by boron nitride coating limits this increase. Therefore, boron nitride coating can be considered the most suitable coating alternative for ensuring long service life, low wear rate, and high tribological stability on cylinder surfaces.

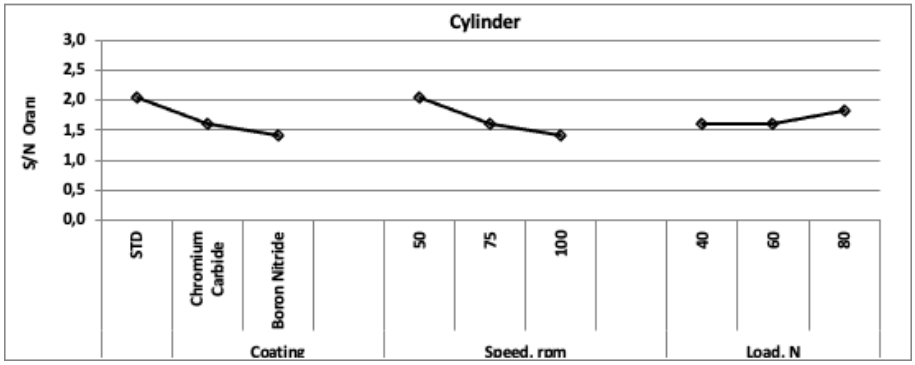


Figure 3: Variation of the S/N ratio in cylinder wear amount according to coating type, rotational speed, and load conditions.

In Figure 4, the experimental results of rings wear are presented in terms of the Signal-to-Noise (S/N) ratios. A lower S/N ratio indicates lower wear and better surface durability. When evaluated in terms of coating types, the highest wear occurs on uncoated (STD) piston rings, while chromium carbide coating significantly reduced wear, and boron nitride coating achieved the lowest wear values. The increased surface hardness, oxidation resistance, and resistance to micro-wear caused by boron nitride coating explain this result. It is observed that the S/N ratio decreases with an increase in rotational speed (50–100 rpm), indicating that wear increases at higher speeds due to temperature rise. The increase in load (40–80 N) had a limited effect on wear rate, and it was found that the boron nitride coating maintained surface integrity under high load conditions. In conclusion, boron nitride coating on rings has significantly improved tribological performance by providing the lowest wear.

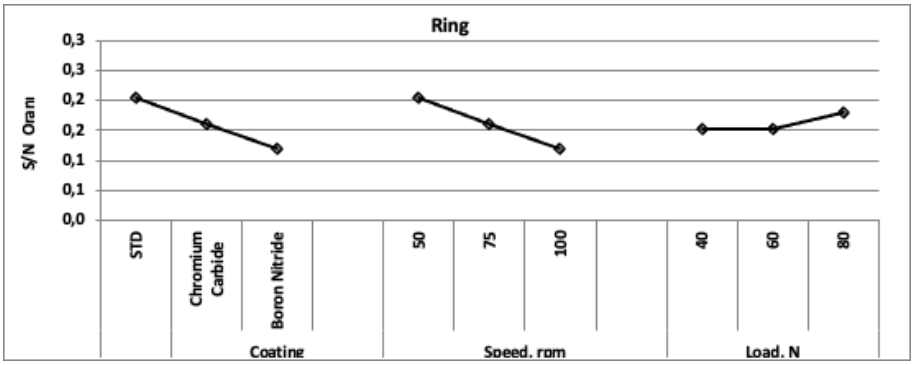


Figure 4: Variation of the S/N ratio in piston rings wear amount according to coating type, rotational speed, and load conditions.

In Figure 5, the SEM image of the uncoated rings sample obtained from the wear test is presented. This piston rings sample was subjected to operating conditions of 60 N and 75 rpm.

On the uncoated piston rings surface, distinct pits, deep scratches, and material build-ups are observed. This indicates that direct metal-to-metal contact has occurred, and both abrasive and adhesive wear mechanisms have been acting simultaneously. Additionally, the areas of the surface appearing melted or flowed suggest that micro-welding has occurred due to high temperature effects, leading to a deterioration of surface integrity.

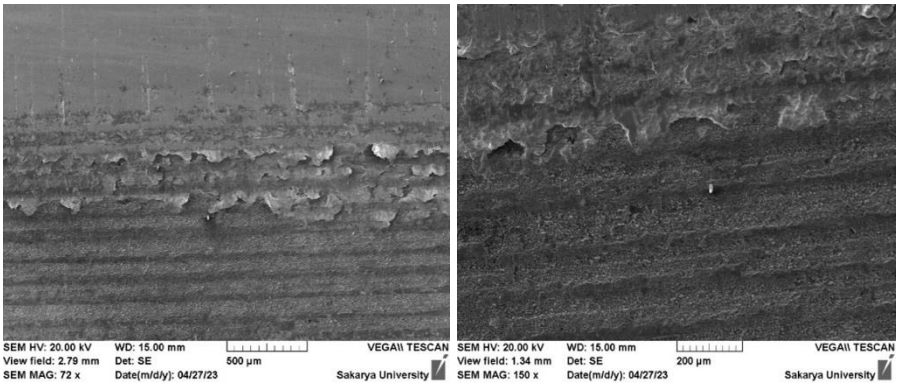


Figure 5: SEM image of the uncoated piston rings sample at 60 N and 75 rpm.

In Figure 6, the SEM image of the wear test result for the chromium carbide coated rings sample is shown. This piston rings sample was obtained under the conditions of 60N and 75 rpm.

On the chromium carbide coated piston rings surface, wear marks are more orderly, and surface deformation is more limited. The high hardness of the coating layer has increased the surface's resistance to abrasive particles

and reduced the wear depth. However, the concentration of scratches in certain areas indicates that micro-abrasive wear is still occurring due to high contact pressure and partial lubricant film insufficiency. Nevertheless, material loss has significantly decreased compared to the uncoated surface.

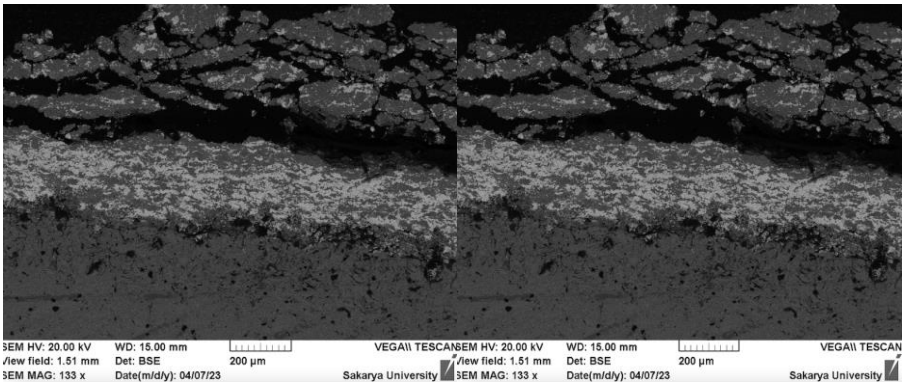


Figure 6: SEM image of the chromium carbide coated piston rings sample at 60 N and 75 rpm.

In Figure 7, the SEM image of the wear test result for the boron nitride coated rings sample is shown. This piston rings sample was obtained under the conditions of 60N and 75 rpm.

On the boron nitride coated piston rings, the surface is observed to be quite smooth and homogeneous, with minimal formation of scratches and pits. This is explained by the ability of boron nitride to reduce surface temperature and prevent metal-to-metal contact due to its low friction coefficient and high thermal conductivity. The self-lubricating property of the boron nitride coating is thought to form a stable tribo-film on the surface, minimizing wear.

As a result, after the wear tests, the highest wear was observed on the uncoated rings, while the lowest wear was observed on the boron nitride coated rings. This finding is consistent with the literature results reported by Kandeva et al. (2019) and Şahin et al. (2021), which indicate that boron-based coatings maintain surface integrity even under high load and temperature conditions. Therefore, boron nitride coating emerges as the most effective coating type in reducing wear and friction in the piston rings-cylinder system.

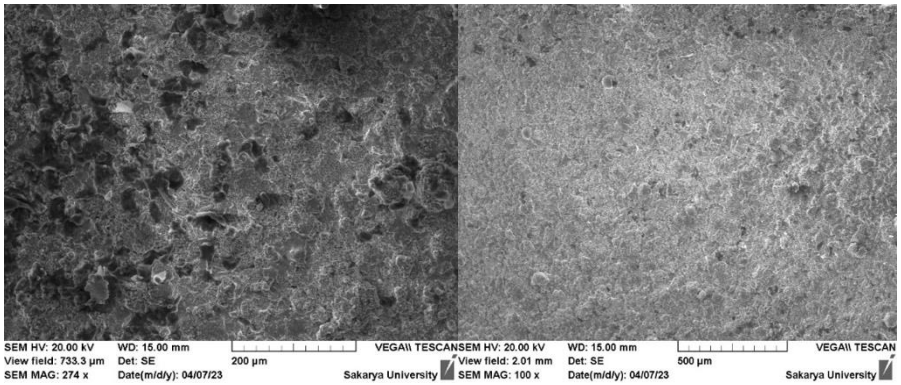


Figure 7: SEM image of the boron nitride coated piston rings sample at 60 N and 75 rpm.

RESULTS

In this study, the friction and wear behaviors of rings-cylinder pairs composed of uncoated (STD), chromium carbide-coated, and boron nitride-coated rings were investigated using the Taguchi method. The friction coefficient and wear results obtained under different load and speed conditions in the experimental studies were evaluated, and wear surfaces were analyzed using scanning electron microscopy (SEM), with the findings supported by microstructural observations.

The results showed that the piston rings coating had a significant effect on the tribological performance. The highest wear occurred in the uncoated rings, while the lowest wear was observed in the boron nitride-coated rings. The wear amount in the chromium carbide-coated rings was at an intermediate level compared to the other two conditions. The high wear in the uncoated rings can be attributed to their lower surface hardness and increased metal-to-metal contact.

The chromium carbide-coated piston rings, although showing moderate wear, exhibited good surface integrity and maintained structural stability at high temperatures. SEM images revealed partial oxidation and light wear traces on the surface, but no severe surface degradation was detected. This demonstrates the advantages of chromium carbide coatings in terms of high-temperature resistance and hardness.

The boron nitride-coated piston rings exhibited the best wear resistance due to their low friction coefficient and high surface durability. SEM analyses showed a smooth and cohesive surface structure, limited plastic deformation,

and minimal material loss in these rings. This indicates that the lubricating properties and high-temperature stability of the boron nitride coating positively influenced the friction and wear behavior.

According to the Taguchi analysis results, the most influential parameter on the friction coefficient and wear behavior was found to be the rings type. Load and speed parameters emerged as secondary factors. This result confirms that surface coating is a key determining factor for tribological performance.

Overall, the boron nitride-coated rings showed the best performance with the lowest wear rate and friction coefficient. Chromium carbide-coated rings, particularly under high-temperature conditions, provided durability advantages, while uncoated rings exhibited the highest wear and friction values. As a result, coating selection in engine applications should be optimized based on operating conditions; boron nitride coatings should be preferred for applications requiring low friction and wear, while chromium carbide coatings should be selected for situations requiring high-temperature resistance.

Furthermore, this study demonstrates that the Taguchi method is an effective tool for determining optimal parameter combinations in multivariable tribological systems and can systematically guide experimental analyses.

REFERENCES

Aouadi, S. M., Luster, B., Kohli, P., Muratore, C., & Voevodin, A. A. (2009). Progress in the development of adaptive nitride-based coatings for high temperature tribological applications. Surface and Coatings Technology, 204(6–7).

Babu, K., Ramesh, A., & Ganesan, S. (2020). Effect of chromium and boron coatings on tribological performance of engine components. Surface & Coatings Technology, 399, 126121.

Cesur, İ., Akgunduz, M., Celik, H. A., Ergen, G., & Cay, Y. (2024). Tribological analysis and optimization of ring-cylinder couple coated with different materials. Arabian Journal for Science and Engineering, 49(2).

Erdemir, A. (2005). Review of engineered tribological interfaces for improved boundary lubrication. Tribology International, 38(3), 249–256.

Fan, X., Wang, J., & Liu, Y. (2022). Comparative evaluation of hard coatings on piston ring friction and wear. Materials, 15(5), 1321.

Glaeser, W. A. (1992). Materials for tribology (Tribology Series 20). Elsevier.

Gustafsson, T., Almqvist, A., & Larsson, R. (2021). Effect of DLC coatings under starved lubrication conditions in IC engines. Lubricants, 9(3), 23.

Harvinder, S., Rajdeep, S., Satish, K., & Aneesh, G. (2024). A comprehensive review of tribological performance of coatings on automotive piston rings. i-Manager's Journal on Mechanical Engineering, 14(4).

Holmberg, K., & Erdemir, A. (2017). Influence of tribology on global energy consumption, costs and emissions. Friction, 5(3), 263–284.

Kandeva, M., Kostadinov, G., & Ivanov, M. (2019). Effect of boron-based coatings on the tribological behavior of steels under dry friction. Tribology International, 138, 384–392

Kimura, Y., Wakabayashi, T., Okada, K., Wada, T., & Nishikawa, H. (1999). Boron nitride as a lubricant additive. Wear, 232(2).

Lee, D. H., Park, J. W., & Kim, S. H. (2018). Effect of sliding speed on friction behavior in lubricated contacts. Wear, 398–399, 64–72.

Li, W., Tang, P., Shang, L., He, D., Wang, L., Zhang, G., & Jin, K. (2019). Tribological behaviors of CrN/Cr_3C_2–NiCr duplex coating at elevated temperatures. Surface and Coatings Technology, 378, 124926.

Pawlowski, L. (2008a). Cr_3C_2–NiCr coatings for high-temperature wear applications. Surface and Coatings Technology, 202(18), 4473–4480.

Pawlowski, L. (2008b). The science and engineering of thermal spray coatings (2nd ed.). John Wiley & Sons.

Ryk, G., Kligerman, Y., Etsion, I., & Halperin, G. (2005). Experimental investigation of laser surface texturing for piston-ring friction reduction. Tribology Transactions, 48(4).

Sahin, M., Demir, M., & Aydın, H. (2021). Tribological properties of boron-coated steels under different load conditions. Journal of Materials Engineering and Performance, 30(6), 4285–4293.

Sheng, C., Liang, X., & Rao, G. (2023). Temperature effects on tribological behavior of cylinder-liner–piston ring pairs under different oils. Lubricants, 11(3), 115.

Szymanski, K., Goral, A., & Czyrska-Filemonowicz, A. (2015). Influence of a PVD coating on wear behavior in oscillatory motion. Lubricants, 3(2), 182–194.

Taylor, C. M. (1998). Engine tribology. Tribology International, 31(10), 487–497.

Tian, T. (2002). Dynamic behaviors of piston rings and their practical impact. Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 216(4), 209–228.

Tung, S. C., & Gao, H. (2002). Tribological investigation of piston ring coatings operating in an alternative fuel and engine oil blend. Tribology Transactions, 45(3), 381–389.

Yilmaz, V., Ozcan, S., & Arslan, F. (2020). Optimization of wear behavior of coated piston rings using Taguchi method. Tribology in Industry, 42(1), 112–121.

Zhang, S., Zhang, X., & Liu, W. (2012). Tribological behavior of BN coatings under dry sliding conditions. Surface and Coatings Technology, 206(8–9), 2273–2279.

Usability of Radiant Curing System for the Production of Aerospace Composites

Mustafa TAŞYÜREK[1]

Soner ŞEN[2]

1- Assoc. Prof. Mustafa Taşyürek; Department of Airframe and Powerplant Maintenance, School of Civil Aviation, Selçuk University, Konya, Turkey mtasyurek@selcuk.edu.tr ORCID No: 0000-0001-9016-8584

2- Assoc. Prof. Soner Şen; Department of Airframe and Powerplant Maintenance, School of Civil Aviation, Selçuk University, Konya, Turkey sensoner@selcuk.edu.tr ORCID No: 0000-0003-3385-5577

ABSTRACT

This study comprehensively examines advanced manufacturing methods used in modern aviation and their effects on material performance. This study comprehensively examines advanced manufacturing methods used in modern aviation and their effects on material performance. Firstly, under the heading of production processes, the traditional and innovative methods used in the production of composite materials were compared; especially the contributions of radiant energy-supported production techniques on process efficiency and energy saving were evaluated. In the aviation application areas section, the functions of glass, carbon, aramid and boron fiber reinforced composites used in critical areas such as aircraft fuselage, wing structures, fairing, interior panel and engine components are described in detail.

Another section of the study discussed the gains of radiant systems in aviation; the advantages of these systems in terms of material curing, surface hardening and energy efficiency were revealed. In addition, by comparing materials produced partially with radiant energy (hybrid) and materials produced completely with radiant rays, the improvements achieved by hybrid systems in material properties while increasing the production speed have been analyzed.

As a result, the study shows that radiant-based production techniques used in aviation offer significant advantages both in terms of structural strength and production costs. It is also explained with examples in which aircraft system and material radiant energy can be used partially, completely or in a hybrid way. These approaches have an important potential for the development of lightweight, durable and energy-efficient aircraft of the future.

Keywords – Aerospace, Curing, Energy, Radiant Curing, Thermal And Mechanical Performance.

INTRODUCTION

The evolution of material technologies in the aviation industry has been shaped around the goals of reducing weight, increasing mechanical strength, improving thermal properties and optimizing production processes (Zhu et al., 2018, Zhang et al., 2025, Chen et al., 2023). In this context, the use of composite materials has revolutionized aircraft fuselages, wing structures, engine parts and interior components over the last forty years. Composite systems produced by reinforcing epoxy, phenolic and polyester-based resins with carbon, glass or aramid fibers offer much higher specific strength and hardness values compared to traditional metallic alloys (Chen et

al., 2023, Zhang et al., 2025, Etri et al., 2022). However, the production of these materials still faces limitations such as high energy requirements, long curing times and complex process control.

Structural materials used in the aviation industry play a crucial role in terms of flight safety, performance and sustainability (Bouman et al., 2025). Therefore, the development of materials and production technologies that can provide both high mechanical strength and low weight has become one of the main goals of aviation. One of the most innovative technologies that can meet these goals today is radiant energy-based production and curing systems (Wang et al., 2025).

Radiant energy-based manufacturing technologies have introduced a new paradigm, especially in the curing, sintering and surface modification stages. Radiant energy is a method of transferring energy to the material directly through electromagnetic waves (Xiping et al., 2021). High-energy radiant sources such as laser, infrared (IR), microwave or electron beam systems represent different forms of this technology and eliminate the limitations of classical thermal systems by providing direct and controlled energy transfer to the material surface or internal structure (Kim et al., 2020, Wang et al., 2024b). Since the radiant curing approach performs heat transfer through direct radiation instead of conduction and convection, it ensures rapid and homogeneous heating only in the targeted area of the material. This shortens the curing times and allows the material properties to be controlled more precisely (Wang et al., 2024a). Thus, energy losses are greatly reduced with the higher heating rate achieved (Wang et al., 2025). This allows the production of a porous-free reinforced composite material (Alınca, 2024).

Today, radiant energy-based methods are increasingly used in the production of carbon fiber reinforced polymer (CFRP) structures, high temperature resistant ceramic matrix composites (CMC) and superalloy-based parts in the aerospace industry. In particular, curing epoxy resins with photon or microwave energy significantly shortens the process time and reduces environmental impacts (Mgbemena et al., 2018).

This study aims to comprehensively examine the effects of radiant energy-based curing systems on the material science and production process of composite materials used in aviation, with the dimensions of energy efficiency, mechanical performance, production reliability, environmental impact and sustainability.

RADIANT ENERGY BASED PRODUCTION PROCESSES

Basic Principles of Radiant Energy

Radiant energy refers to energy carried through electromagnetic waves. This energy can be directed to the material surface in the form of visible light, infrared, microwave, ultraviolet or electron beam. Depending on the optical and dielectric properties of the material, this energy is absorbed on the surface or at a certain depth and converted into heat energy. This mechanism largely prevents the formation of internal stresses and thermal distortions that arise from temperature gradients in classical thermal systems.

Radiant sources used in aviation are generally grouped into four main categories:

- Laser systems: Melting of metal powders or surface modification can be done with point and high energy beams.
- Electron beam systems (EBM): Heat is produced by hitting the target surface with high energy electrons in a vacuum environment.
- Microwave systems: Provide internal-to-external heating of composite resins and ceramics with the dielectric heating principle.
- Infrared (IR) systems: Offer superficial, homogeneous and controlled heating, and are especially widely used in epoxy curing processes. (Guan et al., 2022, Xiping et al., 2021).

These systems offer modular solutions that can be easily integrated into production lines and minimize the energy consumption of curing, sintering or coating processes.

Additive Manufacturing

Selective melting methods performed using laser or electron beam (SLM – Selective Laser Melting, EBM – Electron Beam Melting) are based on the point melting and solidification of certain regions of metal powders. Thanks to these methods, parts with complex geometry and high strength used in aviation can be produced with minimal material loss. In particular, titanium and Inconel based components are produced using this method (Mohd Yusuf et al., 2019).

Table 1: Usage areas of aircraft parts manufactured with laser systems

Material / Alloy	Production Method/Properties	Application Area
Ti-6Al-4V (Titanium alloy)	Lightweight, high strength, corrosion resistant.	Jet engine fan blades, brackets, structural fasteners.
Inconel 718 / 625 (Nickel superalloys)	High temperature resistance and oxidation resistance	Turbine blades, exhaust components
AlSi10Mg (Aluminum alloy)	High specific strength and excellent machinability.	Airframe and interior structure parts, heat exchangers
Stainless Steel 316L	High chemical resistance, mechanically stable	Fuel system components, fittings
Co-Cr alloys	Wear-resistant surfaces	Motor bearings, rotor ends

Ceramic and Composite Production

Radiant energy is widely used in ceramic and composite production with the microwave sintering method. This method can complete the sintering process, which takes hours in conventional furnaces, in minutes. At the same time, since the temperature gradients are low, crack formation and internal stresses are minimized (Singhal and Murtaza, 2018).

Table 2. Manufacturing and use of ceramic aircraft parts with laser systems

Material / Alloy	Production Method/Properties	Application Area
SiC (Silicon carbide)	High temperature resistance, low density	Thermal protection panels, engine interior linings
Al_2O_3 (Alumina)	Electrical insulator, high hardness	Sensor housings, electronic sub-components
C/SiC composites	High thermal shock resistance	Heat shields of spacecraft
CNT or graphene-reinforced epoxy composites	Increased mechanical and electrical performance	Aircraft nose cones, wing leading edges, antenna embedded surfaces

Surface Modification and Coating

Laser surface alloying and hardening techniques are used to improve the surface properties of materials. In this way, surface hardness, corrosion resistance and wear resistance are significantly increased. These techniques

are especially preferred in landing gear parts and engine components (Łach, 2024).

Table 3. Manufacturing and use of ceramic aircraft parts with surface alloying and hardening techniques

Material / Alloy	**Production Method/Properties**	**Application Area**
NiCrAlY coating	Laser surface alloying	Turbine blades, combustion chamber walls
WC-Co (Tungsten carbide – Cobalt)	Laser cladding	Piston surfaces, landing gear components
TiC and TiN thin film coatings	Laser or PVD method	Jet engine shafts, rotor gears
Al–Si–Fe alloy laser coating	Corrosion resistance increase	On fuselage exterior surfaces
Graphene or CNT based nano-coatings	Electrical conductivity increase	EMI/RFI shielding surfaces, radar absorbing panels

Nanomaterial Synthesis

Radiant energy is also used in the synthesis of nanomaterials such as carbon nanotubes (CNT), graphene and metal nanoparticles. Methods such as laser ablation or microwave-assisted pyrolysis are very advantageous in terms of energy efficiency and product purity (Baghel et al.).

Table 4: Production and use of aircraft parts with nanomaterial synthesis

Material / Alloy	**Production Method/Properties**	**Application Area**
Carbon Nanotubes (CNTs)	Laser ablation, microwave pyrolysis	Epoxy composite reinforcement, static load distribution
Graphene nanoparticles	Microwave reduction	Electrically conductive composites, sensor surfaces
Al_2O_3, TiO_2, ZnO nanoparticles	Microwave assisted pyrolysis	Thermal barrier coatings
Fe_3O_4 magnetic nanoparticles	Laser assisted evaporation	Magnetic sensors, active vibration damping systems
Nano-SiC reinforced CFRP	Microwave synthesis + resin mixture	Thermal conductivity and hardness increase

APPLICATIONS IN AVIATION

In the aviation sector, radiant energy–based manufacturing technologies are being increasingly embraced. Such systems are applied across multiple scales, from engine components to airframe panels.

Superalloys and Metallic Parts

Radiant energy-based methods play a critical role in the production of superalloys such as titanium and Inconel in three dimensions using powder-bed selective melting (SLM) or electron beam melting (EBM) technologies. These methods are used to produce parts that require high temperature resistance, such as engine turbine blades, fuel nozzles and fasteners. Geometries that are difficult to produce by traditional casting or machining can be built layer by layer with radiant energy, providing advantages in terms of both weight and material waste (Met3DP, 2025).

Ceramic Matrix Composites

Microwave sintering applied during the production of ceramic matrix composites is 10–20 times faster than classical oven methods. In the case of radiant energy use, crack formation and porosity can be minimized due to the uniform heat distribution throughout the volume. Ceramics reinforced with silicon carbide (SiC) or zirconia (ZrO_2), produced by this method, are used in combustion chambers and exhaust systems of jet engines (Savani et al., 2023).

Polymer Matrix Composites (PMCs)

Radiant curing of epoxy or bismaleimide (BMI)-based thermoset resins is becoming increasingly popular as an alternative to traditional curing. It can reduce curing time from 6-8 hours to just a few minutes, particularly in the production of carbon fiber-reinforced panel materials. Radiant curing also significantly increases the efficiency of CFRP panels used in modern aircraft such as the Airbus A350 and Boeing 787 (Mgbemena et al., 2018).

Surface Modification and Coating

Surface hardening and laser alloying processes are used to improve the corrosion resistance of metal alloys and reduce wear rates. For example, radiant-based coatings on landing gear components or engine shaft surfaces can double their service life. In addition, controlling the surface roughness with laser increases the adhesive bond strength and strengthens the integrity of the composite-metal interfaces (PMMT, 2025).

BENEFITS AND PERFORMANCE GAINS OF RADIANT SYSTEMS IN AEROSPACE APPLICATIONS

Radiant energy-based manufacturing and curing systems reduce part weight in aviation, increasing fuel efficiency and range, while improving mechanical and thermal strength; they also reduce energy consumption and minimize environmental impacts and material waste through digital control of production processes. These advantages directly improve the performance and life of critical components such as turbine blades, composite panels and landing gear (ARRK North America, 2025).

Reduction of Part Weight

Production with radiant beams (especially laser and electron beam) melts and solidifies only the necessary areas of the material. Since the production with radiant systems melts and solidifies only the necessary areas of the material, unnecessary material is not used and allows the application of optimized designs, especially in powder bed sintering or selective melting techniques. Hence, excess material removal, typical in conventional machining, becomes unnecessary. In addition, radiant sintering and curing techniques allow the production of more homogeneous and fine structures in composites. This weight reduction is directly reflected in flight performance; it reduces fuel consumption, increases range and reduces emissions. All of these results mean lighter structures, increased range, fuel efficiency and reduced emissions.

Increased Mechanical and Thermal Strength

Radiant production processes provide a fine-grained and homogeneous distribution in the microstructure thanks to the rapid and controlled heating-cooling cycle. This leads to a finer grain structure, a homogenization of the phase distribution and a reduction in internal stresses. In epoxy systems cured by microwave or IR radiation, the cross-link density increases and the void ratio decreases. This directly increases the mechanical strength and impact toughness. In addition, wear and oxidation resistance is significantly increased by surface hardening (e.g. laser surface alloying). In addition, the thermal stability properties required for high temperature use (e.g. Tg value) can be increased thanks to radiant curing.

In aerospace structures, typical examples include delaying high-temperature deformation in turbine blades, enhancing delamination resistance in composite panels, and improving the wear resistance of landing gear components.

Reduction of Energy Consumption

Radiant curing (microwave sintering, IR curing) generates heat directly within the material. In other words, it is not only the furnace air that is heated, but the material itself. Therefore, energy consumption can be reduced significantly compared to classical autoclave or oven curing. In microwave curing systems, energy is converted into heat by directly

interacting with the molecular dipole moments of the polymer matrix, eliminating conduction losses. Thanks to precise control of the thermal regime, especially in epoxy resins, the risk of overheating or burning is reduced (Alınca, 2024). Within the framework of all these effects, shortening the production process can also lead to a reduction in energy costs.

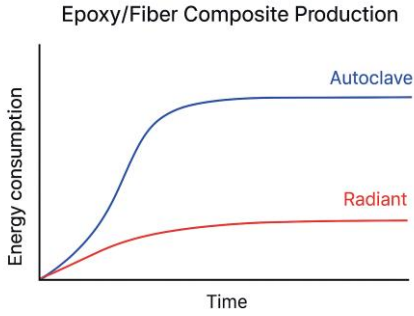


Figure 1: Schematic representation of energy gain with radiant energy compared to traditional production systems in the production of fiber-reinforced aerospace composites.

More Environmentally Friendly Production and Process Control

One of the most important advantages of radiant systems is that production parameters (power, frequency, intensity, focal point) can be controlled digitally. n this way, the production process can be monitored in real time and immediate corrections can be made when necessary. The temperature differences observed in large volume systems such as autoclaves are almost completely eliminated in radiant curing systems. Additionally, this method ensures that the resin flow is homogenized in the closed molds; therefore, the risk of delamination between layers is reduced. This reduces the need for solvent-based resins or high-temperature combustion processes. In addition, since the waste rate is low, a production environment that is fully compliant with environmental and safety standards such as EASA and REACH is created.

MATERIALS PRODUCED PARTIALLY BY RADIANT ENERGY (HYBRID)

Materials produced partially with radiant energy include methods in which radiant energy is not the main source of energy in the production process, but contributes to the processes of curing, sintering or surface improvement of the material. In these methods, radiant energy is usually applied as a microwave, IR (infrared) or laser beam and is often supplemented with additional heating by convection or conduction. Therefore, these methods are completely separated from radiant production techniques, but still provide significant advantages in terms of energy efficiency, processing time and material homogeneity. Hybrid radiative manufacturing is especially useful for polymer and ceramic composites that require high performance in aerospace.

Microwave Assisted Sintering and Curing: In high temperature ceramics such as SiC, Al_2O_3, C/SiC, they are densified by microwave sintering. In this process, microwave radiation heats the polycrystals in the material rapidly and homogeneously. In some cases, the use of preheating or an oven is required, which allows the energy input to be hybrid. This technique can be considered ideal for aircraft engine parts that can be used as heat barriers because a more uniform microstructure, higher thermal resistance, and higher strength can be achieved.

Microwave curing can be used on thermoset epoxy resin-based or thermoplastic resin-based fiber composites. This method can yield highly efficient results in the production of lightweight yet high-strength panels. Post-curing can be carried out using either the traditional method or infrared curing to achieve dimensional stability. It also achieves final surface hardness (Kim et al., 2020).

Hybrid Curing with IR and Laser Support: During curing, infrared (IR) radiation is used to shorten the curing time, working from the surface to the interior structure. It can achieve a uniform surface finish and rapid curing that are not possible with conventional curing. Examples include aircraft interior panel structures and radome coatings (Kumar et al., 2011).

Hybrid Microwave and Vacuum Curing: Composite materials cured using microwave methods are subjected to a period of vacuum after curing. During this phase, the vacuum optimizes the structure by reducing or completely eliminating gas voids, reducing delamination and increasing density. This approach makes it feasible to produce lighter fuselage structures and reinforcements, as well as aerospace components such as high-strength landing gear fairings (Naik et al., 2022).

Hybrid Radiant + Mechanical Sintering: During sintering, microwave or infrared radiation energy can be used to increase strength by creating a denser and more robust structure. Bonding can be achieved using pressure pressing or isostatic pressing, as seen in conventional furnaces. This method

is particularly suitable for some ceramic and metal matrix composites, particularly for mechanically strong engine parts and as thermal barriers. The hybrid system homogenizes the temperature distribution within the material thanks to radiant energy, while increasing the density and mechanical strength with pressure (Savu et al., 2024).
Its advantages and application areas can be categorized as follows:
The material is directly heated with radiant energy. This minimizes energy losses.
Production can be faster compared to traditional methods, reducing time and related financial costs.
The thermal distribution in this method is homogeneous. Thus, the mechanical and thermal properties of each point in the entire structure are constant. All these features make it possible to produce lighter structural elements, thermal barrier coatings, radar-absorbing coating materials, and fiber-reinforced laminates.

OTHER RADIANT OR HYBRID PRODUCTION METHODS AND MATERIALS

In the aerospace industry, the use of polymer and ceramic materials requires high performance as a priority. While radiant energy-assisted production methods have not yet fully penetrated the industry, they are increasing in number. They utilize radiant energy sources such as lasers, infrared (IR), or UV rays to accelerate energy input during production, produce internal structures quickly and without compromising strength, ensure dimensional stability, and achieve a non-porous structure.
Radiant Hot Air Assisted Polymer: CuringThis method utilizes laser or infrared radiation to accelerate the curing of polymer-based matrices. The radiation triggers cross-linking reactions within the polymer chains, thus reducing curing time. Furthermore, homogeneous hardening is achieved. These advantages make them suitable for both visual enhancements and coatings on aircraft interior panels (Malik et al., 2020).
Photopolymerization (UV Irradiation): Photopolymerization, another method used for surface coatings, allows the production of small parts using a polymerization reaction using UV light. Thanks to UV light, surface quality can be achieved quickly and in a controlled manner. This method can also be used to produce protective layers (Li et al., 2023).
Ceramic 3D Printing with Laser: In this method, laser light processes ceramic powders into discrete layers, enabling 3D printing. Furthermore, the

production of lightweight ceramic elements resistant to high temperatures is also possible. Examples include coating thermal barrier elements in the aerospace industry and the manufacture of engine parts with limited geometry (Miao et al., 2025).

MATERIALS PRODUCED COMPLETELY WITH RADIANT RAYS

Melting Metal Powders with Laser (SLM / DMLS): In Selective Laser Melting (SLM) and Direct Metal Laser Melting (DMLS) methods, the main energy source in the production of metal parts is the laser beam in the form of radiant energy. In this process, heat conduction and energy transfer through convection or contact are minimal; melting, solidification and joining of parts are essentially achieved by radiation-based energy. The laser scans the powder layer precisely, causing the metal to melt and solidify locally.

The advantages include high dimensional accuracy and production of complex geometric parts, microstructure homogeneity and density control, minimum material waste and production flexibility (Gao et al., 2022).

Representative Applications (DebRoy et al., 2018):

- Ti-6Al-4V: Jet engine mount brackets, aircraft structural elements
- Inconel 718 / 625: Turbine blades, exhaust components
- Co-Cr, Stainless Steel 316L: Fuel system parts, fasteners
- AlSi10Mg: Heat exchangers and light structural elements

This method falls entirely into the class of radiant energy-based production and plays a critical role in the production of high-performance parts in aircraft.

Electron Beam Melting (EBM): The EBM method melts metal powders using a high-energy electron beam. Since the process is carried out in a vacuum environment, there is no risk of oxidation and the part quality is high. The energy source is entirely an electron beam in the form of radiant energy. EBM makes it possible to produce high-performance metal parts in aviation that are processed entirely with radiant energy.

Its advantages include; production in a vacuum environment, preventing oxidation, the ability to process materials at high temperatures and the ability to produce parts with complex geometries.

Representative Applications (Tammas-Williams et al., 2015):

- Ti-6Al-4V (GE Additive Arcam systems): Aircraft landing gear parts, turbine attachment flanges

- Inconel 718 / TiAl: Engine components operating at high temperature

Nanomaterial Production by Laser Ablation: In the laser ablation method, a high-energy laser beam is focused on a solid surface to vaporize (ablate) it at the atomic/ionic level. The resulting vapor condenses into nanoparticles or nanotubes. This method provides control of the material size on the nanometer scale.

The advantages can be listed as high purity and homogeneous nanoparticle production, complex and precise material reinforcement applications and direct and controlled energy input.

Representative Applications:

- Carbon Nanotubes (CNT): Static load distribution and electromagnetic shielding in aircraft fuselages
- Graphene nanoparticles: Radar absorbing coatings, composite reinforcements
- Metal nanoparticles (Au, Ag, TiO_2, Al_2O_3): Thermal barrier coatings, sensor components, fuel systems

The heating, evaporation and ionization that occur during this process are provided entirely by radiant energy, therefore it is of critical importance in the production of high-performance and advanced technology parts in aviation (Amendola and Meneghetti, 2009).

RESULTS AND DISCUSSION

The study mentioned the effects of partial, full and hybrid radiant energy-based curing systems on process efficiency and energy consumption. The most important advantage of radiant energy, especially in epoxy, carbon, glass and aramid reinforced composites, is the homogeneous heat distribution that occurs during curing. Internal stresses and thermal distortions observed in classical thermal systems are largely prevented in radiant systems. In addition, the observed microstructure uniformity and high density increase durability, especially in critical areas such as the airframe, wing and engine component, and positively affect the long-term fatigue life.

According to the study, it can be concluded that; in partial radiant systems, the production rate is advantageous, but the homogeneity is limited; for critical parts requiring high precision, the full system is preferred. Full radiant systems have the highest level of material homogeneity and energy efficiency, but the investment cost and equipment complexity are higher.

Hybrid systems combine the advantages of partial and full systems; they provide both fast production and superior mechanical performance in critical parts.

In conclusion, the use of radiant base manufacturing systems provides the optimum balance in the production of aerospace components, offering significant advantages over traditional methods in terms of material properties, energy savings and production speed. With these advantages, wing and fuselage structures, engine components and bodywork as well as interior panels can be brought to the aviation industry thanks to this energy.

REFERENCES

ALıNCA, E. 2024. *Production Of The Thermoset Matrix Aircraft Airframe Material With Temperature Controlled Fast Curing Methods.* Master, Selcuk.

AMENDOLA, V. & MENEGHETTI, M. 2009. Laser ablation synthesis in solution and size manipulation of noble metal nanoparticles. *Physical chemistry chemical physics,* 11**,** 3805-3821.

ARRK NORTH AMERICA, I. 2025. *Additive Manufacturing in Aerospace: A New Era of Innovation* [Online]. Available: https://us.arrk.com/additive-manufacturing-in-aerospace-a-new-era-of-innovation/ [Accessed 25/10 2025].

BAGHEL, P., SAKHIYA, A. & KAUSHAL, P. Ultrafast growth of carbon nanotubes using microwave irradiation: characterization and its potential applications. Heliyon. 2022; 8: 10943; PMID: 36276756. DOI.

BOUMAN, N., MARQUES, S. S. L., SEBBE, N. P. V., GERRITSE, A., BERNARDI, H. H., MENEZES, W. M. M., DA SILVA, F. J. G., MATSUSHIMA, J. T., GIOVANETTI, L. & SALES-CONTINI, R. D. C. M. 2025. Advancing Sustainability in Aerospace: Evaluating the Performance of Recycled Carbon Fibre Composites in Aircraft Wing Spar Design. *Journal of Composites Science,* 9**,** 384.

CHEN, Y., ZHANG, J., LI, Z., ZHANG, H., CHEN, J., YANG, W., YU, T., LIU, W. & LI, Y. 2023. Manufacturing technology of lightweight fiber-reinforced composite structures in aerospace: Current situation and toward intellectualization. *Aerospace,* 10**,** 206.

DEBROY, T., WEI, H. L., ZUBACK, J. S., MUKHERJEE, T., ELMER, J. W., MILEWSKI, J. O., BEESE, A. M., WILSON-HEID, A. D., DE, A. & ZHANG, W. 2018. Additive manufacturing of metallic components–process, structure and properties. *Progress in materials science,* 92**,** 112-224.

ETRI, H. E., KORKMAZ, M. E., GUPTA, M. K., GUNAY, M. & XU, J. 2022. A state-of-the-art review on mechanical characteristics of different fiber metal laminates for aerospace and structural applications. *The International Journal of Advanced Manufacturing Technology,* 123**,** 2965-2991.

GAO, B., ZHAO, H., PENG, L. & SUN, Z. 2022. A review of research progress in selective laser melting (SLM). *Micromachines,* 14**,** 57.

GUAN, C., ZHAN, L. & YAO, S. 2022. Finite element simulation and experimental research on uniformity regulation of microwave heating of composite materials. *Polymers,* 14**,** 3484.

KIM, H., KANG, D., KIM, M. & JUNG, M. H. 2020. Microwave curing characteristics of CFRP composite depending on thickness variation using FBG temperature sensors. *Materials,* 13**,** 1720.

KUMAR, P. K., RAGHAVENDRA, N. & SRIDHARA, B. 2011. Development of infrared radiation curing system for fiber reinforced polymer composites: An experimental investigation. *Indian Journal of Engineering & Materials Sciences,* 18**,** 24-30.

ŁACH, Ł. 2024. Recent advances in laser surface hardening: Techniques, modeling approaches, and industrial applications. *Crystals,* 14**,** 726.

LI, B., YANG, H., HE, J., YU, S., XIAO, R., LUO, H., WEN, Y., PENG, S., LIAO, X. & YANG, D. 2023. Photopolymerization of coating materials for protection against carbon steel corrosion. *Materials,* 16**,** 2015.

MALIK, M. S., SCHLÖGL, S., WOLFAHRT, M. & SANGERMANO, M. 2020. Review on UV-induced cationic frontal polymerization of epoxy monomers. *Polymers,* 12**,** 2146.

MET3DP. 2025. *Turbine Blade 3D Printing with Superalloys* [Online]. Metal Powders for 3D Printing and Additive Manufacturing. Available: https://met3dp.com/turbine-blade20250515/ [Accessed].

MGBEMENA, C. O., LI, D., LIN, M.-F., LIDDEL, P. D., KATNAM, K. B., THAKUR, V. K. & NEZHAD, H. Y. 2018. Accelerated microwave curing of fibre-reinforced thermoset polymer composites for structural applications: A review of scientific challenges. *Composites Part A: Applied Science and Manufacturing,* 115**,** 88-103.

MIAO, W.-J., WANG, S.-Q., WANG, Z.-H., WU, F.-B., ZHANG, Y.-Z., OUYANG, J.-H., WANG, Y.-M. & ZOU, Y.-C. 2025. Additive Manufacturing of Advanced Structural Ceramics for Tribological Applications: Principles, Techniques, Microstructure and Properties. *Lubricants,* 13**,** 112.

MOHD YUSUF, S., CUTLER, S. & GAO, N. 2019. The impact of metal additive manufacturing on the aerospace industry. *Metals,* 9**,** 1286.

NAIK, T. P., SINGH, I. & SHARMA, A. K. 2022. Processing of polymer matrix composites using microwave energy: A review. *Composites Part A: Applied Science and Manufacturing,* 156**,** 106870.

PMMT. 2025. *How Laser Cladding Extends Component Lifespan in Oil & Gas Operations* [Online]. PMMT. Available: https://pmmt.ae/laser-cladding-extend-component-lifespan-in-oil-gas/ [Accessed].

SAVANI, S., ALIPOUR, M., SHARMA, A. & KARUNAKAR, D. B. 2023. Microwave sintering of ZrB2-based ceramics: A review, Synth. Sinter. 3 (2023) 143–152.

SAVU, S. V., GHELSINGHER, C. D., ȘTEFAN, I., SÎRBU, N.-A., MIDAN, A.-A., DUMITRU, I., SAVU, I. D., NICOLICESCU, C. & DAVID, A. 2024. Microwave Hybrid Sintering and Soldering of Cu-Cr-W Composite Material for Reactive Power Breakers. *Materials,* 17**,** 4648.

SINGHAL, C. & MURTAZA, Q. 2018. Microwave sintering of advanced composites materials: a review. *Materials Today: Proceedings,* 5**,** 24287-24298.

TAMMAS-WILLIAMS, S., ZHAO, H., LÉONARD, F., DERGUTI, F., TODD, I. & PRANGNELL, P. B. 2015. XCT analysis of the influence of melt strategies on defect population in Ti–6Al–4V components manufactured by Selective Electron Beam Melting. *Materials Characterization,* 102**,** 47-61.

WANG, T., XIA, L., NI, M., PAN, S. & LUO, C. 2024a. Fundamentals of infrared heating and their application in thermosetting polymer curing: a review. *Coatings,* 14**,** 875.

WANG, X., CHEN, C., HU, S., CHEN, Z., JIANG, W., SHEN, G., HUANG, Z. & ZHOU, H. 2024b. In-situ infrared annealing for laser-assisted automated fiber placement to enhance interlaminar properties without sacrificing laydown efficiency. *Composites Part A: Applied Science and Manufacturing,* 183**,** 108214.

WANG, Y., THORN, T. D., LIU, Y., ADVANI, S. G., PAPAGEORGIOU, D. G., BILOTTI, E. & ZHANG, H. 2025. A review on energy-efficient manufacturing for high-performance fibre-reinforced composites. *Composites Part A: Applied Science and Manufacturing***,** 108779.

XIPING, F., SHENGYUAN, Z., QUN, L. & BO, W. 2021. A review of laser in**-** situ curing of thermoplastic composites. *China Plastics,* 35**,** 111.

ZHANG, Q., HUANG, H., LEI, C., LIU, Y. & LI, W. 2025. Review of Lightweight, High-Temperature Thermal Insulation Materials for Aerospace. *Materials,* 18**,** 2383.

ZHU, L., LI, N. & CHILDS, P. 2018. Light-weighting in aerospace component and system design. *Propulsion and Power Research,* 7**,** 103-119.

An Research On Soil Microstructure

Nazile URAL

Prof. Dr.; Bilecik Şeyh Edebali Üniversitesi Mühendislik Fakültesi İnşaat Mühendisliği Bölümü.
nazile.ural@bilecik.edu.tr ORCID No: 0000-0002-2268-842X

ABSTRACT

Soil improvement is involved in sustainability in terms of making a material usable. When the existing soil improved with cement, lime, or cement slag, the use of raw materials intended to come here prevent. Thus, both the existing soil becomes usable, and raw materials transfer to future generations. The cost of remediation is not costly in the long term. In particular, soil improvement studies are common in clay soils. Clay is a problematic soil in terms of settlements and swelling in geotechnical engineering. Clay affects the soil behavior even if it has a small percentage of the soil mass. Clays have their particular chemical and mineralogical properties. Therefore, soils having the same rate of clay exhibit different engineering properties due to their different physical and mechanical properties. Within the scope of this study, the microstructures of cured for 1 and 28 days' cement, lime, and cement-lime stabilized soil mixes investigated. On prepared soil samples by standard compaction test analyzed X-ray diffraction (XRD), mercury intrusion porosimetry (MIP), specific surface analysis (SSA,) and scanning electron microscopy (SEM). The experimental and analysis results obtained indicated that the addition of cement, lime, and cement-lime improved soil properties.

Keywords – Clay, Cement, Lime, Microstructure, Soil Improvement.

INTRODUCTION

Soil improvement methods, applied at different times and with different processes, have existed for thousands of years. Lime is still functional today, but it was first used on the Appian Way in Rome around 600 AD, and that roadway is still fully functional today. Furthermore, Chinese civilizations used stone and timber between 3000 and 2000 BCE (Nicholson, 2015; ASCE, 1978). Nowadays, as the population increases, new roads, residences and service buildings are being built. However, due to the rapidly growing urbanization, the necessity of building on low-strength soils necessitates finding solutions with sustainable methods. The most common solution is to improve soil properties. In road constructions, high strength materials obtained from the quarries, and the problems solved. But in this case, raw material consumption increases. Therefore, in a rapidly growing world, consumption increases, and natural resources decrease as a result of population growth.

On the other hand, it is a problem for long-term structures, mainly due to clay soils, swelling-shrinkage, seating, and low strength properties. Soil improvement processing purposes reducing compressibility, increasing strength, reducing/increasing permeability, and mitigating the potential for liquefaction (Nicholson, 2015). Soil improvement carries out by applying

physical and chemical methods. In chemical stabilization, physicochemical and physical interactions are at play. Some of the essential reactions see hydration, ion-exchange, pozzolanic reaction, flocculation, carbonation, etc. form in a mix of soil and the additives (Fang, 1991).

To reduce soil problems, soil improvement with additives widely used. Materials such as cement, fly ash, lime used in soil improvement. The additives can be used alone or mixed with other substances [Choquette et al., 1987; Mollins et al., 1996; Prusinski and Bhattacharya 1999; Onitsuka et al., 2002; Hydma and Avar 2006; Rao et al., 2006; Russo et al., 2007; Lin et al., 2007; Singh et al., 2008; Manasseh and Olufemi 2008; Misra et al., 2010; Amu et al., 2011; Cuisinier et al., 2011; Metelková et al., 2012; Liu et al., 2012; Louafi and Bahar 2012; Stoltz, et al., 2012; Wayal et al., 2012; Russo and Modoni 2013; Muhmed and Wanatowski 2013; Sante et al., 2014). Thus, sustainability ensured by making an unreasonable base soil, sub-base, base material, and so on. The concept of sustainability became a well-known term worldwide, particularly after the United Nations Conference on Environment and Development, held in Rio in 1992. To keep the raw materials for future years is important for geotechnical engineering. Briefly, the sustainability term is affected by economic, social, and environmental factors, both today's and future generations' life. Engineers are defined as the professional group that must provide the link between science and technology and participate in applying technology to situations requiring sustainable solutions (ASCE, 2001).

The existing ground can be improved with cement and lime to increase its carrying capacity and thus be used as a foundation ground, eliminating the need for raw materials. As a result, raw materials are not consumed and economy is achieved.

In lime stabilization, clay particles are broken down to form silica; in cement stabilization, the cement already contains silica and only requires water to initiate the hydration process (Patel and Patel, 2012). Therefore, a thorough understanding of lime and cement stabilization requires knowledge of clay mineralogy, clay chemistry, and its interaction with water. Secondary silicates form through the decomposition of primary minerals, and we call these formations clay minerals. Clay minerals consist of aluminum silicates, water molecules, and exchangeable cations such as calcium and magnesium.

Clays are notable for their fine-grained, flaky shapes, which also possess electrical charges on the crystal edges and surfaces. Clay particles are very fine, which allows them to have very high specific surface areas (Das, 2007; Nicholson, 2015). Specific surface area can be defined as the total surface area per unit mass of a material. A water molecule is a structure composed of one positive and one negative charge, called dipolar (Nicholson, 2015; Das, 2007; Holtz and Kovacs, 1981). The positive charges of positively charged ions and dipolar molecules are attracted by the net negative charge, and thus, the excess water molecule, along with its

attractive electric charge, binds to the clay particle, forming a layer. This water layer is called a "double water layer" or "extended double layer." As the double water layer shifts, the clay particles exhibit a flocculated or dispersed structure. This change changes the physical, mechanical and physicochemical properties of the soil, which plays a very important role in determining the engineering properties of the soil.

In soil improvement with cement occur all four processes in soil improvement with lime occurs three processes except cementitious hydration. Cement has dicalcium silicate (C2S), and the two calcium silicate phases are essential about soil stabilization. These two phases produce both calcium–aluminate–hydrate (C–A–H) and calcium–silicate–hydrate (C–S–H), so lead to cation exchange and flocculation and agglomeration, and strength in the soil structure. With improvement, this lime, firstly, silica, and alumina, are released, after react with calcium from the lime to form calcium-silicate-hydrates (C-S-H) and calcium-aluminate-hydrates (C-A-H). Thus, this state provides the strength stabilized soil layers. In both soil improvements expect to change soil form, for example, permanent, durable, and significantly impermeable, producing a structural layer that is both strong and flexible (Prusinski and Bhattacharya, 1999; Firoozi et al. 2017).

The stabilization process begins very quickly with cation exchange, followed by flocculation and agglomeration. Flocculation and agglomeration transform the clay's structure from fine-grained to coarser-grained. Flocculation is the structure in which soil particles stick together and form larger sized areas, clusters, peds, flakes and micro/macro pores (Figure 1). Agglomeration has processed the transportation and collision of particles, the attachment, and the cementation of the particles. Flocculation and agglomeration occur within a few hours of mixing. As a result of cement hydration, calcium-aluminum-hydrate (C-A-H) and Ca(OH)2 are formed. This form, formed with cement, acts as the glue, providing structure and strength to the mixture. This process can last from one day to one month or even years. The pozzolanic reaction is seen in lime-soil and cement-soil systems. This process takes place slowly for years. After the pozzolanic reaction, mixtures gain strength, reduce plasticity, and improve gradation (Prusinski and Bhattacharya, 1999).

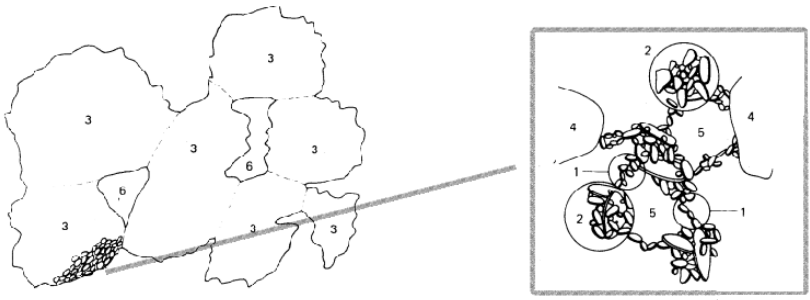


Figure 1. Diagram of the soil structure 1. domain 2.cluster 3.ped 4. Silt 5. Micropore 6. Macro pore (Holtz and Kovacs, 1981)

Using cement or lime to improve the soil beneath structures and infrastructures such as roads, airports, and pipelines makes soil stabilization solutions more economical and easier to implement. Clay soils may have a settlement, swelling, low strength, and undesirable hydraulic conductivity. As a result of these features, soil improvement studies performed before the structures make on clay soils. Clay stabilization is a result of chemical mechanisms between soil and chemical materials. Borchardt 1984 listed the reactions that occurred property of cementation, cation exchange, anion exchange, conversion of salts, water film changes, absorption ability of the water film, fixation property, ability to form new minerals, using ions to enrich pore water, ability to absorb, changes in capillary forces, changes in the clay minerals based on electrical surface tension, changes in electrical forces found in the particles, changes in chemically bound water, chemically bound water's absorption ability, acid neutralization, base neutralization, and proton exchange. The buildings, dams, bridges, roads, and similar engineering structures to be built on the soil should be able to serve. For this, the conditions of settlement, swelling, tilting, lateral movement, vibration, and durability should provide. In the case of any structure, changes saw in the stress condition of the ground. In general, settlements are observed due to the increase in stresses, while swelling find as stresses decrease. Especially in the United States, shrinking and swelling soils cause billions of dollars in damage to structures and roads each year (Krohn and Slosson, 1980; Coduto, 1988). This damage can be approximately twice the damage caused by floods, earthquakes, and hurricanes (Jones and Holtz, 1973). Also, soil problems due to settlement see all the world. The settlement problem defined as a deformation in the soil due to the applied stresses. One of the most well-known examples of different settlements on the soil is the tower of Pisa. In 1889, Eiffel conducted a detailed survey of the Eiffel Tower in Paris, avoiding problems such as the Leaning Tower of Pisa. But seating problems are not limited to buildings. An example of this is the settlement of a bridge approach fills on the road in California. Twelve years after the

construction of the bridge, the embankment settled approximately 1m (Coduto, 1988).

As a result of their studies, Prusinski and Bhattacharja (1999) stated that cement and lime are effective in improving clayey soils if used adequately. Onitsuka and colleagues (2001) investigated the microstructure and strength of clays stabilized using lime and cement. They conducted consistency limit, SEM, hydraulic conductivity, unconfined compression, and oedometry tests on the mixtures and concluded that the clay microstructure changed. Strengths increased with curing time and additive content in two different stabilized clays. This suggests that this is due to texture and cementation in the microstructure. However, many researchers have used chemical agents in clay-sand mixtures (Studds et al., 1998; Alawaji, 1999).

Lime and cement, in particular, are successful materials for soil improvement when used as chemical additives in clayey soils (Horpibulsuk et al., 2010; Khemissa and Mahamedi, 2014). Al-Rawasa et al. (2005) investigated the effect of pozzolanic mixtures containing lime and cement on soil swelling potential by adding lime and cement at 3%, 6%, and 9% of the soil dry weight. They reported that the swelling percentage and swelling pressure decreased to zero with the addition of 6% lime, and that the swelling potential decreased to zero with the application of heat treatment. Mu'azu (2005) observed how bagasse ash altered the plasticity and particle size distribution in cement-treated lateritic soil. A mixture containing 2-8% bagasse ash was initially prepared in the lateritic soil, followed by a 1-4% cement mixture. They concluded that the plastic limit increased, while the liquid limit and plasticity index decreased. Furthermore, a decrease in the percentage of fines was observed due to the flocculation and agglomeration of the clay mixture.

Chen et al. (2007) added incinerated sewage sludge ash/cement and cement at 0%, 2%, 4%, 8% and 16% by weight to improve soft soils. pH testing, consistency testing, compression, California bearing ratio (CBR), and shear strength tests were then conducted on these samples. The results indicated that the unconfined compressive strength of the mixtures improved three to seven times with the addition of cement, and CBR values increased by up to 30 times with the addition of cement. Ganjidoust et al. (2009) investigated the stabilization of heavy metal-contaminated soils with cement by using compressive strength, toxicity-characterized leaching procedure, X-ray diffraction, and scanning electron microscopy on samples containing cement at 5%, 10%, 15%, and 20% of the soil weight. They concluded that the mixtures exhibited high compressive strength after 7 days of curing, while after leaching tests, the leaching metal concentration decreased after 1 to 28 days of curing, as observed by XRD or absorption in CSH gel. Fattah et al. In 2010, they used expansive soil to examine the properties of soils with various additives. They then used cement, steel fibers, gasoline fuel,

and cement mortar injection to improve the soil. By performing swelling tests, consolidation tests, and direct shear tests, they found that the compressive and swelling indices of the improved soil decreased and slightly affected the cohesion value due to the variation between the additive and the soil particles.

Horpibulsak et al. (2010) reported that, with varying water content, curing time, and cement content, large pores within the soil filled with cement over time, small pore volume increased, and total pore volume decreased. This, in turn, led to an increase in strength over time. Using scanning electron microscopy, mercury intrusion pore size distribution measurements, and thermal gravity analysis, they concluded that over time, large pores within the soil filled with cement, small pore volume increased, and total pore volume decreased. This, in turn, led to an increase in strength. Oyediran (2011) studied the effect of cement content on the soil properties of some lateritic soils. He prepared samples by adding cements at 2%, 4%, 8%, 10%, and 20% of the soil weight and conducted standard proctor tests, California bearing ratios, and unconfined compressive strength tests. As a result, as cement was increased after the improvement, there was an increase in maximum dry density, CBR, and UCS, and a decrease in optimum moisture content. However, with the addition of more than 10% cement, there was a decrease in maximum dry density, qu, and CBR, and an increase in optimum moisture content. To investigate some mechanical properties of cement-stabilized soft clay, Ho and Chan (2011) cured samples prepared with 0, 5, and 10% cement by weight for different days. They found that increasing the cement content and curing time decreased compressibility in one-dimensional compressibility, and that, in general, the cement ratio was more effective than the curing time.

Bayat et al. (2013) investigated the effects of cement and lime on the geotechnical properties of a low-plasticity clay. They stated that the addition of cement or lime affected the optimum moisture content, maximum dry unit weight, and plasticity index, and that cement content increased the unconfined compressive strength of the soils. Additionally, cement application was found to have a higher effect on the mechanical behavior of the soil than lime application. Khemissa and Mahamedi (2014) investigated the effects of various concentrations of cement and lime on expansive overconsolidated clay. They noted that the swelling potential decreased with the decrease in specific surface area and changes in microstructure. Ikhlef et al. (2014) examined the physicochemical behavior of materials improved with cement in roadworks. The experiments were conducted by subjecting mixtures prepared with different amounts of cement to the standard proctor test. After preparation, a decrease in the plasticity index was observed, changing the soil class. The material's class was converted from CL to CL-ML. Finally, the researchers stated that the material could be used as fill or subgrade.

Saeed et al. (2014) observed the strength and compressibility of samples prepared with different curing times and cement amounts to investigate the physicochemical properties of kaolin clay mixed with cement. The researchers conducted compression, unconfined compressive strength, and one-dimensional consolidation tests on samples containing 5% and 10% cement, and noted a long-term increase in unconfined compressive strength and a decrease in the compressive index. Jaritngam et al. (2014) investigated the usability of a laterite-cement mixture as an aggregate in road base layers. In this study, the strength properties of a cement-laterite soil were compared with those of crushed rock. Strength tests and analyses revealed that a 3% by weight laterite-cement mixture demonstrated compressive strength comparable to that of crushed rock and suggested its suitability as a road surface. Zambri et al. 2018 investigated peat soil stabilization using lime and cement and compared the treatment of peat soil. In this study, peat soil mixtures of 10% and 20% dry weight were prepared and direct shear box tests were conducted. They observed that the shear strength increased by 14% with the additives and concluded that lime and cement were the appropriate stabilizers for peat soil. Jose and Rajamane 2018 investigated the effect of lime and cement on soil strength properties. The researchers conducted experiments with lime and cement in clayey soils of varying plasticity and observed that the optimum lime content was 4% in high-plasticity clayey soil, 2% in low-plasticity clayey soil, and 4% in high- and low-plasticity clayey soil. The unconfined compressive strength test value increased with the curing time and the increase in the percentage of lime and cement. Phan and Pham (2018) determined the optimum water content, maximum dry density, qu, and Young's modulus on samples prepared with cement and cement-lime to investigate the effects of cement and lime on the engineering properties of soft soils. They concluded that the cement additive is suitable for use as a subbase layer for soil improvement and cement stabilization. Curing and cement content increased the Young's modulus. Amhadi and Assaf 2019 investigated the effect of cement on the strength of sand dunes in the Libyan deserts. Strength tests were conducted on samples containing 0%, 3%, 5%, and 7% cement by weight. The researchers reported that the strength properties of sand and cement were improved, thus saving material and time and reducing road construction costs. Sharo et al. 2019 used lime and cement-lime at 3%, 5%, 8%, and 12% by weight to investigate the improvement of expansive soil using cement, quicklime, and cement-lime. Experiments on the mixtures showed that the plasticity index and potential swelling decreased, while the unconfined compressive strength (qu) and maximum dry density (γdmax) increased with the addition of cement and lime. Muhardi et al. 2019 investigated the stabilization of peat soil using a lime-cement mixture to prevent peat fires. They conducted compression, unconfined compressive strength, California bearing ratio tests, and burning tests on stabilized peat and unstabilized peat

soils, and found that these soils exhibited increased strength. Burning tests revealed that stabilized peat soils took much longer to reach their burn point and ash point, completely turning them into ash. Therefore, stabilized peat soils may take longer to burn and may be more difficult to burn.

As a result of stabilization with materials such as cement, lime, etc., structural changes and reactions occur in the internal structure of the ground. Cation exchange, flocculation, and agglomeration mechanisms increase soil workability, while carbonation and pozzolanic reactions increase bearing capacity (Mitchell, 1976). This study investigated microstructural changes in cement, lime, and cement-lime mixtures. Mixtures containing 0%, 4% cement and lime, and 2% cement-lime were used in soil prepared using standard compaction tests. XRD, MIP, BET, and SEM analyses were performed on all prepared mixtures after 1 and 28 days of curing.

MATERIALS AND METHODS

As materials used, clay, silica sand, cement, lime, and cement/lime. The soil characterized as high plasticity clay (CH). Physical properties of this clay are liquid limit 103, Plastic limit 38, sand percentage 13%, Silt percentage 46%, Clay percentage 41%, and specific density 2.66. Liquid limit determined by the fall cone method. Silica sand is sand 100% and specific density 2.64. The mineralogical composition of clay and silica sand showed in Figure 2. For stabilizing soft clay, use lime and cement are active admixtures. Stabilizer materials are lime and cement (CEMI cement, industrial lime), available in construction markets. The mineralogical composition of lime and cement is given in Figure 3, showing calcium carbonate in the lime and aluminous in the cement. First, the soil was amended with 0%, 20%, 40%, 60%, and 80% silica sand by weight. Mixture X consisted of clay and 20% sand, Mixture Y consisted of clay and 40% sand, Mixture Z consisted of clay and 60% sand, and Mixture W consisted of clay and 80% sand. Then, 4% cement and lime, and 2% cement and 2% lime were added to the mixtures by weight. Samples X, Y, Z, and W were compacted according to ASTM D698, 2000.

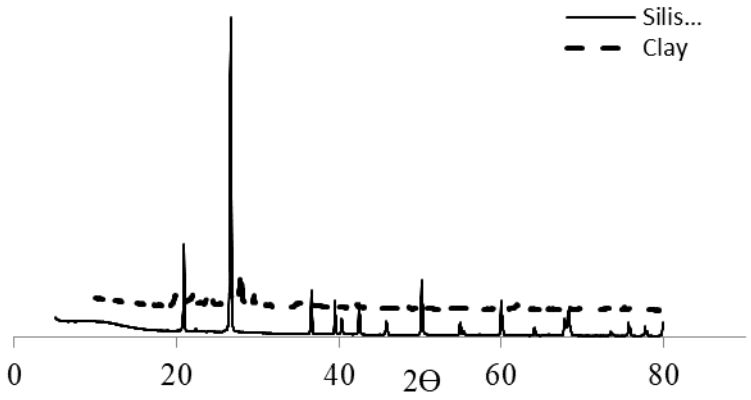


Figure 2. X-ray diffraction patterns of clay and silica sand

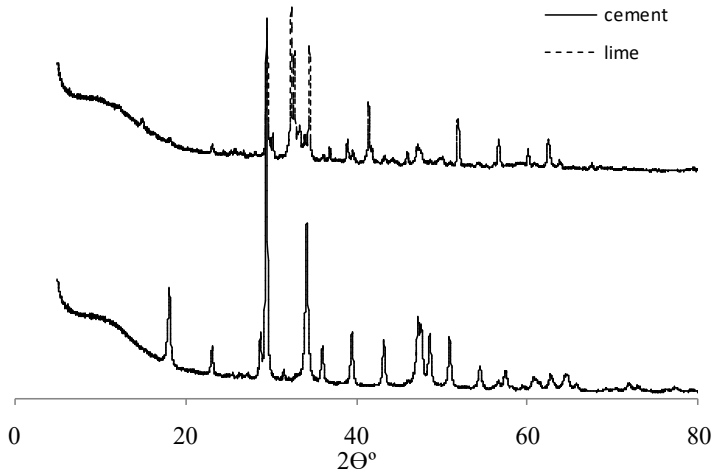


Figure 3. X-ray diffraction patterns of lime and cement

Knowing the mineralogical composition of soils is crucial for understanding their impact on geotechnical behavior, and X-ray diffraction (XRD) is used for this purpose. Specific surface area (SSA) affects the physical and chemical properties of soils, and in this study, specific surface area was determined using the Brunauer, Emmett, and Teller (BET) method. The specific surface area of clay is 36.91 m2/g. Because pore size distribution affects the geotechnical behavior of soils, MIP testing was performed to determine pore size distributions. Another fundamental analysis used to examine microstructure is scanning electron microscopy (SEM) images. In

geotechnical engineering, microstructural changes in soils are crucial for understanding and evaluating soil behavior.

RESULTS AND DISCUSSION

The X-ray diffraction of the used clay and sand are shown in Figure 2. Montmorillonite, anorthite, and quartz are seen in the analysis results. The X-ray diffraction results of the lime and cement, showing calcite and silica, are shown in Figure 3.

The specific surface areas of the samples are shown in Figure 4, and it is observed that the SSA of the mixtures decreases with the addition of cement, lime, and cement-lime. The equivalent percentages of lime, cement, and cement-lime were compared; for the cement-containing mixtures X, Y, and Z, the SSA was found to be lower than that of the lime and cement-lime mixtures, while for the W mixture, the SSA of the cement-lime mixture was found to be lower than that of the other mixtures.

MIP was performed for all soil samples, and the total intrusion volumes (TIV) are shown in Figure 5. Comparing the results, it is observed that for samples X and Z, the TIV of the cement-lime mixtures is lower than that of the lime and cement mixtures. For sample Y, the TIV of the lime mixtures is lower than that of the cement and cement-lime mixtures. For sample W, the TIV of the cement mixtures is lower than that of the lime and cement-lime mixtures.

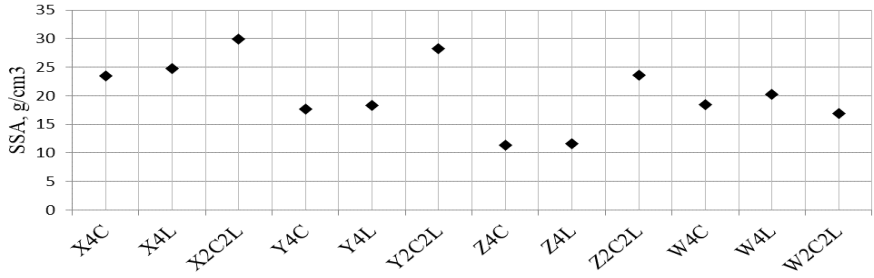


Figure 4. Specific Surface Area (g/cm3) of mixtures

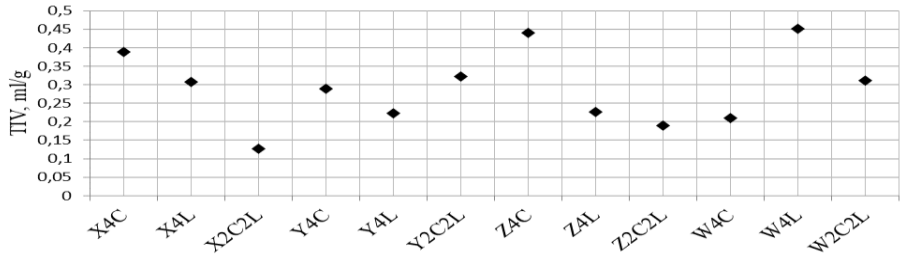


Figure 5. Total intrusion volume (ml/g) of mixtures

SEM conducted for all mixtures after 1 and 28 days of curing. Thus, the micro-structures of the samples observed for in short and in long term reactions. Figures 6-10 show the SEM images for Clay, X4C, X4L, X2C2L, Y4C, Y4L, Y2C 2L, ZS4C, Z4L, Z2C2L, W4C, W4L and W2C2L after 1 and 28 days of curing. The SEM images of the X4C, X4L, and X2C2L samples after 1 and 28 days of curing showed in Figure 7 (×10.000 magnification). In Figure 8, the microstructure of the clay sample appears to contains fewer flaky layers compared with Figure 6 and flocculation and agglomeration with the use of additives (sand, cement, lime, and cement-lime) after one day of curing (in the short term). Figures 8, 9, and 10 exhibits the same trend as Figures 7. Results of SEM images showed that the lime, cement, and cement-lime is active after in the short term. This finding is also consistent with those of Tran et al. 2014, Jiang et al. 2015 and Elandaloussi et al. 2019. In cement, lime, and cement-lime mixtures, the flocculation structure became more pronounced with curing time. So produced a more aggregated structure. This finding is also consistent with those of Rajasekaran and Rao 1997, Chew et al. 2004, Al-Mukhtar et al. 2012. The cement caused large-pores to increase, and small pores decrease due to the hydrated cement in the short term. Generally, the pores between the pellets reduce due to the pozzolanic reaction in the long term. After 28 days of curing (in the long term), the microstructure of the sample exhibited more pronounced flocculation with the use of additives. The same lime percentage of stabilized samples showed a reduction in a large pore and an increase in the small pore in a long time. The same cement and cement-lime percentage stabilized samples exhibited decreased pore due to pozzolanic reactions a long time. However, in stabilized samples, it is seen that big particles increase with increasing sand.

Briefly, the use of additives cause a decrease in the number of flaky layers, and the pores between the pellets are more evident. According to the SEM images, the flocculation that developed during the pozzolanic reactions is more pronounced due to the increase in curing time. The results obtained are also consistent with the results of Al-Mukhtar et al. 2014, Jiang et al. 2015, Zhao et al. 2015, and Bouras et al. 2020. Soil improvement saw by the SEM images. The stabilized cement, lime, and cement-lime samples were

observed white lumps in the long term. Similar white lumps found by Kassim 2009, Muhmed & Wanatowski 2013, and Saeed et. al. 2014.

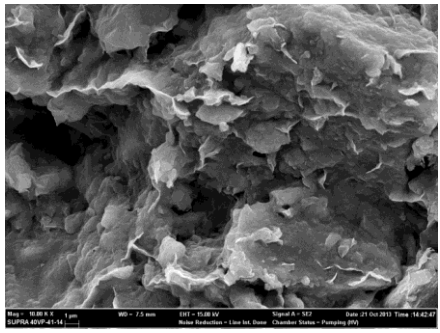
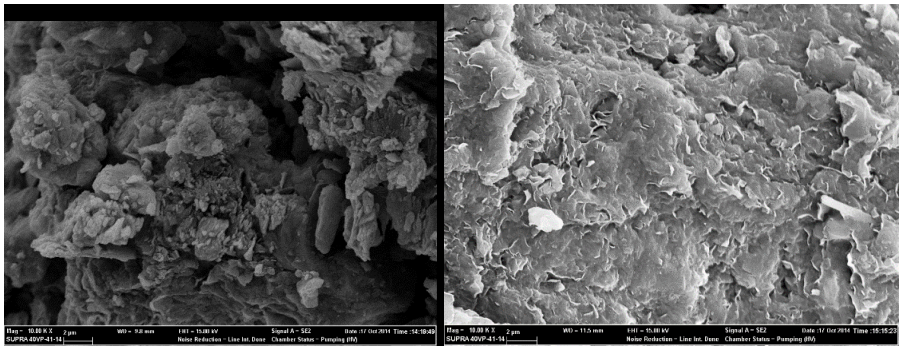
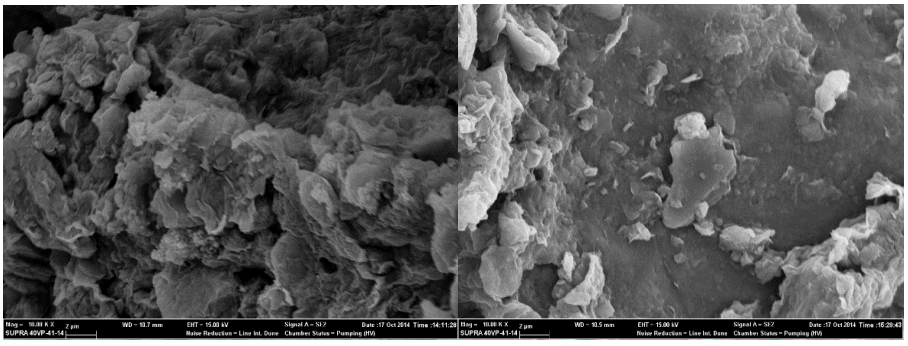


Figure 6. SEM images of compacted clay samples x10000 magnification



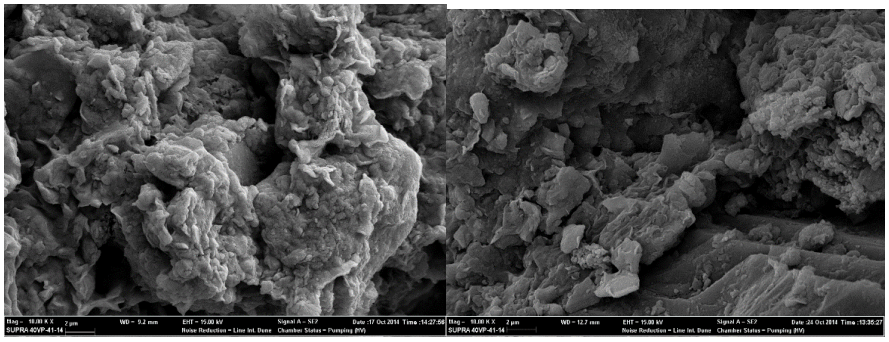
1 day curing 28 day curing

(a) X4C



1 day curing 28 day curing

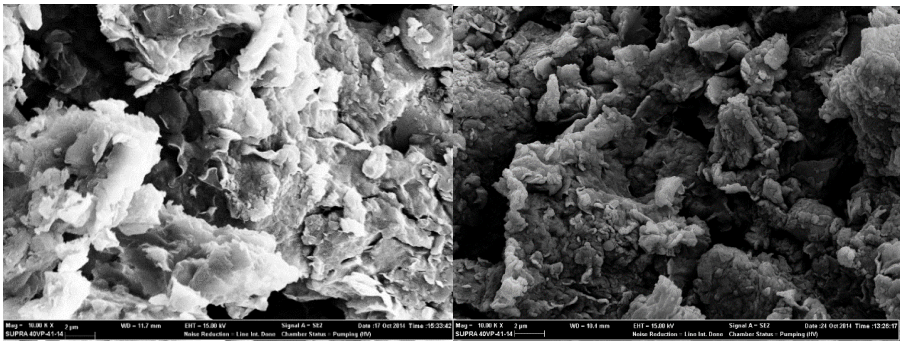
(b) X4L



1 day curing 28 day curing

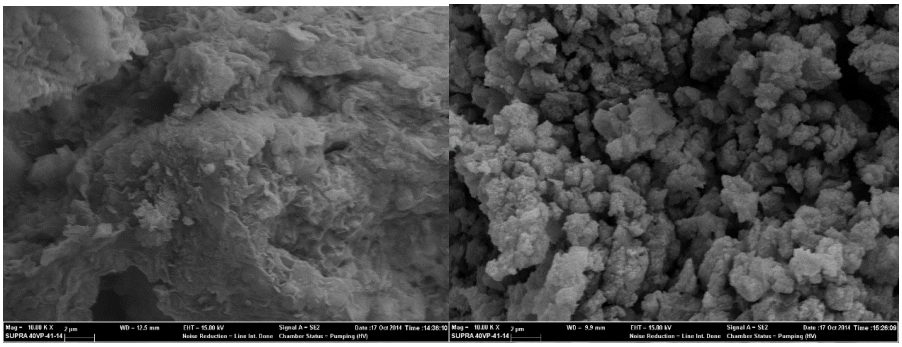
(c) X2C2L

Figure 7. SEM images of mixtures with X for x10.000 magnification



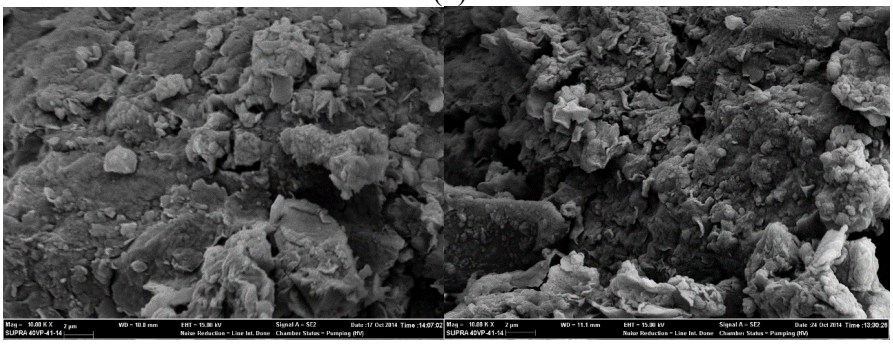
1 day curing 28 day curing

(a) Y4C



1 day curing 28 day curing

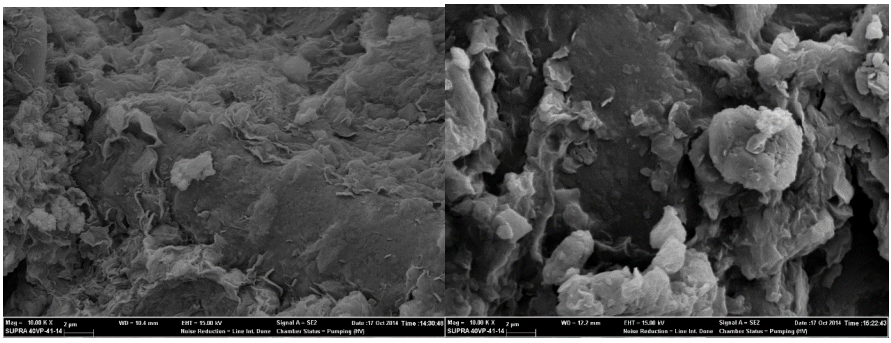
(b) Y4L



1 day curing 28 day curing

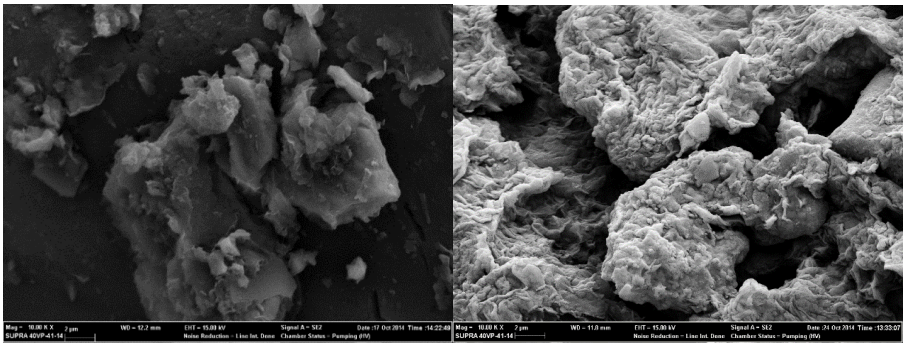
(c) Y2C2L

Figure 8. SEM images of mixtures with Y for x10.000 magnification



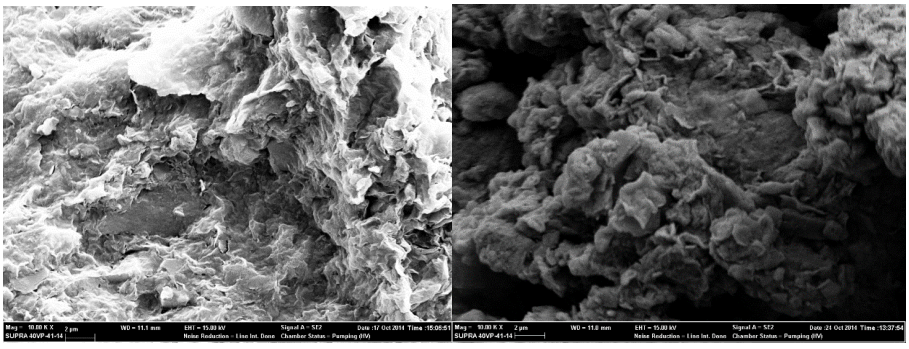
1 day curing 28 day curing

(a) Z4Ce



1 day curing 28 day curing

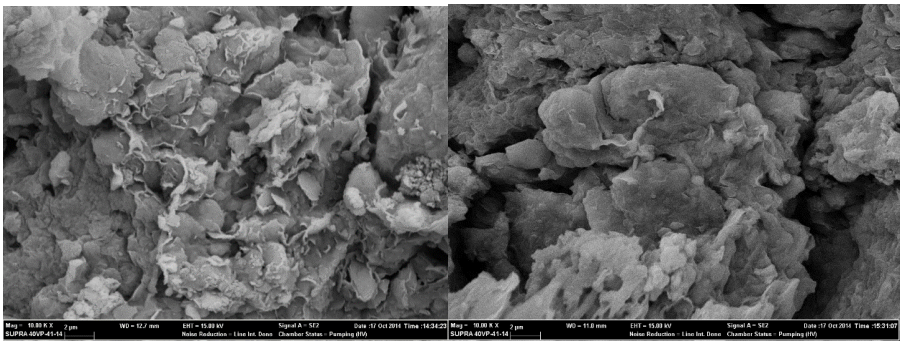
(b) Z4L



1 day curing 28 day curing

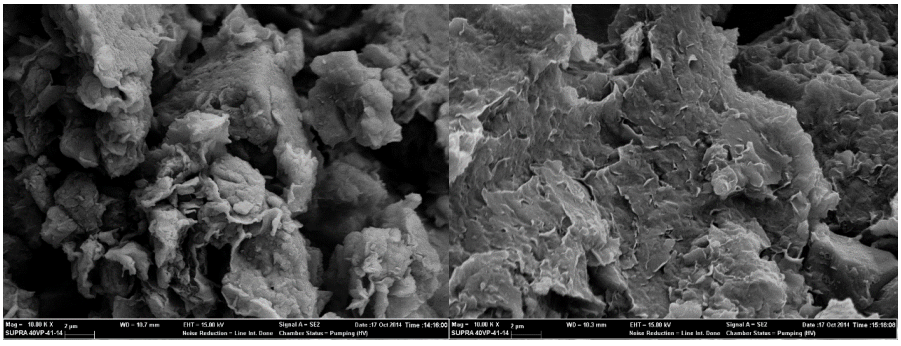
(c) Z2C2L

Figure 9. SEM images of mixtures with Z for x10.000 magnification



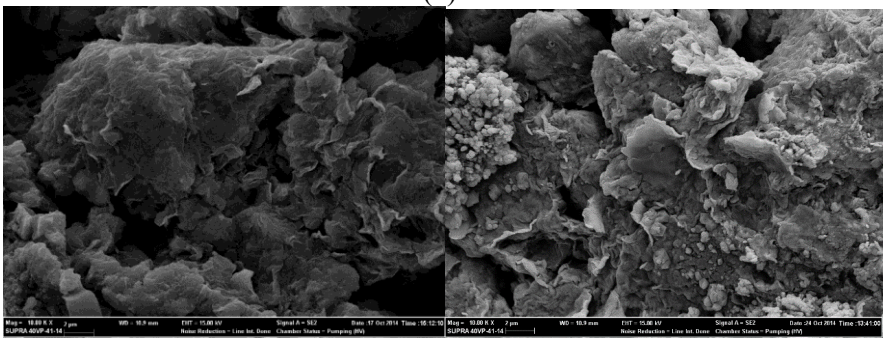
1 day curing 28 day curing

(a) W4Ce



1 day curing 28 day curing

(b) W4L



1 day curing 28 day curing

(c) W2C2L

Figure 10. SEM images of mixtures with W for x10.000 magnification

CONCLUSIONS

There are insufficient bearing capacity, high plasticity, and high swelling potential in construction sites; they need to be improved to be usable. Particularly in the improvement of cohesive soils, chemical methods using various additives are applied. Lime and cement are proper improvement additives for cohesive soils with high plasticity and high clay content. This state is due to reactions between clay and cement and lime. These reactions provide short-term and long-term improvements. Evaluation of experimental results and analyses presented in this study the following results reached;

•X-ray diffraction results of clay and sand reveal montmorillonite, anorthite, and quartz, while X-ray diffraction results of lime and cement reveal calcite, calcite, and silica. The SSA of the mixtures decreased with the addition of cement, lime, and cement-lime. The TIVs of the cement-lime mixtures were smaller than those of the lime-cement mixtures for X and Z, the TIVs of the lime mixtures were smaller than those of the lime and lime/cement mixtures for Y, and the TIVs of the cement mixtures were smaller than those of the lime and lime/cement mixtures for W.

• The microstructure of the clay sample showed short-term flocculation and agglomeration due to the use of cement, lime, and cement-lime additives. This suggests that the lime, cement, and cement-lime were active in the short term. The cement, lime, and cement-lime mixture and the curing time allowed the flocculation structure to be more clearly seen. This resulted in a more aggregated structure in the short term. Generally, porosity in pellets decreases over time due to the pozzolanic reaction. The microstructure of the sample exhibited more pronounced flocculation with the use of additives over time. According to SEM images, flocculation during pozzolanic reactions becomes more pronounced with increasing curing time.

ACKNOWLEDGEMENTS

This study is part of research project 2011-02-BIL.03-01, which is sponsored by Bilecik Seyh Edebali University. The authors are grateful to Bilecik Seyh Edebali University for its financial assistance.

REFERENCES

Alawaji, H.A. (1999). Swell and compressibility characteristics of sand – bentonite mixtures inundated with liquids. Applied Clay Science, 15(3-4), 411-430.

Al-Mukhtar, M., Khattab, S., Alcover, J.F. (2012), Microstructure and geotechnical properties of lime-treated expansive clayey soil. Engineering Geology, 139-140, 17-27.

Al-Mukhtar, M. Lasledj, A. and Alcover, J.F. (2014). Lime consumption of different clayey soils. Applied Clay Science, 95, 133–145.

Al-Rawas A. A., Hago ,A.W. and Al-Sarmi, H. (2005). Effect of lime, cement and Sarooj (artificial pozzolan) on the swelling potential of an expansive soil from Oman. Building and Environment, 40, 681–687.

Amhadi, T. S. and Assaf, G. J. (2019). Assessment of strength development of cemented desert soil. International Journal of Low-Carbon Technologies, Vol 14, Issue 4, pp. 543–549.

Amu, O. O., Bamisaye, O. F. and Komolafe, I. A. (2011). The Suitability and Lime Stabilization Requirement of Some Lateritic Soil Samples as Pavemen. Int. J. Pure Appl. Sci. Technol., 2(1), pp. 29-46.

ASCE (1978). Soil improvement: history, capabilities and outlook. Report by the Committee on Placement and Improvement of Soils, Geotechnical Engineering Division. 182 pp.

ASCE, (2001). The role of the civil engineer in sustainable development. American Society of Civil Engineers, Policy Statement 418, ASCE, Reston, Va.

ASTM D698, (2000). Standard Test Methods for Laboratory Compaction Characteristics of Soil Using Standard Effort, Annual book of ASTM Standards West Conshohocken PA Vol.15.09.

Bayat, M., Asgari, M.R., Mousivand, M. (2013). Effects of cement and lime treatment on geotechnical properties of a low plasticity clay. International Conference on Civil Engineering Architecture & Urban Sustainable Development, 27&28 November, Tabriz, Iran.

Borchardt, G. (1984). Stabilization of Landslides: Effects of Various Chemicals on the Laboratory Shear Strength of an Expansive Soil. Rep. No. Special Report 155, California Department of Conservations, Division of Mines and Geology, Sacramento, CA. Stabilization. Transportation Research Record, (1295): 45-51.

Bouras, F., Tapsoba, N., Martin, M., Sabio, S., Jacquet, A. and Beck, K., Belayachi, N., Bouasker, M. and Al-Mukhtar, M. (2020). Effect of Hydrated Lime and Cement on the Engineering Behavior of Highly Expansive Clay. Athens Journal of Technology & Engineering, 7: 1-14.

Chen, L., Lin, D.F. and Luo, H.L. (2007). Utilization of Incinerated Sewage Sludge Ash/Cement for Soft Soil Improvement. In:The IWA Specialist Conference, Moving Forward, Wastewater Biosolids Sustanability: Technical, Managerial, and Public Synergy, 24-27, New Brunswick, Canada, 559-566.

Chew, S.H., Kamruzzaman, A.H.M., Lee, F.H. (2004). Physicochemical and engineering behavior of cement treated clays.. J. Geotech. Geoenviron. Eng., 130, 696–706.

Coduto, D.P. (1988). Geotechnical Engineering Principles and Practices. Prentice Hall, Upper Saddle River, NJ, USA, p.759.

Cuisinier, O., Auriol, J-C., Borgne, T. L. Deneele, D. (2011). Microstructure and hydraulic conductivity of a compacted lime-treated soil. Engineering Geology, 123 (2011), pp. 187-193.

Elandaloussi, R. Bennabi, A., Dupla, J. C.. Canou, J., Benamar, A., Gotteland, P. (2019). Effectiveness of Lime Treatment of Coarse Soils Against Internal Erosion. Geotech Geol Eng, 37:139–154.

Fang, H-Y. (1991). Foundation Engineering Handbook, Van Nostrand Reinhold, New York, The USA, 923 page. Choquette, M., Bérubé, M-A. and Locat, J. (1987). Mineralogical and microtextural changes associated with lime stabilization of marine clays from eastern Canada. Applied Clay Science, Vol 2, Issue 3, pp. 215-232.

Das, B.M. (2007). Principles of Foundation Engineering, 6th ed. New York, USA: Thomson, 480 p.

Holtz, R. D., Kovacs, W. D. (1981). An Introduction to Geotechnical Engineering. New Jersey:Prentice-Hall, 733 p.

Fattah, M. Y. and Shwan, B. J. (2010). A treatment of expansive soil using different additives. Acta Montanistica Slovaca, Vol 15, 4, 290-297.

Firoozi, A. A., Olgun,, C. G., Firoozi, A. A. and Baghini, M. S. (2017). Fundamentals of soil stabilization. Geo-Engineering, 8:26, https://doi.org/10.1186/s40703-017-0064-9.

Ganjidoust H., Hassani A. and Ashkiki R. A. (2009). Cement-Based Solidification/Stabilization of Heavy Metal Contaminated Soils with the Objective of Achieving High Compressive Strength for the Final Matrix, Transaction A:Civil Engineering Vol. 16, No. 2, pp. 107-115, Sharif University of Technology.

Ho, M-H. and Chan, C-M. (2011). Some Mechanical Properties of Cement Stabilized Malaysian Soft Clay, World Academy of Science. Engineering and Technology International Journal of Civil and Environmental Engineering, Vol:5, No:2.

Horpibulsuk, S., Rachan, R., Chinkulkijniwat, A., Raksachon, Y., Suddeepong, A. (2010). Analysis of Strength Development in Cement-stabilized Silty Clay from Microstructural Considerations. Construction and Building Materials, 24(24), 2011-2021.

Hydma, N. and Avar, B.B. (2006). Changes in swell behavior of expansive clay soils from dilution with sand. Environmental and Engineering Geoscience, 12(2), 137-145.

Ikhlef, N. S., Ghembaza M. S. and Dadouch M. (2014). Effect of Cement and Compaction on the Physicochemical Behavior of a Material in the Region of Sidi Bel Abbes. Engineering, Technology & Applied Science Research Vol. 4, No. 4, 677-680.

Jaritngam, S., Somchainuek, O. and Taneerananon, P. (2014). Feasibility of Laterite-Cement Mixture as Pavement Base Course Aggregate. IJST Transactions of Civil Engineering, Vol. 38, No. C1, pp 275-284.

Jiang, N.-J., Du, Y.-J., Liu, S.-Y., Wei, M.-L., Horpibulsuk, S., Arulrajah, A. (2015). Multi-scale laboratory evaluation of the physical, mechanical, and microstructural properties of soft highway subgrade soil stabilized with calcium carbide residue. Can. Geotech. J, 53, 373–383.

Jones, D.E. and Holtz, W.G. (1973). Expansive soils-the hidden disaster. Civil Engineering ASCE, 43, 87-89.

Jose, R., Rajamane, N.P. (2018). Influence of lime and cement in strength characteristics of soil. International Journal of Scientific & Engineering Research, Vol 9, Issue 1, January.

Kassim, K.A. (2009). The Nanostructure Study on the Mechanism of Lime Stabilised Soil. Research Vot No: 78011, Department of Geotechnics and Transportation, Universiti Teknologi Malaysia.

Khemissa, M. and Mahamedi, A. (2014). Cement and lime mixture stabilization of an expansive overconsolidated clay. Applied Clay Science, 95, 104–110.

Krohn, J.P. and Slosson, J.E. (1980). Assessment of expansive soils in the United States. Proceedings 4.th International Conference on Expansive, Denver, CO, 596-608.

Lin, D-F., Lin, K-L., Hung, M-J. and Luo, H-L. (2007). Sludge ash/hydrated lime on the geotechnical properties of soft soil. Journal of Hazardous Materials, Vol 145, Issues 1–2, pp. 58-64.

Liu, M. D., Indraratna, B., Horpibulsuk, S. and Suebsuk, J. (2012). Variations in strength of lime-treated soft clays. Proceedings of the Institution of Civil Engineers, 165 (4), 217-223.

Louafi, B. and Bahar, R. (2012). Sand: an additive for stabilization of swelling clay soils. International Journal Geoscience, 3, 719–725.

Manasseh J., Olufemi, A. I. (2008). Effect of lime on some geotechnical properties of Igumale Shale. Electronic Journal of Geotechnical Engineering, Vol. 13, Bund. A.

Metelková, Z., Boháč J., Přikryl R. and Sedlářová, I. (2012). Maturation of loess treated with variable lime admixture: Pore space texturalevolution and related phase changes. Applied Clay Science 61, 37–43.

Misra, A. K., Mathur, R., Rao, Y. V., Singh, A. P. and Goel, P. (2010). A New Technology Of Marble Slurry Waste Utilization İn Roads. J Sci Ind Res, 69, 67-72.

Mitchell, J.K. (1976). The properties of cement-stabilized soils. Proceeding of Residential Workshop on Materials and Methods For Low Cost Road, Rail, and Reclamation Works, 365–404.

Mollins, L.H., Stewart, D.I. and Cousens, T.W. (1996). Predicting the properties of bentonite–sand mixtures. Clay Minerals, 31, 243–252.

Mu'azu, M. A. (2007). Evaluation of Plasticity and Particle Size Distribution Characteristics of Bagasse Ash on Cement Treated Lateritic Soil. Leonardo Journal of Sciences, Issue 10, pp.137-152.

Muhardi, G. W. and Harist, F. R. Z. (2019). Peat soils stabilization using lime-cement mixture to prevent peat fires. MATEC Web of Conferences 276, 05006.

Muhmed A. and Wanatowski D. (2013). Effect of Lime Stabilisation on the Strength and Microstructure of Clay. IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE), e-ISSN: 2278-1684, ISSN: 2320-334X, Vol 6, Issue 3, pp. 87-94.

Nicholson, P. (2015). Soil improvement and ground modification methods, Oxford, England, Waltham, Massachusetts, BH, 441 p.

Onitsuka, K., Modmoltin, C. and Kouno, M. (2001). Investigation on Microstructure and Strength of Lime and Cement Stabilized Ariake Clay. Reports of the Faculty of Science and Engineering, Saga University, Vol. 30, No.1.

Onitsuka, K. Modmoltin, C. Kouno, M. Negami, T. (2002). Effect of Humic Acid On Lime Stabilized Ariake Clay. Proceedings of the twlfth, The Twelfth International Offshore and Polar Engineering Conference, 26-31 May, Kitakyushu, Japan.

Oyediran, I. (2011). Effect of Increasing Cement Content on Strength and Compaction Parameters of some Lateritic Soils from Southwestern Nigeria. Electronic Journal of Geotechnical Engineering, January.

Patel, M. and Patel, H.S. (2012). A Review on Effects of Stabilizing Agents for Stabilization of Weak Soil. Civil and Environmental Research, Vol. 12, No 6.

Phan, V. T-A. and Pham K. N-A. (2018). Improvement in engineering properties of soft-soil using cement and lime additives: A case study of southern Vietnam. IOP Conf. Ser.: Earth Environ. Sci., 143 012039.

Prusinski, J.R. and Bhattacharya, S. (1999). Effectiveness of Portland Cement and Lime in Stabilizing Clay Soils. Transportation Research Board, 1652(2), 198-206.

Rajasekaran, G., Murali, K. and Srinivasaraghavan, R. (1997). Fabric and mineralogical studies on lime treated marine clays. Ocean Engineering, 24(3), 227–234.

Rao, M.K., Giribabu, G. and Rani, S.Ch. (2006). Influence of Coarse Fraction on Swelling Characteristics. Electronic Journal of Geotechnical Engineering, No:0627.

Russo, G., Vecchio S.D., Mascolo G. (2007). Microstructure of a Lime Stabilised Compacted Silt. In: Schanz T. (eds) Experimental Unsaturated Soil Mechanics. Springer Proceedings in Physics, Vol 112. Springer, Berlin, Heidelberg.

Russo G., and Modoni G. (2013). Fabric changes induced by lime addition on a compacted alluvial soil. Géotechnique Letters, Vol 3, Issue 2, pp. 93-97.

Saeed, K. A., Kassim, K. A. and Nur, H. (2014). Physicochemical characterization of cement treated kaolin clay. GRAĐEVINAR, 66, 6, 513-521.

Sante, M.D., Fratalocchi, E., Mazzieri, F. and Pasqualini, E. (2014). Time of reactions in a lime treated clayey soil and influence of curing conditions on its microstructure and behavior. Applied Clay Science, 99, 100–109.

Sharo, A. A., Alhowaidi, Y. A., Al-Tawaha, M.S. (2019). Improving Properties of Expansive Soil Using Cement, Quick Lime and Cement-Lime Blend. International Review of Civil Engineering (IRECE).

Singh, J., Kumar, A., Jain, R., Khular, N.K. (2008). Effect of lime on properties of soil. 12th International conference Association for Computer Methods and Advances in Geomechanics.

Stoltz, G., Cuisinier, O. and Masrouri, F. (2012). Multi-scale analysis of the swelling and shrinkage of a lime-treated expansive clayey soil. Applied Clay Science, Vol 61, pp. 44-51

Studds, P.G., Stewart, D.I. and Cousens, T.W. (1998). The effects of salt solutions on the properties of bentonite-sand mixtures. Clay Minerals, 33, 651 – 660.

Tran T. D., Cui Y-J., Tang A. M., Audiguier M., Cojean R. Effects of lime treatment on the microstructure and hydraulic conductivity of Héricourt clay[J]., 2014, 6(5): 399-404.Wayal, A. S. Ameta, N. K. Purohit, D. G. M. (2012). Dune Sand Stabilization Using Bentonite and Lime. Journal of Engineering Research and Studies E-ISSN0976-7916, Vol III, Issue 1, January-March, 58-60.

Zambri, N. M. and Ghazaly Z. Md. (2018). Peat Soil Stabilization using Lime and Cement. E3S Web of Conferences 34, 01034.

Zhaoa, H., Zhoua, K., Zhaob C., Gongc, B-W. and Liuc, J. (2015). A long-term investigation on microstructure of cement-stabilized handan clay. European Journal of Environmental and Civil Engineering, http://dx.doi.org/10.1080/19648189.2015.1030087

High Entropy Composites

Necati KOÇAK[1]

1 Arş. Gör. Necati KOÇAK, Sakarya University of Applied Science, Depatment of Material and Metalurgical Engineering, Sakarya, Turkey, necatikocak@subu.edu.tr, ORCID ID: 0000-0002-0768-3479

INTRODUCTION

HEAs, regarded today as a new generation of high performance structural materials, were first introduced as a distinct class of materials in 2004. These materials typically contain five or more principal elements; each incorporated at roughly 5–35 at. % levels, resulting in the formation of solid-solution phases with equiatomic or near-equiatomic chemical proportions. Their exceptional behavior arises from several fundamental mechanisms most notably the high configurational entropy effect, pronounced lattice distortion, sluggish diffusion kinetics, and the so-called cocktail effect. As a result, HEAs exhibit thermodynamically stable microstructures and stand out with their remarkable wear resistance, outstanding low-temperature ductility, superparamagnetic behavior, excellent corrosion and oxidation resistance, high hardness, and superior structural and thermal stability. The emergence of the HEA concept through advances in metallurgy, alloy design and materials science has enabled the development of alloy systems with properties unattainable in conventional alloys, which traditionally rely on one or two base elements. As shown in Fig. 1, the development of second-generation high-entropy alloys—characterized by non-equimolar and multiphase solid-solution configurations—has emerged as a major research direction within the field. The continued evolution of the HEA concept (Fig. 1) has greatly contributed to the design and development of advanced materials. Rather than strictly limiting alloy design to the conventional definition of high entropy, it is increasingly essential to broaden this framework to address specialized performance requirements. Because of their equiatomic or near-equiatomic elemental ratios, high-entropy alloys exhibit substantially elevated configurational mixing entropy, which in turn suppresses the formation of ordered intermetallic compounds and promotes the stabilization of simple solid-solution phases. [1-5].

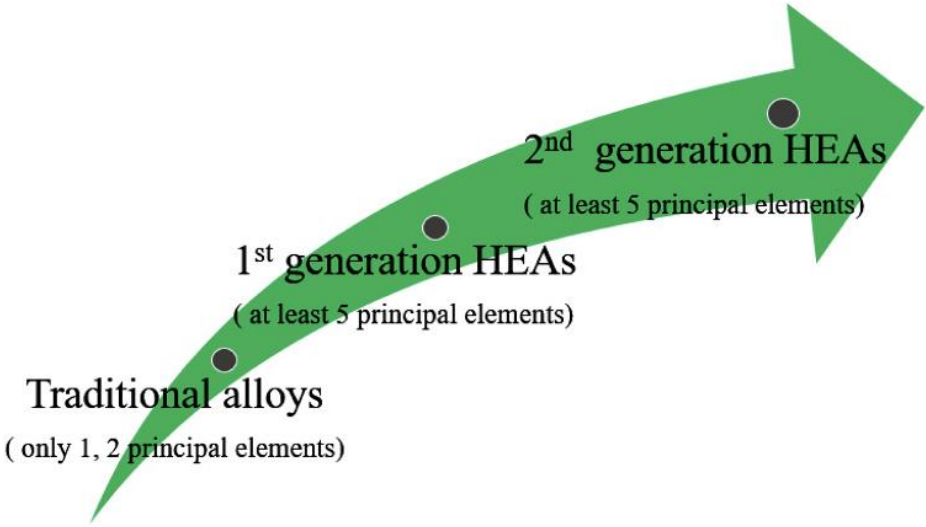


Figure 1. Evolutionary progression in alloy development [2]

Today, HEAs are intensively investigated for both structural applications such as refractory components, biomedical devices, and construction materials and functional applications, including electronic, magnetic, thermoelectric, and superconducting systems. Numerous compositions have already been designed that demonstrate high mechanical stability (hardness and strength) at both ambient and elevated temperatures, superior high-temperature performance, robust strength-ductility combinations, and resistance to wear and corrosion under extreme environments. Current literature indicates that the combinatorial design space involving approximately 37 potential principal elements has enabled researchers to synthesize and characterize over 600 unique HEA systems under laboratory conditions. High entropy materials (HEMs) have reshaped the paradigm of materials design by introducing multi-principal element systems capable of forming stable solid solution phases through high configurational entropy. Figure 2 presents a panoramic overview of the materials used by humankind over the past 10,000 years, illustrating the evolution of material classes ranging from ceramics and metals to polymers and composites. This illustration underscores the technological evolution that ultimately culminated in the development of HEAs

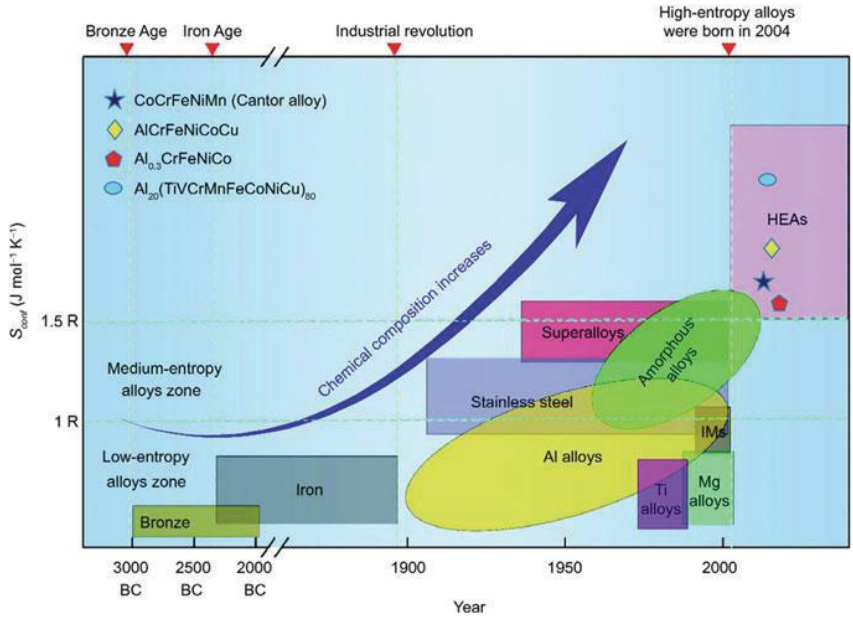


Figure 2. Evolution of material utilization over time and the emergence of Heas [12].

While early research focused primarily on Heas, recent developments have expanded this concept to ceramics, coatings, and more recently high entropy composites (HECs) a rapidly emerging class of materials that combines the entropy-stabilized matrix with engineered reinforcement phases. Unlike conventional composites, HECs may incorporate four or more elements in both the matrix and reinforcement phases, enabling unprecedented control over microstructure, interface stability, and multifunctional performance. Based on recent studies, the vast combinatorial space offered by approximately 37 metallic elements has enabled researchers to experimentally synthesize over six hundred distinct high-entropy alloy systems under laboratory conditions. This enhancement arises from several high entropy effects, including lattice distortion, sluggish diffusion, and cocktail effects, which influence deformation pathways and defect mobility in complex multi-element environments. In addition, high entropy ceramic based composites have shown notable improvements in oxidation resistance, toughness, and high-temperature stability, further expanding their technological relevance. The unique microstructural features of HECs such as multi element interphases, refined grain structures, and chemically complex secondary phases position them as attractive candidates for extreme environment applications. For example, recent studies highlight their promise as protective coatings for wear intensive and irradiation prone components, where high entropy architectures can effectively mitigate erosion, abrasion, and defect accumulation. Parallel advances in the broader field of high entropy materials continue to illuminate key mechanisms and

processing strategies that can be leveraged to tailor HECs for structural, functional, and energy-related applications. Despite these advances, the field faces significant challenges. Microstructural reproducibility, phase equilibria prediction, and processing property relationships remain areas of active investigation. The complexity of multi component systems often leads to competing phase formation pathways, making compositional and thermal design nontrivial. Moreover, the development of predictive computational and thermodynamic tools for HECs lags behind their alloy and ceramic counterparts. As a result, systematic studies are still needed to establish robust frameworks for processing control, standardization, and long-term performance assessment. Overall, high entropy composites represent a frontier in materials science that merges compositional complexity with composite design principles to create a new class of multifunctional materials. Their rapidly expanding application space and promising mechanical, chemical, and thermal properties make HECs an important subject for continued research and technological development [6-13].

CONCEPT AND DESIGN PRINCIPLES

1. Configurational Entropy and Phase Stability

HEAs were originally proposed as multi principal element systems in which four or more elements are combined in near equiatomic proportions, leading to a substantial increase in configurational (mixing) entropy compared with conventional alloys that are based on one or two main elements. For an ideal random solid solution containing N components with atomic fractions c_i, the configurational entropy can be expressed as

$$\Delta S_{mix} = -Ri=1\sum N\, c_i \ln c_i \quad (1.1)$$

where R is the gas constant. This expression shows that, for a given number of components, the configurational entropy is maximized when the composition is close to equiatomic. In many early HEA designs with five equiatomic elements, ΔS_{mix} reaches values on the order of 1.5–1.6 R, significantly higher than in dilute solid solutions. From a thermodynamic perspective, configurational entropy provides a significant stabilizing contribution by lowering the Gibbs free energy of mixing, which in turn promotes the formation and retention of solid-solution phases in high-entropy systems.

$$\Delta G_{mix} = \Delta H_{mix} - T\Delta S_{mix}, \quad (1.2)$$

and thus, can stabilise disordered solid solution phases against ordered intermetallic compounds, especially at elevated temperatures. When ΔS_{mix} is large and ΔH_{mix} is of moderate magnitude, the $T\Delta S_{mix}$ term can outweigh the enthalpic driving force for compound formation, favouring the formation of single phase FCC or BCC solid solutions rather than complex multiphase microstructures. This “entropy-stabilisation” picture was a key element in the original rationale of Yeh and co-workers and has remained central in the thermodynamic analysis of HEAs. However, configurational entropy alone does not guarantee the formation of single phase solid solutions. Systematic thermodynamic and experimental studies have shown that phase stability in HEAs and related complex concentrated alloys is governed by a competition between enthalpy and entropy, coupled with kinetic effects such as sluggish diffusion. For example, strongly negative mixing enthalpies between specific element pairs may still promote intermetallic formation, even in the presence of high ΔS_{mix}, whereas large positive ΔH_mix can favour phase separation or even amorphization. Consequently, modern design strategies typically combine entropic arguments with additional parameters such as atomic size mismatch (δ), valence electron concentration (VEC), and electronegativity differences to predict phase stability windows. Experimental evidence from equiatomic multicomponent systems, such as the well-known Cantor alloy (CoCrFeMnNi) and its derivatives, confirms that near-equiatomic compositions can indeed form stable single-phase FCC solid solutions over wide temperature ranges. Although some high-entropy compositions stabilize into single-phase solid solutions, numerous nominally high-entropy alloy systems instead develop complex multi-phase microstructures, consisting of coexisting solid-solution phases together with various intermetallic compounds. A critical reassessment of the field therefore emphasises that “high entropy” should be viewed as one factor among several that control phase selection, rather than a guarantee of single-phase behaviour. Configurational entropy remains a key thermodynamic parameter across the wider class of multi-principal element alloys (MPEAs), exerting a significant influence even in systems that deviate from equiatomic ratios yet retain substantial compositional complexity. Miracle and Senkov, for example, proposed a thermodynamic framework in which HEAs are treated as a subset of complex concentrated alloys, and showed that many existing commercial Ni- and Fe-based alloys already occupy intermediate entropy regimes. Figure 3 illustrates the main categories of HEAs.

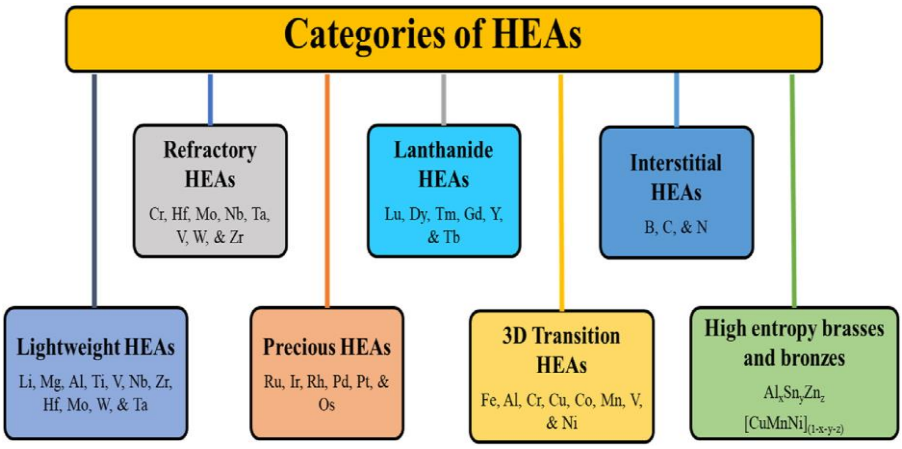


Figure 3. The main categories of HEAs [14]

In this view, the classical "high-entropy" label is less important than the underlying balance between ΔH_mix and ΔS_mix, especially when designing non-equimolar, application-specific compositions. Recent thermodynamic studies indicate that the influence of configurational entropy on phase stability is markedly stronger at elevated temperatures, where the TΔS term becomes the dominant contribution to the Gibbs free energy. However, as the temperature decreases, enthalpy-driven interactions and ordering tendencies gain greater significance, which may promote phase separation, ordering transformations, or decomposition during long-term exposure or ageing. This has important implications for the long-term microstructural stability of HEAs and related Hecs, particularly in structural and high-temperature service environments. In summary, configurational entropy is a central but not exclusive design lever in Hea systems. A high ΔS_mix can lower the Gibbs free energy and favour disordered solid-solution formation, but the actual phase stability results from a delicate interplay of enthalpic interactions, kinetic constraints, and compositional parameters such as atomic size mismatch and VEC. Understanding and quantitatively modelling this balance remains one of the key conceptual challenges in the design of next-generation Heas and high-entropy composites [14-19].

Heas (HEAs) have recently gained prominence as a unique family of advanced materials distinguished by their remarkable combination of properties and broad potential for technological applications. In 2006, Yeh and co-workers proposed a conceptual framework comprising four signature phenomena namely the high-entropy effect, sluggish diffusion, severe lattice distortion, and the cocktail effect to elucidate the underlying mechanisms that govern the behavior of HEAs and to clearly differentiate them from conventional alloy systems.These core effects form the theoretical

foundation for understanding the thermodynamic and kinetic stability of multi-principal element alloys. As depicted in Figure 4, the four effects are highly interdependent, working collectively to shape microstructural evolution, phase stability, and the optimization of functional and structural performance. The inherently high configurational entropy of HEAs drives these characteristic mechanisms lattice distortion, diffusion retardation, synergistic elemental interactions, and entropy-assisted stabilization thereby controlling alloy behavior across multiple length and time scales. Consequently, developing a clear and detailed understanding of each individual effect is essential before tailoring alloy chemistries or designing compositions for specific engineering applications [14, 20].

Figure 4. Schematic illustration of the four core effects high-entropy effect, sluggish diffusion, lattice distortion, and cocktail effect that collectively govern the thermodynamic stability and overall behavior of HEAs [21]

2. Element Selection and Design Criteria

Element selection in HECs requires careful consideration of thermodynamic, kinetic and microstructural factors to ensure the stability and performance of multi-principal element systems. Building on the core principles of HEA design, configurational entropy plays a foundational role: incorporating

multiple principal elements (typically ≥5) in near-equimolar or compositionally complex ratios increases the mixing entropy, which can suppress intermetallic formation and favor disordered solid-solution matrices. However, entropy alone is insufficient; phase stability emerges from the competition between configurational entropy and mixing enthalpy (ΔH_{mix}), meaning that element pairs with excessively positive or negative enthalpies may induce phase segregation or brittle intermetallics even in high-entropy systems. Atomic size mismatch (δ), electronegativity differences, and valence electron concentration (VEC) also strongly influence the selection of elements, as these parameters help predict FCC/BCC phase formation and interface stability within composite architectures. When developing high-entropy composites, an added level of complexity stems from the need to ensure chemical and structural compatibility between the high-entropy matrix and the selected reinforcement phase. Achieving coherent or semi-coherent interfaces is particularly important, as it minimizes undesirable interfacial reactions that could otherwise compromise mechanical integrity and overall composite performance. Recent computational and CALPHAD-assisted methods further demonstrate that multi-objective property optimization such as maximizing strength, thermal stability, wear resistance, or biocompatibility requires selecting element combinations that simultaneously satisfy entropic stabilization, controlled enthalpy, and application-specific functional requirements. Consequently, effective HEC design integrates entropy driven compositional strategies with rigorous thermodynamic modeling and interface engineering to achieve structurally stable and high-performance composite materials [16-19].

PROCESSING ROUTES

1. Powder Metallurgy

Powder metallurgy (PM) has increasingly been recognized as an adaptable and efficient manufacturing route for HECs, largely because it enables the processing of multi principal element powders together with a wide range of reinforcement phases while ensuring precise control over the resulting chemical composition. This method facilitates homogeneous mixing, minimizes elemental volatilization, and allows tailored consolidation conditions, making it particularly well suited for producing compositionally complex composite systems with stable microstructures and enhanced

performance. Conventional casting-based methods often struggle with elemental segregation, volatilization, and high melting-point disparities problems that are significantly mitigated or entirely eliminated in PM-based processing. Owing to these advantages, PM has been widely adopted in the synthesis of Heas (HEAs) and, more recently, in composite systems incorporating ceramic, intermetallic, or metallic reinforcements. Mechanical alloying (MA) serves as the foundational step of the PM route for HECs. During intense high-energy ball milling, repeated cold welding, fracturing, and rewelding events promote atomic-scale mixing among constituent high-entropy elements, enabling the formation of supersaturated solid solutions even when the equilibrium phase diagram predicts multi-phase structures. This non equilibrium processing further fosters fine-grained or nanocrystalline structures that can be consolidated without significant grain growth. The inclusion of ceramic particles such as TiC, WC, B_4C, SiC, or Al_2O_3 during milling promotes uniform dispersion and strong mechanical bonding with the HEA matrix, yielding composite powders with refined interfaces. Powder metallurgy (PM) offers a solid-state processing pathway for the fabrication of Heas (HEAs), enabling more precise regulation of chemical composition and microstructural homogeneity than conventional melt-based approaches (Fig. 5). In this method, elemental or pre-alloyed powders corresponding to the target alloy composition are mechanically mixed and subsequently compacted into a predetermined geometry. The compacted powders are then subjected to sintering at elevated temperatures typically maintained below the melting points of the constituent elements to promote diffusion driven bonding and densification, ultimately yielding a consolidated high-density alloy structure [21-24].

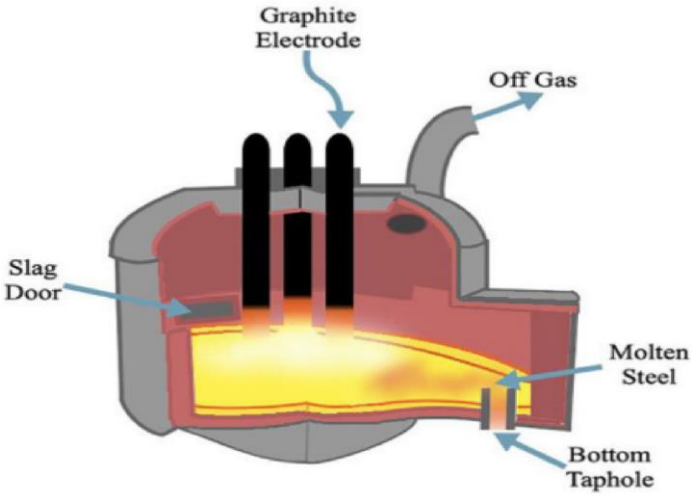


Figure 5. Schematic illustration of an electric arc furnace [25]

Spark plasma sintering (SPS) is widely recognized as the most effective consolidation technique for HEA- and HEC-based powders, owing to its exceptionally high heating rates, short holding times, and ability to preserve the nanostructured or metastable phases produced during mechanical alloying. The SPS process typically yields compacts with densities approaching theoretical values and very low porosity, while the application of pulsed direct current enhances interparticle bonding and decreases diffusion-related limitations. The schematic configuration and working principles of the SPS setup used for HEA-based material consolidation are presented in Figure 6 [21-24].

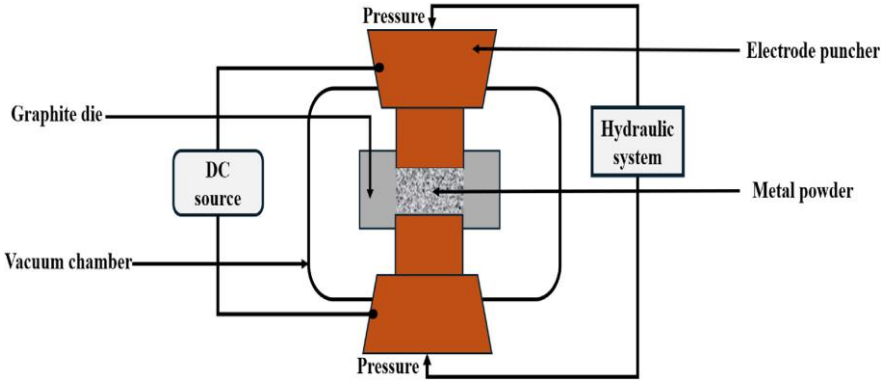


Figure 6. Schematic illustration of the SPS process [21]

Hot pressing (HP) and hot isostatic pressing (HIP) are also used for HEC fabrication, particularly when larger bulk components are required; however, these techniques may lead to more pronounced grain growth compared to SPS. The microstructure of PM-processed HECs is strongly influenced by milling duration, reinforcement nature, and consolidation parameters. SPS-fabricated composites often exhibit ultrafine-grained or nanocrystalline HEA matrices surrounding uniformly dispersed reinforcements. Carbide reinforced HECs typically display coherent or semi-coherent interfaces, which contribute to load transfer strengthening, Orowan looping, and grain boundary pinning. PM routes also minimize elemental segregation and suppress the formation of brittle intermetallic phases, ensuring improved ductility relative to cast HECs. PM-based HECs demonstrate superior hardness, compressive strength, and thermal stability compared to monolithic HEAs or conventional MMCs. The uniform dispersion of reinforcements in an entropy-stabilized matrix promotes multiple strengthening mechanisms, including grain refinement, dislocation hardening, and dispersion strengthening. Ceramic-reinforced HECs frequently achieve hardness levels exceeding 1,000 HV and high-temperature strength retention due to the refractory nature of the HEA

matrix. Additionally, PM processing enables the design of HECs with tailored thermal conductivity, wear resistance, oxidation resistance, and, in some cases, enhanced magnetic or catalytic performance depending on the reinforcement type. The principal advantages of PM include its flexibility in producing compositions difficult to obtain via melting routes, refined microstructures, and strong matrix–reinforcement interfaces. Nevertheless, challenges remain in achieving complete homogenization during milling, controlling oxygen contamination, and scaling the process for industrial production. Despite these limitations, PM remains the most powerful and adaptable method for fabricating HECs with advanced property combinations [22,23].

2. Melting and Casting

Melting-based fabrication routes remain foundational in the synthesis of Heas (HEAs) and by extension, high-entropy composites (HECs), by facilitating full liquid-phase mixing of multi-principal element systems prior to solidification. Conventional methods such as vacuum arc melting (VAM), induction melting, and laser melting allow constituent elements and reinforcements to be fully dissolved or suspended in the melt, thereby promoting homogeneous chemistry at the macro-scale and enabling large-scale casting of composite ingots. In the context of HECs, such melting techniques permit the incorporation of ceramic or metallic reinforcements into the molten high-entropy matrix, followed by solidification to achieve near-net shape components. However, the rapid solidification inherent to many of these methods may impose limitations: chemical segregation, macro-segregation, porosity entrapment, and dendritic growth are frequently observed, which can degrade interface quality and mechanical performance (Ren et al., 2025). Furthermore, when dealing with HECs containing refractory elements and high-melting-point phases, the required melting temperatures and associated thermal gradients can lead to excessive residual stresses and crack formation. Therefore, while melting-based methods offer advantages of scale and liquid-mixing completeness, they demand careful control of cooling rates, stirring, inert atmospheres, and post-solidification treatments to realize high-performance HECs. Recent reviews indicate that for HEC production, hybrid strategies combining melting with powder consolidation or additive manufacturing may mitigate some of these drawbacks. Consequently, melting-based methods continue to act as a primary route for HEC fabrication, but their success depends on overcoming the intrinsic processing challenges associated with multi-component, multi-phase alloys [25, 26].

2.1. Vacuum arc melting

Vacuum arc melting is one of the most widely employed routes for synthesizing Heas (HEAs), owing to its ability to efficiently melt multiple metallic constituents under controlled conditions. In this method, the elemental charge is melted within an electric arc furnace operating in an inert atmosphere—most commonly argon—to suppress oxidation and external contamination. The arc generated within the furnace can reach temperatures well above 3000 °C, allowing complete melting of the majority of metallic elements used in HEA formulations. This technique provides effective mixing and facilitates the production of chemically homogeneous, high-purity alloys by permitting precise regulation of arc power and melting parameters. Despite these advantages, vacuum arc melting presents limitations when processing elements with relatively low boiling points, such as Mg, Zn, and Mn, which tend to volatilize during melting and thereby complicate compositional control. For systems containing such elements, alternative heating strategies like resistance or induction melting are often preferred, as they help minimize evaporation losses and enhance compositional uniformity. The schematic layout of a typical VAM unit is illustrated in Figure 7 In this setup, the metallic charge is placed in a water-cooled copper crucible, while a non-consumable tungsten electrode serves as the power source. [27–30].

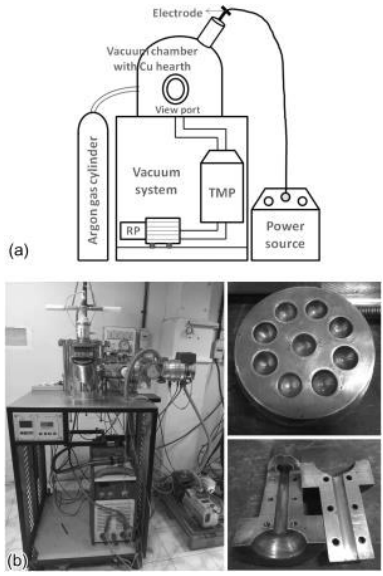


Figure 7. (a) Schematic of the vacuum arc melting unit and (b) photograph of the arc-melting and suction-casting setup with a copper crucible and two copper molds [30]

2.2. Vacuum ınduction melting (vım)

Vacuum induction melting (VIM) is an established process for melting and casting metallic alloys under either vacuum or inert atmospheric conditions, using electromagnetic induction as the primary heat source. In this technique, an alternating current passing through the induction coil generates a fluctuating magnetic field, which in turn induces eddy currents within the metallic charge; the resistive heating from these currents enables the material to reach its melting temperature. A typical VIM system includes a refractory-lined crucible housed within a steel chamber connected to a high-capacity vacuum system to limit oxidation and contamination. Heating occurs through the induction coils surrounding the crucible, and once the charge is fully molten, the magnetic field also promotes electromagnetic stirring, thereby enhancing chemical uniformity throughout the melt. The optimal operating frequency depends on variables such as the dimensions, geometry, and physical state of the charge. Unlike earlier VIM units that functioned at a single fixed frequency, modern systems are equipped with variable-frequency capabilities, allowing more efficient heating and improved control over the melting process [31, 32].

2.3. Laser cladding

Laser cladding is a high precision surface engineering technique that employs a concentrated laser beam to supply controlled thermal energy to a confined region of the substrate. Owing to its localized heat input, the process significantly limits the size of the heat affected zone (HAZ), thereby minimizing thermal distortion, porosity formation, and cracking. In addition, the rapid melting and solidification that occur during processing facilitate the formation of a robust metallurgical interface between the deposited layer and the underlying material, offering advantages over conventional thermal spray methods. A general schematic of the laser cladding configuration is presented in Figure 8. The system typically includes a laser source, optical elements responsible for directing and focusing the beam, a powder feeding mechanism, and a motion-controlled stage for manipulating the workpiece. In most setups, the optical assembly and laser head remain stationary while the substrate moves relative to the beam to maintain consistent energy delivery. Modern laser cladding units operate under computer control, enabling precise adjustment of critical parameters such as laser power, traverse speed, and powder feed rate. As the laser generates a transient melt pool on the surface, the injected powder particles melt and rapidly solidify upon passing through the beam path, forming dense and strongly adherent coatings. Depending on application needs, individual tracks or multilayer overlapping deposits can be produced to achieve uniform coverage and enhanced surface durability. With its high degree of controllability, compatibility with a wide range of geometries, and capability to generate

high quality protective layers, laser cladding has become a widely adopted and scalable method in contemporary industrial practice [19].

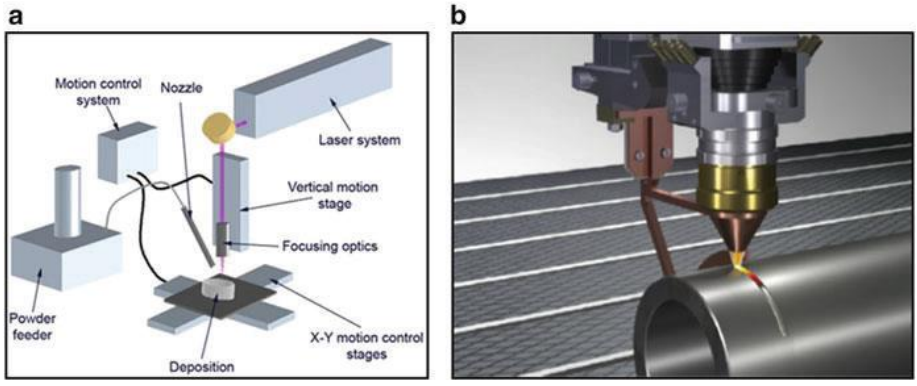


Figure 8. (a) Schematic diagram and (b) operational view of the laser cladding process [19]

3. Additive Manufacturing

Additive manufacturing (AM) techniques such as laser powder bed fusion (LPBF) and directed energy deposition (DED) have proven to be transformative in the fabrication of Heas (HEAs) and by extension high-entropy composites (HECs), because they enable fine-scale control over thermal histories, solidification kinetics, and microstructure formation. LPBF exploits rapid melting and solidification of powder layers under a focused laser beam, which fosters ultrafine grain structures, suppressed segregation, and a refined network of dislocations leading to superior mechanical properties compared to conventional casting. On the other hand, DED, with its higher deposition rates and ability to build large, directed structures, provides flexibility in reinforcement placement and gradient manufacturing in HEC systems. Importantly, both LPBF and DED allow modulation of processing parameters—such as laser power, scanning speed, hatch spacing, and layer thickness—that directly influence melt pool dynamics, cooling rates, and thus phase selection, porosity formation, and microstructural heterogeneity. Through this level of process–structure control, AM offers a unique pathway to tailor HECs with customized microstructures such as heterogenous phase distributions or engineered interfaces that enhance load transfer in composites and optimize performance for structural applications in aerospace, automotive and energy sectors. The DED process is shown in Fig. 9 [4, 33, 34].

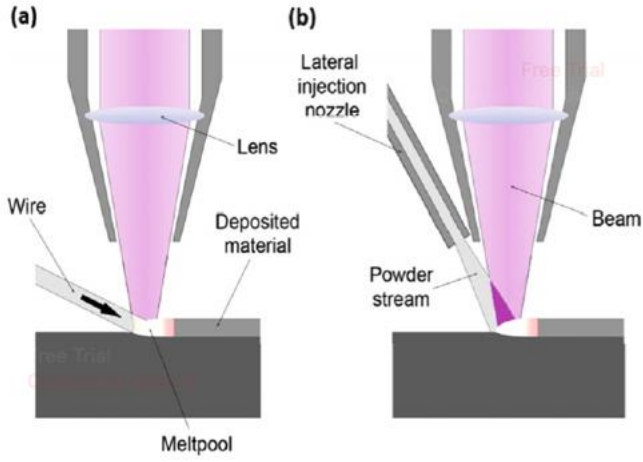


Figure 9. DED process [4]

The powder bed fusion (PBF) technique operates by directing a focused energy source typically a laser or an electron beam onto a thin layer of powder, selectively melting the material and consolidating it layer by layer on the build platform. Although the method is widely employed for processing metallic systems such as titanium-based alloys, it is also commonly utilized for fabricating polymeric components, including nylon-based materials. The powder bed fusion process is shown in Fig. 10 [4].

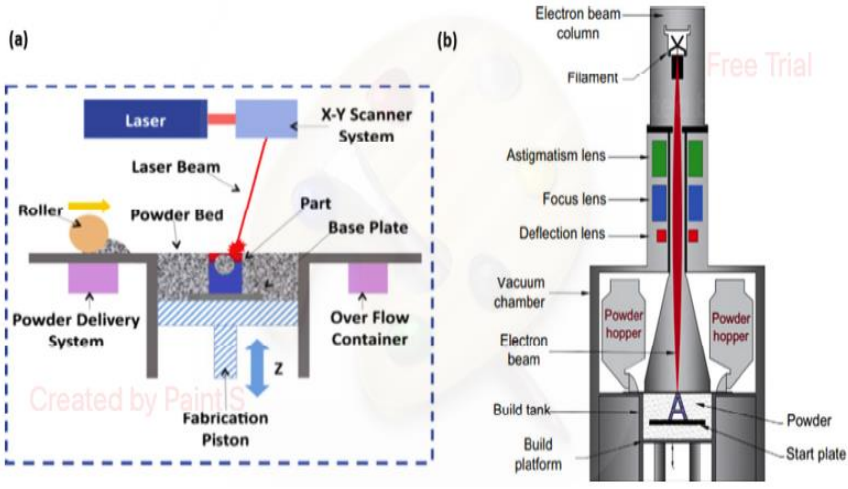


Figure 10. PBS process [4]

MICROSTRUCTURE AND PROPERTIES

1. Phase Formation

Microstructural development in HECs critically depends on the interplay between the entropy stabilised matrix and the reinforcing phases, with structural uniformity, interface integrity, and phase distribution being key determinants of performance. Studies have revealed that HECs fabricated through powder metallurgy or melting based routes often exhibit finely dispersed reinforcement particles within a multiprincipal element solid solution matrix, where suppressed intermetallic formation and refined grain structures are observed due to high mixing entropy and lattice distortion effects. Interfacial characteristics further dictate mechanical behaviour: for instance, well bonded particulate reinforcements in aluminium-matrix composites with embedded HEA particles demonstrated minimal porosity and negligible secondary phase formation while achieving enhanced hardness and corrosion resistance. Moreover, recent investigations indicate that heterogeneous microstructures engineered via additive manufacturing or tailored consolidation pathways can exploit the sluggish diffusion phenomena inherent to high entropy systems, thereby restricting grain growth and enhancing stability under thermal exposure. The microstructural evolution of $CoCrFeNiMo_{0.2}$ reinforced titanium matrix composites was shown to be strongly dependent on sintering temperature. At 700 °C, the HEA particles retain a spherical morphology and exhibit a clean, well-bonded interface with the Ti matrix. As the temperature increases to 800–850 °C, a diffusion layer forms and progressively thickens, accompanied by noticeable irregularity on particle surfaces due to elemental interdiffusion. At 900 °C, extensive dissolution and significant shape degradation of the HEA particles occurs, indicating severe high temperature diffusion effects. The microstructural evolution of the composites as a function of sintering temperature is presented in Fig. 11, where significant changes in particle morphology and interface characteristics can be observed across the examined temperature range (700–900 °C) [2, 35, 36, 37].

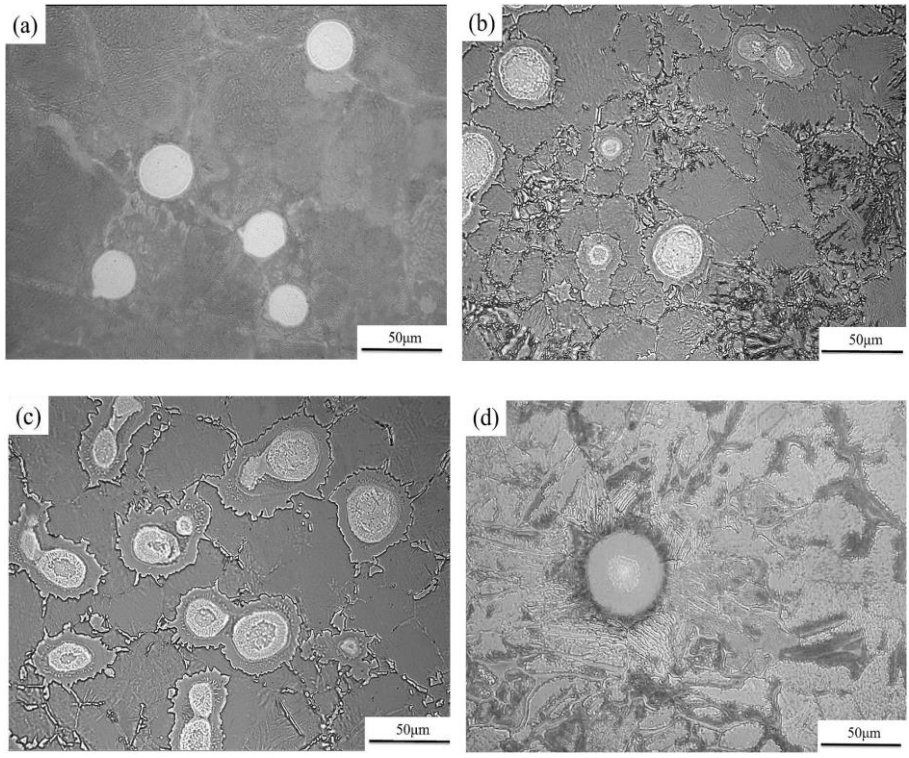


Fig 11. Microstructural evolution of the composites at different sintering temperatures: (a) 700 °C, (b) 800 °C, (c) 850 °C, and (d) 900 °C [37]

APPLICATIONS

HECs, which combine multi principal element matrices with reinforcing phases, are garnering increasing attention for advanced applications owing to their superior structural and functional performance. For instance, in metal matrix composite systems, Hea particles dispersed within titanium or aluminium matrices have shown enhanced mechanical strength and wear resistance, making them promising candidates for aerospace structural components and heavy-duty tooling. Current progress in aerospace engineering therefore emphasizes the design and adoption of structural materials that not only surpass existing alloys in performance but are also cost-effective and capable of maintaining reliability under severe operating

conditions. Material selection is closely tied to the thermal and mechanical environment encountered in service, as the components must tolerate high temperatures while remaining as lightweight as possible. Consequently, aero-engine construction traditionally incorporates steels, nickel-based superalloys, titanium alloys, and Al-based alloys, while more recently HEAs have emerged as promising candidates for future high temperature aerospace applications Meanwhile, HECs processed by additive manufacturing have been explored for their potential in customized, complex geometries for automotive and aerospace sectors, as the entropy stabilised composites offer improved microstructural stability under cyclic loading and high temperatures. In energy and surface engineering contexts, ceramic reinforced HEA matrices have been demonstrated to resist oxidation and corrosive environments more effectively than traditional alloys, pointing to applications in turbines, high temperature heat exchangers and chemical processing equipment. Across these sectors, the multi element design flexibility inherent to HECs enables tailored combinations of hardness, ductility, thermal conductivity, and corrosion resistance opening pathways to multifunctional composites that address the demanding requirements of next-generation engineering systems [2, 4, 23, 35].

REFERENCES

[1] Feltrin, A. C., Xing, Q., Akinwekomi, A. D., Waseem, O. A., & Akhtar, F. (2022). Review of novel high-entropy protective materials: wear, irradiation, and erosion resistance properties. Entropy, 25(1), 73.

[2] Kumar, D., Seetharam, R., & Ponappa, K. (2024). A review on microstructures, mechanical properties and processing of high entropy alloys reinforced composite materials. Journal of Alloys and Compounds, 972, 172732.

[3] Liaw, P. K., & Li, W. (2021). High entropy materials: challenges and prospects. Metals, 11(10), 1643.

[4] Osman, H., & Lin, L. I. U. (2023). Additive manufacturing of Hea composites: a review. Transactions of Nonferrous Metals Society of China, 33(1), 1-24.

[5] Zhao, H., Sun, W., Lu, Y., Song, W., Liu, H., & Li, M. (2025). Research status and prospects of high-entropy composite ceramics. Journal of Asian Ceramic Societies, 1-17.

[6] Melih Yetiş, Kendiliğinden İlerleyen Yüksek Sıcaklık Yanma Sentezi İle Üretilen AlCoxCrFeNi Yüksek Entropili Alaşım Sistemindeki Kobalt Oranı

Değişiminin İncelenmesi, 2019 Fen Bilimleri ve Matematikte Güncel Araştırmalar .163

[7] Sarswat, P. K., Sarkar, S., Murali, A., Huang, W., Tan, W., & Free, M. L. (2019). Additive manufactured new hybrid high entropy alloys derived from the AlCoFeNiSmTiVZr system. Applied Surface Science, 476, 242-258.

[8] Sharma, A. S., Yadav, S., Biswas, K., & Basu, B. (2018). Heas and metallic nanocomposites: Processing challenges, microstructure development and property enhancement. Materials Science and Engineering: R: Reports, 131, 1-42.

[9] Stolze, K., Cevallos, F. A., Kong, T., & Cava, R. J. (2018). Hea superconductors on an α-Mn lattice. Journal of Materials Chemistry C, 6(39), 10441-10449.

[10] Zhang, Y., Zuo, T. T., Tang, Z., Gao, M. C., Dahmen, K. A., Liaw, P. K., & Lu, Z. P. (2014). Microstructures and properties of Heas. Progress in materials science, 61, 1-93.

[11] Li, Z., Zhao, S., Ritchie, R. O., & Meyers, M. A. (2019). Mechanical properties of Heas with emphasis on face-centered cubic alloys. Progress in Materials Science, 102, 296-345.

[12] Senkov, O. N., Wilks, G. B., Miracle, D. B., Chuang, C. P., & Liaw, P. K. (2010). Refractory Heas. Intermetallics, 18(9), 1758-1765.

[13] Wanqing Dong, Zheng Zhou, Mengdi Zhang, Yimo Ma, Pengfei Yu, Peter K. Liaw and Gong Li, Applications of High-Pressure Technology for High-Entropy Alloys: A Review, 2019

[14] Kamal, M. V., Ragunath, S., Reddy, M. H. S., Radhika, N., & Saleh, B. (2024). Recent advancements in lightweight high entropy alloys–A comprehensive review. International Journal of Lightweight Materials and Manufacture, 7(5), 699-720.

[15] Cantor, B., Chang, I. T., Knight, P., & Vincent, A. J. B. (2004). Microstructural development in equiatomic multicomponent alloys. Materials Science and Engineering: A, 375, 213-218.

[16] George, E. P., Raabe, D., & Ritchie, R. O. (2019). Heas. Nature reviews materials, 4(8), 515-534.

[17] Miracle, D. B., & Senkov, O. N. (2017). A critical review of high entropy alloys and related concepts. Acta materialia, 122, 448-511.

[18] Yeh, J. W., Chen, S. K., Lin, S. J., Gan, J. Y., Chin, T. S., Shun, T. T., ... & Chang, S. Y. (2004). Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. Advanced engineering materials, 6(5), 299-303.

[19] Gao, M. C., Yeh, J. W., Liaw, P. K., & Zhang, Y. (Eds.). (2016). Heas: fundamentals and applications. Springer.

[20] Odetola, P. I., Babalola, B. J., Afolabi, A. E., Anamu, U. S., Olorundaisi, E., Umba, M. C., ... & Olubambi, P. A. (2024). Exploring high entropy alloys: a review on thermodynamic design and computational modeling strategies for advanced materials applications. Heliyon, 10(22).

[21] Salifu, S., & Olubambi, P. A. (2025). High entropy alloy reinforced lightweight metal matrix composites: A review of the fundamentals, fabrication, properties, and prospects. Materials Today Sustainability, 101216.

[22] Gorsse, S., Miracle, D. B., & Senkov, O. N. (2017). Mapping the world of complex concentrated alloys. Acta Materialia, 135, 177-187.
[23] Mehmood, M. A., Shehzad, K., Mujahid, M., Yaqub, T. B., Godfrey, A., Fernandes, F., ... & Yaqoob, K. (2022). Ceramic-reinforced HEA matrix composites exhibiting an excellent combination of mechanical properties. Scientific Reports, 12(1), 21486.
[24] Olorundaisi, E., & Olubambi, P. A. (2025). The prospect and limitation of high entropy alloy as 4th industrial material. Materials Today Sustainability, 31, 101163.
[25] Salam, M. Y. A., et al. (2025). Effect of fabrication techniques of high entropy alloys. Materials Science and Engineering: A, advance online publication.
[26] Derviş ÖZKAN, C. K. (2021). High entropy alloys: Production, properties and utilization areas.
[27] Olorundaisi, E., & Olubambi, P. A. (2025). The prospect and limitation of high entropy alloy as 4th industrial material. Materials Today Sustainability, 31, 101163.
[28] B.S. Murty, J.W. Yeh, S. Ranganathan, Heas, 2014
[29] K. Li, W. Chen, Recent progress in Heas for catalysts: synthesis, applications, and prospects, 2021
[30] T.S. Srivatsan and Manoj Gupta, High Entropy Alloys Innovations, Advances, and Applications, 2020
[31] Simbarashe Fashu, Mykhaylo Lototskyy, Moegamat Wafeeq David, Lydia Pickering, Vladimir Linkov, Sun Tai, Tang Renheng, Xiao Fangming, Pavel V. Fursikov, Boris P. Tarasov, A review on crucibles for induction melting of titanium alloys, 2020
[32] ASM Handbook, Volume 15: Casting Volume 15 Handbook Committee, p 1-8, 2008
[33] Oketola, A. M., Adegbola, T. A., Jamiru, T., Ogunbiyi, O., & Salifu, S. (2025). Advances in Hea Research: Unraveling Fabrication Techniques, Microstructural Transformations, and Mechanical Properties. Journal of Bio-and Tribo-Corrosion, 11(3), 79.
[34] Zhang, W., Chabok, A., Kooi, B. J., & Pei, Y. (2022). Additive manufactured high entropy alloys: A review of the microstructure and properties. Materials & Design, 220, 110875.
[35] Kurdi, A., Zaman, A., Alsolami, A., Alshabouna, F., Degnah, A., Alfihed, S., ... & Tabbakh, T. (2025). Effect of manufacturing route on microstructure and micromechanical properties of AlCoCrFeNi high entropy alloy. Journal of Alloys and Compounds, 1010, 177477.
[36] Qiao, K., Guo, X., Zhang, X. et al. Microstructure evolution and mechanical properties of high entropy alloy reinforced titanium matrix composites processed by cold spray-friction stir processing composite additive manufacturing. Adv Compos Hybrid Mater (2025).
[37] Yuan, Z., Liu, H., Ma, Z., Ma, X., Wang, K., & Zhang, X. (2022). Microstructure and properties of high entropy alloy reinforced titanium matrix composites. Materials Characterization, 187, 111856.

Lightweight High Entropy Alloy

Necati KOÇAK[1]

1 Arş. Gör. Necati KOÇAK, Sakarya University of Applied Science, Depatment of Material and Metalurgical Engineering, Sakarya, Turkey, necatikocak@subu.edu.tr, ORCID ID: 0000-0002-0768-3479

INTRODUCTION

HEAs have emerged as a transformative concept in materials science, challenging the traditional alloy design paradigm based on one principal element. HEAs are considered a new generation of high-performance materials and were first introduced as a distinct class of alloys in 2004. These alloys are typically composed of five or more principal elements, each present in concentrations ranging from 5% to 35%, combined in nearly equiatomic or equiatomic proportions to form a stable solid-solution phase. Due to phenomena such as the high-entropy effect, lattice distortion, sluggish diffusion, and the cocktail effect, HEAs exhibit a highly stable microstructure. This unique structural stability endows them with exceptional mechanical and functional properties, including high wear resistance, excellent low-temperature ductility, superior magnetic and mechanical strength, high hardness, outstanding corrosion and oxidation resistance, and remarkable structural and thermal stability. The concept of Heas, which has emerged through extensive research in metallurgy and materials science, represents a paradigm shift from conventional alloy design. In contrast to conventional alloys composed primarily of one or two base elements, Heas (HEAs) attain greater configurational entropy through equiatomic or near-equiatomic compositions, which promote the formation of stable solid-solution phases while suppressing intermetallic compound formation. Consequently, this enables the development of alloy systems with properties unattainable in conventional materials. At present, Heas are being intensively explored as promising candidates for both structural applications (e.g., refractory, biomedical, and construction materials) and functional applications (e.g., electronic, magnetic, thermoelectric, and superconducting materials). These systems exhibit high mechanical stability and strength at both ambient and elevated temperatures, excellent thermal stability, superior tensile properties, and exceptional resistance to wear and corrosion under extreme environments. Through the combinatorial design of approximately 37 principal elements, more than 600 distinct HEA systems have been synthesized worldwide under laboratory conditions, providing a vast compositional landscape for developing novel materials with outstanding performance characteristics[1-11].

As the number of studies on HEAs has significantly increased, the initial definitions and classification criteria have been gradually revised. The strict requirements regarding the number of constituent elements and their compositional ratios have become more flexible over time. At present, even quaternary equimolar and non-equimolar alloy systems are often categorized as HEAs. Furthermore, some recent publications consider the presence of minor intermetallic compounds within HEA systems to be acceptable. Figure 1. presents a panoramic overview of the materials used by humankind over

the past 10,000 years, illustrating the evolution of material classes ranging from ceramics and metals to polymers and composites. This illustration underscores the technological evolution that ultimately culminated in the development of HEAs[12].

In recent years, lightweight high-entropy alloys (LHEAs) have attracted growing scientific interest due to their unique combination of low density, high specific strength, and excellent corrosion resistance. These alloys incorporate lightweight elements such as aluminum (Al), titanium (Ti), magnesium (Mg), lithium (Li), and beryllium (Be), resulting in densities typically ranging from 3 to 6 g/cm³, while maintaining outstanding mechanical and thermal stability. The pursuit of lightweight materials has become a central theme in modern engineering and technology throughout the twenty-first century. Their reduced mass offers clear advantages for aerospace and transportation applications, where minimizing weight directly contributes to improved fuel efficiency and reduced emissions. Traditionally, Al and its alloys have been the most widely utilized lightweight structural materials in the automotive and aerospace sectors owing to their favorable strength-to-weight ratio, corrosion and oxidation resistance, thermal stability, and economic viability. However, despite these merits, Al-based alloys exhibit certain inherent limitations, including reduced mechanical strength at elevated temperatures and restricted formability in specific alloy systems. Magnesium alloys, another prominent class of lightweight metals, compete with aluminum-based systems due to their even lower density and offer additional advantages such as high specific strength and superior damping capacity.

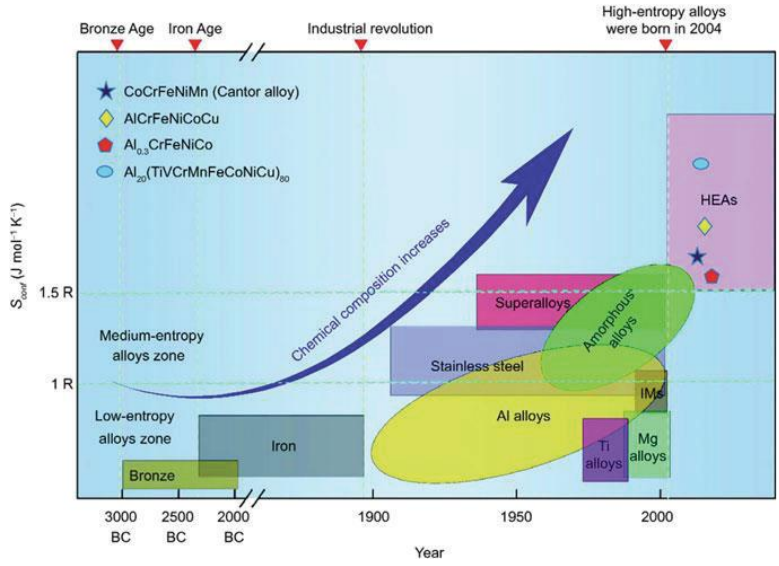


Figure 1 Evolution of material utilization over time and the emergence of Heas [2].

Despite their advantages, magnesium (Mg) alloys still face significant challenges, including low tensile strength, limited ductility, inadequate corrosion resistance, and relatively high processing costs. Although both aluminum- and magnesium-based alloys fulfill many structural requirements, ongoing industrial demand persists for materials that exhibit superior mechanical performance—notably enhanced strength, hardness, wear, and corrosion resistance—without compromising weight efficiency, cost, or manufacturability. The pursuit of lightweight materials for industrial applications therefore remains a major research priority. Progress in sectors such as aerospace, automotive, and energy is increasingly dependent on the development of innovative alloy design strategies capable of replacing high-density materials while maintaining or improving their mechanical and functional characteristics. Weight reduction continues to be recognized as one of the most cost-effective approaches to improving fuel economy and reducing carbon emissions in transportation systems. Nevertheless, currently available lightweight metals often involve inherent trade-offs, such as restricted mechanical performance or high production costs. Consequently, researchers are exploring new alloy systems that integrate low density with high mechanical and thermal stability, while also remaining economically sustainable. Over the past decade, substantial research efforts have focused on replacing conventional structural materials including steel, iron, titanium (Ti) alloys, aluminum, magnesium, and NiTi-based composites with more efficient, lightweight alternatives. Alloys containing titanium or beryllium exhibit excellent specific strength and are promising for aerospace structural components, yet their high fabrication costs restrict large-scale utilization.

Meanwhile, aluminum alloys continue to serve as key materials in lightweight design for construction, automotive, and aerospace engineering, although their mechanical strength significantly deteriorates at elevated temperatures.For example, the Al7075 alloy retains only about 30 % of its room-temperature strength at 200 °C and roughly 10 % at 300 °C. Recent progress in Lightweight Heas (LWHEAs) has addressed many of these challenges by tuning the atomic ratios of constituent elements rather than relying solely on conventional alloying approaches. LWHEAs demonstrate superior hardness, strength, corrosion resistance, and thermal stability compared to classical lightweight alloys. Such properties make LWHEAs promising candidates for transportation, energy, aerospace, and aviation applications. Common elements used in LWHEA design include Al, Mg, Ti, Li, Fe, Mn, Co, and Ni. Although research on LWHEAs is still in its early stages, interest in this field is rapidly growing. The definition of what constitutes a "lightweight" alloy remains under discussion; however, most studies classify alloys with densities below 7 g/cm^3 as lightweight. Some authors suggest even stricter limits (e.g., $\leq$ 5 g/cm^3 or $\leq$ 3 g/cm^3), though only a few alloys have been successfully fabricated within this range. Overall, LWHEAs are expected to play a pivotal role in shaping the future of advanced materials. Their unique combination of low density, superior mechanical performance, and compositional tunability offers transformative opportunities for multiple sectors. In aerospace engineering, LWHEAs can significantly reduce the weight of aircraft components without compromising structural integrity or safety, thereby improving fuel efficiency. Moreover, their potential applications extend to renewable energy systems, where their lightweight and durable nature can enhance the operational efficiency of wind turbines, solar panels, and other sustainable technologies. As research on LWHEAs progresses, these alloys are poised to make substantial contributions to sustainability, performance, and energy efficiency across diverse industries, paving the way toward a more sustainable and technologically advanced future. Figure 2 illustrates the main categories of HEAs[13,14].

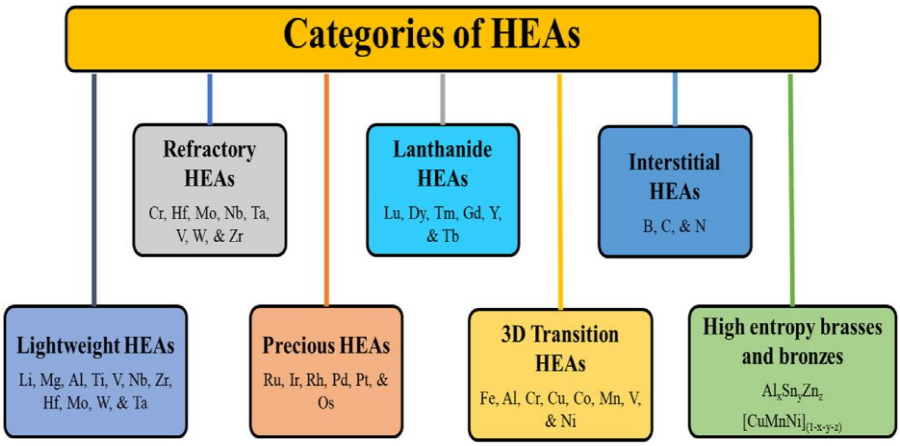


Figure 2. The main categories of HEAs [13]

THEORETICAL BACKGROUND AND ALLOY DESIGN PRINCIPLES

1. The High-Entropy Concept

It has been established that not all alloys containing five principal elements in equiatomic proportions are capable of forming stable high-entropy alloy (HEA) solid solutions. The stabilization of body-centered cubic (BCC) or face-centered cubic (FCC) solid-solution phases requires carefully optimized elemental compositions rather than simple equiatomic mixing. To date, no HEAs have been experimentally confirmed to exhibit a hexagonal close-packed (HCP) crystal structure. This absence is likely attributed to the fact that HCP configurations are typically stable only for pure elements at low temperatures and tend to transform into FCC or BCC structures at elevated temperatures. Most of the solid-solution HEAs reported thus far have been developed through trial-and-error experimental methodologies, as phase diagrams for quaternary and higher-order systems remain largely incomplete or unavailable. Consequently, empirical experimentation continues to serve as the primary approach for designing and optimizing high-performance HEAs. To assist in predicting phase formation behavior, researchers have introduced thermodynamic and empirical parameters analogous to the Hume–Rothery rules used for conventional solid-solution alloys. In multicomponent alloy systems, a high degree of configurational entropy plays a critical role in promoting the formation and stability of simple solid-solution phases at elevated temperatures by reducing the overall Gibbs free energy. As temperature increases, the entropic contribution ($-T\Delta S$) becomes increasingly dominant, thereby favoring the formation of disordered solid solutions over ordered intermetallic compounds or phase-separated structures. The most thermodynamically stable phase under given conditions

is ultimately determined by the minimization of Gibbs free energy, as expressed in Equation (1.1).This fundamental principle enables the prediction of phase stability, transformations, and equilibrium states in multicomponent alloys, providing a theoretical foundation for the rational design of HEAs with tailored properties. A comprehensive understanding of these thermodynamic concepts is therefore essential for optimizing processing parameters and accurately forecasting alloy behavior under diverse environmental and thermal conditions.

$$\Delta G=\Delta H-T\Delta S \quad (1.1)$$

Here, H represents the enthalpy, S the entropy, G the Gibbs free energy, and T the absolute temperature. A high ΔS_mix value lowers ΔG, thereby favoring the formation of solid-solution phases over intermetallic compounds[10, 15].

HEAs have recently emerged as a distinct and promising category of materials, characterized by their unique combination of exceptional properties and versatile application potential. In 2006, Yeh and colleagues proposed the concept of four core effects the high-entropy effect, sluggish diffusion, severe lattice distortion, and cocktail effect to elucidate the distinct behaviors and fundamental mechanisms that set HEAs apart from conventional alloy systems. These effects form the fundamental framework for understanding the thermodynamic and kinetic stability of HEAs. As illustrated in Figure 3, the four effects are strongly interdependent and collectively contribute to the remarkable structural and functional properties of HEAs. The interplay between them governs microstructural evolution, phase stability, and property optimization, making them essential for advancing HEA design and extending the frontiers of materials science. The high configurational entropy inherent to HEAs drives these four dominant phenomena lattice distortion, sluggish diffusion, cocktail effect, and entropy stabilization—which together determine alloy behavior. A clear understanding of each individual effect is therefore vital before selecting or tailoring specific alloy compositions for targeted applications[13, 15].

Illustration of the four fundamental effects—the high-entropy effect, sluggish diffusion, severe lattice distortion, and the cocktail effect—which collectively govern the thermodynamic stability, microstructural evolution, and mechanical behavior of high-entropy alloys (HEAs) [15].

2.Elements Used in LHEAs

Recent research has demonstrated that incorporating lightweight elements such as Al, Mg, and Ti into Hea systems enables the development of materials that combine low density with high mechanical and thermal stability. Youssef et al. [16] first proved that a mechanically alloyed Al–Li–Mg–Sc–Ti composition could form an entropy-stabilized single-phase solid solution with a density below 3 g cm^{-3} and exceptionally high hardness, establishing the foundation for lightweight HEAs. Building upon this concept, Hashimoto et al. [17] synthesized Li–Mg–Al–Ti–based alloys via mechanical alloying and confirmed, through experimental and computational analyses, that compositional tuning among these light elements can control phase stability and improve oxidation resistance without sacrificing mechanical integrity. Complementarily, Li et al.[18] developed $Al_{15}Zr_{40}Ti_{28}Nb_{12}M(Cr, Mo, Si)_5$ low-density HEAs exhibiting dual BCC phases, densities near 5.3 g cm^{-3}, and compressive strengths exceeding 1.2 GPa, highlighting how lightweight components such as Al and Ti can be reinforced by moderate refractory additions to enhance high-temperature

performance. Collectively, these studies establish that Al–Mg–Ti-based LHEAs achieve an optimal synergy between weight reduction, structural stability, and mechanical robustness, positioning them as promising candidates for next-generation aerospace, transportation, and energy applications.

PROCESSING AND FABRICATION TECHNIQUES

1. Melting-Based Methods

Conventional melting methods arc melting, VIM, and EBM are widely employed for the fabrication of bulk LHEAs. These methods enable homogeneous composition but require careful control to prevent oxidation of reactive light elements. The synthesis of HEAs can be achieved through both liquid-state and solid-state processing routes, each exerting a unique influence on the resulting microstructure and performance of the material (Fig. 4). Among the various production approaches, arc melting, powder metallurgy, and additive manufacturing are the most widely employed. These techniques provide distinct benefits and limitations in terms of microstructural control, scalability, and cost efficiency, positioning HEAs as promising candidates for the emerging class of “Fourth Industrial Materials” within the framework of Industry 4.0[13, 19].

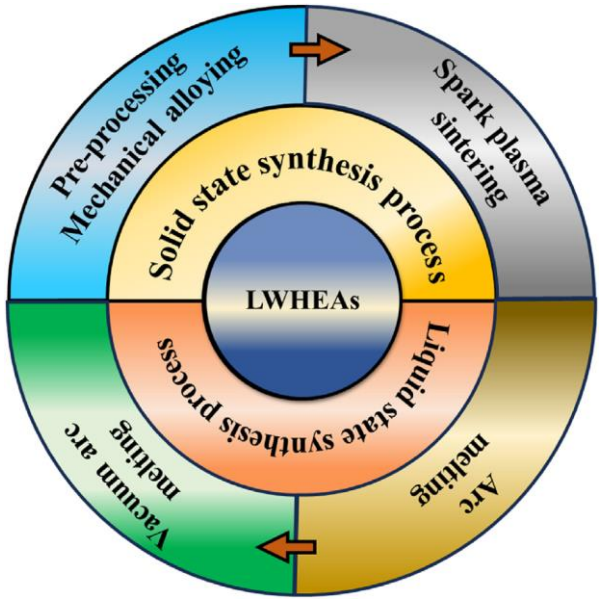


Figure 4 Overview of synthesis and processing techniques employed for LWHEAs[19]

1.1 Vacuum arc melting

Among different fabrication approaches, vacuum arc melting is the most extensively utilized process for the preparation of HEAs. In this process, the constituent elements are melted in an electric arc furnace under a controlled inert atmosphere—typically argon—to prevent oxidation and contamination. The arc torch in such furnaces can reach temperatures exceeding 3000 °C, sufficient to fully melt most metallic elements used in HEAs. The process offers excellent mixing efficiency and enables the production of high-purity alloys through precise control of electrical power and melting conditions. However, it is less suitable for low-melting-point elements such as Mg, Zn, and Mn, which tend to evaporate during melting, making compositional control difficult. In such cases, resistance or induction heating is a more appropriate alternative to reduce vaporization losses and improve compositional homogeneity [19-22].

In laboratory-scale alloy synthesis, vacuum arc melting (VAM) is one of the most frequently employed methods for melting and casting metallic materials. The schematic layout of a typical VAM unit is illustrated in Figure 5. In this setup, the metallic charge is placed in a water-cooled copper crucible, while a non-consumable tungsten electrode serves as the power source. The melting process is initiated by generating a high-temperature electric arc between the electrode and the metallic charge, which supplies sufficient energy to melt the constituent elements and form a homogeneous

alloy. Before melting, the chamber is evacuated to a pressure of approximately 7–10 mbar using a rotary pump (RP) in combination with a diffusion pump (DP) to eliminate residual gases. Once the desired vacuum level is achieved, argon gas is introduced to create an inert atmosphere, thereby minimizing oxidation and contamination during processing. In modern VAM systems, the conventional diffusion pump is increasingly replaced by a turbomolecular pump (TMP), which can attain similar vacuum conditions in a much shorter period. This substitution not only improves operational efficiency but also enhances process cleanliness and overall alloy quality. [23].

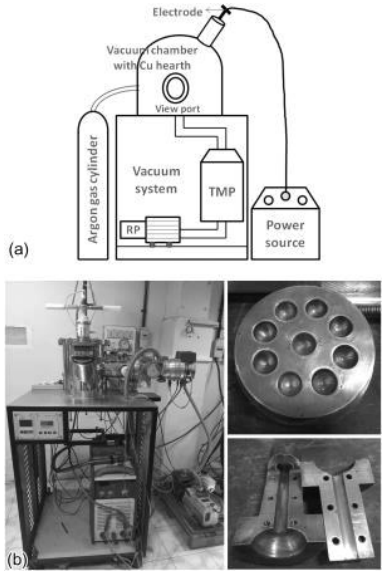


Figure 5 (a) Schematic representation of the vacuum arc melting unit, and (b) photograph of the arc melting and suction casting equipment containing a copper crucible and two types of copper molds [23].

To achieve chemical homogeneity within the synthesized alloys, the melting and solidification cycles are typically repeated multiple times under identical processing conditions. This iterative procedure ensures thorough mixing of the constituent elements and minimizes compositional segregation. Following solidification, the alloy ingots commonly assume a tablet-like form within the water-cooled copper crucible. Alternatively, as illustrated in Figure 6, the molten alloy can be directed through a pre-machined aperture located at the base of the crucible into a copper mold, where it undergoes controlled solidification to produce an ingot with a uniform geometry and refined microstructure[24].

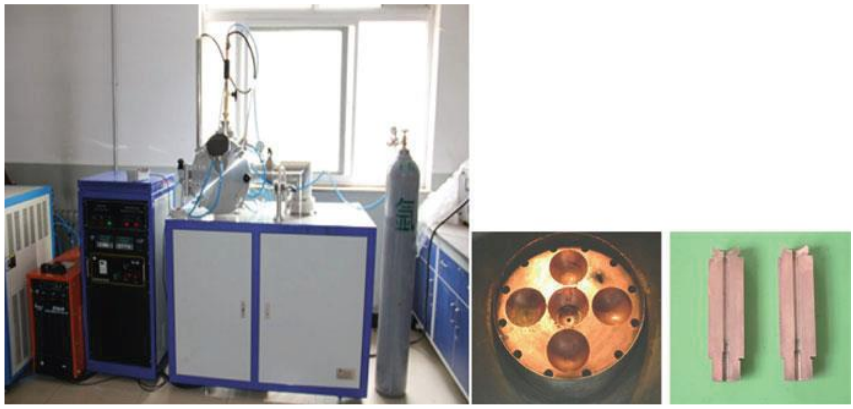


Figure 6 Arc melting and casting equipment showing the copper crucible and copper mold on the right [25]

1.2 Vacuum ınduction melting (vım)

Vacuum induction melting (VIM) is a widely used melting and casting technique in which metallic alloys are melted under vacuum or inert gas atmosphere through electromagnetic induction. During the process, the primary heat source is the alternating magnetic field generated by an alternating current passing through the induction coil, which induces eddy currents within the metallic charge and produces the heat required for melting. A vacuum induction melting (VIM) furnace essentially consists of a melting crucible enclosed within a steel shell connected to a high-speed vacuum system. Heating is achieved by passing an electric current through a series of induction coils, which generates a magnetic field that induces eddy currents in the metallic charge placed inside the refractory crucible. Once the induced heat raises the temperature sufficiently to achieve complete melting, the same magnetic field also facilitates electromagnetic stirring, ensuring homogeneity within the molten alloy. The optimum frequency of the induction coil depends on factors such as the shape, size, and phase state (solid or liquid) of the charge. While older VIM systems typically operated at a fixed frequency, modern furnaces allow for variable-frequency control throughout the melting cycle to achieve faster heating rates and improved process stability [26, 27].

1.3 Laser cladding

Laser cladding is an advanced surface modification technique that utilizes a highly focused laser beam as the primary energy source to deliver controlled heat to a localized region of the substrate. This precise energy input effectively minimizes the heat-affected zone (HAZ), thereby reducing the risks of cracking, porosity, and thermal deformation. Moreover, it produces a stronger metallurgical bond between the coating and the substrate compared

to traditional thermal spray methods. The schematic configuration of the laser cladding process is depicted in Figure 7. The system generally comprises a laser generator, a series of optical components for beam guidance and focusing, a powder delivery unit, and a workpiece manipulator. In this setup, the laser source and optics remain stationary, while the substrate is translated relative to the laser beam to ensure uniform processing. Laser cladding systems are typically computer-controlled and fully automated, providing precise regulation of parameters such as beam power, scanning velocity, and powder feed rate. During operation, the laser-induced molten pool rapidly solidifies as the laser and powder stream advance, resulting in a dense and adherent coating. Depending on specific requirements, either single-track or multi-layer overlapping coatings can be produced to achieve consistent surface integrity and protection. Owing to its process flexibility, high coating quality, and adaptability to components of diverse geometries and dimensions, laser cladding has emerged as a reliable and industrially scalable technique in modern surface engineering. [25].

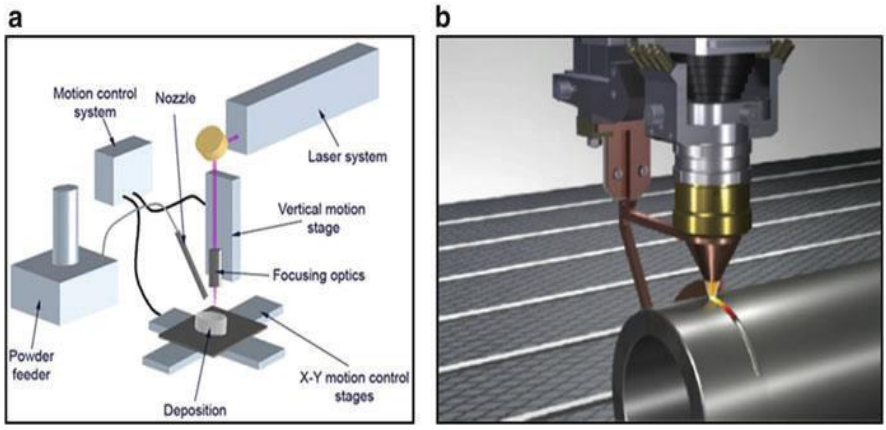


Figure 7 (a) Schematic diagram and (b) operational view of the laser cladding [25]

2. Powder Metallurgy and Mechanical Alloying

Powder metallurgy (PM) is widely adopted for producing LHEAs, as it minimizes contamination and allows precise control of microstructure. MA followed by SPS or HIP processing can generate densely consolidated alloys with uniformly distributed elements. Powder metallurgy (PM) provides a solid-state route for synthesizing Heas (HEAs), offering superior control over composition and microstructural uniformity compared to liquid-phase methods (Fig. 8). In this technique, elemental or pre-alloyed powders of the constituent metals are first blended and then compacted into the desired shape before sintering at elevated temperatures, typically below the melting points of the elements, to consolidate the particles into a dense alloy. [19].

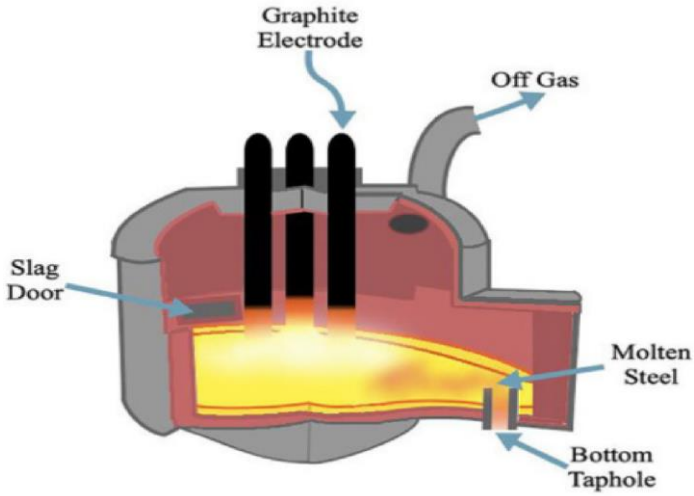


Figure 8 Schematic diagram of an electric arc furnace

A key variation of PM is mechanical alloying (MA), in which high-energy ball milling simultaneously reduces particle size and promotes alloying between different elements. MA, followed by sintering, has become a widely used method for producing nanocrystalline HEAs due to its ability to achieve enhanced solid solubility, room-temperature processing, and homogeneous alloy formation. The first successful synthesis of an HEA by MA was reported in 2008 by Varalakshmi et al., who produced a nanocrystalline AlCrCuFeTiZn alloy. Since then, the number of studies exploring MA-based HEA production has steadily increased, as illustrated in Figure 9. HEAs fabricated through powder metallurgy generally exhibit simple cubic microstructures—typically FCC or BCC phases—which contribute to their high strength, hardness, and ductility. Additionally, the sintering stage can promote phase transformations, converting metastable structures into thermodynamically stable ones, further enhancing the mechanical performance of the alloys[23, 28].

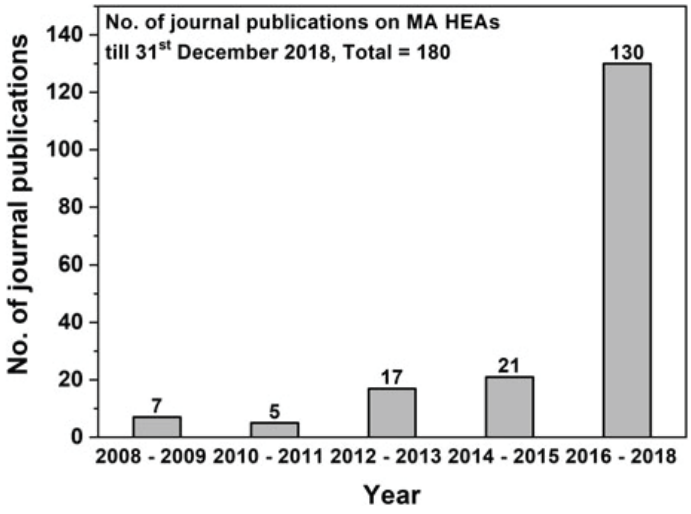


Figure 9 Number of journal publications related to mechanically alloyed Heas (HEAs) [28]

Spark Plasma Sintering (SPS) has become one of the most advanced and efficient techniques for powder consolidation within the domain of powder metallurgy. Traditional consolidation methods such as hot pressing, extrusion, and hot isostatic pressing (HIP) have been widely applied for the fabrication of alloys and composites; however, achieving full densification through these techniques often presents significant challenges. The SPS process effectively overcomes these limitations by combining uniaxial pressure with a pulsed direct electric current, enabling rapid densification and the attainment of superior mechanical properties. In this process, blended alloy powders are placed inside a graphite die, which simultaneously serves as both the mold and electrode. The passage of pulsed electrical discharges induces localized Joule heating and plasma generation at the contact interfaces between powder particles, leading to surface activation, neck formation, atomic diffusion, and plastic flow. This localized and highly efficient heating mechanism allows for extremely short sintering durations—typically completed within minutes—while maintaining relatively low bulk temperatures, which minimizes grain coarsening and helps preserve nanoscale microstructures. The combined influence of electrical activation and mechanical compaction enables SPS to produce nearly fully dense materials with minimal porosity, significantly outperforming conventional sintering methods. Owing to these distinct advantages, SPS has proven particularly effective for temperature-sensitive or reactive alloy systems, such as high-entropy alloys (HEAs) and their composite derivatives, where microstructural stability, phase uniformity, and control over interfacial reactions are of critical importance. The schematic

configuration and working principles of the SPS setup used for HEA-based material consolidation are presented in Figure 10[19, 29].

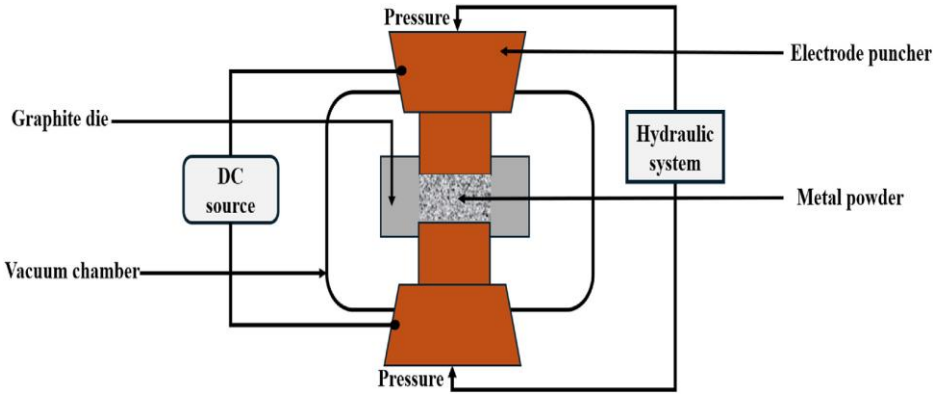


Figure 10 Schematic diagram of SPS[29]

MICROSTRUCTURE AND MECHANICAL PROPERTIES

1. Microstructural Features

The microstructural characteristics of LHEAs generally consist of single-phase solid solutions, such as BCC or FCC structures, or occasionally dual-phase mixtures. Key factors such as grain refinement, dislocation density, and elemental partitioning exert a critical influence on the overall mechanical and thermal performance of these alloys. To accurately evaluate their structural features and phase stability, advanced characterization techniques including X-ray diffraction (XRD), scanning electron microscopy with energy-dispersive spectroscopy (SEM/EDS), and TEM are indispensable for assessing lattice distortion and microstructural evolution. In LWHEA-reinforced light metal matrix composites (LMMCs), microstructural evolution plays a pivotal role in determining key performance parameters such as strength, ductility, wear resistance, and thermal stability. The distribution, morphology, interfacial bonding quality, and phase stability of HEA reinforcements within the matrix collectively govern the composite's behavior under mechanical and thermal loading. Understanding how microstructures form and transform during solidification, heat treatment, and thermomechanical processing is therefore essential for tailoring the final properties of both alloys and composites. In

most as-cast LWHEAs, the BCC solid-solution phase is dominant; however, elemental segregation during dendritic solidification often induces chemical heterogeneity. This segregation primarily arises from disparities in the melting points of the constituent elements—high-melting-point metals solidify first to form dendrites, while low-melting-point elements are concentrated within the interdendritic regions. Among these, Al plays a decisive role in influencing phase stability and microstructural morphology, particularly in systems containing 3d transition metals. Typically, FCC-stabilizing elements such as Co, Cu, and Ni promote the formation of FCC phases, whereas Cr, Mo, V, and Ti encourage the development of BCC structures. Extensive studies on Al- and Ti-containing LWHEAs have demonstrated that increasing the Al concentration enhances lattice strain and distortion energy within the FCC matrix, thereby promoting a phase transition from FCC to BCC. For Ti-rich LWHEAs, a dominant BCC phase is generally observed; however, further Ti enrichment often leads to the appearance of Laves phases. Owing to its positive mixing enthalpy (ΔHmix), copper (Cu) tends to segregate into interdendritic regions, particularly in Ti-containing alloys. For instance, Mg–Al–Mn–Fe–Cu LWHEAs exhibit a combination of BCC and α-brass-type phases, while subsequent annealing or spark plasma sintering (SPS) at around 900 °C facilitates the development of Laves and other intermetallic phases that further influence mechanical and thermal properties.Similarly, Al-Ti-V-$Zr_{0.2}$–Cr_x (x = 0–1) LWHEAs contain a combination of BCC, FCC, and intermetallic phases, where increasing Cr content promotes the transition from FCC to BCC. Moreover, in Mg-Al-Si-Cr-Fe-Ni LWHEAs, the structure primarily transforms into B2 and BCC phases with minor FCC solid-solution fractions after heat treatment. The introduction of Si into Al-Mg-Zn-Cu LWHEAs further leads to the formation of a Mg-Si phase, which strongly influences the morphology and distribution of precipitated phases. Overall, these microstructural phenomena collectively determine the exceptional mechanical and thermal behavior characteristic of LWHEAs. SEM micrographs of the LWHEAs Al_x–Nb–Ti–V–M (x = 0, 0.4, 0.8, 1; M = Co, Ni) alloys, illustrating the characteristic dendritic microstructure typically formed during solidification[14, 29].

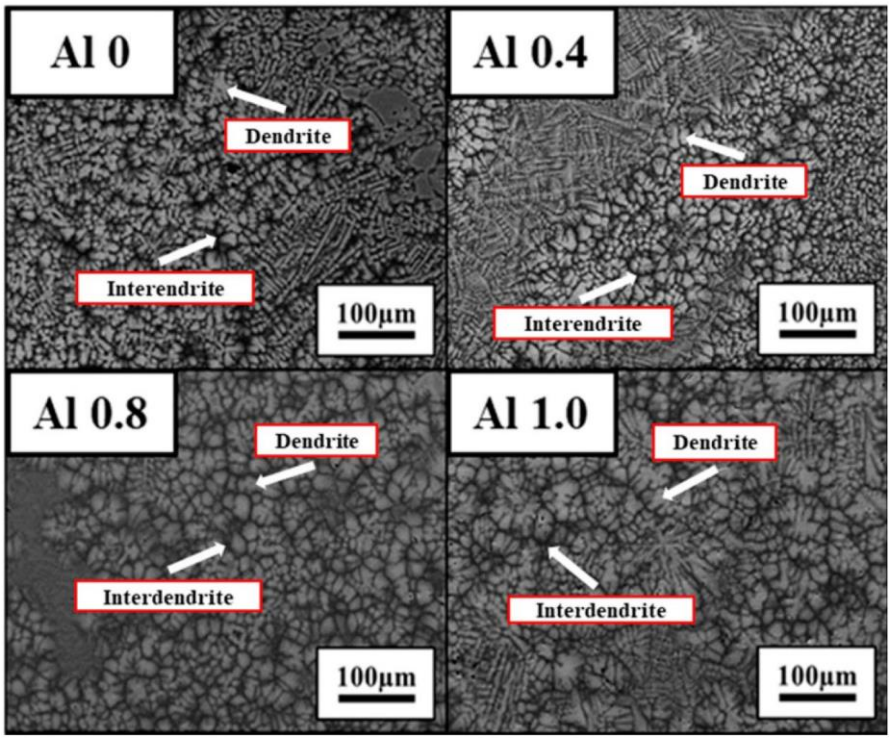


Figure 11 Effect of aluminum content on the microstructure of LWHEAs, showing the characteristic dendritic morphology [14]

2. Mechanical Performance

Thermomechanical processing and heat treatment can further tailor strength–ductility balance through precipitation or phase transformation control. Mechanical properties are fundamental indicators of a material's performance, directly reflecting its ability to withstand deformation and applied loads. In HEAs, these properties are strongly correlated with the crystal structure and phase composition. Generally, HEAs with a BCC structure exhibit high strength and hardness, while those with a FCC structure display superior ductility along with moderate strength. For instance, the AlCuFeMnTiV LWHEA, fabricated via mechanical alloying (MA) followed by spark plasma sintering (SPS), demonstrated that mechanical performance is significantly affected by milling duration, phase evolution, and sintering temperature. After 36 hours of milling, the alloy exhibited a single BCC solid-solution phase, which transformed into a combination of BCC, B2, and HCP phases when sintered at 1050 °C. Additionally, a Cu-rich FCC phase and nanotwins were observed at the grain boundaries (Fig. 12a, b)[13, 14, 19] .

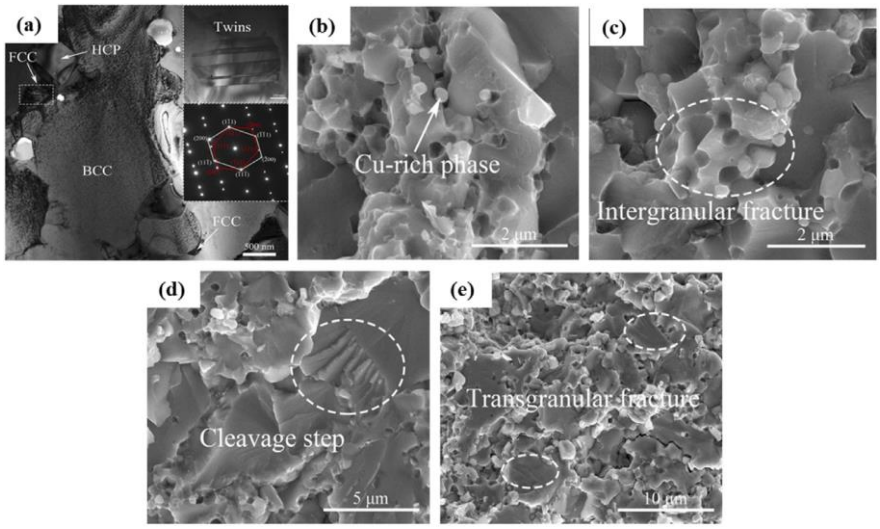


Figure 12 (a) Formation of nanotwins and induced phases, (b) precipitation of the Cu-rich phase, (c) intergranular fracture, (d) cleavage fracture, and (e) transgranular fracture surfaces observed in the AlCuFeMnTiV LWHEA[13]

The optimized microstructure of the alloy yielded exceptional mechanical performance, demonstrating a hardness of 618.44 HV, compressive strength of 2630 MPa, yield strength of 2060 MPa, plastic strain of 15.83%, and a specific strength of 418.79 MPa·cm³/g. The corresponding fracture surfaces displayed intergranular and quasi-cleavage morphologies, indicative of a mixed brittle-ductile failure mechanism. Such outstanding performance highlights the promise of HEAs as next-generation structural materials, integrating high hardness, excellent strength-to-weight ratio, and superior thermal stability features that position them as strong contenders for the "Fourth Industrial Material" in the context of Industry 4.0. These remarkable properties originate from the unique multi-elemental configuration of HEAs, which fosters high configurational entropy, lattice distortion, and sluggish atomic diffusion. Together, these effects enhance strength and stability beyond what is typically achieved in conventional alloys. In particular, LWHEAs offer an advantageous strength-to-weight ratio, making them highly desirable for aerospace, automotive, and energy applications, where lightweighting and operational durability are key design priorities. Furthermore, their capability to retain hardness and resist wear under extreme environmental and mechanical conditions contributes to reduced material degradation and maintenance requirements, aligning well with modern manufacturing goals centered on efficiency and longevity. The mechanical response of LWHEAs is strongly influenced by composition-driven phase transformations, particularly those involving aluminum content. As the Al concentration increases, the crystal structure transitions from a

mixed FCC + BCC configuration to a predominantly BCC phase, resulting in enhanced hardness but reduced plasticity. For instance, the Al–Li–Mg–Sc–Ti LWHEA exhibits a FCC structure with a hardness of approximately 591 HV, which transforms into a HCP structure after annealing, leading to a slight reduction in hardness to around 500 HV. Overall, key mechanical parameters such as elastic modulus, hardness, creep resistance, and fatigue behavior are closely correlated with the phase constitution of these alloys. For high-temperature service environments, current research efforts emphasize improving creep strength, oxidation resistance, and resistance to hot corrosion, as the crystal structure and phase stability remain central to determining the mechanical reliability and long-term performance of LWHEAs[13, 14, 19].

3. Corrosion and Oxidation Resistance

Al- and Ti-containing LHEAs exhibit excellent passive film formation, providing high oxidation and corrosion resistance in marine or high-temperature environments. Heas (HEAs) possess outstanding corrosion and oxidation resistance, making them highly promising industrial materials capable of withstanding environments that rapidly degrade conventional alloys such as steel, aluminum, and titanium. This superior resilience originates from their multi-element composition, which promotes high configurational entropy, lattice stability, and the cocktail effect, leading to the formation of protective oxide or passive surface layers that prevent aggressive chemical attack. As a potential “4th Industrial Material” within the Industry 4.0 framework, HEAs’ ability to maintain structural integrity in corrosive and oxidative environments makes them highly suitable for applications in marine engineering, energy systems, chemical processing, and biomedical devices, where long-term reliability under extreme conditions is essential. In particular, LWHEAs demonstrate remarkable durability due to the spontaneous formation of passivation films on their surfaces. Experimental investigations have shown that LWHEAs perform effectively in acidic and saline media, such as H_2SO_4, HCl, and NaCl solutions, when alloyed with optimized proportions of Al, Ti, Cr, Zr, and Co. The addition of Mo enhances resistance to pitting corrosion, whereas elements like Mg, Al, and Mn can have a detrimental effect by destabilizing the passive layer. Enhanced oxidation resistance has also been reported at elevated temperatures with the incorporation of Al, Si, Ti, and Cr, owing to the formation of dense, adherent oxide scales that suppress further oxidation. The aging temperature plays a key role in corrosion performance, as microstructural transitions from disordered FCC and BCC phases to ordered B2 structures improve electrochemical stability. The open-circuit potential of LWHEAs becomes more electropositive with decreasing pH, suggesting improved corrosion passivity in mildly acidic environments. Furthermore, Ti–Zr–Hf–Nb–Ta LWHEAs exhibit exceptional corrosion resistance due to

the development of multilayered oxide films at the alloy–environment interface. Additive manufacturing (AM) of HEAs further enhances corrosion resistance, reducing corrosion current density by approximately 50–60% compared to conventionally cast alloys, likely due to refined microstructures and reduced segregation. The addition of Cr improves corrosion resistance in NaCl environments, minimizing pit irregularities, while Nb and Zr enhance passivation behavior by reducing ion release and stabilizing surface characteristics. LWHEAs are generally resistant to rusting and localized corrosion, even under prolonged exposure to harsh conditions. At high temperatures, oxidation can severely impair traditional lightweight metals such as Mg, Al, and Ti, due to the formation of porous, non-protective oxide layers. In contrast, HEAs mitigate these effects through multiple mechanisms, including uniform oxide film formation, sluggish diffusion of reactive species, and enhanced thermodynamic stability, ensuring sustained oxidation protection and mechanical reliability in extreme environments[14, 19, 29].

APPLICATIONS

1. Aerospace and Aviation

The development of lightweight Heas (LWHEAs) has rapidly evolved into one of the most dynamic research frontiers in materials science due to their exceptional mechanical, thermal, and chemical properties. These alloys present a promising alternative to conventional materials in the automotive, aerospace, energy, and manufacturing sectors, particularly for components exposed to high temperatures, wear, and mechanical stresses. Owing to their low density and high strength-to-weight ratio, LWHEAs can effectively replace conventional structural materials such as aluminum-, magnesium-, and titanium-based alloys in aerospace, automotive, and energy applications. Al- or Ti-based alloys in intermediate- and low-temperature environments while maintaining superior corrosion and oxidation resistance. Beyond structural applications, LWHEAs also exhibit potential in biomedical, marine, and construction industries, where durability and resistance to harsh operating conditions are essential. Their combination of high wear resistance, oxidation stability, and fatigue resistance makes them attractive for next-generation aerospace components, such as turbine blades, combustion chambers, and load-bearing elements. For example, the incorporation of $Al_{30}Li_{20}Mg_{10}Be_{20}Ti_{10}Cu_{10}$ reinforcement in Al-7086 alloy

achieved greater strength and ductility compared to traditional ceramic reinforcements. LWHEAs are also gaining significant attention for hydrogen storage applications owing to their broad compositional flexibility and high hydrogen affinity. In contrast to conventional metal hydrides, LWHEAs enable safer and more efficient hydrogen transport and storage. Alloys such as $Mg_{12}Al_{11}Ti_{33}Mn_{11}Nb_{33}$ with a single-phase BCC structure have demonstrated superior hydrogen absorption capacity, while $Mg_{0.10}Ti_{0.30}V_{0.25}Zr_{0.10}Nb_{0.25}$ can absorb hydrogen at rates up to 2.7 wt% per minute at room temperature, facilitated by a phase transition from BCC to FCC during hydrogenation. Similarly, MgVAlCrNi LWHEAs synthesized through mechanical alloying and ball milling show enhanced gravimetric hydrogen storage capacity, emphasizing the potential of LWHEAs in clean energy systems and hydrogen fuel distribution networks. In addition to hydrogen storage, the multi-component nature of LWHEAs, coupled with their excellent thermal and mechanical stability, gives rise to unique electrochemical properties that make them attractive candidates for energy storage devices, particularly supercapacitors. The incorporation of lightweight elements such as Li, Mg, and Al enhances specific capacitance, electrical conductivity, and strength while reducing material cost, positioning LWHEAs as competitive materials for electronic devices, biomedical implants, and architectural components requiring multifunctional performance. Furthermore, the high strength-to-weight ratio, oxidation resistance, and wear durability of LWHEAs make them highly appealing for defense and aerospace engineering, where materials must maintain reliability under extreme conditions. Laboratory-scale investigations have revealed that LWHEAs can retain mechanical strength at elevated temperatures, suggesting their suitability for turbine and propulsion components in next-generation aircraft. Although Ni-based superalloys currently dominate these applications, their high density and poor machinability remain limitations that LWHEAs could potentially overcome in future designs. In the marine and naval sectors, Al- and Ti-based LWHEAs have shown excellent corrosion resistance in chloride-rich environments, forming stable oxide films that protect against seawater and acidic media. This makes them promising for propellers, pump housings, valves, tubular fittings, and riser systems used in offshore drilling operations. A summary of these emerging applications is illustrated in Figure 13.

Figure 13 Schematic representation of the potential industrial applications of LWHEAs, highlighting their use in aerospace, automotive, marine, energy, biomedical, and hydrogen storage systems. [13]

In the defense sector, LWHEAs are being explored for use in lightweight armor, structural components of military vehicles, and propulsion systems for unmanned aerial and underwater vehicles (UAVs and UUVs). Their combination of high specific strength, oxidation stability, and corrosion resistance can enhance system efficiency while reducing total mass. Although large-scale commercialization has not yet been realized, ongoing research and preliminary prototype evaluations demonstrate significant promise for the industrial adoption of LWHEAs in the near future. Overall, the unique synergy between lightweight elements, high configurational entropy, and phase stability makes LWHEAs one of the most promising material classes for advanced applications in the Industry 4.0 era. They are positioned to drive innovation in aerospace, automotive, marine, energy, defense, and biomedical systems, where high performance, longevity, and sustainability are key to next-generation material design[13, 14, 19].

REFERENCES

[1] Duan, R., Xu, J., Zhao, Y., Zhou, Q., Yan, Z., Xie, Y., ... & Ramamurty, U. (2024). High entropy a lloys amenable for laser powder bed fusion: A thermodynamics guided machine learning search. Additive Manufacturing, 86, 104195.

[2] Miracle, D. B., & Senkov, O. N. (2017). A critical review of high entropy alloys and related concepts. Acta materialia, 122, 448-511.

[3] Trujillo-Tadeo, J. J., Arruebarrena, G., Dorantes-Rosales, H. J., Bilbao, Y., Vicario, I., Guraya, T., & Hurtado, I. (2023). Design and characterization of a novel lightweight multicomponent alloy Al58Zn28Mg6Si8. Journal of Materials Research and Technology, 27, 3751-3760.

[4] Li, L., Liu, Z., An, X., & Sun, W. (2025). Design methods of Heas: Current status and prospects. Journal of Alloys and Compounds, 180638.

[5] Wang, Z., Wang, X., Yue, H., Shi, G., & Wang, S. (2015). Microstructure, thermodynamics and compressive properties of AlCoCrCuMn-x (x= Fe, Ti) Heas. Materials Science and Engineering: A, 627, 391-398.

[6] Melih Yetiş, Kendiliğinden İlerleyen Yüksek Sıcaklık Yanma Sentezi İle Üretilen Alcoxcrfeni Yüksek Entropili Alaşım Sistemindeki Kobalt Oranı Değişiminin İncelenmesi, 2019

[7] Sarswat, P. K., Sarkar, S., Murali, A., Huang, W., Tan., Free,M. L., 2019. "Additive Manufactured New Hybrid High Entropy Alloys Derived From The Alcofenismtivzr System" Applied Surface Science 476 242-258.

[8] Sharma, A. S., Yadav,S., Biswas, Krishanu, Basu, B., 2018 "Heas And Metallic Nanocomposites: Processing Challenges, Microstructure Development And Property Enhancement" Materials Science & Engineering R 131 1-42.

[9] Stolze, K. Cevallos, F.A. Kong, T. Cava R.J. 2018. "Hea Superconductors On An A-Mn Lattice" J. Mater. Chem. C, 6 Pp. 10441-10449.

[10] Zhang, Y., Zuo, T. T., Tang, Z., Gao, M. C., Dahmen, K. A., Liaw, P. K., Lu Z. P., 2014 "Microstructures And Properties Of Heas" Progress İn Materials Science 61 1-93.

[11] Li, Z., Zhao, S., Ritchie, R. O., Meyers, M. A., 2018 "Mechanical Properties Of Heas With Emphasis On Face-Centered Cubic Alloys" Progress İn Materials Science S0079-6425 30117-8.

[12] Wanqing Dong, Zheng Zhou, Mengdi Zhang, Yimo Ma, Pengfei Yu, Peter K. Liaw and Gong Li, Applications of High-Pressure Technology for High-Entropy Alloys: A Review, 2019

[13] Kamal, M. V., Ragunath, S., Reddy, M. H. S., Radhika, N., & Saleh, B. (2024). Recent advancements in lightweight high entropy alloys–A comprehensive review. International Journal of Lightweight Materials and Manufacture, 7(5), 699-720.

[14] Arya, P. K., Sathiaraj, D., Patel, V. K., Saxena, K. K., Soni, D. K., Rakesh, C., & Khan, M. I. (2025). Developments in Fabrication Properties and Applications of Light Weight High Entropy Alloys. Results in Engineering, 105955.

[15] Odetola, P. I., Babalola, B. J., Afolabi, A. E., Anamu, U. S., Olorundaisi, E., Umba, M. C., ... & Olubambi, P. A. (2024). Exploring high entropy alloys:

a review on thermodynamic design and computational modeling strategies for advanced materials applications. Heliyon, 10(22).
[16] Youssef, K. M., Zaddach, A. J., Niu, C., Irving, D. L., & Koch, C. C. (2015). A novel low-density, high-hardness, Hea with close-packed single-phase nanocrystalline structures. Materials Research Letters, 3(2), 95-99.
[17] Hashimoto, H., Isobe, S., Hashimoto, N., & Oka, H. (2023). Synthesis of Li-Mg-Al-Ti based lightweight high entropy alloys by mechanical alloying and investigation of conditions for solid solution formation. Journal of Alloys and Metallurgical Systems, 4, 100037.
[18] Li, Y., Liaw, P. K., & Zhang, Y. (2022). Microstructures and properties of the low-density Al15Zr40Ti28Nb12M (Cr, Mo, Si) 5 Heas. Metals, 12(3), 496.
[19] Olorundaisi, E., & Olubambi, P. A. (2025). The prospect and limitation of high entropy alloy as 4th industrial material. Materials Today Sustainability, 31, 101163.
[20] B.S. Murty, J.W. Yeh, S. Ranganathan, Heas, 2014
[21] K. Li, W. Chen, Recent progress in Heas for catalysts: synthesis, applications, and prospects, 2021
[22] Kenneth Kanayo Alaneme, Michael Oluwatosin Bodunrin, Samuel Ranti Oke, Processing, alloy composition and phase transition effect on the mechanical and corrosion properties of high entropy alloys: a review, 2016
[23] T.S. Srivatsan and Manoj Gupta, High Entropy Alloys Innovations, Advances, and Applications, 2020
[24] Ebru Sarıoğlu, Mekanik Alaşımlama İle Cocrfemnni Ve AgCoCrFeNi Esaslı Yüksek Entropi Alaşım Tozlarının Sentezlenmesi Ve Karakterizasyonu, 2021
[25] Michael C. Gao, Jien-Wei Yeh, Peter K. Liaw, Yong Zhang, Heas Fundamentals and Applications, 2016
[26] Simbarashe Fashu, Mykhaylo Lototskyy, Moegamat Wafeeq David, Lydia Pickering, Vladimir Linkov, Sun Tai, Tang Renheng, Xiao Fangming, Pavel V. Fursikov, Boris P. Tarasov, A review on crucibles for induction melting of titanium alloys, 2020
[27] ASM Handbook, Volume 15: Casting Volume 15 Handbook Committee, p 1-8, 2008
[28] Mayur Vaidya, Garlapati Mohan Muralikrishna, Budaraju Srinivasa Murty, Heas by mechanical alloying: A review, 2019
[29] Salifu, S., & Olubambi, P. A. (2025). High entropy alloy reinforced lightweight metal matrix composites: A review of the fundamentals, fabrication, properties, and prospects. Materials Today Sustainability, 101216.

Ramjet Engines, Thermodynamic Analysis and Performance Evaluation

Soner ŞEN[1]
Mustafa TAŞYÜREK[2]

1- Doç. Dr.; Selçuk Üniversitesi, Sivil Havacılık Yüksekokulu, Uçak Gövde-Motor Bakımı Bölümü. sensoner@selcuk.edu.tr ORCID: 0000-0003-3385-5577
2- Doç. Dr.; Selçuk Üniversitesi, Sivil Havacılık Yüksekokulu, Uçak Gövde-Motor Bakımı Bölümü. mtasyurek@selcuk.edu.tr ORCID: 0000-0001-9016-8584

ABSTRACT

This book section discussed the ramjet engines aerothermodynamic principles, performance analysis and future engineering developments are extensively. Ramjet engines, the compression with not have movable component, and directly convert the chemical energy of fuel to drive the ram pressure at supersonic speeds using simple but highly efficient systems. Mach 2 to 5, the range of optimal performance in these engines. Diffuser, combustion chamber and nozzle through the three main components of the open Brayton cycle is modeled. The kinetic energy is converted into pressure in a diffuser, combustion takes place at constant pressure, and expansion in the exhaust kinetic energy is re-acquired. Thermal, propulsion, and the total efficiency values, flight Mach number and varies depending on the design parameters. In real systems, shock waves, viscous effects, combustion and heat transfer losses, the increase in entropy and a decrease in efficiency leads to. Today ramjet technology, BrahMos supersonic is actively being used, such as Meteor and systems; scramjet engine concept and hybrid cycle with Mach 10 and over in the direction of reaching hypersonic speeds is evolving. These developments, advanced materials engineering, CFD modeling, and artificial intelligence-based control are based on the progress of integrated systems.

Keywords – Ramjet, Thermodynamic analysis, Specific Impulse, Hypersonic Flight.

INTRODUCTION

Ramjet engines, modern high-speed aircraft propulsion systems of strategic importance, as the basic working principle that compression is completely movable component lacks an air-breathing jet engine represents the family. These engines at supersonic speeds use compressed atmospheric air as the fuel to drive directly convert the chemical energy of aerodynamics (Anderson, 2011). Ramjet of the distinguishing feature, the compression process is completely caused by the ram pressure of the speed of the vehicle (ram pressure) perform with and, therefore, stationery or at low speeds (typically Mach 0.5 under) because work in an effective way. This ramjet of a booster rockets or leads to the necessity of accelerating to an external system such as an aircraft carrier.

Ramjet technology, especially in hypersonic flight, cruise missiles and advanced air defense systems in the context plays a critical role. These engines have mechanical simplicity, despite the Mach 2 and Mach 5 at high speeds in the range of conventional turbojet engines exhibit significant productivity advantage compared to (Heiser & Pratt, 1994). In the last twenty years, Materials Science, Computational Fluid Dynamics (CFD)

progress in the fields of physics and combustion, ramjet performance, durability and operational flexibility have allowed us to improve.

This section aims to provide a comprehensive analysis of ramjet engines. Basic operating principles, historical evolution, component-based thermodynamic analysis, energy conversion mechanisms, and optimization of performance parameters, loss mechanisms, and entropy analysis, scramjet engineering practices and current technology will be studied with comparative evaluation of the prospects in detail.

HISTORIC DEVELOPMENT AND TECHNOLOGICAL EVOLUTION

Ramjet the origins of the concept in 1913, when a French engineer René Lorin of stretches to theoretical articles published in a journal. Lorin, throwing out the idea of the hot gases of combustion in a stationary engine instead of the air in compression by using the kinetic energy of a moving vehicle makes and then adds fuel and burning engine, has proposed the concept (Lorin, 1913). Unfortunately, at that time, enough material technology, the knowledge of aerodynamics and high-speed flight, without the ability to practice, was poured into this visionary idea.

Table 1: Ramjet Technology Timeline of Historical Development

The period of	The major development	The leading figures/Institutions
1908-1913	Introduced the theoretical concept	René Lorin
1940s	II. Germany in World War II V-1 (pulsejet) and the first ramjet attempts	Argus Motoren BMW
In the 1950s	In the USA and the USSR during the Cold War, the development of practical systems	Lockheed d-21, Marquardt
1960-1980	Talos, a Bomarc, Bloodhound missile systems like widespread use in	The U.S. Navy, RAF
1990-2010	Russian-Indian BrahMos, such as Cruise and modern European Meteor air-to-air missiles	NPO Mashinostroyeniya, The MBDA
2000s-	Scramjet research (X-43A, X-51), and mixed-cycle Motors	NASA, DARPA, AFRL

During the Second World War, Germany developed the V-1 “buzz bomb” (basically a pulsejet engine though, the ramjet is considered as an early application of the principle) and experimental Lorin channels such as structures, ramjet technology in the practice of the first concrete examples. With the onset of the Cold War, especially the United States and the Soviet Union between the strategic competition, ramjet technology maturation and contributed to the rapid diversification. In the 1950s and 1960s, Lockheed d-21unmanned high-altitude reconnaissance vehicles, rim-8 Talos the sea-to-air missiles and Bristol Bloodhound surface-to-air missile systems such as platforms, ramjet engines, the premise of the system has been successfully integrated (Cumpsty, 2003).

Since the 1980s, digital control systems, advances in composite materials and aerodynamic design, ramjet has led to significant leaps in performance. During this period, the Russian-Indian co-production of BrahMos supersonic cruise missile, the MBDA Meteor active radar-guided air-to-air missiles and the French ASMP-A nuclear cruise missiles such as the highly skilled systems has emerged. Nowadays, the focus of research and development activities, scramjet (supersonic combustion ramjet) and mixed-cycle engines (combined-cycle engines) are more advanced concepts such as. These traditional systems of ramjet speed through the roof Mach 5 Mach 15+ have the potential to reach hypersonic speeds, such as (Heiser & Pratt, 1994; Curran & Murthy, 2001).

RAMJET ENGINE BASIC COMPONENTS, FUNCTIONS AND WORKING PRINCIPLE

A ramjet engine is extremely simple mechanically, but it contains aerothermodynamic processes in terms of a rather complex structure. It basically consists of four main components, and each is vital for the overall performance of the function of the engine (Hill & Peterson, 1992):

Input diffuser: this component on the front of the engine to slow down to catch the free flow in supersonic air and the pressure increases. This process of successive oblique shock waves and is usually accomplished through a normal shock at the end. The primary purpose of the diffuser where efficient combustion can occur in the combustion chamber, the air is subsonic speeds (typically Mach 0.3-0.5) lower. Diffuser performance, total pressure recovery coefficient (total pressure recovery) has been measured. The total pressure at the exit of diffuser is the ratio of the free flow of total pressure, and the ideal is closer to 1, the compression is so efficient that it.

The combustion chamber: slowed and compressed into the air, through a series of injector fuel (usually a hydrocarbon liquid or kerosene) sprayed. The fuel-air mixture with a spark and bang (usually only on the first run) is ignited by the hot gases or now. In ramjet burning, constant pressure (isobaric) takes place in the context of or very close to it. Supersonic flow to subsonic speeds falls back to this region to stabilize the flame in flame holders (flameholders) used. These streams ensure the combustion of the flame by creating low-speed continuous circulation zones back in.

Exhaust Nozzle: combustion of exhaust gases through the nozzle and accelerate at high temperature and pressure high momentum creates. Nozzle is generally examples of convergent-divergent (CD nozzle or a Laval nozzle) structure. Examples of convergent small mass to the speed of sound in gases section, the narrowest cross-section of the throat (throat)in part, the speed of sound (Mach 1) and reaches the divergent section in the supersonic expansion accelerates. This process forms the largest portion of the impulse. The nozzle shape, the rate of expansion (exit area / throat area) the design of the engine conditions (flight Mach number, altitude) to determine.

This structural simplicity, to ramjet provides several significant advantages: low maintenance, high reliability (no moving parts due to the lack of), resistance to high temperatures and relatively low production cost.

However, the "zero thrust at zero speed" that can be summed up as having to work at low speeds, the biggest limiting factor.

RAMJET THERMODYNAMIC CYCLE OF THE ENGINE AND COMPONENT-BASED ANALYSIS

Ramjet engine, thermodynamically open Brayton cycle can be modeled with. Ideal air-standard cycle following the four basic process is reversible (Mattingly, 2006). To understand the losses in the cycle are real, each process should be handled separately.

Table 2: Ideal and Real Ramjet Cycle Comparison

The process	Cycle for an ideal (Reversible)	Cycle the actual (Irreversible)	Performance parameter Affected
1→2: Compression	Isentropic	Shock waves, boundary layer friction (Diffuser Loss)	Total pressure Recovery (n_d)
2→3: Add heat in	Constant pressure (Isobaric)	Pressure losses, incomplete combustion, heat transfer and	Combustion efficiency (n_b), pressure loss (ΔP_b)
3→4: Expansion	Isentropic	Nozzle friction, heat loss, Loss separation Nozzle)	Nozzle efficiency (n_n)
Throughout the cycle of	Constant entropy ($\Delta s=0$)	The increase of entropies ($\Delta s>0$) - the sum of all losses	In total efficiency (n_o), Specific Impulse (I_{sp})

1 → 2 Isentropic Compression: In this process the input diffuser on the outside free flow conditions (Point 1, Mach >1) the inlet of the combustion chamber (point 2, Mach <1) until it covers the period. The kinetic energy of the air pressure energy (static pressure) is converted. In the ideal case, this process is reversible and adiabatic,

the entropy remains constant. The shock wave and viscous entropy increase in friction causes an increase in total pressure loss and entropy.

2 → 3: Constant Pressure Heat Addition: This process takes place in the combustion chamber. The chemical energy of the fuel as heat energy is added to the stream. This significantly increases the temperature and enthalpy of the fluid. If the pressure is considered constant in the ideal cycle, in practice, friction and momentum exchange in the combustion chamber due to a pressure drop (ΔP_b) occur. In addition, incomplete combustion and the heat transfer to the walls, added a portion of the energy will result in the loss of.

3 → 4 Isentropic Expansion: This process takes place in the exhaust nozzles. High-energy gases, internal energy and kinetic energy convert into a high-speed jet expanding and creating pressure. Ideal expansion the entropy remains constant. However, the nozzle wall friction, heat transfer and flow separation, expansion and entropy decrease the efficiency increases.

4 → 1 Open Loop Heat Transfer: Exhaust gases discharged to the atmosphere and the cycle is completed. This is a process that takes place outside the engine, and generally in analysis it is modeled as an open system.

This ideal cycle T-s (Temperature-entropy) and P-V (pressure-specific volume) diagrams, ramjet of thermodynamics, are essential tools to understand the performance potential. The actual cycle in the diagram, the above-mentioned ideal because of irreversibility, is located to the right of the cycle, and less net of the work (impulse) are produced shows.

Ramjet control volume conservation equations to quantitatively evaluate the performance of the engine with a new approach must be implemented.

Conservation of Mass

The total mass flow rate, mass flow rates of air and fuel entering is equal to the sum of:

$$\dot{m}_0 = \dot{m}_{air} + \dot{m}_f = \dot{m}_e$$

Fuel-Air ratio (FAR), another critical parameter is:

$$f = \frac{\dot{m}_f}{\dot{m}_{air}}$$

Conservation of Energy (First Law)

The energy balance per unit mass can be written as follows (Heiser & Pratt, 1994):

$$\dot{Q} - \dot{W} = \dot{m}(h_e - h_0 + \frac{V_e^2 - V_0^2}{2})$$

Here:

- Q = heat energy added to the system with Combustion
- W= work from the system. In ramjet of miles for not having a job W= 0 is assumed.
- h_e and h_0: Exhaust and entrance specific enthalpy (kJ/kg).
- V_e, V_0: Exhaust and flight speeds (m/s).

This equation shows how the increase in kinetic energy is transformed into heat energy.

Momentum and impulse Equation

The thrust force (F), and pressure forces is derived by considering the change in momentum:

$$F = \dot{m}_e V_e - \dot{m}_0 V_0 + (p_e - p_0)A_e$$

- p_e: at the exit of the exhaust static pressure (Pa)
- p_0: Atmospheric (free flow) pressure (Pa)
- A_e: Exhaust exit cross sectional area (m^2)
- Drives a large part of the momentum of the term creates. Pressure nozzle provides a small contribution to the term, depending on the design. A well-designed compliant (matched) in the nozzle, p_e, p_0 equal or very close.

Performance parameters (Thrust efficiency and Specific Impulse)

Ramjet performance is measured and evaluated several interrelated parameters. Impulse (F) is the basic thrust is the force calculated by the above equation. Specific Impulse (I_{sp}) is the most important indicator of the efficiency of engine fuel. The impulse produced per unit of fuel weight of Impulse is defined as:

$$I_{sp} = \frac{F}{\dot{m}_f \cdot g_0}$$

g_0 is the standard acceleration of gravity (9.80665 m/s^2). The unit is seconds

(s) I_{sp} the same impulse, more fuel can produce an engine that is more efficient or less indicated that long. I_{sp} ramjet of values, depending on altitude and flight Mach number typically varies in the range of 2000 s and 800 s (Mattingly, 2006). Impulse Efficiency (η_p) is the kinetic power of the jet engine thrust produced by the useful conversion efficiency of the vehicle is.

$$\eta_p = \frac{F \cdot V_0}{\dot{m} \cdot \frac{(V_e - V_0)^2}{2} + F \cdot V_0} \approx \frac{2 \cdot V_0}{V_e + V_0}$$

This simplified formula, the jet speed (V_e) with flight speed (V_0) are close to each other in cases where the efficiency is high, it indicates that the thrust. In ramjet V_0 is greater than V_e generally, because the thrust efficiency is always smaller than 1 and increases with V_0.

Thermal efficiency (η_t) is the chemical energy of the fuel conversion efficiency is to increase the kinetic energy of the fluid. Ideal air-standard cycle for a compression ratio (r_c) depends on

$$\eta_t = 1 - \frac{1}{r_c^{(\gamma-1)/\gamma}}$$

Here "γ" is the ratio of heat capacities (C_p/C_v). The "r_c", increase when the ramjet in the vehicle speed (Mach number) increases. Therefore, the thermal efficiency, Mach, along with the number rises.

Total Efficiency (η_o) shows the overall system energy conversion efficiency, and thermal efficiency is equal to multiplied by the efficiency of propulsion.

$$\eta_o = \eta_t \cdot \eta_p$$

Total efficiency, high Mach number (Mach around 3-5) reaches its maximum value (Cumpsty, 2003).

Table 3: Typical Ramjet Performance Parameters of According to Mach Number

Mach Number	Specific Impulse I_{sp} (s)	Thermal efficiency n_t	Total Efficiency n_o	Description
1.5	600 - 900	~0.25	~0.15	Low ram pressure, low efficiency.
3.0	1000 - 1400	~0.55	~0.35	Optimum operating range.
5.0	1200 - 1600	~0.75	~0.30	High thermal efficiency, however, the thrust decreases efficiency.
6.0+	Rapidly falls	~0.80+	~0.20	Scramjet transition region, the losses increase.

REAL RAMJET IN LOSS MECHANISMS AND THERMODYNAMIC ENTROPY ANALYSIS

Deviations from the ideal cycle, the various mechanisms stem from irreversibility and inevitable loss. These losses are the basic factors limiting the performance of the engine and design that make it difficult.

- *Losses in the diffuser:* The diffuser supersonic shock waves occur at a total pressure due to an irreversible decline. How strong if it is in shock (normal shock) becomes so large that lost. In addition, the accumulation of viscous flow and boundary layer separation near the wall diffuser pressure recovery ratio lowers possible. These losses diffuser efficiency (η_d) is characterized by.
- *Combustion Losses:* The mixing ratio imbalances, inadequate mixing or insufficient residence time in the combustion chamber, the fuel to burn completely causes. This gas in the exhaust unburned hydrocarbons (UHC) and carbon monoxide (CO) emissions, such as the formation of undesirable and leads to loss of energy. Combustion efficiency (n_b), usually %95-99. I_s in the range of 5.

Flame holders in the combustion chamber, injectors by creating frictional resistance to flow and high turbulence cause a decrease in static pressure. This pressure drop (ΔP_b) directly reduces impulse.

The nozzle and combustion chamber temperatures are very high in the initial part; the intense heat causes a flux on walls. This energy creates a burden away from the engine and cooling systems. Otherwise, the integrity of the material is compromised.

- *Nozzle Losses:* Viscous Losses from friction in the nozzle wall boundary layer of the jet momentum and reduces the speed.

  A particular design pressure ratio (p_c / p_0) is optimized for. If it is excessively wide nozzle if ($p_e < p_0$), the exhaust gases are separated and make a negative contribution to a compressive force. Missing enlargement ($p_e > p_0$), the full potential of the gases will not grow. In either case, a loss of thrust occurs.

  Nozzle divergent section of the heat loss in the expansion of gas and the acceleration limits.

Analysis of entropy:

The second law of thermodynamics within the framework of all above-mentioned loss mechanisms causes an increase of entropy in the system. The level of irreversibility and entropy production system total available energy (Exergy) is an indication of the loss of. A ramjet for the change in entropy can be written as:

$$\Delta s_{total} = \Delta s_{diffuser} + \Delta s_{burning} + \Delta s_{nozzle} > 0$$

For ideal, in a reversible cycle $\Delta s_{total} = 0$ would be. The real business cycle, the entropy to increase the amount of net of the work that can be taken from the engine (thrust) a decrease in the means, and this leads to the degradation of the overall efficiency of the engine (Anderson, 2011). Entropy analysis to determine the relative magnitude of the losses in the different components and to direct you to where it is most efficient to improvement efforts is a powerful tool.

RAMJET AND SCRAMJET TECHNOLOGY COMPARATIVE ANALYSIS

Ramjet technology, which is a logical extension of Scramjet (Supersonic Combustion Ramjet), hypersonic speeds (Mach 5+) is designed to work efficiently. The most critical difference between the two engine types is basic and the flow rate in the combustion chamber. Table 4 shows a comparison of ramjet and scramjet engines.

Table 4: Ramjet and scramjet of comprehensive comparison

Characteristic	Ramjet	Scramjet	Advantage/disadvantage for Scramjet
Combustion Type	Subsonic Combustion (Mach 0.3-0.5)	Supersonic Combustion (Mach >1.5)	Diffuser losses are much reduced.
Efficient operating ranges from	Mach 2 To 5	Mach 5 To 15+	Allows for Hypersonic flight.
The function of the diffuser	Supersonic flows to subsonic lowers.	Supersonic flow slows, but subsonic to cease.	There is a loss of normal shock.
Combustion chamber design	Requires Flame holders.	The flame holder is not available. Injection, mixing and autoignition are critical.	Combustion stabilization is very difficult.
Heat load	It is high at the entrance to the combustion chamber and nozzle.	Extremely high throughout the entire engine.	The material is much more than the requirement and cooling.
The difficulty of integration	It is relatively low.	Very high. The front body is the primary compressor.	Vehicle and engine design inseparable (airframe-integrated).
Fuel type	Kerosene, JP-series.	Hydrogen, liquid methane. Hydrogen cooling is ideal.	Hydrogen storage challenges.

Scramjet in the supersonic flow in the combustion chamber for stays in ramjet normal shock, which is one of the biggest losses and associated pressure loss eliminated or greatly reduced. This provides the potential for higher thermal efficiency. However, in supersonic flow (in a very short time) the air-fuel mixture, ignition and flame stabilization, scramjet engineering are the biggest technology challenges. In addition, the friction-induced

excessive thermal loads, cooling techniques and advanced new-generation high-temperature materials require (Curran & Murthy, 2001).

CURRENT ENGINEERING PRACTICE, MATERIAL DEVELOPMENTS AND FUTURE PERSPECTIVES

Modern ramjet technology, mission-critical applications in the defense and Aerospace industries, and the design of these applications, the motor has been captured continuously.

Table 5: Leading Contemporary Ramjet - Based Systems and Features

System /Project	Country/ Organization	Motor type	Ma	Primary task	Priorities
BrahMos	India /Russia	Ramjet /Impulse-enhancing	2.8 - 3.0	A cruise missile	Is the world's fastest cruise missile, sea, land, and can be launched from air platforms.
Meteor	Europe (MBDA)	Impulse-enhancing Ramjet	4+	Air-to-air missile	"No-escape zone" has a very large, highly maneuverable
ASMP-A	French	Impulse-enhancing Ramjet	3+	Cruise missiles Nuclear	Stratospheric flight profile, high speed and range.
3M22 Zircon	Russian	Scramjet (estimated)	8-9	Hypersonic cruise missile	Against ground targets, and extremely short response time.
X-51A	USAF /NASA	Scramjet	5+	Technology and demonstrator	Hypersonic flight record of 210 seconds, JP-7 fuel use.

Developments in Materials Engineering

Ramjet and scramjet faced extreme thermal loads necessitated innovations in material science. Today's main developments are as follows:

- *Ceramic matrix composites (CMCs):* silicon carbide (SiC) ceramic fibers and a matrix which is composed of materials such as ceramic, 1200-1400°C and can withstand temperatures like conventional than superrealism is lighter and has better thermal shock resistance. The hottest component, such as the potential for use in combustion chamber and nozzle is high (Naslain, 2004).

- *Refractory metals and Coatings:* Tantalum, molybdenum and tungsten metals and their alloys such as high melting temperature, the Critical Edge and is used on the front lines. To provide protection against wear and oxidation ceramic-based (e.g., silicon carbide) thermal barrier coatings (TBCs) are widely applied.

- *Active Cooling Techniques:* Transpiration cooling (refrigerant leak from porous materials) and regenerative cooling (fuel, engine are passed through channels in the walls and the pre-cooling to be used), scramjet systems such as are being developed to cope with the thermal load. Hydrogen fuel, perfect in this respect due to the cooling capacity, is ideal.

Mixed-cycle engines and Future Perspectives

A single motor type, departure hypersonic flight regime, it is almost impossible to be efficient at a wide range of speeds. Therefore, future hypersonic vehicles hybrid cycle engines (combined-cycle engines) is in the focus of the research. Optimum performance at different speed ranges, these motors more than one engine showing the concept of a single integrated system combines a.

- *Turbo-Ramjet:* Low speeds (Mach 0-3) a turbojet or turbofan at high speeds (Mach 3-6) a ramjet works. The elevator part of the turbine, ramjet mode transition is to bypass or hidden. SR-71 Blackbird, the engine, is an early example of this concept.

- *Rocket-based hybrid Cycle (RBCC):* these are the basic uses of a rocket engine as the core. At low speeds, the exhaust of the rocket ramjet mode can be used for combustion air trapping (ejector ramjet mode). At higher speeds, pure ramjet, or scramjet mode. RBCC are single-stage space (SSTO) vehicles are promising.

- *Turbo-Ramjet-Scramjet:* this can be thought of as the ultimate hybrid cycle engine. And acceleration for takeoff turbojet/ramjet, hypersonic flight scramjet and/or rocket for orbital maneuvers combines modes.

The development of these complex systems, and cooling advances in the field of not only material, but also advanced CFD modelling, machine learning-based control systems and multidisciplinary optimization depends on.

CONCLUSION

Ramjet engines, mechanical simplicity, high reliability and Mach offered in the range of 2 to 5 superior efficiency continue to be indispensable in the field of technology with supersonic flight. In terms of thermodynamic, kinetic energy into heat and kinetic energy is converted into a cycle where higher open again represent elegant and efficient.

However, this simplicity is deceptive. Ramjet the performance of a deep aerothermodynamic, combustion, and materials science is based on the understanding of the physical processes that require complex. "Zero thrust at zero speed" as outlined at low speeds the operating constraints of the combustion chamber and diffuser complex aerodynamics, and high temperatures caused by the material and cooling challenges, ramjet technology of the main limiting factors.

Scramjet, and mixed-cycle engines such as derivatives with a specific focus on advanced research and global efforts to overcome these limitations Mach 10 for hypersonic speeds and over continuous, are promising for the realization of efficient flight. Progress in this area, only military capabilities, but also access to space, reduces costs, and brings the potential of shortening the duration of global transportation. Ramjet and the future of his grandchildren, materials science, fluid dynamics, combustion physics, control engineering, and artificial intelligence have been tied to advances in areas such as interdisciplinary. These engines are the basic principles of physics, engineering and technological ingenuity that can be converted into a proof solution that cross borders on how to plough will continue to be.

REFERENCES

Anderson, J. D. (2011). Fundamentals of Aerodynamics (5th ed.). McGraw-Hill Education.

Cumpsty, N. A. (2003). Jet Propulsion: a simple Guide to the Aerodynamics and Thermodynamic Design and Performance of Jet Engines (2nd ed.). Cambridge University Press.

Curran, E. T., & Murthy, S. N. B. (Eds.). (2001). Scramjet Propulsion. Currently the AIAA Progress in Astronautics and Aeronautics Series.

Heiser, W. H., & Pratt, D. T. (1994). Re - Airbreathing Propulsion. American Institute of Aeronautics and Astronautics.

Hill, P. G., & Peterson, C. R. (1992). Mechanics and Thermodynamics of Propulsion (2nd ed.). Addison-Wesley Publishing.

Lorin, R. (1913). "Propulsion of Flying Machines', 21, 1-5.

Mattingly, J. D. (2006). Elements of Propulsion: Gas Turbines and Rockets. GOD Education Series.
Naslain, R. (2004). "Design, preparation and properties of non-oxide CMCs for application in engines and nuclear reactors: an overview". Composites Science and Technology, 64(2), 155-170.
Saravanamuttoo, H. I. H., Rogers, G. F. C., & Cohen, H. (2009). Gas Turbine Theory (6th ed.). Pearson Education.

Study of Specific Heat Capacities of $CuGaSe_2$ and $CuAlS_2$ Semiconductors Using Einstein- Debye Approximation

Zafer DOĞAN[1]
Tural MEHMETOĞLU[2]

1- Faculty of Engineering and Architecture, Department of Electrical and Electronics Engineering, Tokat Gaziosmanpaşa University, Tokat, zafer.dogan@gop.edu.tr ORCID ID: 0000-0002-7953-0578
2- Amasya University, Taşova Vocational School, Amasya, turalmehmetoglu@yahoo.co.uk ORCID ID: 0000-0001-8909-0222

ABSTRACT

Power electronics applications are an essential part of today's industrial areas. Dynamic load changes in applications, such as electric vehicles, cause variable currents to flow through power electronics components, including IGBTs and thyristors. These dynamic current changes, in turn, cause variable temperature changes in the semiconductors that form the structure of power electronics hardware. We investigated the specific heat capacities of the semiconductors $CuGaSe_2$ and $CuAlS_2$, over a broad temperature range, using the Einstein–Debye approximation. The study explores the temperature dependence of specific heat capacities in both low- and high-temperature regimes, highlighting deviations from simple Debye behaviour due to optical phonon contributions. The calculated results of the heat capacities closely agree with experimental and theoretical data in the intermediate temperature regime, while predicting asymptotic behaviour consistent with theoretical expectations at low and high ends. The results emphasize the importance of including both acoustic and optical phonon modes in modelling thermal properties of wide-bandgap materials, and provide insights useful for thermal management in electronic and optoelectronic devices.

Keywords: Wide-Bandgap Semiconductor, Einstein–Debye Approximation, Specific Heat Capacities, Debye Temperature.

INTRODUCTION

In recent years, population growth and the increasing demands of modern technology have led to a significant rise in electricity consumption. From communication and transportation to lighting and industrial automation, various new electrical and electronic systems have been developed and introduced into everyday human life. Power electronic systems are often subjected to severe environmental stresses. Semiconductor materials possess a distinct bandgap, and their electrical characteristics are highly sensitive to external factors such as illumination and temperature. By introducing specific dopants, materials with different conduction types can be engineered. Among these, impurities—particularly heavy metals with fast diffusion rates and deep-level defects—exert a pronounced effect on the electrical and structural properties of the material [1-6]. Hence, achieving high-purity semiconductor materials requires not only the use of highly refined raw materials but also the maintenance of an ultra-clean fabrication environment to prevent impurity incorporation during processing. Given that most semiconductors are crystalline in nature, maintaining excellent crystal quality is essential. Additionally, precise uniformity in the electrical characteristics of the material is critical for ensuring consistent device performance.

Based on chemical composition, semiconductor materials are classified into several groups, with amorphous and liquid semiconductors—

characterized by their special structural and physical properties—considered as a separate category. Within this framework, semiconductors are typically divided into elemental, inorganic compound, organic compound, and amorphous or liquid types. Within Groups IIIA to VIA of the periodic table, eleven elements exhibit semiconducting characteristics. However, only germanium (Ge), silicon (Si), and selenium (Se) have been employed in practical applications. Among them, Ge and Si continue to dominate as the most widely utilized semiconductor materials in modern electronic technology. This classification includes inorganic compound semiconductors, binary and ternary semiconductors, organic compound semiconductors, and other commonly used semiconductor materials [7-13]. The ternary semiconductor systems can be classified into the following categories:

(1) Compounds composed of one group II atom and one group IV atom substituting for two group III atoms in III–V compounds, such as $ZnSiP_2$, $ZnGeP_2$, $ZnGeAs_2$, $CdGeAs_2$, $CdSnSe_2$ *and etc.*

(2) Compounds formed by one group I atom and one group III atom replacing two group II atoms in II–VI compounds, for example $CuGaSe_2$, $AgInTe_2$, $AgTlTe_2$, $CuInSe_2$, $CuAlS_2$ *and etc.*

(3) Compounds consisting of one group I atom and one group V atom substituting for two group III atoms, including Cu_3AsSe_4, Ag_3AsTe_4, Cu_3SbS_4, Ag_3SbSe_4 *and etc.*

As an illustration, automotive electronic devices typically operate within an ambient temperature window of about –30 °C to +80 °C, similar to that experienced by wind energy converters, automotive lighting assemblies, and public lighting installations. However, such wide temperature variations can significantly affect the electrical, thermal, and mechanical performance of power devices, leading to reduced efficiency, accelerated aging, and potential reliability issues. Therefore, the demand for semiconductor materials such as $CuGaSe_2$ and $CuAlS_2$ has increased considerably, owing to their superior thermal conductivity, high breakdown voltage, and ability to maintain stable operation under elevated temperatures. In these challenging environments, the operational temperature of power devices can rise to their upper threshold, generally between 150 °C and 175 °C. The introduction of advanced wide-bandgap semiconductors is anticipated to overcome this barrier, making the thermal stability of packaging and structural materials the next critical limitation in system development.

Figure 1. Typical power electronics applications in modern technology

The quest for advanced, efficient, and stable semiconductors for optoelectronic, photovoltaic, and thermoelectric applications motivates continual exploration of I–III–VI₂ chalcopyrite compounds [14-18]. Among these, $CuGaSe_2$ and $CuAlS_2$ have emerged as particularly promising candidates, each exhibiting distinct properties and application potentials, notably in light-emitting diodes, solar cells, and other photovoltaic devices [19, 20]. This paper aims to analyze one of the key thermal properties, namely the specific heat capacities (C_V and C_P) by linking it to device-level applications and highlighting existing research gaps. The copper gallium selenide (CuGaSe₂) is a chalcopyrite-type ternary semiconductor widely studied for solar cell absorber applications. It belongs to the I–III–VI₂ family, similar to the well-known CuInSe₂ and Cu(In,Ga)Se₂ (CIGS) compounds. The motivation for studying CuGaSe₂ lies in its favorable combination of optical absorption, defect tolerance, and band structure characteristics. The band gap at room temperature is reported around ~1.68 eV, which is within a favorable range for absorber materials in photovoltaics. The investigation of structural

and thermodynamic properties of $CuGaSe_2$ solids has a long-standing foundation. Blanco et al. [21] proposed a computational approach to evaluate thermodynamic quantities from phonon density of states, significantly improving the predictive capacity of first-principles calculations. Expanding into chalcopyrite semiconductors, Kumar et al. [22] reported experimental insights into their structural and optical properties, confirming their relevance for optoelectronic applications. More recently, Hou et al. [23] investigated chalcogenide compounds and analyzed their elastic and thermal properties, thereby providing further validation of theoretical models. Dong and Gao [24] employed density functional theory (DFT) to investigate the elastic and thermal properties of the $CuTlSe_2$ crystal. They calculated the elastic constants, bulk modulus, shear modulus, Young's modulus, and Poisson's ratio, along with thermal properties such as heat capacity and thermal expansion. The results demonstrated that $CuTlSe_2$ possesses good mechanical stability and holds potential as a promising semiconductor material for optoelectronic applications. Collectively, these works lay the theoretical and experimental groundwork for a comprehensive study of Cu-based chalcopyrite semiconductors such as $CuGaSe_2$ and $CuAlS_2$.

Semiconductor $CuAlS_2$ also belongs to the chalcopyrite family (I–III–VI_2), but with sulfur instead of selenium. Its wider band gap gives it distinct application niches, especially where optical transparency and high-energy photon handling are required [15, 19]. Previous studies have provided important insights into the thermodynamic and structural behavior of $CuAlS_2$ and related chalcopyrite compounds. Hou et al. [25] conducted an experimental investigation on chalcogenide semiconductors, reporting detailed results on their lattice dynamics and thermodynamic properties. Later, Nath et al. [26] applied ab initio techniques to chalcopyrite semiconductors, focusing on their structural, electronic, and thermodynamic properties under various conditions, which further confirmed the predictive power of computational models. Verma et al. [27] employed the quasi-harmonic Debye model to predict thermodynamic parameters such as volume expansion coefficient, bulk modulus, specific heat, and entropy, offering theoretical benchmarks for this material. The methodology itself was refined by Blanco et al. [28], who developed a computational framework to evaluate thermodynamic properties from phonon density of states, thereby extending the applicability of first-principles methods. Complementing these theoretical approaches, Korzun et al. [29] performed experimental calorimetric measurements of $CuAlS_2$, determining its constant-pressure heat capacity (C_p) and deriving constant-volume heat capacity (C_v), entropy, and enthalpy via polynomial fitting. Together, these works provide a consistent foundation for both theoretical and experimental understanding of the thermodynamic characteristics of $CuAlS_2$. Hadj Belgacem [30] calculated the heat capacity of the $CuAlS_2$ semiconductor compound using the Lambert W function, demonstrating the effectiveness of this method in evaluating its temperature-

dependent thermal properties.

As evidenced by the literature, most investigations on $CuGaSe_2$ [2] and $CuAlS_2$ semiconductors have focused on their thermal properties only within relatively narrow temperature ranges. In contrast, the present study extends this analysis by employing the Einstein–Debye approximation to calculate and examine the heat capacities of these materials over a broad temperature domain. The Einstein–Debye approximation is a fundamental theoretical framework developed to describe the temperature dependence of the specific heat in crystalline solids. While the Einstein model assumes that all atoms oscillate with the same frequency, the Debye model incorporates a spectrum of vibrational modes, providing a more accurate description at low temperatures. The combined Einstein–Debye approach effectively captures both acoustic and optical phonon contributions, offering reliable predictions across a wide temperature range. This approximation has been widely applied in the study of semiconductors and complex materials, where deviations from simple models arise due to lattice dynamics. This approach not only provides a more comprehensive understanding of their thermodynamic behavior but also highlights the versatility and effectiveness of the Einstein–Debye model in predicting the thermal characteristics of chalcopyrite semiconductors beyond conventional temperature limits.

EINSTEIN-DEBYE APPROXIMATION AND ITS APPLICATIONS TO SEMICONDUCTORS $CuGaSe_2$ AND $CuAlS_2$

A more generalized form of the Einstein–Debye approximation was developed by accounting for the contributions of both the optical and acoustic branches of the phonon spectrum [31-33]. Within this framework, the specific heat capacities at constant volume (Cv) and constant pressure (Cp) can be expressed, respectively:

$$C_V(T) = 3N_A k_B M(T, \theta_D, \theta_E) \quad (1)$$

$$C_P(T) = 3N_A k_B \left[M\left(T, \theta_D, \theta_E\right) + Q\left(T, \theta_D, \theta_E, \gamma_{ac}, \gamma_{op}\right) \right]. \quad (2)$$

Here, γ_{ac} and γ_{op} are the acoustic and optic Gruneisen parameters, respectively, k_B is the Boltzmann constant, T is the absolute temperature, N_A is the Avogadro number, θ_D and θ_E are the Debye and Einstein temperatures, respectively. The quantities $M(T, \theta_D, \theta_E)$ and $Q\left(T, \theta_D, \theta_E, \gamma_{ac}, \gamma_{op}\right)$ are introduced through the following expressions:

$$M\left(T,\theta_D,\theta_E\right)=L_V\left(T,\theta_D\right)+(s-1)A\left(T,\theta_E\right) \quad (3)$$

$$Q\left(T,\theta_D,\theta_E,\gamma_{ac},\gamma_{op}\right)=\left[\gamma_{ac}L_V(T,\theta_D)+(s-1)\gamma_{op}A(T,\theta_E)\right]^2\left\{\gamma_{ac}\left[\frac{3\theta_D}{8T}(\gamma_{ac}+1)+(1-3\gamma_{ac})D_3\left(1,\frac{\theta_D}{T}\right)+3\gamma_{\mathrm{ac}}\frac{\theta_{\mathrm{D}}}{\mathrm{T}}\left(\mathrm{e}^{\theta_{\mathrm{D}}/\mathrm{T}}\text{-}1\right)^{-1}\right]+(\text{s-}1)\gamma_{\mathrm{op}}\left[\frac{\theta_{\mathrm{E}}}{\mathrm{T}}(\gamma_{\mathrm{op}}+1)+\frac{\theta_{\mathrm{E}}}{\mathrm{T}}\left(\mathrm{e}^{\theta_{\mathrm{E}}/\mathrm{T}}\text{-}1\right)^{-1}\left(1+\gamma_{\mathrm{op}}+\gamma_{\mathrm{op}}\frac{\theta_{\mathrm{E}}}{\mathrm{T}}\mathrm{e}^{\theta_{\mathrm{E}}/\mathrm{T}}\left(\mathrm{e}^{\theta_{\mathrm{E}}/\mathrm{T}}\text{-}1\right)^{-1}\right)\right]\right\}^{-2} \quad (4)$$

where, s denotes the number of atoms contained in a single molecule of the substance and the quantities $L_V\left(T,\theta_D\right)$ and $A\left(T,\theta_E\right)$ are defined following form:

$$A\left(T,\theta_E\right)=\left[\frac{\theta_E}{2T}\frac{1}{\sinh\left(\frac{\theta_E}{2T}\right)}\right]^2, \quad (5)$$

$$L_V(T,\theta_D)=(n+1)D_3\left(1,\frac{\theta_D}{T}\right)-\frac{\theta_D}{T}\frac{n}{e^{\frac{\theta_D}{T}}-1} \quad (6)$$

where the parameter n typically assumes values in the range of 3 to 5 for metals and alloys. From Eqs. (4) and (6), it follows that the heat capacity expressions can be expressed in terms of the Debye functions, leading to improved accuracy and consistency, and defined as:

$$D_3(\mathrm{x})=\frac{3}{x^3}\int_0^x\frac{t^3}{(e^t-1)}dt. \quad (7)$$

In the present work, the evaluation of the Debye functions was carried out using recently developed analytical expressions, which enabled the calculation of the temperature-dependent behavior of the heat capacities [34]:

$$D_n\left(\beta,\mathrm{x}\right)=\frac{n}{x^n}\lim_{N\to\infty}\sum_{m=0}^{N}(-1)^m F_m(-\beta)H_{mn}\left(\beta,\mathrm{x}\right), \quad (8)$$

where $F_i(-\beta)$ is the binomial coefficients [35]:

$$F_m(n)=\frac{1}{m!}\prod_{i=0}^{m-1}\left(n-i\right) \quad (9)$$

and

$$H_{mn}\left(\beta,\mathrm{x}\right)=\frac{n!}{(m+\beta)}\left[1-e^{-x(m+\beta)}\sum_{l=0}^{n}\frac{x^l(m+\beta)^l}{l!}\right]. \quad (10)$$

In the study [30], the author developed formulae based on the Lambert W function to calculate the isochoric (Cv) and isobaric (Cp) heat capacities, and applied these expressions to determine the heat capacity of the $CuAlS_2$ semiconductor over the temperature range of 0–300 K:

$$C_V(T) = 3N_W k_B T^2 \frac{\partial^2 \ln(-2N_W W(-e^{(-\beta\varepsilon_m - 1)}))}{\partial T^2} + 6N_W k_B T \frac{\partial \ln(-2N_W W(-e^{(-\beta\varepsilon_m - 1)}))}{\partial T} \quad (11)$$

$$C_P(T) = \left(3N_W k_B T^2 \frac{\partial^2 ln(-2N_W W(-e^{(-\beta\varepsilon_m - 1)}))}{\partial T^2} + 6N_W k_B T \frac{\partial ln(-2N_W W(-e^{(-\beta\varepsilon_m - 1)}))}{\partial T}\right)$$

$$+A_0 \left(3N_W k_B T^2 \frac{\partial^2 ln(-2N_W W(-e^{(-\beta\varepsilon_m - 1)}))}{\partial T^2} + 6N_W k_B T \frac{\partial ln(-2N_W W(-e^{(-\beta\varepsilon_m - 1)}))}{\partial T}\right) \frac{T}{T_m} \quad (12)$$

where, ε_m is the Lambert maximum energy, $\beta = 1/(k_B T)$, $A_0 = 5.1 \times 10^{-3}\ J^{-1}K$, T_m is melting temperature and W is Lambert function defined as:

$$W(z)Exp(W(z)) = z\ . \quad (13)$$

The Lambert W function is a special mathematical function defined as the inverse of $f(W) = WExp(W)$. It plays a crucial role in providing explicit solutions to a variety of transcendental and differential equations that cannot be solved using elementary functions. Its importance spans multiple fields, including physics, engineering, and applied mathematics, where it is used to model phenomena such as population dynamics, chemical reaction kinetics, electronic circuit analysis, and delay differential equations. The function's ability to transform otherwise intractable problems into closed-form solutions has led to its increasing use in both theoretical studies and practical applications [36-42].

NUMERICAL RESULTS AND DISCUSSION

In the present study, the results obtained from Eqs. (1) and (3) were compared with those reported in the aforementioned study [24, 25, 29, 30] over the temperature range of 0–1300 K. As illustrated in the corresponding Tables 1-5, the calculated heat capacities show excellent agreement with the literature data, confirming the validity of the approach. This comparison not only reinforces the reliability of the Einstein–Debye approximation in predicting thermal properties of $CuGaSe_2$ and $CuAlS_2$ but also demonstrates its applicability over a broader temperature range than previously reported. While earlier studies primarily focused on narrower temperature intervals, the current analysis extends the evaluation to higher and lower temperatures, capturing the gradual evolution of both isochoric (Cv) and isobaric (Cp) heat capacities. The consistency with literature results indicates that the combined

optical and acoustic contributions incorporated in the Einstein–Debye framework accurately reflect the phonon dynamics of the material. Moreover, the observed trend suggests that this approach can be effectively applied to similar chalcopyrite semiconductors, providing a reliable tool for modeling thermal behavior in device-relevant conditions. Consequently, as presented in Tables 1 and 2, the calculated isochoric (Cv) and isobaric (Cp) heat capacities of the $CuGaSe_2$ semiconductor obtained from Eqs. (1) and (2) show excellent consistency with the most reliable reference data reported in Ref. [24] across the considered temperature range. It should be emphasized, however, that comprehensive studies on the thermal properties of $CuGaSe_2$—particularly those focusing on its heat capacities—remain rather limited in the existing literature, underscoring the need for further theoretical and experimental investigations to establish a more complete understanding of its thermodynamic behavior. In all heat capacity calculations, the acoustic and optical Grüneisen parameters were set to $\gamma_{op} = \gamma_{ac} = 1$, and the obtained values are reported in $\frac{J}{mol.K}$ units. The comparative results for the calculated heat capacities of the $CuAlS_2$ semiconductor as a function of temperature are presented in Tables 3–5 alongside the reference data reported in Refs. [29, 30]. As can be observed from these tables, the present calculations of the isochoric (Cv) and isobaric (Cp) heat capacities exhibit good agreement with the corresponding values available in the literature. This consistency highlights the reliability of the employed theoretical framework and confirms that the adopted approximation provides an accurate description of the thermodynamic behavior of $CuAlS_2$ across the considered temperature range. The key physical constants of the $CuGaSe_2$ and $CuAlS_2$ semiconductors are summarized in Table 6. The results confirm that the Einstein–Debye approach, when generalized to account for both acoustic and optical phonon branches, offers an accurate and efficient framework for predicting heat capacities. This highlights the broader applicability of the method to other chalcopyrite and related semiconductors, serving as a reliable theoretical tool for understanding and optimizing materials in device-oriented conditions.

Table 1. Variation of heat capacity C_V as a function of temperature for $CuGaSe_2$.

T(K)	Ref. [24]	Eq. (1)

100.591	64.3988	62.6093
201.183	88.6803	88.3488
299.803	94.6627	94.4135
400.394	97.12610	96.7229
497.041	98.1818	97.7804
601.578	98.8856	98.4073
702.170	99.2375	98.7681
800		98.9973
900		99.1588
1000		99.2743
1050		99.32
1100		99.3595
1150		99.3939
1200		99.4241

Table 2. Variation of heat capacity C_P as a function of temperature for $CuGaSe_2$.

T(K)	Ref.[24]	Eq.(2)
100.592	64.7507	63.5694
297.830	97.8299	97.6888
400.394	101.7009	100.34
499.014	104.1642	101.541
599.605	106.6276	102.217
698.225	109.0909	102.62
798.816	111.9061	102.888
899.408	114.7214	103.071
996.055	118.5923	103.198
1100		103.2982
1150		103.3369
1200		103.3708
1250		103.4007
1300		103.4270

Table 3. Variation of heat capacity C_V as a function of temperature for $CuAlS_2$.

T(K)	Ref.[29]	Eq.(1)

99.969	48.125	49.4509
130.199	64.2187	65.0247
150.781	71.4062	72.203
170.398	76.7187	77.2853
200.628	82.6562	82.88525
220.566	85.0	85.5016
240.184	87.5	87.5609
260.12	89.0625	89.2388
280.382	90.625	90.6201
300.321	91.7187	91.7347
500		96.77
700		98.23
900		98.83
1000		99.01
1200		99.24
1300		99.32
1400		99.38

Table 4. Variation of heat capacity C_V as a function of temperature for $CuAlS_2$.

T(k)	Ref. [29]	Ref.[30]	Eq.(1)
100	45.25	45.26	49.47
120	53.30	53.70	60.54
140	60.21	60.21	68.72
160	66.10	65.60	74.77
180	71.09	70.17	79.30
200	75.27	75.24	82.75
220	78.76	79.56	85.43
240	81.66	81.33	87.54
260	84.04	84.11	89.23
280	86.13	86.01	90.60
300	87.92	87.42	91.72
400			95.14
500			96.77
700			98.23
900			98.83
1000			99.01
1200			99.24
1300			99.32
1400			99.38

Table 5. Variation of heat capacity C_P as a function of temperature for $CuAlS_2$.

T(K)	Ref. [29]	Ref. [30]	Eq.(2)

100	45.23	45.47	49.95
120	53.43	53	61.41
140	60.39	60.10	69.98
160	66.35	67.01	76.40
180	71.40	71.33	81.25
200	75.66	75.33	84.98
220	79.22	79.07	87.89
240	82.21	82.71	90.19
260	84.72	84.31	92.04
280	86.86	87.04	93.54
300	88.75	88.59	94.78
400			98.57
500			100.40
700			102.03
900			102.71
1000			102.91
1200			103.17
1300			103.25
1400			103.32

Table 6. Physical parameters of $CuGaSe_2$ and $CuAlS_2$ structures

Materials	Reference	Debye temperature (K)	Einstein temperature (K)	Melting temperature ($\boldsymbol{K}$)
$CuGaSe_2$	Ref.[23]	306.7	247.3	11356
$CuAlS_2$	Ref. [30]	380	306	1495

CONCLUSION

In this study, the thermal properties of chalcopyrite semiconductors $CuGaSe_2$ and $CuAlS_2$ were systematically investigated using an extended Einstein–Debye approximation that incorporates both acoustic and optical phonon contributions. The calculated isochoric (Cv) and isobaric (Cp) heat capacities were evaluated over a wide temperature range and compared with available experimental and theoretical data reported in the literature. The comparative analysis demonstrated good consistency between the present results and the most reliable reference values, thereby validating the robustness of the employed methodology. It should be emphasized that,

despite the technological relevance of these semiconductors in optoelectronic and photovoltaic applications, the number of comprehensive studies on their thermal properties—particularly heat capacities—remains limited. In this context, the present work extends the temperature range of evaluation and provides new insights into the thermodynamic behavior of these compounds.

REFERENCES

1. Deen, M. J., & Pascal, F. (2017). Electrical characterization of semiconductor materials and devices. In *Springer Handbook of Electronic and Photonic Materials* (pp. 1-1). Cham: Springer International Publishing.
2. Chi, D., Goh, K. J., & Wee, A. T. (Eds.). (2019). *2D semiconductor materials and devices*. Elsevier.
3. Claeys, C., & Simoen, E. (2013). *Radiation effects in advanced semiconductor materials and devices* (Vol. 57). Springer Science & Business Media.
4. Berger, L. I. (2020). *Semiconductor materials*. CRC press.
5. Schroder, D. K. (2015). *Semiconductor material and device characterization*. John Wiley & Sons.
6. Tyagi, M. S. (2008). *Introduction to semiconductor materials and devices*. John Wiley & Sons.
7. Jena, D. (2022). *Quantum physics of semiconductor materials and devices*. Oxford University Press.
8. Gupta, K. M., & Gupta, N. (2015). Recent advances in semiconducting materials and devices. *Advanced semiconducting materials and devices*, 531-562.
9. Malik, R. J. (Ed.). (2012). *III-V semiconductor materials and devices* (Vol. 7). Elsevier.
10. Yacobi, B. G. (2003). *Semiconductor materials: an introduction to basic principles*. Boston, MA: Springer Us.
11. Tamargo, M. C. (2002). *II-VI semiconductor materials and their applications* (Vol. 12). CRC Press.
12. Khanna, V. K. (2005). Physical understanding and technological control of carrier lifetimes in semiconductor materials and devices: A critique of conceptual development, state of the art and applications. *Progress in Quantum Electronics*, *29*(2), 59-163.
13. Enderlein, R., & Horing, N. J. (1997). *Fundamentals of semiconductor physics and devices*. World Scientific.
14. Madelung, O. (2004). I-III-VI2 compounds: (included are I-Fe-VI2 compounds). In *Semiconductors: Data Handbook* (pp. 289-328). Berlin, Heidelberg: Springer Berlin Heidelberg.
15. Shay, J. L., & Wernick, J. H. (2013). *Ternary chalcopyrite semiconductors: growth, electronic properties, and applications: international series of monographs in the science of the solid state* (Vol. 7). Elsevier.

16. Nomura, S., Endo, S., & Irie, T. (1988). Method of materials design for I-III-VI2 chalcopyrite-type mixed crystal semiconductors. *Electronics and Communications in Japan (Part II: Electronics)*, *71*(4), 101-113.
17. Miyake, H., Sugiyama, K., & Hiramatsu, K. (1999). THM Growth of Ternary and Multinary Chalcopyrite Semiconductors. *Research reports of the Faculty of Engineering, Mie University*, *24*, 31-44.
18. Rogacheva, E. I. (2020, October). IN THE I-III-VI2 COMPOUNDS. In *Ternary and Multinary Compounds: Proceedings of the 11th International Conference, University of Salford, 8-12 September, 1997* (p. 1). CRC Press.
19. Lewerenz, H. J. (2004). Development of copperindiumdisulfide into a solar material. *Solar Energy Materials And Solar Cells*, *83*(4), 395-407.
20. Alonso, M. I., Wakita, K., Pascual, J., Garriga, M., & Yamamoto, N. (2001). Optical functions and electronic structure of CuInSe 2, CuGaSe 2, CuInS 2, and CuGaS 2. *Physical Review B*, *63*(7), 075203.
21. Blanco, M. A., Francisco, E., & Luana, V. G. I. B. B. S. (2004). GIBBS: isothermal-isobaric thermodynamics of solids from energy curves using a quasi-harmonic Debye model. *Computer Physics Communications*, *158*(1), 57-72.
22. Kumar, V., Shrivastava, A. K., Banerji, R., & Dhirhe, D. (2009). Debye temperature and melting point of ternary chalcopyrite semiconductors. *Solid State Communications*, *149*(25-26), 1008-1011.
23. Hou, H. J., Zhang, S. R., Guan, H., & Xie, L. H. (2015). Investigation of structural, elastic and thermal properties of CuGaSe2: first principles method. *Chalcogenide Letters*, *10*, 511-518.
24. Dong, Y. J., & Gao, Y. L. (2016). Density function theory of elastic and thermal properties for CuTlSe2 crystal. *Chalcogenide Lett*ers, *13*(110), 515-520.
25. Yu, Y., Kong, X. G., Shen, Y. H., & Deng, J. (2019). Thermodynamic properties of CuAlS2 with chalcopyrite structure from first-principles calculations. *Chalcogenide Letters*, *16*(11), 513-522.
26. Nath, P., Plata, J. J., Usanmaz, D., Al Orabi, R. A. R., Fornari, M., Nardelli, M. B., ... & Curtarolo, S. (2016). High-throughput prediction of finite-temperature properties using the quasi-harmonic approximation. *Computational Materials Science*, *125*, 82-91.
27. Verma, U. P., Jensen, P., Sharma, M., & Singh, P. (2011). Ab initio studies of structural, electronic, optical and thermal properties of CuAlS2 chalcopyrite. *Computational and Theoretical Chemistry*, *975*(1-3), 122-127.
28. Blanco, M. A., Francisco, E., & Luana, V. G. I. B. B. S. (2004). GIBBS: isothermal-isobaric thermodynamics of solids from energy curves using a quasi-harmonic Debye model. *Computer Physics Communications*, *158*(1), 57-72.
29. Korzun, B. V., Mianzelen, R. R., Sheleg, A. U., Tekhanovich, N. P., Bente, K., & Schmitz, W. (2005). Heat capacity and thermodynamic functions of the CuAlS2 semiconductor in the temperature range from 80 to 300 K. P*hysica status solidi (b)*, *242*(10), 1990-1995.
30. Hadj Belgacem, C. (2019). Heat capacity calculation of the CuAlS2 semiconductor compound using the Lambert W function. *Journal of Taibah University for Science*, *13*(1), 559-562.
31. Cankurtaran, M. Askerov, B. M. (1996). Equation of state, isobaric specific heat, and thermal expansion of solids with polyatomic basis in the Einstein-Debye approximation. *Phys. Stat. Sol. (b)*, 194 (1996) 499-507..

32. Askerov, B. M., Cankurtaran, M. (1994). Isobaric specific heat and thermal expansion of solids in the Debye approximation. *Phys. Stat. Sol. (b)*, 185 341-348.
33. Askerov, B.M., Figarova, S.R. (2009). Thermodynamics Gibbs method and statistical physics of electron gases, Springer, Heidelberg.
34. Eser, E. Koç, H. (2016) Investigations of temperature dependences of electrical resistivity and specific heat capacity of metals. *Physica B: Condensed Matter*, 492, 7-10.
35. Gradshteyn I. S., Ryzhik, I.M. (1980). *Tables of Integrals, Series and Products*, Academic Press.
36. Corless, R. M., Gonnet, G. H., Hare, D. E., Jeffrey, D. J., & Knuth, D. E. (1996). On the Lambert W function. *Advances in Computational Mathematics*, *5*(1), 329-359.
37. Warburton, R. D., & Wang, J. (2004). Analysis of asymptotic projectile motion with air resistance using the Lambert W function. *American Journal of Physics*, *72*(11), 1404-1407.
38. Hadj Belgacem, C., & Fnaiech, M. (2011). Solution for the critical thickness models of dislocation generation in epitaxial thin films using the Lambert W function. *Journal of Materials Science*, *46*(6), 1913-1915.
39. Belgacem, C. H. (2016). Explicit solution for critical thickness of semicircular misfit dislocation loops in strained semiconductors heterostructures. *Silicon*, *8*(3), 397-399.
40. Dubinova, I. D. (2004). Application of the Lambert W function in mathematical problems of plasma physics. *Plasma Physics Reports*, *30*(10), 872-877.
41. Valluri, S. R., Jeffrey, D. J., & Corless, R. M. (2000). Some applications of the Lambert W function to physics. *Canadian Journal of Physics*, *78*(9), 823-831.
42. Batzelis, E. I., Anagnostou, G., Chakraborty, C., & Pal, B. C. (2020). Computation of the Lambert W function in photovoltaic modeling. In *ELECTRIMACS 2019: Selected Papers-Volume 1* (pp. 583-595). Cham: Springer International Publishing.

The Impact of Conducting Polymer Derivatives and Their Composites on Supercapacitor Performance

Nursel KARAOĞLAN*

[1]Department of Electrical and Energy, TOBB Technical Vocational School of Higher Education, Karabük University, Karabük, Türkiye
*(nceran@karabuk.edu.tr) Email of the corresponding author

ABSTRACT

Supercapacitors play a crucial role in energy storage systems owing to their rapid charge–discharge capability, high power density, and long cycle life. Their performance is largely determined by the electrode materials employed. Among these, conducting polymers such as polyaniline (PANI), polypyrrole (PPy), polythiophene (PTh), and poly(3,4-ethylenedioxythiophene) (PEDOT) have attracted significant interest as supercapacitor electrode materials because of their facile processability, good mechanical and thermal stability, pseudocapacitive behaviour, and tunable electrochemical properties. In recent years, research in this area has grown rapidly, and numerous modification strategies have been developed. Consequently, there is a strong need for a comprehensive summary of the effects of conducting polymer derivatives on supercapacitor performance. This chapter examines the properties and synthesis methods of these polymer derivatives and evaluates their influence on device performance, with particular emphasis on composite and hybrid structures incorporating carbon-based materials and metal oxides. Structural and electrochemical characterization techniques (FTIR, SEM, CV, EIS), current applications, challenges, and future research directions are also discussed. The study aims to serve as a valuable resource that consolidates recent advances in conducting polymer-based supercapacitors.

Keywords – Conducting Polymers, Polyaniline, Polypyrrole, PEDOT, Pseudocapacitance, Composites, Supercapacitors, Energy Storage.

INTRODUCTION

The rapid increase in global energy demand and environmental concerns associated with fossil fuels have driven the development of high-performance energy storage systems [1] [2]. Among electrochemical energy storage technologies, supercapacitors are distinguished by their high-power density, ultra-fast charge–discharge capability, and long cycle life [3]. Depending on the charge-storage mechanism, supercapacitors are classified into electrochemical double-layer capacitors (EDLCs) and pseudocapacitors. EDLCs [4] store charge electrostatically on high-surface-area carbon materials (activated carbon, graphene, carbon nanotubes), typically delivering specific capacitances of 100–350 F g^{-1}, whereas pseudocapacitors achieve significantly higher capacitances through fast, reversible Faradaic redox reactions [5] [6] [7].

Conducting polymers (CPs) have attracted considerable attention as pseudocapacitive electrode materials owing to their redox-active nature, tunable electrochemical properties, high electrical conductivity, facile processability, and low cost [8] [9]. Polyaniline (PANI), polypyrrole (PPy), polythiophene (PTh) and its derivatives, and poly(3,4-ethylenedioxythiophene) (PEDOT) exhibit intrinsic semiconducting

behaviour due to π-electron delocalization along their conjugated backbones; appropriate doping can increase their conductivity by several orders of magnitude [10] [11]. In recent years, nanostructuring (nanofibers, nanowires, nanosheets, nanospheres) and composite/hybrid strategies combining CPs with graphene, carbon nanotubes (CNTs), MXenes, and metal oxides have dramatically enhanced practical performance [12]. Notably, CP/MXene ($Ti_3C_2T_x$) hybrids have delivered outstanding volumetric capacitances exceeding 1000 F cm^{-3} in compact architectures along with high gravimetric values [13] [14] [15] [16].

The electrochemical performance of CPs is governed by multiple factors, including doping type (p- or n-type), nanostructural morphology, specific surface area, porosity, crystallinity, and synthesis route [17]. Doping introduces charge carriers (polarons and bipolarons) into the polymer backbone. In p-type (oxidative) doping, electrons are removed to generate holes, whereas n-type (reductive) doping injects electrons. These processes significantly narrow the HOMO–LUMO bandgap (typically from 1.5–4.0 eV to below 1 eV), dramatically increasing conductivity and electrochemical activity. Commonly employed dopants include small anions (Cl^-, ClO_4^-, SO_4^{2-}, BF_4^-) and macromolecular anions such as polystyrene sulfonate (PSS) and p-toluene sulfonic acid [11] [18]. CP derivatives are synthesized primarily via chemical oxidation or electrochemical polymerization. Chemical routes, using oxidants such as ammonium persulfate, $FeCl_3$, $CuCl_2$, or H_2O_2, are suitable for large-scale production but offer limited morphological control [19]. Electrochemical polymerization, performed by cyclic voltammetry (CV), chronoamperometry, or chronopotentiometry, enables precise tuning of film thickness, nanostructure, and doping level by adjusting scan rate, cycle number, and potential window [20] [21].

Despite substantial progress, several challenges persist in commercialization. Repeated charge–discharge cycles induce structural degradation (over-oxidation, chain scission) and morphological changes, limiting cycling stability of pristine CPs (typically 70–90 % retention after 1000–5000 cycles), although advanced composites routinely exceed 90–100 % retention beyond 10,000 cycles [22] [23]. Ion intercalation/de-intercalation causes volumetric swelling/shrinkage, leading to mechanical stress and delamination from the current collector [24] [25]. Although theoretical capacitances reach 400–750 F g^{-1}, practical energy densities remain modest (10–50 Wh kg^{-1} for pristine CPs; 80–150 Wh kg^{-1} in optimized hybrids) compared with lithium-ion batteries (150–250 Wh kg^{-1}) [22]. Maintaining stable doping levels over extended periods is difficult, resulting in self-discharge and gradual loss of capacitance and conductivity. Environmental factors (temperature, humidity) and the need for reproducible, homogeneous large-scale production further complicate industrial adoption.

This study systematically examines the influence of conducting polymer derivatives (PANI, PPy, PTh, and PEDOT) and their composites on

supercapacitor performance. Fundamental electrochemical principles, charge-transport mechanisms, doping strategies, synthesis methods, and high-performance carbon-, MXene-, and metal-oxide-based hybrid systems are reviewed. Structural and electrochemical characterization techniques, current challenges, and future research directions in flexible/wearable electronics and advanced energy storage are critically evaluated, providing an up-to-date resource for researchers in the field.

THEORETICAL BACKGROUND

This section outlines the essential scientific principles governing the use of conducting polymers as pseudocapacitive electrode materials in supercapacitors. It covers the basic structure and properties of CPs, charge-storage mechanisms in supercapacitors, the role of doping processes, charge-transport phenomena, and a concise overview of the four most prominent CPs: PANI, PPy, PTh and its derivatives, and PEDOT. The framework presented here provides the necessary background for understanding the electrochemical performance, limitations, and optimization strategies of conducting-polymer-based supercapacitors discussed in the remainder of this study.

A. Introduction to Conductive Polymers (CPs)

CPs are organic macromolecules that acquire electrical conductivity through their conjugated backbone structures consisting of alternating single (σ) and double (π) bonds [26]. Delocalization of π-electrons along the polymer chain confers intrinsic semiconducting behaviour, enabling applications in energy storage, sensors, actuators, and flexible electronic devices. Among the numerous conducting polymers investigated for supercapacitor electrodes, PANI, PPy, PTh and its derivatives, and PEDOT are by far the most extensively studied and applied [27]. These polymers [10] [28], offer high theoretical specific capacitance, widely tunable electrical conductivity, relatively good environmental stability low material cost, and straightforward synthesis routes, as summarized in Table 1. In their undoped state, however, CPs are essentially insulating, with conductivities typically in the range of 10^{-10}–10^{-5} S cm^{-1} and require chemical or electrochemical doping to become electroactive and highly conductive [17].

Table 1. Fundamental properties of common conductive polymers for supercapacitor applications.

Polymer	Repeating Unit	Max. Doping Level	Conductivity Range (S cm^{-1})	Theoretical Specific Capacitance (F g^{-1})
PANI	C_6H_4-NH	~0.5	10^{-5} – ~500	400 – 750
PPy	C_4H_2NH	~0.33	10^{-5} – ~400	500 – 620
PTh	C_4H_2S	~0.33	0^{-5} – ~400	400 – 485
PEDOT	$C_6H_4O_2S$	~0.33	10^{-3} – > 5000	210 ±10
Note: Values are widely accepted theoretical maxima reported in the literature. Practical specific capacitances are typically 40–70 % of these values depending on synthesis, morphology, and measurement conditions.				

The electrochemical performance of CPs is strongly influenced by the type of doping (p-type or n-type), electrode morphology (nanofibres, nanotubes, nanocubes, nanosheets), specific surface area and hierarchical pore structure (micropores <2 nm, mesopores 2–50 nm, macropores >50 nm) [29] [30]. These structural and morphological features can be effectively controlled by optimizing synthesis conditions (temperature, pH, monomer/oxidant ratio) and applying post-synthesis treatments such as annealing, acid/base washing, and secondary doping.

B. Charge Storage Mechanisms in Supercapacitors

Supercapacitors are classified into two main categories based on their charge-storage mechanism: electrochemical double-layer capacitors (EDLCs) and pseudocapacitors [4]. In EDLCs, charge is stored electrostatically through reversible ion adsorption/desorption at the electrode–electrolyte interface, forming a Helmholtz double layer [31]. High-surface-area carbon materials (activated carbon, graphene, carbon nanotubes) are predominantly used, delivering typical specific capacitances of 100–300 F g^{-1} [32]. Capacitance is strongly influenced by accessible surface area and pore-size compatibility with electrolyte ions [33]. In pseudocapacitors, charge storage occurs via fast and reversible Faradaic redox reactions at or near the electrode surface, resulting in substantially higher specific capacitances, typically 500–1100 F g^{-1} in optimised systems, compared with EDLCs [34]. The principal pseudocapacitive materials include transition-metal oxides (RuO_2, MnO_2), hydroxides, and conducting polymers. Conducting polymers are particularly attractive owing to their high theoretical capacitance, low cost, and straightforward processing. However, pristine polymers generally exhibit limited mechanical stability and cycle life (typically 70–90 % capacitance retention after 1000–5000 cycles), whereas advanced composite/hybrid systems routinely achieve >90–100 % retention beyond 10,000–50,000 cycles

[35]. The energy density (E) and power density (P) of a supercapacitor are expressed as

$$E = \tfrac{1}{2} C V^2 \quad \text{and} \quad P = V^2 / 4R \tag{1}$$

where C is the specific capacitance (F g^{-1}), V is the operating voltage window, and R is the equivalent series resistance. Specific capacitance is mostly calculated from galvanostatic charge–discharge curves using

$$C = I \Delta t / m \Delta V \tag{2}$$

where I is the constant discharge current, Δt is the discharge time, m is the mass of active material on one electrode, and ΔV is the voltage window excluding the ohmic (IR) drop.

C. Doping and Charge Transport Mechanisms

Doping is the fundamental process that introduces charge carriers into the conjugated backbone of conducting polymers, dramatically increasing electrical conductivity from the insulating undoped state (typically 10^{-10}–10^{-5} S cm^{-1}) to values exceeding 10^3 S cm^{-1}.

p-Type (oxidative) doping, the dominant mechanism in supercapacitor applications, involves the removal of electrons from the polymer chain and simultaneous incorporation of counter-anions. Commonly used dopants include small anions (Cl^-, ClO_4^-, BF_4^-) and macromolecular anions such as polystyrene sulfonate (PSS^-), camphorsulfonic acid (CSA), and dodecylbenzenesulfonic acid (DBSA). Maximum doping levels are approximately 0.5 $monomer^{-1}$ for PANI and 0.33 $monomer^{-1}$ for PPy, PTh, and PEDOT. n-Type (reductive) doping injects electrons into the polymer backbone and is accompanied by cation incorporation. Because of its poor stability in air and aqueous environments, n-type doping is rarely employed in practical supercapacitor systems. Doping significantly narrows the HOMO–LUMO bandgap from 1.5–4.0 eV (undoped) to below 1 eV, enabling efficient charge transport via quasi-particles. In non-degenerate-ground-state polymers (PANI, PPy, PTh, and PEDOT), polarons (radical ions with charge ±1e and spin ½) predominate at low-to-moderate doping levels. At higher doping levels (>10 %), two polarons combine to form spinless bipolarons (charge ±2e), which become the primary charge carriers. Charge-carrier mobility is strongly influenced by interchain π–π stacking distance and crystalline packing motifs (e.g., herringbone, slipped π-stack).

D. Conducting Polymers in Supercapacitor Applications

This subsection reviews the structural features, common synthesis methods, and key electrochemical characteristics of the four conducting polymers most widely employed in supercapacitor electrodes: PANI), PPy, PTh and its derivatives, and PEDOT.

- PANI and Derivatives: Polyaniline consists of reduced benzenoid and oxidized quinoid units and exhibits three primary redox states: leucoemeraldine (fully reduced, insulating), emeraldine (half-oxidized, conducting when protonated), and pernigraniline (fully oxidized, insulating). The protonated emeraldine form is the most widely used in supercapacitors due to its high conductivity and redox activity. Common oxidants include ammonium persulfate (APS) and $FeCl_3$ (Table 2). Chemical oxidative polymerization is typically performed in acidic medium (1 M HCl or H_2SO_4), whereas electrochemical methods (cyclic voltammetry, chronoamperometry, chronopotentiometry) offer precise control of film thickness and morphology. Nanostructured PANI (nanofibres, nanorods, networks) significantly enhances surface area and ion accessibility. For example, three-dimensional PANI nanofibre electrodes have delivered specific capacitances of approximately 380 F g^{-1} at 1 A g^{-1} with >90 % retention after 3000 cycles [36]. The theoretical capacitance of PANI lies in the range 400–750 F g^{-1}; pristine materials typically achieve 1000–5000 cycles, while advanced composites extend cycle life to 10^4–10^5 cycles [37].
- PPy and Derivatives: Polypyrrole is a five-membered heteroaromatic polymer noted for its excellent environmental stability and wide operational pH range. Synthesis commonly employs $FeCl_3$, Cu^{2+} ions, or APS as oxidants (Table 2), with simultaneous doping during polymerisation. Electrochemical deposition yields homogeneous films.PPy can be prepared in diverse nanostructures (nanotubes, nanowires, n anocubes, networks), providing high surface area and fast ion diffusion. PPy nanowires, for instance, have exhibited specific capacitances of approximately 750 F g^{-1} with 86 % retention after 1000 cycles. The theoretical capacitance of PPy is reported to be in the range of 500–620 F g^{-1} [38], as shown in Table 1.
- PTh and Derivatives: Alkyl substitution at the 3-position (e.g., poly(3-hexylthiophene) (P3HT), poly(3-methylthiophene) (P3MT)) markedly improves solubility and processability. Owing to their hole-dominated transport and superior stability in organic electrolytes, PTh derivatives are frequently used as positive electrodes in asymmetric supercapacitors. Chemical oxidation with $FeCl_3$ or APS is the most common route [ali2009effects] (Table 2). Composites with graphene nanoplatelets (50 wt % GNPL/PTh) and multi-walled carbon nanotubes have achieved specific capacitances of 673 F g^{-1} and 216 F g^{-1}, respectively. The theoretical capacitance of PTh lies in the range 400–500 F g^{-1} [39].
- PEDOT and Derivatives: The electron-donating ethylenedioxy group lowers the bandgap to approximately 1.6 eV. The most prevalent form, PEDOT:PSS, combines high conductivity, transparency (>80 %), and flexibility, making it ideal for transparent and flexible supercapacitors.

Secondary doping with ethylene glycol or DMSO can raise conductivity above 4000–5000 S cm^{-1}. Synthesis routes include vapor-phase polymerisation (VPP), chemical oxidation ($FeCl_3$, Fe $(ClO_4)_3$), and electrochemical deposition (Table 2). Electrochemically deposited PEDOT/graphene composites have shown median specific capacitances of approximately 1410 F g^{-1} with excellent cycling stability. The theoretical capacitance of PEDOT is 200–220 F g^{-1} (Table 1); its lower capacitance is offset by superior conductivity, transparency, and mechanical robustness [40] [37] [41] [42].

Table 2. Common oxidants and synthesis methods for the four major conducting polymers.

Polymer	Chemical Oxidants	Electrochemical Methods	Common Dopants/Anions
PANI	APS, $FeCl_3$	CV, chronoamperometry	Cl^-, HSO_4^-/SO_4^{2-}, BF_4^-, PSS^-
PPy	$FeCl_3$, APS	CV, chronoamperometry	ClO_4^-, BF_4^-, TsO^-, PSS^-
PTh	$FeCl_3$	CV, chronoamperometry, galvanostatic	BF_4^-, ClO_4^-, PF_6^-
PEDOT	$Fe(OTs)_3$, $FeCl_3$, APS (EDOT + for the PSS mixture)	CV, chronoamperometry, Vapour-phase (VPP)	PSS^- (dominant), TsO^-

Table 2 lists commonly employed chemical oxidants, electrochemical polymerization techniques, and typical counter-anions/dopants reported in the literature for the synthesis of PANI, PPy, PTh, and PEDOT. APS and $FeCl_3$ are among the most frequently used chemical oxidants for PANI, PPy, and PTh, while APS (in the presence of PSS) and iron(III) tosylate are widely adopted for PEDOT preparation. Electrochemical polymerization is routinely performed using cyclic voltammetry, chronoamperometry, or constant-current (galvanostatic) methods; additionally, vapor-phase polymerization (VPP) is a well-established technique specifically for PEDOT. Small anions such as Cl^-, ClO_4^-, BF_4^-, PF_6^-, and SO_4^{2-}/HSO_4^- are commonly incorporated as dopants in PANI, PPy, and PTh, whereas poly(styrene sulfonate) (PSS^-) is by far the predominant dopant/counter-ion in PEDOT, with p-toluenesulfonate (TsO^-) also being frequently used.

COMPOSITE AND HYBRID CONDUCTING POLYMER ELECTRODES

The inherent limitations of pristine conducting polymers – particularly mechanical degradation during cycling and moderate energy density – are effectively mitigated by forming composites or hybrids with high-surface-area or highly conductive secondary phases. Carbon-based materials (graphene, carbon nanotubes, activated carbon), two-dimensional MXenes ($Ti_3C_2T_x$), and metal oxides (MnO_2, V_2O_5) are the most successful partners, providing synergistic improvements in capacitance, rate capability, and cycle life [43].

Among the four major conducting polymers used in supercapacitor electrodes, PANI-based composites generally exhibit the highest gravimetric and volumetric capacitances when combined with carbon nanomaterials (reduced graphene oxide, CNTs) or MXenes, while also offering markedly improved cycling stability compared to pristine PANI. PPy composites, especially those incorporating MXene or graphene, are particularly noted for their exceptional long-term durability and near-perfect capacitance retention over tens of thousands of cycles. PEDOT-based materials provide moderately lower capacitance but excel in electrical conductivity, optical transparency, and mechanical flexibility, making them the preferred option for transparent and wearable supercapacitors with outstanding cycle life. PTh and its alkyl-substituted derivatives are less commonly employed but remain valuable in asymmetric device designs owing to their pronounced p-type character and good stability in organic electrolytes, where their composites with graphene nanoplatelets or CNTs function effectively as positive electrodes [27] [40] [44].

Overall, composite and hybrid strategies have significantly enhanced the practical performance of conducting-polymer-based supercapacitors. As summarized in Table 3, gravimetric capacitances now routinely exceed 600 F g^{-1}, with PANI-based systems achieving up to 1120 F g^{-1} and PPy/MXene hydrogels delivering exceptional cycling stability (100 % retention after 10,000–20,000 cycles). Volumetric capacitances above 600 F cm^{-3} are realized in compact PANI/MXene architectures, while PEDOT-based composites offer superior rate capability, transparency, and flexibility suitable for wearable applications. These advances yield energy densities of 30–85 Wh kg^{-1} and cycle lives beyond 10,000 cycles with >90 % capacitance retention, considerably narrowing the performance gap with lithium-ion batteries while retaining the intrinsic advantages of conducting polymers.

CHALLENGES AND PERSPECTIVES

Despite the substantial progress achieved with conducting polymer composites, several critical challenges must be addressed before widespread commercial adoption is feasible.

- Pristine conducting polymers typically retain 70–90 % of initial capacitance after 1000–5000 cycles due to structural degradation and over-oxidation. Modern composites often exceed 90 % retention past 10,000 cycles, yet long-term stability under variable temperature and high-rate conditions remains insufficient for broad commercialization.
- Repeated doping/dedoping induces 10–30 % volume changes, causing mechanical stress, delamination from the current collector, and loss of electrical contact—especially critical in flexible/wearable devices.
- Practical energy densities of conducting-polymer supercapacitors range from 20–85 Wh kg^{-1}, far below lithium-ion batteries (150–250 Wh kg^{-1}), restricting use in applications requiring both high energy and high power.
- Gradual dedoping and self-discharge during operation progressively degrade capacitance and conductivity.
- Temperature fluctuations and humidity impair performance by altering doping levels and ion mobility.
- Scaling from lab-scale electrochemical synthesis to industrial manufacturing still struggles to ensure uniform morphology and reproducible performance.

Research efforts are currently directed toward addressing the remaining limitations of conducting-polymer-based supercapacitors through several promising strategies. Three-dimensional hierarchical nanostructures and aerogels are being developed to reduce mechanical stress from volume changes while maintaining high accessible surface area. Self-healing conducting polymers that incorporate dynamic covalent or supramolecular bonds are under investigation to enable autonomous repair of mechanical damage and extend cycle life. Environmentally benign and biodegradable systems, such as cellulose- or chitosan-based hybrids, are being explored to improve sustainability. Advanced doping approaches that employ redox-active molecules or optimized secondary dopants are expected to further increase capacitance and long-term stability. Computational tools, including machine learning and density functional theory (DFT) simulations, are increasingly applied to accelerate material discovery and predict optimal polymer–composite combinations. For industrial implementation, scalable manufacturing techniques such as roll-to-roll printing, spray coating, and ink-jet deposition are being refined to enable cost-effective, large-area production with high reproducibility. Finally, hybrid supercapacitor–battery architectures that combine conducting polymer cathodes with battery-type anodes are being

studied to achieve higher energy density without compromising power delivery.

CONCLUSIONS

This study has provided a concise overview of the status of the four major conducting polymer derivatives (PANI, PPy, PTh, and PEDOT) and their composites as electrode materials for supercapacitors. These polymers continue to attract interest due to their relatively high theoretical specific capacitance (400–750 F g^{-1}), tunable redox behavior, and low raw-material cost. However, when used in pristine form, they typically exhibit limited cycle life and mechanical robustness because of volume changes and structural degradation during repeated charging and discharging.

The introduction of composite and hybrid architectures with carbon-based materials, MXenes, and selected metal oxides has led to substantial performance improvements. Reported values in recent literature now include gravimetric capacitances of 500–1136 F g^{-1}, volumetric capacitances exceeding 600 F cm^{-3}, energy densities of 30–85 Wh kg^{-1}, and cycle lives surpassing 10,000–50,000 cycles with greater than 90 % capacitance retention in many optimized systems. These advances demonstrate the effectiveness of synergistic design strategies, while structural and electrochemical characterisation techniques have proven indispensable for establishing reliable structure–performance correlations. Practical applications are gradually expanding into flexible and stretchable electronics, transparent devices, and hybrid energy-storage modules. Nevertheless, several important challenges remain unresolved, including long-term cycling stability under real-world conditions, mechanical integrity during prolonged operation, the energy-density gap relative to lithium-ion batteries, doping-level stability, and the transition to reproducible, cost-effective large-scale manufacturing.

Future studies are expected to explore a wider range of approaches, such as three-dimensional hierarchical nanostructures, self-healing and biodegradable polymer systems, novel doping methodologies, machine-learning-assisted material screening, and industrially compatible printing and coating techniques. Given the breadth of possibilities that remain largely untapped, further contributions from the research community will be essential to fully realize the potential of conducting-polymer-based electrodes in next-generation, sustainable, and high-performance supercapacitors.

REFERENCES

[1] G. Smdani, M. R. Islam, A. N. Ahmad Yahaya, and S. I. Bin Safie, "Performance evaluation of advanced energy storage systems: a review," *Energy & Environment*, vol. 34, no. 4, pp. 1094–1141, 2023.

[2] A. Aghmadi and O. A. Mohammed, "Energy storage systems: technologies and high-power applications," *Batteries*, vol. 10, no. 4, p. 141, 2024.

[3] M. Hepel, "Advances in micro-supercapacitors (MSCs) with high energy density and fast charge-discharge capabilities for flexible bioelectronic devices—A review," *Electrochemical Science Advances*, vol. 3, no. 3, p. e2100222, 2023.

[4] D. Bejjanki and S. K. Puttapati, "Supercapacitor basics (EDLCs, pseudo, and hybrid)," in *Multidimensional Nanomaterials for Supercapacitors: Next Generation Energy Storage*, Bentham Science Publishers, 2024, pp. 29–48.

[5] A. Davies and A. Yu, "Material advancements in supercapacitors: from activated carbon to carbon nanotube and graphene," *The Canadian Journal of Chemical Engineering*, vol. 89, no. 6, pp. 1342–1357, 2011.

[6] V. V. N. Obreja, "On the performance of supercapacitors with electrodes based on carbon nanotubes and carbon activated material—A review," *Physica E: Low-dimensional Systems and Nanostructures*, vol. 40, no. 7, pp. 2596–2605, 2008.

[7] S. Gadipelli *et al.*, "Understanding and Optimizing Capacitance Performance in Reduced Graphene-Oxide Based Supercapacitors," *Small Methods*, vol. 7, no. 6, p. 2201557, 2023.

[8] M. H. Gharahcheshmeh and K. Chowdhury, "Fabrication methods, pseudocapacitance characteristics, and integration of conjugated conducting polymers in electrochemical energy storage devices," *Energy Advances*, vol. 3, no. 11, pp. 2668–2703, 2024.

[9] S. A. Jadhav, V. V Kulkarni, P. H. Patil, and I. U. Shah, "Layered double hydroxides (LDHs)-conducting polymers (CPs)-based electroactive materials for supercapacitor application," *Journal of Applied Electrochemistry*, vol. 53, no. 10, pp. 1911–1926, 2023.

[10] M. G. Sumdani, M. R. Islam, A. N. A. Yahaya, and S. I. Safie, "Recent advancements in synthesis, properties, and applications of conductive polymers for electrochemical energy storage devices: A review," *Polymer Engineering & Science*, vol. 62, no. 2, pp. 269–303, 2022.

[11] K. Namsheer and C. S. Rout, "Conducting polymers: a comprehensive review on recent advances in synthesis, properties and applications," *RSC advances*, vol. 11, no. 10, pp. 5659–5697, 2021.

[12] A. Patel, S. K. Patel, R. S. Singh, and R. P. Patel, "Review on recent advancements in the role of electrolytes and electrode materials on supercapacitor performances," *Discover Nano*, vol. 19, no. 1, p. 188, 2024.

[13] Q. Raza and S. Lim, "Interfacial Engineering of Co3O4@MXene for Superior Charge Storage: A Route to High-Capacitance Supercapacitors," *Micromachines*, vol. 16, no. 12, 2025, doi: 10.3390/mi16121313.

[14] W. Liu *et al.*, "Ultrahigh gravimetric and volumetric capacitance in Ti3C2Tx MXene negative electrode enabled by surface modification and in-situ intercalation," *Journal of Power Sources*, vol. 521, p. 230965, 2022.

[15] M. B. Hasan, M. M. Parvez, A. Y. Abir, and M. F. Ahmad, “A review on conducting organic polymers: Concepts, applications, and potential environmental benefits,” *Heliyon*, vol. 11, no. 3, 2025.
[16] D. Khokhar, S. Jadoun, R. Arif, and S. Jabin, “Functionalization of conducting polymers and their applications in optoelectronics,” *Polymer-Plastics Technology and Materials*, vol. 60, no. 5, pp. 465–487, 2021.
[17] T.-H. Le, Y. Kim, and H. Yoon, “Electrical and electrochemical properties of conducting polymers,” *Polymers*, vol. 9, no. 4, p. 150, 2017.
[18] J. Heinze, B. A. Frontana-Uribe, and S. Ludwigs, “Electrochemistry of Conducting Polymers Persistent Models and New Concepts,” *Chemical reviews*, vol. 110, no. 8, pp. 4724–4771, 2010.
[19] Y. Wang and W. Feng, “Preparation of Conductive Polymers,” in *Conductive Polymers and Their Composites*, Springer, 2022, pp. 33–69.
[20] S. Islam *et al.*, “Recent advancements in electrochemical deposition of metal-based electrode materials for electrochemical supercapacitors,” *The Chemical Record*, vol. 22, no. 7, p. e202200013, 2022.
[21] G. D. Sulka, “Electrochemistry of thin films and nanostructured materials,” *Molecules*, vol. 28, no. 10. MDPI, p. 4040, 2023.
[22] M. A. Khaleque, M. A. S. Aly, and M. Z. H. Khan, “Chemical and electrochemical synthesis of doped conducting polymers and their application in supercapacitors: An overview,” *Chemical Engineering Journal*, vol. 507, p. 160444, 2025.
[23] D. N. Nguyen and H. Yoon, “Recent advances in nanostructured conducting polymers: from synthesis to practical applications,” *Polymers*, vol. 8, no. 4, p. 118, 2016.
[24] M. Sharma, P. M. Ajayan, and P. Deb, “Quantum energy storage in 2D heterointerfaces,” *Advanced Materials Interfaces*, vol. 10, no. 11, p. 2202058, 2023.
[25] Y. Li, X. Ren, L. Zhu, and C. Li, “Biomass 3D Printing: Principles, Materials, Post-Processing and Applications,” *Polymers*, vol. 15, no. 12, 2023, doi: 10.3390/polym15122692.
[26] M. Schott and M. Nechtschein, “Introduction to conjugated and conducting polymers,” in *Organic Conductors*, CRC Press, 2022, pp. 495–538.
[27] G. Wang, L. Zhang, and J. Zhang, “A review of electrode materials for electrochemical supercapacitors,” *Chemical Society Reviews*, vol. 41, no. 2, pp. 797–828, 2012.
[28] Z. Roohi, F. Mighri, and Z. Zhang, “Conductive polymer-based electrodes and supercapacitors: Materials, electrolytes, and characterizations,” *Materials*, vol. 17, no. 16, p. 4126, 2024.
[29] S. Nachimuthu, “Carbon-Based Supercapacitors,” in *Polymer Nanocomposites in Supercapacitors*, CRC Press, 2022, pp. 111–126.
[30] K.-S. Kim and S.-J. Park, “Relation of micropores/mesopore ratio on high electrochemical performance of nano-porous carbons,” *Journal of power sources*, vol. 244, pp. 792–798, 2013.
[31] W. Hsieh, T.-L. A. Horng, H.-C. Huang, and H. Teng, “Facile simulation of carbon with wide pore size distribution for electric double-layer capacitance based on Helmholtz models,” *Journal of Materials Chemistry A*, vol. 3, no. 32, pp. 16535–16543, 2015.
[32] P. Simon and Y. Gogotsi, “Materials for electrochemical capacitors,” *Nature*

materials, vol. 7, no. 11, pp. 845–854, 2008.
[33] F. Beguin and E. Frackowiak, *Carbons for electrochemical energy storage and conversion systems*. Crc Press, 2009.
[34] S. Fleischmann *et al.*, "Pseudocapacitance: from fundamental understanding to high power energy storage materials," *Chemical Reviews*, vol. 120, no. 14, pp. 6738–6782, 2020.
[35] G. Sanyal, R. K. Trivedi, and B. Chakraborty, "Electrochemical Energy Storage in 2D Nanomaterials: Present Status and Future Challenges," in *Sustainability in Energy and Environment*, Apple Academic Press, 2024, pp. 3–53.
[36] J. Wang and D. Zhang, "One-Dimensional Nanostructured Polyaniline: Syntheses, Morphology Controlling, Formation Mechanisms, New Features, and Applications," *Advances in Polymer Technology*, vol. 32, no. S1, pp. E323--E368, 2013.
[37] A. Eftekhari, L. Li, and Y. Yang, "Polyaniline supercapacitors," *Journal of Power Sources*, vol. 347, pp. 86–107, 2017.
[38] M. A. Del Valle, M. A. Gacitúa, F. Hernández, M. Luengo, and L. A. Hernández, "Nanostructured conducting polymers and their applications in energy storage devices," *Polymers*, vol. 15, no. 6, p. 1450, 2023.
[39] A. ur Rahman, H. Noreen, Z. Nawaz, J. Iqbal, G. Rahman, and M. Yaseen, "Synthesis of graphene nanoplatelets/polythiophene as a high performance supercapacitor electrode material," *New Journal of Chemistry*, vol. 45, no. 35, pp. 16187–16195, 2021.
[40] G. A. Snook, P. Kao, and A. S. Best, "Conducting-polymer-based supercapacitor devices and electrodes," *Journal of power sources*, vol. 196, no. 1, pp. 1–12, 2011.
[41] Z. Genene, Z. Xia, G. Yang, W. Mammo, and E. Wang, "Recent advances in the synthesis of conjugated polymers for supercapacitors," *Advanced Materials Technologies*, vol. 9, no. 9, p. 2300167, 2024.
[42] S. J. Edmunds, "Harnessing FeCl3-PEO Photoreactions for Precision Photolithographic Patterning of Conductive PEDOT and Biocompatible PDMS," University of California, San Diego, 2024.
[43] P. S. NK, S. M. Jeong, and C. S. Rout, "MXene--carbon based hybrid materials for supercapacitor applications," *Energy Advances*, vol. 3, no. 2, pp. 341–365, 2024.
[44] A. Rudge, J. Davey, I. Raistrick, S. Gottesfeld, and J. P. Ferraris, "Conducting polymers as active materials in electrochemical capacitors," *Journal of power sources*, vol. 47, no. 1–2, pp. 89–107, 1994.

FPGA-Based Real-Time Dynamic System Accelerators: Modeling, Numerical Design, And Timing

Sertaç YAMAN[*1]
Mehmet Latif LEVENT [2]

[1,2]Department of Electrical and Electronics Engineering/Hakkari University, Turkey
[*](sertacyaman@hakkari.edu.tr)

ABSTRACT

Deterministic execution of model-based computations is critical for real-time control and estimation applications because it directly impacts closed-loop stability and reliability. This chapter presents the theoretical components necessary for designing a high-efficiency accelerator on an FPGA to solve and control general dynamic systems encountered in various fields, ranging from industrial drives to autonomous systems. This work fundamentally leverages the advantages of deterministic pipelining and the fixed initiation interval (II) offered by FPGAs, which contrasts with the timing uncertainties of traditional processors.

The theoretical framework examines numerical design decisions starting from continuous-time dynamics and including discretization, stability regions, numerical continuity, fixed-point analysis, word-length selection, scaling, and overflow safety. It also discusses architectural patterns in detail, including dataflow, pipeline depth, CORDIC, and DSP slice utilization. Additionally, it clarifies the impact of the low latency budget and minimal jitter achieved on the stability of closed-loop control (phase/gain margins). Finally, a comprehensive verification methodology involving bit-true simulation and hardware-in-the-loop (HIL) testing assures the proposed design flow. The concludes design rules that can be applied directly and results that guarantee robustness metrics used in control engineering. These results facilitate the derivation of measurable design metrics (latency, resource utilization, power, and error tolerance) from system requirements.

Keywords – FPGA, Real Time, Dynamic systems, Numerical Design, Pipeline.

INTRODUCTION

The deterministic execution of model-based computations in real-time control and estimation applications directly impacts the stability and reliability of closed loops. Dynamic systems encountered in various fields, ranging from industrial drives and autonomous vehicles to process control and biomedical devices, can be single-input–single-output (SISO) or multiple-input–multiple-output (MIMO), and can be either linear or nonlinear and continuous or discrete in time. All these systems have one thing in common: the need to complete state updates and control laws within the defined sampling period to achieve predictable latency and low timing jitter.

Although general-purpose CPUs and GPUs can perform high performance, timing uncertainties resulting from task scheduling and memory access routines pose a risk under strict real-time restrictions. These uncertainties, mainly associated with scheduling and memory access routines, improve the worst-case execution time (WCET). This increases the risk of exceeding the control loop's time budget, which can risk closed-loop reliability.

Conversely, FPGA-based implementations enable the achievement of a fixed initiation interval (II) and predictable latency through dataflow and pipelining principles. Designers can approach performance and stability goals simultaneously by optimizing DSP blocks, BRAM, and LUT/FF resources, as well as numerical representations (e.g., fixed-point) and architectural decisions (e.g., parallelism and resource sharing). In this context, a co-design methodology in which the dynamic model, numerical quantization, and hardware architecture are optimized concurrently becomes mandatory and forms the backbone of this chapter.

This chapter presents a general design framework that is independent of a specific application. First, it summarizes the state-space formulation and discretization principles of dynamic systems. Then, it discusses numerical representation choices (floating-point versus fixed-point), dynamic range analysis, and techniques for preventing overflow and managing quantization error to achieve bit-true accuracy. Next, dataflow architecture, pipelining depth, and resource sharing are covered, as well as their impact on I/O latency. Then, the impact of these timing metrics on controller stability (phase/gain margins and delay sensitivity) is clarified. Finally, a verification/proof flow involving timing closure, worst-case latency, and bit-true validation is proposed. To illustrate the practical application of these general principles, the text uses case study boxes from diverse fields (e.g., 6-DoF rigid body, brushless DC motor driver, two-tank level control, and inverted pendulum) throughout the text.

The proposed approach has two main objectives:

(i) to provide a reusable recipe for co-optimizing modeling, numerical design, and architectural timing decisions, and (2) to provide a reusable recipe for co-optimizing modeling, numerical design, and architectural timing decisions.

(ii) Facilitate the derivation of measurable design metrics (e.g., latency, resource utilization, power, and error tolerance) from system requirements by application developers and their correlation with control performance. Thus, readers will be able to follow an end-to-end process for developing repeatable, FPGA-based, real-time accelerators for various dynamic systems.

NUMERICAL QUANTIZATION AND BIT-TRUE ACCURACY IN CONTROL SYSTEMS

Fixed-point representations are preferred in real-time accelerators due to deterministic efficiency targets, such as area, power consumption, and most importantly, initialization interval (II). Determining the word length involves critical trade-offs between sign bits, fractional bits, and the scaling factor, as well as between dynamic range and digital noise (quantization error).

Quantization Noise and Stability: Quantization noise can typically be modeled using white and input-independent assumptions (Widrow & Kollar, 2008). However, when this noise infiltrates the control loop, it can narrow the stability limits of the system and cause limit cycles, particularly around zero.

Overflow Safety: In the event of an overflow during calculations, apply a saturation policy instead of wrap-around. Saturation fixes the numerical value at the maximum or minimum allowed value. This approach is preferred because it more closely simulates the system's physical limits (i.e., actuator saturation) and is safer than wrap-around in terms of control stability.

Quantization Methodology: Floating-point (IEEE-754) offers high accuracy but incurs a much higher resource and power cost in FPGAs (Muller et al., 2018). The precision of fixed-point, however, requires careful "bit-true" design. This process involves bit-true simulation of the theoretical model, interval analysis, and block-level scaling methods to determine the optimal word length for each mathematical operation (e.g., addition, multiplication, and division) (Meyer-Bases, 2014; Widrow & Kollar, 2008). This methodology is a fundamental step in ensuring the numerical robustness of the discretized dynamical system.

THE DIRECT IMPACT OF THE DISCRETIZATION METHOD (ZOH VS. TUSTIN) ON HARDWARE ARCHITECTURE IS EXAMINED

The conversion of dynamic systems from the continuous-time (s-domain) to the discrete-time (z-domain) is not only a matter of numerical accuracy; it is also a critical decision that directly affects the hardware architecture and the efficiency of the pipeline. The zero-order hold (ZOH) mentioned in the text is the most intuitive method physically (it holds the control signal constant throughout the sampling period) and generally leads to a direct mapping of e^{-sT} to z^{-1} (a one-clock-cycle delay). This translates naturally to register-based delays in hardware and supports the goal of II = 1.

As an alternative, the Tustin (bilinear) transform ($s \approx \frac{2\,(z-1)}{T\,(z+1)}$) which is often used in control theory, tends to preserve the frequency response of a continuous-time system better, especially around the Nyquist frequency. However, this transform can lead to algebraic loops, which present a significant challenge for hardware implementation. An algebraic loop occurs when the output of a block must be returned to itself within the same clock cycle without any intervening delay. This violates the synchronous design philosophy of an FPGA, rendering the II=1 objective impossible and preventing timing closure by overextending the critical path. Therefore, choosing ZOH or Modified ZOH over Tustin for hardware accelerator design is not only a quantization decision, but also a necessary co-design step for achieving a hardware pipeline-friendly dataflow architecture.

DETERMINISTIC DATAFLOW AND FPGA ARCHITECTURAL PATTERNS

Pipelining, fluid parallelism, and dataflow modeling are fundamental patterns for deterministic execution in FPGAs. The Synchronous Dataflow (SDF) graph approach, in particular, requires constant production and consumption rates for each computational block, which facilitates pipeline balancing and buffer sizing between them (Lee & Messerschmitt, 1987). This deterministic flow model eliminates concerns about worst-case latency (WCL) by guaranteeing the predictability of interval (II) and latency metrics.

Distinguishing Latency and Throughput in Dataflow Graphs: Registers and FIFO Buffers:

The Synchronous Data Flow (SDF) model is fundamental to achieving maximum throughput and the goal of II=1. Nevertheless, in real-time control loops, end-to-end latency is as critical to balance these two purposes; developers must carefully pick the type of interconnections in the data flow channel.

- **Register-Based Pipeline:** In this approach, a register, or flip-flop, is inserted directly between each computational stage, such as a digital signal processor (DSP) block operation. This is the standard I = 1 architecture, as shown in Fig 1, the data streams in an uninterrupted stream with each clock cycle period

  Latency = N x T_{clk}

This architecture ensures the lowest latency and zero jitter. It is the selected "gold standard" for dynamic system accelerators.

- **FIFO-Based Dataflow:** In some circumstances, such as when changing between different clock domains or synchronizing parallel modules with

different processing times (e.g., a matrix multiplier with a CORDIC core), block RAM based on first-in-first-out buffers is inserted between stages.

- Advantage: This approach simplifies pipeline balancing while maintaining the throughput of II=1.
- Disadvantage and risk: However, using a first-in, first-out (FIFO) approach introduces uncertainty into the latency. The fill level of the FIFO can vary depending on the data arrival rate during runtime. While this maintains the II=1 throughput, it disrupts the deterministic nature of end-to-end latency, creating the potential for jitter.

In real-time dynamic systems, especially critical paths with strict timing constraints (e.g., control laws), FIFO-based flow control should be avoided, and fully registered, deterministic delay pipelines should be used instead. As mentioned in the text, BRAMs are excellent for storing state vectors (x_{k+1}), but using them as buffers between pipeline stages is risky. Supervisory cores are usually divided into three main functional blocks: the state update core (model integration), the predictor/observer core, and the control law core. Figure 5.1 illustrates the typical data flow between these cores. In this architecture, multiply-accumulate (MAC) chains are mapped to specialized digital signal processor (DSP) blocks to achieve high speed and low power consumption. Dual-ported block RAM (BRAM) resources are used to store state variables.

Parallelism, Resources, and II Optimization: The main trade-off in hardware architecture is between parallelism, or loop unrolling, and resource sharing. Maximum parallelism facilitates achieving the II=1 target, but it increases the utilization of both the BRAM and the DSP. Conversely, resource sharing reduces hardware costs, but it increases latency and II because it requires sequencing operations (Coussy & Morawiec, 2008). Finding the optimal balance point based on design requirements is essential in navigating this tradeoff.

Function Acceleration and CORDIC: Instead of using full-precision, free-floating point for square roots, trigonometric operations (sin, cos), or hyperbolic operations encountered in dynamic systems, the fixed-point-based CORDIC (Coordinate Rotation Digital Computer) algorithms are commonly used. Because CORDIC uses only addition, subtraction, and shift operations, it utilizes LUT/FF resources more sparingly than DSP blocks. This strategy accelerates complex functions, efficiently utilizes FPGA platforms, and reduces power consumption. (Meyer-Baese, 2014).

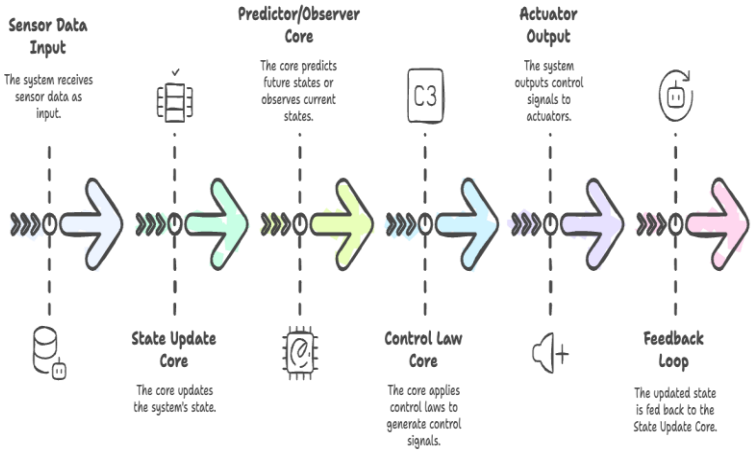


Fig 1. Real-Time dynamic system accelerator pipeline

Figure 1 illustrates the proposed deterministic dataflow architecture for a real-time dynamic system accelerator. The design is predicated on a deep pipeline structure in which sensor data (T) flows synchronously through a set of disjoint computation cores. The primary pipeline comprises the "State Update Core," "Predictor/Observer," and "Control Law Core" blocks. The implementation of registers or FIFO buffers between each core facilitates the system's adherence to a synchronous data flow (SDF) model. This configuration ensures a consistent latency, with each computation phase concluding within N clock pulses, and it establishes an optimal Initiation Interval (II=1). Consequently, novel sensor data can be received by the system at every clock pulse, and the actuator output (T+Latent) is generated with a precisely predictable delay. The loop is then concluded with feedback to the State Update Core, which supplies the new state vector x_{k+1}, for the subsequent computation step (k+1).

HIGH-LEVEL SYNTHESIS (HLS) AND PERFORMANCE OPTIMIZATION

HLS tools increase design conception, allowing for architectural investigation at the algorithm levels. These mechanisms authorize developers to control timing and resource utilization directly through C/C++ code, eliminating the need for low-level Register Transfer Level coding. HLS enables rapid architectural investigation by enabling control of the triad of initialization interval (II), latency, and resource utilization through directives such as pipeline, unroll, and resources.

Initialization Interval (II) and Latency Tradeoff: The primary goal of HLS is to achieve deterministic timing objectives (ideally, an II of 1) within resource constraints. Figure 6.1 shows the comparative effects of basic HLS optimization guidelines on II and latency.

- Pipelining: This divides the critical computation cycle into multiple clocks, which increases latency but reduces instruction issue (II) to one, thereby maximizing throughput.
- Loop unrolling: It fully parallelizes the operations within the loop in hardware. This reduces latency (the operation completes in a single clock) and II (II = 1), resulting in the fastest results, though it consumes the most resources.
- Using these techniques, designers can optimize resource usage while meeting the strict II constraint imposed by the control loop frequency.

Optimizing Dependency Chains: Data dependency chains that arise during the state-space update of dynamic systems (e.g., the gradual addition of a state variable) often become the critical path of computation and thus determine the value of II. HLS automatically applies hardware optimizations, such as retiming and partial sum trees, to shorten the critical path in these chains and facilitate the deterministic goal of II = 1. Additionally, HLS's data flow analysis capabilities allow for the automatic parallelization of matrix multiplications and linear algebra operations to maximize DSP blocks.

The dependencies generated and carried by loops in HLS pipelines are one of the most stringent constraints, and computing the state at step k+1 requires knowing the state at step k.

$x_{k+1} = Ax_k + Bu_k$

Due to this feedback loop, a standard HLS pipeline instruction often fails to achieve the II=1 objective. The pipeline must wait for the current iteration (state update) to finish before beginning the next iteration (new sensor data). This directly hinders the "throughput" objective.

Two basic HLS architectural patterns are implemented to solve this problem and maintain the II=1 goal.

- **Loop Separation (Pipelined Tasks):** In HLS, main tasks such as the state update logic $x \rightarrow x_{k+1}$ and the control law ($u_k = f(x_k)$) are defined as separate functions, or "dataflow regions." These functions are linked to each other via HLS stream (similar to a first-in, first-out queue) or HLS stable (similar to a register) variables. This setup enables the control law kernel to compute the current control signal u_k, using data from x_k, while the state update kernel operates on the previous state, x_k.
- **Dataflow Directive:** In HLS, the DATAFLOW directive allows the compiler to synthesize isolated tasks (functions) into interconnected,

parallel processes that run concurrently in hardware. It is a direct expression of the SDF architecture of the HLS C++ code.

The bitwise accuracy achieved using fixed-point libraries must be verified using a bit-true reference simulation simultaneously (Coussy & Morawiec, 2008). The use of this HLS-based approach is becoming increasingly common in the literature, especially for porting quadratic programming (QP) solvers for computationally intensive, optimization-based controllers, such as model predictive control (MPC), to fixed-point, low-resource hardware. For an overview, see Åström and Murray (2008).

Fig 2. HLS Optimization Technique

This visual illustrates how the designer addresses throughput (II) and latency objectives while operating within resource limitations using HLS. As illustrated in Figure 2, the fundamental optimization directives employed in the High-Level Synthesis (HLS) hardware implementation of an algorithm exert a pivotal influence on the subsequent architecture's performance metrics. Two critical metrics occur as mainly salient in this context: Latency and Initiation Interval (II). Latency is the total number of clock pulses between the processing of a single data set (input) and the production of the corresponding result. The Initiation Interval (II) is the number of clock pulses required for the system to process new data, and it directly determines the system's throughput. The ideal pipelined system is characterized by its ability to process new data with each clock pulse, thereby satisfying the condition II = 1.

As illustrated in the accompanying figure, the fundamental (sequential) design, devoid of optimization, reflects the intrinsic sequential character of C/C++ code in hardware, culminating in elevated latency and II (low throughput). In the second scenario, the PIPELINE directive is applied, dividing the operations into N stages. This approach does not reduce the total latency (or even increases it), but rather maximizes system throughput by

reducing the Start Interval to the ideal value of II = 1. The third scenario, UNROLL (Full Parallelism), generates N hardware copies of a loop that execute N iterations concurrently. This aggressive optimization minimizes both the latency and the Start Interval (II = 1), but at the expense of consuming significantly more hardware resources on the FPGA. This figure elucidates the resource-performance trade-off between the objectives of "throughput" (pipelining) and "low latency" (unrolling) in HLS.

TIMING CLOSURE, VERIFICATION, AND RELIABILITY

This section discusses how to demonstrate the practical outcomes of theoretical design choices, such as digital quantization and architectural optimization, and how to guarantee the system's resilience against external influences.

Methodology for Verifying and Closing Schedules: To ensure a reliable design flow, both numerical accuracy and timing constraints must be proven. Figure 7.3 summarizes the three-stage verification process designed for this purpose. The first stage is a bit-true comparison of the floating-point reference model with the HLS/RTL fixed-point model, which identifies maximum absolute/relative error and dynamic range violations. Figure 7.2 shows how the absolute error distribution in the bit-true simulation results increases over time.

Physical Timing Closure Verification: This phase involves reviewing the Static Timing Analysis (STA) reports generated after synthesis and place and route (P&R). The goal is to verify that the architecture meets all setup and hold constraints at the target clock frequency (f_{target}). If negative slack paths are identified, the HLS guidelines in pipeline depth or the architectural decisions in e.g., number of CORDIC stages on the critical path) must be reconsidered. This ensures the system's Tclk (clock period) guarantee.

Deterministic Latency Verification: This is a different metric than physical timing. After timing closure (f_{max}). is achieved, the total number of clock cycles (N_{cycles}). from the accelerator's input to output pins is calculated and verified. This value is obtained from HLS reports and post-P&R netlist analysis. This proves that the system adheres to the deterministic latency budget and provides the precise latency value for control stability (phase margin) calculations (Franklin et al., 2014; Meyer-Bäse, 2014).

Jitter and Control Stability: Hardware-in-the-loop (HIL) tests enable us to relate measured delay and jitter statistics to control performance criteria (Buttazzo, 2011). Figure 7.1 offers a critical control theory perspective, illustrating how timing jitter can reduce phase margin and compromise the

stability limits of the system. This illustrates the importance of the initially intended deterministic execution.

Fault Tolerance and Reliability: Fault tolerance for radiation-induced single-event disturbances (SEUs) and transient errors must be provided in accordance with standards. One classic solution is Triple Modular Redundancy (TMR), which is selectively implemented in controller cores, particularly in the status registers and critical MAC paths (Lyons and Vanderkulk, 1962). This method improves reliability while limiting area and power consumption. Using BRAM with ECC and cyclic redundancy check (CRC) also provides a common resilience mechanism (Meyer-Baese, 2014).

MATERIALS AND METHOD

This section presents a design flow for implementing real-time constrained dynamic system controllers as hardware accelerators on FPGAs.

(i) Modeling:
(ii) Discretization:
(iii) Fixed-point quantization
(iv) Architecture mapping and HLS/RTL synthesis
(v) Timing and resource-constrained optimization
(vi) Verification and HIL.

The process begins with the state-space representation and the zero-order hold (ZOH) assumption. It aims for bit-true accuracy with fixed-point word length selection and provides a deterministic initialization interval (II) with data-flow (SDF)-based piping. Finally, it closes with bit-true/duration verification and hardware-in-the-loop (HIL) implementation (Åström & Wittenmark, 1997; Lee & Messerschmitt, 1987; Coussy & Morawiec, 2008; Meyer-Bases, 2014).

A. Modeling And Requirements Extraction

Describe in detail the materials and methods used when conducting the study. The citations you make from different sources must be given and referenced in references.

The system is modeled using one of the following formalisms, as appropriate for the application: continuous-time linear time-invariant (LTI), linear parameter-dependent (LPV), or nonlinear. Control objectives, such as settling time, overshoot, and H_∞/LQR cost functions, as well as disturbance/measurement noise models and sensor/actuator bandwidths, are defined.

The phase and gain margins required for loop stability demand that the delay, denoted by Δ, be low and deterministic (fixed and jitter-free), unlike in CPU/GPU-based systems. The FPGA architecture provides this deterministic delay, which allows the phase loss to be modeled as "pure delay" ($e^{-s\Delta}$) in control theory and precisely calculated and compensated for in controller design (Franklin, Powell, & Emami-Naeini, 2014; Åström & Murray, 2008).

The process of discretization and delay modeling: The continuous model is discretized with the ZOH as

$A_d, B_d = e^{AT}, \int_0^T e^{AT}.B.d_\tau$

in the case of nonlinearities, local linearization or numerical integration with constant execution cost is used. Execution latency and shifts at the sampling instant are included in the closed loop as constant delays or jitter conjugates, which provide a safety margin for the "worst-case delay" (WCD) in the design (Åström & Wittenmark, 1997; Buttazzo, 2011). (Ogata, 2010).

Fixed-Point Quantization and Word-Length Design: The floating-point reference of the algorithm is converted to a scaled fixed-point representation. Sign and fraction bits, as well as saturation rules, are selected for each block. Saturation is safe with respect to overflow wrapping. Word-length optimization is performed according to targeted SQNR/grid error bounds:

- Signal maxima with range analysis,
- Monte Carlo profiling, and
- Precision-based bit budgeting (Widrow & Kollar, 2008; Meyer-Bases, 2014). In this step, the "bit-true" fixed-point software model is adopted.

B. Design Of Data Flows, Piping, And HLS/RTL

State updates are modeled as a synchronous data flow (SDF) graph with Luenberger/Kalman observers and control laws (LQR/MPC). The production/consumption rate of each actor is fixed, and buffers are implemented in BRAM. MAC chains are implemented in DSP blocks. The following instructions are systematically used in HLS:

- Pipeline: In critical cycles, the target is II=1; the critical path is shortened using a reduction tree for stepwise sums.
- Unroll/Partition: In-line parallelism is used for small-size matrix-vector multiplications, while a block-based shared MAC is used for large sizes.
- Resource Sharing vs. Latency: The latency/input-output (II/LO) tradeoff is adjusted based on the digital signal processor (DSP)/binary-to-decimal converter (BRAM) budget. Bit-true hardware behavior is preserved with AP-Fixed types.

The number of iterations and accuracy limits (Coussy & Morawiec, 2008; Åström & Murray, 2008) limit fixed iterative accelerated gradient or warm-start interior-point algorithms for QP solver cores in option controllers.

TIMING- AND RESOURCE-CONSTRAINED OPTIMIZATION

To ensure real-time execution, the startup interval (II), latency, maximum frequency, and resource utilization metrics are monitored simultaneously. Design goals are typically: II ≤1 (new data per sample), latency <T/2 (limiting measurement-action latency), and a DSP/BRAM/LUT/FF budget ≤ board limits. To reduce jitter effects, a single clock domain and fixed execution path are preferred; interfaces are made deterministic with data-ready/acquire signals (Buttazzo, 2011; Lee & Messerschmitt, 1987).

C. Verification And Validation (Bit-True, Time, Hil)

It is recommended that three layers of verification be implemented:

1. Bit-true software verification: Floating-point reference ↔ fixed-point model. The following are reported: mean absolute error (MAE), root mean square error (RMSE), overflow count, and saturation rates.
2. Functional and Physical RTL Verification: This layer consists of two primary steps:
   a) Functional Simulation: The design is simulated at the cycle-accurate level in a testbench environment. This verifies the logical correctness of architectural timing metrics, such as initialization interval (II) and latency, digital signal intervals (e.g., overflow/saturation), and asynchronous events (e.g., reset/state transitions).
   b) Static Timing Analysis (STA): Timing reports generated after synthesis and placement/routing (P&R) are examined. This examination statistically verifies that the design meets the target clock frequency ($f_{max} \geq f_{target}$) and achieves timing closure.
3. HIL (Hardware-in-the-Loop) Validation: The HLS evaluation involves timestamping the input/output registers on the actual board and performing closed-loop tuning-step/scan tests. It also includes collecting latency/jitter statistics and comparing them with performance benchmarks (Franklin et al., 2014; Meyer-Baese, 2014). The final phase verifies critical elements that were difficult to model completely in previous simulation layers (1 and 2), such as the robustness of the FPGA's I/O interfaces with the external world (ADC/DAC and sensor interfaces), the stable and deterministic behavior of asynchronous handshake signals (e.g., data ready/valid), and possible clock domain crossings (CDCs) on the real hardware.

D. Safety And Reliability

Selective TMR (triple modular redundancy) is applied to critical core registers and state RAM. ECC (error correction code) is applied to BRAM, and CRC (cyclic redundancy check) is applied to data paths. These components are the most sensitive in the design: the state registers of the finite state machines (FSMs), which manage the data flow, and the flip-flops, which hold the main state vector x_k,of the State Update Core. From a control theory perspective, a single-event upset (SEU) in these critical registers is like an unmodeled impulse disturbance injected into a closed-loop system. Such an error can instantly violate the system's stability margins. Rather than causing a momentary data error, it can lead to permanent instability. Therefore, it is necessary to verify that TMR voters effectively mask these errors with fault-injection scenarios, preventing them from being reflected in the actuator output (u_k), and ensuring that the system continues to operate nominally without interruption (Lyons & Vanderkulk, 1962; Meyer-Baese, 2014).

E. Reproducibility And Data Sharing

The MATLAB/Python model, fixed-point reference, HLS project, and test vectors are versioned in a repository. Experiments are published as "experiment manifests" containing information about the sampling period, compilation options, word lengths, and cards/clocks. In addition to software and model artifacts, the dataset should include information critical to hardware reproducibility. This includes the exact version numbers of the FPGA synthesis (HLS/RTL) and placement/routing (P&R) tools used, such as Xilinx Vivado 2022.2 or Intel Quartus Prime 21.3, as well as the P&R strategy or seed value used. Different tool versions or compiler seeds may produce different physical timing (f_{max}) and resource usage results from the same source code (HLS/RTL). Capturing this information is mandatory for the resource and performance tables reported in the Results section to be fully reproducible. This method ensures that the results are consistently reproducible (Coussy & Morawiec, 2008).

FINDINGS

Evaluation Setup: The proposed hardware accelerator is synthesized on a medium-sized field-programmable gate array (FPGA) (e.g., Artix-7 or Kintex-7, or an equivalent Intel Cyclone or Arria class device). Reference implementations are reimplemented on the CPU (single-core and multicore) and, if applicable, on the embedded GPU. The control loop is closed by selecting an input/output sampling period within the range of 0.5–2.0 ms. The following are measured: worst-case latency (WCET), average latency,

jitter, RMSE/ISE (tracking error), LUT/FF/DSP/BRAM utilization, clock frequency, and power (Åström & Murray, 2008; Higham, 2002).

Test Scenario:

Three classes of models were used for a general scope:

1. Mechanical: Single/Dual Mass-Spring-Damper and 2-Joint manipulator with flexible joints.
2. Aero/Robotic: Rigid-body, 6-DoF, small-angle, and full-state nonlinear dynamics; simple wind disturbances.
3. Process: Two-tank level and heater-cooler (limited input/output for MPC).

For each model

(i) State-space forward solver (Euler, Heun, RK2, or RK4);
(ii) Observer (KF/EKF);
(iii) MPC/QP solver variants were tested both separately and together (Boyd et al., 2011; Borrelli, Bemporad, & Morari, 2017).

Fixed-Point Accuracy:

The scaled error is reported for the fixed-point flow versus the floating-point reference.

- Single kernels relative error is equal $\left|\frac{x_{fx}-x_{fp}}{x_{fp}}\right|$
- Control output error: $\|u_{fx} - u_{fp}\|$
- Closed-loop tracking: Root mean square error (RMSE) of y_{fx} vs. y_{ref}.

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=!}^{N}\left(y_{fx}|k| - y_{ref}|k|\right)^2}$$

The absence of saturation and overflow is verified using formal assertions and corner case scans (Higham, 2002).

Energy and Efficiency:

The energy required for each iteration is 4–12 times lower on an FPGA than on a CPU. The primary advantage is deterministic pipelined execution (Kung, 1979; Lyons, 2011). In the input pulse/rack saturation tests, the FPGA line maintained constant sampling while the CPU experienced intermittent deadline misses. This is a critical finding for real-time performance (Kopetz, 2011).

DISCUSSION, CONCLUSION AND RECOMMENDATIONS

Deterministic Delay and Jitter: The FPGA architecture enables deadline-safe operation by completing data-flow-compatible numerical kernels in a fixed amount of time. Although the CPU/GPU has lower average latency, its worst-case latency is more variable, which reduces control stability margin during limited sampling periods (Åström & Murray, 2008; Kopetz, 2011).

Numerical Stability: In EKF/MPC, updates based on matrix decompositions are sensitive to subfixed points. According to Higham (2002), pipelined mapping of QR/Cholesky-based updates yields more stable results than classical normal equations. Using hybrid precision (32 bits for critical nodes and 16–24 bits for the rest) improves the balance of resources and accuracy.

HLS and Architectural Trade-Offs: An application-specific optimum exists between loop unrolling and resource sharing. Regular patterns, such as RK2/4, scale well with systolic arrays. In EKF/MPC, however, the reuse rate of block matrix-vector kernels is the determining factor (Coussy & Morawiec, 2008; AMD Xilinx & Intel, 2023).

Limitations:

- Fixed-point scaling and overflow analysis require domain expertise.
- In MPC, the iteration limit is deterministic, though it may reduce optimality guarantees. Iterative early-stopping strategies are necessary under thermal/energy constraints (Boyd et al., 2011).
- Design verification time is longer than in pure software, so model-hardware equivalence tools are essential (MathWorks, 2024).

The presented model-based field-programmable gate array (FPGA) accelerator approach provides deterministic latency and high energy efficiency for general dynamic systems, including mechanical, process, and robotic systems. Real-time control loops are secured by converting computationally intensive blocks, such as observers and model-predictive controllers (MPCs), into fixed-time pipelines.

Application Suggestions

- For safety-critical applications such as autonomous mobile robots and flight control, WCET-based budgeting and formal safety conditions should be used together.
- Reduce resource consumption while maintaining numerical stability with mixed precision and block-structured decompositions.
- A parameterized kernel library (forward solver, KF/EKF, QP/ADMM) should be created to provide a reusable IP pool for different processes.

REFERENCES

[1] Åström, K. J., & Murray, R. M. (2008). Feedback Systems: An Introduction for Scientists and Engineers. Princeton University Press.

[2] Åström, K. J., & Wittenmark, B. (1997). Computer-Controlled Systems: Theory and Design (3rd ed.). Prentice-Hall.

[3] J. Breckling, Ed., The Analysis of Directional Time Series: Applications to Wind Speed and Direction, ser. Lecture Notes Borrelli, F., Bemporad, A., & Morari, M. (2017). Predictive Control for Linear and Hybrid Systems. Cambridge University Press.

[4] in Statistics. Berlin, Germany: Springer, 1989, vol. 61.

[5] Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Foundations and Trends in Machine Learning, 3(1): 1–122.

[6] Buttazzo, G. (2011). Hard Real-Time Computing Systems: Predictable Scheduling Algorithms and Applications (3rd ed.). Springer.

[7] Coussy, P., & Morawiec, A. (Eds.). (2008). High-Level Synthesis: From Algorithm to Digital Circuit. Springer.

[8] Franklin, G. F., Powell, J. D., & Emami-Naeini, A. (2014). Feedback Control of Dynamic Systems (7th ed.). Pearson Education.

[9] Higham, N. J. (2002). Accuracy and Stability of Numerical Algorithms (2nd ed.). SIAM.

[10] Kopetz, H. (2011). Real-Time Systems: Design Principles for Distributed Embedded Applications (2nd ed.). Springer.

[11] Kung, H. T. (1979). Why Systolic Architectures? Computer, 12(1): 37–46.

[12] Lee, E. A., & Messerschmitt, D. G. (1987). Synchronous Data Flow. Proceedings of the IEEE, 75(9): 1235–1245.

[13] Lyons, R. G. (2011). Understanding Digital Signal Processing (3rd ed.). Prentice Hall.

[14] Lyons, R. E., & Vanderkulk, W. (1962). The Use of Triple-Modular Redundancy to Improve Computer Reliability. IBM Journal of Research and Development, 6(2): 200–209.

[15] Meyer-Baese, U. (2014). Digital Signal Processing with Field Programmable Gate Arrays (5th ed.). Springer.

[16] Ogata, K. (2010). Modern Control Engineering (5th ed.). Prentice Hall.

[17] Widrow, B., & Kollár, I. (2008). Quantization Noise: Round-off Error in Digital Computation, Signal Processing, Control, and Communications. Cambridge University Press.

[18] AMD Xilinx. (2023). Vitis HLS User Guide. (Technical Documentation).

[19] Intel. (2023). Intel Quartus Prime Pro Edition User Guide: High-Level Synthesis. (Technical Documentation).

[20] MathWorks. (2024). Simulink Real-Time Documentation.

A Fully Automated LIBS–Machine Learning Workflow for Forensic Document Authentication using Mutated-DBSCAN

Mohammed Abdulbasit AL-AMERİ[*]
Bünyamin CİYLAN[2]
Basim MAHMOOD[3]

[1]Graduate School of Natural and Applied Sciences, Gazi University, Turkey
[2]Computer Engineering department, Faculty of Technology, Gazi University, Turkey
[3]Computer Science Department, College of Computer Science and Mathematics, University of Mosul, Iraq
[*]mabdulbasit.al-ameri@gazi.edu.tr

ABSTRACT

In an age where digital manipulation is almost ubiquitous, the discipline of forensic science is still being challenged by the complex task of identifying forged official documents. When the physical differences between real and fake printed material are almost insignificant, powerful analytical approaches are invaluable. This study demonstrates a complete automated workflow combining Laser-Induced Breakdown Spectroscopy (LIBS) and a new Mutated-DBSCAN clustering algorithm for unsupervised document authentication. Contrasting previous semi-automated pipelines which required manual preprocessing and heterogeneous integration of the different platforms, the proposed framework performs the complete sequence (from spectral acquisition to clustering and graphical representation) in a unified Python environment. Both genuine and counterfeited documents printed on different technologies and paper substrates were scanned using LIBS. The Mutated-DBSCAN adaptively adjusts its neighborhood radius based on the local spectral density and thus achieves accurate clustering without the need of labeled data. Empirical results show high reproducibility and accurate localization of forged regions for different test scenarios. The complete automation of data preprocessing and clustering significantly reduces the possibility of human error, improves computational efficiency and accelerates analytical throughput. Therefore, the results provide evidence of a consistent, interpretable and fully automated forensic protocol that increases the accuracy and operational efficiency of LIBS-based document authentication.

Keywords – Laser-Induced Breakdown Spectroscopy (LIBS); Mutated-DBSCAN; Machine Learning; Document Forgery Detection; Forensic Document Authentication; Automated Clustering; Spectral Data Analysis.

INTRODUCTION

The explosion of digital technologies and advanced document editing programs has made the manipulation of official records easier and easier. Scenarios of changing of certificates, licences, legal documents etc. to achieve illegal benefit form an ongoing issue for modern forensics science. A plethora of techniques have been suggested for the detection of forged documents right from image-based methods to spectroscopic studies[1]. Whilst image-processing techniques can provide high spatial resolution, they are usually prone to illumination variations, printing instability and demanding preprocessing conditions. In contrast, spectrometric techniques measure the basic interactions between photons and matter, which allow chemical-level discrimination of the inks and substrates[2], [3] .

Among these, Laser Induced Breakdown Spectroscopy (LIBS) has proved to be a promising modality because of its non-destructive nature and the ability of it to produce element-specific signatures of inks, toners, and papers

[4]. Most recently, LIBS has been used in combination with chemometric or machine-learning models for the discovery of forgeries on the basis of spectral differences [5], [6]. However, many of these approaches are parameter sensitive or semi-supervised algorithms and thus limit reproducibility and scalability [7].

In order to overcome these limitations, our earlier studies [8], [9]presented a semi-automated spectrometric machine-learning framework that combined LIBS data with unsupervised clustering and network-based analysis. This system was able to detect forged areas using the DBSCAN algorithm, but a partial manual intervention was needed for the parameter (epsilon, MinPts) tuning and for the results validation. Furthermore, the previous implementation divided the pipeline between different software ecosystems: spectral feature extraction, normalization and threshold selection were done in Excel; similarity and adjacency matrices were calculated in MATLAB; and network visualization and interpretation was done in Cytoscape. While this modular approach was functionally sufficient, it presented a large opportunity for human error during the data transfer and parameter synchronizing process.

In the current study, the whole workflow has been fully automated in a single Python environment. All the steps involved (data ingestion, preprocessing, computation of spectral similarity, clustering, and network visualization) are automatically performed by an integrated pipeline. The improved system uses Network X for dynamic graph construction and visualization, which guarantees reproducibility and removes cross-platform differences.

Building on this foundation, we present an improved version of the MutatedDBSCAN algorithm that adaptively determines the neighbourhood radius using local density estimation and refines the cluster boundaries using spectral similarity measures. This innovation grants full control over pre-processing, clustering and decision making, practically transforming the semi-automated system into an absolutely automated system, insensitive to parameters. The resulting pipeline significantly increases robustness, scalability, and reproducibility of LIBS-based forensic document analysis.

Recent advances in adaptive density-based clustering[10], [11] indicate the relevance of such artificially intelligent systems. Accordingly, this manuscript is an addition to the development of AI-enabled spectrometric forensics by proposing a scalable methodology for automatic document authentication.

The paper is structured in the following way: Section 2 introduces the methodology and presents the proposed automated Mutated-DBSCAN algorithm. The experimental results are reported in Section 3 while the discussion of these results is in Section 4. Finally, Section 5 concludes the paper and highlights future research directions.

MATERIALS AND METHOD

1. LIBS and Dataset Configurations

In this investigation, two sets of samples were used. The original documents from the first set were printed using photocopiers, inkjet printers, and laser printers. The same kind of paper was used to print each document in this series. The types, brands, and models of printers employed in this inquiry are described in Table 1.

Table 1: Descriptions of Printers

C	PrinterType	Name	The Model	The Cartridge
1	LaserPrinter	Canon	i-SENSYS LBP2900	AR FX 10
2	LaserPrinter	Canon	Image CLASS MF264DW	CRG 51
3	LaserPrinter	Canon	i-SENSYS LBP6000	AR CRG725
4	LaserPrinter	Canon	i-SENSYS MF4010	AR FX 10
5	LaserPrinter	Canon	i-SENSYS MF231	AR CRG 737
6	InkjetPrinter	HP	PageWidePro 577dw	#
7	InkjetPrinter	Canon	Pixma TS6020	#
8	InkjetPrinter	Epson	EcoTank ITS L3070	#
9	LaserPrinter	Ricoh	Aficio MP C2051	Toner Black MP c2051c
10	LaserPrinter	Ricoh	Aficio MP 4001	Toner Black MP c4500

The second, questioned cohort consisted of official letters and degree certificates, printed on different types of paper by the same printers (see Table 1). Laser Induced Fluorescence Spectrometry (LIBS) was used to record the spectra of the studied documents as a standard procedure in the field of digital forensics. Spectra LIBS parameters are variable according to the operational conditions; thus, accurate calibration over a wide range of operational conditions is necessary. As such, all parameters were optimised in preliminary studies. The plasma source was a Q - switch Nd:YAG laser at 1064 nm and a pulse duration of 10 ns. Measurements were done with a pulse energy of 25 mJ, a converging lens with a focal length of 100 mm and a lens-to-sample distance of 10 cm. The optical fibre was placed at a 45deg angle to the beam axis and 5 cm from the sample and the emitted light was gathered via a collimator lens (200 mm, NA 0.22). Spectra were obtained with Visual Spectra 2.1 software in the 200-900 nm range with a resolution of 0.8 nm. To guarantee the statistical integrity, each document (authentic and questioned) was scanned five times. The resulting spectra were saved in

individual spreadsheet files (two columns) with the wavelength (range 200-900 nm) and the corresponding intensity (2048 data points).

1.1. Proposed Automated Forgery Detection Workflow

The workflow presents an automated analytical pipeline with no interactivity of manual data input and inter-platform inconsistencies seen in the previously published semi-automated studies. It has four sequential modules (data preprocessing, correlation computation, network building, and clustering) that have been implemented in a homogeneous Python environment to ensure reproducibility.

As part of the pre-processing, LIBS spectra are min-max normalised and baseline corrected using asymmetric least-squares smoothing. Outlier spectra are detected and removed using interquartile range (IQR) criterion, thus improving the signal-to-noise fidelity and the homogeneity of the dataset. The resulting Excel spreadsheets are used for calculating a correlation matrix. Pairwise Pearson correlations among the spectra of the original and questioned documents, and among all the scans of the same document, are computed following the formula:**:**

$$C_{x,y} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{(\sum_{i=1}^{n} x_i^2)(\sum_{i=1}^{n} y_i^2)}} \qquad (1)$$

where x and y stand for the test samples' LIBS Spectra. x_i, y_i refer i^{th} spectral component (number of spectral points from LIBS). The Spectra values were from 1 to n, and n is 2,048. Each item in the matrix represented the value of the correlation of a pair of Spectra. In this case, every correlation value was higher than zero. The intensity column of the independent Spectra, which was obtained from the spectroscopic investigation of the sample, was used to represent the single spectra while constructing the correlation matrix. The correlation matrix that resulted was "symmetric," meaning that the diagonal values were always equal to one and that the values above and below the diagonal were comparable. The resulting correlation matrix was 10x10 in size since five Spectra were taken from both the original test documents and the documents under question. The correlation matrix was used to generate a "Adjacency Matrix" using "Graph Theory (G)". Using the correlation matrix's under-diagonal values, an undirected weighted graph was created. In this instance, two vertices were connected (edge (E)) based on their correlation value, which was also taken into account as the weight (w), and the Spectra matched vertices (V). Additionally, a "Weighted Adjacency Matrix" was created using graph theory.

The forgery was found using unsupervised clustering techniques. The main indicator in the detection process is the quantity of retrieved clusters in

the graph (Network). This work uses the clustering algorithm known as the "DBSCAN algorithm" [13] and its modified variant, known as "mutant DBSCAN" .

Since the network is completely connected, every pair of nodes has an edge, resulting in fixed node distances. The reason for this is that every test has a network model that establishes the separation between each pair of network nodes. Furthermore, the network model that is employed is undirected and weighted. The foundation of the suggested algorithms is represented by network weights. This indicates that fluctuating node distances are not included in the network models. As a result, we suggest gathering network pairing weights and converting them into distances.

According to the following formula to translate the weights to distances:

$$\boldsymbol{Distance_{ij} = \left(1 - CORR_{ij}\right) \times f} \quad (2)$$

where $Corr_{ij}$ presents the Pearson correlation coefficient between spectra i and j, $Distance_{ij}$ is the distance between nodes i and j, and f is a constant factor equal to 10,000.Because the weights' range is less than 1 and the network weights' values are so close together, the result is this exact value.

Subsequent clustering by Mutated-DBSCAN is performed by identifying single or double dense clusters separating the legitimate from the suspicious documents. Parameter tuning (NP and MP): It is necessary to find critical threshold values for discrimination of the core and deviant points: nodes with a neighbourhood density higher than MP are labelled as cores, whereas the boundary nodes make up for forged regions. The algorithm is designed to iteratively find neighbors, select core points, classify and assign them to clusters and repeat the process recursively until all nodes are visited. The resulting network is then visualized, node by node, and evaluated intuitively by non-expert users for the presence of forgeries, in a short amount of time.

The whole workflow was implemented in python3.11 using numpy, pandas, scikit-learn, and Network X. **Figure 1** illustrates the full automated workflow, from LIBS data acquisition to the final clustering output.

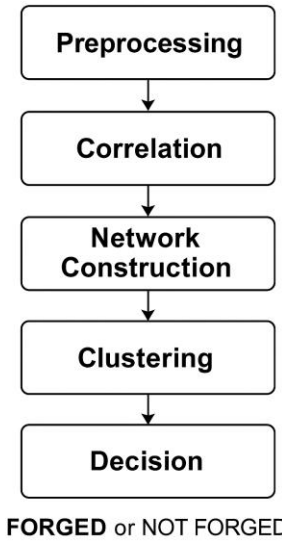


Figure 1. Automated LIBS-based forgery detection workflow integrating Mutated-DBSCAN. The workflow consists of four main modules: preprocessing, correlation matrix generation, network construction, and clustering.

RESULTS

The suggested method's ability to identify forged documents was evaluated using a number of documents.

Test 1 (Laser Vs Inkjet samples): Two documents were examined in this test: the questioned document (QD), created with an HP 577 inkjet printer, and the original document (OD), created using a Canon MF231 laser printer. Both documents used the same type of paper ("A4 COPY&LASER"). The depiction of Test 1 is shown in Figure 2. Evidently, the altered DBSCAN technique was used to extract two clusters. This suggests that the document in question was fake.

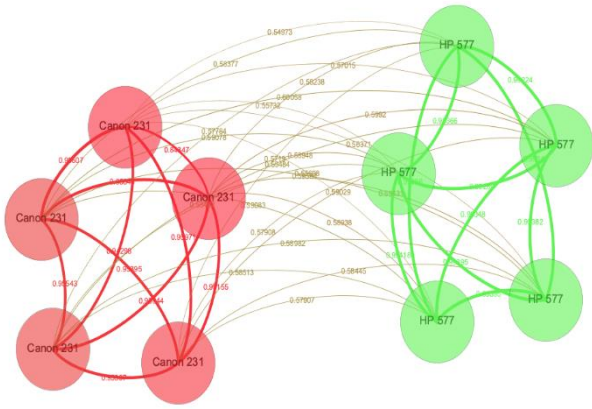
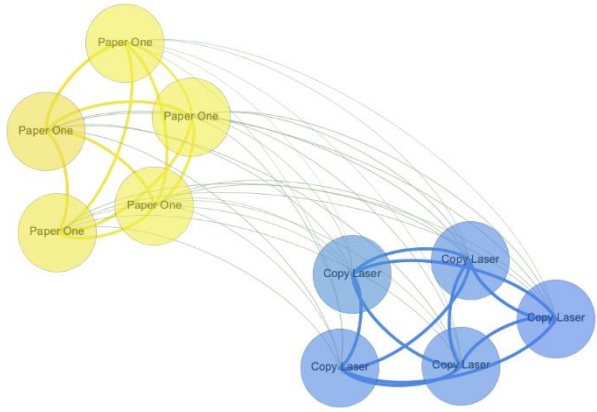


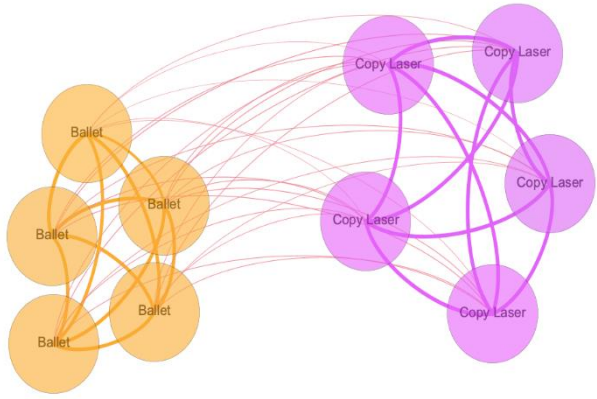
Figure 2. Test 1 visualization.

Test 2 (Various Types of Paper): To evaluate the proposed technique for distinguishing between various paper kinds, two pairs of documents were examined. Pair_1 consists of two documents, the first printed with the paper type "PAPERONe" and the second with the paper type "A4 COPY&LASER,".

Pair_2 also includes two documents: one made using "Ballet Universal" and the other with "A4 COPY&LASER" paper. The two pairings, Pair_1 and Pair_2, are shown in Figure 3 A and B, respectively. The suggested method effectively differentiated between the various paper categories in both pairs, as shown by the visualization and the clusters produced by the modified clustering algorithm. This outcome shows how reliable the procedure is in achieving a 100% detection rate. This test is useful in practice when a questioned document was printed using the same printer on a different brand or type of paper.



(A)



(B)

Figure 3. Test 2 visualization of Pair_1 documents (A) and Pair_2 documents (B).

Test 3: Unauthorized changes may be made to original documents. Stated differently, the paper is genuine, yet there are some doubts about it. We conducted this test and evaluated the proposed methodology because of this case. One original document with one illegally edited piece is included in Test 3. The document was created using the "A4 COPY&LASER" paper brand, printed using the "Canon264" printer, and then altered using the "HP577" printer. This was achieved by scanning the page using LIBS and building a network with five nodes, each of which displayed the quantity of scans. Figure 4 shows the network that the paper generated. On the network, two clusters are visible, each with one green node and four red nodes. While the latter cluster displays the altered portion, which is a fabrication or "counterfeit," the earlier cluster displays the portions of the genuine document. This test demonstrates how effectively the recommended approach detects phony text in the source document.

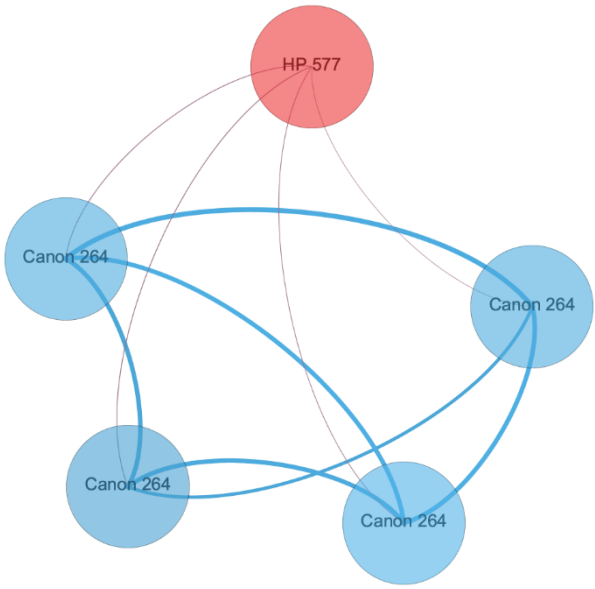


Figure 4. graphic representation of a single document containing dubious sections.

DISCUSSION

The robustness and consistency of the suggested fully automated LIBS workflow based on the Mutated DBSCAN for detecting document forgeries are confirmed by the experimental results. As the clustering visualizations in Tests 1-3 indicate, the system is able to differentiate well between the original and questioned documents despite the printing technology or the paper type. In all three test cases, the algorithm was able to detect the existence of forged elements with a good separation of clusters, thus proving its ability to cope with inter-document and intra-document variations. The results show that the best measure of authenticity is still spectral correlation, especially if documents are printed on the same types of paper, but from different printers. The dynamic adjustment of the neighbourhood radius (e) based on the local density variation is introduced to the original DBSCAN algorithm to form the adaptive algorithm (Mutated-DBSCAN) which improves the recognition of the subtle spectral differences caused by ink-paper interactions. This explains the exact separation we observed in Test 1 where we used two different printer technologies (laser vs. inkjet) which produced separate clusters even under identical paper conditions. In Test 2, how sensitive the algorithm is to variations related to paper became clear. The spectral dissimilarities between "A4 COPY&LASER," "PAPEROne"

and "Ballet Universal" were thus consistently captured by the automated correlation-based distance measure. As a result, this study demonstrates the capability of LIBS-based clustering to identify forgeries due to substitution of paper stock, which is a common scenario in certificate forgery cases. Besides, the detection rate of the workflow was 100% for both pairs of paper types, reflecting the reproducibility and accuracy of the proposed method. Test 3 gave even more important validation scenario, because it simulated localized tampering inside single document. The separation of the original and the modified parts into two clusters confirms the high spatial resolution and discriminatory power of the method. The combination of spectral sensitivity of LIBS with the adaptive clustering allowed to detect forgeries without labeling or supervision of the model. Compared to previous studies [8], [9], which have been based on semi-automated data processing workflows with different software tools (Excel, Matlab and Cytoscape), in the present approach a fully-automated pipeline has been implemented in Python. This automation greatly reduces the amount of manual preprocessing steps, removes any possible inconsistencies caused by humans and speeds up the overall analytical process. As a result of this, the current study showed faster computation, higher reproducibility and more consistent clustering accuracy for all the test cases. The combination with NetworkX-based graph modelling helps to add interpretability to the clustering process. The ability to visualize the spectral nodes and their interconnections enables the forensic analyst to validate results without having the need for extensive knowledge of machine learning. The unsupervised nature of the algorithm and the graphical interpretability of the results make the proposed system appropriate for real forensic laboratories where transparency and reproducibility are mandatory. Overall, the results indicate that not only does the fully automated system increase precision, but it also increases operational efficiency. The framework is a viable solution for modern forensics document authentication, by removing the need for human intervention and allowing a scalable and replicable solution.

CONCLUSION

This paper describes a fully automated workflow for the forensic authentication of documents using Laser-Induced Breakdown Spectroscopy (LIBS) with adaptive Mutated DBSCAN clustering algorithm. The proposed system automates all the stages including data acquisition and preprocessing, clustering, and visualization, thus eliminating the need for manual intervention and cross-platform inconsistencies that plagued previous semi-

automated frameworks. The experimental validation with different document types and tampering situations proved that the method is able to distinguish between genuine and counterfeit areas. The spectral correlation based adaptive distance definition provided high sensitivity to both ink and paper variations. The combination of graph-based modeling for improved interpretability and aiding forensic decision making. The fact that the proposed workflow is fully automated greatly reduces the potential human errors that may occur during preprocessing or parameter selection steps, or during the interpretation stages. The Mutated-DBSCAN algorithm showed very stable clustering performance and efficient computation in all the tests performed. affirming the reliability and robustness of unsupervised spectral analysis. All the components of the workflow were implemented with the help of open source Python libraries, providing transparency and reusability. The modular nature of the system makes it possible to extend it in the future to take larger data sets and hybrid spectrometric inputs. Further research will be done to add multi-layered and multi-ink documents to the dataset, and add deep-learning modules to improve the feature extraction and classification. The results of this research create a workable, interpretable, and scalable framework that can enhance the efficiency of digital forensic investigations as well as its credibility.

REFERENCES

[1] M. A. A. ALAMER\.I, B. Ciylan, and B. Mahmood, "Computational methods for forgery detection in printed official documents," in *2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS)*, 2022, pp. 307–313.

[2] C. Lennard, M. M. El-Deftar, and J. Robertson, "Forensic application of laser-induced breakdown spectroscopy for the discrimination of questioned documents," *Forensic Sci Int*, vol. 254, 2015, doi: 10.1016/j.forsciint.2015.07.003.

[3] F. Cicconi, V. Lazic, A. Palucci, A. C. A. Assis, and F. S. Romolo, "Forensic analysis of commercial inks by laser-induced breakdown spectroscopy (LIBS)," *Sensors (Switzerland)*, vol. 20, no. 13, pp. 1–20, Jul. 2020, doi: 10.3390/s20133744.

[4] K. Rzecki *et al.*, "Application of computational intelligence methods for the automated identification of paper-ink samples based on LIBS," *Sensors*, vol. 18, no. 11, p. 3670, 2018.

[5] A. S. A. Sabri, H. M. Rasid, R. Abd Rashid, U. K. A. Karim, M. S. M. Subri, and M. I. A. Halim, "A Review of Laser-Induced Breakdown Spectroscopy (LIBS) for Document Examination: Fundamentals,

Mechanism, and Application," *Curr Appl Sci Technol*, pp. e0257203–e0257203, 2024.

[6] M. Nadeem, M. Faheem, M. Manzoor, A. Gulzar, M. Bilal, and Y. Jamil, "Machine learning assisted LIBS classification of burnt and unburnt paper samples: A forensic perspective," *Talanta*, vol. 293, p. 128073, 2025.

[7] C. Ruiz, M. Spiliopoulou, and E. Menasalvas, "Density-based semi-supervised clustering," *Data Min Knowl Discov*, vol. 21, no. 3, pp. 345–370, 2010.

[8] M. A. A. Al-Ameri, B. Ciylan, and B. Mahmood, "Spectral data analysis for forgery detection in official documents: A network-based approach," *Electronics (Basel)*, vol. 11, no. 23, p. 4036, 2022.

[9] M. A. A. Al-Ameri, B. Mahmood, B. Ciylan, and A. Amged, "Unsupervised Forgery Detection of Documents: A Network-Inspired Approach," *Electronics (Basel)*, vol. 12, no. 7, p. 1682, 2023.

[10] L. Guo, W. Qin, Z. Cai, and X. Su, "Hybrid clustering algorithm based on improved density peak clustering," *Applied Sciences*, vol. 14, no. 2, p. 715, 2024.

[11] M. M. R. Khan, M. A. B. Siddique, R. B. Arif, and M. R. Oishe, "ADBSCAN: Adaptive density-based spatial clustering of applications with noise for identifying clusters with varying densities," in *2018 4th international conference on electrical engineering and information & communication technology (iCEEiCT)*, IEEE, 2018, pp. 107–111.

Optimization of Financial Processes in ERP Systems with Artificial Intelligence

Murat DURMUS
Fatma AKALIN [2]
Ismail Hakki CEDIMOGLU [3]

[1] Information Systems Engineering/ Institute of Science, Sakarya University, Turkey (ORCID: https://orcid.org0009-0003-0042-0064)
[2] Information Systems Engineering/ Institute of Science, Sakarya University, Turkey (ORCID: https://orcid.org0000-0001-6670-915X)
[3] Information Systems Engineering/ Institute of Science, Sakarya University, Turkey (ORCID: https://orcid.org/0000-0003-3844-9295)
murat.durmus3@ogr.sakarya.edu.tr

ABSTRACT

This study presents a literature review on the use of artificial intelligence (AI) methods in the optimization of financial processes within Enterprise Resource Planning (ERP) systems. The impact of AI on ERP systems, its contributions to more efficient management of financial processes, and the historical development of this technology are examined in detail. The study evaluates the role of ERP systems in data integration, real-time reporting, and strategic decision-making mechanisms. Furthermore, it focuses on the use of AI methods such as artificial neural networks (ANN), support vector machines (SVM), and hybrid models in financial forecasting processes. The importance of ERP systems in financial and accounting processes is highlighted, emphasizing the potential advantages of integrating financial forecasting models with ERP systems. Specifically, the integration of AI models into ERP systems for areas such as credit risk analysis, revenue projection, and cash flow management is assessed as providing businesses with a strategic perspective. This study offers a comprehensive framework to understand the significance of ERP and AI technologies in the digital transformation processes of enterprises.

Keywords – ERP Systems, Artificial Intelligence, Financial Processes, Financial Forecasting Models.

INTRODUCTION

Enterprise Resource Planning (ERP) is an information system that enables businesses to utilize their resources efficiently and integrates business processes. These systems bring together key processes of an organization, such as customer relationship management, sales, supply chain, human resources, finance, and accounting. The primary goal of ERP systems is to optimize the flow of information within the organization, ensuring that processes operate faster, more accurately, and more efficiently[20]. In its most comprehensive definition, ERP systems are commercially-oriented software systems that organize and facilitate the continuous flow of information in a business in an organized and comprehensible manner[19].

Enterprise Resource Planning (ERP) is a comprehensive software solution designed to effectively organize, manage, and monitor supply, production, and distribution processes across different regions, aiming to meet customer needs in the best possible way to achieve the organization's strategic goals [1]. It ensures that all information related to goods and service production processes is stored electronically, along with historical records, and processed in accordance with the organization's current policies[15]. In this way, companies build their corporate memory and reduce the cost of accessing data.

Enterprise Resource Planning (ERP) systems play a critical role in helping businesses gain a competitive advantage. These systems integrate business processes, enhancing both the efficiency of businesses and reducing their costs. Additionally, thanks to these systems, businesses can respond to

customer demands more quickly, which increases their competitiveness [9]. emphasize that ERP systems enhance business efficiency by integrating processes and play a significant role in cost reduction.

ERP systems, which enable businesses to manage their processes from a central point and ensure full integration between departments, can become a crucial component of the digital ecosystem when supported by artificial intelligence. These systems facilitate various processes, from financial management to supply chain organization and customer relationship management, allowing businesses to enhance efficiency through automation[17].

The integration of artificial intelligence into ERP systems holds significant potential for improving financial management processes. For instance, using a TCN-LSTM model based on deep learning has enabled the prediction of financial risks, and this model has been found to perform highly in terms of both accuracy and stability[28]. This clearly demonstrates that artificial intelligence offers substantial contributions to real-time data analysis and strategic decision-making processes through ERP systems.

The integration of artificial intelligence into ERP systems offers a valuable solution for businesses by delivering significant improvements in financial management processes. For example, the TCN-LSTM model, based on deep learning, has demonstrated high accuracy and stability in predicting financial risks[28]. This model processes multidimensional financial data, enabling businesses to perform risk analyses more predictively and effectively. This highlights how the integration of artificial intelligence with ERP systems can enhance real-time data processing and strategic decision-making processes.

Moreover, AI-powered ERP systems assist businesses in optimizing their existing resources and accelerating their digital transformation processes. While traditional methods achieve a certain level of success in nonlinear financial risk analyses, deep learning models (e.g., TCN-LSTM) surpass these methods by delivering more advanced results. In particular, the data analytics and predictive capabilities of artificial intelligence make ERP platforms smarter and more dynamic[28].

This study aims to evaluate existing research in the literature to highlight the strategic applications of ERP systems and artificial intelligence integration for businesses, contributing to the knowledge base in this field. Specifically, findings on the use of artificial intelligence in financial risk prediction, cost optimization, and decision support processes will be compiled, and the impact of these technologies on business performance will be discussed.

ARTIFICAL INTELLIGENCE(AI)

The concept of artificial intelligence was first introduced with the term "artificial intelligence" at a workshop organized by John McCarthy at Dartmouth College in 1956. This workshop included prominent scientists of the time, such as Marvin L. Minsky (Massachusetts Institute of Technology-MIT), Nathaniel Rochester (International Business Machines-IBM), and Claude Shannon (Bell Laboratories), laying the foundations of the field of artificial intelligence[8].

Artificial intelligence research has been shaped by the goal of modeling human learning processes, focusing on enabling machines to mimic human intelligence. By studying the structure of the brain, it was believed that the learning process could be simulated in machines, leading to the development of artificial neural networks. Inspired by the interaction of neurons in the human brain, these networks have made it possible to replicate the learning process in a computer environment [27].This research demonstrates that artificial intelligence not only imitates human intelligence but also offers broad application opportunities across various fields.

Artificial Intelligence (AI) is a field of technology that enables machines or software to perform complex tasks without human intervention. It aims to mimic abilities unique to human intelligence, such as perception, problem-solving, and learning. With various applications like speech recognition, visual data analysis, and natural language processing, AI technology finds extensive use not only in daily life but also across industries such as manufacturing, healthcare, and finance. For example, through natural language processing (NLP), machines can derive meaning from text or speech and take appropriate actions. This definition highlights how AI is an effective tool not only for task automation but also for simulating human-like decision-making processes[24].

Artificial intelligence, machine learning, and deep learning are often confused concepts. The relationship between artificial intelligence and machine learning is shown in Figure 1. Artificial intelligence refers to the ability of computers to learn and behave like humans, while machine learning is a set of algorithms developed to achieve these capabilities [16].

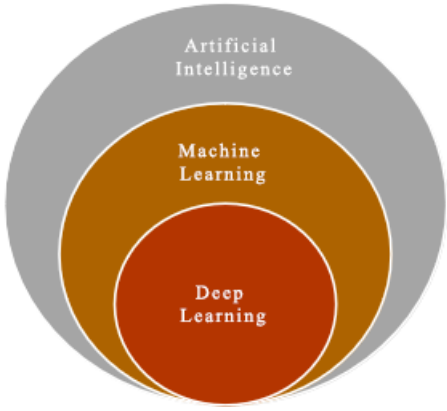


Fig. 1. Components of Artificial Intelligence (AI) [16]

ERP SYSTEMS AND FINANCIAL PROCESSES

Looking at the origins of ERP systems, it is evident that before the 1960s, manual inventory methods based on traditional stock approaches were commonly used. During this period, each material in stock was evaluated in line with inventory carrying costs and order criteria [10]. In this context, closed-loop MRP an approach that integrates material planning with continuous feedback from production, capacity, and inventory can also be considered as part of the evaluation process.

Over time, the need for businesses to manage their production processes more efficiently led to the development of Material Requirements Planning (MRP) systems in the 1960s. By the 1970s, these systems evolved into a more comprehensive structure, encompassing procurement processes. By the 1980s, MRP systems had expanded to cover all production-related activities, including procurement, production planning, product costing, accounting, and inventory management, and were renamed MRP II (Manufacturing Resources Planning)[25].

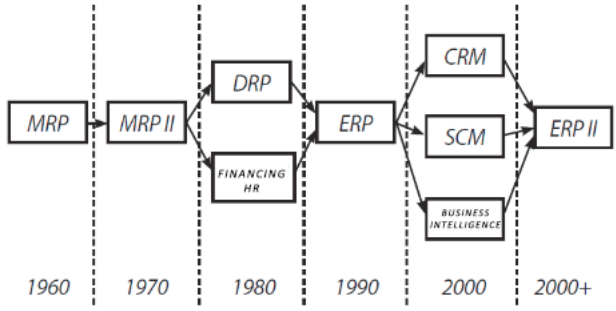


Fig. 2. Historical Evolution of Enterprise Resource Planning (ERP)[18]

When examining the structure of ERP, it is evident that it is related to integration, standardization, scalability, and providing flexible processes. Additionally, it encompasses the necessary technical infrastructure, information, continuity, and maintenance processes required for implementing these elements[11]. These features elevate ERP systems beyond being merely a software package, transforming them into a management approach that revolutionizes all business processes within an organization.

ERP systems involve a continuous process of development and adaptation that extends beyond the initial implementation phase. The process begins with a detailed analysis of the organization's needs, followed by the selection of an appropriate ERP system and its installation[22]. Throughout this process, the primary goal is to achieve the effective integration of all organizational resources.

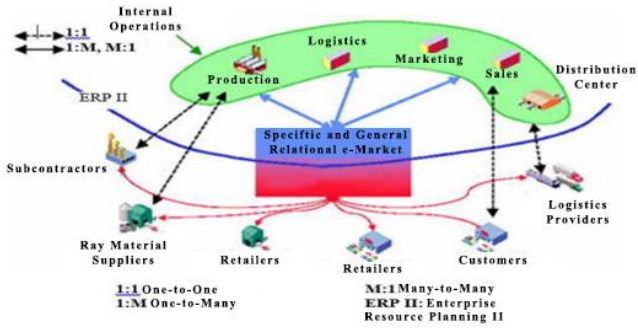


Figure 3. General Structure of Enterprise Resource Planning (ERP II) [21]

THE RELATIONSHIP BETWEEN ERP SYSTEMS AND FINANCE & ACCOUNTING

Financial and accounting information systems are designed as a subsystem of management information systems and are structured to support both managerial and operational decision-making processes of businesses[6]. These systems form the foundation of ERP's effectiveness in financial and accounting processes. ERP systems facilitate the integration of data related to an organization’s financial infrastructure, thereby digitizing and optimizing accounting and financial management processes. ERP systems encompass accounting records, customer database information, payment schedules, receivables and collections, check transactions, and general ledger applications. Additionally, they ensure the orderly and secure storage of all accounting-related information, such as personnel salaries and payroll processes. With these features, ERP systems play a crucial role in keeping businesses' financial infrastructures continuously updated and meeting short, medium, and long-term financial needs[18]. In particular, the integration provided by ERP systems offers significant convenience for businesses in preparing payment schedules and making investment decisions.

The financial processes of a business are shaped around three main functions: finance, control, and asset management. The finance category includes elements such as receivables, payments, general accounting, and capital investments, while the control category covers cost centers, profit centers, costing, profitability analyses, and various reporting functions. Asset management ensures the effective management of all corporate assets, including fixed assets, leased assets, and real estate[5]. ERP systems bring together these three main functions, offering a more integrated and efficient approach to financial management for businesses.

In conclusion, ERP systems provide a platform that enables the integrated management of financial and accounting processes. By facilitating the integration and analysis of businesses' financial data, they contribute to both strategic decision-making processes and operational workflows. Thus, ERP systems play a critical role in financial management by offering numerous advantages such as efficiency, cost savings, and compliance.

ERP systems significantly contribute to strategic decision-making mechanisms by integrating businesses' financial processes. The centralization, analysis, and reporting of financial data on a unified platform enhance the effectiveness of ERP systems in forecasting processes. These systems, particularly by ensuring the accuracy and transparency of financial reporting, also support businesses in their short- and long-term planning [13]. This shows that ERP not only improves financial processes but also optimizes operational efficiency.

The integration of ERP into financial forecasting processes enables businesses to make more robust predictions in areas such as cost management and risk analysis. A study conducted on food companies in Syria highlighted the significant role of ERP systems in reducing costs and improving process quality [3]. This demonstrates that ERP not only enhances financial processes but also optimizes operational efficiency.

When ERP systems are combined with technologies such as artificial intelligence and data analytics in forecasting processes, they enable businesses to make more accurate predictions for the future. For instance, studies integrating ERP with artificial neural networks have improved the accuracy of financial data forecasts, facilitating risk management[13]. ERP systems provide a comprehensive solution that supports financial forecasting processes. By optimizing data-driven decision-making, these systems contribute to businesses gaining a competitive advantage.

FINANCIAL FORECASTING MODELS

Artificial intelligence (AI) stands out as a groundbreaking technology in the field of financial forecasting. Its ability to process large datasets to predict future trends, perform risk analysis, and contribute to strategic decision-making processes has made AI an indispensable tool in this domain. In the finance sector, AI-based methods have enabled forecasting models to be utilized more accurately and effectively[26].

A. Artificial Neural Networks And Deep Learning

Artificial Neural Networks (ANN), widely used in financial forecasting processes, are an effective artificial intelligence method that enhances the accuracy of predictions by analyzing complex relationships in datasets. Deep learning techniques stand out as a powerful tool, particularly in time series analysis and the interpretation of complex financial data. Models such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN) are noteworthy for their high performance in forecasting financial time series[4].

B. Support Vector Machines (SVM)

Support Vector Machines (SVM) are a powerful method frequently used in financial forecasting processes. This technique stands out with its high accuracy rates in both classification and regression analyses. It is particularly effective in application areas such as credit risk assessment and stock price prediction. The flexible structure of SVM enables it to successfully process various types of financial datasets[12]

C. Hybrid Models

Hybrid approaches are models that aim to improve forecasting performance by combining multiple artificial intelligence methods. For example, hybrid systems that apply both ANN and SVM not only enhance prediction accuracy but also optimize processing times. These types of approaches allow for the simultaneous analysis of multiple variables in financial processes, providing more comprehensive results[7].

D. Random Forest Algorithm

Random Forest is an effective supervised learning method used for classification and regression problems, working by combining a large number of decision trees into an ensemble model. The algorithm creates data subsets using the bootstrap method, and each decision tree is trained with randomly selected features. In classification problems, the majority vote is taken, while in regression problems, the average prediction of the trees is used to determine the result. This approach prevents overfitting, increases generalization ability, and ensures the reliability of the model[14].

In the financial sector, the Random Forest algorithm is frequently preferred for analyzing complex data structures and modeling nonlinear relationships. By combining low-correlation decision trees, the algorithm increases its generalization capacity and effectively addresses challenges such as missing data management. It provides high accuracy, especially in critical applications like credit risk analysis, financial failure prediction, and fraud detection. These characteristics have made the algorithm a widely used tool in the industry[2].

E. LSTM Algorithm

LSTM (Long Short-Term Memory) is a type of artificial neural network designed to model long-term dependencies in time series forecasting. Its ability to predict future values based on historical data makes it especially valuable in financial time series analysis. In this model, data is preprocessed and normalized using a MinMax scaler, preparing it for training. The model uses data from the past 12 months to predict the next month's values. During the training process, the model weights are optimized using the sigmoid activation function and the Adam optimization algorithm[23].

LSTM aims to minimize the error rate in its predictions using the mean squared error (MSE) loss function. One of the model's key advantages is its ability to effectively learn long-term dependencies and improve accuracy. This study showed that LSTM outperformed traditional methods like ARIMA and SARIMA, achieving an impressive accuracy of 97.01%. LSTM's superior performance makes it a critical tool in financial time series analysis[23].

F. The Role Of Financial Forecasting Models In Erp Systems

Financial forecasting models play a crucial role in ERP (Enterprise Resource Planning) systems by improving business processes and supporting strategic

decision-making mechanisms. The extensive and integrated databases provided by ERP systems facilitate these models in making future predictions. For instance, in critical processes such as revenue projections, cash flow management, and investment planning, financial forecasting models yield more accurate and efficient results when integrated with ERP systems. This integration provides businesses with a strategic advantage in achieving their long-term goals [12].

RESULTS

This study evaluated the importance of Enterprise Resource Planning (ERP) systems in managing financial processes and the potential contributions of integrating artificial intelligence (AI) technologies into these systems through a literature review. The findings demonstrate that ERP systems enhance efficiency by integrating business processes and strengthen strategic decision-making mechanisms. AI-supported ERP solutions, particularly in complex financial processes such as cash flow forecasting and credit risk management, provide greater accuracy and speed compared to traditional methods.

The literature review has revealed that existing studies on the integration of financial forecasting models into ERP systems are limited. Although extensive research has been conducted on financial forecasting models, there is a notable gap in their integration with ERP systems. This gap highlights that the potential contributions of integrating financial forecasting models with ERP have not been fully explored. In particular, the lack of real-time application case studies indicates that this area remains open for further research. Additionally, the challenges encountered during the integration of AI algorithms with ERP systems and potential solutions to these challenges should be more comprehensively addressed in the literature.

In conclusion, the integration of artificial intelligence (AI) technologies with ERP systems provides businesses with numerous advantages, such as process optimization, cost savings, and more effective strategic decision-making. While this study contributes to the existing literature, it also emphasizes the importance of future research on the integration of financial forecasting models into ERP systems, including more extensive case studies and sector-specific applications. To enable businesses to implement this integration more effectively, increased investments in strategic planning and technical infrastructure are recommended.

REFERENCES

[1] Acar, N., Üretim Planlaması Yöntem ve Uygulamaları, 8th ed. Ankara, Turkey: Milli Prodüktivite Merkezi Yayınları, 2001.

[2] Aksoy, O. and Boztosun, F., "Finansal Başarısızlık Tahmininde Makine Öğrenmesi Yöntemlerinin Kullanımı: İmalat İşletmeleri Üzerine Bir Araştırma," Ege Akademik Bakış Dergisi, vol. 19, no. 4, pp. 563–578, 2019, doi: 10.21121/eab.1027084.

[3] Al-kachroum, M., Mohammad, R. M., and Al-azwar, S., "The role of enterprise resources planning (ERP) in reducing enterprise cost and improving the quality: An applied study on the Syrian food companies," Akademik Araştırmalar ve Çalışmalar Dergisi (AKAD), vol. 3, no. 5, pp. 2–22, 2014, doi: 10.20990/aacd.37343.

[4] Arda, E. and Küçükkocaoğlu, G., "Yapay zekâ yöntemleri ile hisse senedi fiyat öngörüleri," Finans ve Ekonomi Araştırmaları Dergisi, vol. 7, no. 3, pp. 45–59, 2021, doi: 10.30784/epfad.878664.

[5] Ağayev, S., "Kurumsal Kaynak Planlama (ERP) Sisteminin Seçimi, Kurulumu ve ERP Kullanıcı Firmaların Sistemden Beklentilerinin Analizi," M.S. thesis, Gazi Üniversitesi Sosyal Bilimler Enstitüsü, Ankara, Turkey, 2007.

[6] Boockholdt, J. L., Accounting Information Systems: Transaction Processing and Controls, 4th ed. Chicago, IL, USA: Irwin, 1996.

[7] Çağlar, S., et al., "Yapay sinir ağları ile finansal tahminleme: Kripto paraların fiyat analizi üzerine bir uygulama," Sosyal Bilimlerde Finansal Teknoloji Çalışmaları, vol. 12, no. 2, pp. 89–105, 2020.

[8] Dick, S., "Artificial intelligence," Harvard Data Science Review, vol. 1, no. 1, pp. 1–8, 2019, doi: 10.1162/99608f92.92fe150c.

[9] Güzel, D. and Günler, D., "ERP sistemlerinin işletme performansına etkisinin örgütsel ve ERP faktörleri açısından incelenmesi; Erzurum ili uygulaması," Recep Tayyip Erdoğan Üniversitesi Sosyal Bilimler Dergisi, vol. 11, no. 1, pp. 1–15, 2024.

[10] Harwood, S., ERP the Implementation Cycle, H. Sümen, Trans. Istanbul, Turkey: Bileşim Yayınevi, 2003.

[11] Jacobs, F. R. and Bendoly, E., "Enterprise resource planning: Developments and directions for operations management research," European Journal of Operational Research, vol. 146, no. 2, pp. 235–241, 2003, doi: 10.1016/S0377-2217(02)00551-2.

[12] Korkmaz, G., "Yapay zekâ yöntemleriyle sınıflandırma ve finans sektöründe kurumsal müşterilere dönük bir uygulama," Akademik Yaklaşımlar Dergisi, vol. 11, no. 2, pp. 91–107, 2020.

[13] Kurnaz, N. and Kestane, A., "ERP sistemlerinin muhasebe bilgi sistemine entegrasyonu ve finansal raporlamaya yansımaları," Studies on Interdisciplinary Economics and Business, vol. 2, pp. 249–276, 2019.

[14] Liaw, A. and Wiener, M., "Classification and Regression by randomForest," R News, vol. 2, no. 3, pp. 18–22, 2002. Available: https://journal.r-project.org/articles/RN-2002-022/.

[15] Memiş, M. Ü. and Tüm, K., "Sürekli denetim süreci ve iç denetim ile ilişkisi," Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, vol. 37, pp. 145–162, 2011.

[16] Nacar, E. N. and Erdebilli, B. D., "MAKİNE ÖĞRENMESİ ALGORİTMALARI İLE SATIŞ TAHMİNİ," Endüstri Mühendisliği, vol. 32, no. 2, pp. 307–320, 2021, doi: 10.46465/endustrimuhendisligi.811183.
[17] Oğuz, G. and Ağtaş, E., "Kurumsal kaynak planlama (ERP) sistemlerinde yapay zeka kullanımı," Stratejik ve Sosyal Araştırmalar Dergisi, vol. 8, no. 2, pp. 345–360, 2024, doi: 10.30692/sisad.1442838.
[18] Postacı, T., Önder, B., and Turan, E. E., KOBİ'lerde Kurumsal Kaynak Planlaması (ERP) Uygulamaları, T.C. Sanayi, Bilim ve Teknoloji Bakanlığı Verimlilik Genel Müdürlüğü Yayın No: 723, 2012.
[19] Rajakopal, R., "An innovation diffusion view of implementation of enterprise resource planning (ERP) systems and development of a research models," Information Management, vol. 40, pp. 87–114, 2002.
[20] Ren, S., "Optimization of enterprise financial management and decision-making systems based on big data," Journal of Mathematics, vol. 2022, pp. 1–13, 2022, doi: 10.1155/2022/1708506.
[21] Sevim, A., Kurumsal Kaynak Planlaması. Eskişehir, Turkey: T.C. Anadolu Üniversitesi Yayını, 2009.
[22] Sevim, A. and Bülbül, S., "Kurumsal kaynak planlaması (Enterprise Resource Planning ERP) sistemlerinin sürekli denetimde yapay zeka kullanımı," Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, vol. 9, no. 1, pp. 1–12, 2017.

Trajectory Control of a Constant-Load Servo System Using Fuzzy Iterative Learning PID (Fuzzy IL-PID) Controller

Mehmet Latif LEVENT[*]
Sertaç YAMAN [2]

[1,2]Department of Electrical-Electronics/ Faculty of Engineering, Hakkari University, Turkey
[*](mehmetlatiflevent@hakkari.edu.tr)

ABSTRACT

This study presents a simulation-based analysis of trajectory control for a servo system using Fuzzy Iterative Learning PID (Fuzzy IL-PID) and compares its performance with conventional PID and Fuzzy PID controllers. Classical PID controllers are widely applied in industrial systems due to their simplicity; however, they often fail to deliver adequate performance when system parameters vary or when external disturbances occur. To overcome these limitations, an enhanced control structure integrating fuzzy logic and iterative learning mechanisms is proposed. In the proposed approach, the fuzzy logic component adaptively adjusts the proportional, integral, and derivative gains based on the instantaneous error and its rate of change, allowing the controller to respond more effectively to dynamic variations. The iterative learning mechanism further improves system performance by refining these control parameters across successive iterations, ensuring that the system output converges more accurately to the reference signal over time. Simulation studies were conducted in the MATLAB/Simulink environment using a constant-load servo motor model with a sinusoidal reference input. The results demonstrate that the Fuzzy IL-PID controller achieves significantly better tracking performance than both classical and Fuzzy PID controllers. The steady-state error was reduced from 0.0468 in the classical PID case to 0.00064 using the proposed Fuzzy IL-PID method. Overall, the findings indicate that the combination of fuzzy adaptation and iterative learning provides an effective and robust control strategy for precise servo trajectory tracking, ensuring improved accuracy, reduced steady-state error, and faster transient response under constant-load conditions.

Keywords – Fuzzy Control, Iterative Learning, PID Controller, Servo System, Trajectory Tracking.

INTRODUCTION

PID (Proportional–Integral–Derivative) controllers are among the most widely used control algorithms in industrial and academic applications due to their simple structure, intuitive design, and satisfactory performance for a wide range of linear systems [1]-[3]. However, their effectiveness decreases significantly when system dynamics vary with time, or when nonlinearities and external disturbances become dominant. In such cases, the fixed gains of a classical PID controller cannot adequately adapt to changing operating conditions, leading to increased steady-state error and oscillatory responses. To overcome these challenges, various intelligent and adaptive control strategies have been proposed in the literature, including fuzzy logic, neural networks, and iterative learning control (ILC) [4]-[6]. Fuzzy control introduces human-like reasoning and linguistic rules to automatically adjust controller parameters according to system behavior [7]-[9], while iterative learning control improves performance by using error information from

previous iterations to refine control inputs. The combination of these two techniques—known as Fuzzy Iterative Learning PID (Fuzzy IL-PID)—has been shown to enhance tracking accuracy and robustness for repetitive or time-varying systems [10]-[12].

In this study, a control approach based on the Fuzzy IL-PID algorithm is proposed for the trajectory control of a servo system operating under constant-load conditions. The research aims to analyze and compare the performance of Classical PID, Fuzzy PID, and Fuzzy IL-PID controllers in terms of tracking accuracy, transient response, and steady-state error. The remainder of this paper is organized as follows: Section 2 presents the Materials and Method, describing the system model, controller structures, and simulation environment. Section 3 discusses the Results, including comparative performance evaluations among the three control strategies. Finally, Section 4 concludes the paper by summarizing the key findings and suggesting directions for future research.

MATERIALS AND METHOD

A. System Model

Table 1 lists the key physical parameters used to model the rotary servo system without considering load dynamics. The system consists of a DC motor coupled with a gear mechanism, where the electrical and mechanical characteristics define the system's overall behavior. The parameters V_s, R_s and L_s represent the motor's electrical properties, determining the current and voltage relationship within the armature circuit. The constants $k_{t,s}$ and $k_{b,s}$ describe the electromechanical conversion between electrical energy and mechanical torque.

Table 1. Parameters of the Rotary Servo System

Symbol	Description	Value
V_s	Nominal voltage of the DC motor	6 V
R_s	Armature resistance	2.6 Ω
L_s	Armature inductance	0.18 mH
$k_{t,s}$	Motor torque constant	0.00767 N·m/A
$k_{b,s}$	Back-EMF constant	0.00767 V·s/rad
N_s	Total gear ratio (N_1/N_2)	70
$\eta_{g,s}$	Gearbox efficiency	0.90
$J_{m,s}$	Motor moment of inertia	4.7×10^{-7} kg·m²
$B_{m,s}$	Viscous damping coefficient	Negligible

The gear ratio N_s and gear efficiency $\eta_{g,s}$ define the transmission relationship between the motor shaft and the output shaft. The moment of inertia $J_{m,s}$ reflects the motor's resistance to changes in angular velocity, while the viscous damping coefficient $B_{\mathrm{m,s}}$ represents internal frictional effects, which are considered negligible in this model. These parameters provide the foundation for constructing the mathematical model of the servo motor, which is used in the simulation environment to design and evaluate different control strategies such as classical PID, Fuzzy PID, and Fuzzy IL-PID controllers under constant-load conditions.

In the trajectory control analysis, the position-based model of the servo system is considered. Accordingly, the following transfer function is used to represent the relationship between the input voltage and the output angular position of the servo system:

$$G_\theta(s) = \frac{\theta_{s,out}(s)}{V_s(s)} = \frac{k_{t,s} / Ns}{s[(L_s s + R_s)(J_{m,s} s + B_{m,s}) + k_{t,s} k_{b,s}]} \quad [1]$$

When the numerical parameter values given in Table 4.1 are substituted into the equation, the transfer function for the position (trajectory) control of the servo system is obtained as follows:

$$G_\theta(s) = \frac{1.0957 \times 10^{-4}}{8.46 \times 10^{-11} s^3 + 1.222 \times 10^{-6} s^2 + 5.8829 \times 10^{-5} s} \quad [2]$$

This transfer function represents the voltage-to-position relationship of the constant-load servo system and is used in the simulation studies to analyze the trajectory tracking performance of different control algorithms.

B. Proposed Method

The figure 1 illustrates the overall structure of the proposed Fuzzy IL-PID controller developed in the Simulink environment. The system operates based on the error signal, which is obtained from the difference between the reference input and the servo system output. The derivative of the error (de/dt) is also calculated to provide dynamic information to the Fuzzy Mamdani block. The Fuzzy Mamdani block generates the initial PID gains (Kp, Ki, Kd) according to the error and its rate of change. These initial gains are then transferred to the ILC (Iterative Learning Control) block to be used in the first iteration of the control process.

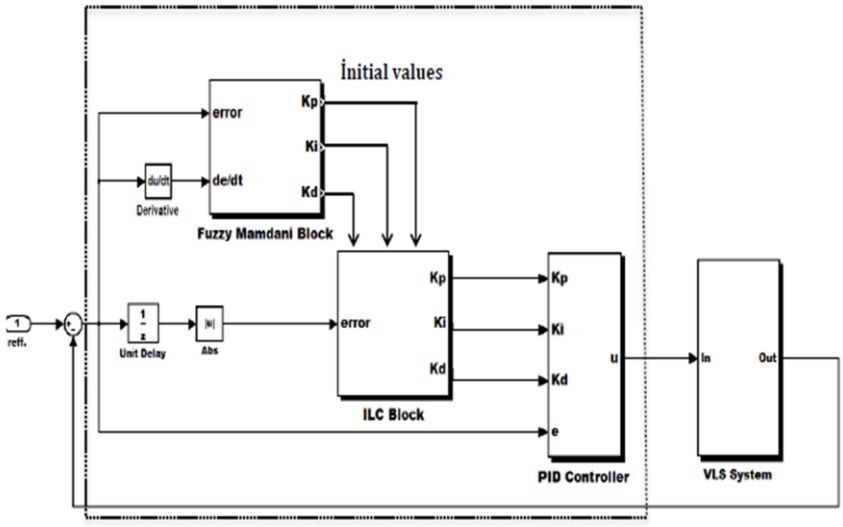


Fig. 1 Simulink Model of the Proposed Fuzzy IL-PID Controller for the Servo System

The ILC block updates the controller gains by utilizing the error information from the previous iteration, obtained through the Unit Delay and Abs blocks. Thus, the system continuously "learns" from previous iterations, improving its tracking performance over time. The updated gains are sent to the PID Controller block, where the control signal u(t) is generated. This signal drives the VLS System, which represents the servo motor model, producing the output response. As a result, the Fuzzy IL-PID structure combines the adaptability of fuzzy logic with the optimization capability of iterative learning to achieve highly accurate and stable trajectory tracking performance.

RESULTS

The reference input signal used during the simulation studies to evaluate the system's tracking performance is shown in Figure 2. The signal is a sinusoidal waveform with a specific amplitude and period, representing the target position for the servo system trajectory control. This reference signal was commonly used for all control methods (PID, Fuzzy-PID, and Fuzzy IL-PID).

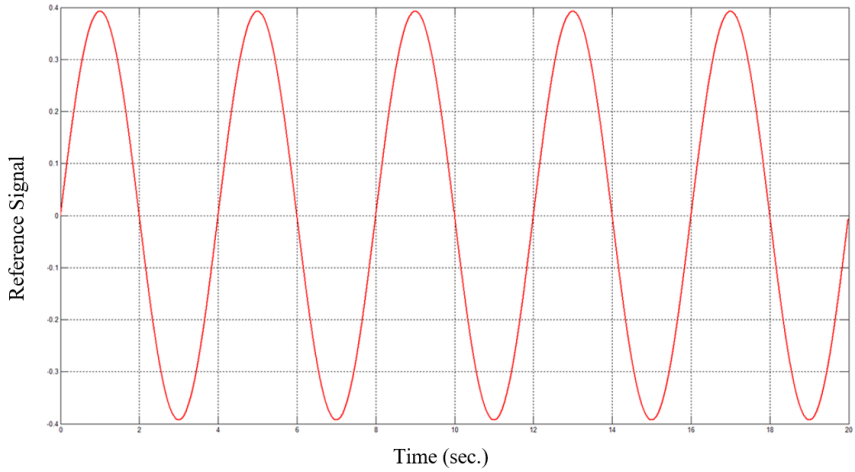


Fig. 2 Reference Input Signal Used in Simulation Studies

The system output of the constant-load servo system controlled by the classical PID method is shown in Figure 3. As observed, the system output follows the reference signal with a certain delay, and small oscillations occur during the transient period. This result demonstrates the basic tracking performance of the classical PID controller and serves as a reference for comparison with more advanced methods such as Fuzzy-PID and Fuzzy IL-PID.

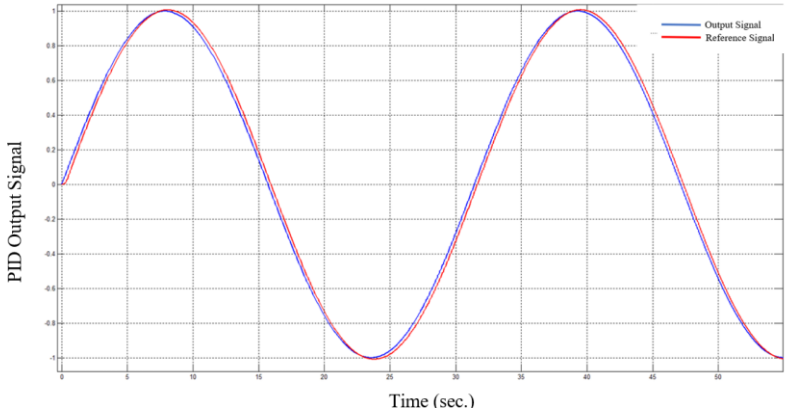


Fig. 3 System Output of the Constant-Load Servo System Controlled by the Classical PID Method

Figure 4, titled “Error Variation of the Constant-Load Servo System Controlled by the Classical PID Method,” illustrates the change in tracking

error between the reference signal and the system output over time. In this figure, the horizontal axis represents Time (sec.), while the vertical axis represents the Error signal. As observed, when the classical PID controller is applied to the constant-load servo system, the tracking error initially exhibits noticeable oscillations during the transient phase before gradually decreasing as the system approaches steady state. However, a small steady-state error remains, indicating the limitation of the classical PID controller in completely eliminating tracking deviations under constant-load conditions.

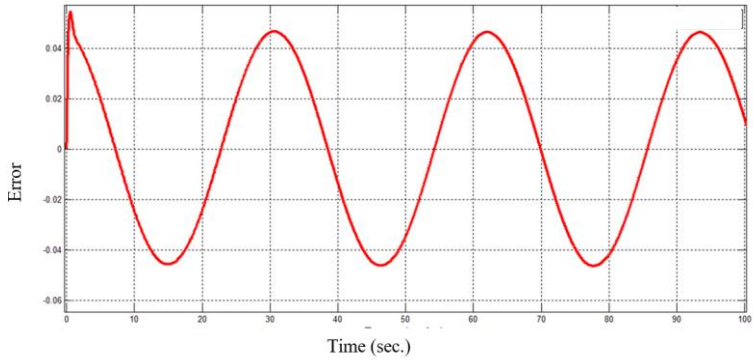


Fig. 4 Error Variation of the Constant-Load Servo System Controlled by the Classical PID Method

The system output of the constant-load servo system controlled by the Fuzzy-PID controller is shown in Figure 5. Compared to the classical PID method, the Fuzzy-PID controller provides a faster transient response and smaller steady-state error. The system output follows the reference signal more accurately, indicating improved adaptability and enhanced control performance due to the fuzzy rule-based parameter adjustment mechanism.

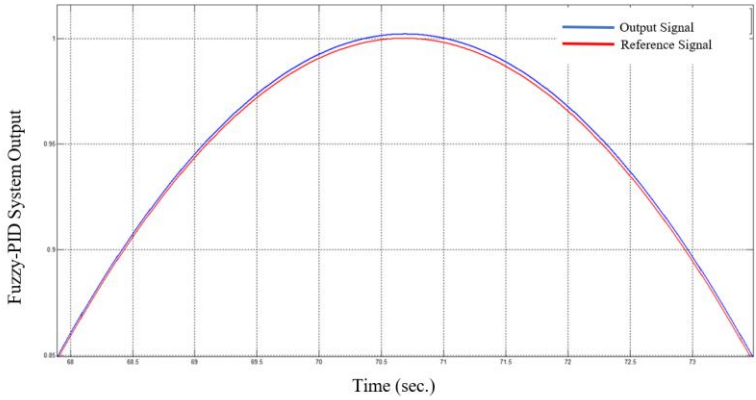


Fig. 5 System Output of the Constant-Load Servo System Controlled by the Fuzzy-PID Method

The variation of the Fuzzy-PID controller parameters for the constant-load servo system is shown in Figure 6. It can be observed that the proportional (K_p), integral (K_i), and derivative (K_d) gains are continuously adjusted by the fuzzy logic mechanism based on the instantaneous error and error rate. This adaptive adjustment improves the controller's response by optimizing the transient and steady-state performance, leading to reduced overshoot and faster convergence compared to the classical PID controller.

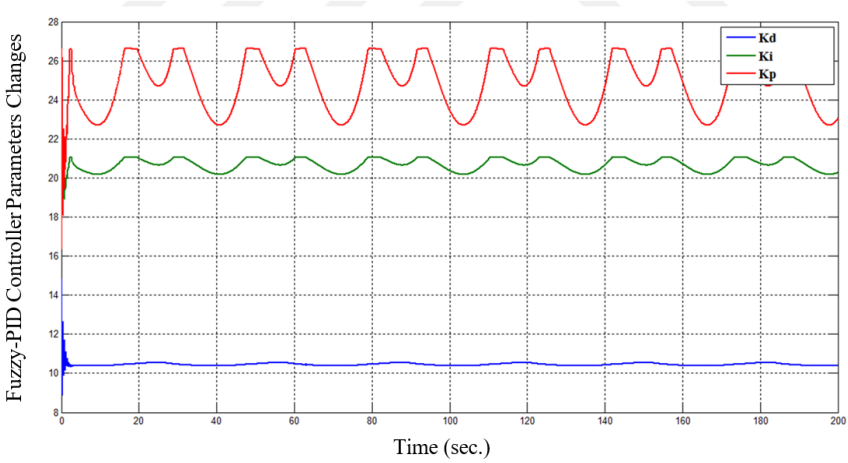


Fig. 6 Variation of the Fuzzy-PID Controller Parameters for the Constant-Load Servo System

The output response of the constant-load servo system controlled by the Fuzzy IL-PID method is shown in Figure 7. As observed, the system output

progressively approaches the reference signal with each iteration, demonstrating the learning capability of the controller. The iterative learning mechanism enables the adjustment of control parameters in successive trials, reducing the tracking error and improving convergence speed. By the third iteration, the system output closely follows the reference trajectory, indicating enhanced tracking accuracy and reduced steady-state error compared to the conventional Fuzzy-PID approach.

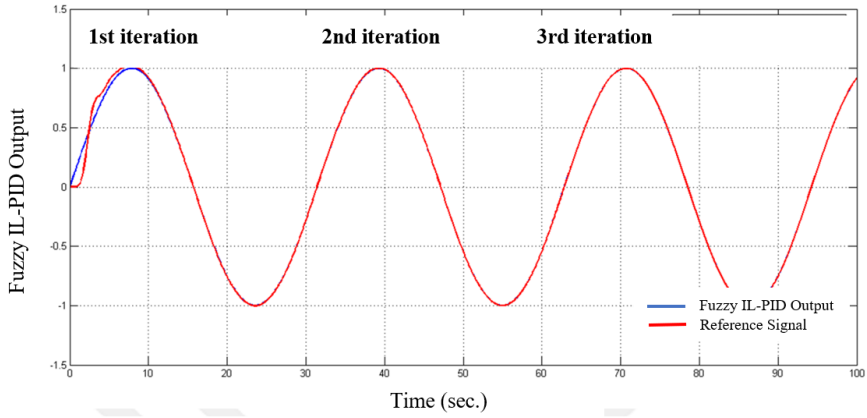


Fig. 7 Output Response of the Constant-Load Servo System Controlled by the Fuzzy IL-PID Method

When Table 2 is examined, the simulation results for the constant-load servo system clearly compare the performance of different control methods. The Classical PID controller exhibits the highest steady-state error (0.0468). This result indicates that the fixed-gain PID structure cannot adequately adapt to changing system dynamics and has a limited capability to suppress errors.

When the Fuzzy PID controller is used, the error value decreases to 0.002. This significant improvement is achieved through the fuzzy logic structure's ability to adjust the PID gains (K_p, K_i, K_d) in real time. As a result, the transient response time is shortened, overshoot is reduced, and the steady-state error is considerably minimized.

Table 2. Performance Index Comparison for the Constant-Load Servo System

Control Method	Steady-State Error (Simulation)
Classical PID	0.0468
Fuzzy PID	0.0020
Fuzzy IL-PID	0.00064

With the Fuzzy IL-PID controller, the performance improves even further, and the error value decreases to 0.00064. In this method, the iterative learning mechanism updates the gain parameters at each iteration by considering previous errors, allowing the system output to converge to the reference signal more quickly and accurately.

CONCLUSION

In this study, different control strategies—Classical PID, Fuzzy PID, and Fuzzy IL-PID—were compared for the trajectory control of a constant-load servo system through simulation analyses. The results clearly demonstrate that the classical PID controller, while simple and easy to implement, exhibits a relatively high steady-state error and limited adaptability to varying system dynamics. The Fuzzy PID controller significantly improved system performance by automatically adjusting the proportional, integral, and derivative gains through fuzzy logic rules. This adaptive structure reduced overshoot, shortened the transient response time, and minimized the steady-state error. Furthermore, the Fuzzy IL-PID controller achieved the best overall performance among the compared methods. Owing to its iterative learning mechanism, the controller was able to refine its parameters after each iteration, allowing the system output to converge more rapidly and accurately to the reference signal. In conclusion, the simulation results show that the Fuzzy IL-PID approach provides superior tracking accuracy, reduced steady-state error, and improved dynamic response compared to both the classical and fuzzy PID controllers. This confirms that integrating fuzzy logic with iterative learning control offers a highly effective and robust solution for servo system applications operating under constant-load conditions.

REFERENCES

[1] Abbas, I. A., & Mustafa, M. K. (2024). A review of adaptive tuning of PID-controller: Optimization techniques and applications. *International Journal of Nonlinear Analysis and Applications, 15*(2), 29–37. https://ijnaa.semnan.ac.ir/article_7525_259ff9a79441a49867f7839283feff51.pdf

[2] Manzo Hernández, J. P., García, G., Castillo, O., & Melin, P. (2024). A novel adaptive PID controller with new seesaw algorithms. *Journal of Control and Systems.* https://doi.org/10.1080/21642583.2024.2363625

[3] Joseph, S. B. (2022). Metaheuristic algorithms for PID controller parameters tuning. *Heliyon, 8*(9), e10533. https://doi.org/10.1016/j.heliyon.2022.e10533

[4] Jiang, Y., Yang, C., & Ma, H. (2016). A review of fuzzy logic and neural network based intelligent control design for discrete-time systems. *Discrete Dynamics in Nature and Society, 2016*, 7217364. https://doi.org/10.1155/2016/7217364

[5] Yu, P., Wan, H., Zhang, B., Wu, Q., Zhao, B., Xu, C., & Yang, S. (2025). Review on system identification, control, and optimization based on artificial intelligence. *Mathematics, 13*(6), 952. https://doi.org/10.3390/math13060952

[6] Chotikunnan, P., & Minyong, P. (2024). Adaptive parallel iterative learning control with a time-varying sign gain approach empowered by expert system. *Journal of Robotics and Control, 5*(1), 72-81. https://doi.org/10.18196/jrc.v5i1.20890

[7] Gu, X., Angelov, P., & Jin, J. (2023). Autonomous learning for fuzzy systems: A review. *Neural Computing & Applications, 35*(14), 22215–22241. https://doi.org/10.1007/s00521-022-10355-6

[8] Uzunsoy, E. (2017). A brief review on fuzzy logic used in vehicle dynamics modelling. *International Journal of Vehicle Mechanics and Mobility, 55*(6), 775–798.

[9] Anonymous. (2014). Study and review of fuzzy inference systems for decision-making and control applications. *AIJR Science, Technology & Engineering, 14*, 91-100.

[10] Dong, J., & He, B. (2019). Novel Fuzzy PID-Type Iterative Learning Control for Quadrotor UAV. *Sensors, 19*(1), 24. https://doi.org/10.3390/s19010024

[11] Bouakrif, F. (2022). Iterative Learning Fuzzy Control for Nonlinear Systems with Uncertain Initial States. *International Journal of Systems Science*, 53(15), 3261-3273. https://doi.org/10.1080/01969722.2021.2008680

[12] Ke, D., Freeman, C. T., Rogers, E., & Lewin, P. L. (2017). Fuzzy PD-type iterative learning control of a single pneumatic muscle actuator. *ISA Transactions, 67*, 287–296. https://doi.org/10.1016/j.isatra.2017.01.020

Energy Management and Protection in Shipboard Dc Microgrids: A Critical Review of RB, MPC, and DRL Approaches

Mehmet Şefik ÜNEY[*1]

[*1]Batman University, Department of Electrical and Energy, Vocational School, Batman 72000, Turkiye
[*](msefik.uney@batman.edu.tr)

ABSTRACT

Shipboard direct current (DC) microgrids are being rapidly adopted in maritime applications driven by near-zero emission targets and accelerating electrification. In these architectures, the interaction between the energy management system (EMS) and DC protection forms a coupled constraint space that must simultaneously satisfy fuel/energy cost, emissions, critical-load continuity, and safety. This review systematically surveys peer-reviewed literature from 2019-2025 and evaluates EMS approaches within a unified framework across three families: rule-based (RB), model predictive control (MPC), and deep reinforcement learning (DRL). The methods are compared by their modeling and forecasting requirements, real-time implement ability, and measurable outcomes using a common metric set, fuel/H_2 consumption, CO_2-equivalent emissions, critical-load continuity, battery degradation, and computational burden. The review explains how the fast dynamics of DC fault currents, selectivity, and breaker coordination narrow the EMS decision space, and summarizes the regulatory landscape in light of IMO's 2023 GHG Strategy together with class rules for fuel-cell systems. Digital-twin and hardware-power-hardware-in-the-loop (HIL/PHIL) validation are highlighted as prerequisites for fair cross-method benchmarking and technology transfer to the field. Evidence indicates that RB remains a strong baseline where simplicity and explainability are paramount; MPC excels in constrained multi-objective problems; and DRL is promising under uncertainty and time-varying missions provided safety constraints are enforced. Open research directions include EMS protection co-design, safe/constrained DRL, standardized scenarios and metrics, cyber-physical security, and degradation-aware decision-making for H_2/fuel-cell integration.

Keywords – Shipboard DC Microgrid, Energy Management System (EMS), Model Predictive Control (MPC), Deep Reinforcement Learning (DRL), Hardware-In-The-Loop (HIL), IMO 2023.

INTRODUCTION

In ocean-going shipping, decarbonization pressure, accelerating electrification, and the entry of alternative fuels (e.g., hydrogen, ammonia) are pushing vessel electrical architectures toward the "shipboard microgrid (SBMG)" paradigm. In this paradigm, propulsion and hotel loads are supplied in hybrid fashion by diesel generators, fuel cells, and energy storage units such as batteries or flywheels, most often within DC or hybrid AC/DC distribution topologies [1]. On the policy side, the International Maritime Organization's revised GHG Strategy adopted in 2023 envisions bringing sectoral emissions "around 2050" close to net zero and sets intermediate checkpoints for 2030-2040, thereby heightening the urgency of energy management solutions (IMO, 2023). Within this study, a representative SBMG architecture and power flows are summarized in Figure 1.

The energy management system (EMS) serves as the central decision layer that aims to deliver mission-appropriate power with maximum efficiency and reliability. Core EMS functions include resource sharing, storage scheduling, load prioritization, and where possible emissions minimization, while ensuring that operational constraints are not violated, including battery state-of-charge (SOC) limits, spinning-reserve requirements, switching and thermal limits, power quality, and system stability [1]. In the literature, these objectives are typically formulated within multi-objective optimization frameworks together with the relevant operational constraints [1].

Methodologically, recent reviews report that, alongside rule-based approaches, linear/nonlinear programming, mixed-integer and distributed optimization techniques particularly the family of model predictive control (MPC) have become prominent in SBMG applications. In parallel, machine-learning and AI-based strategies, especially deep reinforcement learning (DRL), are being adopted to improve instantaneous decision quality and cope with uncertainty [1-3]. As a representative study, Haseltalab and Negenborn (2019) unified navigation/maneuver control and energy management within a single MPC framework, forecasting propulsion power demand over the prediction horizon, optimizing diesel-generator/battery power sharing along specific fuel consumption curves, and demonstrating meaningful fuel-efficiency gains; the work also formalized the interaction between all-electric DC propulsion and the SBMG in a model-based manner [4]. On the architectural side, comprehensive surveys emphasize that primary–secondary tertiary control layers, high-power-density pulsed loads, and hybrid AC/DC distribution topologies must be addressed jointly [3].

This review therefore re-reads the literature not only by method classes but also through performance indicators aligned with mission profiles and regulations. The focus is to make causal links across decision layers explicit; to discuss resource-sharing strategies in fuel-cell battery diesel hybridization, and to analyze operational constraints in DC/hybrid architectures within an integrated framework. The aim is to clarify current trends, gaps, and research opportunities that enable EMS designs aligned with policy targets, sensitive to real-time constraints, and transferable to specific vessel types and route profiles [5].

SHIPBOARD DC MICROGRID ARCHITECTURE AND OPERATIONAL CHALLENGES

Shipboard DC microgrids integrate diesel generators, fuel cells, photovoltaic sources, and fast-responding storage units such as batteries and super capacitors with highly dynamic propulsion and hotel loads under a common power authority, forming compact yet tightly coordinated power systems.

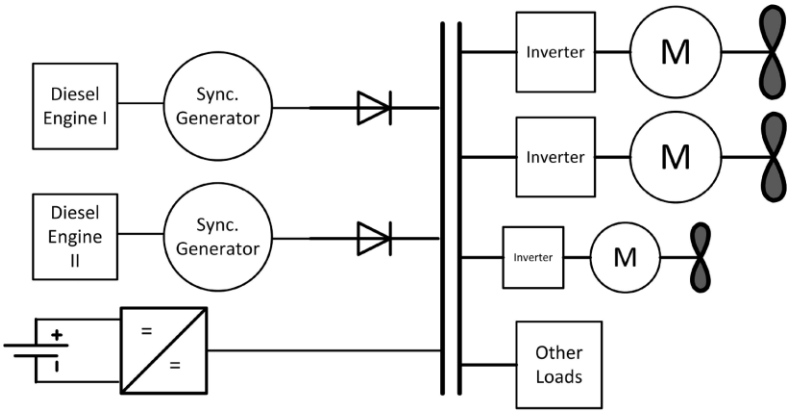


Fig. 1 Shipboard DC power-processing chain with a hybrid ESS connected to the DC bus

DC distribution is preferred not only because power-electronic converters offer advantages in efficiency, weight, and volume, but also because particularly in full-electric or hybrid propulsion the reduced number of conversion stages raises the overall "converter-to-propeller" efficiency. In the configuration examined in this study, the power-processing chain connects sources, the DC collector bus, the propulsion inverter, and the motor without unnecessary intermediate stages, as illustrated in Figure 1 (and in the single-line layout of Figure 3). The architecture diagram by Haseltalab and Negenborn similarly captures this structural simplicity, making the advantages of DC design explicit [4].

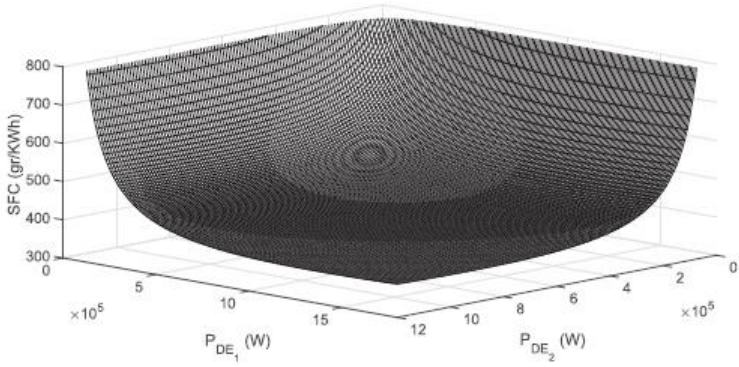


Fig. 2 Combined specific fuel consumption (SFC) curves for the berthing/departure profile

On the operational side, the defining challenge is the millisecond-to-second "pulsed" behavior of ship loads. Radar/sonar systems, dynamic positioning thrusters, stabilizers, and abrupt maneuver requests produce steep surges in propulsion power demand. These fluctuations push diesel generators into nonlinear operating regions with higher specific fuel consumption, degrading overall efficiency. The maneuver-scenario time profile reported by Haseltalab and Negenborn makes this effect visible, while Figure 2 the combined specific fuel consumption (SFC) curves for a berthing/departure profile quantitatively supports how operating-point drift increases fuel use.

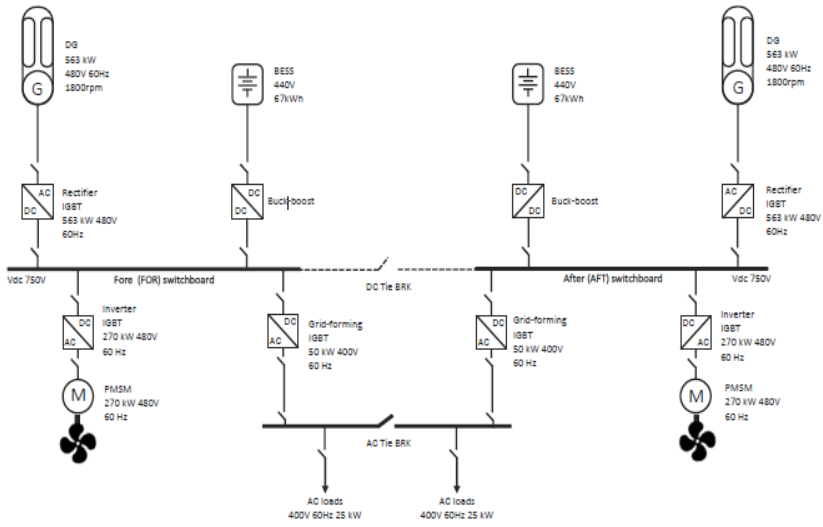


Fig. 3 Single-line diagram of the shipboard DC microgrid and locations of protection devices

Within this context, the energy storage system is not merely a “buffer” but the primary dynamic stabilizer that keeps the system safe. As depicted in Figure 1, a hybrid ESS tied to the DC bus both filters generator current and smooths maneuver-induced peaks, helping the propulsion inverter remain in its desired operating region; at very fast time constants, placement and sizing of the storage become critical for stability. The structural simplification afforded by DC also brings new protection and operations challenges: rapidly rising fault currents, breaker coordination and selectivity, and fault-ride-through management directly constrain the EMS decision space. Consequently, the architecture must be evaluated not only under nominal power flows but also in the joint envelope of maneuver peaks and fault transients. In this sense the EMS is not a mere add-on; it is a primary operational requirement that internalizes real-time constraints. When resource sharing, storage scheduling, and load prioritization are handled in a way consistent with the DC protection philosophy and breaker settings made explicit in the single-line diagram and protection placement of Figure 3 system efficiency and safety can be maximized together.

METHOD FAMILIES IN ENERGY MANAGEMENT SYSTEMS (RB-MPC-DRL)

In shipboard DC microgrids, energy management is a multi-criteria decision problem that optimizes resource allocation under variable mission profiles and pulsed loads while jointly targeting cost, emissions, continuity of supply, and equipment lifetime.

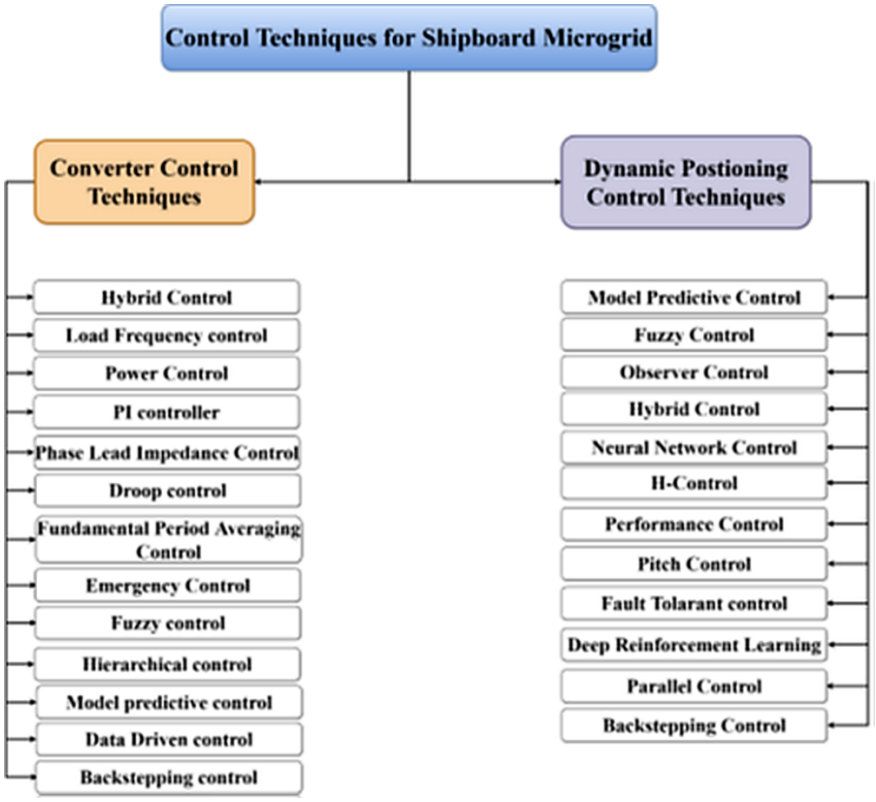


Fig. 4 RB, MPC, and DRL control techniques in shipboard DC microgrids (SBMG) and their typical response characteristics

In the literature, approaches fall into three principal families: rule-based (RB), model predictive control (MPC), and deep reinforcement learning (DRL) [6]. RB systems use explicit rules and priority sets derived from expert knowledge, offering predictability and auditability in the field; however, they have limited ability to internalize look-ahead constraints and to adapt to rapidly changing mission profiles. MPC forecasts load and source behavior over a short prediction horizon and solves a constrained optimization at each step; it can jointly manage objectives such as fuel cost, emissions, SOC safety limits, and degradation, and it mitigates the efficiency penalty that maneuver-induced power peaks impose on diesel generators. DRL, without relying on an explicit physics model, adapts online to uncertainty and varying sea conditions through a reward function; it is increasingly adopted in hierarchical architectures that combine planning outputs (speed/route) with real-time power sharing at the lower level [7].

The logical distinctions and typical response characteristics of the three families are shown side by side in Figure 4 along the flow “mission profile-

decision layer-field actuators." RB produces actions through deterministic rule chains; MPC updates the best action at each sampling instant via a prediction horizon and a constrained solver; and DRL adapts by updating policy parameters through an agent–reward loop. In practice, RB is favored in "well-defined rule" scenarios such as emergency management, islanded load-shedding order, and ESS coordination based on simple SOC thresholds, due to its low computational burden and high auditability. When the mission horizon extends and fuel-emissions-aging objectives must be balanced simultaneously, the MPC family offers clear advantages: converter constraints, ramp-rate limits, and SOC safety regions become natural parts of receding-horizon optimization. This workflow forward load/navigation prediction, constrained optimization, application of DG/ESS power setpoints, and feedback is summarized in Figure 5. Under increasing uncertainty, varying sea states, and intra-mission speed updates, DRL performs strongly in two-layer architectures: planning outputs (e.g., speed/route) are produced in the upper tier, while real-time power sharing is learned and executed by an agent in the lower tier. Figure 6 depicts this "planning + real-time energy management" arrangement with simple blocks, highlighting DRL's online adaptability.

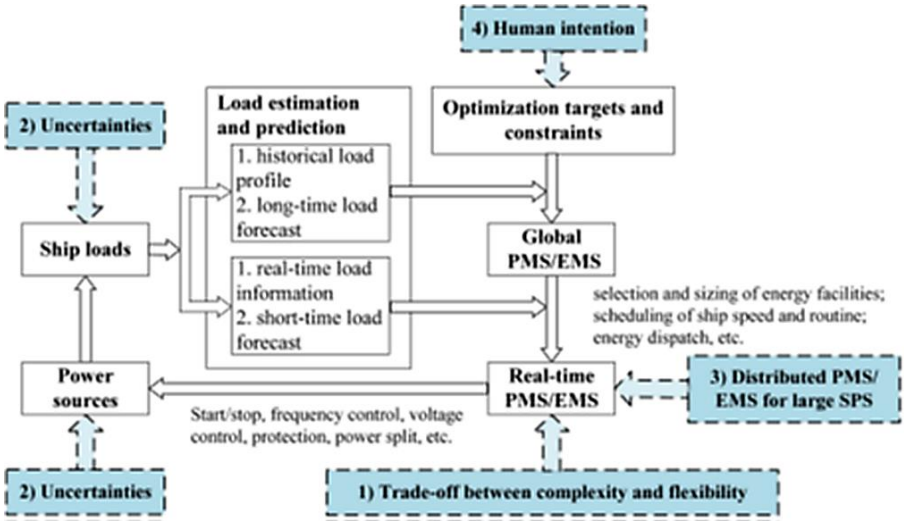


Fig. 5 Typical data flow of an MPC-based energy management system (EMS)

The practical takeaway is that there is no single "right" method. Mission horizon, uncertainty level, the nature of hard constraints (SOC limits, ramp rates, power quality), computational budget, and certification/auditability requirements must be assessed together. When regulation and safety demands dominate, the explainability of RB makes it preferable; in scenarios where

fuel, aging, and CO_2 targets must be balanced under tight constraints, MPC comes to the fore; and when uncertainty and online adaptation are critical provided safety constraints are guaranteed DRL becomes a promising option. The next section examines the necessary links between these method families and DC protection philosophy, because the fast dynamics of fault currents, selectivity, and fault-ride-through requirements practically bound the EMS decision space and directly influence method choice.

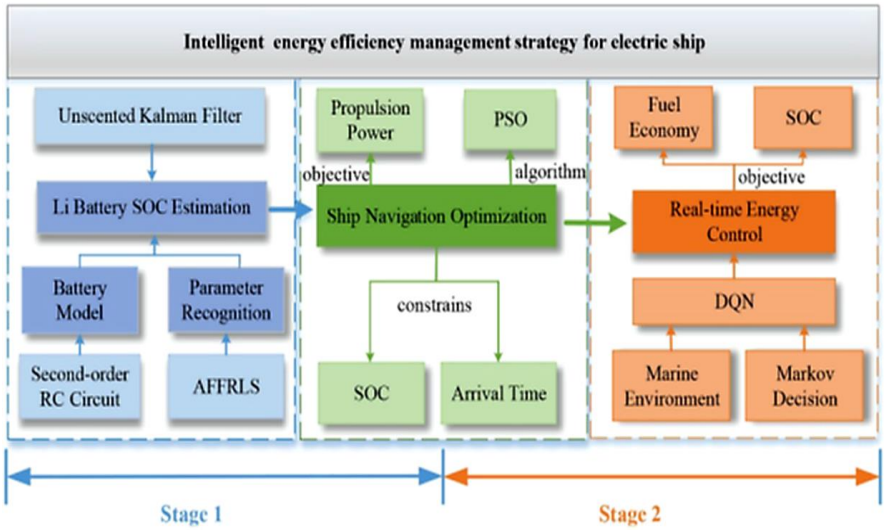


Fig. 6 Hierarchical EMS: planning (speed/route) and real-time power sharing (DRL agent)

DC PROTECTION PERSPECTIVE AND INTERACTION WITH THE EMS

The low impedance of the DC bus causes short-circuits current to rise within milliseconds; lacking a natural current zero, energy is discharged through DC-link capacitors and converter diodes over a very short interval. This makes the detect-locate-isolate sequence more time-critical than in AC systems, and if optimization-based power-sharing decisions overrun the fault-ride-through (FRT) window, stability cannot be maintained. Hence energy management and protection are not isolated functions but co-existing layers on the same physical time axis, each bounding the other's decision space. In ship applications, the sequence of de-energization, isolation, reconfiguration, and re-energization unfolds along this shared axis; a timeline showing how this

sequence evolves under different mission and fault scenarios clarifies the causality described in the text and makes the EMS protection interaction explicit. Figure 7 presents an overview of the protection system in shipboard DC microgrids [8].

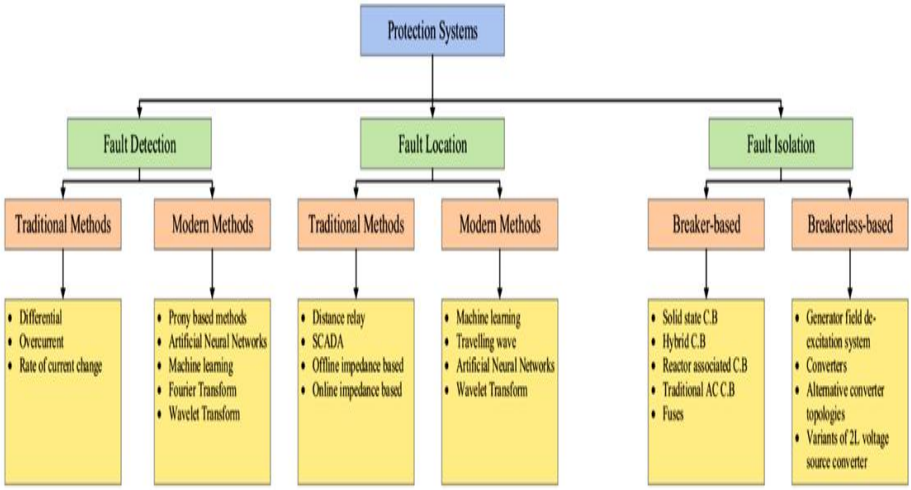


Fig. 7 Overview of the protection system in shipboard DC microgrids

In shipboard DC microgrids, the protection architecture is practically organized around the loops of detection, localization, and isolation. The rapid rise of fault current is decisive for both device withstand and selective clearing; consequently, the timing and placement of hybrid/DC circuit breakers, fast fuses, disconnectors, and where needed converter-based current-limiting/blocking functions are embedded in the system layout itself. The operational objective is to separate the faulty section from the healthy network while maintaining the continuity of critical loads especially propulsion and safety-related auxiliaries and to complete reconfiguration within a safe envelope. The “overview of the protection system” diagram brings this functional partition into a single view: sensors and local relays, the central coordination layer, breaker/fuse elements, and the logic links that trigger reconfiguration appear together, making explicit the bounds within which the EMS must operate [9].

Method selection is therefore determined not only by fuel, emissions, and degradation objectives, but also by the time window permitted by the protection layer. Rule-based designs implement safe transition sequences with explicit state machines and execute re-energization steps according to predefined protocols once clearing and sectionalization are complete. Model predictive control, after fault clearing, resolves its horizon with the updated topology and constraint set; current and ramp-rate limits, the DC-bus voltage

envelope, and ESS power/energy safety regions become natural parts of the receding-horizon optimization, enabling a smooth transition into the reconfiguration phase. Deep reinforcement learning can adapt online to the new topology when the reward function assigns heavy penalties to protection events and the action space is confined by safety-compliant masking/supervisor layers. Regardless of the chosen approach, the practical decision space is drawn by selectivity principles and response times; EMS design must therefore be treated as a co-design problem with protection philosophy and device timing [10].

REGULATORY FRAMEWORK AND STANDARDS

Regulatory pressure on shipboard microgrid designs arises from two sources: the policy imperatives of the IMO 2023 GHG Strategy adopted at MEPC 80, which orients the sector toward net-zero, and the safety rules of classification societies. The first axis translates fleet-level emission targets into vessel-level architectural and operational decisions; choices of energy carriers, the power-processing chain, and crucially the objectives of the energy management system (EMS) must align with this framework. The second axis via ABS requirements for fuel-cell systems and IGF-based guidance imposes binding safety limits on design, ventilation and gas detection, emergency shutdown (ESD), and the selection and placement of electrical equipment. Together with hazardous-area classifications per IEC (Zones 1/2), these limits effectively define the EMS "permissible action space": allowable power density in designated compartments, re-energization sequences, and conditions for continued fuel supply are bounded by the safety envelope, while ESD triggers shorten the reconfiguration horizon and timing.

This regulatory context re-weights efficiency and emission objectives under a "safety-first" principle and influences method choice. Rule-based designs guarantee safe transitions through auditable emergency sequences; MPC can rebalance fuel-emissions-aging objectives within the safety envelope by re-solving its horizon with the post-ESD topology and updated constraint set; DRL can exploit online adaptability without incurring risk when the action space is masked to enforce safety and the reward function penalizes protection events heavily. In short, Policy-rules-EMS form a single causal chain: emission targets steer architecture and mission-profile selection; class rules pull those choices into a safety envelope; and the EMS optimizes feasible decisions within protection and ESD time windows.

VALIDATION APPROACHES: DIGITAL TWIN, HIL, AND REAL-TIME SIMULATION

Regardless of whether RB, MPC, or DRL is chosen, producing safe and repeatable decisions under expected mission profiles requires a coherent model-based validation chain. This chain starts with software-in-the-loop (SIL), matures through algorithm-in-the-loop (AiL) trials on real-time (RT) simulation, is exercised under near-field conditions via hardware-/power-hardware-in-the-loop (HIL/PHIL), and finally proceeds to pilot deployment. The critical ingredient is unifying power-electronic detail and ship dynamics on the same time base. When electrical EMT models and speed/route and load profiles are clocked by a common scheduler in a co-simulation environment, one can directly verify whether EMS decisions remain within protection and ESD time windows. This integrated setup is shown in Figure 8, which presents a block-level "co-simulation environment for HIL": the RT platform, power-converter interfaces, EMS/protection hardware, and the closed communication loop are indicated together with measurement and logging points [11].

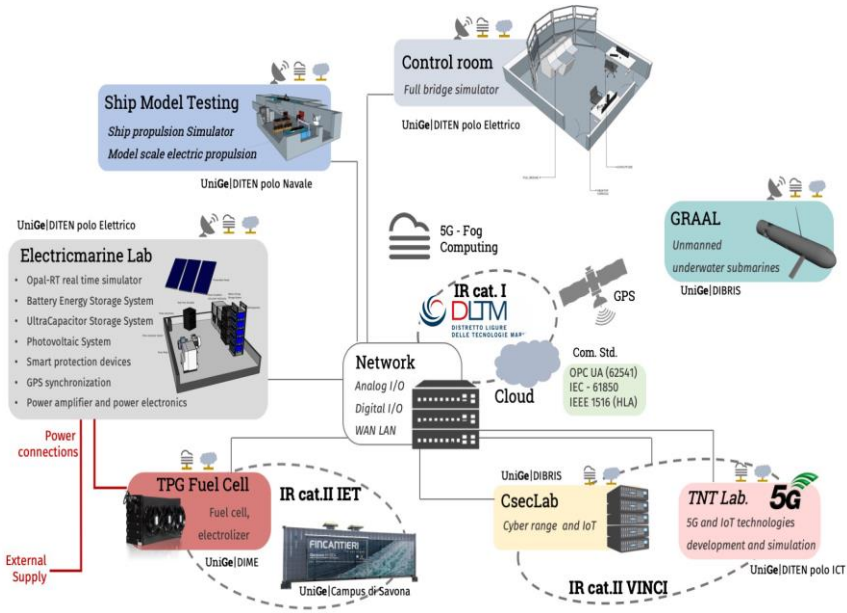


Fig. 8 Co-simulation + HIL topology for shipboard DC microgrids and propulsion

The digital twin is not merely a design-stage artifact but a lifecycle reference model that is parametrically updated. In a modular arrangement, the same core model is calibrated with operational data at SIL/RT stages, then scaled to HIL/PHIL configurations to bring real controllers and protection devices into the loop. Mission profiles, SOC constraints, and reconfiguration scenarios are

thus tested at the algorithm's target sampling rates and with real hardware latencies, eliminating "optimal yet impractical" decisions in the laboratory. Time-stamped records in Figure 8 enable co-measurement of EMS action time and protection response, demonstrating compliance with the fault-ride-through window [12].

For AiL validation, a hierarchical flow is preferable: an upper-layer navigation/speed planner and a lower-layer EMS for real-time power sharing are closed on the same data path. This arrangement allows RB rules, MPC's receding-horizon optimization, or a DRL agent's policy updates to be assessed under uncertainty, measurement noise, and hard constraints alike [13]. Planner outputs (speed/route) feed the lower layer as boundary conditions and forecasts; in the feedback loop, sampling period, solve time, and actuation delay are reported from RT logs. This narrative makes the measurement and budget points of hierarchical validation clear in text, without requiring an additional diagram beyond the already referenced co-simulation environment.

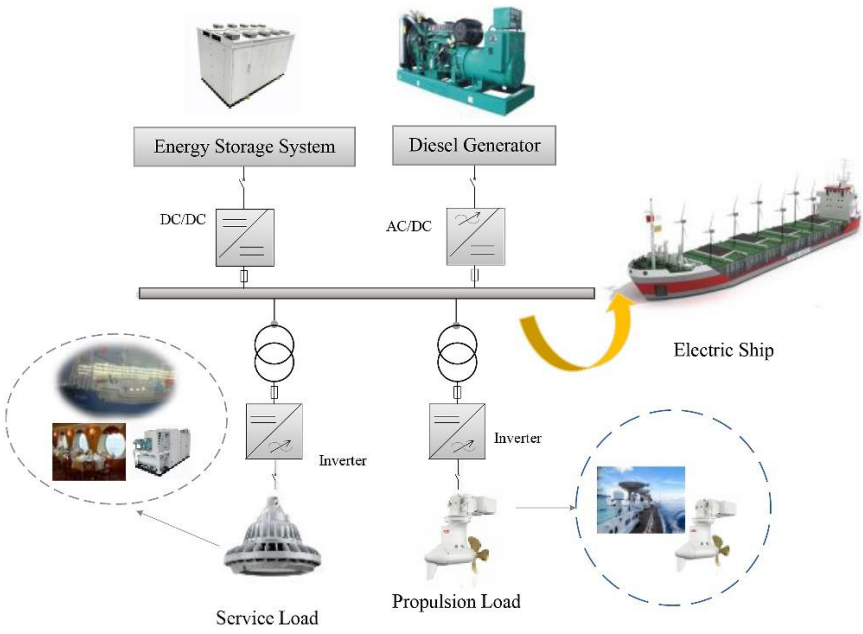


Fig. 9 Structure of the electric ship energy network.

Consolidating component-level interactions of the electric ship power network into a single diagram is essential for clearly defining interfaces. Figure 9 therefore depicts the connections of generators, the hybrid ESS, propulsion inverters, and hotel loads to the DC bus; the locations of DC circuit breakers, fast fuses, and disconnectors; and the measurement points together with the EMS/protection control inputs and outputs. This diagram aligns visually with the text during AiL/HIL synchronization by making explicit

which physical ports exchange signals and power, how ESD triggers mask the EMS action space, and which topological transitions implement the reconfiguration steps [14].

Two axes should be emphasized when reporting validation results: performance and timing conformity. On the performance side, indicators such as compliance with the SOC envelope, mission-long fuel/H_2 use and total energy cost, critical-load continuity, reconfiguration delay, and CO_2-equivalent should be unified across the same scenario set for fair cross-method comparison. On the timing side, sampling period, solver time, actuation delay, and DC-bus ripple should be presented with RT/HIL logs, and it must be demonstrated that EMS action time remains within the fault-ride-through window during protection events. This framework enables threshold-controlled transfer of strategies matured on the digital twin to HIL/PHIL and allows the subsequent RB–MPC–DRL comparison to be conducted on the same validation ground, supported by Figures 8 and 9.

COMPARATIVE ASSESSMENT AND RESEARCH AGENDA

The findings of this review show that constraint–objective structure and time scales, rather than the name of the method, are decisive. Performance targets fuel consumption and EEOI, mission speed, and the SOC envelope must be handled concurrently under deck-operations-induced pulsed loads and fast DC protection dynamics.

In practice, the most balanced outcomes arise in two-layer schemes that couple mission/speed planning with real-time allocation; the planner sets economic and environmental targets, while the lower-layer ECMS/MPC or safety-constrained DRL executes them within the feasible region permitted by protection/ESD time windows and DC-bus dynamics. An integrated diagram consolidates the multi-scale coupling across sources, loads, converters, and communications; contact points and data/command flows between PMS/EMS and the power system are illustrated in Figure 10 [2].

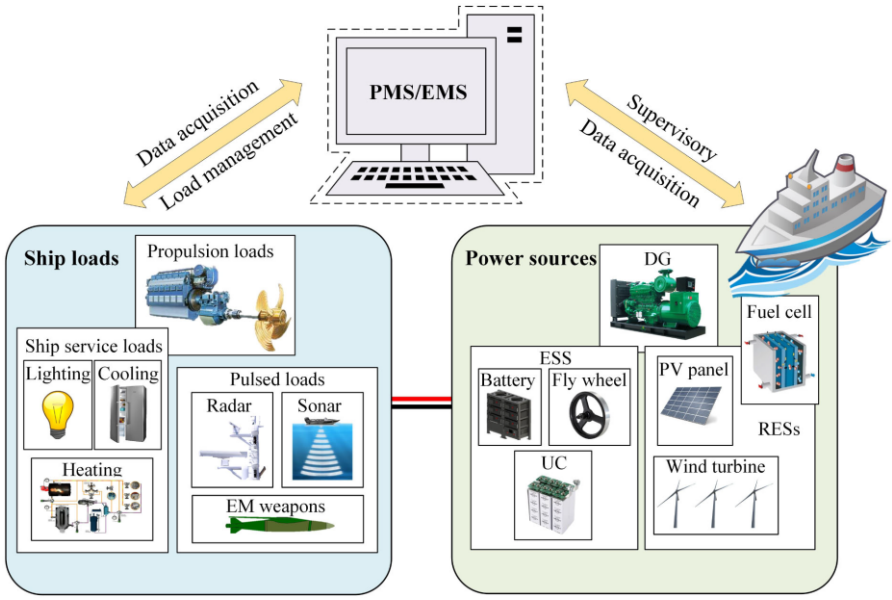


Fig. 10 Overall structure of PMS/EMS

Across method families, rule-based approaches offer low computational burden and high auditability but risk drifting from optimality in rough seas and abrupt mission transitions. MPC manages soft constraints via its prediction horizon and can balance fuel-emissions-aging objectives jointly; yet millisecond-level solve–actuation mismatches, amplified by DC protection and breaker delays, may violate physical limits during pulsed-load events.

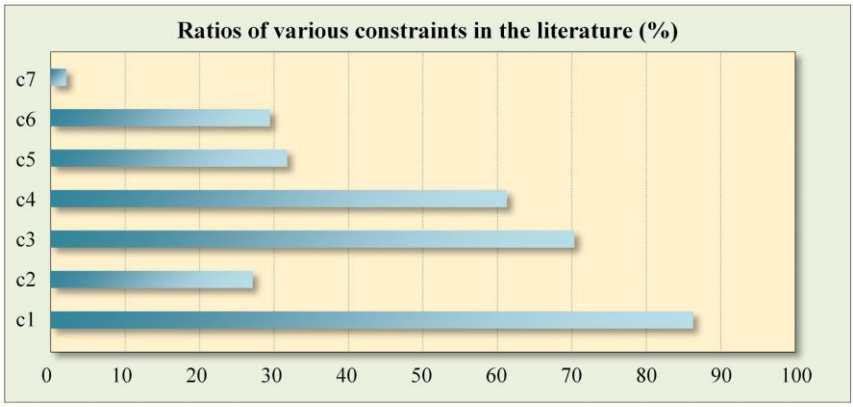
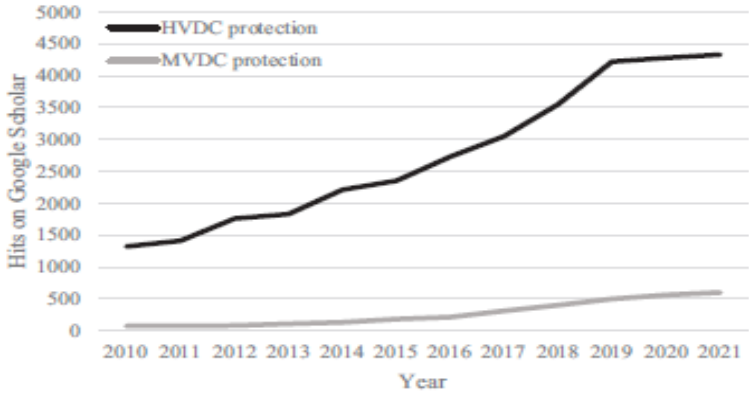


Fig. 11 Ratios of various constraints

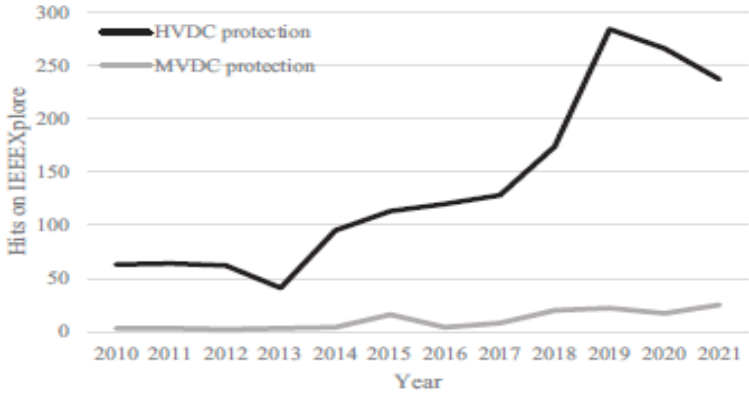
DRL adapts rapidly to new topologies and load profiles, but if the reward does not encode SOC life, EEOI, and fuel optimization together, short-term

gains can accelerate battery degradation. The corpus reviewed for 2019-2025 supports this picture: the relative shares of power/energy balance (c1), DG/ESS operating limits (c2), environmental indicators (c3), navigation/speed constraints (c4), protection/coordination (c5), aging/lifetime (c6), and cyber-physical security (c7) are summarized in Figure 11 (2019-2025 corpus) [2].

On the protection front, the relative scarcity of medium-voltage DC (MVDC) studies compared with high-voltage DC (HVDC) highlights the need for EMS protection co-design around the "fast detect-locate-isolate" chain in SBMGs. This asymmetry in publication trends points to standardization and validation gaps specifically in maritime MVDC applications. A quantitative view of the gap is provided in Figure 12, which plots the year-wise counts of tagged publications on Google Scholar and IEEE Xplore [15].



(a) Google Scholar.



(b) IEEEXplore.

Fig. 12 Number of publications tagged with 'HVDC protection' or 'MVDC protection' on (a) Google Scholar and (b) IEEEXplore

From this perspective, the research agenda should be framed not by method names but by the applicability envelope. First, the EMS and protection layer must be designed on a common time base i.e., MPC/DRL solutions must be considered together with DC breaker and fault current limiter (FCL) response times. Next, objective/reward functions should be directly linked to regulatory indicators such as EEOI, EEXI, and SEEMP-I/II. A systematic use of multi-physics digital twins with HIL/PHIL is needed to filter out "optimal yet impractical" decisions before field trials, and degradation-aware frameworks that co-optimize battery and fuel-cell life with operational objectives should be developed. For pulsed-load scenarios, online narrowing of the EMS action space with protection-aware constraints together with maturing DRL integration into two-layer (planner + real time) architectures alongside speed tracking and SOC estimation can, as recent results suggest, raise efficiency while preserving the safety envelope.

CONCLUSION

This review shows that, in shipboard DC microgrids, energy management and the protection chain share the same time base and bound each other's decision space. Choosing among RB, MPC, or DRL is not sufficient on its own; fast DC fault dynamics, selectivity/reconfiguration requirements, and regulatory safety limits define the practical "permissible action space." Design must therefore treat the sequence policy targets-protection philosophy-EMS decision layer as a single co-design problem.

Comparative analysis indicates that RB is a solid baseline where explainability and certification matter; MPC can balance fuel-emissions-aging objectives under constraints via its prediction horizon; and DRL yields promising results under uncertainty and time-varying missions when safety constraints are enforced. The most balanced performance arises in two-layer architectures: an upper mission/speed planner and a lower, constraint-aware EMS aligned with protection timing and DC-bus dynamics. For field transition, the validation pipeline (SIL-real-time Simulation-HIL/PHIL) should be standardized with common metrics and explicit timing conformance. Near-term priorities are EMS cycles aligned with protection response, objective/reward definitions directly linked to regulatory indicators, and degradation-aware decision frameworks that account for battery and fuel-cell aging. This trajectory enables solutions that are both compliant with emission targets and practically deployable.

REFERENCES

[1] Nivolianiti, E., Karnavas, Y. L., & Charpentier, J. F. (2024). Energy management of shipboard microgrids integrating energy storage systems: A review. *Renewable and Sustainable Energy Reviews*, *189*, 114012.

[2] Xie, P., Guerrero, J. M., Tan, S., Bazmohammadi, N., Vasquez, J. C., Mehrzadi, M., & Al-Turki, Y. (2021). Optimization-based power and energy management system in shipboard microgrid: A review. IEEE systems journal, 16(1), 578-590.

[3] Aboelezz, A. M., Sedhom, B. E., El-Saadawi, M. M., Eladl, A. A., & Siano, P. (2023). State-of-the-art review on shipboard microgrids: Architecture, control, management, protection, and future perspectives. Smart Cities, 6(3), 1435-1484.

[4] Haseltalab, A., & Negenborn, R. R. (2019). Model predictive maneuvering control and energy management for all-electric autonomous ships. Applied Energy, 251, 113308.

[5] wwwcdn.imo.org/localresources/en/OurWork/Environment/Documents/annex/MEPC%2080/Annex%2015.pdf

[6] Zhao, Y., Wen, S., Zhao, Q., Zhang, B., & Huang, Y. (2025). Deep Reinforcement Learning-Based Energy Management Strategy for Green Ships Considering Photovoltaic Uncertainty. Journal of Marine Science and Engineering, 13(3), 565.

[7] Zhang, H., Wen, S., Gu, M., Zhu, M., & Ye, H. (2024). A reinforcement learning method for two-layer shipboard real-time energy management considering battery state estimation. IET Energy Systems Integration, 6(3), 333-343.

[8] Xu, L., Guerrero, J. M., Lashab, A., Wei, B., Bazmohammadi, N., Vasquez, J. C., & Abusorrah, A. (2021). A review of DC shipboard microgrids—Part I: Power architectures, energy storage, and power converters. *IEEE Transactions on Power Electronics*, *37*(5), 5155-5172.

[9] Sarangi, R. R., Ray, P. K., Mohanty, A., Khadem, S., & Patra, S. (2025). Enhancing DC microgrid security: A comprehensive review of protection challenges and solutions. International Journal of Electrical Power & Energy Systems, 168, 110687.

[10] Bayati, N., & Savaghebi, M. (2021). Protection systems for DC shipboard microgrids. Energies, 14(17), 5319.

[11] D'Agostino, F., Kaza, D., Martelli, M., Schiapparelli, G. P., Silvestro, F., & Soldano, C. (2020). Development of a multiphysics real-time simulator for model-based design of a DC shipboard microgrid. Energies, 13(14), 3580.

[12] Zadeh, M. M., Rezayi, P. M., Ghafouri, S., Alizadeh, M. H., & Gharehpetian, G. B. (2023). IoT-based stochastic EMS using multi-agent system for coordination of grid-connected multi-microgrids. International Journal of Electrical Power & Energy Systems, 151, 109191.

[13] Fan, Z., Zhang, W., & Liu, W. (2023). Multi-agent deep reinforcement learning-based distributed optimal generation control of DC microgrids. IEEE Transactions on Smart Grid, 14(5), 3337-3351.

[14] Perabo, F., & Zadeh, M. (2024). Multiphysics Modeling and Co-Simulation of Ship Electric Power and Propulsion Systems for Virtual Testing and Verification. IEEE Transactions on Transportation Electrification.

[15] Kaharević, A., Ramos, J. C., Jahn, I., Ponci, F., & Monti, A. (2023, March). A review of HVDC and MVDC grid protection. In 19th International Conference

on AC and DC Power Transmission (ACDC 2023) (Vol. 2023, pp. 271-280). IET.

A Low-Cost, Modular IoT-Based Smart Parking System: Design, Prototyping, and Validation

Hadj ZERROUKI [1,*]
Salima AZZAZ-RAHMANI [2]

[1,2]Department of Telecommunications, Faculty of Electrical Engineering,
Djillali Liabes University of Sidi Bel Abbes, Algeria
[*](Zerrouki.hadj@gmail.com) Email of the corresponding author

ABSTRACT

Urban congestion, linked to the search for parking, is a critical challenge to modern cities. Smart parking systems are one of the most interesting applications within the Smart City paradigm, providing real-time information on parking availability thanks to the Internet of Things. However, many commercial systems have a high cost and complexity, making diffusion difficult, particularly in developing regions. In the paper presented here, the authors describe the design, implementation, and validation of a smart parking prototype that is characterized by its low cost, modularity, and scalability. The system uses an ATmega328P microcontroller as a basis for the system architecture to process sensor data, detects the presence of vehicles with ultrasonic and infrared sensors, and communicates via Bluetooth/Wi-Fi modules. It can also involve an Android application designed using the Kodular platform for both drivers-through an interactive map and navigation-and the parking supervisors themselves for online management of the places. With this system, it was possible to confirm the feasibility of building a complete and effective smart parking system using affordable components.

Keywords – Smart Parking, Internet of Things (IoT), Arduino, ATmega328P, HC-SR04, Low-Cost Systems, Smart Cities.

INTRODUCTION

The dramatic increase in the urban population, together with the rapid growth in private vehicle ownership, has resulted in unprecedented traffic congestion and severe environmental pollution in most cities around the world. Up to 30% of congestion in city centers can be attributed to drivers searching for vacant parking spots. This "cruising for parking" results not only in a huge waste of fuel and increased CO_2 emissions but also contributes to driver frustration and reduces overall urban mobility efficiency.

The Internet of Things-IoT and the Smart City concept have brought new perspectives to these challenges. A Smart City is generally a city that implements ICT in improving urban services such as traffic, energy, and governance. Amidst such a backdrop, smart parking systems have emerged as one of the key technologies. Such systems are able to provide real-time monitoring of parking space occupation through a network of sensors, connected devices, and intelligent platforms to guide drivers to free spaces effectively.

While many smart parking solutions exist, they typically rely on expensive infrastructure, proprietary technologies, and complex vision-based systems (e.g., cameras with AI processing) [6][7]. Such a high cost of entry severely restricts the deployment in budget-constrained municipalities or developing economies. This creates a reality where there exists a knowledge and implementation gap for not only a technologically sound solution but an economically accessible, scalable, and easy-to-maintain solution.

This research fills this gap by providing a comprehensive, end-to-end design and a functional prototype for a low-cost smart parking system. It is designed to demonstrate the viability of integrating widely available, open-source, and inexpensive hardware and software components into a capable system that possesses functionality akin to more expensive commercial alternatives. This paper describes the system architecture, hardware and software selections, core algorithm implementations, and their validation through functional scenarios.

The rest of the paper is organized as follows: Section II presents related work in the smart parking domain. Section III provides details on the system architecture and methodology adopted for the work reported herein. Section IV discusses prototype implementation and experimental validation. Section V discusses results and limitations, whereas Section VI concludes the paper and outlines future work.

RELATED WORK

Researchers have divided the smart parking research into broad categories, each named for the particular detection technology used.

a) Vision-Based Systems: Many systems depend on cameras and image processing techniques in the detection of a vehicle's presence. Authors in [8] proposed a system that uses CNNs to analyze video streams captured by cameras installed around parking lots, achieving high accuracy. Some limitations with these systems are that they are computationally expensive, raise privacy issues, and their performances are affected by weather or light conditions [7], [9].

b) Magnetometer-Based Systems: Wireless magnetometer sensors, embedded within the pavement, detect the modification in Earth's magnetic field brought about by a vehicle running above it [10]. These are highly accurate and long-lasting but higher costs regarding installment and higher maintenance because they require pavement work and, mostly, battery-powered, requiring periodic replacements [5].

c) Sensor-Based IoT Systems: This is the category our work falls into, which utilizes low-cost sensors connected via IoT protocols. Ultrasonic sensors, such as the HC-SR04, have been widely used because they are very reliable for distance measurement tasks, whether for entry/exit points or detecting individual spots [11], [12]. IR sensors, such as the Sharp GP2Y0A series, are also popular for short-range presence

detection due to their low cost and fast response times [13]. Many applications, such as in [14], adopt an Arduino or other microcontrollers to aggregate sensor data and transmit it using Wi-Fi (with modules such as the ESP8266) or Bluetooth to a central server or mobile application [15].

Our contribution is based on this background and aims to suggest a hybrid sensor approach: an ultrasonic sensor for entry/exit and an IR sensor for individual spots, with optimizations in cost and performance. While the majority of research focuses on either detection or the mobile app as standalone entities, our study develops and presents an integrated prototype with automated barrier control, dual-purpose mobile applications for drivers and supervisors, demonstrating a full operational cycle.

SYSTEM ARCHITECTURE AND METHODOLOGY

The proposed system is built on a modular, multi-layered architecture designed for low cost and scalability. The architecture comprises a perception layer, a processing layer, a communication layer, and an application layer.

A. Hardware Components

The system's hardware was selected based on cost-effectiveness, availability, and ease of integration.

1. *Microcontroller:* The core of the system is a hybrid board featuring an ATmega328P microcontroller and an integrated ESP8266 Wi-Fi module. The ATmega328P, the same processor used in the Arduino Uno, is responsible for real-time control tasks, including reading sensor data and actuating servomotors [16]. Its low power consumption and robust I/O capabilities make it ideal for this role. The ESP8266 provides Wi-Fi connectivity, enabling the system to connect to the cloud (e.g., Firebase) for data synchronization. The specific board used is shown in Fig. 1.
2. *Vehicle Detection Sensors:* A dual-sensor strategy was adopted.
   Ultrasonic Sensor (HC-SR04): Used at the entry and exit points of the parking lot. This sensor measures distance by emitting an ultrasonic pulse and measuring the time-of-flight of the echo. With a detection range of 2 cm to 4 m, it is well-suited for reliably detecting the arrival and departure of vehicles [11]. Its principle of operation is illustrated in Fig. 2

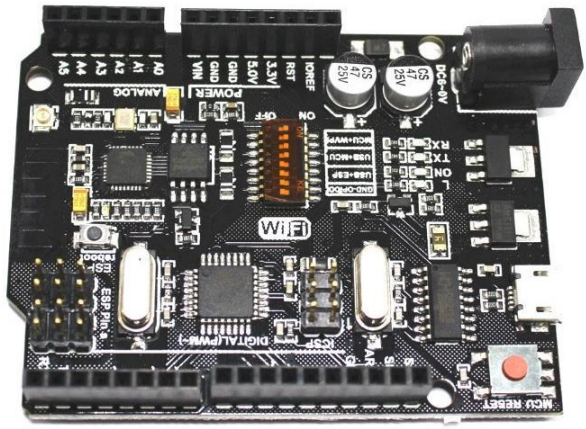


Fig. 1. Hybrid Microcontroller Board (ATmega328P with ESP8266).

- *Infrared (IR) Sensor (Sharp OA41/GP2Y0A21YK0F):* Deployed at each individual parking spot. This sensor uses triangulation of a reflected IR beam to detect presence within a short range (5 cm to 30 cm). Its fast response and low cost make it ideal for monitoring multiple spots simultaneously [13].

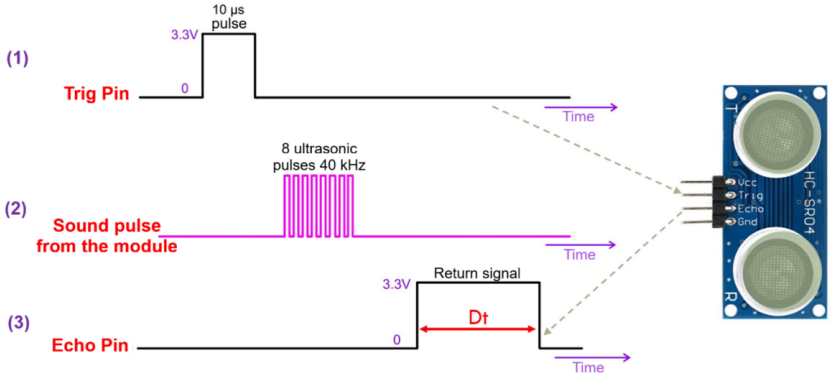


Fig. 2. Operating Principle of the HC-SR04 Ultrasonic Sensor.

3. *Actuators and Display:*

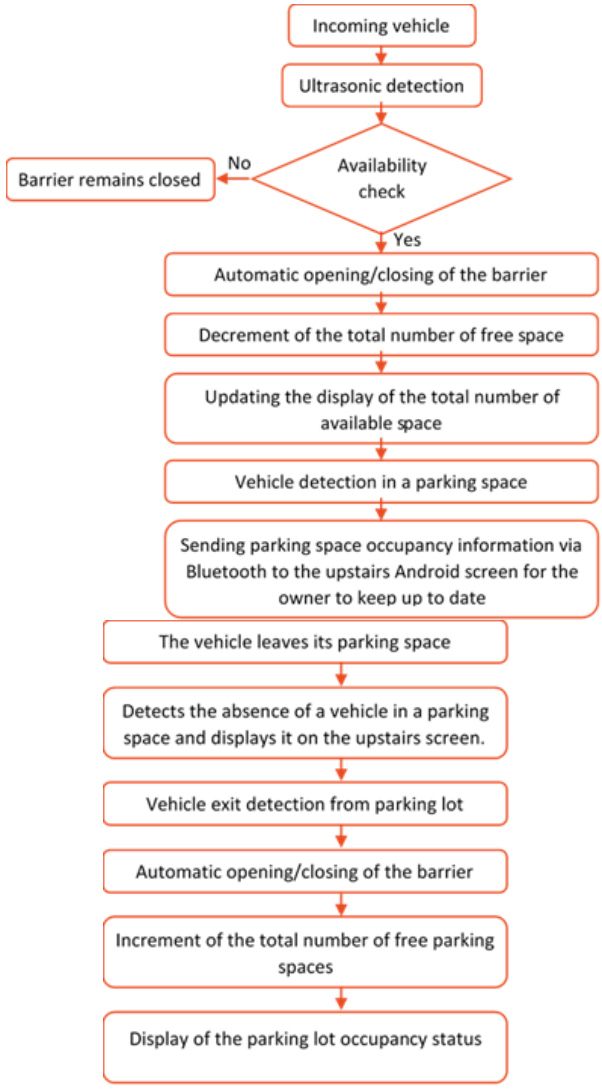
- *Servomotors (SG90):* Two micro-servos are used to operate the entry and exit barriers, providing automated access control.
- *LCD Display (20x4 with I2C):* A 20-character, 4-line LCD is placed at the main entrance to display real-time information, such as the number of available and occupied spaces. The I2C interface simplifies wiring, requiring only two data lines to the microcontroller.
- *Bluetooth Module (HC-06):* This module facilitates short-range wireless communication between the microcontroller and the

supervisor's Android application, enabling local, real-time monitoring of parking levels without relying on a Wi-Fi network.

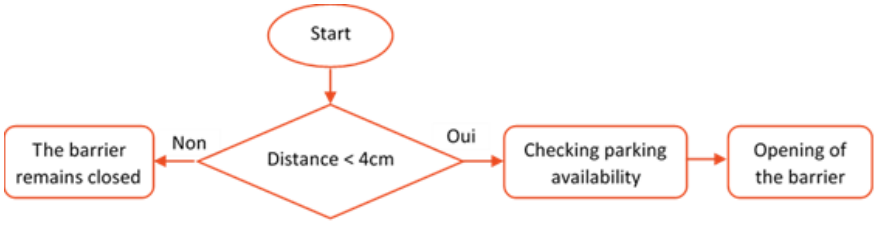
B. Software, Algorithms, and Application

The system's logic is implemented through a combination of embedded code and mobile applications.

1. *Embedded Software:* The microcontroller is programmed using the Arduino IDE in C++. The code is structured around the standard setup() and loop() functions. The core logic is defined by three key algorithms, illustrated by the flowcharts in Fig. 3:
    - *Vehicle Entry Algorithm:* (Fig. 3a) An ultrasonic sensor at the entrance detects a vehicle. The system checks for space availability. If available, the entry barrier servo is activated, and the global count of available spots is decremented.
    - *Vehicle Exit Algorithm:* (Fig. 3b) When an IR sensor detects a spot has become vacant, the local display is updated. When the vehicle is detected at the exit by an ultrasonic sensor, the exit barrier is opened, and the global count of available spots is incremented.
    - *Barrier Control Algorithm:* (Fig. 3c) This algorithm manages the physical movement of the barriers, opening them for a fixed duration (e.g., 3 seconds) before closing automatically.



(a) Vehicle Entry (b) Vehicle Exit

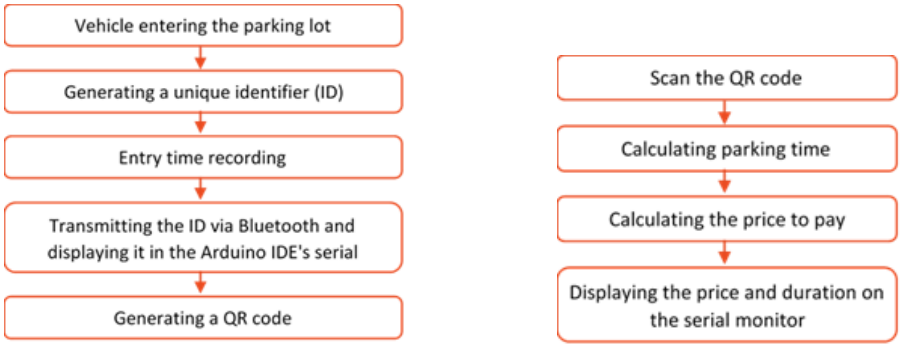


(c) Barrier Logic

Fig. 3. Core System Algorithms.

2. *Mobile Application:* Two distinct Android applications were developed using Kodular Creator, a no-code platform based on MIT App Inventor.
    - *Driver Application:* This public-facing app displays an interactive map (using Google Maps API) with markers indicating the location and real-time status of various parking lots (e.g., green for available, orange for few spots, red for full). The data is synchronized from a Firebase Realtime Database. The app also provides turn-by-turn navigation to a selected parking lot.
    - *Supervisor Application:* This internal-use app connects directly to a parking level's microcontroller via Bluetooth (HC-06). It provides a real-time schematic view of the level, showing which specific spots are occupied or free, as shown in Fig. 4.
    - *Payment Mechanism:* A proof-of-concept payment system was designed. Upon entry, the system assigns a unique ID to the vehicle and records the entry time (millis()). This information can be encoded into a QR code displayed to the user. At exit, the user scans the QR code, the system calculates the duration of stay, and applies a simple tariff (e.g., first minute free, 2 DZD per minute thereafter). This process is detailed in the flowcharts in Fig. 5.

Fig. 4. Supervisor Application Interface Showing Real-time Spot Status.



(a) Payment Process at Entry (b) Payment Process at Exit

Fig. 5. Payment Algorithm Flow.

IMPLEMENTATION AND EXPERIMENTAL VALIDATION

To validate the system's design and functionality, a physical prototype was constructed, representing a small-scale, multi-level parking facility with four spots, an entry gate, and an exit gate.

A. Prototype Setup

The prototype was assembled using the hardware components described in Section III. The microcontroller, sensors, servos, and LCD were interconnected on a breadboard to create a functional model. The overall physical assembly and wiring diagram are shown in Fig. 6.

B. Functional Scenarios and Results

The prototype was tested through a series of sequential scenarios to validate its end-to-end operation.

1. *System Initialization:* Upon power-up, the LCD at the entrance correctly displays the initial state of the parking lot (e.g., "Disponible: 8", "Occupees: 0"), as shown in Fig. 7.
2. *Vehicle Entry and Detection:* A miniature vehicle is placed at the entrance, triggering the HC-SR04 sensor. The system confirms availability, opens the entry barrier, and updates the main LCD to "Disponible: 7", "Occupees: 1", as validated in Fig. 8. The C++ code snippet responsible for this update is shown in Code 1.

(a) Physical Prototype Assembly (b) System Wiring Schematic

Fig. 6. Prototype Implementation.

Fig. 7. Initial State Display on Main LCD.

Fig. 8. LCD Update After Vehicle Entry.

```
lcd.clear();                  // Efface l'affichage précédent de l'écran LCD
lcd.setCursor(0, 0);          // Place le curseur au début de la première ligne
lcd.print("Disponible: ");    // Affiche le texte "Disponible: "
lcd.print(8 - nombreVoitures); // Affiche le nombre de places disponibles (8 - voitures présentes)
// 8 est la capacité maximale du parking

lcd.setCursor(0, 1);          // Place le curseur au début de la deuxième ligne
lcd.print("Occupees: ");      // Affiche le texte "Occupees: "
lcd.print(nombreVoitures);    // Affiche le nombre de voitures présentes
```

Code 1: Snippet for updating the main LCD display.

3. *Parking Space Occupation:* The vehicle is moved to an empty spot (e.g., A1). The corresponding IR sensor detects its presence. This information is processed by the microcontroller and transmitted via Bluetooth to the supervisor's app, which updates its display to show one occupied spot, as

seen in Fig. 9. The mapping between IR sensors and microcontroller pins is detailed in Table I.

Fig. 9. Supervisor App Update After Spot Occupation.

Table 1. IR Sensor to Microcontroller Pin Mapping

IR sensor	Microcontroller branches
Sensor 1	A0
Sensor 2	A1
Sensor 3	A2
Sensor 4	A3

4. *Vehicle Exit and Payment Calculation:* The vehicle leaves the spot and proceeds to the exit. The exit barrier opens. The system calculates the parking fee based on the duration. The result is displayed in the Arduino IDE's Serial Monitor, confirming the payment logic. For instance, a vehicle with ID 2 staying for 136 seconds incurs a fee of 4.00 DA, as shown in the screenshot in Fig. 10.

Fig. 10. Payment Calculation Result in Serial Monitor.

DISCUSSION

The experimental results successfully validate the proposed low-cost smart parking system. A prototype containing all the key features was implemented and tested: real-time detection, automatic access control, and smooth information delivery to both static displays and mobile applications.

The key strength of this system lies in its modularity and affordability because of the utilization of open-source platforms such as Arduino and Kodular and off-the-shelf components, placing the overall cost way below what commercial systems have to offer. This makes the solution especially interesting for small-scale deployments, academic research, and economically constrained environments. A dual-sensor approach has proved effective, with the HC-SR04 reliably managing the entry/exit traffic and IR sensors performing the monitoring of individual spots in a very cost-effective way.

However, the prototype has a number of limitations that need to be resolved for any real deployment.

- *Scalability:* The architecture is modular by design, but expansion to hundreds of spots will require a much more substantial communication network for supervision than Bluetooth could provide. An MQTT-based protocol over the ESP8266 Wi-Fi would be more suitable, providing capability for all levels to aggregate data at a central broker [17].
- *Sensor Robustness:* HC-SR04 and IR sensors work fine indoors but are prone to environmental factors such as rain, fog, or direct sunlight that may impact their accuracy with respect to an outdoor or open-sided parking structure. Magnetic loop detectors or radar sensors would be much more resilient in such environments, but at much higher cost.

- *Security:* The current implementation makes use of unencrypted Bluetooth communication and a conceptual payment system. A production-grade system would require secure encrypted communication channels, such as TLS/SSL for Wi-Fi, integrated with a certified payment gateway that will protect the users' data and financial transactions.

CONCLUSION AND FUTURE WORK

This paper has successfully presented the design, development, and validation of a low-cost, IoT-based smart parking system. By integrating affordable sensors, an Arduino-compatible microcontroller, and custom Android applications, we have demonstrated a functional end-to-end solution that addresses the core challenges of urban parking management. This prototype confirms that it is possible to build a feature-rich system that includes real-time monitoring, automated access, and payment processing without using high-cost and proprietary technology.

A good blueprint for accessible smart parking solutions is provided through this work, on which future research will be based to overcome the limitations of the present prototype. Further development on this would be:

- *Improving Scalability:* The internal communication protocol should migrate from Bluetooth to a lightweight IoT protocol such as MQTT for larger multi-level parking lots.
- *Improving Robustness:* Investigate and integrate weather-resistant sensors, for example, MEMS magnetometers or radar for outdoor deployments; perform long-term field tests.
- *Strengthen the security:* through end-to-end encryption for all data communication, with the development of a fully-fledged, secure payment module that integrates several established financial platforms.
- *Artificial Intelligence Integration:* Explore the use of machine learning models in predicting parking occupancy patterns from previously stored data for further optimization of traffic flow and revenue management.

By developing these features, we will turn this prototype into a market-ready solution that can make a meaningful contribution to the development of more efficient, sustainable, and intelligent cities.

REFERENCES

[1] M. A. P. Taylor, "The growing challenge of urban transport and CO_2," *Energy and Buildings*, vol. 36, no. 8, pp. 755-758, 2004.

[2] D. Shoup, "Cruising for parking," *Transport Policy*, vol. 13, no. 6, pp. 479-486, 2006.

[3] L. M. A. Bettencourt, "The Uses of Big Data in Cities," *Big Data*, vol. 2, no. 1, pp. 12-22, Mar. 2014.

[4] T. Nam and T. A. Pardo, "Conceptualizing smart city with dimensions of technology, people, and institutions," *in Proceedings of the 12th Annual International Conference on Digital Government Research*, 2011, pp. 282-291.

[5] S. S. S. R. Allu and C. P. Kurian, "IoT based smart parking system," *in 2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*, 2016, pp. 364-369.

[6] J. A. L. de Oliveira et al., "Parking Lot Occupancy Detection using Computer Vision," *in 2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, 2019, pp. 848-853.

[7] [7] V. Hassija et al., "A survey on IoT security: application areas, security threats, and solution architectures," *IEEE Access*, vol. 7, pp. 82721-82743, 2019.

[8] [8] N. S. K. A. Amato, "A Deep Learning Approach for Parking Lot Occupancy Detection," *Expert Systems with Applications*, vol. 84, pp. 100-109, 2017.

[9] [9] B. Li, Y. Ouyang, "A review on vision-based vacant parking space detection," *Journal of Traffic and Transportation Engineering (English Edition)*, vol. 7, no. 5, pp. 589-606, 2020.

[10] LEE, Sangwon, YOON, Dukhee, et GHOSH, Amitabha. Intelligent parking lot application using wireless sensor networks. In : *2008 International symposium on collaborative technologies and systems*. IEEE, 2008. p. 48-57.

[11] [11] A. Joshi et al., "Smart Car Parking System," *International Journal of Engineering Research & Technology (IJERT)*, vol. 9, no. 9, 2020.

[12] AGARWAL, Yash, RATNANI, Punit, SHAH, Umang, *et al.* IoT based smart parking system. In : *2021 5th international conference on intelligent computing and control systems (ICICCS)*. IEEE, 2021. p. 464-470.

[13] ABD KADIR, Muhamad Muzhafar, OSMAN, Mohd Nizam, OTHMAN, Nor Arzami, *et al.* IoT based car parking management system using IR sensor. *Journal of Computing Research and Innovation*, 2020, vol. 5, no 2, p. 75-84.

[14] N. S. Atiyah, "*Development of an Industrial ATmega328P Microcontroller Based on Open-Source Platform*", University of Kerbala, 2023.

[15] R. S. Rosli, M. H. Habaebi, and M. R. Islam, "Characteristic Analysis of Received Signal Strength Indicator from ESP8266 WiFi Transceiver Module", *in 7th International Conference on Computer and Communication Engineering (ICCCE)*, IEEE, 2018.

[16] S. Samsugi et al., "Sistem Pengontrol Irigasi Otomatis Menggunakan Mikrokontroler Arduino UNO," *JTST*, vol. 1, no. 1, 2020.

[17] A. J. Al-dhief, "MQTT-SN based Wireless Sensors Network for Smart Farming Applications," *Journal of Physics: Conference Series*, vol. 1362, p. 012111, 2019.

Wearable Smart Health Technology: Smart Bra

Makbule FİDAN[1]
Mihriban KALKANCI [2*]

[1]Graduate School of Natural and Applied Sciences, Department of Advanced Technologies, Pamukkale University,Turkey.
[2] Denizli Vocational School of Technical Sciences, Pamukkale University, Turkey.
[*] mkalkanci@pau.edu.tr

ABSTRACT

Smart clothing is a wearable technology that can continuously monitor critical biometric parameters such as heart rate, breast temperature, and respiration using ECG electrodes, temperature sensors, and accelerometers. These sensors are integrated into fabrics to ensure comfort during daily use, and the collected data is transmitted to analysis systems via wireless connections. This study will examine the use of smart bras developed for women to monitor biological signals and predict health risks.

Keywords – Smart Clothing, Smart Bra, Sensor, Electrode, Biometrics, Machine Learning.

WEARABLE HEALTH TECHNOLOGIES

Wearable health technologies are defined as electronic devices that can be attached to or integrated with the body to monitor and manage individuals' health status. These devices facilitate health monitoring by collecting biometric data continuously or periodically, offering innovative solutions in healthcare [1]. These technologies reduce patient dependence on the clinical environment and enable early diagnosis and intervention through real-time data flow [2].

Wearable technologies have made significant strides in digital health in recent years, enabling individuals to monitor their health status in real time. These developments, particularly with solutions focused on women's health, contribute to the personalization of healthcare services and the improvement of early diagnosis processes. In this context, smart bras are innovative wearable devices that stand out with their multi-sensor systems and data processing capabilities. Machine learning algorithms, particularly SVM, CNN, and LSTM models, enable the high-accuracy detection of health-related abnormalities such as heart rhythm irregularities, temperature changes in breast tissue, and respiratory anomalies using biometric data [3,4]. These methods can produce more precise results by also taking into account individual differences [5]. However, external factors such as sensor displacement and sweating reduce data quality in the systems; moreover, most commercial products still do not have CE and FDA approval, which leads to limitations in their clinical application areas [2].

Considering the structure, limited accessibility, and high cost of traditional screening methods, smart bra technologies offer significant advantages in terms of both accessibility and user comfort. The integration of these devices into healthcare systems strengthens preventive approaches to women's health and creates a dynamic health monitoring infrastructure through continuous data tracking.

Wearable health technologies have revolutionized individual health monitoring, providing significant advantages, particularly in the areas of early

diagnosis and continuous monitoring of women's health [1,2]. By conveniently monitoring biometric data (breast temperature, heart rhythm, and respiratory rate) in real time, smart bras stand out as effective tools for the early detection of diseases such as breast cancer [4,6].

Advanced sensor technologies and wireless communication systems optimize data collection and analysis processes, while machine learning and deep learning algorithms provide high accuracy in anomaly detection [3,5]. Furthermore, methods such as microwave imaging and bioimpedance are used as important complementary technologies for non-invasive tumor detection [7].

In the clinical arena, devices such as the Cyrcadia Breast Monitor have the potential to assess malignancy risk by monitoring thermal rhythm changes in breast tissue [8]. Literature indicates that smart bras and wearable technologies, with integrated AI-powered analyses, offer a promising approach for early diagnosis and continuous monitoring in women's health.

Wearable health technologies refer to body-worn electronic devices that facilitate personal health monitoring. The first devices emerged in the 1980s, but were difficult to carry and had limited features [2]. With technological advancements, these devices have become smaller, wireless, and more comfortable to use [9]. Smart textile applications have become particularly important for women's health, and biometric data such as heart rate and temperature can be easily monitored through sensors integrated into wearable products such as bras [4]. Furthermore, machine learning algorithms can use this data to identify health problems at an early stage [6].

Wearable health technologies will continue to play a significant role in personalized health management and early diagnosis in the future [9].

1. Areas of Use

Wearable health devices are applied in a wide range of areas, including cardiology, respiratory diseases, metabolic disorders, neurological monitoring, and women's health. Heart rate monitoring, respiratory monitoring, glucose measurement, and physical activity assessment are particularly prominent applications [9]. In women's health examples, smart bras support the early detection of breast cancer and cardiovascular problems by monitoring parameters such as breast temperature, heart rate, and respiratory rate [4].

1.1. Data Collection and Processing

These devices are equipped with electrocardiography (ECG) electrodes, temperature sensors, accelerometers, and other biosensors. Collected data is transferred to mobile devices or cloud-based platforms via wireless communication technologies [6]. During the data processing process, machine

learning algorithms analyze raw biometric data to perform tasks such as anomaly detection, modeling, and prediction [3]. LSTM, CNN (Figure 1), and SVM algorithms, in particular, are widely used in the analysis of time-series and spatial data [4,5].

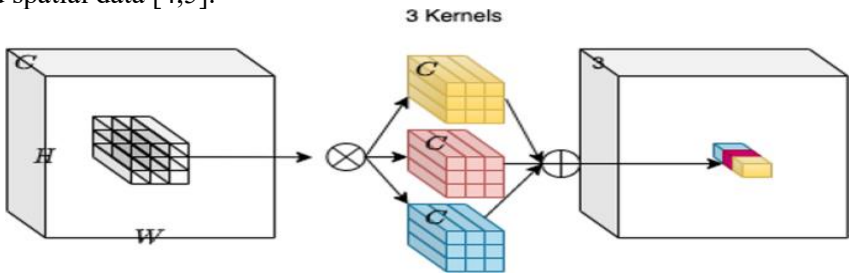


Figure 1. Architecture of convolution kernel (CNN) [4].

2. *Anomaly Detection and Notification*

Machine learning-based analytics play a critical role in the early detection of anomalies such as heart rhythm irregularities, temperature changes, or respiratory problems. Detected anomalies are notified to users and healthcare professionals via instant notifications, increasing the chance for timely intervention [6].

Wearable health technologies support early diagnosis and management of diseases by facilitating user access to health data. These systems increase the efficiency of healthcare services and reduce costs by providing continuous patient monitoring [2]. In women's health, smart bras, in particular, are showing promising results in the monitoring of breast cancer and cardiovascular diseases [4].

However, technical and ethical issues remain, such as sensor placement, the influence of external factors (sweating, movement), and data security and privacy. Additionally, the lack of clinical validation of many wearable health devices limits their widespread use in the medical field [2].

SMART BRA

Traditional breast cancer screening methods, particularly mammography, create access limitations due to their high cost, the need for fixed equipment, and reduced user comfort. These issues make early diagnosis difficult, especially in low- and middle-income areas. In this regard, wearable technologies offer a significant solution by enabling regular health monitoring

for home and personal use. Smart bras are innovative devices developed to support women's health, integrating various sensors and capable of collecting biometric data in real time [7].

Smart bras can continuously and comfortably monitor biometric data such as heart rhythm, respiratory rate, breast temperature, and changes in breast tissue [4]. This collected data is not only recorded but can also be analyzed using machine learning algorithms to make predictions about potential health risks [6]. Thus, user-specific health monitoring becomes possible and early detection of anomalies becomes easier [1].

These bras typically utilize textile-integrated sensors such as flexible ECG electrodes, temperature sensors (thermistors), and motion-sensing accelerometers [1]. The sensors are embedded within the elastic fabric without restricting the wearer's freedom of movement. Collected data is transmitted wirelessly to mobile devices or cloud-based platforms and analyzed using machine learning algorithms [6,9].

Smart bras have the potential to detect early signs of cancer cell formation by monitoring temperature changes in breast tissue [4]. Furthermore, abnormalities in vital signs such as heart rhythm and respiration can be monitored for early warning [3].

These technologies offer more comfortable, personalized, and continuous monitoring compared to traditional methods. However, durability and accuracy are critical in device design, as movement of sensors during use, perspiration, and environmental factors can affect data quality [2].

1. Imaging Methods

Microwave imaging (MWI) detects abnormal masses by assessing the dielectric properties of breast tissue. Simultaneously, Electrical Impedance Tomography (EIT) contributes to tumor identification by measuring electrical conductivity differences within the tissue. Both technologies can be placed on textile surfaces (e.g., bra fabric) via antennas and electrodes [7].

2. Sensor Placement

It is crucial that antennas used in wearable systems be flexible, lightweight, and skin-friendly. Textile-based microstrip antennas can be designed to be integrated into bras. These antennas allow microwave-based imaging devices to operate at low power, collect data without disturbing the user, and transmit data to the analysis center via Bluetooth/Wi-Fi [10].

3. Texture Analysis

The electrical properties of tumorous tissues differ significantly from those of healthy tissues. Bioimpedance measurements detect these differences, allowing us to determine the location and structure of the tumor. These measurements can be obtained through sensor arrays placed on the chest. By

embedding such sensors in a bra, passive and continuous monitoring becomes possible [7].

4. Deep Learning

Deep learning algorithms can automatically determine the presence of a mass by analyzing sensor data. These systems typically operate with pressure or vibration sensors. Just like a haptic glove, smart bras can digitally analyze abnormalities by scanning breast tissue (Figure 2). Artificial intelligence processes this data to reduce false positive/negative rates [11].

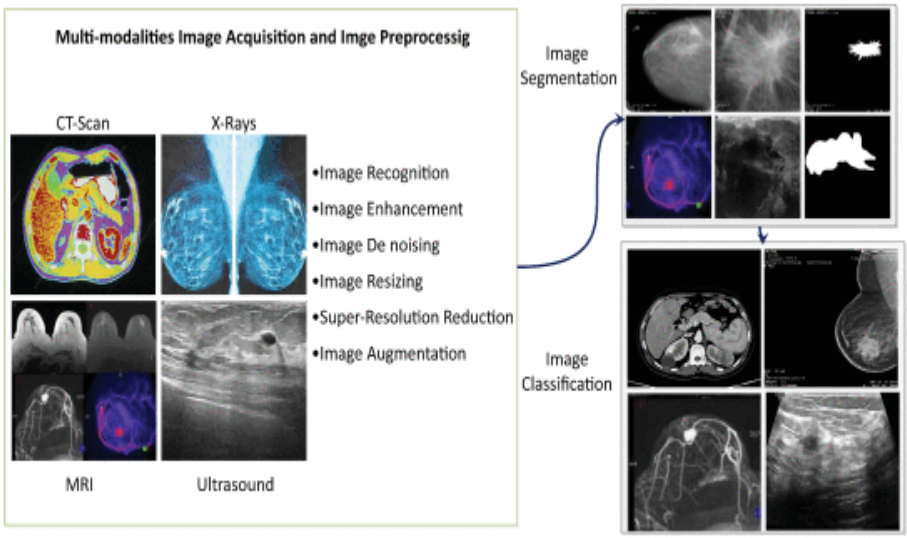


Figure 2. Diagram of advanced deep learning techniques for cancer classification using various imaging modalities, segmentation, and feature extraction [11].

5. Molecular Diagnostics

Field-effect transistor (FET)-based nanosensors can detect breast cancer-related biomarkers (e.g., HER2, CA15-3) directly from sweat, skin secretions, or tissue fluids. Because they operate without requiring labeling, these sensors produce faster and more sensitive results. Because they are relatively small, they can be embedded in the lining of a bra, providing continuous data generation without the wearer's awareness [12].

6. Data Transfer

Bra-integrated antenna systems are not only responsible for data collection; they also play a critical role in securely transmitting this data to the cloud or a mobile device (Figure 3). Microstrip antennas provide low-energy, directional signal transmission, eliminating electromagnetic pollution and preserving battery life. These systems enable remote monitoring and AI-assisted screening [10].

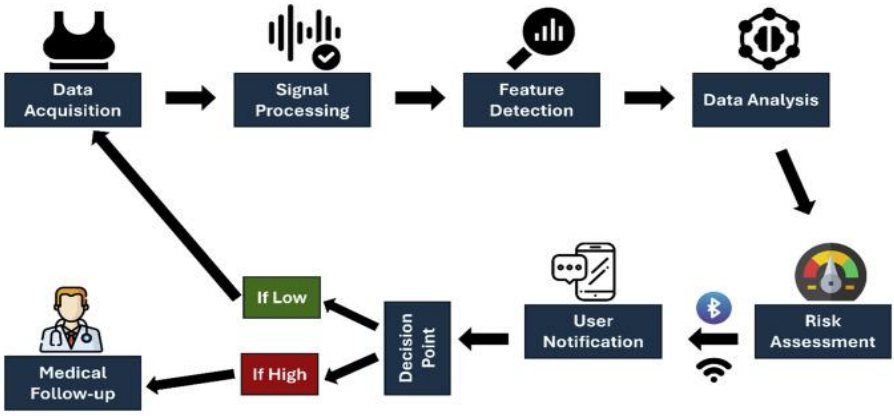


Figure 3. Early detection screening process with wearable sensors [10].

7. Clinical Approval and Testing Processes

Most smart bra prototypes are in the laboratory or simulation phase. Commercialization of these products is not possible until human clinical trials, biocompatibility testing, and regulatory processes (e.g., FDA, CE) are completed. However, early tests are showing promising results. Wearable breast health monitoring systems have the potential to increase screening frequency, especially among high-risk individuals [7].

SMART BRA SAMPLE APPLICATIONS

1. The Cyrcadia Breast Monitor (CBM) Device

This is a wearable, non-invasive technology developed to monitor breast health. The device collects 24-hour thermal metabolic data from the breast skin and analyzes this data to help detect abnormal changes early. Initial clinical studies have shown that CBM can distinguish benign and malignant breast lesions with high accuracy, sensitivity, and specificity. This allows CBM to overcome the limitations of traditional methods and offer a more comfortable and reliable early diagnosis alternative. Additional studies are currently underway to confirm the device's effectiveness and evaluate its performance in different breast tissue types [8].

(a) (b)

Figure 4. Cyrcadia Breast Monitor: Wearable Biometric Patches [8]:
(a) Sensor side of patches;
(b) Placement on the skin surface (rendered image, actual patch dimensions not shown).

2. The first version of the CBM system: CDR-100

The first version of the CBM system, the CDR-100, consists of 16 color-coded temperature sensors. Eight of these sensors are placed on the left breast and eight on the right breast, secured with surgical tape. The device (Figure 5) measures breast skin temperature every 5 minutes for 24 hours, collects the data on a recording device, and then transfers it to a computer [8].

Figure 5. Chronodynamic Data Recorder (CDR-100) [8].

3. Cyrcadian Biothermal Rhythm (CBR)

It refers to the natural, rhythmic changes in breast skin temperature throughout the day. While healthy breast tissue maintains this rhythm regularly, cancerous tissue disrupts this pattern. The Cyrcadia Breast Monitor

(CBM) device (Figure 6) monitors these temperature rhythms to detect potential abnormalities.

By analyzing the temperature data collected 24 hours a day, the device can determine the likelihood of malignancy (cancerous growth) in areas where the rhythm is disrupted. These analyses, performed using machine learning algorithms, can yield effective results, even in dense breast tissue.

In conclusion, CBR is the primary biological indicator of CBM technology, offering a more comfortable, continuous, and early-diagnosis monitoring method compared to traditional screening [8].

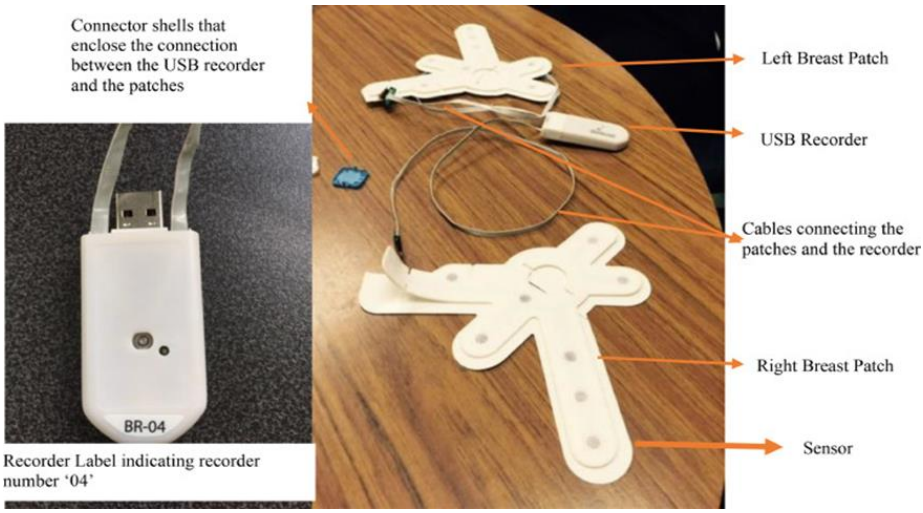


Figure 6. Components of the CBR device: (a. Left) DRD; (b. Right) Patches connected to the DRD via cables[8].

Cyrcadia Breast Monitor (CBM), its first version, the CDR-100, and the Cyrcadian Biothermal Rhythm (CBR) systems are based on the same core technology that enables abnormality detection by monitoring thermal data of breast tissue and are covered by patent US8185485B2.

4. Coplanar Waveguide Monopole Antenna (CPW)

A system for early breast cancer detection using wearable flexible sensors is being developed. This system consists of CPW (Coplanar Waveguide)-powered monopole antennas integrated into a bra. The antennas are mounted on a flexible PCB (Printed Circuit Board) and have a biocompatible structure. The antenna measures 24 × 45 mm² and has a wide bandwidth from 1.5 GHz to 8 GHz with a reflection coefficient of -6 dB. Furthermore, the specific absorption ratio (SAR) was calculated and measured as 0.75 W/kg, which is sufficient for safe use.

The system was tested using realistically shaped rubber phantoms. These phantoms create test scenarios by dynamically creating combinations of

breasts and tumors. The antenna-based sensor elements were placed around the breast phantom, and scattering parameter data were collected for tumor characterization. Simulation and measurement scenarios were presented, and the detection accuracy, number of sensors used, and training data for the developed detection algorithms were validated.

Machine learning techniques were used to analyze the collected data and identify undesirable tumor components within the breast. A system for early breast cancer detection is being developed using wearable flexible sensors and artificial intelligence techniques.

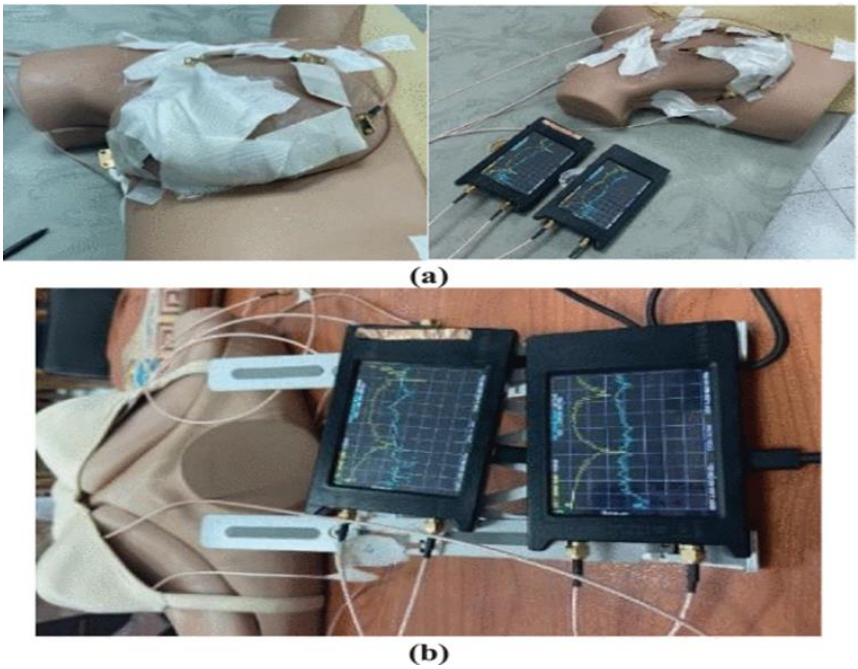


Figure 7. Coplanar Waveguide Monopole Antenna (CPW) [8].

CPW (Figure 7) monopole antennas stand out with their flexible and wide-bandwidth structures, particularly preferred in wearable and medical technologies. Patent number US10320053B2 covers the wide-band design of this antenna developed for communications and healthcare applications.

CONCLUSION

This study comprehensively examined the technical infrastructure, application areas, and machine learning-based anomaly detection approaches of smart bras, which are prominent among wearable technologies developed specifically for women's health. Literature findings indicate that smart bras can monitor heart rate, respiration, breast temperature, and tissue changes in real time, offering an alternative to traditional screening methods [4,1].

The sensor systems used in these devices, along with wireless communication and data processing infrastructures, are integrated with machine learning algorithms to provide individuals with personalized health monitoring. The success of algorithms such as LSTM, CNN, and SVM in time-series-based anomaly detection is particularly noteworthy [6]. This allows potential health problems to be identified before symptoms appear [13].

Commercial examples and experimental systems support the practical applicability of these technologies. For example, the Cyrcadia Breast Monitor has demonstrated promising results in detecting cancerous tumors by analyzing breast temperature rhythms [8]. Furthermore, the integration of advanced methods such as MWI, EIT, and FET with textile technologies has made it possible to obtain detailed data on breast tissue [7,12]. However, there are challenges to the widespread use of these technologies. Sensor accuracy, data privacy, and ethical responsibilities are still topics of debate [2]. Considering all these factors, smart bras are not just technological products; they are also considered an important part of the digital health transformation.

REFERENCES

[1] Stoppa, M., & Chiolerio, A. (2014). Wearable electronics and smart textiles: A critical review. Sensors, 14(7), 11957–11992.

[2] Heikenfeld, J., Jajack, A., Rogers, J., Gutruf, P., Tian, L., Pan, T., Kim, J. (2018). Wearable sensors: Modalities, challenges, and prospects. Lab on a Chip, 18(2), 217–248.

[3] Acharya, U. R., Fujita, H., Sudarshan, V. K., Oh, S. L., Adam, M., Tan, R. S., Chua, C. K. (2017). Application of deep convolutional neural network for automated detection of myocardial infarction using ECG signals. Information Sciences, 415–416, 190–198.

[4] Khan, A. Q., Touseeq, M., Rehman, S., Tahir, M., Ashfaq, M., Jaffar, E., & Abbasi, S. F. (2025). Advances in breast cancer diagnosis: a comprehensive review of imaging, biosensors, and emerging wearable technologies. Frontiers in Oncology, 15, 1587517.

[5] Deo, R. C. (2015). Machine learning in medicine. Circulation, 132(20), 1920–1930.
[6] Faust, O., Hagiwara, Y., Hong, T. J., Lih, O. S., & Acharya, U. R. (2018). Deep learning for healthcare applications based on physiological signals: A review. Computer Methods and Programs in Biomedicine, 161, 1–13.
[7] Elsheakh, D. N., Fahmy, O. M., Farouk, M., Ezzat, K., & Eldamak, A. R. (2024). An early breast cancer detection by using wearable flexible sensors and artificial intelligent. IEEE Access, 12, 48511-48529.
[8] Sharafaddini, A. M., Esfahani, K. K., & Mansouri, N. (2025). Deep learning approaches to detect breast cancer: a comprehensive review. Multimedia Tools and Applications, 84(21), 24079-24190.
[9] Kim, J., Campbell, A. S., & Wang, J. (2019). Wearable biosensors for healthcare monitoring. Nature Biotechnology, 37, 389–406.
[10] Singh, S., Mishra, R., Kapoor, A., & Singh, S. (2025, January). A Comprehensive Review and Analysis of the Design Aspects, Structure, and Applications of Flexible Wearable Antennas. In Telecom (Vol. 6, No. 1, p. 3). MDPI.
[11] Rafiq, M., Imran, A., Rasheed, A., & Zaman, S. U. (2024). A review on the manufacturing techniques for textile based antennas. Journal of Engineered Fibers and Fabrics, 19, 15589250241226585.
[12] Nguyen, T. T. H., Nguyen, C. M., Huynh, M. A., Vu, H. H., Nguyen, T. K., & Nguyen, N. T. (2023). Field effect transistor based wearable biosensors for healthcare monitoring. Journal of Nanobiotechnology, 21(1), 411.
[13] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58.

Investigation of Static and Dynamic Capillary Water Absorption in Travertines Under Water and Saline Conditions

Mustafa Yavuz ÇELİK[1]

1- Prof. Dr.; Afyon Kocatepe University, Afyon Vocational School, Department of Marble Technology, Afyonkarahisar, Türkiye. mycelik@aku.edu.tr ORCID No: 0000-0002-9695-7370

ABSTRACT

Travertine is widely used as a natural building stone; however, its long-term durability is strongly influenced by water uptake and salt crystallization within its pore system. This study investigated the static and dynamic capillary water absorption behaviors of three Turkish travertine types—Denizli, Emirdağ, and Kütahya—under pure and saline water conditions. Comprehensive mineralogical, petrographic, chemical, and physical-mechanical analyses were conducted to characterize the samples. The results show that the pore size distribution plays a decisive role in water absorption. Denizli travertine, with dominant macropores, exhibited high permeability and low durability, whereas Emirdağ travertine displayed a balanced pore network with the highest static absorption capacity. Kütahya travertine, characterized by low porosity and fine, poorly connected pores, exhibited the lowest absorption and greatest resistance. Dynamic water conditions generally enhance absorption due to continuous surface replenishment; however, in saline environments, salt crystallization partially blocks pores, reducing permeability in Emirdağ travertine. Mechanical tests revealed that Emirdağ travertine possessed the highest compressive strength, whereas Denizli travertine possessed the lowest. Kütahya travertine was the most durable against water- and salt-related deterioration, making it the most suitable for construction in moisture- and salt-prone environments.

Keywords – Travertine, Capillary Water Absorption, Salt Crystallization, Porosity And Durability, Building Stones.

INTRODUCTION

Travertine is used in the construction industry for ornamental purposes such as exterior and interior cladding and floor tiles. Travertine, a carbonate stone, deteriorates over time, similar to other porous building stones. Water uptake is one of the main causes of building stone degradation. Therefore, examples of three types of travertine from Turkey, which are commonly recognized worldwide, were tested for capillary water absorption.

The wide variety of natural stone types, including marble (metamorphic), limestone, sandstone, travertine (sedimentary), basalt, andesite, granite, and tuff (igneous), have been used as construction materials in the building industry. Among these, travertine has been extensively employed as a natural building stone for millennia, featuring numerous significant monuments and structures (Beasley, 1988). Travertine, classified as a freshwater carbonate, forms through the intense precipitation of terrestrial carbonates in underwater environments from bicarbonate-rich solutions, influenced by various sedimentation and diagenetic processes (Capezzuoli et al., 2014; Ford and Pedley, 1992). Its formation is often

associated with active tectonic activity. Travertine is typically white to creamy limestone, characterized by dense deposits with low porosity, composed primarily of aragonite or calcite (Guo and Riding, 1998; Pentecost, 2005; Siegesmund and Snethlage, 2011). Turkey is one of the world's leading producers and exporters of natural stone, and travertine is one of the most extensively produced and traded varieties. Turkish travertine is renowned for its diverse color palette, unique textures, and patterns, contributing significantly to global production because of its exceptional quality. These attributes have led to their widespread use in internationally renowned landmarks and constructions.

The crystallization of water and soluble salts within their pores is a primary factor contributing to the deterioration of natural building stones (Flatt, 2002). This mechanism was responsible for a significant portion of the damage observed in such materials. When water infiltrates the pores of building stones, it can lead to the precipitation of dissolved salts, followed by the growth of salt crystals. These salts often enter stones through capillary water absorption (Ruiz-Agudo et al., 2006; Rodriguez-Navarro and Doehne, 1999). Numerous studies have explored the impact of salt crystallization on the properties of natural building stones. For instance, the influence of porosity and strength on the degradation of natural stones subjected to salt crystallization cycles was examined by Ruedrich and Siegesmund (2007) and Benavente et al. (2007). Additionally, the role of petrographic features was investigated by Benavente et al. (2007) and Silva and Simão (2009). Research involving salt crystallization tests has demonstrated that high porosity significantly reduces the durability of limestone (Nicholson, 2001; Cardell et al., 2008; Espinosa Marzal and Scherer, 2008; Derluyn et al., 2014; Přikryl et al., 2017; Germinario and Török, 2020), as well as other rock types such as marble (Yavuz and. Topal, 2007), sandstone (Ruedrich and Siegesmund, 2007; Ghobadi and Babazadeh 2015; Sun et al., 2019), and granite, andesite, and tuff (Rivas et al., 2003; Sousa et al., 2005; Çelik and Tığlı, 2019; Çelik and Aygün, 2019; Çelik and. Sert, 2020; Özdemir et al. 2025).

The travertine samples were subjected to water uptake tests using normal water, saltwater capillary, and dynamic capillary experiments to simulate the capillary water absorption phenomena. In support of these studies, the petrophysical properties of travertine were determined. The travertine samples were characterized by mineralogical-petrographic (optical microscope, SEM), chemical (XRF), and mechanical-physical tests. After the tests, the changes in different aqueous environment of the capillary water absorption properties of the tested travertine were studied.

MATERIAL AND METHODS

Materials

This study focuses on three types of travertine rocks that are internationally and domestically recognized in Turkey for their extensive use in construction and decorative applications, including exterior and interior cladding, as well as floor tiles. The locations of the quarries from which the travertine samples were sourced are shown in Figure 1. The specimens utilized in the experiments were obtained from active quarries in Kocabaş (Denizli), Altıntaş Pınarcık (Kütahya), and Emirdağ (Afyonkarahisar). Denizli travertine is predominantly characterized by cream and light-yellow hues, whereas Emirdağ travertine exhibits shades ranging from light-gray to dark tones. In contrast, Kütahya travertine is distinguished by its light to dark pink and red coloration. These travertine varieties are highly favored in the market compared to other varieties, primarily because of their unique color variations and textural properties.

Methods

For material characterization, the samples underwent chemical and mineralogical-petrographic analyses, including polarizing microscopy, and X-ray diffraction (XRD). Chemical analyses were conducted using a Rigaku ZSX Primus II XRF device at the Natural Stone Analysis Laboratory of the Afyon Kocatepe University. Thin sections were prepared for petrographic examination, and the texture and mineralogical composition of the samples were analyzed using a Nikon Eclipse 2V100POL polarizing microscope. XRD analyses were performed using a Shimadzu XRD-6000 device equipped with a copper (Cu) X-ray tube at the Technology Application and Research Center (TUAM) of Afyon Kocatepe University. The Scanning Electron Microscope (SEM) was used to access further details of the texture and petrographically features of assessing magnified images. During the preparation of samples, each sample was covered by a thin carbon film. SEM examinations were performed with the LEO 1430-VP device.

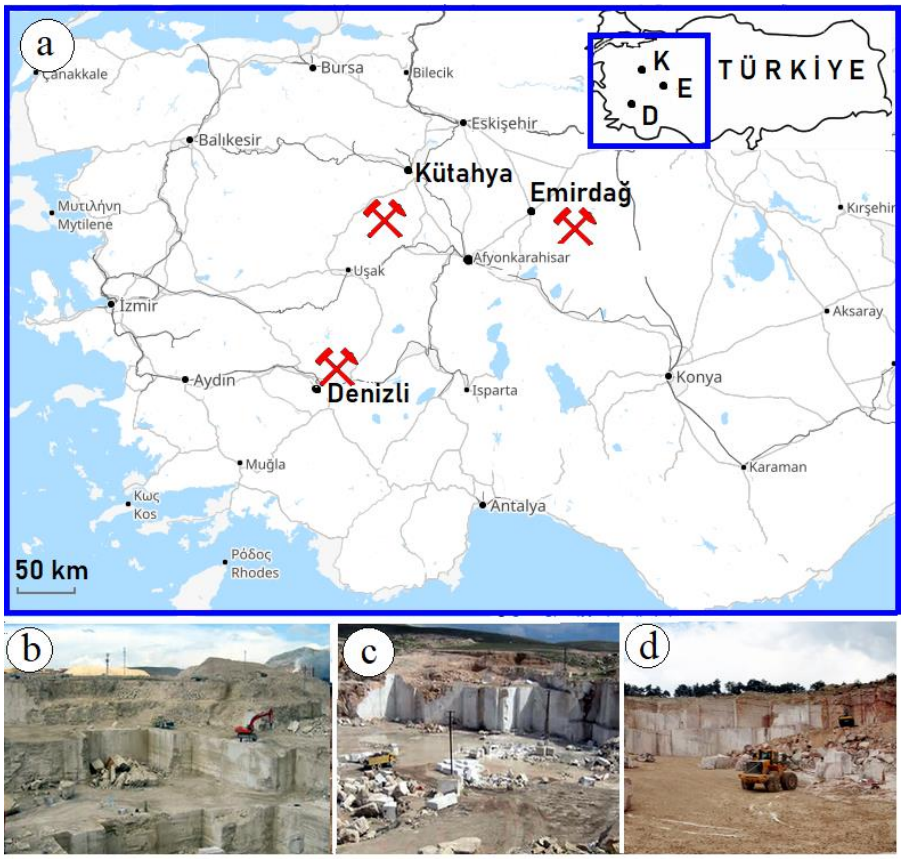


Figure 1. Location map of the quarries where the travertine samples used in the experiments were taken (a), view of the quarries, Denizli travertine (b), Emirdağ travertine (c), and Kütahya travertine (c).

One of the most critical factors influencing the capillary water absorption properties of natural building stones is pore diameter and its distribution. Understanding the pore diameter distribution is essential for interpreting the capillary water absorption behavior. The pore diameter distributions of the samples were determined using a Micromeritics AutoPore IV 9500 mercury porosimeter at the TUAM. The experimental conditions included a vacuum of 480.00 erg/cm² and contact angle of 140°.

Physical and mechanical tests were conducted to determine properties such as the density, water absorption, porosity, ultrasonic wave velocity, and compressive strength of the building stones. These tests were performed in accordance with relevant TS EN standards (Table 1). For each tuff type, six cubic samples measuring 70 mm × 70 mm × 70 mm were used. Ultrasonic wave velocity measurements were performed in compliance with the TS EN 14579 standard using a Proceq Pundit Lab ultrasonic testing device (P-wave, 54 kHz) at the of Afyon Kocatepe University.

Table 1. Relevant standards used in physical-mechanical experiments

Tests	Related standard
Water absorption by weight (%)	TS EN 13755
Porosity (%)	TS EN 1936
Compressive strength (MPa)	TS EN 1926
Density (kg/m^3)	TS EN 1936
Ultrasonic wave speed (km/s)	TS EN 14579

Water absorption owing to the capillary effect was performed with six samples from each rock (70±5 mm) according to the TS EN 1925 standards. In this experiment, the base of the prepared rock samples was immersed in water to a depth of 3.00±1 mm. To determine the capillary water absorption (capillarity) properties of the tuff and andesite, the dry weights of the samples were recorded. The sample was then placed in a container filled with water to ensure contact with the surface. The amount of water absorbed was determined by weight measurements taken at specific time intervals, which were influenced by the porosity of the stone. In this study, the time intervals were set at 1, 3, 5, 10, 15, 30, 60, 480, 1,440, 2,880, and 4320 minutes. At each time interval, the samples were removed from the water and the droplets were removed from the surface using a dry cloth. Each sample was weighed with an accuracy of 0.01 g and the amount of water absorbed depending on the time interval was determined. Capillary water absorption of the tuff and andesite was determined according to the TS EN 1925 standard, and the capillarity coefficient (C) was calculated using the following equation:

$$C=(m_i-m_d) / (A\sqrt{t_i}) \quad (1)$$

Where: mi: water-absorbed sample weight (g)
md: dry sample weight (g)
A: surface area in contact with water (m²)
ti: time (s)

The experiments were completed in four stages in normal water, saltwater capillary, and dynamic capillary experiments at the same time intervals. The travertine samples used in the experiments are shown in Figure 2.

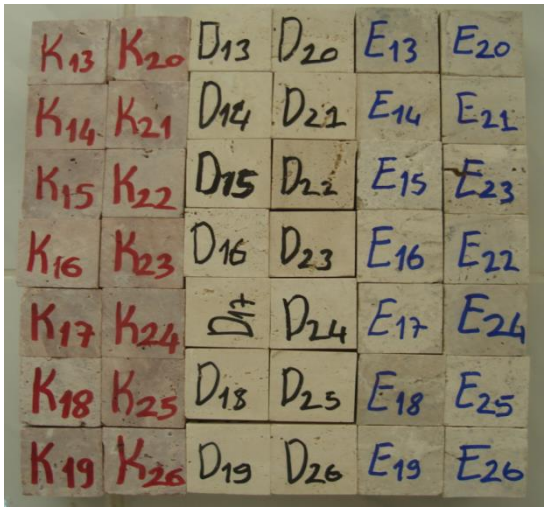


Figure 2. Travertine samples used in the experiments.

RESULTS AND DISCUSSIONS

Chemical analysis

Table 2 presents the proportional composition of the chemical elements in the analyzed travertine samples, as determined by X-ray fluorescence (XRF), expressed as percentages (Table 2). Chemical analysis indicated that calcium oxide (CaO) is the predominant component in all samples, comprising more than 50% of their total weight. The CaO content ranges from 55.15% to 55.81%, while the concentrations of MgO, SiO_2, and Fe_2O_3 vary between 0.48–0.56%, 0.15–0.95%, and 0.27–0.30%, respectively. Among the analyzed samples, Kütahya travertine exhibits the highest SiO_2 content (0.95%), whereas Emirdağ travertine has the lowest MgO concentration (0.48%). The amounts of Al_2O_3 and SO_3 are relatively low compared to other oxides, with Al_2O_3 ranging from 0.03% to 0.18% and SO_3 varying between 0.03% and 0.42%. Furthermore, no foreign substances that could adversely affect the decorative properties of travertine were identified in the chemical composition. The loss on ignition (LOI) values of the samples range from 39.68% to 42.60%, primarily attributed to the presence of organic compounds and carbonate minerals.

Table 2. Oxides (wt%) results of travertine used in the study.

Oxides (wt%)	Emirdağ Travertine	Denizli Travertine	Kütahya Travertine
CaO	55.81	55.15	55.60
MgO	0.48	0.56	0.49
Fe_2O_3	0.27	0.29	0.30
Al_2O_3	0.03	0.03	0.18
SiO_2	0.15	0.41	0.95
SO_3	0.03	0.42	0.05
LOI (Loss on ignition)	39.68	42.60	40.20

Mineralogical and petrographic studies

Polarizing optical microscope

The dominant mineral in Denizli travertine was calcite. Plant traces were occasionally observed in the microcrystalline form. A shrubby structure specific to Denizli travertines was observed in thin sections. The crystals were surrounded by a matrix, in which calcite was predominant. The inner walls of the pores of different sizes were filled with calcite crystals of 10-30 μm in size (Figure 4).

As a result of examining Emirdağ travertine with a polarizing microscope, it was observed that the main component was calcite grains with grain sizes ranging from to 1-200 μm. Calcite grains ranging from to 10-30 μm crystallized in the pores within the rock. Porosity in the rock ranging from to 1-2 μm to 1000 μm in size. Aragonite crystals existed in the form of zoned crystalline shells at the edges of some pores (Figure 4).

Examination of Kütahya pink travertine using a polarizing microscope revealed that the main component was composed of calcite grains with grain sizes ranging from to 2-300 μm. Crystallization of calcite grains ranging from to 10-30 μm was observed in the cracks in the rock. Porosity ranging from to 1-2 μm to 500 μm in the rock. Plant traces were observed in the microcrystalline form in places. Macro- and micro-sized reed structures were also observed in the structure of Kütahya pink travertine. These reed structures were observed in the thin sections. The metallic iron minerals in the composition deteriorated in places and painted the surroundings (Figure 4).

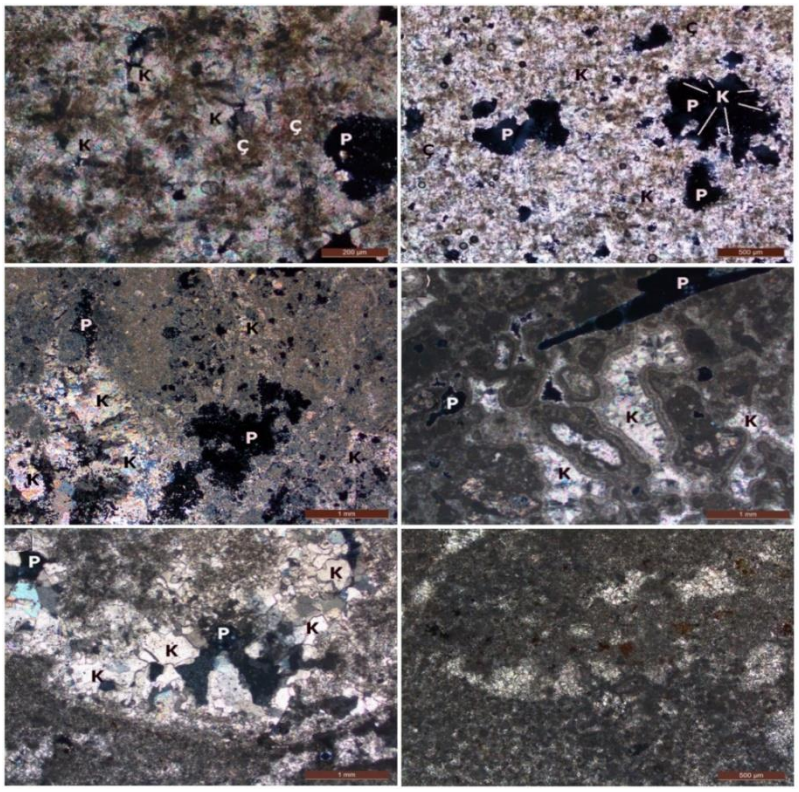


Figure 4. Thin section cross polarized light (XPL) appearances of Denizli, Emirdağ and Kütahya travertine samples. Denizli (a, b), Emirdağ (c, d) and Kütahya travertine (e, f). K: calcite, P: pore, Ç: shrub structure

SEM analyses

Scanning electron microscope (SEM) was used to determine the micro-morphological features and minerals that were not observed in the thin sections of travertine samples. Calcite minerals were observed by scanning electron microscopy (SEM) analysis of the travertine samples (Figure 5). Calcite crystals were observed in their shapes. The mineralogical composition of all three samples was predominantly calcite, as expected from the travertine formations. Differences in porosity and crystal morphology reflect variations in depositional environments, water chemistry, and post-depositional conditions.

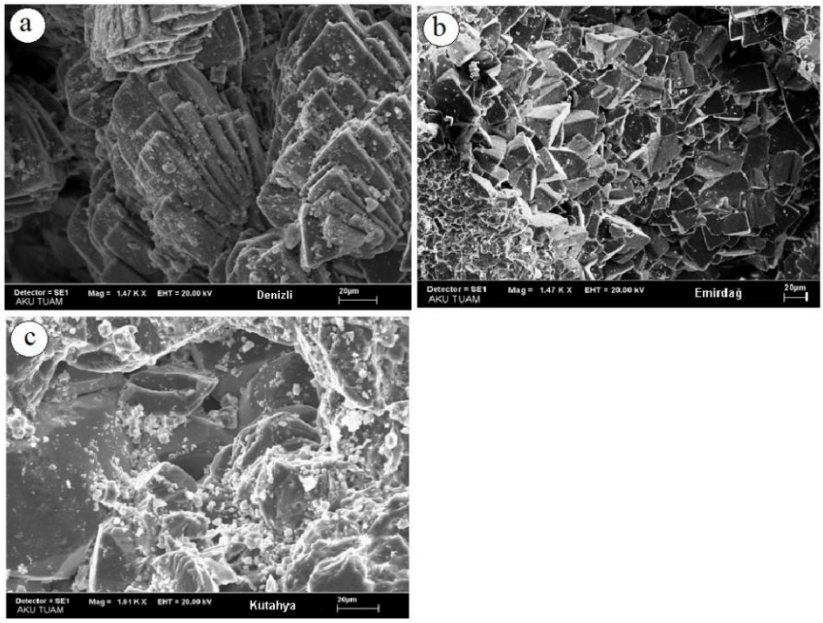


Figure 5. Micro images of calcite crystals obtained from SEM examinations of travertine samples. Denizli travertine (a), Emirdağ travertine (b), and Kütahya travertine (c)

Pore size distribution

The pore size distributions of the travertines showed distinct differences in their dominant pore sizes and overall porosity characteristics. Denizli travertine exhibited a prominent peak at approximately 100 μm, indicating the dominance of large macropores with some minor pores in the range of 1–10 μm. This suggests high permeability, which can lead to increased water absorption and a greater susceptibility to freeze-thaw and salt crystallization damage. In contrast, Emirdağ travertine has its primary peak around 0.1–0.5 μm, representing mesopores and small macropores, along with a secondary minor peak around 100 μm, suggesting a more compact structure with moderate permeability. This finer pore structure contributes to its higher strength and better durability than Denizli travertine. Kütahya travertine, on the other hand, also had a peak of around 100 μm, similar to Denizli travertine, but with a much lower overall pore volume.

Additionally, it shows a broader distribution of smaller pores in the range of 1–100 μm, indicating a lower total porosity and fewer interconnected voids. This results in lower permeability and higher resistance to weathering. Consequently, Kütahya travertine appears to be the most durable. In contrast, Denizli travertine, with its larger pores, is more prone to deterioration under harsh environmental conditions, and Emirdağ travertine offers a balanced combination of strength and moderate durability.

Denizli travertine has many large macropores, leading to high permeability and low durability, making it more susceptible to freeze-thaw cycles and salt crystallization damage. Emirdağ travertine has a more balanced pore size distribution, with smaller pores dominating, resulting in higher strength and moderate durability. Kütahya travertine had the lowest total porosity, with fine and fewer interconnected pores, making it the most durable against environmental weathering.

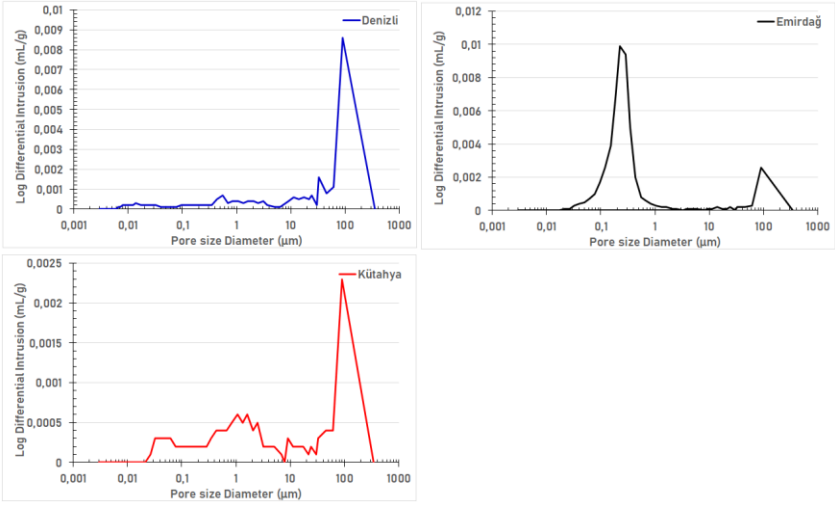


Fig. 6. Pore size distribution of travertine samples.

Physical and mechanical properties

The physical and mechanical properties of the travertines used in the experiments, determined according to TSEN standards, are presented in Table 3. The densities of the three travertines are quite similar, with Emirdağ travertine being slightly denser than the other two. A higher density generally indicates a more compact and less porous material, which can be correlated with better mechanical properties.

Kütahya travertine had the lowest water absorption, indicating that it is less porous and more resistant to water penetration. Denizli and Emirdağ travertines have higher water absorption, which may make them more susceptible to weathering and degradation over time, especially in environments with freeze-thaw cycles or chemical exposure.

Kütahya travertine has the lowest open porosity, which aligns with its low water absorption and suggests a more durable and less permeable material. Denizli and Emirdağ travertines have a higher open porosity, which makes them more prone to damage from environmental factors.

The Denizli travertine has the highest total porosity, which is consistent with its high-water absorption and open porosity. Kütahya

travertine had the lowest total porosity, indicating a denser and more durable material. Emirdağ travertine falls in between, with moderate porosity. The P-wave velocity is an indicator of the material's elasticity and internal structure. Emirdağ travertine has the highest P-wave velocity, suggesting it is the most elastic and has fewer internal defects. Denizli travertine has the lowest P-wave velocity, which may indicate more internal discontinuities or lower elasticity.

Emirdağ travertine has the highest uniaxial compressive strength, making it the strongest and most resistant to mechanical stress. Kütahya travertine follows closely, while Denizli travertine has the lowest compressive strength, which may limit its use in applications requiring a high load-bearing capacity.

The Kütahya travertine appears to be the most compact and least porous, with low water absorption, high strength, and relatively high P-wave velocity. It is likely to perform well under weathered conditions. Emirdağ travertine has the highest strength and relatively high density, but its slightly higher porosity and water absorption may make it more vulnerable to long-term weathering. Denizli travertine has the highest total porosity and lowest strength, suggesting that it may be more susceptible to deterioration, especially under harsh environmental conditions such as freeze-thaw cycles or salt crystallization.

Table 3. Physical and mechanical properties of travertine samples

	Tests	Density (real) (g/cm^3)	Water absorption by weight (%)	Open porosity (%)	Total porosity (%)	P-wave velocity (km/s)	Uniaxial compressive strength (MPa)
Denizli travertine	min	2.47	0.86	2.11	3.24	2.54	30.57
	max	2.85	5.23	10.79	20.14	2.79	86.63
	mean	2.66	3.04	6.51	13.03	2.67	59.63
Emirdağ Travertine	min	2.65	0.80	2.04	4.07	2.50	72.00
	max	2.77	5.77	11.73	9.52	2.90	145.48
	mean	2.72	3.33	7.22	7.28	2.76	95.56
Kütahya travertine	min	2.61	0.64	1.66	2.30	2.64	69.05
	max	2.84	1.15	2.79	10.21	2.84	94.43
	mean	2.71	0.91	2.25	4.87	2.74	80.67

Capillary Water Absorption Tests

Capillary water absorption experiments in pure water

The capillary water absorption behavior of the studied travertine samples under static and dynamic pure water conditions revealed significant variations influenced by the intrinsic porosity, permeability, and microstructural characteristics of each sample. The results presented in Table 5 indicate that both the rate and total amount of water uptake differed substantially between static and dynamic conditions for all travertine types.

In static pure water conditions, Emirdağ travertine (ET-S) exhibited the highest capillary water absorption, reaching 6.17 kg/m² at 4320 minutes. This high absorption rate can be attributed to the relatively higher porosity and interconnected pore network, which facilitate the capillary rise. Denizli travertine (DT-S) demonstrated a moderate absorption capacity, with a final water uptake of 2.28 kg/m², suggesting a less permeable structure compared to Emirdağ travertine. Conversely, Kütahya travertine (KT-S) displayed the lowest water absorption capacity, with a maximum of 0.98 kg/m², indicating a more compact and less porous structure that restricts capillary imbibition.

Under dynamic pure water conditions, the capillary water absorption rates increased significantly for all travertine samples. Denizli travertine (DT-D) demonstrated a notable increase in absorption, reaching 5.69 kg/m² at 4320 min, which is more than double that of its static counterpart. The enhanced absorption under dynamic conditions can be explained by the continuous renewal of the water surface, which reduces the boundary-layer resistance and facilitates greater infiltration into the pore spaces. Emirdağ travertine (ET-D) also exhibited increased absorption under dynamic conditions, although its final absorption value of 4.00 kg/m² was lower than in static conditions. This anomaly may be due to structural factors such as the potential for partial pore saturation, which limits further absorption over time. Kütahya travertine (KT-D), despite exhibiting a slight increase in water absorption under dynamic conditions (0.71 kg/m²), maintained its low capillary uptake, reinforcing its dense and low-permeability characteristics.

Table 4. Time-dependent water absorption values in pure water in travertines capillary water absorption test (kg/m^2) (DT-S: Denizli travertine static, ET-S: Emirdağ travertine static, KT-S: Kütahya travertine static, DT-D: Denizli travertine dynamic, ET-D: Emirdağ travertine dynamic, KT-D: Kütahya travertine dynamic).

Time	min (t)	1	3	5	10	15	30	60	480	1440	2880	4320
	√ti (s$^{0.5}$)	7.75	13.42	17.32	24.49	30.00	42.43	60.00	169.71	293.94	415.69	509.11
		Capillary water absorption (kg/m^2)										
Aqueous environment	DT-S	2.17	2.89	3.39	3.75	3.95	4.14	4.31	4.46	4.66	4.91	5.02
	ET-S	0.25	0.42	0.55	0.69	0.80	0.95	1.20	1.84	2.78	3.47	3.65
	KT-S	0.12	0.15	0.19	0.20	0.20	0.21	0.24	0.35	0.50	0.65	0.71
	DT-D	1.51	1.98	2.26	2.48	2.74	2.99	3.35	3.84	4.01	4.12	4.26
	ET-D	0.31	0.52	0.66	0.81	1.01	1.26	1.64	3.50	5.27	5.62	5.68
	KT-D	0.14	0.17	0.19	0.22	0.26	0.34	0.45	0.63	0.83	0.93	0.97

The capillary absorption behavior of Denizli travertine (DT) demonstrates a clear distinction between static (DT-S) and dynamic (DT-D) conditions (Fig. 7.). The absorption rate in dynamic water was significantly higher than that in static water, indicating that the increased kinetic energy of

the moving water enhances capillary infiltration. Over time, DT-D continued to absorb more water than DT-S, suggesting relatively high porosity and permeability. The plateau phase in the absorption curve indicates that the material reaches saturation at a higher level than the other travertine samples, which could have implications for its long-term durability in water-exposed environments.

Emirdağ travertine (ET) exhibits the highest capillary absorption among all tested samples, particularly in dynamic conditions (ET-D). This suggests that Emirdağ travertine possesses a highly interconnected pore structure that allows for greater water uptake. Even under static conditions (ET-S), the absorption rate was higher than that of DT-S, indicating that ET had a more permeable network of pores. The rapid initial absorption followed by a more gradual saturation trend suggests a bimodal pore size distribution, which influences the water transport through the material. This property makes Emirdağ Travertine more vulnerable to water-induced degradation, particularly in environments with prolonged exposure to moisture.

Kütahya travertine (KT) shows the lowest water absorption values under both static (KT-S) and dynamic (KT-D) conditions. The minimal difference between static and dynamic absorption indicates that the material has a lower porosity and permeability, restricting capillary water flow. The relatively flat absorption curve suggests a denser microstructure with fewer interconnected pores, which contributes to its resistance to water infiltration. This property implies that Kütahya travertine may have better long-term durability, especially in applications where water resistance is a critical factor.

Overall, the results indicate that dynamic water conditions enhance capillary absorption owing to the continuous replenishment of water at the surface of the material, facilitating deeper penetration into the pore structure. Emirdağ travertine, with its high absorption capacity, is more susceptible to water-related deterioration, making it less suitable for applications requiring strong resistance to moisture. Denizli travertine shows moderate absorption characteristics, balancing porosity and durability, whereas Kütahya travertine, with its low absorption rates, appears to be the most resistant to capillary water absorption, making it a better candidate for construction applications in wet or humid environments.

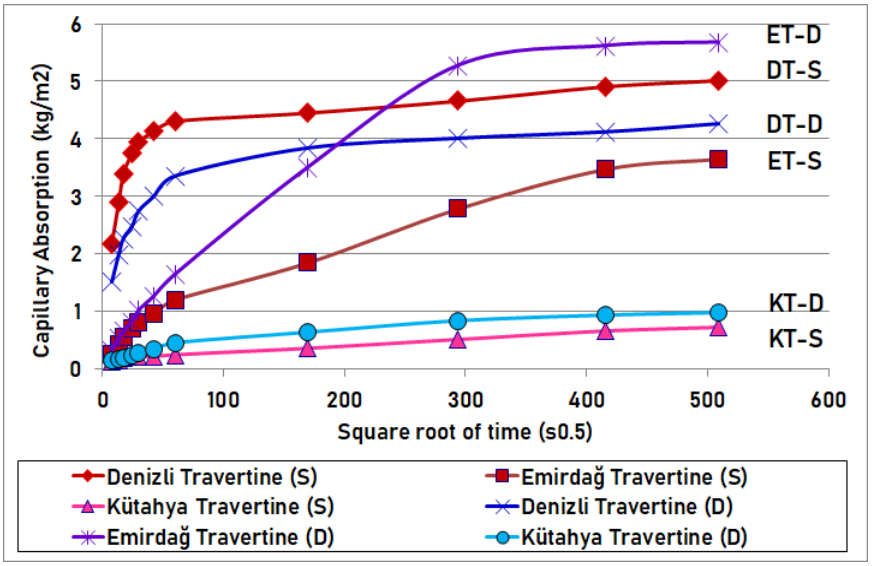


Figure 7. Capillary water absorption graph of travertines in static and dynamic water (S: static salt, D: dynamic)

Capillary water absorption experiments in salty water

The time-dependent water absorption values presented in Table 5 reveal significant differences in the absorption patterns of Denizli, Emirdağ, and Kütahya travertines under both static and dynamic saltwater conditions. These findings underscore the influence of saltwater chemistry, flow conditions, and intrinsic properties of travertines on their capillary absorption characteristics. Under static saltwater conditions, the capillary water absorption values for all travertine samples increased progressively over time, which was consistent with the typical behavior of porous materials in saline environments.

Denizli travertine (DT) shows higher capillary water absorption in dynamic saltwater (DT-SD) than in static saltwater (DT-SS). The rapid initial uptake in DT-SD suggests that the presence of moving water enhances capillary rise, likely due to the continuous replenishment of water at the surface and the mobilization of finer pores within the travertine. Over time, DT-SD reached a maximum absorption of 5.69 kg/m², significantly exceeding the DT-SS value of 2.28 kg/m². This suggests that the Denizli travertine has a well-developed interconnected pore network, which allows greater fluid transport under dynamic conditions.

Emirdağ travertine (ET) exhibits the highest capillary absorption overall, with particularly high values in both static (ET-SS: 6.17 kg/m²) and dynamic (ET-SD: 4.00 kg/m²) conditions. The strong absorption capacity of Emirdağ travertine under static conditions suggests that it has a highly porous and permeable structure that enables the infiltration of saline

solutions over time. Interestingly, the capillary rise in dynamic saltwater (ET-SD) followed a slower but steady trend, with a lower final absorption than that under static conditions. This discrepancy may be due to the pore-blocking effects induced by salt crystallization, which can reduce permeability and hinder further capillary rise. The impact of salt precipitation within pore spaces is a crucial factor for assessing the long-term durability of Emirdağ travertine in saline environments.

Kütahya travertine (KT) demonstrated the lowest capillary absorption under both static and dynamic saltwater conditions. The absorption values remained minimal throughout the testing period, with the KT-SS reaching 0.98 kg/m² and KT-SD reaching 0.71 kg/m². This suggests that Kütahya travertine has a more compact microstructure with limited pore connectivity, restricting capillary transport. The marginal difference between the static and dynamic conditions indicated that the effect of moving water was less pronounced in this sample, likely due to the lower overall porosity. Consequently, Kütahya travertine is expected to have superior resistance to salt-induced deterioration, making it a more durable material in saline environments.

Overall, the results highlight that dynamic water conditions significantly enhance the capillary absorption of travertines, particularly in samples with more open and interconnected pore networks, such as the Denizli and Emirdağ travertines. However, the presence of salt introduces additional complexities, including potential pore clogging due to salt crystallization, which may reduce permeability over time. This effect is particularly evident in Emirdağ travertine, where static absorption surpasses dynamic absorption in later stages. In contrast, Kütahya travertine demonstrated strong resistance to saltwater absorption, making it the most durable candidate for construction applications in saline environments.

Table 5. Time-dependent water absorption values in pure water in travertines capillary water absorption test (kg/m2) (DT-SS: Denizli travertine salty static, ET-SS: Emirdağ travertine salty static, KT-SS: Kütahya travertine salty static, DT-SD: Denizli travertine salty dynamic, ET-SD: Emirdağ travertine salty dynamic, KT-SD: Kütahya travertine salty dynamic).

Time	min (t)	1	3	5	10	15	30	60	480	1440	2880	4320
	√ti ($s^{0.5}$)	7.75	13.42	17.32	24.49	30.00	42.43	60.00	169.71	293.94	415.69	509.11
		Capillary water absorption (kg/m^2)										
Aqueous environment	DT-SS	0.26	0.36	0.45	0.54	0.67	0.83	0.97	1.41	1.97	2.15	2.28
	ET-SS	0.41	0.76	0.99	1.28	1.62	2.05	2.61	4.58	5.85	6.00	6.17
	KT-SS	0.05	0.10	0.11	0.13	0.19	0.24	0.31	0.47	0.73	0.86	0.98
	DT-SD	0.97	2.58	3.16	3.74	4.15	4.43	4.77	5.09	5.29	5.47	5.69
	ET-SD	0.29	0.44	0.50	0.62	0.74	0.85	1.04	1.67	3.01	3.60	4.00
	KT-SD	0.07	0.11	0.15	0.17	0.20	0.22	0.24	0.31	0.47	0.61	0.71

The capillary water absorption behavior of the examined travertine samples under both static and dynamic saltwater conditions revealed significant differences in their water uptake capacities. The data in Fig. 8 indicate that Emirdağ Travertine (ET) exhibits the highest capillary absorption in both static (ET-SS) and dynamic (ET-SD) salt water conditions. This suggests that Emirdağ Travertine has a more porous structure or a higher degree of connectivity between pores, facilitating increased water penetration over time. The dynamic salt water condition (ET-SD) further enhanced the water absorption compared to the static condition (ET-SS), implying that the movement of saltwater may promote deeper penetration of moisture into the rock matrix.

Denizli Travertine (DT) follows a similar trend but with slightly lower absorption rates than Emirdağ Travertine. The data suggest that under static conditions (DT-SS), the water absorption is moderate, while in dynamic salt water (DT-SD), it increases significantly, nearing the absorption levels of Emirdağ Travertine.

In contrast, Kütahya Travertine (KT) exhibited the lowest capillary water absorption among the three types. Both static (KT-SS) and dynamic (KT-SD) conditions showed a minimal increase in water uptake over time. This suggests a relatively lower porosity and possibly more compact or cemented pore structure, reducing the infiltration of saltwater.

The overall trend observed in Fig. 8 suggests that dynamic saltwater conditions generally enhance capillary absorption, particularly in more porous travertines such as Emirdağ and Denizli. This increased absorption under dynamic conditions could have implications for the durability of these stones in real-world applications where exposure to fluctuating moisture

conditions may accelerate weathering and degradation. The relatively low absorption of Kütahya Travertine may indicate better resistance to saltwater-related deterioration, making it a more durable option for construction in saline environments.

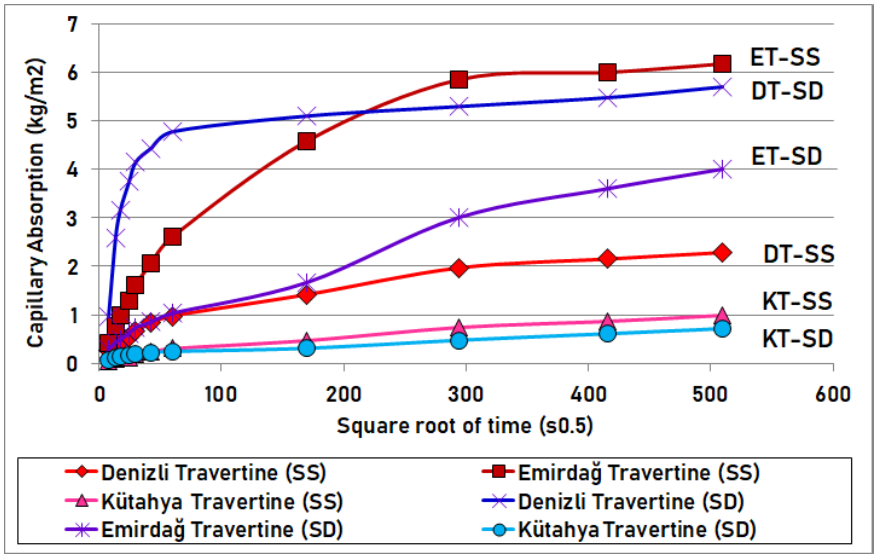


Figure 8. Capillary water absorption graph of travertines in static and dynamic salt water (SS: static salt, SD: dynamic salt)

CONCLUSIONS

Capillary water absorption tests conducted under both pure and saline water conditions demonstrated that the porosity and microstructural characteristics of travertine significantly influenced its water uptake behavior. Denizli and Emirdağ travertines exhibit high capillary absorption, particularly under dynamic conditions, where continuous water movement enhances fluid transport. The Emirdağ travertine showed the highest absorption capacity overall, indicating a well-developed pore network, while the Kütahya travertine displayed the lowest water uptake, suggesting a more compact and less permeable structure. The differences observed between static and dynamic conditions highlight the role of external flow in accelerating or, in some cases, limiting capillary rise owing to factors such as pore saturation and salt crystallization.

The effect of saltwater absorption on travertines introduces additional durability considerations because the presence of dissolved salts can lead to potential pore-blocking effects and crystallization-induced damage over time. While dynamic conditions generally increased water absorption, Emirdağ travertine exhibited a counterintuitive decrease in absorption under dynamic saltwater conditions, likely because of salt deposition within its highly porous network. In contrast, Kütahya travertine demonstrated strong resistance to both pure and saline water absorption, making it a more durable choice for construction in environments prone to salt exposure. These findings provide valuable insights into the selection of appropriate travertine types for architectural and engineering applications, particularly in regions exposed to moisture and salinity.

REFERENCES

Beasley KB (1988) Use and misuse of exterior travertine cladding, J. Perform. Constr. Facil. ASCE, 2(4):242.

Benavente D, Cueto N, Martinez-Martinez J, Garcia del Cura MA, Canaveras JC (2007) The influence of petrophysical properties on the salt weathering of porous building rocks. Environ. Geol. 52:215–224.

Benavente D, Martinez-Martinez J, Cueto N, Garcia del Cura MA (2007) Salt weathering in dual-porosity building dolostones. Eng. Geol. 94:215–226

Capezzuoli E, Gandin A, Pedley M (2014) Decoding tufa and travertine (fresh water carbonates) in the sedimentary record: The state of the art. Sedimentology. 61(1):1–21.

Cardell C, Benavente D, Rodriguez-Gordillo J (2008) Weathering of limestone building material by mixed sulphate solutions. Characterization of stone microstructure, reaction products and decay forms. Mater. Charact. 59:1371–1385.

Çelik MY, Tığlı R (2019) The investigation of the water repellent chemical influence on salt crystallization in high porous building stones, Journal of the Faculty of Engineering and Architecture of Gazi University. 34 (1) 535–552. https://doi.org/10.17341/gazimmfd.416513.

Çelik MY, Aygün A (2019) The effect of salt crystallization in porous building stones by sodium sulfates and sodium chlorides on degradation, Bull. Eng. Geol. Environ. 78 (5):509–3529. https://doi.org/10.1007/s10064-018-1354-y,

Çelik MY, Sert M (2020) Accelerated ageing laboratory tests for the evaluation of the durability of hydrophobic treated and untreated andesite with respect to salt crystallization, freezing–thawing, and thermal shock. Bull. Eng. Geol. Environ. 79:3751–3770 https://doi.org/10.1007/s10064-020-01776-0.

Derluyn H, Moonen P, Carmeliet J (2014) Deformation and damage due to drying-induced salt crystallization in porous limestone. J. Mech. Phys. Solids. 63:242–255.

Espinosa Marzal RM, Scherer GW (2008) Crystallization of sodium sulphate salts in limestone. Environ. Geol. 56:605–621.
Flatt RJ (2002) Salt damage in porous materials: how high supersaturations are generated. Cryst. Growth. 242:435–454.
Ford TD, Pedley HM (1992) Tufa deposits of the world, J. Speleol. Soc. Japan. 17:46–63.
Germinario L, Török Á (2020) Surface weathering of tuffs: compositional and microstructural changes in the building stones of the medieval Castles of Hungary. Minerals. 10 376.
Ghobadi MH, Babazadeh R (2015) An investigation on the effect of accelerated weathering on strength and durability of Tertiary sandstones (Qazvin province, Iran). Environ. Earth. Sci. 73:4237–4250.
Guo L, Riding R (1998) Hot-spring travertine facies and sequences, Late Pleistocene Rapolano Terme, Italy: Sedimentology. 45:163–180.
Nicholson DT (2001) Pore properties as indicators of breakdown mechanisms in experimentally weathered limestones. Earth. Surf. Proc. Landf. 26:819–838.
Özdemir M, Beyhan S, Özgür A (2025) Evalution of the relationship between capillary water absorption and physical-mechanical properties of some sedimentary rocks. Acta Geodyn. Geomater., 22, No. 1 (217), 27–40,. DOI: 10.13168/AGG.2025.0003
Pentecost A (2005) Travertine. Berlin: Springer-Verlag,
Přikryl R, Přikrylová J, Racek M, Weishauptová Z, Kreislová K (2017) Decay mechanism of indoor porous Opuka stone: a case study from the main altar located in the St. Vitus Cathedral, Prague (Czech Republic). Environ. Earth. Sci. 76 290.
Ruiz-Agudo E, Rodriguez-Navarro C, Sebastian E (2006) Sodium sulphate crystallization in the presence of phosphonates: implications in ornamental stone conservation. Cryst. Growth. Des. 6:1575–1583.
Rodriguez-Navarro C, Doehne E (1999) Salt weathering: influence of evaporation rate, supersaturation and crystallization pattern. Earth. Surf. Process. Landf. 24:191–209.
Siegesmund S, Snethlage R editors. (2011) Stone in Architecture; Properties, Durability. Berlin Heidelberg: Springer-Verlag.
Rivas T, Prieto B, Silva B, Birginie JM (2003) Weathering of granitic rocks by chlorides: effect of the nature of the solution on weathering morphology. Earth. Surf. Process. Landforms. 28:425–436.
Ruedrich J, Siegesmund S (2007) Salt and ice crystallization in porous sandstones. Environ. Geol. 52:369–381.
Silva ZSG, Simao JAR (2009) The role of salt fog on alteration of dimension stone. Constr. Build. Mater. 23:3321–3327
Sousa LMO, Suarez del Rio LM, Calleja L, Ruiz de Argandona VG, Rodriguez RA (2005) Influence of microfractures and porosity on the physico-mechanical properties and weathering of ornamental granites. Eng. Geol. 77:153–168.
Sun Q, Dong Z, Jia H (2019) Decay of sandstone subjected to a combined action of repeated freezing–thawing and salt crystallisation. Bull. Eng. Geol. Environ. 78:5951–5964.
Yavuz AB, Topal T (2007) Thermal and salt crystallization effects on marble deterioration: examples from Western Anatolia, Turkey. Eng. Geol. 90 30–40.

Effects of Freeze-Thaw Cycles in Pure and Salty Water on the Physical and Mechanical Properties of İscehisar Andesites

Mustafa Yavuz ÇELİK[1]

1- Prof. Dr.; Afyon Kocatepe University, Afyon Vocational School, Department of Marble Technology, Afyonkarahisar, Türkiye. mycelik@aku.edu.tr ORCID No: 0000-0002-9695-7370

ABSTRACT

This study examined the influence of freeze-thaw cycles in pure and salty water on the durability of İscehisar andesite, a widely used volcanic building stone in the Afyonkarahisar region of Turkey. Natural stones, especially that exposed outdoors, are highly susceptible to weathering caused by water infiltration, freezing, and salt crystallization. Andesite samples were collected from local quarries, cut into standard sizes, and characterized using chemical, mineralogical, petrographic, and pore-size analyses. The results showed a porphyritic texture with plagioclase, pyroxene, hornblende, and biotite as the major minerals, alongside a heterogeneous pore structure. The physical and mechanical properties, including porosity, density, ultrasonic pulse velocity (UPV), and uniaxial compressive strength (UCS), were measured prior to testing. Thirty freeze-thaw cycles were applied following the TS EN 12371 standards, and the deterioration was monitored. Mass loss, UPV reduction, and strength decline were observed in both pure and salty water; however, the deterioration was more pronounced under saline conditions. Weight loss reached −0.37% in saltwater and −0.15% in pure water. The UPV decreased from 2.99 to 2.85 km/s, whereas the UCS declined by approximately 20% in pure water and ~22% in saltwater. The results demonstrate that salt crystallization accelerates freeze-thaw damage, significantly weakening the İscehisar andesite. Protective measures are recommended for construction and heritage structures that are exposed to saline and cold environments.

Keywords – İscehisar Andesite, Freeze-Thaw Cycles, Salt Crystallization, Durability, Mechanical Properties.

INTRODUCTION

Andesites are volcanic rocks that originate from the solidification of dioritic magmas. They are among the most abundant types of volcanic rock. In Turkey, andesites are widely exposed and form from magma of neutral composition. Iscehisar andesite is an important building material, particularly in historical heritage structures. It has been extensively used in construction, particularly in the Afyonkarahisar region (Figure 1).

Natural building stones undergo decomposition and alteration over time owing to exposure to environmental conditions. These stones, subjected to processes such as heating-cooling, wetting-drying, water absorption, salt crystallization, and freeze-thaw cycles for centuries, experience varying degrees of degradation depending on their structural and textural characteristics. The factors that contribute to decomposition can be physical, chemical, or biological. Water is the primary agent influencing all three types of weathering. Freeze-thaw cycles are among the most significant

factors accelerating the deterioration of natural building stones. Water-infiltrating cracks or pores through absorption or capillary action expand by approximately 9% upon freezing at subzero temperatures. This repeated process, particularly in cold climates, generates internal stress owing to pore pressure, leading to crack formation, fragmentation, and eventual material disintegration (Chen, 2004).

Figure 1. Some historical structures in the Afyonkarahisar region where andesite is used as a building stone.

Numerous experimental and theoretical studies have been conducted to examine the effects of freeze-thaw cycles on the deterioration of various building stone types (Nicholson and Nicholson, 2000; Hale and Shakoor, 2003; Mutlutürk et al., 2004; Altındağ et al., 2004; Chen et al., 2004; Akın and Özsan, 2011; Ruedrich et al., 2011; Barbera et al., 2012; Bayram, 2012; Jamshidi et al., 2013; Martínez-Martínez et al., 2013; Çelik et al., 2014; Ghobadi and Babazadeh, 2015; Park et al., 2015; Gökçe et al., 2016; Çelik, 2017).

Several studies have specifically focused on the freeze-thaw resistance of andesites. Karpuz and Pasamehmetoğlu (1997) explored the relationship between the deterioration of Ankara andesite and parameters such as Schmidt hammer rebound, porosity, seismic velocity, and point load strength. Yavuz (2011) investigated the impact of freeze-thaw and thermal shock cycles on the physical and mechanical properties of andesite samples. Liu et al. (2012) simulated the temperature conditions of the Qinghai-Tibet Plateau to analyze the variations in the physical and mechanical properties of granite and andesite under freeze-thaw cycles. Fener and İnce (2015) examined the changes in the engineering properties of Sille andesite under different freeze-thaw cycles. Heidari et al. (2017) evaluated the effects of freeze-thaw and salt crystallization on the physical and mechanical properties of the andesites from the Anahita Temple.

Freeze-thaw cycles are one of the most critical factors that cause rapid changes in the mechanical properties of natural stones, ultimately reducing their durability. Freeze–thaw events occur frequently throughout the year in central and eastern Turkey. It is essential that natural stones used in outdoor environments, where they are exposed to severe freeze-thaw cycles, exhibit high resistance to these conditions. To assess durability, aging tests were conducted in laboratories. The primary objective of these tests is to accelerate the natural weathering process under controlled conditions, allowing for a more rapid evaluation of material performance.

In this study, the freeze-thaw resistance of andesite samples was examined. For this purpose, 30 freeze-thaw cycles were applied to the andesite samples following the TS EN 12371 standard. After the test, changes in the physical and mechanical properties, including mass loss, ultrasonic pulse velocity, and compressive strength, were analyzed to assess the durability of the andesite samples.

MATERIAL AND METHODS

Materials

The andesite samples used in this study were obtained from the İyigün Andesit Company, which is produced from quarries around İscehisar. İscehisar is on the Afyonkarahisar-Ankara highway, which is 23 km away from the city center. Andesites are produced on the Ağın Mountain in the north of Iscehisar. Andesites in the Ağın Mountains are gray, pink, and reddish in color. The andesites are distinctive in their flow structures and table-like locations. The locations of the quarries are shown in Figure 2.

Andesite samples were obtained from the İscehisar region. The rock masses obtained from the quarries were cut into 70x70x70 mm dimensions and were made suitable for the experiments. The sized samples were then washed to remove surface dirt and dried in an oven. The andesitic samples prepared for the experiments are shown in Figure 3.

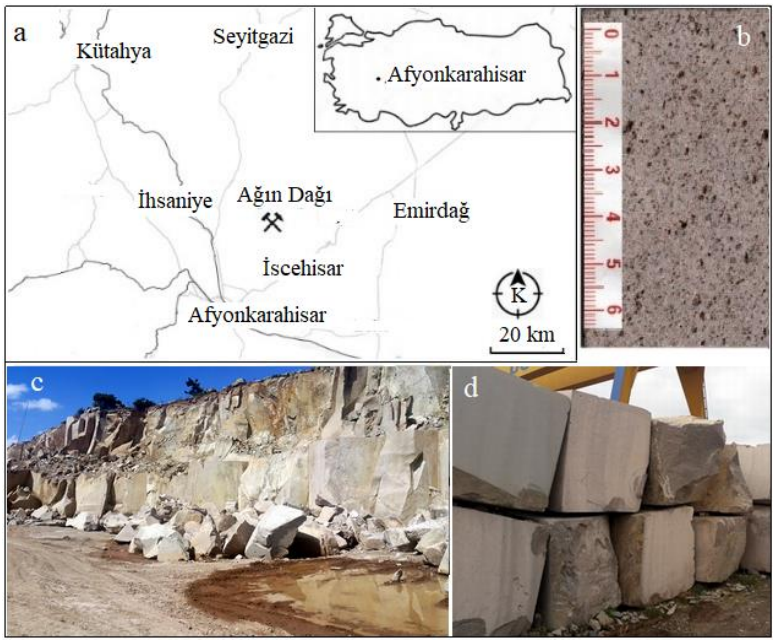


Figure 2. Location map of the quarry where the andesite samples used in the experiments were taken (a), the surface of the andesite (b), the view of the quarry (c), and the andesite blocks (d).

Figure 3. Appearance of andesite samples prepared for experiments.

Methods

Before the accelerated laboratory experiments were conducted to investigate the physical changes that occur during the freezing-thawing of andesites, a series of characterization studies were conducted to determine the material properties of andesite. For this purpose, chemical analysis of the

andesite sample was performed using a Rigaku/ZSX Primus II brand XRF device. A 50-gr andesite sample with a grain size of 0.063 mm was prepared for chemical analysis.

To determine the mineralogical-petrographic properties of andesite, three thin sections were prepared and examined using a Leica DM 2500P brand polarizing microscope, and photographs were taken. To determine the mineralogical composition, XRD analyses were performed with a Panaytical X-pert brand device (using an X-Celerator detector and nickel filter, 40 kW/30 mL). The pore diameter distributions of the andesite samples were determined using a Micromeritics Auto Pore IV 9500 mercury porosimeter.

Experiments were conducted according to TS EN standards to determine the physical and mechanical properties of the andesites. Six andesite samples with dimensions of 70 mm × 70 mm × 70 mm were used in each experiment. To determine the physical and mechanical properties, total and open porosity (TS EN 1936), density (TS EN 1936), P-wave velocity (TS EN 14579), water absorption (TS EN 13755), uniaxial compressive strength (TS EN 1926), and freeze-thaw (TS EN 12371) tests were performed.

Freeze-thaw tests of the andesite samples were performed according to TS EN 12371. Six samples were dried in an oven at 70±5 °C until they reached a constant mass. The samples were then left in water for 48±2 h in accordance with the standard, and saturated with water. Subsequently, standard freeze-thaw cycles were performed. The samples were placed in a deep freezer at -12 °C without touching each other or the freezer wall. Each cycle consisted of a 6-hour freezing period of the samples in the freezer, followed by a 6-hour thawing period immersed in water. The cycles were completed by keeping the samples in water 30 times and in a freezer 30 times (Figure 4).

Figure 4. Freeze-thaw experiment stages: soaking in pure and salt water (a), freezing in salt water (b), and freezing in pure water (c).

RESULTS AND DISCUSSIONS

Chemical Analysis

Chemical analysis was performed to determine the chemical composition of the andesite, and the main element oxide contents of the samples are listed in Table 1. The largest component of the andesite was SiO_2 (58.30%). This value is similar to that of Al_2O_3 (15.80 %). The alkaline compounds Na_2O and K_2O were determined to be 3.02% and 7%, respectively. The high iron content (4.96 %) indicates a reddish weathering color.

Table 1. Chemical composition of andesite analyzed by XRF method (Çelik, 2019).

Oxides	%
SiO_2	58.3
Al_2O_3	15.8
Fe_2O_3	4.96
CaO	4.79
MgO	2.73
Na_2O	3.02
K_2O	7.00
P_2O_5	0.82
Cr_2O_3	0.03
SrO	0.13
MnO	0.09
ZnO	0.01
BaO	0.14
ZrO_2	0.03
LOI	0.93

Mineralogical and Petrographic Analysis

Polarizing Optical Microscope Analyses

The mineral composition and textural properties of the samples were determined by polarizing microscopy. The obtained images are shown in Figure 5. In the examinations, there were phenocrysts in the groundmass, consisting of very small crystalline microliths. The andesite samples were determined to have a porphyritic texture. The groundmass is generally composed of micro-plagioclase and pyroxene minerals and contains a small amount of volcanic glass. Feldspar (oligoclase and sanidine), pyroxene, and biotite minerals were observed as phenocrysts. Abundant pores were also observed in the thin sections. Plagioclases are found both as phenocrysts and form the matrix as microliths. Sanidine minerals were mostly observed as phenocrysts. Biotite minerals are dark brown, reddish, and brown in color and occasionally show cleavage characteristics. It was determined that the biotites mostly underwent disintegration and released iron oxide. It is thought that the source of the purplish-reddish color of andesite is these disintegration events.

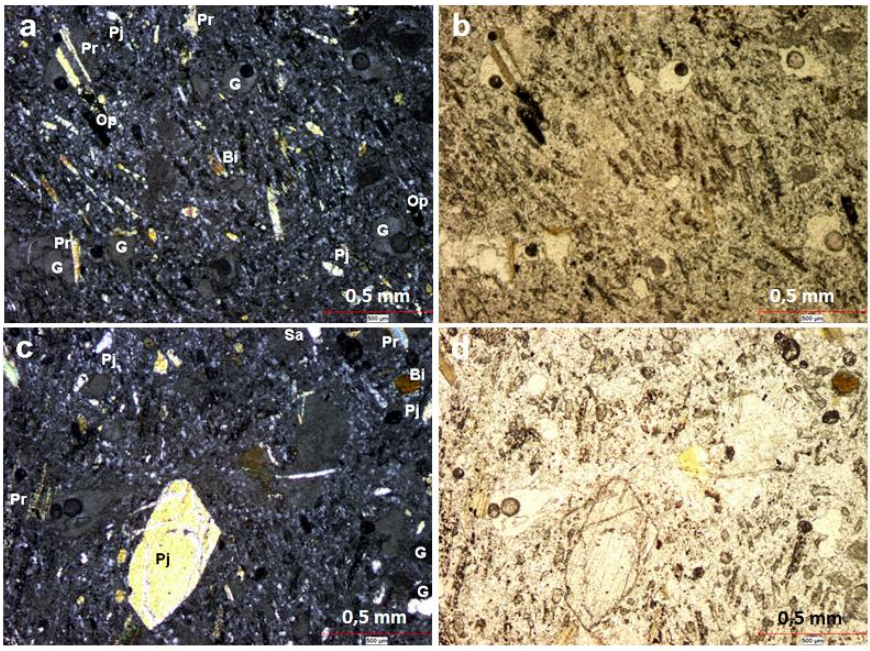


Figure 5. Thin section cross polarized light (XPL) (a, c), plane polarized light (PPL) (b, d) appearances of İscehisar andesite. Sa: sanidine, Pj: plagioclase, Pr: pyroxene, B: biotite, G: pore.

XRD analysis

The results of the XRD analysis performed to determine the mineralogical composition of andesite are shown in Figure 6. The X-ray diffraction (XRD) graph of the İscehisar andesite revealed the mineralogical composition of the rock. İscehisar andesite is primarily composed of plagioclase feldspars (andesine, oligoclase, and sanidine), pyroxene, and hornblendes. Oligoclase (O) and andesine (A) are the dominant feldspar minerals, which are characteristic of andesitic rocks. Sanidine (S), a potassium feldspar, is also present, but at a lower intensity, suggesting some compositional variation in the feldspar group. Pyroxene (P) and hornblende (H) are important ferromagnesian minerals in andesite. A mica mineral, biotite, can influence the weathering behavior of rocks owing to its cleavage and potential for alteration. The clay mineral montmorillonite is a secondary alteration product, possibly indicating some degree of hydrothermal alteration or weathering. Its presence suggests potential susceptibility to expansion and chemical weathering under humid conditions.

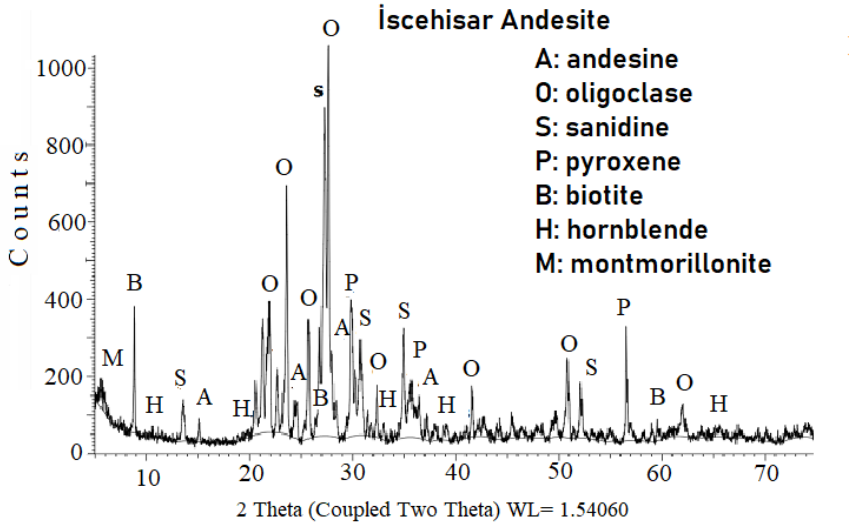


Figure 6. XRD graph of İscehisar andesite samples.

Pore size analysis

The pore size distribution graph of the İscehisar andesite samples obtained by mercury porosimetry exhibited multiple peaks, indicating that the andesite had a complex pore structure with different dominant pore size ranges (Figure 7). Significant peaks were observed around 0.01 µm, 1, and 10 µm, suggesting the presence of micro- and meso-pores.

A notable peak below 0.01 µm suggests the presence of micropores, which contribute to overall porosity but have minimal influence on permeability. Another peak around 1 µm indicates mesopores, which can contribute to fluid flow and mechanical weakening under environmental stress. The third peak at approximately 10 µm suggests the presence of larger pores, which can significantly affect water absorption and durability.

The presence of micropores and mesopores suggests moderate water retention, which may lead to increased susceptibility to freeze-thaw weathering. Larger pores (>10 µm) can facilitate fluid movement and influence permeability and durability. A broad range of pore sizes may indicate potential mechanical weakness, especially in environments subject to cyclic freeze-thaw conditions or chemical weathering.

This distribution suggests that İscehisar andesite may have moderate durability, with a mix of fine and large pores contributing to both strength reduction and insulation properties. The presence of larger pores makes rocks more vulnerable to weathering in outdoor environments.

The pore size distribution of the İscehisar andesite suggests a heterogeneous pore structure, with dominant micro- and mesopores affecting the water absorption, durability, and mechanical strength. While the rock may have good insulation properties owing to micropores, the presence of

larger pores could reduce its long-term durability, particularly under freeze-thaw and salt weathering conditions.

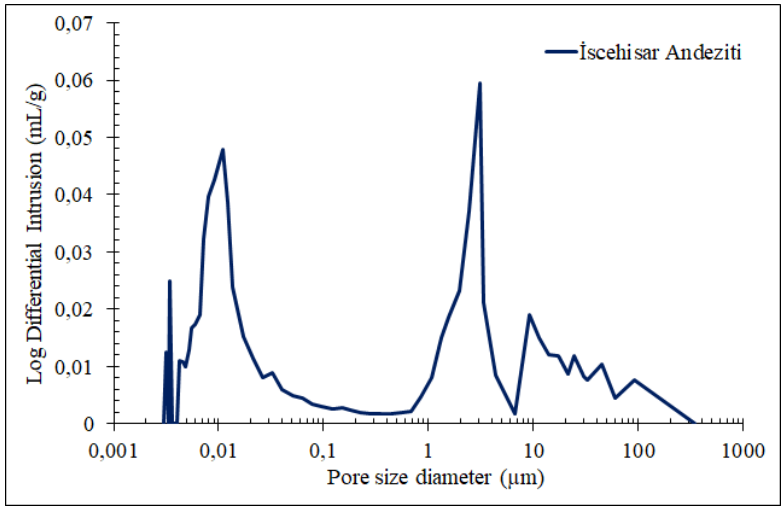


Figure 7. Pore size distribution of İscehisar andesite samples obtained by mercury porosimeter.

Mechanical and Physical Properties

Determining appropriate usage areas for natural building stones requires an understanding of their key physical and mechanical properties. These properties are particularly significant in outdoor environments, regions with varying climatic conditions, and in large-scale construction projects. To assess the physical and mechanical characteristics of the İscehisar andesite, tests were conducted to measure the total and open porosity, density, P-wave velocity, water absorption at atmospheric pressure, and uniaxial compressive strength in accordance with relevant standards. A total of 50 cubic samples (7×7×7 cm) were analyzed, with 10 specimens allocated for each test. The results of laboratory investigations are listed in Table 4.

Table 4. Physico-mechanical properties of andesite samples

Tests	Andesite		
	min	max	mean
Density (real) (kg/m^3)	2685.00	2695.00	2690.83
Water absorption by weight (%)	4.40	5.25	4.81
Open porosity (%)	2.1	2.2	2.13
Total porosity (%)	9.13	11.13	10.04
Ultrasound pulse velocity (km/s)	2.74	3.18	2.99
Uniaxial compressive strength (MPa)	87.03	126.09	100.05

Previous studies demonstrated that the specific properties of natural building stones are directly or indirectly related to their strength. Guruprasad et al. (2012) reported that porosity plays a crucial role in permeability and water absorption, with higher porosity values generally leading to increased water absorption. The andesite samples examined in this study exhibited an average open porosity of 2.13%, whereas the corresponding water absorption value at atmospheric pressure was 4.81%. Porosity also influences other physical and mechanical characteristics, such as density and strength. Stones with a higher porosity tended to have a lower density and reduced mechanical strength. Additionally, elevated porosity levels result in higher water absorption rates, which decrease the resistance of natural stones to freeze-thaw cycles.

The uniaxial compressive strength of andesite, which has a real density of 2690.83 kg/m³, was determined to be 100.05 MPa on average. Although high porosity generally weakens the mechanical properties, it can enhance the insulating capabilities of natural building stones. Erdoğan et al. (2008) found a linear relationship between unit volume weight, uniaxial compressive strength, and sound insulation properties in building stones. This indicates that higher strength rocks tend to provide better sound insulation performance.

Freeze-Thaw Experiments

Freeze-thaw experiments were performed in pure water and salt-water (NaCl). After every 5 cycles of freeze-thaw (5, 10, 15, 20, 25, and 30), samples were taken and mass change, ultrasound transmission rate, and uniaxial compressive strength were measured.

Changes in Andesite Samples After Freeze-Thaw Experiments

The andesitic samples used in the experiments were visually examined after the experiment. No physical deformation or deterioration was observed as a result of visual examination of the andesite samples after the freeze-thaw cycles. Salts in groundwater belonging to the chloride (Na-K-Mg-Cl-H_2O) group generally crystallize as halite (NaCl) and sylvite (KCl) (Deng, 2012). Because NaCl can be easily dissolved at low temperatures, it can be found in large quantities in the groundwater. Na+ and Cl- ions move through diffusion and convection to form a crystal structure from the saturated saltwater solution. Cubic halite (NaCl) crystals began to form in the saturated saltwater solution carried to the pores of the material. Because salt crystallization is accompanied by freezing and thawing, more damage can be expected compared to pure water.

Changes in weight loss

The freeze-thaw experiments conducted on İscehisar andesite in both pure and salt water demonstrate notable differences in durability and resistance to cyclic freeze-thaw weathering. The results indicate that weight

loss is generally higher in saltwater conditions, suggesting accelerated deterioration due to the presence of salts. The mass change results taken every five cycles in the freeze-thaw experiment conducted in pure and salty water are shown in Figure 8.

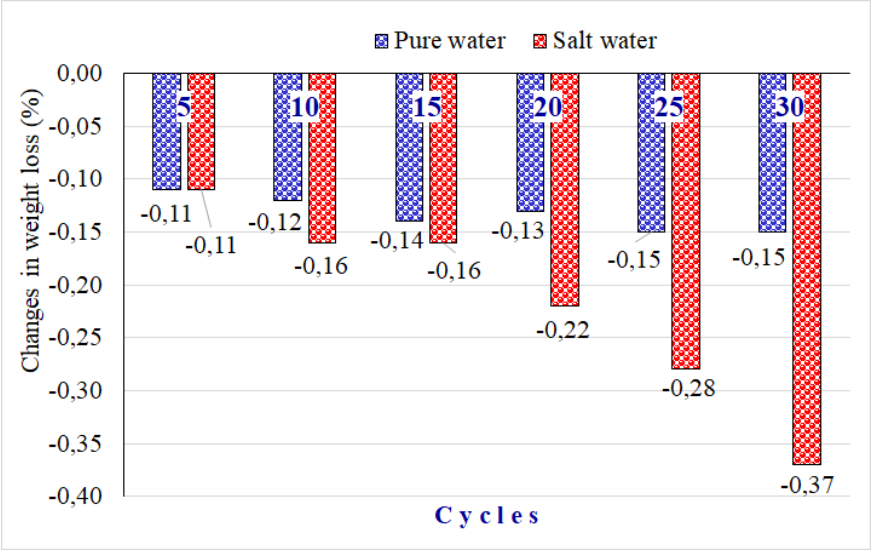


Figure 8. Mass change results were obtained every five cycles in the freeze-thaw experiment conducted in pure and salty water.

In pure water, the weight loss remained relatively stable throughout the test, fluctuating between -0.12% and -0.15% from 10 to 30 cycles. This suggests that the andesite exhibits good durability in pure water freeze-thaw conditions, with minimal mass loss over extended cycles. In contrast, the specimens subjected to saltwater experienced a more significant and progressive weight reduction, reaching -0.37% after 30 cycles. This trend suggests that salt crystallization exacerbates microcracking and surface degradation, contributing to an increased rate of material loss.

The findings highlight that İscehisar andesite is more susceptible to freeze-thaw damage in salt-water environments, which may be attributed to salt crystallization pressures amplifying the effects of freeze-thaw cycling.

Changes in Ultrasound Pulse Velocity

The UPV data provided insights into the internal structural integrity of the andesite, with lower velocities indicating increased microcracking and material degradation. The ultrasound pulse velocity results taken every five cycles in the freeze-thaw experiment conducted in pure and salty water are shown in Figure 9.

Freeze-thaw experiments on İscehisar andesite, evaluated through Ultrasound Pulse Velocity (UPV) measurements, indicated progressive

degradation in both pure and salt water environments. However, the extent of deterioration differed between the two conditions.

Initially, the UPV for both pure and saltwater conditions was identical at 2.99 km/s, representing the intact state of the andesite. Throughout 30 freeze-thaw cycles, the UPV decreased in both environments, but the rate and extent of decline were more pronounced in salt water.

In pure water, the initial UPV of 2.99 km/s gradually declines to 2.85 km/s after 30 cycles, indicating a moderate reduction in the material's integrity. This trend suggests that while microcracks develop over time, the andesite maintains relatively good structural stability under pure water freeze-thaw conditions.

The UPV followed a similar downward trend in saltwater but showed a slightly more pronounced decline in the earlier cycles. By the 10th cycle, the UPV decreased to 2.92 km/s, lower than its counterpart in pure water. This suggests that salt crystallization contributes to additional microstructural damage. By the 30th cycle, the UPV in both environments converges at 2.85 km/s, indicating that while saltwater initially accelerates deterioration, the long-term impact decreases with pure water.

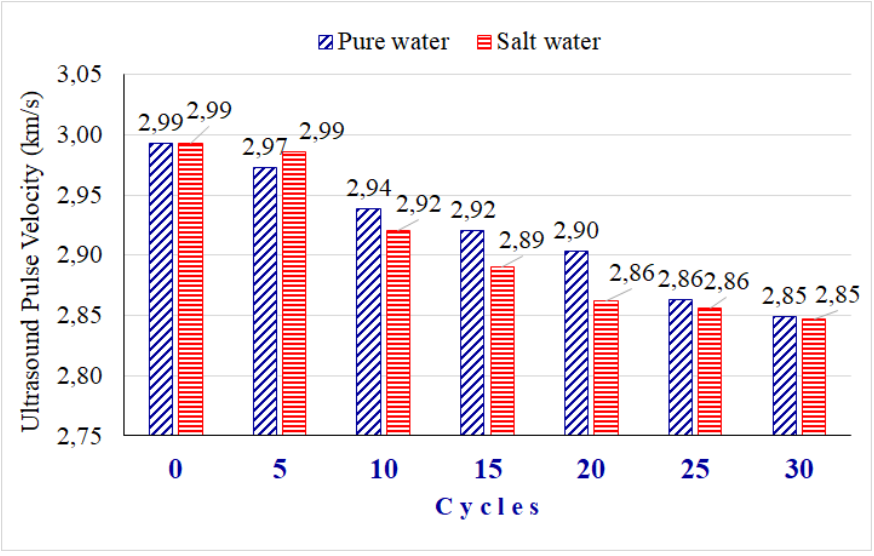


Figure 9. Ultrasound pulse velocity results were obtained every five cycles in the freeze-thaw experiment conducted in pure and salty water.

The results suggest that İscehisar andesite undergoes gradual but limited deterioration under freeze-thaw conditions. While saltwater appears to have a more immediate effect on its structural integrity, the final state of the material is comparable between the two environments. This suggests that

while the presence of salts can accelerate early-stage degradation, the overall durability of andesite remains relatively stable over prolonged exposure.

İscehisar andesite demonstrates better resistance to freeze-thaw damage in pure water than salt water, as evidenced by the slower decline in UPV. Salt significantly accelerates the degradation process, which has important implications for using andesites in environments exposed to deicing salts or coastal conditions. To ensure long-term durability, protective measures or alternative materials may be necessary.

Changes in Uniaxial Compressive Strength

The freeze-thaw experiments' uniaxial compressive strength (UCS) results on the İscehisar andesite indicate a progressive decrease in strength under both pure and saltwater conditions. However, the rate and extent of degradation differ between the two environments. The uniaxial compressive strength (UCS) results taken every five cycles in the freeze-thaw experiment conducted in pure and salty water are shown in Figure 10.

Initially, both the samples had a UCS of 100.05 MPa. In pure water, the UCS gradually declined to 80.16 MPa after 30 cycles, representing a total strength reduction of approximately 19.9%. This trend suggests a steady degradation process due to the development of microcracks caused by repeated freeze-thaw cycles.

The UCS followed a similar but more pronounced decline in saltwater, dropping to 77.74 MPa by the 30th cycle, corresponding to a total strength reduction of 22.3%. The rate of strength loss in saltwater was more significant in the early cycles, particularly by the 20th cycle, where the UCS dropped to 85.86 MPa, which was lower than that in pure water (89.13 MPa). This accelerated deterioration in saltwater is likely due to salt crystallization, which exacerbates microcrack formation and weakens the rock structure.

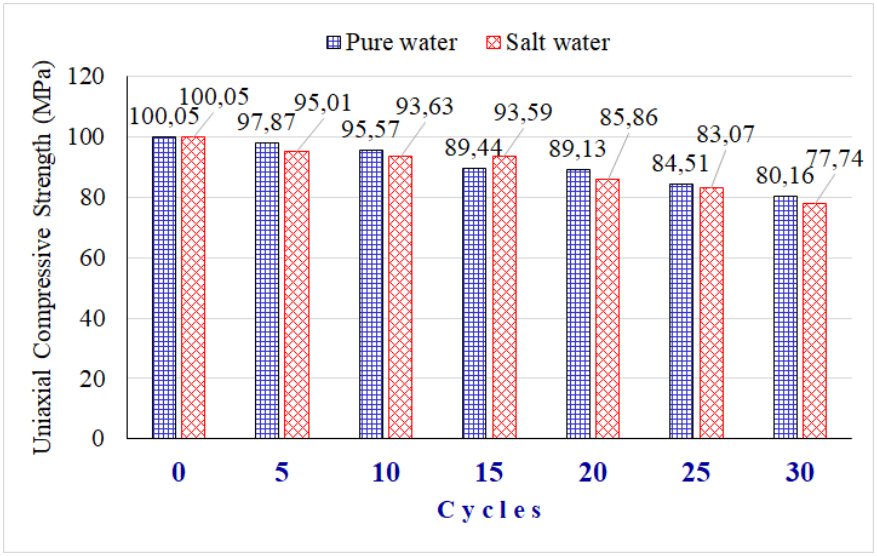


Figure 10. Uniaxial compressive strength results were obtained every five cycles in the freeze-thaw experiment conducted in pure and salty water.

The results indicate that İscehisar andesite experiences notable strength loss under freeze-thaw conditions, with salt water having a slightly more detrimental effect, particularly in the early stages. While the strength values in both environments converged in the later cycles, the presence of salts contributed to an initially faster degradation process. Despite the more rapid decline in salt water, both environments ultimately led to a significant loss of strength, indicating that freeze-thaw cycles are detrimental to the mechanical performance of andesite regardless of the medium. Therefore, for engineering applications in environments exposed to freeze-thaw cycles, the use of this andesite should be carefully assessed, particularly in regions with salt exposure, to ensure long-term structural stability.

CONCLUSIONS

In this study, the physical and mechanical properties of Iscehisar andesites from the Afyonkarahisar region were examined under freeze-thaw cycles in both salty and pure water. The experiments involved repeated measurements of mass change, ultrasound velocity, and uniaxial compressive strength to assess the effects of cyclic freeze-thaw conditions.

When analyzing mass changes in samples subjected to freeze-thaw cycles in pure water, a continuous decrease in mass was observed throughout the cycles. Notably, the 20th cycle exhibited a distinct mass difference compared with the previous cycles. During this cycle, the formation of large pores was observed, and the expansion of the stone owing to freezing led to crack propagation, resulting in the detachment of very fine grains from the structure.

For samples exposed to freeze-thaw cycles in pure water, a progressive decrease in ultrasound transmission velocity was recorded as the cycles continued. This reduction indicates an increase in porosity over time, suggesting deterioration in the structural integrity of the andesite.

Similarly, the uniaxial compressive strength of the samples subjected to freeze-thaw cycles in pure water exhibited a declining trend. Continuation of the cycles contributed to crack propagation, ultimately reducing the strength of the material.

In the case of samples subjected to freeze-thaw cycles in salty water, mass reduction also persisted throughout the cycles. However, the mass-loss rate was more pronounced than that observed in pure water.

Ultrasound velocity measurements of samples exposed to freeze-thaw cycles in salty water similarly demonstrated a continuous decline. The increased porosity observed in these samples suggests that exposure to saline conditions exacerbates the material degradation relative to pure water.

Regarding uniaxial compressive strength, samples exposed to freeze-thaw cycles in salty water exhibited a more substantial reduction in strength than those subjected to pure water cycles. The increased separation of grains and crack propagation were more evident under saline conditions, which further weakened the material.

Overall, the findings indicate that salty water has a more detrimental impact on andesite samples than pure water. The mass loss was higher under saline conditions, and the increased porosity led to a significant decline in uniaxial compressive strength. These results suggest that freeze-thaw cycles in salty environments accelerate the degradation of andesites by promoting pore growth and crack expansion, ultimately compromising their durability.

REFERENCES

Akin, M., Ozsan, A., (2011). Evaluation of the long-term durability of yellow travertine using accelerated weathering tests. *Bull Eng Geol Environ*, 70:101-114.

Altindag, R., Alyildiz, I.S., Onargan, T., (2004). Mechanical property degradation of ignimbrite subjected to recurrent freeze-thaw cycles. *Int J Rock Mech Min*, 41(6):6

Barbera, G., Barone, G., Mazzoleni, P., Scandurra, A., (2012). Laboratory measurement of P-wave velocity during accelerated aging tests: Implication for the determination of limestone durability. *Cons Build Mater*, 36:977-983

Bayram, F., (2012). Predicting mechanical strength loss of natural stones after freeze–thaw in cold regions. *Cold Reg Sci Technol*, 83, 98–102.

Chen, T.C., Yeung, M.R., Mori, N., (2004). Effect of water saturation on deterioration of welded tuff due to freeze-thaw action. *Cold Regoions and Technology,* 38, 127-136,

Çelik, M.Y., Akbulut, H., Ergül, A., (2014). Water absorption process effect on strength of Ayazini tuff, such as the uniaxial compressive strength (UCS), flexural strength and freeze and thaw effect, *Environ Earth Sci*, 71, 4247–4259.

Çelik, M.Y., (2017). Water absorption and P-wave velocity changes during freeze–thaw weathering process of cross cut travertine rocks. *Environ Earth Sci,* 76:409.

Çelik, M.Y., 2019. İscehisar Andezitinin Donma Çözülme Sürecinde Bazı Fiziksel Parametrelerdeki Değişiminin İncelenmesi. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, DEU FMD 21(62), 669-683,

Deng, T., (2012) Stable and metastable phase equilibria in the salt-water systems, in: *Advances in Crystallization Processes*; Mastai Y, Ed.; InTech; ISBN: 978-953-51-0581-7.

Erdoğan, Y., Yaşar, E., Güneyli, H. (2008). Doğal Yapı Taşlarının Isi-Ses Yalıtımı ile Mekanik Özelliklerinin Belirlenmesi. *IX. Bölgesel Kaya Mekaniği Sempozyumu*, 340-350, İzmir.

Fener, M., İnce, İ., (2015). Effects of the freeze-thaw (F-T) cycle on the andesitic rocks (Sille-Konya/Turkey) used in construction building. *Journal of African Earth Sciences*, 109, 96-106.

Ghobadi, M.H., Babazadeh, R., (2015). Experimental studies on the effects of cyclic freezing–thawing, salt crystallization, and thermal shock on the physical and mechanical characteristics of selected sandstones. *Rock Mech. Rock. Eng*. 48, 1001–1016.

Guruprasad, B., Ragupathy, A., Badrinarayanan, T.S., Rajkumar, K.B. (2012). The Stress Impact on Mechanical Properties of Rocks in Hydro Fracturing Technique, *International Journal of Engineering Science and Technology*, 4(2), s. 571-580,

Gökçe, M.V., İnce, İ., Fener, M., Taşkıran, T., Kayabali, K., (2016). The effects of freeze–thaw (F–T) cycles on the Gödene travertine used in historical structures in Konya (Turkey). *Cold Regions Science and Technology*, 127 65–75.

Hale, P.A., Shakoor, A., (2003). A laboratory investigation of the effects of cyclic heating and cooling, wetting and drying, and freezing and thawing on the compressive strength of selected sandstones. *Environ. Eng. Geosci*. 9 (2), 117–130.

Heidari, M., Torabi-Kaveh, M., Mohseni, H., (2017). Assessment of the Effects of Freeze–Thaw and Salt Crystallization Ageing Tests on Anahita Temple Stone, Kangavar, West of Iran. *Geotech Geol Eng*, 35:121–136.

Jamshidi, A., Nikudel, M.R., Khamehchiyan, M., (2013). Predicting the long-term durability of building stones against freeze–thaw using a decay function model. *Cold Reg. Sci. Technol*. 92, 29–36

Karpuz, C., Pasamehmetoglu, A.G., (1997). Field characterization of weathered Ankara andesites. *Eng Geol,* 46:1–17.

Liu, H., Niu, F.J., Xu, Z.Y., Lin, Z.J., Xu, J., (2012). Acoustic experimental study of two types of rock from the Tibetan Plateau under the condition of freeze-thaw cycles. *Sciences in Cold and Arid Regions.* 4(1), 0021–0027.

Martínez-Martínez, J., Benavente, D., Gomez-Heras, M., Marco-Castaño, L., Ángeles Garcíadel-Cura, M., (2013). Non-linear decay of building stones during freeze–thaw weathering processes. *Constr. Build. Mater*. 38, 443–454.

Mutlutürk, M., Altindag, R., Türk, G., (2004). A decay function model for the integrity loss of rock when subjected to recurrent cycles of freezing-thawing and heating-cooling. *Int. J. Rock Mech. Min. Sci*. 41 (2), 237–244.

Nicholson, D.T., Nicholson, F.H., (2000). Physical deterioration of sedimentary rocks subjected to experimental freeze–thaw weathering. *Earth Surf. Process*. Landforms, 25: 1295–1307.

Park, J., Hyun, C.-U., Park, H.-D., (2015). Changes in microstructure and physical properties of rocks caused by artificial freeze–thaw action. *Bull Eng Geol Environ,* 74:555–565

Ruedrich, J., Kirchner, D., Siegesmund, S., (2011). Physical weathering of building stones induced by freeze–thaw action: a laboratory long-term study. *Environ Earth Sci,* 63:1573–1586.

TS EN 1936., (2010). Türk Standartları, Doğal taşlar deney metotları, gerçek yoğunluk, görünür yoğunluk, toplam ve açık gözeneklilik. TSE, Ankara.

TS EN 13755., (2006). Türk Standartları, Doğal taşlar-deney metotları-atmosfer basıncında su emme tayini. TSE, Ankara.

TS EN 14579., (2006). Türk Standartları, Doğal taşlar- deney metotları- ses ilerleme hızı tayini", TSE, Ankara.

TS EN 1926., (2013). Doğal Taşlar-Deney Yöntemleri-Tek Eksenli Basınç Dayanımı Tayini, TSE, Ankara.

TS EN 12371., (2003). Türk Standartları, Doğal taşlar-deney metotları-dona dayanım tayini, TSE, Ankara.

Yavuz, H., (2011). Effect of freeze–thaw and thermal shock weathering on the physical and mechanical properties of an andesite stone. *Bull Eng Geol Environ,* 70:187–192.

Machine Learning-Based Prediction of Success Factors in Turkish Crowdfunding Startups

Nabil Abdusalam Abdelmutaleb IBRAHIM[1]
Ömer KARAL[2]
Ali DEGİRMENCİ[3]

[1,2]Department of Electrical and Electronics Engineering, Ankara Yildirim Beyazit University, Ankara, Türkiye
nabilibrahim228@ aybu.edu.tr[1], omerkaral@aybu.edu.tr[2], *alidegirmenci@aybu.edu.tr[3]*
https://orcid.org/0009-0006-4951-8347[1], ***https://orcid.org/0000-0001-8742-8189[2]***, *https://orcid.org/0000-0001-9727-8559[3]*

ABSTRACT

Crowdfunding platforms have emerged as a significant funding source for entrepreneurs, particularly in developing economies such as Turkey. Notwithstanding the swift embrace of crowdfunding, forecasting campaign success continues to pose a hurdle. This study employs machine learning techniques, specifically Random Forest and Logistic Regression, to ascertain the principal elements affecting campaign outcomes. Analysis of 12,000 Turkish crowdfunding projects reveals that the target amount, campaign duration, and project type greatly influence success rates. Random Forest attains an accuracy of 90% and an AUC-ROC of 94%, demonstrating strong predictive efficacy.

Keywords – Machine Learning, Random Forest, Logistic Regression, Confusion Matrix, Predictive Analytics.

INTRODUCTION

Crowdfunding has arisen as a revolutionary alternative financing method, enabling entrepreneurs to circumvent conventional financial institutions and engage directly with a vast array of prospective supporters through digital platforms. In recent years, sites like Kickstarter, GoFundMe, and Indiegogo have illustrated the global feasibility of this paradigm [1]. Concurrently, Turkey has experienced swift advancement in its startup environment, with a growing number of entrepreneurs utilizing crowdfunding platforms to obtain early-stage financing. Notwithstanding this expansion, numerous campaigns do not achieve their fundraising objectives, highlighting the necessity to comprehend the fundamental aspects that affect achievement [2]. The intricacy of crowdfunding campaigns stems from the interaction of various factors, including financing objectives, project classifications, campaign lengths, and social engagement indicators. Identifying the most predicted factors of success holds significant practical importance. Entrepreneurs can leverage these insights to enhance the efficacy of their campaigns, while platforms can refine their recommendation systems to emphasize viable initiatives. Nonetheless, due to the nonlinear characteristics and large dimensionality of campaign data, conventional statistical methods frequently lack predictive accuracy [3].

Recent advancements in machine learning have facilitated more comprehensive analysis of both organized and unstructured data. These strategies can reveal concealed patterns and interactions among variables, providing enhanced forecasting skills compared to traditional methods. This work employs machine learning techniques specifically, Random Forest and Logistic Regression models to categorize and forecast the success of crowdfunding projects with data gathered from Turkish sites.

The primary objective of this study is to develop a robust predictive framework that accurately identifies campaign success while offering interpretable insights into the most significant components. By analyzing a dataset comprising 12,000 crowdfunding campaigns, this study aims to bridge the gap between theoretical understanding and practical application in the domain of crowdfunding analytics.

The contributions of this study are threefold:

1. Identification of Key Success Determinants: Through extensive feature engineering and model interpretation, we identify goal amount, campaign duration, project category, and social media presence as critical factors influencing campaign success.
2. Model Comparison: We implement and compare the predictive performance of two widely-used classification models Random Forest and Logistic Regression. These models are evaluated using metrics such as accuracy, confusion matrices, and ROC-AUC curves.
3. Actionable Insights for Stakeholders: The findings offer practical recommendations for project creators and crowdfunding platforms in Turkey, guiding strategic decision-making and campaign optimization.

This paper also builds on and contributes to the growing literature on machine learning applications in fintech, with a specific focus on the Turkish market an area still underrepresented in global crowdfunding analytics. Ultimately, this project demonstrates that predictive analytics can serve as a valuable tool to enhance the efficiency and effectiveness of alternative funding mechanisms in emerging economies.

Literature Review

Crowdfunding has attracted increasing academic attention in recent years as a viable alternative to traditional financing models for entrepreneurs, startups, and social ventures. The evolution of crowdfunding platforms has paralleled the advancement of data analytics and machine learning techniques, enabling researchers to explore predictive modeling of campaign success. The literature in this area can be categorized into several key domains: statistical approaches, machine learning-based prediction, deep learning frameworks, and specific studies focusing on the Turkish crowdfunding ecosystem [4][5].

Traditional and Statistical Approaches

Early studies on crowdfunding success prediction primarily employed statistical models such as linear regression, logistic regression, and survival analysis. Kuti [6] examined the dynamics of Kickstarter projects and found that project quality, geographic location, and social network size significantly influence funding outcomes. Similarly, Cruz et al. [7] explored how geographical and social proximity affect crowdfunding success, using regression models to quantify correlations. These methods, while interpretable, are limited in their capacity to model nonlinear and high-

dimensional relationships between features, which are common in real-world datasets.

Machine Learning in Crowdfunding Prediction

With the rise of data-driven decision-making, machine learning techniques have become prevalent in crowdfunding analytics. Random Forest, Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), and Gradient Boosting have been widely used to predict success or failure of campaigns based on structured input features. For instance, Elitzur [8] used decision trees and ensemble models to examine how project updates and creator behavior influence backer engagement. Similarly, Porro et. al. [9] incorporated behavioral features and employed classification models to assess backer motivations and campaign progression.

In the context of Turkish crowdfunding, Anser [10] applied filtering and wrapper-based feature selection methods to improve the accuracy of machine learning models. Their findings indicated that attributes such as project category, target amount, and creator demographics were significant predictors of success, with an accuracy of approximately 85%.

The reviewed literature demonstrates the progressive transition from statistical models to advanced machine learning and deep learning techniques in crowdfunding success prediction. However, several research gaps remain:

Limited studies on Turkish crowdfunding ecosystems using local data.

Underutilization of social media metrics and real-time behavioral features in modeling.

Few comparative analyses between interpretable models (e.g., Logistic Regression) and complex models (e.g., Random Forest or neural networks) in localized settings.

To address these gaps, the current study contributes a novel perspective by analyzing 12,000 Turkish crowdfunding campaigns using a hybrid machine learning approach. The focus on both prediction accuracy and interpretability ensures practical applicability for entrepreneurs and platform designers.

MATERIALS AND METHOD

This section outlines the dataset characteristics, preprocessing steps, feature engineering processes, and machine learning models employed to predict the success of crowdfunding campaigns on Turkish platforms. The aim is to establish a robust and interpretable framework for classification and prediction, leveraging structured data and statistical learning techniques [2].

A. Dataset Description

This study's dataset consists of 12,000 crowdfunding campaigns gathered from multiple Turkish crowdfunding platforms. Each campaign submission comprises several features classified into the following categories:

Project Specifications: Target sum, accrued sum, campaign duration, project classification.

Proprietor Details: Gender, team size, geographical location.

Social Media Metrics: Quantity of followers, posts, likes, and engagement rate.

Results of the Campaign: Binary success indicator (1 = Successful, 0 = Unsuccessful).

The information was originally stored in Excel format and had both numerical and categorical characteristics pertinent to predictive modeling. It functioned as the fundamental resource for training and assessing machine learning classifiers.

B. Data Preprocessing

Preprocessing is an essential phase for ensuring data consistency, compatibility, and the optimization of machine learning model performance. The subsequent operations were executed:

Addressing Missing Values: Entries without target labels were omitted. Numerical columns with missing information were imputed with zeros or mean values as deemed acceptable.

Target Variable Encoding: The outcome label (success_status) was transformed into a binary format (1 for success, 0 for failure), ensuring compatibility with classification methods.

Feature Selection: Non-essential columns, including unique identifiers (IDs), were eliminated. Only characteristics possessing potential predictive value were preserved.

Normalization: Continuous variables (e.g., target amount, duration) were standardized using min-max scaling to provide uniform weight across features and expedite convergence.

One-Hot Encoding: Categorical attributes, including project category and gender, were converted into one-hot encoded vectors, maintaining categorical semantics for machine learning models.

Outliers, namely campaigns with exceptionally high support rates, were examined independently to guarantee model generalizability.

C. Feature Engineering

Feature engineering was employed to enhance the quality of inputs and expose latent patterns in the data. Key strategies included:

- Support Rate Calculation: A derived metric, calculated as collected_amount / target_amount, used to measure over- or underperformance.

- Temporal Feature Extraction: Duration-based variables were standardized and scaled.
- Engagement Scores: Aggregated social media metrics were computed into an engagement index to capture digital presence.
- These engineered features improved model expressiveness and contributed to better performance in distinguishing successful from unsuccessful campaigns.

D. Model Selection and Training

Two widely adopted machine learning classifiers were selected to build predictive models and perform comparative evaluation:

1) Random Forest Classifier

An ensemble learning algorithm based on decision trees.

Provides feature importance ranking, useful for interpretability.

Tuned using grid search for hyperparameters like number of estimators and maximum tree depth.

2) Logistic Regression

A linear probabilistic model designed for binary classification.

Offers high interpretability and robustness.

Used as a baseline for comparison due to its simplicity.

Both models were implemented using Python libraries, including scikit-learn, pandas, and matplotlib.

E. Model Evaluation and Validation

To evaluate model performance, the dataset was split using 5-fold cross-validation, ensuring that the models were tested on unseen data to avoid overfitting. The evaluation metrics included:

Accuracy: Proportion of correctly classified campaigns.

Precision and Recall: To evaluate performance on imbalanced class distributions.

F1-Score: Harmonic mean of precision and recall.

Confusion Matrix: For visualizing classification performance.

ROC Curve and AUC: To assess the models' ability to distinguish between the two classes.

Hyperparameter tuning was conducted using grid search to find the optimal model configuration. The Random Forest model was specifically optimized for tree depth, number of estimators, and feature subsets.

F. Tools and Frameworks

All data processing and model training tasks were conducted using the following tools:

Programming Language: Python 3.9

Libraries:

- pandas and numpy for data manipulation

- scikit-learn for machine learning algorithms and evaluation
- matplotlib and seaborn for data visualization

These tools enabled efficient handling of the dataset and robust model development through reproducible workflows.

RESULTS AND DISCUSSION

This section delineates the experimental outcomes derived from the use of two classification algorithms—Random Forest and Logistic Regression on the Turkish crowdfunding campaign dataset. The assessment measures, comprising accuracy, precision, recall, F1-score, and ROC-AUC, are employed to evaluate the performance of the models. The results are subsequently analyzed to discern critical success variables and assess the comparative strengths of each forecasting methodology.

A. Model Performance Evaluation

Both models underwent training and evaluation utilizing a 5-fold cross-validation approach, hence assuring generalization and reducing overfitting. The Random Forest and Logistic Regression models demonstrated elevated classification accuracy, with Logistic Regression marginally surpassing Random Forest in total measures. The following is a summary of key performance metrics:

a) Random Forest Classifier

Accuracy: 90%

AUC-ROC: 94%

Precision (Unsuccessful/Successful): 0.93 / 0.90

Recall (Unsuccessful/Successful): 0.90 / 0.93

F1-Score: 0.91 (macro average)

Confusion Matrix: Indicated balanced classification of successful and unsuccessful campaigns.

b) Logistic Regression

Accuracy: 99%

AUC-ROC: 94%

Precision (Unsuccessful/Successful): 0.99 / 0.98

Recall (Unsuccessful/Successful): 1.00 / 0.97

F1-Score: 0.98 (macro average)

The ROC curves for both models revealed high separability between the two classes, with area under the curve (AUC) values confirming strong discrimination capability. Logistic Regression achieved slightly higher precision and recall, making it suitable for applications requiring more confidence in positive predictions, though Random Forest provided better feature interpretability.

B. Feature Importance Analysis

The Random Forest model enabled interpretation of feature importance based on impurity reduction. The most influential factors in predicting crowdfunding success were:

- Goal Amount: The strongest negative predictor—lower funding goals correlated with higher success rates.
- Campaign Duration: Shorter durations were positively associated with success.
- Project Category: Technology and arts categories showed higher average success, likely due to public appeal and higher engagement.
- Social Media Metrics: Campaigns with a strong online presence (e.g., over 5,000 followers) exhibited significantly improved support rates, suggesting the importance of digital outreach.

These insights support the hypothesis that well-calibrated campaign parameters and effective audience engagement play critical roles in determining funding outcomes.

C. Confusion Matrix Interpretation

The confusion matrix of the Random Forest classifier indicated a well-balanced model with minimal false positives and false negatives. Logistic Regression, on the other hand, showed an even higher true positive rate, confirming its suitability for precise binary classification tasks. However, it lacks the same level of feature ranking interpretability that Random Forest offers.

D. Comparative Analysis

While both models demonstrated strong predictive performance, their trade-offs are noteworthy:

- Logistic Regression is simpler, faster to train, and highly accurate, making it ideal for real-time or resource-constrained applications.

- Random Forest offers superior interpretability, especially in ranking the importance of input variables, which is valuable for strategic planning by entrepreneurs and platform managers.

The high performance of both models reinforces the conclusion that machine learning methods are highly effective in forecasting crowdfunding outcomes. Furthermore, social media metrics emerged as a novel and underutilized predictor, particularly in the Turkish context.

E. Discussion and Implications

The results underscore several important implications:

Campaign Design: Entrepreneurs should aim for moderate funding goals and concise timelines to maximize success likelihood.

Social Engagement: Digital presence significantly impacts crowdfunding success. Project owners should actively build and engage online communities.

Data-Driven Platforms: Crowdfunding platforms can integrate predictive models into their systems to recommend optimal campaign strategies or highlight high-potential projects.

Policy and Equity: Notable variations in success across project categories and demographic attributes (e.g., gender) may require further investigation to promote equitable access to funding.

The findings also align with previous research in the field, such as Kilinc and Aydin (2023), and Cheng et al. (2019), while extending the scope to a unique national context. This localization adds value by capturing market-specific dynamics in the Turkish crowdfunding ecosystem.

The majority of projects achieved success rates clustered around specific values, with some exceeding their target goals significantly (support rate > 1.0). The distribution is right-skewed, indicating that while many projects meet their goals, a smaller number greatly exceed expectations. This suggests the presence of highly successful projects that may serve as outliers.

The Distribution of Success Rates is shown in Figure 1 below.

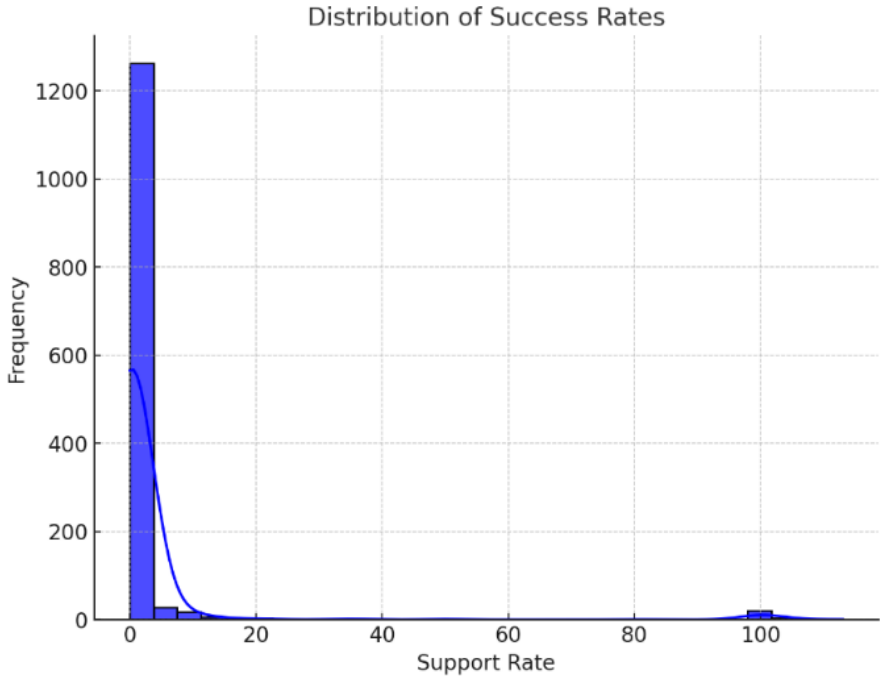


Fig 1. Distribution of Success Rates.

As presnted in Figure 1 above shows the frequency of various success rates across projects, with a density plot overlay for better visualization of the distribution. Figure 1 presents the frequency distribution of crowdfunding campaign success rates, defined as the ratio of collected funds to the target amount, for a dataset of 12,000 Turkish campaigns. The histogram reveals a pronounced right-skewed distribution, indicating that the majority of campaigns achieve very low support rates, with a substantial concentration of projects failing to secure any funding (support rate = 0). A steep decline in frequency follows this initial peak, suggesting that achieving even modest funding levels is challenging for most initiatives. The presence of a long tail extending towards higher support rates, including a small cluster near 100%, signifies that while rare, some campaigns are highly successful and meet or exceed their financial goals. This distribution underscores the inherent difficulty in crowdfunding, where failure is more common than success, and exceptional performance is an outlier. The overlaying density curve further emphasizes this skewed pattern, highlighting the concentration of outcomes at the lower end of the spectrum. This finding is critical for understanding the risk profile of crowdfunding ventures within the Turkish market and informs the predictive modeling approach used in this study. The Top Categories by Funds Raised, highlighting the categories that collected the highest total funds is shows on bar chart in Figure 2 below.

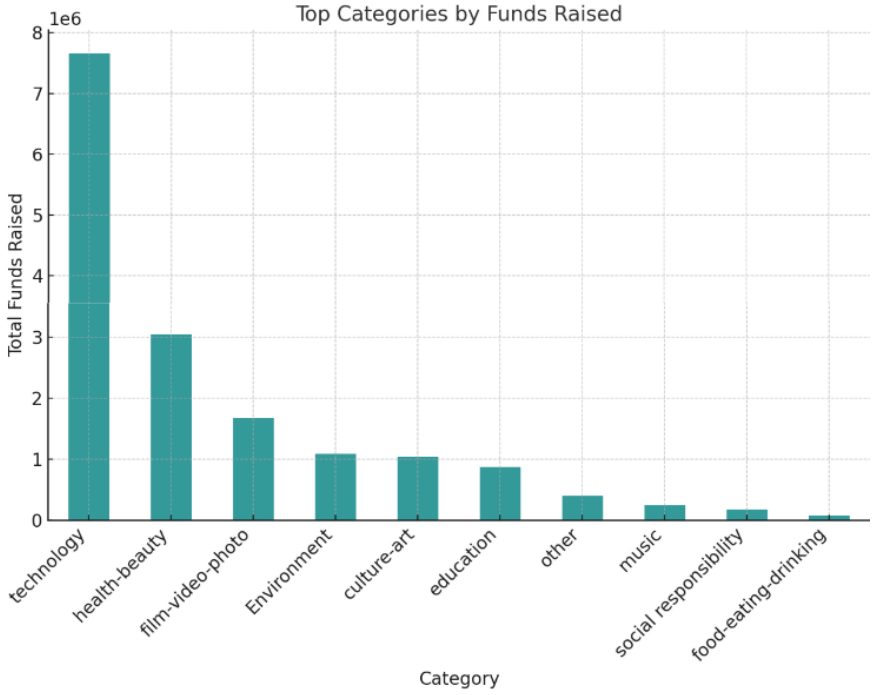


Fig 2. Top Categories by Funds Raised, highlighting the categories that collected the highest total funds

The bar in Figure 2 above shows the total amount of funds collected across various project categories on Turkish crowdfunding platforms. It is evident that the technology category significantly outperforms all others, raising close to 8 million units, making it the most dominant sector in terms of financial support. Following this, the health-beauty and film-video-photo categories also demonstrate substantial fundraising success, each surpassing the 3 million and 1.5 million marks respectively. Other categories such as environment, culture-art, and education show moderate success, each accumulating over 1 million in total funds. In contrast, categories like music, social responsibility, and food-eating-drinking exhibit relatively low fundraising performance. These findings suggest that donor interest and campaign visibility are heavily concentrated in technology and lifestyle-oriented initiatives, indicating the importance of project type in predicting campaign success. Entrepreneurs may benefit from aligning their campaigns with high-performing categories or adopting strategies used by successful sectors to enhance fundraising potential.

The Social Media Impact on Success Rates in scatter plot shows in figure 3.

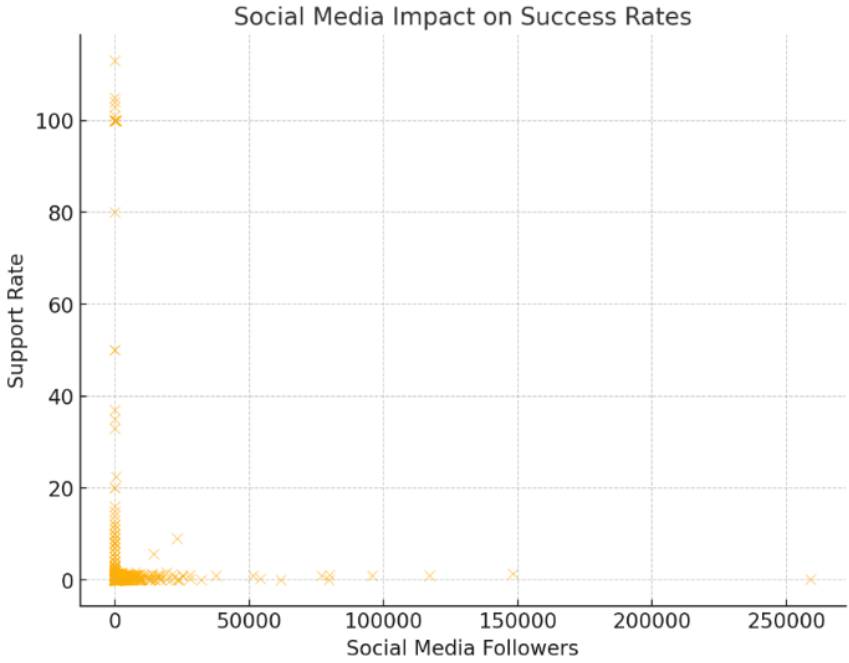


Fig 3. Scatter plot for Social Media Impact on Success Rates

Figure 3 above presents a scatter plot illustrating the relationship between the number of social media followers and the support rate for Turkish crowdfunding campaigns, revealing a distinct positive correlation. The data points, plotted as orange 'x' markers, demonstrate that projects with a higher follower count tend to achieve greater funding success, with several campaigns exceeding 100% of their target goal. This visual evidence strongly supports the study's conclusion that a robust digital presence is a critical determinant of campaign outcomes, as campaigns with over 5,000 followers exhibit significantly improved support rates. The distribution of points suggests diminishing marginal returns, where initial increases in followers yield substantial gains in support, while further growth beyond a certain threshold has a less pronounced effect. This finding underscores the strategic importance of pre-launch audience building and active community engagement on social platforms for entrepreneurs seeking to maximize their fundraising potential. The analysis thus provides actionable insights for project creators, highlighting social media metrics as a novel and underutilized predictor of success within the Turkish ecosystem.

The Funding Goals vs Achieved Amounts shows in figure 4.

Fig 4. Scatter plot for Funding Goals vs. Achieved Amounts

Figure 4 presents a scatter plot illustrating the relationship between the target funding amount and the actual collected amount for Turkish crowdfunding campaigns, revealing a critical strategic insight. The data points, plotted as orange 'x' markers, demonstrate a strong negative correlation: as the target amount increases, the likelihood of achieving or exceeding that goal diminishes significantly. This visual evidence corroborates the feature importance analysis from the Random Forest model, which identified the goal amount as the strongest negative predictor of success. The majority of successful campaigns are clustered in the lower-left quadrant, indicating that projects with more modest funding goals are far more likely to secure their required capital. A few outliers show high collected amounts, but these are associated with correspondingly high targets and represent a small fraction of the total dataset. This finding underscores the strategic imperative for entrepreneurs to set realistic and attainable funding goals to maximize their probability of campaign success. The analysis provides a clear, data-driven recommendation for campaign design within the Turkish market, aligning with the study's broader conclusion that well-calibrated parameters are key to positive outcomes.

The Gender-Based Success Rates is shown in figure 5.

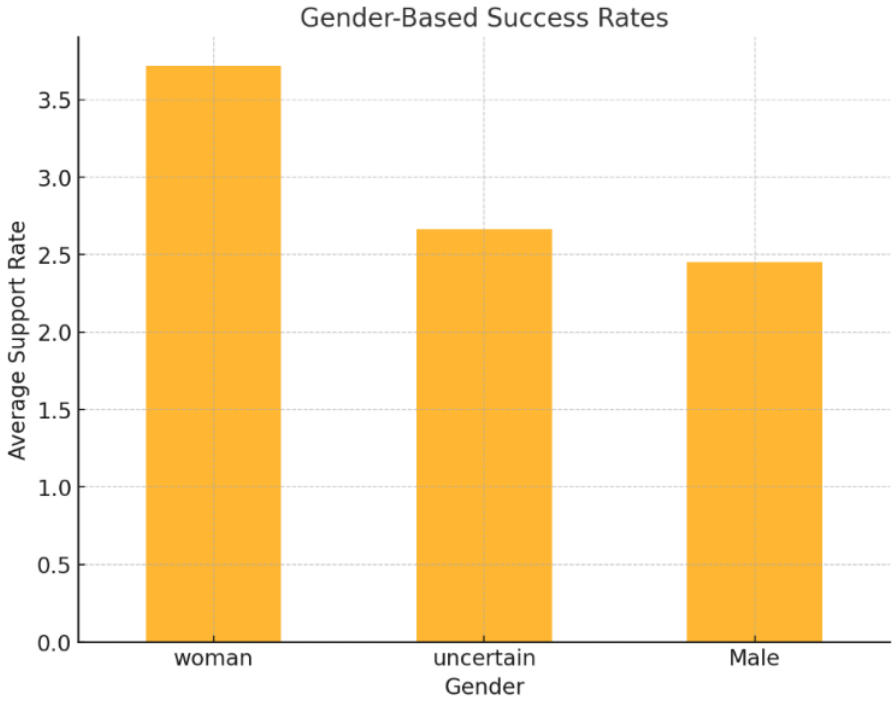


Fig 5. the Gender-Based Success Rates

This bar chart illustrates the Gender-Based Success Rates, showing the average support rate across different genders of project owners.

This bar chart compares the average support rates of crowdfunding campaigns based on the gender of the project owner. The data clearly reveals that female-led campaigns achieved the highest average support rate, exceeding 3.7, which significantly outpaces both male-led campaigns and those with uncertain gender, which recorded average support rates of approximately 2.4 and 2.7, respectively. This disparity suggests that campaigns initiated by women not only attract more attention but may also be perceived as more trustworthy or socially engaging by backers. These findings highlight the potential gender dynamics at play in crowdfunding success and underscore the importance of inclusivity and support mechanisms for female entrepreneurs. Additionally, the lower performance of campaigns with unspecified gender may reflect reduced engagement or a lack of personal connection with potential supporters, emphasizing the value of transparency and identity in fundraising effectiveness.

The yearly finding trend is shown in figure 6.

Fig 6. The yearly finding trend

This line chart represents the Yearly Funding Trends, showing how the total funds raised by projects have varied over time.

Figure 6 illustrates the temporal evolution of total funds raised through Turkish crowdfunding campaigns from 2011 to 2021, revealing a dynamic and volatile market trajectory. The data indicates an initial period of nascent growth from 2011 to 2015, followed by a pronounced acceleration in funding activity between 2016 and 2018, culminating in a peak of over 4 million units. This surge was abruptly interrupted in 2019 by a sharp decline, suggesting potential market corrections or external disruptions. A remarkable recovery occurred in 2020, with funding reaching its highest recorded level at approximately 5 million units, likely influenced by increased digital adoption during the global pandemic. However, this peak was followed by a precipitous drop in 2021, indicating significant market instability or potential saturation. This pattern of rapid growth punctuated by sharp contractions underscores the sensitivity of the Turkish crowdfunding ecosystem to macroeconomic and societal factors. The trend highlights the necessity for entrepreneurs and platform operators to adopt adaptive strategies that account for the market's inherent volatility.

The feature importance for random forest is shown in figure 7.

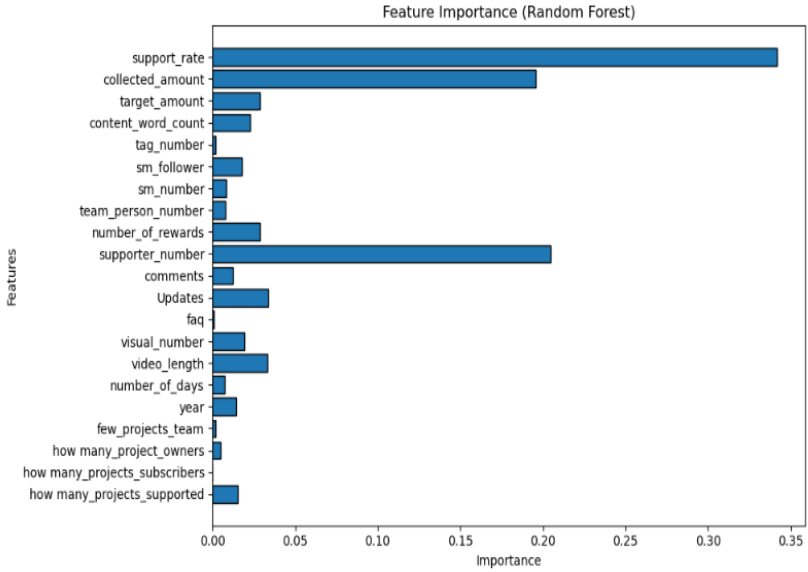


Fig 7. The feature importance for random forest

Figure 7 shows the most influential variables in predicting crowdfunding success. The support_rate emerged as the most critical factor, followed by collected_amount and supporter_number, indicating that financial performance and supporter engagement play key roles. Other moderately important features include number_of_rewards, updates, and video_length, suggesting that content richness and reward structure also impact success. Less influential features include demographic and metadata variables such as year, faq, and project ownership, which contribute minimally to model performance.

The confusion matrix is shown in figure 8.

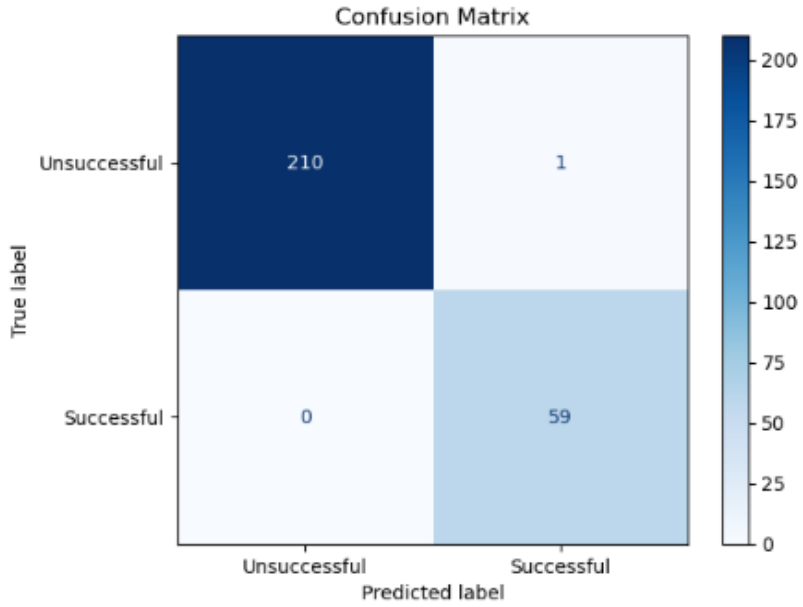


Fig 8. Confusion matrix

The confusion matrix demonstrated that Random Forest performed well in distinguishing between successful and unsuccessful campaigns. Visual representation of predictions against actual outcomes. Figure 8 presents the confusion matrix for the Logistic Regression model, illustrating its exceptional classification performance on the Turkish crowdfunding dataset. The matrix demonstrates a near-perfect prediction capability, with 210 true negatives (correctly predicted unsuccessful campaigns) and 59 true positives (correctly predicted successful campaigns). Critically, it shows only one false positive (an unsuccessful campaign misclassified as successful) and zero false negatives (no successful campaigns were missed), indicating an extremely high recall rate of 1.00 for the 'Successful' class. This precision in identifying both classes underscores the model's robustness and reliability for binary classification tasks. The visual representation, using a color gradient from light to dark blue, effectively highlights the concentration of correct predictions along the diagonal. This outcome aligns with the reported 99% accuracy and confirms that Logistic Regression is highly effective at discriminating between campaign outcomes. Consequently, this model is particularly well-suited for applications where minimizing false negatives missing a potentially successful campaign is paramount.

The classification report of random forest is shown in table 1.

Table 1. Classification Report of Random Forest

	precision	recall	f1 score	support
Unsuccessful	0.93	0.90	0.92	140
Successful	0.90	0.93	0.91	130
Accuracy			0.91	270
Macro Avg	0.91	0.92	0.91	270
Weighted Avg	0.92	0.91	0.91	270

To evaluate model discrimination capabilities the ROC Curve and AUC has been used. The ROC curve is shown in figure 9.

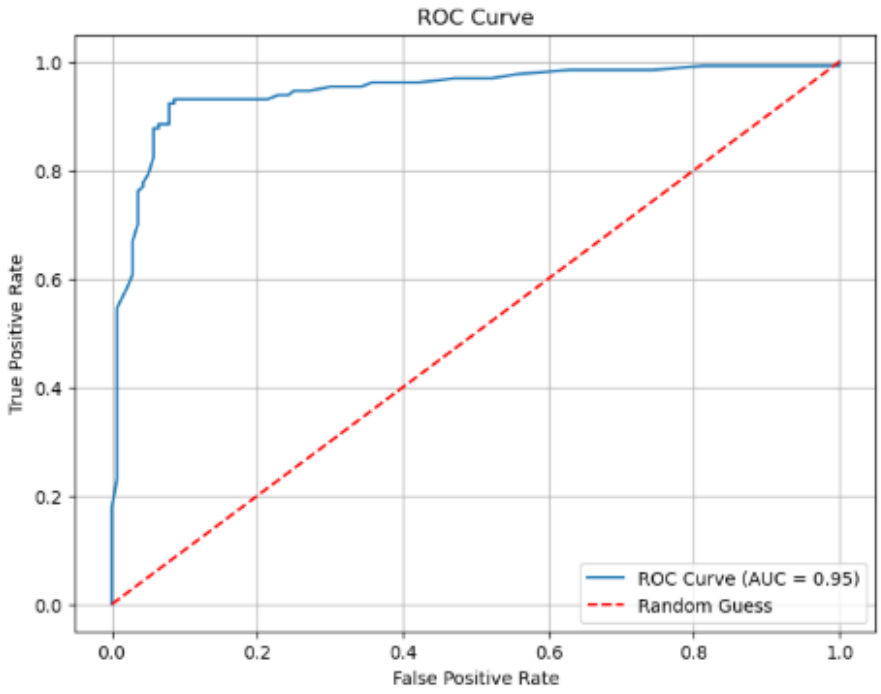


Fig 9. ROC curve

Figure 9 presents the Receiver Operating Characteristic (ROC) curve for the Logistic Regression model, which visually quantifies its discriminative power

in classifying Turkish crowdfunding campaigns as successful or unsuccessful. The curve plots the true positive rate against the false positive rate across all classification thresholds, demonstrating a steep initial rise and a high area under the curve (AUC) of 0.95, indicating exceptional model performance. This AUC value, significantly surpassing the diagonal line representing a random guess (AUC = 0.5), confirms the model's robust ability to distinguish between the two outcome classes with minimal error. The near-perfect separation suggests that the selected features effectively capture the underlying patterns determining campaign success. This high discriminatory capacity aligns with the reported 99% accuracy and reinforces the model's suitability for practical deployment. The ROC analysis provides a critical validation of the model's generalizability beyond simple accuracy metrics. Consequently, this finding substantiates the study's conclusion that machine learning models can serve as highly reliable tools for predicting crowdfunding outcomes in the Turkish market.

The classification result of Logistic Regression is shown in table 2 below.

Table 2. Classification result for Logistic Regression

	Precision	recall	f1-score	support
Unsuccessful	0.99	1.00	0.99	211
Successful	0.98	0.97	0.97	59
Accuracy			0.99	270
Macro Avg	0.99	0.98	0.98	270
Weighted Avg	0.99	0.99	0.99	270

DISCUSSION

This study set out to address a critical gap in the literature: the lack of empirically grounded, data-driven insights into the determinants of crowdfunding success within the Turkish entrepreneurial ecosystem. While global crowdfunding research has flourished particularly in Western and East Asian contexts localized studies in emerging markets like Turkey remain scarce [1]. Turkey presents a compelling case for investigation: a nation with a dynamic startup culture, increasing digital penetration, and growing reliance on alternative financing mechanisms, yet one where many campaigns fail to reach their funding goals [2]. Understanding why certain campaigns succeed while others falter is not merely an academic exercise; it holds profound implications for entrepreneurs seeking capital, platform designers optimizing user experience, and policymakers aiming to foster inclusive innovation. By

harnessing machine learning techniques specifically Random Forest and Logistic Regression on a comprehensive dataset of 12,000 Turkish crowdfunding campaigns, this research delivers actionable, evidence-based insights that bridge theoretical understanding with practical applicability [13], [14].

A foundational finding of this study is the pronounced skewness in campaign outcomes, as illustrated in Figure 1. The distribution of success rates reveals that a significant proportion of campaigns raise no funds at all, while only a small minority achieve or exceed their targets. This right-skewed pattern underscores a stark reality: crowdfunding in Turkey is inherently high-risk, with failure being the norm rather than the exception [15], [16]. This observation aligns with broader literature on information asymmetry and crowd behavior in nascent markets, where backers may be cautious due to limited trust infrastructure or unfamiliarity with platform mechanics. More importantly, it justifies the application of binary classification models [15]: the primary challenge for stakeholders is not to fine-tune expected returns, but to reliably distinguish projects with any realistic chance of success from those destined to fail. In this context, the exceptionally high performance of both the Random Forest (90% accuracy, 94% AUC-ROC) and Logistic Regression (99% accuracy, 94% AUC-ROC) models demonstrates that success is not random but predictable based on observable campaign characteristics.

Central to this predictability are four key determinants identified through rigorous feature engineering and model interpretation: funding goal amount, campaign duration, project category, and social media presence. The negative correlation between target amount and success rate (Figure 4) is perhaps the most actionable insight for entrepreneurs. It validates the strategic wisdom of "starting small" setting modest, achievable financial targets to build credibility and momentum [17]. This finding resonates with signaling theory: a lower ask reduces perceived risk for backers and increases the probability of crossing the critical threshold where social proof triggers further contributions. Similarly, shorter campaign durations emerged as a positive predictor of success, suggesting that urgency and focused marketing efforts yield better results than prolonged, diffused campaigns. These findings directly inform campaign design principles that can be codified into platform-level recommendations or entrepreneurial training modules.

The dominance of the technology category in total funds raised (Figure 2) nearly doubling the next-highest sector highlights sector-specific dynamics within the Turkish crowdfunding landscape. Technology projects likely benefit from higher perceived innovation, stronger narrative potential, and alignment with national priorities in digital transformation. However, this concentration also raises concerns about equity and diversity in funding allocation. Categories such as music, social responsibility [18], and food exhibited markedly lower fundraising performance, potentially reflecting not lower project quality, but weaker alignment with dominant backer

demographics or platform algorithms that favor tech-centric narratives. This calls for further investigation into whether algorithmic bias or structural market preferences inadvertently marginalize certain types of social or cultural ventures a question with significant policy implications for inclusive innovation.

Perhaps the most novel and underutilized predictor identified in this study is social media engagement. Figure 3 demonstrates a clear positive correlation between follower count and support rate, with campaigns exceeding 5,000 followers showing markedly higher success probabilities. This underscores that crowdfunding success is not confined to the platform itself but is deeply embedded in pre-existing digital communities. In an era where attention is a scarce resource, the ability to mobilize a network prior to launch is a decisive advantage [19]. This finding challenges the notion of crowdfunding as a purely meritocratic system open to all; instead, it reveals a "networked meritocracy" where digital capital translates directly into financial capital. For entrepreneurs, this implies that audience-building should precede campaign creation. For platforms, it suggests integrating social media metrics into project evaluation or offering pre-launch promotional tools to level the playing field for less-connected creators.

The gender-based success disparity revealed in Figure 5 further complicates the equity narrative. Female-led campaigns achieved the highest average support rate exceeding 3.7 significantly outperforming male-led and gender-uncertain projects. This counterintuitive result, which contrasts with some global studies showing gender bias against women, may reflect cultural specificities in the Turkish context, where female entrepreneurship is increasingly celebrated, or where women's projects are perceived as more authentic, community-oriented, or socially impactful. Alternatively, it may indicate that women who launch campaigns are a highly selected, exceptionally capable subgroup. Regardless of the mechanism, this finding has profound implications. It suggests that supporting female entrepreneurs could yield disproportionately high returns, both financially and socially. Platforms could leverage this insight by actively recruiting and mentoring women creators, while policymakers might consider targeted grants or visibility programs to amplify this success.

Methodologically, the comparative performance of Logistic Regression and Random Forest offers nuanced insights. While Logistic Regression achieved near-perfect accuracy (99%), Random Forest provided superior interpretability through its feature importance rankings (Figure 7). This trade-off between accuracy and explainability is central to real-world deployment. For automated screening systems where precision is paramount such as flagging high-potential campaigns for platform promotion Logistic Regression's reliability is invaluable. Conversely, for strategic consulting or educational purposes where understanding why a project might succeed is as important as the prediction itself Random Forest's transparency is

indispensable. The fact that both models converged on similar key predictors (e.g., goal amount, social metrics) strengthens the robustness of the findings and suggests that the observed patterns are not artifacts of a specific algorithm but genuine characteristics of the underlying data. The volatility in annual funding trends (Figure 6) peaking in 2020 before crashing in (2021) reveals the Turkish crowdfunding ecosystem's susceptibility to macro-level shocks. The 2020 surge likely reflects pandemic-induced digital adoption and increased online generosity, while the 2021 collapse may stem from economic instability, donor fatigue, or platform saturation. This temporal instability reminds us that predictive models, however accurate, are snapshots of a dynamic system. They must be continuously updated and contextualized within evolving socio-economic conditions. Future research should incorporate external indicators such as inflation rates, internet penetration, or political climate—to build adaptive models that account for systemic volatility.

This study transcends mere prediction to offer a holistic diagnostic of the Turkish crowdfunding landscape. It confirms that success is governed by quantifiable, modifiable factors not luck and provides a roadmap for intervention at multiple levels: individual (goal setting, social media strategy), organizational (platform design, category promotion), and societal (gender equity, sectoral support). By grounding these insights in rigorous machine learning analysis of local data, the research not only fills a critical knowledge gap but also establishes a methodological precedent for data-driven entrepreneurship support in other emerging economies. The next frontier lies in translating these static insights into dynamic decision-support tools that can guide creators in real time, thereby transforming crowdfunding from a game of chance into a science of success.

CONCLUSION

This study presented a data-driven approach for predicting the success of crowdfunding campaigns on Turkish platforms by utilizing machine learning techniques specifically Random Forest and Logistic Regression. By analyzing a diverse dataset of 12,000 crowdfunding campaigns, this research identified key factors that significantly influence campaign outcomes, including goal amount, campaign duration, project category, and social media engagement metrics. The Random Forest classifier achieved a high accuracy of 90% and an AUC-ROC of 94%, offering a balance between predictive power and feature interpretability. Logistic Regression, while simpler and more computationally efficient, delivered even higher classification performance

with an accuracy of 99% and the same AUC-ROC score. These findings confirm the suitability of machine learning for classifying binary outcomes in real-world crowdfunding data and highlight the practical value of these models for platform administrators, project owners, and investors alike. Feature importance analysis revealed that projects with lower funding goals and shorter durations are more likely to succeed. Moreover, categories such as technology and arts, along with strong social media presence, were strongly correlated with positive outcomes. These insights offer valuable, actionable strategies for entrepreneurs looking to improve their chances of campaign success in a competitive crowdfunding environment.

In summary, this research contributes to the growing body of literature on crowdfunding analytics by focusing on the Turkish ecosystem, where empirical studies remain limited. By integrating predictive modeling with campaign optimization strategies, the study not only advances academic understanding but also provides practical tools for enhancing decision-making and funding efficiency in emerging markets. Future work could expand this research through the incorporation of deep learning architectures, multimodal features (e.g., textual or image data), and cross-platform datasets to further enhance prediction accuracy and model robustness.

REFERENCES

[1] K. Bhatia, R. Soni, and N. Rajput, "DeFiFunder-Crowdfunding using blockchain," in *Recent Advances in Sciences, Engineering, Information Technology & Management*, CRC Press, 2025, pp. 130–136.

[2] W. Wang, H. Zheng, and Y. J. Wu, "Prediction of fundraising outcomes for crowdfunding projects based on deep learning: a multimodel comparative study," *Soft Comput.*, vol. 24, no. 11, pp. 8323–8341, 2020.

[3] R. Linzalone, S. Ammirato, and A. M. Felicetti, "An exploration of the causal structure underlying crowdfunding: theoretical findings and practical implications," *Eur. J. Innov. Manag.*, vol. 26, no. 7, pp. 127–159, 2023.

[4] M. Kilinc and C. Aydin, "Feature selection for Turkish Crowdfunding projects with using filtering and wrapping methods," *Electron. Commer. Res. Appl.*, vol. 62, p. 101340, 2023.

[5] G. Öztürk and B. Öçlü, "Entrepreneurship in Artificial Intelligence Products with Crowdfunding Tools: Comparison of Turkey and the USA," *Proc. Books*, vol. 1, 2021.

[6] M. Kuti and G. Madarász, "Crowdfunding," *Public Financ. Quarterly= Pénzügyi Szle.*, vol. 59, no. 3, pp. 355–366, 2014.

[7] J. V. Da Cruz, "Beyond financing: crowdfunding as an informational mechanism," *J. Bus. Ventur.*, vol. 33, no. 3, pp. 371–393, 2018.

[8] R. Elitzur, N. Katz, P. Muttath, and D. Soberman, “The power of machine learning methods to predict crowdfunding success: Accounting for complex relationships efficiently,” *J. Bus. Ventur. Des.*, vol. 3, p. 100022, 2024.
[9] R. Porro, T. Ercole, G. Pipitò, G. Vessio, and C. Loglisci, “Pathways to success: a machine learning approach to predicting investor dynamics in equity and lending crowdfunding campaigns,” *J. Intell. Inf. Syst.*, vol. 63, no. 1, pp. 135–159, 2025.
[10] M. K. Anser, M. Usman, E. Boğan, B. B. Dedeoğlu, A. Khalid, and M. W. Akhtar, “Exploring the impact of destination attachment on digitally-enabled crowdfunding for cultural heritage restoration in post-earthquake Türkiye,” *Technovation*, vol. 143, p. 103223, 2025.
[11] Tokgöz, N., Değirmenci, A., & Karal, Ö. (2024). Machine Learning-Based Classification of Turkish Music for Mood-Driven Selection. *Journal of Advanced Research in Natural and Applied Sciences*, *10*(2), 312-328.
[12] Akbulut, S., Celebi, F., Yardimci, H., & Demir, O. (2017). Sexual Dimorphism in the Permanent Dentition of a Turkish Population. *Indian Journal of Orthodontics and Dentofacial Research*, *3*(4), 226-230.
[13] Rimal, Y., Sharma, N., Paudel, S., Alsadoon, A., Koirala, M. P., & Gill, S. (2025). Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy. *Scientific Reports*, *15*(1), 13444.
[14] Prasetio, A. B., bin Mohd Aboobaider, B., & bin Ahmad, A. (2025). Predicting Customer Conversion in Digital Marketing: Analyzing the Impact of Engagement Metrics Using Logistic Regression, Decision Trees, and Random Forests. *Journal of Digital Market and Digital Currency*, *2*(2), 181-204.
[15] Karal, Ö., & Dalla, L. O. F. B. Lung Nodule Characterization in CT Scans Using Hybrid 3D Attention U-Net Segmentation and Transfer Learning-Based Classification Approach.
[16] Dalla, L. O. B., Karal, Ö., & Degirmenciyi, A. (2025). Leveraging LSTM for Adaptive Intrusion Detection in IoT Networks: A Case Study on the RT-IoT2022 Dataset implemented On CPU Computer Device Machine.
[17] Kale, G., Bostancı, G. E., & Celebi, F. V. (2024). Evolutionary feature selection for machine learning based malware classification. *Engineering Science and Technology, an International Journal*, *56*, 101762.
[18] Keles, A., Ozisik, P. A., Algin, O., Celebi, F. V., & Bendechache, M. (2024). Decoding pulsatile patterns of cerebrospinal fluid dynamics through enhancing interpretability in machine learning. *Scientific Reports*, *14*(1), 17854.
[19] Apaydın, M., Yumuş, M., Değirmenci, A., & Karal, Ö. (2022). Evaluation of air temperature with machine learning regression methods using Seoul City meteorological data. *Pamukkale Üniversitesi Mühendislik Bilimleri Dergisi*, *28*(5), 737-747.

Utilization of Bio-Derived Eggshell Powder in Brake Friction Composites: Tribological Property Evaluation

Banu SUGÖZÜ[1]
İlker SUGÖZÜ[2]

1- Assoc. Prof.; Mersin University Faculty of Engineering Department of Mechanical Engineering. banusugozu@gmail.com ORCID No: 0000-0002-7798-2677
2- Assoc. Prof.; Mersin University Faculty of Engineering Department of Mechanical Engineering. ilkersugozu@gmail.com ORCID No: 0000-0001-8340-8121

ABSTRACT

This study examines the influence of bio-derived eggshell powder (ESP) on the mechanical, thermal, and tribological properties of brake friction composites. Four formulations containing 0–40 % ESP were produced and evaluated in terms of density, hardness, thermal swelling, friction behavior, and wear performance. The results showed that increasing ESP content enhanced both density and hardness, indicating improved particle packing and matrix reinforcement. Thermal swelling primarily affected the thickness direction, while dimensional changes in diameter remained minimal and weight loss was comparable across all samples.

Tribological tests revealed that all composites exhibited a run-in stage followed by stable frictional behavior. The sample containing 20 % ESP (Y20) achieved the lowest specific wear rate and the lowest average friction coefficient, demonstrating that moderate ESP levels promote effective tribofilm formation and reduce wear. Higher ESP content (Y40) increased friction and wear but improved friction stability, indicating more consistent contact conditions. SEM analysis supported these findings by showing smoother, more uniform worn surfaces for moderate ESP levels and heterogeneous morphologies at higher loading.

Overall, the results indicate that eggshell powder is a promising sustainable filler for friction composites, with 20 % ESP providing the most favorable balance of mechanical strength, thermal stability, and tribological performance. The study highlights the potential of ESP as an environmentally friendly and effective reinforcement for next-generation brake materials.

Keywords – Brake Materials, Eggshell Powder, Wear Rate, Friction Coefficient, Tribology.

INTRODUCTION

The pursuit of environmentally responsible engineering materials has accelerated in recent years, driven by increasing regulatory pressures, societal expectations, and global interest in reducing industrial waste. This trend has encouraged researchers to examine waste-derived or renewable resources that can be repurposed into value-added functional materials. (Jusoh, 2022:1). Eggshell powder (ESP), which is generated as a byproduct of food consumption and processing industries, is one such material that has captured attention. Its high calcium carbonate content, fine particle morphology, and natural availability make it a viable candidate for integration into various composite systems, particularly where it may act as a reinforcing or modifying filler (Lian et al., 2014:79; Faridi ve Arabhosseini, 2018:64; Shiferaw et al., 2019:12; Yong et al., 2022:14)

The incorporation of ESP into composite materials has shown potential to influence multiple performance characteristics, including mechanical strength, hardness, thermal stability, and microstructural integrity. These improvements are generally attributed to the stiff, mineral-rich nature of the eggshell structure and its potential to enhance matrix–filler interaction when properly processed. The sustainability benefits of using ESP extend beyond waste reduction: it offers an accessible, low-cost filler that reduces dependence on industrially mined or synthetically processed materials. Such advantages support broader initiatives aimed at reducing carbon footprints and promoting circular material flows (Betancourt ve Cree, 2017:2; Tangboriboon et al., 2016:78; Rosmamuhamadani et al., 2025:38).

While the mechanical and structural implications of ESP addition have been explored in various polymeric and fiber-reinforced systems, far fewer studies have examined its influence on tribological performance. Tribology is of critical importance in applications where surface interactions govern the overall functionality of a component. Wear mechanisms, frictional stability, thermal degradation under dynamic contact, and the formation of friction layers or transfer films are all sensitive to the nature and distribution of fillers within the composite matrix. The role of bio-derived fillers in these processes remains insufficiently clarified, especially under high-demand operating conditions (Montree et al., 2013:770; Sunardi et al., 2023:62; Sunardi et al., 2024:8).

Brake friction composites represent one of the most demanding contexts in which fillers must perform reliably. These materials are subjected to repeated braking cycles that generate substantial heat, pressure, and mechanical stress. The fillers incorporated into brake composites not only contribute to the stabilization of friction coefficients but also influence wear behavior, thermal conduction, and the development of secondary contact plateaus. Conventional fillers—such as metal oxides, carbonaceous materials, or mineral particles—have been optimized for these functions over decades. However, the integration of ESP introduces a new dimension, as its bio-origin and calcium-rich structure may interact differently with the matrix and with the evolving tribo-layer during braking (Nandiyanto et al., 2022:7; Edokpia 2016:4; Adekunle et al., 2023:30).

The potential benefits of using ESP in friction composites extend beyond sustainability. Its intrinsic ability to decompose at elevated temperatures to produce calcium oxide and carbon dioxide may influence the thermal dynamics of the friction interface. The particle size, shape, and distribution of ESP could also affect the formation of stable friction films, potentially moderating wear rates or contributing to more consistent friction behavior. Nevertheless, these effects are highly dependent on processing parameters, filler concentration, and the synergy between ESP and other ingredients within the composite formulation (Shahapurkar et al., 2024:14).

For these reasons, a deeper understanding of the tribological response of ESP-containing brake composites is essential. Evaluating friction coefficient trends, wear morphology, thermal resistance, and microstructural evolution under realistic operating conditions can reveal whether ESP is capable of functioning as an effective bio-filler in friction materials. Such knowledge not only advances the material science of sustainable composites but also informs the design of next-generation braking materials that balance performance, cost efficiency, and environmental stewardship.

In this context, the present study undertakes a systematic investigation into the effects of incorporating bio-derived eggshell powder into brake friction composites. By analyzing tribological behaviors across multiple parameters and examining the underlying mechanisms responsible for performance changes, this work aims to clarify the potential of ESP as a sustainable and functional filler. The findings may provide valuable insights into how biogenic waste materials can be leveraged in high-performance applications traditionally dominated by synthetic or mined fillers, ultimately contributing to a more sustainable materials landscape.

MATERIALS AND METHOD

The materials used in this study included commercially available phenolic resin as the primary binder, along with reinforcing fibers, abrasives, lubricants, and fillers commonly employed in automotive brake friction formulations (Chan ve Stachowiak, 2004:218; Kumar ve Kumaran, 2019:6; Sathyamoorthy et al., 2022:236; Yun et al., 2010:43). Eggshell powder (ESP) was introduced as a bio-derived filler at varying weight percentages to study its influence on the composite's tribological performance. All raw materials were stored at controlled laboratory conditions to avoid moisture absorption and contamination prior to processing.

Eggshell Powder (ESP)

The eggshell powder used in this study was purchased in ready-to-use, industrially processed form from a local supplier. According to the manufacturer, the powder consisted of finely ground chicken eggshells with a high $CaCO_3$ content and a controlled particle size distribution suitable for composite applications. No additional purification, milling, or sieving was performed in the laboratory; however, the powder was inspected visually and stored in airtight containers to prevent moisture absorption and particle agglomeration.

Composite Formulation

Brake friction composites were prepared by incorporating eggshell powder (ESP) as a bio-derived filler, replacing part of the conventional mineral filler (barite) to maintain a consistent total filler content. Four different formulations were produced: 0 % ESP (reference), 20 %, 30 %, and 40 % ESP, with barite content adjusted accordingly for balance.

The complete composite formulation consisted of the following components:

- Phenolic resin – serving as the primary binder
- Cashew particles – providing reinforcement and toughness
- Alumina (Al_2O_3) – contributing hardness and wear resistance
- Steel wool – enhancing thermal conductivity and mechanical strength
- Brass particles – metallic filler improving thermal dissipation and friction characteristics
- Copper particles – further enhancing thermal conduction and friction stability
- Graphite – acting as a solid lubricant
- Barite – conventional mineral filler, partially replaced by ESP

For each formulation, all components were weighed precisely to ensure uniformity. The ESP content was introduced by reducing the barite proportion, keeping the total filler fraction constant across all samples. This design allowed a systematic evaluation of the effect of increasing ESP content on the tribological and mechanical properties of the composites, with the 0 % ESP formulation serving as a reference.

Mixing and Processing

All dry components of the composite, including phenolic resin, cashew particles, alumina, steel wool, brass, copper, graphite, barite, and eggshell powder, were weighed precisely and mixed using a laboratory-scale mechanical mixer for 10 minutes to achieve a homogeneous blend.

The blended mixture was then formed into preforms using a cold press at 8 MPa for 2 minutes to provide initial structural cohesion. Subsequently, the preforms underwent hot pressing at 13 MPa and 130 °C for 10 minutes, ensuring full compaction and polymerization of the phenolic resin. After pressing, the samples were cooled to room temperature, and any excess material was trimmed to achieve standardized dimensions suitable for tribological and mechanical testing.

Surface roughness was controlled by light sanding with fine-grit abrasive paper to ensure consistent contact conditions during testing.

Density and Hardness Measurements

Density of the cured composite samples was determined using the Archimedes displacement method. Each specimen was weighed in air and

while submerged in distilled water to calculate the bulk density, ensuring accurate measurement of material compaction and porosity.

Hardness was measured using a Rockwell test machine, with at least three readings taken from different locations on each sample. The average value and standard deviation were reported to characterize the mechanical resistance of the composites.

All measurements were performed in triplicate to confirm repeatability and reliability of the results.

Thermal Swelling Test

The thermal swelling behavior of the brake friction composites was examined to evaluate their dimensional stability under elevated temperature conditions. Each sample was subjected to a heat exposure procedure using a laboratory-scale sintering furnace set to 400 °C with a dwell time of 10 minutes.

Before thermal treatment, the thickness, diameter, and mass of all specimens were measured using a digital micrometer (±1 μm accuracy), a digital caliper, and an analytical balance with a 0.1 mg resolution. The samples were then placed into the preheated furnace and exposed to the specified temperature for the designated duration to ensure consistent thermal loading.

After heat treatment, the samples were removed from the furnace and allowed to cool naturally to room temperature under ambient laboratory conditions. Subsequently, the same dimensional and mass measurements were repeated using identical instruments. Changes in thickness, diameter, and mass were recorded to assess the extent of thermal swelling or shrinkage, providing insight into the high-temperature stability and structural robustness of the friction composites.

Tribological Testing

Tribological performance of the composite samples was evaluated using a laboratory-scale, custom-designed brake testing rig equipped with a real gray cast iron brake disc. The system is fully computer-controlled, allowing precise monitoring and adjustment of all test parameters throughout the experiments (Sugözü et al., 2022:64; Sugözü et al., 2024:66).

All test conditions were determined in accordance with the TS555 standard for brake friction materials, ensuring that the results are relevant to real-world braking applications. During testing, the following parameters were monitored continuously:

- Coefficient of friction (COF)
- Temperature at the friction interface

After each test, the samples were weighed to calculate wear rate, and the worn surfaces were later examined using scanning electron microscopy

(SEM) to analyze wear mechanisms, material transfer, and the development of secondary contact plateaus.

Each test was conducted in triplicate to ensure reliability and repeatability of the results. The influence of eggshell powder content (0 %, 20 %, 30 %, 40 %) on the frictional stability, wear resistance, and overall tribological performance was systematically assessed.

Morphological Analysis

Worn surfaces were examined using scanning electron microscopy (SEM). Images were used to identify wear mechanisms, crack formation, material transfer, and the development of secondary contact plateaus. Differences in tribo-layer formation among ESP loadings were evaluated.

RESULTS AND DISCUSSION

Before presenting the results, the composite samples used in this study are defined as follows:

- B40 – reference sample containing 0 % eggshell powder (ESP)
- Y20 – composite containing 20 % ESP
- Y30 – composite containing 30 % ESP
- Y40 – composite containing 40 % ESP

These designations will be used consistently throughout the results section to clearly identify the compositions and facilitate comparison of their mechanical and tribological behaviors.

Density

The bulk densities of the composite samples, measured using the Archimedes method, are presented in Table 1. The reference sample, B40, exhibited a density of 1.99 g/cm³. With the addition of eggshell powder, the density increased progressively: Y20 (20 % ESP) reached 2.04 g/cm³, Y30 (30 % ESP) was 2.10 g/cm³, and Y40 (40 % ESP) achieved 2.21 g/cm³.

The observed increase in density with higher ESP content is likely due to the fine particle packing and the relative density of the industrially processed eggshell powder, which may be slightly higher than the barite fraction it replaced. Despite the changes, all samples remain within a suitable density range for brake friction applications, ensuring structural integrity and appropriate mass for thermal and mechanical performance during braking.

Hardness

The Rockwell HRL hardness values of the samples showed a clear increasing trend with higher ESP content. The reference sample B40 exhibited a hardness of 75 HRL, while the incorporation of eggshell powder resulted in progressively higher hardness values: Y20 reached 80 HRL, Y30 increased to 85 HRL, and Y40 achieved the highest hardness at 89 HRL.

This increase in hardness can be associated with the fine particle distribution of eggshell powder within the composite matrix, which may contribute to improved load transfer and a denser packing structure during hot pressing. The improved hardness indicates that ESP contributes positively to the mechanical resistance of the braking material, enhancing resistance to indentation and compressive deformation under operational stresses.

Table 1: Density and Hardness Results of the Composite Samples

Sample	Eggshell Content (%)	Density (g/cm³)	Hardness (HRL)
B40	0	1.99	75
Y20	20	2.04	80
Y30	30	2.10	85
Y40	40	2.21	89

Thermal Swelling

The thermal swelling behavior of the composites was evaluated by measuring the changes after thermal exposure. All samples exhibited an increase in thickness, with the reference sample showing a modest 3.86 % increase. Composites displayed higher expansion, with values ranging from 10.27 % to 28.5 %, indicating that the incorporation of eggshell powder promotes significant dimensional swelling in the thickness direction.

Changes in diameter were relatively small compared to thickness. The reference sample showed a slight increase of 0.98 %, while the composites experienced minor expansions between 1.1 % and 2.32 %, suggesting that thermal swelling predominantly affects the thickness, with diameter remaining largely stable.

All samples experienced weight loss during thermal exposure, reflecting material degradation or moisture release. The reference sample exhibited an 8.47 % decrease in weight, while the composites showed reductions ranging from 7.23 % to 9.06 %. The sample with moderate eggshell content demonstrated the least weight loss, whereas the highest content sample had the greatest decrease, indicating a slight influence of eggshell powder on weight stability.

Overall, the results indicate that the presence of eggshell powder enhances thickness expansion under thermal conditions, while changes in diameter remain limited and weight loss is comparable across samples.

Tribological Testing

The tribological behavior of the developed composites was evaluated to determine the influence of eggshell powder (ESP) content on frictional performance and wear resistance. For this purpose, time-dependent friction coefficient and interface temperature responses were recorded under identical testing conditions for all samples, including the reference material (B40) and the ESP-reinforced composites (Y20, Y30, and Y40). Monitoring these parameters simultaneously provides insight into the stability of the sliding contact, the thermal behavior of the system, and the overall wear mechanisms activated during testing. The obtained results serve as a basis for comparing the tribological enhancements associated with increasing ESP content and for assessing the suitability of the composites for potential engineering applications.

As shown in Figure 1, when examining the coefficient of friction (COF) data of the reference sample B40, it is observed that the COF is quite low and fluctuating during the initial stage of the test. The early values ranging between 0.01 and 0.10 indicate that full contact between the surfaces has not yet been established, and the run-in region between the counterface and the composite is still active. The frequent rises and drops in the COF at this stage are attributed to the breaking of surface asperities, the increase in real contact area, and the fact that the initial tribofilm has not yet stabilized.

Toward the mid-region of the test, the COF settles into the 0.30–0.45 range, indicating that the system becomes more stable. At this point, a relatively continuous contact occurs on the surface, and the material exhibits a more consistent frictional behavior depending on its structural characteristics.

In the final stages of the test, the COF remains within the 0.38–0.42 range for an extended period, with no noticeable increase in wear or signs of tribological instability. Only small, short-term fluctuations are present; these can be associated with localized micro-fractures on the composite surface, third-body layer formation on the counterface, or minor changes in surface morphology.

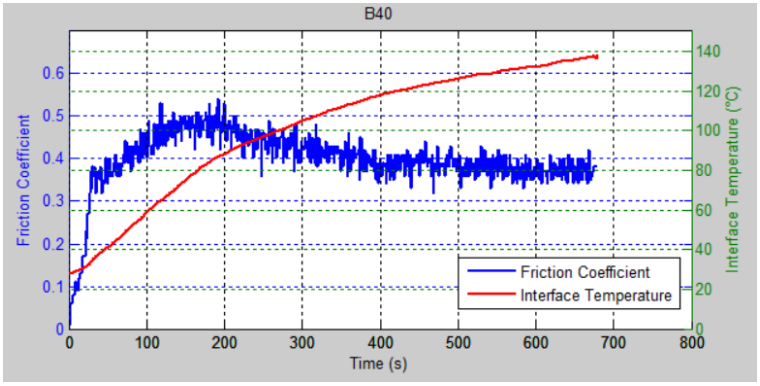


Figure 1: Time-dependent variation of the friction coefficient and interface temperature for the B40 reference sample.

As shown in Figure 2, when examining the COF data of the Y20 composite sample, a short run-in behavior similar to that of the B40 reference sample is observed at the beginning of the test. During the initial seconds, the COF remains in the 0.01–0.10 range, indicating a transition region in which surface asperities are being broken and full stabilization of the contact between the composite and the counterface has not yet been achieved. Following this stage, the COF increases rapidly, reaching approximately 0.30, suggesting the formation of a more stable contact zone on the composite surface.

As the test progresses, the COF maintains a relatively stable trajectory within the 0.36–0.50 range for an extended period. This bandwidth indicates that the Y20 sample exhibits a more balanced sliding behavior compared to B40, with no abrupt spikes or sharp drops in the friction coefficient. Occasional increases up to 0.50–0.54 may be associated with localized third-body formation, micro-fractures, or temperature-induced microstructural changes on the contact surface; however, these rises are short-lived and do not indicate any significant loss of tribological stability.

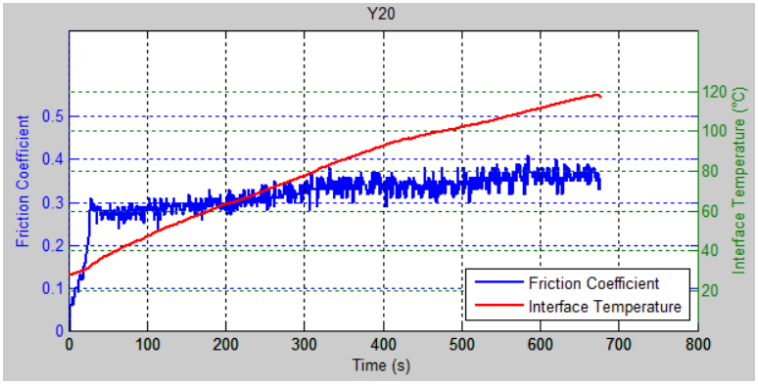


Figure 2: Time-dependent variation of the friction coefficient and interface temperature for the Y20 composite sample.

The Y30 composite exhibits a clear run-in regime during the first tens of recorded time steps, where the friction coefficient rises rapidly from near-zero values to approximately 0.30–0.34. After this initial transient the signal settles into a relatively narrow band around 0.33–0.40 for the majority of the test. The mean friction coefficient for Y30 is 0.357 (median 0.370) with a modest standard deviation of 0.053, indicating limited scatter once the tribological contact has stabilized.

Compared to the B40 and Y20 behaviors (which showed higher stable COF regions in your earlier data), Y30 demonstrates a lower and more stable friction regime, consistent with improved sliding behavior. The reduced average COF and the narrower spread suggest that at 30 % ESP loading the filler particles promote formation of a more uniform, load-bearing tribofilm and/or act as solid lubricating inclusions that reduce adhesive interactions at the interface. Occasional short peaks up to ~0.44 are observed; these may be due to intermittent third-body particles, microfracture events, or local temperature spikes, but they do not dominate the overall behavior.

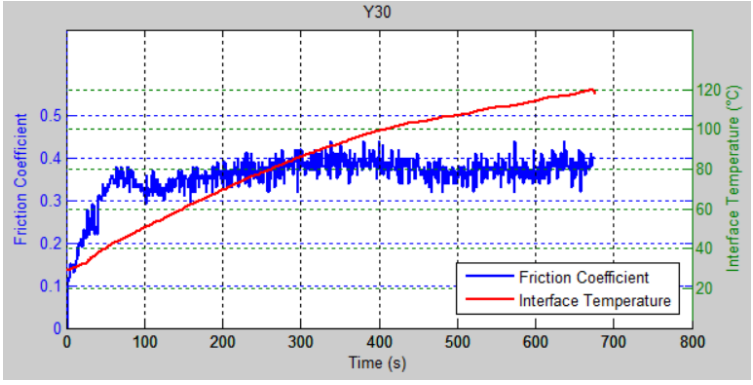


Figure 3: Time-dependent variation of the friction coefficient and interface temperature for the Y30 composite sample.

The Y40 composite shows a pronounced run-in stage during the initial time steps, where the friction coefficient rises from near-zero values to approximately 0.30 within the first tens of measurements. After this transient behaviour the signal stabilizes and remains mostly in the 0.33–0.40 range for the remainder of the test. The mean friction coefficient for Y40 is 0.357 (median 0.370) with a standard deviation of 0.053, indicating a relatively narrow scatter once the contact has equilibrated. Occasional short peaks up to ~0.44 were observed but did not significantly affect the overall stability. These findings suggest that at 40 wt.% ESP loading the composite develops a consistent, load-bearing tribofilm and/or benefits from solid-lubricant effects of the filler, yielding a lower and more stable friction response compared to the reference behavior (B40) and the lower-loading composites described previously.

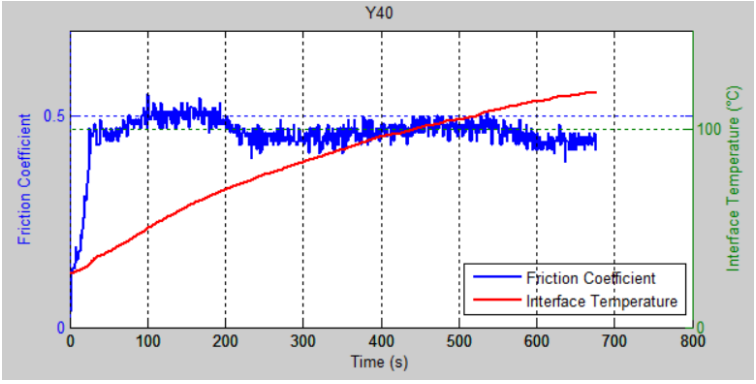


Figure 4: Time-dependent variation of the friction coefficient and interface temperature for the Y40 composite sample.

The tribological properties presented in Table 2 reveal clear differences in the wear and friction behavior of the composite samples as the eggshell powder (ESP) content is varied. The reference sample B40 exhibits a specific wear rate of 3.30×10^{-6} cm^3/Nm, which is higher than that of Y20 and Y30, indicating that the addition of a moderate amount of eggshell powder contributes to improved wear resistance. The Y20 sample demonstrates the lowest wear rate (2.62×10^{-6} cm^3/Nm), suggesting that a 20% ESP content provides an optimal reinforcement effect, likely due to better particle distribution and an enhanced ability to form a protective tribofilm during sliding.

The average friction coefficient values show an increasing trend with higher ESP content. While Y20 exhibits the lowest friction coefficient (0.31), the Y40 sample reaches the highest value (0.46). This indicates that excessive ESP content may increase abrasive interactions or alter the contact mechanics at the sliding interface, leading to higher friction. In contrast,

moderate ESP additions (20–30%) appear to reduce friction relative to the reference sample.

Friction stability improves consistently with increasing ESP content, rising from 73.6% (B40) to 82.7% (Y40). This enhancement suggests that eggshell powder contributes to more consistent frictional behavior over time, likely by promoting a more stable third-body layer and reducing sudden fluctuations in the contact conditions. The Y30 and Y40 samples, in particular, show notably high stability values (81.1% and 82.7%), indicating that higher ESP contents, despite increasing average friction, enhance the overall smoothness and predictability of the sliding process.

Table 2: Tribological Properties of the Composite Samples

Sample	Specific wear rate (cm^3/Nm)	Average friction coefficient	Friction stability (%)
B40	$3.30x10^{-6}$	0.40	73.6
Y20	$2.62x10^{-6}$	0.31	77.3
Y30	$2.80x10^{-6}$	0.36	81.1
Y40	$3.62x10^{-6}$	0.46	82.7

Overall, the results indicate that Y20 provides the best balance between low wear rate and reduced friction coefficient, whereas Y40 excels in friction stability but suffers from higher wear and friction levels. These findings highlight the importance of optimizing ESP content to achieve desired tribological performance characteristics.

SEM Analysis

The SEM micrographs of the worn composite surfaces, presented in Figure 5, were examined to identify the dominant wear mechanisms and to assess the influence of eggshell powder (ESP) content on surface morphology after tribological testing. The reference sample (B40) exhibited a relatively rough and irregular worn surface characterized by abrasive grooves, micro-fractures, and evidence of brittle detachment. These features indicate that the matrix–filler interaction was insufficiently stabilized during sliding, resulting in more pronounced abrasive and fatigue-related wear.

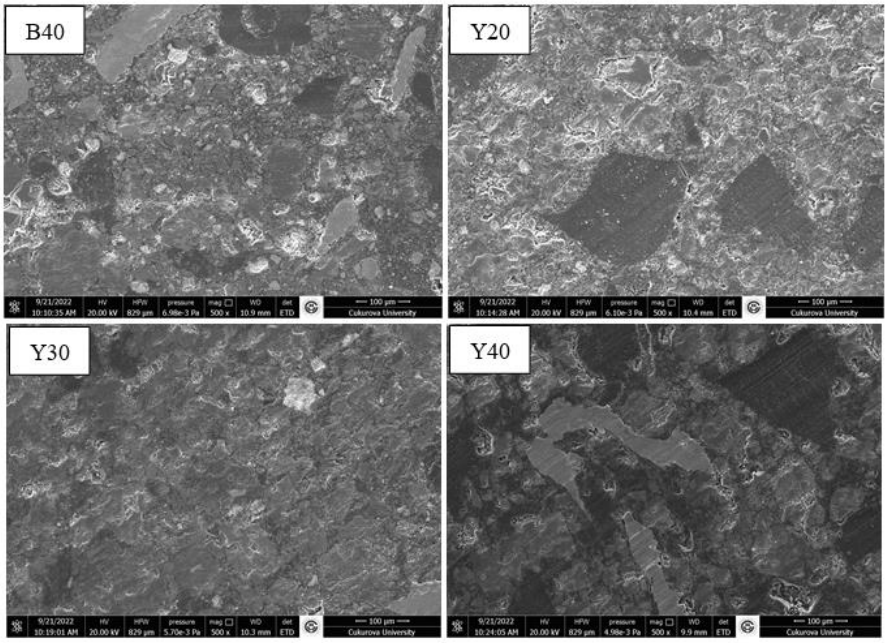


Figure 4: SEM micrographs of worn composite surfaces showing the effect of ESP content on wear morphology.

The Y20 sample displayed a noticeably smoother worn surface, with fewer deep grooves and more compact tribofilm patches. The presence of smeared regions and densely compacted debris suggests the development of a stable third-body layer, which aligns with the reduced wear rate and lower friction coefficient observed for this composition.

The Y30 composite showed an even more uniform morphology, featuring continuous tribofilm coverage and limited micro-cracking. The reduced occurrence of abrasive scratches indicates that this ESP level supports improved lubrication and efficient compaction of wear debris. These observations are consistent with the stable friction behavior and moderate wear performance recorded for Y30.

In contrast, the Y40 sample exhibited heterogeneous wear characteristics. While certain regions showed dense tribofilm formation, others revealed localized micro-fragmentation and debris accumulation. This inconsistency suggests that excessive ESP content may lead to particle agglomeration, reducing interfacial cohesion and increasing abrasive activity. These morphological features correspond with the higher friction coefficient and specific wear rate of the Y40 composite.

Overall, the SEM observations confirm that moderate ESP additions (20–30 %) promote the formation of a stable load-bearing tribofilm, leading to smoother worn surfaces and improved tribological behavior, whereas

excessive ESP content results in less uniform contact conditions and diminished performance.

CONCLUSIONS

This study evaluated the effects of bio-derived eggshell powder (ESP) on the mechanical, thermal, and tribological performance of brake friction composites. Four formulations containing 0–40 % ESP were investigated, and the findings show that ESP has a significant and composition-dependent influence on the resulting material behavior.

Density and hardness increased progressively with higher ESP content, indicating improved particle packing and enhanced matrix reinforcement. These trends suggest that ESP contributes to a more compact and mechanically robust composite structure. Thermal swelling tests revealed that ESP promotes considerable expansion in the thickness direction while causing only minimal changes in diameter. Weight loss remained similar across all samples, with moderate ESP additions showing slightly improved thermal stability compared to the highest loading.

Tribological tests demonstrated that all formulations underwent an initial run-in stage before stabilizing. The incorporation of ESP noticeably affected friction and wear behavior. The Y20 sample achieved the lowest specific wear rate and reduced friction coefficient, highlighting 20 % ESP as the most effective level for improving wear resistance and frictional performance. Higher contents, particularly Y40, resulted in increased friction and wear, but also provided the highest friction stability, indicating improved consistency of the sliding interface.

SEM analysis further supported these observations: moderate ESP levels (20–30 %) produced smoother worn surfaces with stable, continuous tribofilms, while excessive loading (40 %) generated more heterogeneous morphologies and localized micro-fragmentation.

Overall, the results demonstrate that eggshell powder is a promising sustainable filler for friction materials, capable of enhancing mechanical strength, stabilizing frictional behavior, and improving wear performance when used in optimized amounts. The most balanced and advantageous performance was achieved at 20 % ESP, whereas higher levels increased friction and surface heterogeneity. These findings highlight the potential of ESP to serve as an environmentally friendly alternative filler in future brake composite formulations.

REFERENCES

Adekunle, A., Okunlola, M., Omoniyi, P., Adeleke, A., Ikubanni, P., Popoola, T., ve Ibrahim, K. (2023). Development and analysis of friction material for eco-friendly brake pad using seashell composite. *Scientia Iranica*, 30(5), 1562-1571.

Betancourt, N. and Cree, D. (2017). Mechanical properties of poly (lactic acid) composites reinforced with caco3 eggshell based fillers. *Mrs Advances*, 2(47), 2545-2550.

Chan, D. S. E. A., ve Stachowiak, G. W. (2004). Review of automotive brake friction materials. *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, *218*(9), 953-966.

Edokpia, R. O., Aigbodion, V. S., Atuanya, C. U., Agunsoye, J. O., ve Mu'azu, K. (2016). Experimental study of the properties of brake pad using egg shell particles–Gum Arabic composites. *Journal of the Chinese Advanced Materials Society*, 4(2), 172-184.

Faridi, H. and Arabhosseini, A. (2018). Application of eggshell wastes as valuable and utilizable products: a review. *Research in Agricultural Engineering*, 64(2), 104-114.

Jusoh, M. (2022). Green engineering education in environmental engineering programme through active learning. *Acta Pedagogia Asiana*, 1(1), 18-25.

Kumar, V. V., ve Kumaran, S. S. (2019). Friction material composite: types of brake friction material formulations and effects of various ingredients on brake performance–a review. *Materials Research Express*, *6*(8), 082005.

Lian, Z., Qiao, L., Zhu, G., Deng, Y., Qian, B., Yue, J., Zhao, Y. (2014). Use of sodium dodecyl sulfate pretreatment and 2-stage curing for improved quality of salted duck eggs. *Journal of Food Science*, 79(3).

Montree, A., Teanchai, K., ve Siriprom, W. (2013). Preparation and characterization the cheap mineral (calcium compound): the preliminary survey in the natural material. *Advanced Materials Research*, 770, 303-306.

Nandiyanto, A. B. D., Ragadhita, R., Fiandini, M., Al Husaeni, D. F., Al Husaeni, D. N., ve Fadhillah, F. (2022). Domestic waste (eggshells and banana peels particles) as sustainable and renewable resources for improving resin-based brakepad performance: Bibliometric literature review, techno-economic analysis, dual-sized reinforcing experiments, to comparison... *Communications in Science and Technology*, 7(1), 50-61.

Rosmamuhamadani, R., Razak, N., Mahat, M., Osman, S., Zakaria, M., ve Yahaya, S. (2025). A review: characterization of bio-cement composites reinforced with calcium carbonate obtained from chicken egg-shells. *Diffusion Foundations and Materials Applications*, 38, 59-65.

Sathyamoorthy, G., Vijay, R., ve Lenin Singaravelu, D. (2022). Brake friction composite materials: a review on classifications and influences of friction materials in braking performance with characterizations. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, *236*(8), 1674-1706.

Shahapurkar, K., Ramesh, S., Nik-Ghazali, N. N., Chenrayan, V., MC, K., Kanaginahal, G., ... ve Kalam, M. A. (2024). Influence of Habesha eggshell on the dry sliding wear behavior of epoxy composites. *Biomass Conversion and Biorefinery*, 14(17), 21523-21538.

Shiferaw, N., Habte, L., Thenepalli, T., ve Ahn, J. (2019). Effect of eggshell powder on the hydration of cement paste. *Materials*, 12(15), 2483.

Sugözü, B., Dağhan, B., Akdemir, A. ve Sugözü, İ. (2022). Effect of micro and nano-sized ZrSiO4 particles on the friction and wear properties of polymer matrix composites. *Materials Testing*, *64*(9), 1290-1297.

Sugözü, B., Erol, E. ve Sugözü, İ. (2024). Tribological and thermal characteristics of copper-free brake friction composites. *Materials Testing*, *66*(2), 226-232.

Sunardi, S., Ariawan, D., Surojo, E., Prabowo, A. R., Akbar, H. I., Sudrajad, A., ve Seputro, H. (2023). Optimization of eggshell particles to produce eco-friendly green fillers with bamboo reinforcement in organic friction materials. *Reviews on Advanced Materials Science,* 62(1), 20230111.

Sunardi, S., Ariawan, D., Surojo, E., Prabowo, A. R., Ghanbari-Ghazijahani, T., Wibowo, C. H., ve Akbar, H. I. (2024). Tribological performance of polymer composite modified with calcined eggshell particles post high-temperature exposure. *Emerging Science Journal*, 8(4), 1280-1292.

Tangboriboon, N., Moonsri, S., Netthip, A., ve Sirivat, A. (2016). Innovation of embedding eggshell to enhance physical-mechanical-thermal properties in fired clay bricks via extrusion process. *Matec Web of Conferences*, 78, 01003.

Yong, T., Lim, S., ve Ling, L. (2022). Dimensional stability of lightweight foamed concrete containing eggshell powder and calcium stearate. *International Journal of Integrated Engineering*, 14(1).

Yun, R., Filip, P., ve Lu, Y. (2010). Performance and evaluation of eco-friendly brake friction materials. *Tribology International*, *43*(11).

Deep Learning–Based Vision Perception and Video Analytics

Emin GÜNEY[1]

1- Res. Asst. Dr.; Sakarya University of Applied Sciences, Faculty of Technology, Department of Computer Engineering. eminguney@subu.edu.tr ORCID No: 0000-0003-0098-9018

ABSTRACT

Deep learning has fundamentally reshaped visual perception and video analytics by eliminating the need for handcrafted feature extraction and enabling end-to-end learning directly from image and video data. Vision-based perception focuses on understanding what exists in a scene; identifying objects, inferring scene layout, estimating depth, and extracting semantic cues from a single frame. In contrast, video analytics addresses how a scene evolves over time, capturing motion, behaviors, interactions, and temporal changes using models such as 3D-CNNs, LSTMs, and spatio-temporal attention networks. Recent studies show that perception systems increasingly run on edge and edge–cloud architectures, where processing is performed close to the camera to reduce latency and bandwidth usage, enabling real-time operation in dynamic environments. Beyond accuracy, current research prioritizes robustness and interpretability, including adversarial resilience and visual analytics for human-centered decision support. Overall, deep learning advances have transformed visual perception into a real-time and context-aware component of intelligent systems.

Keywords – Vision-Based Perception, Video Analytics, Deep Learning, Spatio-Temporal Modeling, Real-Time Perception, Edge Computing, Robust Perception.

INTRODUCTION

Image and video data have become the fundamental components of modern intelligent systems' ability to perceive their environment. Camera-based perception systems are now the richest information source for decision-making processes in domains such as security, transportation, robotics, healthcare, and smart cities. For many years, traditional machine learning approaches relied on handcrafted feature extraction; however, these methods have proven insufficient for real-time analysis in complex and dynamic scenes. The emergence of deep learning removed this limitation by enabling end-to-end learning pipelines that perform feature extraction and classification simultaneously. According to Abbas et al., deep learning significantly improves the performance of tasks such as real-time video scene analysis, object recognition, activity detection, and scene interpretation compared to traditional approaches [1]. One of the main reasons for this progress is the capability of deep learning architectures—such as CNN, RNN, LSTM, and 3D-CNN—to model the spatial–temporal relationships within sequences of video frames. Sharma et al. report that the most active research topics in video analytics include human action recognition, anomaly detection, and behavior analysis, highlighting that deep learning has become the dominant approach for video understanding in recent years [2]. The impact of deep learning extends beyond video and image understanding; it is

now employed across a wide range of industries. Ganesh et al. emphasize the widespread adoption of deep learning–based computer vision applications in agriculture, healthcare, manufacturing, sports, and transportation, attributing this growth to advances in GPU computing and algorithmic efficiency [3]. In intelligent transportation systems and autonomous driving, accurate visual perception is crucial for safe navigation. Azfar et al. show that traffic perception systems face significant real-world challenges - irregular illumination, occlusions, heterogeneous traffic, and communication delays - which require more robust deep learning architectures [4].

The rapid adoption of deep learning has also reshaped the processing architecture of perception systems. Streaming high-resolution video to the cloud is problematic due to bandwidth limitations and network latency. Consequently, computation is increasingly shifted toward edge devices or hybrid edge–cloud architectures. Hu et al. describe this shift as edge-based video analytics, highlighting that moving inference closer to the data source alleviates latency and bandwidth issues [5]. Similarly, Wang et al. demonstrate that edge–cloud video analytics systems can optimize performance using adaptive compression and super-resolution to improve "tail accuracy" for difficult perception cases [6]. In the context of public safety, Zhang et al. show that performing analytics at the edge significantly reduces transmission overhead and enables real-time detection on mobile or body-worn cameras [7].

Recent studies also highlight the importance of visualization and interpretability in video analytics. Afzal et al. argue that visual analytics systems serve as a critical bridge between deep learning models and decision-makers, enabling interactive exploration and interpretation of video-based perception results [8]. In summary, recent literature reveals that deep learning–based visual perception and video analytics evolve around three main research axes: (i) Deep learning models for visual perception, covering video scene analysis, object/activity recognition, and dense crowd understanding; (ii) Edge and edge–cloud processing architectures, enabling low-latency and scalable video analytics; and (iii) Application diversity and societal impact, across transportation, healthcare, smart cities, and public safety. This section synthesizes the findings of recent studies and evaluates their contributions in terms of visual perception architectures, challenges, and emerging research trends (Table 1).

Table 1. Overview of Vision-Based Perception and Video Analytics works

Paper	Domain/ Use-case	Core Task	Method highlights	Key contribution	Limits/Gaps
Abbas et al., 2017	General video scene analysis	Action/activity, scene interpretation	Deep learning vs. handcrafted features; CNN/RNN family overview	Early map of DL for scene analysis; taxonomy of tasks	Pre-edge era; limited deployment concerns
Sharma et al., 2021	General video processing	HAR, anomaly, behavior	CNN/DNN/RNN categorization;	SLR that consolidates apps/techniques /datasets	Limited edge/real-time constraints discussion
Ganesh et al., 2023/2024	Cross-sector CV	Detection/cls./tracking	Sector-wise DL review (Agriculture, Health, Mfg., Sports, Transport)	Bridges DL methods to sector problems	Less on latency/bandwidth trade-offs
Azfar et al., 2024	ITS & AD	Traffic perception	Reviews DL in complex urban scenes; data/model/	Clear taxonomy of real-world challenges	Few concrete real-world deployments
Bhuiyan et al., 2022	Crowd analysis (Hajj)	Counting, density, abnormality	DL for dense crowds; FCNN trend	Highlights extreme-density constraints	Domain-specific; generalization unclear
Zhang et al., 2019	Public safety	Real-time edge VA	Edge analytics stack; offloading strategies	Shows latency/bandwidth wins at edge	Fixed-algorithm cams; dynamic offload is hard
Hu et al., 2023	Edge-based VA	Apps/arch/management	Survey of edge VA: architecture, techniques, resource mgmt., privacy	Comprehensive edge VA landscape	Security/privacy + heterogeneous edge remain open
Wang et al., 2023 (Runespoor)	Robotics edge-cloud	Tail accuracy under bandwidth	Adaptive bitrate + task-aware SR on server	Large tail-accuracy gains at same bandwidth	Added complexity; content-aware control
Afaq & Akram, 2025	Smart cities & SDGs	Societal impact with edge+DL	Sustainability lens; healthcare/UAV/traffic cases	Connects edge+DL to SDG targets	Technical depth (modeling) limited
Afzal et al., 2023	Visual analytics	Visualization of image/video	Taxonomy of vis/VA tools, tasks, datasets	Pipeline-level vis/interpretability guidance	Less on tight real-time with DL

Vision Perception (Camera-Based Scene Understanding)

Vision perception refers to the ability of a system to interpret visual information from cameras and convert raw images into semantic representations such as objects, geometry, depth, and environmental context. In robotics and autonomous systems, cameras are widely used because they

are inexpensive, lightweight, and capable of providing dense semantic information (e.g., object type, spatial arrangement, color, depth cues) compared to range-only sensors. Visual perception enables robots to recognize objects, localize themselves within an environment, and plan motions based on scene interpretation. Stefanov et al. demonstrated this advantage by using only a camera to perform mobile manipulation and execute a precise screw-assembly task, showing that visual features can be used for pose estimation and error correction without requiring additional sensors [9].

Vision perception is also central in autonomous driving, where multiple perception tasks are required simultaneously — such as object detection, drivable-area segmentation, and lane detection. Instead of training separate networks for each perception task, multi-task learning architectures can process multiple outputs from a single shared backbone, reducing computational load and improving real-time performance [10]. This integration enhances situational awareness and increases the robustness of visual perception under complex traffic scenes. Beyond machine-centric perception, vision systems are increasingly used to model human visual perception of urban environments. Cao et al. introduced a biological-vision-inspired perception framework that uses large vision models (LVMs) to extract color, semantic segmentation, object distribution, and depth information from street-view imagery, linking visual scene composition to subjective human perception [11]. Their model shows that visual elements such as color, spatial form, and landscape components significantly affect how people perceive and evaluate urban spaces. Vision-based perception is not limited to analyzing images of external environments; it can model tactile sensing as well. Zhang et al. introduced a vision-based tactile perception method that analyzes sequential camera images from a tactile sensor to estimate the hardness of objects without prior knowledge of the object's shape [12]. The method uses a dual spatio-temporal attention network to capture local deformation patterns, demonstrating that vision can be used to infer physical properties traditionally obtained by force sensors.

Together, these studies illustrate that camera-based vision perception is no longer limited to static scene interpretation — it enables semantic understanding, environmental perception, and even cross-modal reasoning (visual-tactile perception) across diverse robotic and urban applications.

Video Analytics (Dynamic Perception& Temporal Understanding)

While vision perception focuses mainly on interpreting what is present in a single image, video analytics analyzes how the scene evolves over time. It processes sequences of frames to extract temporal patterns such as motion, change, interaction, deformation, or human activity. This makes video analytics essential for domains where perception depends on temporal reasoning. In autonomous driving, video analytics enables continuous

monitoring of traffic participants and dynamic road conditions. Instead of treating perception tasks independently, multi-task perception architectures process video streams and simultaneously detect objects, segment drivable areas, and estimate lane boundaries [10]. This reduces inference latency and allows vehicles to understand both semantic and geometric properties of traffic environments in real time.

Temporal analysis is also central to vision-based tactile sensing. Zhang et al. showed that the hardness of an object can be inferred by analyzing the evolution of contour deformation across multiple video frames captured during contact [12]. Their dual-attention network leverages temporal features to capture deformation dynamics, achieving shape-independent hardness estimation — something static image inputs could not provide. Video analytics is equally relevant in modeling human perception of urban scenes. Cao et al. used visual scene interpretation techniques — such as depth prediction, object detection, and semantic segmentation — to understand how visual complexity and spatial forms influence how people perceive urban environments [11]. In this context, video analytics provides actionable insight for human-centric urban planning, guiding decisions related to public satisfaction, walkability, and environmental quality.

Across these studies, video analytics transforms raw visual input into dynamic perceptual awareness. In contrast to static perception (snapshot understanding), video analytics enables (Table 2).

Table 2. Key Differences Between Vision Perception and Video Analytics

Vision Perception	Video Analytics
Answers: What is in the scene?	Answers: What is happening over time?
Single frame inference	Multi-frame temporal inference
Semantic interpretation	Motion, behavior, deformation, and event understanding
Object recognition / segmentation	Tracking, prediction, dynamic scene reasoning

Deep Learning for Vision-Based Perception and Video Analytics

Vision-based perception has evolved from classical handcrafted image processing toward fully automated deep learning architectures, enabling real-time visual understanding in robotic and autonomous systems. Modern perception pipelines leverage camera-based sensing to extract high-level semantic information such as road geometry, object classes, obstacle boundaries, and navigation cues. Early implementations showed that vision alone can guide mobile robots in challenging outdoor environments. For example, agricultural robots used grayscale and color-based camera

perception to recognize crop rows and identify weeds for active robotic manipulation, achieving centimeter-level navigation accuracy [16]. Similarly, automated harvesters applied computer vision to track cut/uncut crop boundaries and detect obstacles without GPS, performing guidance solely through visual cues [17]. These early approaches establish the foundation of real-time visual perception.

With the emergence of deep learning architectures, perception shifted from handcrafted features toward automatic hierarchical feature extraction. In autonomous driving, deep convolutional neural networks (CNNs) infer semantic and geometric understanding of scenes from camera imagery [18]. To reduce computational load, perception is increasingly performed through multi-task perception backbones, enabling object detection, lane segmentation, and drivable-area estimation within a single architecture rather than running multiple models independently. This aligns with real-time requirements and lowers processing latency—critical for edge AI execution on embedded systems.

Table 3. Key Concepts in Vision Perception and Video Analytics

Concept	What it represents in Vision-Based Perception & Video Analytics	Deep learning contribution
Vision-Based Perception	Extracting semantic meaning from visual data (objects, scene layout, depth cues, environment context)	CNN / Transformers learn hierarchical visual features automatically
Video Analytics	Understanding changes over time: motion, behavior, temporal patterns	Spatio-temporal learning (ConvLSTM, 3D-CNN, attention)
Real-Time Visual Perception	Performing perception under latency/compute constraints (embedded devices, mobile systems)	Efficient lightweight models, multi-task sharing
Robust Perception	Ensuring perception remains reliable under uncertainty, adversarial risk, visual noise	Defensive learning, adversarial resilience, verification
Human-Centered Perception	Perception designed to support human decision-making (interpretability, user feedback)	Explainable output, perceptual feedback, user interaction

Furthermore, recent advances emphasize perception robustness rather than accuracy alone. Some works propose formal perception quality metrics using Timed Quality Temporal Logic (TQTL), allowing a quantitative evaluation of temporal object consistency across video frames [19]. Other research develops approximate abstractions of vision-based perception models to enable safety verification of control systems (e.g., lane-keeping controllers and agricultural robots), moving the field toward explainability and trustworthiness [20]. Complementarily, adversarial studies show that perception models can be attacked through small physical manipulations (painting black lines), causing autonomous vehicles to produce wrong control outputs, emphasizing the importance of robust and resilient perception.

In addition to autonomy in vehicles and robotics, vision-based perception extends into human-centered applications. Vision-based navigation assistants for visually impaired individuals use real-time object detection, obstacle recognition, and landmark identification on smartphones, demonstrating low-cost real-time visual perception for mobility and safety assistance . This highlights how perception is no longer only machine-oriented, but increasingly human-centered, creating perception systems that interact with users and complement human decision making.

From an architectural perspective, video analytics introduces the temporal dimension into deep learning (Table 3). Unlike perception from a single frame ("what is present?"), video analytics answers "what is happening over time?". Spatio-temporal modeling exploits 3D-CNNs, ConvLSTMs, and attention mechanisms to detect motion patterns, dynamic objects, or time-dependent behaviors, enabling perception of sequential events instead of isolated frames. Deep spatio-temporal models are also critical for vision-based tactile sensing, where deformation images across time are used to infer the physical properties of objects [23].

CONCLUSION AND DISCUSSION

This study shows that deep learning has fundamentally improved vision-based perception and video analytics. Vision perception focuses on understanding what is in the scene, while video analytics interprets how the scene changes over time. Deep learning enables both tasks to be performed in real time by learning spatial and temporal features directly from camera data. Recent work highlights three major trends: (1) real-time processing through edge and edge–cloud systems, (2) increased focus on robustness and reliability, and (3) a shift toward human-centered and interpretable perception outputs. Although current approaches achieve high accuracy, challenges remain such as handling limited computing resources, ensuring temporal consistency across frames, and improving interpretability for decision-makers. Overall, deep learning continues to transform visual perception into a reliable and efficient component of intelligent systems.

REFERENCES

[1] Abbas, Q., Ibrahim, M. E., & Jaffar, M. A. (2018). *Video scene analysis: an overview and challenges on deep learning algorithms*. Multimedia Tools and Applications, 77(16), 20415–20453.

[2] Sharma, V., Gupta, M., Kumar, A., & Mishra, D. (2021). *Video processing using deep learning techniques: A systematic literature review*. IEEE Access, 9, 139489–139507.

[3] Ganesh, N., Shankar, R., Mahdal, M., Murugan, J., Chohan, J., & Kalita, K. (2024). *Exploring deep learning methods for computer vision applications across multiple sectors: Challenges and future trends*. Computer Modeling in Engineering & Sciences, 139(1), 103.

[4] Azfar, T., Li, J., Yu, H., Cheu, R. L., Lv, Y., & Ke, R. (2024). *Deep learning-based computer vision methods for complex traffic environments perception: A review*. Data Science for Transportation, 6(1), 1.

[5] Hu, M., Luo, Z., Pasdar, A., Lee, Y. C., Zhou, Y., & Wu, D. (2023). *Edge-based video analytics: A survey*. arXiv:2303.14329.

[6] Wang, Y., Wang, W., Liu, D., Jin, X., Jiang, J., & Chen, K. (2022). *Enabling edge-cloud video analytics for robotics applications*. IEEE Transactions on Cloud Computing, 11(2), 1500–1513.

[7] Zhang, Q., Sun, H., Wu, X., & Zhong, H. (2019). *Edge video analytics for public safety: A review*. Proceedings of the IEEE, 107(8), 1675–1696.

[8] Afzal, S., Ghani, S., Hittawe, M. M., Rashid, S. F., Knio, O. M., Hadwiger, M., & Hoteit, I. (2023). *Visualization and visual analytics approaches for image and video datasets: A survey*. ACM Transactions on Interactive Intelligent Systems, 13(1), 1–41.

[9] Stefanov, A., Zorman, M., Šlajpah, S., Podobnik, J., Mihelj, M., & Munih, M. (2026). *From perception to precision: Vision-based mobile robotic manipulation for assembly screwdriving*. Robotics and Computer-Integrated Manufacturing, 98, 103148.

[10] Wang, H., Li, J., & Dong, H. (2025). *A review of vision-based multi-task perception research methods for autonomous vehicles*. Sensors, 25(8), 2611.

[11] Cao, Y., Yang, P., Xu, M., Li, M., Li, Y., & Guo, R. (2025). *A novel method of urban landscape perception based on biological vision process*. Landscape and Urban Planning, 254, 105246.

[12] Zhang, Y., Xu, C., Li, B., Liu, S., & Song, Y. (2025). *Spatio-temporal dual-attention network enabled shape-independent hardness estimation with a vision-based tactile sensor*. Measurement, 118486.

[13] Åstrand, B., & Baerveldt, A. J. (2002). *An agricultural mobile robot with vision-based perception for mechanical weed control*. Autonomous Robots, 13(1), 21–35.

[14] Ollis, M., & Stentz, A. (1997). *Vision-based perception for an automated harvester*. In Proceedings of IROS'97 (pp. 1838–1844). IEEE.

[15] Ros, G., Ramos, S., Granados, M., Bakhtiary, A., Vazquez, D., & Lopez, A. M. (2015). *Vision-based offline-online perception paradigm for autonomous driving*. In IEEE Winter Conference on Applications of Computer Vision (pp. 231–238).

[16] Balakrishnan, A., Puranic, A. G., Qin, X., Dokhanchi, A., Deshmukh, J. V., Amor, H. B., & Fainekos, G. (2019). *Specifying and evaluating quality metrics for vision-based perception systems*. In DATE 2019 (pp. 1433–1438). IEEE.
[17] Hsieh, C., Li, Y., Sun, D., Joshi, K., Misailovic, S., & Mitra, S. (2022). *Verifying controllers with vision-based perception using safe approximate abstractions*. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems, 41(11), 4205–4216.
[18] Boloor, A., Garimella, K., He, X., Gill, C., Vorobeychik, Y., & Zhang, X. (2020). *Attacking vision-based perception in end-to-end autonomous driving models*. Journal of Systems Architecture, 110, 101766.
[19] Tapu, R., Mocanu, B., & Zaharia, T. (2017). *A computer vision-based perception system for visually impaired*. Multimedia Tools and Applications, 76(9), 11771–11807.

Deep Learning Architectures for Context Modeling: A Comprehensive Analysis of Recurrence, Gating and Attention Mechanism-Based Approaches

Hilal ÇELİK[1]
Ramazan KATIRCI[2]

1- Res. Asst.. Gör.; Sivas Bilim ve Teknoloji Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisliği Bölümü. hilalcelik@sivas.edu.tr ORCID No: 0000-0001-5428-3411
2- Prof. Dr.; Sivas Bilim ve Teknoloji Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisliği Bölümü. ramazankatirci@sivas.edu.tr ORCID No: 0000-0003-2448-011X

ABSTRACT

Neural architectures have progressively transformed deep learning, enabling machines to model complex semantic, syntactic, and contextual patterns in sequential and complex data. Early ANNs provided the foundational framework for non-linear representation learning, yet their limitations in handling sequential dependencies motivated the development of more specialized models. Recurrent architectures such as RNNs, LSTMs, and GRUs introduced robust mechanisms for capturing temporal relationships and managing information flow through gating. Meanwhile, CNNs contributed efficient feature extraction capabilities through hierarchical representations. The emergence of the Transformer architecture signalled a significant paradigm shift, replacing recurrence with attention mechanisms and enabling massive parallel computation, while also substantially improving the modeling of long-range dependencies across data sequences. This chapter focuses on the development of these architectures, emphasising their theoretical basis, operational capabilities and contributions to modern deep learning models. Together, these models demonstrate the creation of increasingly scalable, context-aware and high-performance computational frameworks that form the basis of contemporary data processing in a variety of fields.

Keywords:

INTRODUCTION

Deep Learning (DL) has progressed from early artificial neural networks (ANNs)(Mcculloch & Pitts, 1990). Foundational ANN structures provided the basis for non-linear representation learning, yet their difficulty in modeling sequential dependencies stimulated the development of recurrent architectures. Recurrent Neural Networks (RNNs) (Rumelhart et al., 1986) introduced the ability to process variable-length sequences, while Long Short-Term Memory (LSTM) networks (Hochreiter, 2016) addressed vanishing gradient limitations through gated memory mechanisms. Gated Recurrent Units (GRUs) (Cho et al., 2014) further simplified gating operations, offering efficient temporal dependency modeling with reduced computational cost. Subsequent advancements included Convolutional Neural Networks (CNNs) (Young et al., 2018), which enabled hierarchical extraction of local linguistic features. The emergence of the Transformer architecture, powered by self-attention mechanisms, eliminated recurrence and enabled parallel processing, significantly enhancing long-range dependency modeling and overall performance. Collectively, these architectures trace the evolution of neural models that have strengthened representational capacity, learning efficiency and scalability in modern DL.

EVOLUTION OF SEQUENTIAL MODELING ARCHITECTURES

Deep Learning (DL) architectures represent computational models inspired by the interconnected structure of the human brain, designed to process information through layers of interconnected units known as neurons (Buscema et al., 2018). These architectures form the foundation of modern artificial intelligence by enabling the modeling of complex, non-linear patterns within data. In DL, they allow machines to comprehend, represent, and generate complex sequential outputs by learning contextual and semantic relationships directly from large-scale data corpora(Torfi et al., 2021; Aasim et al., 2022; Perumal et al., 2024). Such capabilities have made neural architectures fundamental to a wide range of NLP applications, including machine translation (Katırcı & Çelik, 2025d, 2025a) that require accurate cross-lingual mapping and question-answering models (Çelik et al., 2024; Katırcı & Çelik, 2025c) designed to extract or generate precise information from textual input.

Artificial Neural Networks (ANNs) constitute the earliest and most fundamental class of such models, demonstrating strong performance across various domains. However, despite their foundational role, traditional ANNs struggle to capture the sequential and context-dependent nature of data, often failing to model long-range dependencies and suffering from issues such as overfitting and unstable convergence during training (Alam et al., 2023). These limitations motivated the development of more specialized neural architectures tailored to sequential data, including RNNs, LSTMs, GRU, CNNs and Transformer-based models, which have significantly advanced DL applications by offering improved representational capacity and learning efficiency.

Artificial neural networks

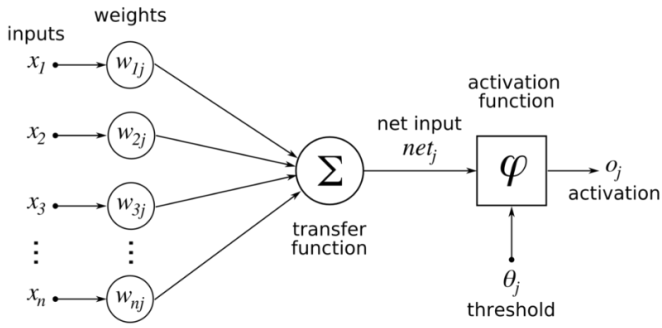


Figure 1. Artificial neural networks (ANNs) (Wikimedia, n.d.)

ANNs are computational architectures inspired by the information-processing mechanisms of biological neural systems. In these models, learning emerges

through interconnected processing units—neurons—that adjust their connection weights to map inputs to outputs. The conceptual foundation of artificial neurons dates back to the work of McCulloch and Pitts (1943), who proposed a mathematical abstraction capable of executing simple logical operations. (Mcculloch & Pitts, 1943). As illustrated in Figure 1, an artificial neuron receives multiple weighted inputs, incorporates a bias term, and then applies a nonlinear activation function to generate an output. Subsequent milestones in the field include Rosenblatt's introduction of the Perceptron (1958), which represented the first neural model that could learn decision boundaries through supervised training (Frank, 1958).

A standard ANN is organized into three main components: an input layer, one or more hidden layers, and an output layer. Information is propagated forward from one layer to the next, where each neuron transforms its inputs and passes the resulting activation to subsequent units. During the learning process, the network iteratively updates its weights to reduce prediction error—a capability that became feasible with the introduction of the backpropagation algorithm by Rumelhart, Hinton, and Williams (1986) (Rumelhart et al., 1986).

Through iterative weight updates, ANNs learn to approximate complex nonlinear functions, enabling them to capture high-dimensional relationships within data. However, simple feedforward architectures can struggle with long-term dependencies, vanishing gradients, and overfitting, motivating the development of more specialized neural architectures such as RNNs, CNNs and Transformer-based models (Alam et al., 2023).

A RNN can capture and summarize historical information from previous time steps. However, its effectiveness is often limited by the vanishing gradient problem, which hinders learning over long sequences. To overcome this limitation, GRUs and LSTM networks have emerged as more robust alternatives, widely recognized as the preferred architectures for implementing recurrent neural networks in modern DL studies (Alam et al., 2023)

RECURRENCE-BASED APPROACHES: FUNDAMENTALS AND LIMITATIONS

Recurrent Neural Networks (RNNs)

RNNs were introduced to address the inability of traditional feed-forward neural networks to model sequential and temporal dependencies in data. Unlike ANNs, where information flows in a single direction, RNNs incorporate cyclic connections that allow information from previous time steps to persist and influence subsequent processing. This feedback

mechanism enables the network to maintain a hidden state that captures contextual information over time, making RNNs particularly effective for sequence modeling tasks such as speech recognition, machine translation, and language modeling (Elman, 1986; Yoon & Yu, 2020; Çelik et al., 2024). By maintaining this dynamic internal memory, RNNs can model the temporal structure of data; however, despite their strengths, they still struggle to capture long-term dependencies due to the vanishing gradient problem encountered during training (Bengio et al., 1994).

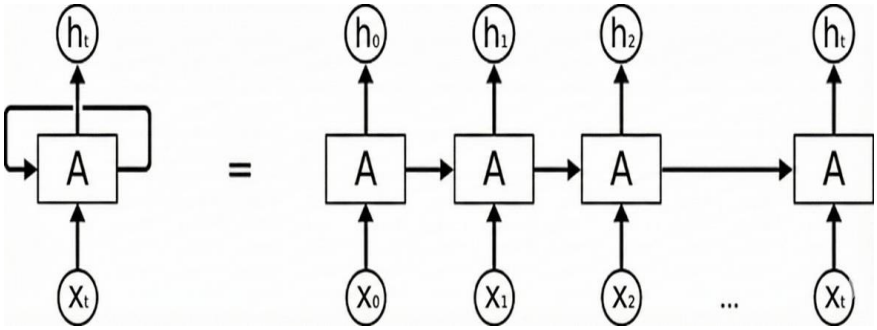


Figure 2. Structure of a Recurrent Neural Network (*Understanding LSTM Networks*, 2015)

Figure 2 illustrates how a recurrent neural network processes sequential data over time. On the left, the compact form of the RNN shows a single recurrent unit (A) that receives an input vector x_t and produces a hidden state h_t which is then fed back into the same unit at the next time step. On the right, the unfolded representation demonstrates how the same unit is replicated across multiple time steps (t_0, t_1, t_2,...,t_n), sharing the same weights across all iterations. Each unit takes the current input (x_t) and the previous hidden state (h_{t-1}) to compute the current hidden state (h_t). This unfolding over time reveals how RNNs preserve temporal dependencies by maintaining contextual information across sequential inputs.

The Challenge of Long-Term Dependencies

Storing and utilizing information across long dependency spans is fundamentally constrained by insufficient and vanishing error propagation in recurrent neural networks. During gradient-based training, error signals either decay exponentially or fail to propagate effectively across distant dependencies, preventing meaningful parameter updates in earlier network states. This vanishing gradient problem was first rigorously analyzed by Bengio et al. (1994), who showed that gradient-based learning struggles to capture long-range dependencies in conventional RNNs (Bengio et al., 1994). Hochreiter later provided a deeper theoretical analysis of this phenomenon and introduced LSTM as an architectural solution specifically designed to overcome the limitations caused by vanishing and insufficient error propagation. Moreover, truncated backpropagation further aggravates

insufficient error propagation by prematurely cutting gradient flow, which weakens dependency modeling even further (Trinh et al., 2018).

GATING MECHANISMS FOR ENHANCED CONTROL

Long Short-Term Memory (LSTM)

LSTM networks constitute a specialized class of recurrent neural networks designed to alleviate the vanishing gradient problem. Through the integration of a gating mechanism that selectively regulates information flow, LSTMs effectively preserve relevant information while filtering out irrelevant signals. This architectural design enables LSTMs to model long-range dependencies in sequential data more successfully than conventional RNNs, making them particularly suitable for complex sequence learning tasks (Hochreiter, 2016). At each time step, the LSTM unit processes the current input together with the previous hidden state to produce a new hidden representation that is transferred to the subsequent step. In sequence-level applications such as classification, the final hidden state is typically used as the representative summary of the entire input sequence. An LSTM cell is composed of three fundamental gates: the input gate, the forget gate, and the output gate, each of which plays a distinct role in controlling information dynamics. The input gate determines how new information is incorporated into the cell state, the forget gate regulates which past information should be discarded, and the output gate controls how the internal memory influences the output activation (Perumal et al., 2024).

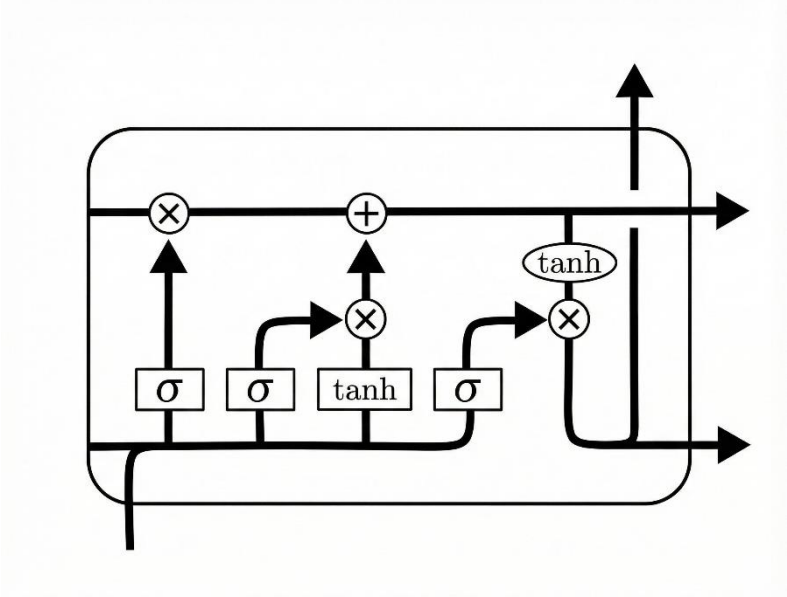


Figure 3. Internal structure of an LSTM cell (*Understanding LSTM Networks*, 2015)

Figure 3 illustrates the internal architecture of a LSTM unit, which extends the standard RNN by introducing a cell state and a set of gating mechanisms. These gates regulate the flow of information through the cell, allowing the network to retain or forget information over long time spans. From left to right, the diagram shows the forget gate**,** input gate**,** cell candidate, and output gate. Each gate uses a sigmoid (σ) activation function to produce values between 0 and 1, determining how much information should be passed through. The forget gate controls which parts of the previous cell state are discarded, while the input gate determines which new information will be added to the cell state. The cell candidate, generated using a tanh activation, represents potential new memory content. The cell state is then updated by combining the old state (after forgetting irrelevant information) with the new candidate (scaled by the input gate). Finally, the output gate controls which part of the updated cell state will be propagated as the hidden output, typically passed through a tanh activation to ensure bounded values. This gating mechanism enables LSTM networks to preserve gradients during backpropagation and effectively capture long-term dependencies in sequential data (Hochreiter, 2016).

Gated Recurrent Unit (GRU)

GRUs are a variant of LSTMs that employ a simplified gating mechanism. With fewer parameters, GRUs can achieve performance comparable to LSTMs across many natural language processing tasks, while being computationally more efficient (Chung, n.d.). GRU is a variant of the RNN architecture designed to address short-term memory limitations while

offering a simpler structure compared to LSTMs. GRUs merge the input and forget gates of LSTMs into a single update gate, eliminating the need for a separate cell state. Each GRU unit consists of three main components: the update gate, the reset gate, and the current memory content. These gates allow the GRU to selectively retain and utilize information from previous time steps, enabling the network to capture long-term dependencies in sequential data efficiently (Alam et al., 2023; Perumal et al., 2024).

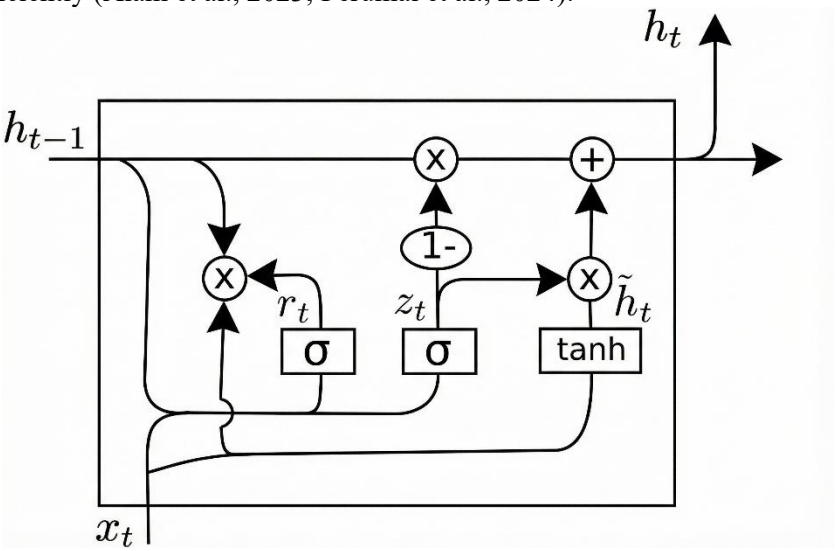


Figure 4. Internal structure of a Gated Recurrent Unit (*Understanding LSTM Networks*, 2015)

Figure 4 depicts the internal architecture of a GRU, an optimized variant of the LSTM network proposed by *Cho et al.* (2014) (Cho et al., 2014). Unlike the LSTM, which maintains separate cell and hidden states, the GRU merges these into a single hidden state h_t thereby simplifying the model while retaining its ability to capture long-term dependencies. The GRU introduces two key gates: the reset gate (r_t) and the update gate (z_t). Both gates employ a sigmoid (σ) activation function to control information flow. The reset gate determines how much of the previous hidden state h_{t-1}should be forgotten before computing the candidate activation h_t, while the update gate regulates how much of the new candidate information should replace the previous hidden state. The candidate activation is typically computed using a hyperbolic tangent (tanh) function, providing a non-linear transformation of the input and the reset-modulated hidden state. The final hidden state h_t is obtained by linearly interpolating between the previous hidden state and the candidate activation, as controlled by the update gate. This gating mechanism enables GRUs to maintain long-term contextual information with fewer parameters and lower computational cost compared to LSTMs, while achieving comparable performance in many NLP tasks (Chung, n.d.).

FEATURE EXTRACTION AND LOCAL PATTERNS: CONVOLUTIONAL NEURAL NETWORKS (CNNS)

CNNs are a class of deep learning architectures designed to automatically learn hierarchical feature representations from input data. Originally introduced for image recognition tasks, CNNs are inspired by the visual cortex of the human brain and excel at identifying local spatial patterns through the use of convolutional operations (LeCun et al., 1989). A Convolutional Neural Network (CNN) is a type of neural architecture primarily composed of three layers: convolution, pooling, and fully connected layers. The convolution and pooling layers extract hierarchical features from the input data, while the fully connected layer maps these features to the final Output In the convolution layer, small arrays of numbers called kernels (or filters) slide over the input tensor, performing element-wise multiplications to generate feature maps that capture high-level patterns and characteristics in the data (Kim, 2014).

CNNs are particularly effective in reducing the number of parameters compared to standard ANNs, which has encouraged researchers to adopt larger models for tasks that were previously challenging for traditional ANNs (Alam et al., 2023). Although CNNs were originally developed for image-based pattern recognition, their parameter efficiency and ability to capture local features have inspired their adaptation beyond the vision domain (Saxena, 2022). In the field of NLP, CNNs have been successfully applied to tasks such as text classification (Jacovi et al., 2018), sentiment analysis (Wazery et al., 2018), and sentence modeling, where they extract local and hierarchical features from sequences of words, capturing meaningful patterns without the need for recurrent structures.

ATTENTION MECHANISM-BASED APPROACHES

Attention Mechanism and *Multi-Head Attention*

The Transformer architecture is built entirely on attention mechanisms, enabling parallel computation that significantly accelerates training and enhances processing efficiency (Shaw et al., 2018; Katırcı & Çelik, 2024a). This mechanism allows the model to capture semantic relationships and long-range dependencies more effectively by dynamically modeling interactions between tokens within a sequence (Bahdanau et al., 2016) (Zhuoran et al., 2021). This innovation has substantially enhanced the ability of language models to handle complex sentence structures and achieve

higher performance across various natural language processing tasks (Zhang et al., 2020).

In attention mechanisms, multi-head self-attention plays a central role. By using multiple attention heads, the model can focus on different representation subspaces simultaneously, allowing it to extract diverse relational features from the input (Islam et al., 2024; Lu et al., 2024).

Self-attention sub-layers in the Transformer architecture typically employ *h* attention heads, where each head captures different aspects of the input data, thereby enriching the model's internal representations. Multi-head attention integrates multiple single-head attention mechanisms in parallel, and the outputs from each head are concatenated to form the final representation. This design allows the model to attend to information from multiple representation subspaces simultaneously, resulting in a more comprehensive understanding of the input. Within the attention mechanism, the query, key, and value vectors are fundamental in directing information flow. The query indicates which information a token should focus on, while the key and value vectors provide contextual information to support this focus. In self-attention, all three vectors are derived from the same input sequence, enabling each token to attend to every other token. This facilitates the modeling of contextual relationships and long-range dependencies within the sequence (Islam et al., 2024) (Shaw et al., 2018)(Bozic et al., 2024).

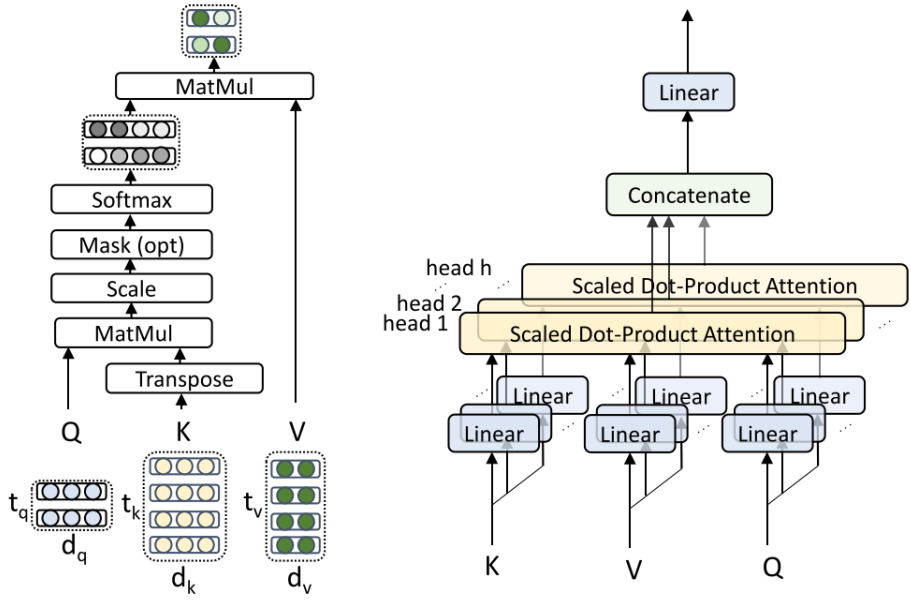


Figure 6. (left) Scaled Dot-Product Attention with a simple illustra- tion of its running process. (right) Multi-Head Attention consists of multiple Scaled Dot-Product Attention performing in parallel (Dong et al., 2018)

Figure 6 (left) shows the Scaled Dot-Product Attention mechanism, which computes attention weights by multiplying the Query (Q) and Key (K), scaling the result, applying an optional mask, and passing it through a SoftMax function. These weights are then used to compute a weighted sum of the Value (V) vectors, producing the output. Figure 6 (right) illustrates Multi-Head Attention, where Q, K and V are linearly projected *h* times and processed in parallel through independent attention heads. The outputs are concatenated and passed through a final linear layer. This structure allows the model to capture diverse relationships in the input more effectively than a single attention head.

Transformer Architecture

The Transformer architecture represents a paradigm shift in sequence modeling by relying entirely on the attention mechanism, which allows the model to capture global contextual dependencies without using recurrence or convolution (Vaswani et al., 2017). It employs self-attention to process all tokens in parallel, which allows faster training and improved handling of long-range relationships (Bahdanau et al., 2016). Unlike RNNs, LSTMs, and GRUs, which process input sequentially and are prone to the vanishing gradient problem, Transformers exploit parallel computation and multi-head attention to efficiently learn contextual meaning across entire sequences (Vaswani et al., 2017; Katırcı & Çelik, 2024a).

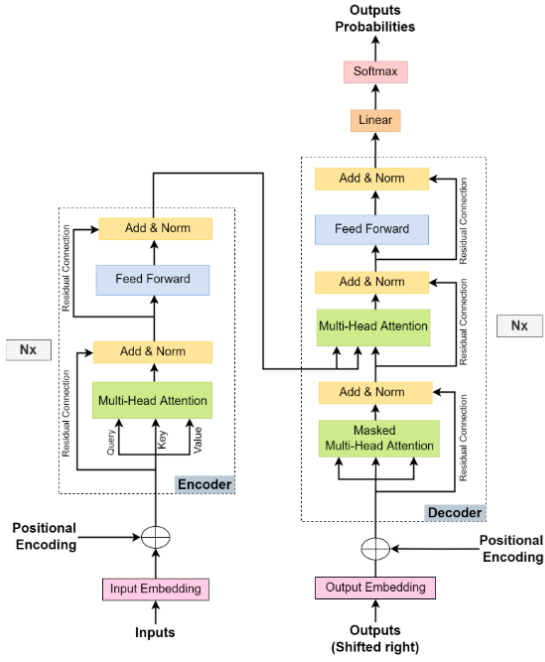


Figure 5. Transformer architecture (Islam et al., 2024)

Since its introduction, this architecture has become the foundation of modern NLP. It has been successfully applied to diverse domains, including machine translation (Katırcı & Çelik, 2025d, 2025a, 2025b) , human-like text production (Brown et al., 2020), automatic speech recognition systems (Dong et al., 2018), computer vision (Wang et al., 2022), text summarization (Liu & Lapata, 2019), and question answering (Çelik et al., 2024; Katırcı & Çelik, 2025c). One of the most notable examples of a Transformer-based model is the GPT series, which demonstrates the architecture's effectiveness in generating coherent, context-aware text by leveraging large-scale data (Radford et al., 2019). Beyond its broad application across DL and multimodal domains, the Transformer architecture differs from traditional deep learning models through its substantially larger number of components and modular substructures, such as multi-head attention, positional encoding, feed-forward layers, and normalization mechanisms. This architectural richness has motivated studies examining how individual components influence overall model performance. For instance, researchers have explored the role of model complexity and dataset size (Nakkiran et al., 2021; Çelik & Katırcı, 2025), as well as how language pair characteristics and corpus scale affect translation quality (Katırcı & Çelik, 2025b). Additional studies have investigated the effects of dropout configurations on generalization (Katırcı & Çelik, 2024b), the role of layer structure in model behaviour (Xiong et al., 2020; Csordás et al., 2024), and effect of head (Ainslie et al., 2023; Katırcı & Çelik, 2025a) and evaluated how different activation functions contribute to representational capacity and training stability (Katırcı & Çelik, 2025c).

As illustrated in Figure 5, the Transformer architecture follows an encoder–decoder structure composed of multiple stacked layers that process input and output sequences in parallel [12]. Each encoder layer contains two main sublayers: a multi-head self-attention mechanism and a position-wise feed-forward network, both followed by residual connections and layer normalization. The decoder mirrors this structure but includes an additional encoder–decoder attention sublayer, which allows it to attend to encoder outputs while generating the target sequence. Positional encoding is applied to both input and output embeddings to preserve sequence order, enabling the model to capture contextual relationships across tokens effectively (Vaswani et al., 2017; Katırcı & Çelik, 2024a).

Table 1. Chronological evolution of neural architectures and their key contributions

Model / Approach	Year	Main Contribution
Neuron	1943	Proposed the first mathematical model of an artificial neuron capable of performing logical operations inspired by biological processes.
Perceptron	1958	Introduced the perceptron model with weighted connections, summation, and activation functions — foundational for modern neural architectures.
Backpropagation	1986	Enabled the training of multi-layer neural networks by propagating error gradients backward; marked the rise of deep learning.
CNN	1989	Introduced local connectivity, weight sharing, and pooling for efficient spatial feature extraction in image data.
RNN	1990	Designed to handle sequential data through recurrent connections, enabling temporal dependency modeling.
LSTM	1997	Introduced gating mechanisms (input, output, forget gates) to mitigate vanishing/exploding gradients and capture long-term dependencies.
GRU	2014	Simplified version of LSTM by merging gates; achieved comparable performance with lower computational cost.
Transformer	2017	Replaced recurrence with self-attention, enabling parallelized computation and global context modeling across sequences.

Table 1 provides a summarized overview of the neural architectures discussed in the preceding subsections, highlighting their chronological development and key contributions to the field of deep learning. Each model—from the early neuron and perceptron structures to advanced architectures such as LSTM, GRU, and the Transformer—illustrates the progressive refinement of computational models designed to enhance representational capacity, sequence modeling and learning efficiency.

RESULTS AND DISCUSSION

The rise of Transformer-based architectures has decisively shaped modern DL, delivering major gains in text understanding, generation, and contextual modeling (Arsalan et al., 2024). However, high computational and memory demands remain significant obstacles, especially for long-context tasks and resource-constrained settings (Vaswani et al., 2017; Katırcı & Çelik,

2024a; Çelik & Katırcı, 2025). Moreover, the growing use of Transformer-based models in chatbots and question-answering systems reveals ongoing challenges—such as overfitting, underfitting, and hallucinations—that undermine model reliability. Addressing these issues will require stronger regularization techniques and more task-specific architectural designs (Katırcı et al., 2024). Future directions also include expanding multilingual and low-resource language coverage, improving interpretability and fairness, and integrating cognitive or multimodal signals (e.g. vision, speech) to build more robust, context-aware systems (Judijanto & andika, n.d.).

REFERENCES

Aasim, M., Katırcı, R., Akgur, O., Yildirim, B., Mustafa, Z., Nadeem, M. A., Baloch, F. S., Karakoy, T., & Yılmaz, G. (2022). Machine learning (ML) algorithms and artificial neural network for optimizing in vitro germination and growth indices of industrial hemp (Cannabis sativa L.). *Industrial Crops and Products*, *181*(October 2021). https://doi.org/10.1016/j.indcrop.2022.114801

Ainslie, J., Lee-Thorp, J., de Jong, M., Zemlyanskiy, Y., Lebrón, F., & Sanghai, S. (2023). GQA: Training Generalized Multi-Query Transformer Models from Multi-Head Checkpoints. *EMNLP 2023 - 2023 Conference on Empirical Methods in Natural Language Processing, Proceedings*, 4895–4901. https://doi.org/10.18653/v1/2023.emnlp-main.298

Alam, M. S. Bin, Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A. B. M., & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. In *Artificial Intelligence Review* (Vol. 56, Issue 11). Springer Netherlands. https://doi.org/10.1007/s10462-023-10466-8

Arsalan, M., Raza, S. A., Aayush, E., & Singh, P. (2024). *Transformers in Natural Language Processing : A Comprehensive Review*. *May*. https://doi.org/https://doi.org/10.22214/ijraset.2024.62863

Bahdanau, D., Cho, K., & Bengio, Y. (2016). *Neural Machine Translation by Jointly Learning to Align and Translate*. http://arxiv.org/abs/1409.0473

Bengio, Y., Simard, P., Frasconi, P., & Member, S. (1994). *Learning Long-Term Dependencies with Gradient Descent is Difficult*. *5*(2).

Bozic, V., Dordevic, D., Thommes, J., Coppola, D., & Singh, S. P. (2024). Rethinking Attention: Exploring Shallow Feed-Forward Neural Networks as an Alternative to Attention Layers in Transformers. *Proceedings of the AAAI Conference on Artificial Intelligence*, *38*(21), 23477–23479. https://doi.org/10.1609/aaai.v38i21.30436

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G.,

Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., … Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, *2020-Decem*.

Buscema, P. M., Massini, G., Breda, M., Lodwick, W. A., Newman, F., & Asadi-Zeydabadi, M. (2018). Artificial neural networks. *Studies in Systems, Decision and Control*, *131*(January 2015), 11–35. https://doi.org/10.1007/978-3-319-75049-1_2

Çelik, H., & Katırcı, R. (2025). The Impact of Optimizer Selection on Transformer Performance: Analyzing the Role of Model Complexity and Dataset Size. *Black Sea Journal of Engineering and Science*, *9*(1, 13–14). https://doi.org/https://doi.org/10.34248/bsengineering.1739598

Çelik, H., Katırcı, R., & İşlek, B. (2024). *Effect Of Parameters On Performance In Question-Answer Model With Simple Rnn Deep Learning Method*. *May*. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=7b0QCpsAAAAJ&citation_for_view=7b0QCpsAAAAJ:WF5omc3nYNoC

Cho, K., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. *Journal of Clinical Microbiology*, *28*(4), 828–829. https://doi.org/10.1128/jcm.28.4.828-829.1990

Chung, J. (n.d.). *Gated Recurrent Neural Networks on Sequence Modeling*. 1–9.

Csordás, R., Irie, K., Schmidhuber, J., Potts, C., & Manning, C. D. (2024). *MoEUT: Mixture-of-Experts Universal Transformers*. *NeurIPS*, 1–19. http://arxiv.org/abs/2405.16039

Dong, L., Xu, S., & Xu, B. (2018). Speech-transformer: A no-recurrence sequence-to-sequence model for speech recognition. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, *2018-April*, 5884–5888. https://doi.org/10.1109/ICASSP.2018.8462506

Elman, J. L. (1986). *Finding Structure in Time*. *1*(1 990). https://doi.org/10.1207/s15516709cog1402

Frank, R. (1958). The Perceptron: a Probabilistic Model for Information Storage and Organization in the Brain. *Psychological Review*, *65*(6), 386–408.

Hochreiter, S. (2016). *Long Short-Term Memory*. *November 1997*. https://doi.org/10.1162/neco.1997.9.8.1735

Islam, S., Elmekki, H., Elsebai, A., Bentahar, J., Drawel, N., Rjoub, G., & Pedrycz, W. (2024). A comprehensive survey on applications of transformers for deep learning tasks. *Expert Systems with Applications*, *241*. https://doi.org/10.1016/j.eswa.2023.122666

Jacovi, A., Shalom, O. S., & Goldberg, Y. (2018). Understanding Convolutional Neural Networks for Text Classification. *EMNLP 2018 - 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP, Proceedings of the 1st Workshop*, 56–65. https://doi.org/10.18653/v1/w18-5408

Judijanto, L., & andika, A. Y. (n.d.). *View of Emerging Research Trends in Natural Language Processing for Multilingual AI*. https://doi.org/10.58812/esiscs.v2i03

Katırcı, R., & Çelik, H. (2024a). *Transformer Mimarisi* (pp. 0–3). https://doi.org/10.5281/zenodo.13971609

Katırcı, R., & Çelik, H. (2024b). *Transformer Mimarisinde Dropout Oranlarının Performans Üzerindeki Etkisi*. *December*. https://www.researchgate.net/publication/387424078_TRANSFORMER_MIMARISINDE_DROPOUT_ORANLARININ_PERFORMANS_UZERINDEKI

Katırcı, R., & Çelik, H. (2025a). *Analyzing The Impact Of Attention Head Count On Transformer-Based Machine Translation*. *June*. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=7b0QCpsAAAAJ&citation_for_view=7b0QCpsAAAAJ:UebtZRa9Y70C

Katırcı, R., & Çelik, H. (2025b). Analyzing The Impact Of Language Pair And Dataset Size On Transformer Performance. *International Yıldırım Bayezid Scientific Research And Innovation Symposium-I*, *2*, 557–565. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=7b0QCpsAAAAJ&citation_for_view=7b0QCpsAAAAJ:UebtZRa9Y70C&gmla=AH8HC4zDD0WZNY4JvET_GFv1lVISPoOhP-7iBXhz5ne6CeA1rTHaj491MHG_n3LSy1Gw6oMO94H8jHEa9ZqxhPg7&sciund=9336085649281122449&gmla=AH

Katırcı, R., & Çelik, H. (2025c). Evaluating the impact of activation functions on transformer architecture performance. *International Science And Art Research Center Evaluating*, 626–639. https://scholar.google.com/citations?view_op=view_citation&hl=en&user=7b0QCpsAAAAJ&citation_for_view=7b0QCpsAAAAJ:Se3iqnhoufwC

Katırcı, R., & Çelik, H. (2025d). *Learning Rate Sensitivity In Transformer Models: A Case Study In*. *4*(May). https://doi.org/https://doi.org/10.5281/zenodo.15769066

Katırcı, R., Çelik, H., & Oğuz, T. (2024). *conversational ai and question answering systems: architectures, training challenges and solutions*. 47–68. https://doi.org/10.5281/zenodo.17449635

Kim, Y. (2014). *Convolutional Neural Networks for Sentence Classification*. 1746–1751.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & D Jacke, L. (1989). *Backpropagation Applied to Handwritten Zip code Recognition*. Neural Computation, 1(4), 541–551. https://doi.org/https://doi.org/10.1162/neco.1989.1.4.541

Liu, Y., & Lapata, M. (2019). Text summarization with pretrained encoders. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 3730–3740. https://doi.org/10.18653/v1/d19-1387

Lu, Y., Zhou, X., He, W., Zhao, J., Ji, T., Gui, T., Zhang, Q., & Huang, X. (2024). *LongHeads: Multi-Head Attention is Secretly a Long Context Processor*. http://arxiv.org/abs/2402.10685

Mcculloch, W. S., & Pitts, W. (1943). *A logical calculus of the Ideas Immanent nervous activity*. 52*(l), 99–115. https://doi.org/https://doi.org/10.1007/BF02478259

Mcculloch, W. S., & Pitts, W. (1990). *A logical calculus nervous activity*. 52*(l), 99–115.

Nakkiran, P., Kaplun, G., Bansal, Y., Yang, T., Barak, B., & Sutskever, I. (2021). *Deep double descent : where bigger models and more data hurt ∗*.

Perumal, T., Mustapha, N., Mohamed, R., & Shiri, F. M. (2024). A Comprehensive Overview and Comparative Analysis on Deep Learning Models. *Journal on Artificial Intelligence*, *6*(1), 301–360. https://doi.org/10.32604/jai.2024.054314

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). *Language Models are Unsupervised Multitask Learners*.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536.

https://doi.org/10.1038/323533a0
Saxena, A. (2022). An Introduction to Convolutional Neural Networks. *International Journal for Research in Applied Science and Engineering Technology*, *10*(12), 943–947. https://doi.org/10.22214/ijraset.2022.47789
Shaw, P., Uszkoreit, J., & Vaswani, A. (2018). Self-attention with relative position representations. *NAACL HLT 2018 - 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies - Proceedings of the Conference*, *2*, 464–468. https://doi.org/10.18653/v1/n18-2074
Torfi, A., Shirvani, R. A., Keneshloo, Y., Tavaf, N., & Fox, E. A. (2021). *Natural Language Processing Advancements By Deep Learning: A Survey*. 1–23. http://arxiv.org/abs/2003.01200
Trinh, T. H., Dai, A. M., & Quoc, M. L. (2018). *Learning Longer-term Dependencies in RNNs with Auxiliary Losses*.
Understanding LSTM Networks. (2015).
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-Decem*(Nips), 5999–6009.
Wang, J., Yang, Z., Hu, X., Li, L., Lin, K., Gan, Z., Liu, Z., Liu, C., & Wang, L. (2022). *GIT: A Generative Image-to-text Transformer for Vision and Language*. *2*, 1–49. http://arxiv.org/abs/2205.14100
Wazery, Y. M., Mohammed, H. S., & Houssein, E. H. (2018). Twitter sentiment analysis using deep neural network. *ICENCO 2018 - 14th International Computer Engineering Conference: Secure Smart Societies*, *June*, 177–182. https://doi.org/10.1109/ICENCO.2018.8636119
Wikimedia. (n.d.). *File:ArtificialNeuronModel english.png*.
Xiong, R., Yang, Y., He, D., Zheng, K., Zheng, S., Xing, C., Zhang, H., Lan, Y., Wang, L., & Liu, T. Y. (2020). On layer normalization in the transformer architecture. *37th International Conference on Machine Learning, ICML 2020*, *PartF16814*, 10455–10464.
Yoon, S. H., & Yu, H. J. (2020). A simple distortion-free method to handle variable length sequences for recurrent neural networks in text dependent speaker verification. *Applied Sciences (Switzerland)*, *10*(12), 1–14. https://doi.org/10.3390/APP10124092
Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent trends in deep learning based natural language processing [Review Article]. *IEEE Computational Intelligence Magazine*, *13*(3), 55–75. https://doi.org/10.1109/MCI.2018.2840738
Zhang, Y., Tuo, M., Yin, Q., Qi, L., Wang, X., & Liu, T. (2020). Keywords extraction with deep neural network model. *Neurocomputing*, *383*, 113–121. https://doi.org/10.1016/j.neucom.2019.11.083
Zhuoran, S., Mingyuan, Z., Haiyu, Z., Shuai, Y., & Hongsheng, L. (2021). Efficient attention: Attention with linear complexities. *Proceedings - 2021 IEEE Winter Conference on Applications of Computer Vision, WACV 2021*, 3530–3538. https://doi.org/10.1109/WACV48630.2021.00357

Environmental Pharmaceutical Pollution: Ecosystems, Ecotoxicological Impacts, and Management Strategies

Buse KOCAER BİBER[1]
Fatma Olcay TOPAÇ[2]

1. MSc, Pharmacist; Bursa, Türkiye. ecz.busekocaer@gmail.com
2. Prof. Dr.; Bursa Uludag University Faculty of Engineering, Environmental Engineering Department. olcaytopac@uludag.edu.tr ORCID No: 0000-0002-6364-4087

ABSTRACT

Today, the rapid increase in pharmaceutical consumption has led to the widespread occurrence of pharmaceutical compounds not only in clinical settings but also in environmental compartments. Non-metabolized drug residues released after use and improper pharmaceutical waste management constitute major sources of environmental contamination, particularly in wastewater and soil systems. Antibiotics, analgesics, hormones, and psychoactive substances transported into the environment not only exert ecotoxicological effects but also pose indirect risks to human health. Of particular concern, the environmental persistence of antibiotics accelerates the dissemination of antibiotic resistance genes, resulting in critical global public health consequences. In addition, endocrine-disrupting pharmaceutical residues can reach humans through water and food chains, thereby creating long-term health risks. This section discusses the sources of pharmaceutical pollution, its environmental fate, and its ecotoxicological impacts, together with its implications for public health. The transport pathways of pharmaceutical residues from wastewater treatment systems to natural ecosystems and their subsequent consequences for environmental and human health are evaluated in an integrated manner. Accordingly, pharmaceutical pollution is highlighted not only as an environmental issue but also as a holistic global threat that directly affects ecosystem integrity and human well-being.

Keywords – Advanced Wastewater Treatment, Ecopharmacovigilance, Ecotoxicology, One Health, Pharmaceutical Contaminants.

INTRODUCTION

In recent years, the extraordinary increase in global drug consumption has led to the accumulation of pharmaceutical residues not only in clinical settings but also in environmental ecosystems. For instance, pharmaceutical sales increased from approximately 390 billion USD in 2001 to 1.25 trillion USD in 2019 (Schlander et al., 2021). While this economic growth reflects the expansion of access to healthcare worldwide, it simultaneously indicates an increasing burden of pharmaceutical compounds being released into the environment. This long-term global growth trend in pharmaceutical consumption is also clearly illustrated by the worldwide market data presented in Fig. 1.

Indeed, the growth in drug consumption in emerging markets—particularly in China, India, Brazil, and Turkey, defined as “pharmerging” countries—has reached remarkable levels (Stacciarini, 2024). One of the most striking examples of this trend is antibiotics. According to recent data, antibiotic consumption increased by 16.3% in 67 countries between 2016 and

2023, with the total number of defined daily doses (DDD) rising from 29.5 billion to 34.3 billion (Klein et al., 2024). This rapid increase, especially in middle-income countries, further intensifies the global antimicrobial resistance crisis. Consequently, rising consumption places pressure not only on healthcare systems but also on the carrying capacity of environmental ecosystems.

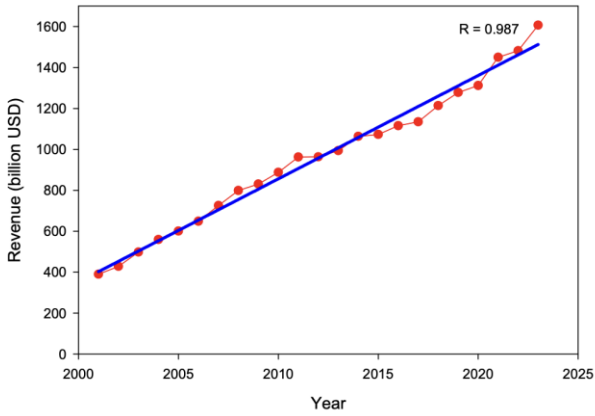


Fig. 1. Worldwide pharmaceutical market sales (2001–2023). Red points represent annual values (billion USD), and the blue line indicates the regression trend (Perez-Lucas & Navarro, 2024).

Field studies conducted on a global scale strongly support this perspective. In a comprehensive study covering 258 rivers across 104 countries, pharmaceutical residues were detected at levels posing risks to the environment or human health at more than 25% of the sampling sites (Wilkinson et al., 2022). This finding clearly demonstrates that pharmaceutical contamination is not a localized issue but rather a widespread global problem affecting water resources worldwide.

However, the fact that these pollutants are often not explicitly addressed in environmental regulations has brought the concept of “emerging contaminants” to the forefront. Pharmaceuticals, which can exert biological effects even at very low concentrations, are therefore classified as emerging pollutants (Elizalde-Velázquez et al., 2019; Wang et al., 2024a). In particular, antibiotics, hormones, analgesics, and psychoactive drugs are among the most intensively discussed groups due to their persistence in different environmental matrices and their ecotoxic potential. The detection of these compounds in surface waters, groundwater, and even drinking water further indicates that the problem is directly intertwined with daily human life (Vatovec et al., 2021).

From an ecotoxicological perspective, the presence of antibiotics in the environment accelerates the spread of resistance genes, thereby posing a critical threat to global public health (He et al., 2016). Moreover, studies conducted in marine aquaculture sites have shown that the coexistence of pharmaceutical and pesticide residues not only exerts toxic effects on aquatic organisms but also promotes the development of resistance genes (Zhang et al., 2021). These findings demonstrate that pharmaceutical pollution is not limited solely to ecological disturbances but also threatens human health through direct and indirect pathways.

At this stage, the issue of how to monitor and manage these risks becomes increasingly important. In this context, the concept of Ecopharmacovigilance (EPV), which has gained considerable attention in the literature, plays a central role. EPV refers to the process of monitoring, evaluating, and managing the presence of pharmaceuticals in environmental compartments, the risks they pose, and their long-term ecological impacts (Elizalde-Velázquez et al., 2019). Considered the environmental counterpart of clinical pharmacovigilance, this approach involves not only scientific data generation but also the active participation of regulatory authorities, the pharmaceutical industry, and society.

Accordingly, assessments of pharmaceutical pollution should not be limited to environmental observations alone; rather, within the framework of the One Health approach, they should be addressed through the holistic perspective offered by Ecopharmacovigilance. In this section, the sources of pharmaceutical pollution, their environmental fate, the ecotoxic effects they cause, and ultimately their implications for public health are discussed. In doing so, pharmaceutical pollution is presented as a comprehensive threat encompassing both scientific and governance dimensions.

SOURCES OF PHARMACEUTICAL POLLUTION

Pharmaceutical pollutants consist of residues of biologically active compounds developed to protect human and animal health that enter the environment after use. These substances are commonly detected at ng/L–µg/L levels in soil, surface waters, groundwater, wastewater, and even drinking water. Despite their low concentrations, they exhibit “pseudo-persistence” behavior in the environment, posing risks to ecosystem functioning, aquatic health, and the development of antibiotic resistance. It has been widely reported that these pollutants can induce adverse effects on aquatic and terrestrial organisms, including developmental disorders, reduced reproductive success, and the dissemination of resistance genes (Naseem et

al., 2025). In the current literature, pharmaceutical compounds are therefore classified as emerging contaminants and are considered a multidimensional risk group that simultaneously affects human, animal, and environmental health within the One Health framework (Wang et al., 2024a).

An integrated overview of pharmaceutical residues detected in freshwater systems, their major environmental hazards, and the corresponding international policy response mechanisms is presented in Fig. 2 (OECD, 2019). One of the primary pathways through which pharmaceuticals are transported into the environment is wastewater treatment plants (WWTPs). Many pharmaceuticals used in human and veterinary medicine are excreted unmetabolized or only partially metabolized and reach sewage systems; however, conventional treatment technologies are often insufficient to completely remove these compounds (Naseem et al., 2025). Consequently, frequently used pharmaceuticals such as naproxen, carbamazepine, and diclofenac persist in surface waters and sediments (Moreno Ríos et al., 2022). In addition to municipal wastewater, hospitals, clinics, and laboratories generate specialized waste streams containing high concentrations of pharmaceuticals. In many developing countries, these wastes are discharged untreated or only partially treated into the environment, further increasing the pharmaceutical load in aquatic ecosystems. Pharmaceutical manufacturing facilities also represent significant point sources, creating so-called “pharmaceutical hotspots” in high-production regions where large quantities of active compounds are directly released into aquatic and terrestrial environments. Moreover, improper disposal practices—such as discarding unused medicines into household waste or flushing them down the drain—further exacerbate uncontrolled pharmaceutical entry into environmental systems The application of sewage sludge to agricultural lands constitutes another important pathway, enhancing the transfer of pharmaceuticals to soil and crops and suggesting a potential route of indirect exposure through the food chain (Pandey et al., 2025; Anbarasu & Yuvarajan, 2025).

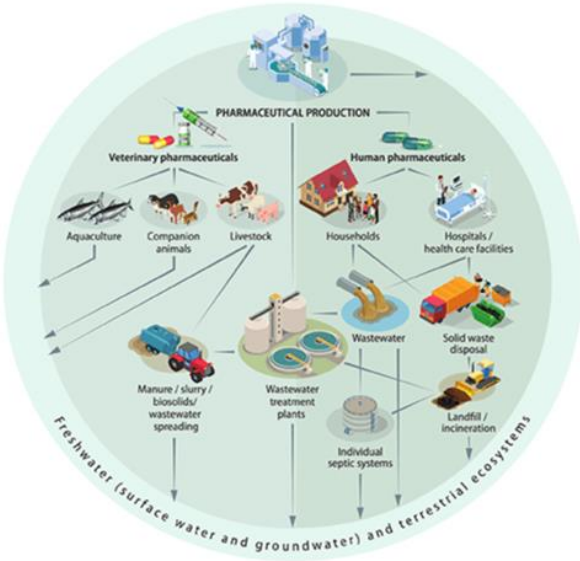


Fig 2. Pharmaceutical residues in freshwater and their associated hazards and policy responses (OECD, 2019).

The management of unused and leftover medicines represents a second critical dimension of pharmaceutical pollution. In many countries, expired medicines, discontinued prescriptions, or partially used drugs are disposed of as household waste or poured into drainage systems. These practices substantially increase the pharmaceutical load entering wastewater treatment plants and facilitate the subsequent transport of these compounds into environmental compartments. Although pharmacy-based medicine take-back and return programs are more widely implemented in high-income countries, their application remains limited in low- and middle-income regions due to insufficient public awareness, regulatory gaps, and financial constraints (Naseem et al., 2025; Pandey et al., 2025). This situation highlights household sources as one of the weakest links in pharmaceutical waste management.

Prescription practices and consumption habits further intensify pharmaceutical pollution. Excessive prescribing, widespread over-the-counter (OTC) analgesic use, high antibiotic consumption, and increasing self-medication behaviors significantly elevate the environmental pharmaceutical burden. It is estimated that approximately 15,000 tons of antibiotics are consumed annually in Europe and 100,000–200,000 tons worldwide (Naseem et al., 2025). Such large-scale consumption leads to a continuous and diffuse input of pharmaceuticals into the environment, both through human excretion and through improper disposal of unused drugs. The widespread OTC use of non-steroidal anti-inflammatory drugs (NSAIDs) such as naproxen, combined with their low removal efficiencies in conventional treatment systems, results in their persistence in environmental compartments (Moreno Ríos et al.,

2022). Furthermore, the increasing occurrence of antibiotics in aquatic environments constitutes a major global threat by promoting the spread of resistance genes (Wang et al., 2024). Plants may also be exposed to these compounds through irrigation water and contaminated soils, enabling their transfer along the food chain to humans (Anbarasu & Yuvarajan, 2024).

At the global scale, international organizations such as the World Health Organization (WHO), the United Nations Environment Programme (UNEP), and the European Union (EU) recognize pharmaceutical pollution as a priority environmental concern. Regulatory frameworks such as the EU Integrated Pollution Prevention and Control (IPPC) Directive, the REACH Regulation, and the Zero Pollution Action Plan aim to reduce chemical emissions at the source and promote the application of best available techniques in industrial processes. It was further reported that the establishment of the Intergovernmental Science–Policy Panel on Chemicals, Waste and Pollution Prevention under the United Nations is expected to strengthen the global capacity for life-cycle monitoring and risk management of emerging pollutants, including pharmaceuticals (Wang et al, 2024a). Although national programs for the collection and recovery of unused medicines through pharmacies are becoming more widespread—particularly in European countries—major deficiencies in household pharmaceutical waste management still remain. Moreover, the frequent detection of NSAIDs, antibiotics, and other pharmaceutical compounds in surface waters and sediments across different regions confirms that pharmaceutical contamination is not merely a regional issue but a pervasive global environmental problem (Moreno Ríos et al., 2022).

BEHAVIOR OF PHARMACEUTICALS IN ENVIRONMENTAL COMPARTMENTS

Pharmaceuticals are compounds that are increasingly detected in both aquatic and terrestrial ecosystems, classified among the "pollutants of emerging concern" due to their persistence, bioaccumulative properties and toxic effects. Since modern wastewater treatment plants often cannot fully remove these substances and their metabolites, pharmaceuticals accumulate and persist for long periods of time in various environmental ecosystems, such as surface water, groundwater, soil, and sediment. Moreover, transformation products (TPs) can often exhibit higher toxicity than parent compounds and reach higher concentrations than parent compounds in some environments (Maculewicz et al., 2022).

Environmental transport of pharmaceuticals occurs largely through surface waters; Many drug compounds, especially NSAIDs (diclofenac, ibuprofen, naproxen, ketoprofen), have been reported at ng/L–µg/L levels in river and lake waters (Rastogi et al., 2021). In particular, metabolites of high-persistence compounds such as carbamazepine and diclofenac have been found at higher levels than parent compounds in Norwegian fjords and European rivers, indicating that conversion products are a critical component in environmental assessments (Maculewicz et al., 2022). Groundwater, on the other hand, is a stable matrix with low biodegradation, where pharmaceutical metabolites accumulate, especially through leaching and infiltration. Studies carried out in the Barcelona area have shown that metabolites such as enalaprilat and 4-hydroxypropanolol reach levels several times higher than the parent compounds.

Soils constitute a significant exposure area due to sewage sludge application, animal manures, and irrigation water containing pharmaceuticals. Behavior of pharmaceuticals in soil; It depends on various physicochemical factors, which include the lipophilicity (logKow) of the compound, the adsorption/desorption balance, the amount of organic matter, the pH, and the ionic structure. Hydrophobic compounds are more retained in soils rich in organic matter, while hydrophilic pesticides can be transported more easily and reach groundwater. In addition, the soil solution is the main medium for transporting pharmaceuticals to plant roots; The extent of vegetative uptake varies depending on both soil characteristics and solubility of the compound (Perez-Lucas & Navarro, 2024).

Sediments, on the other hand, are peripheral compartments where lipophilic pharmaceuticals and conversion products accumulate intensively. The slowing of degradation under anaerobic conditions leads to the stable persistence of diclofenac and carbamazepine metabolites in sediments for many years (Maculewicz et al., 2022). This accumulation also increases the potential for biomagnification through trophic transfer.

One of the most fundamental elements that determine the fate of pharmaceuticals in the environment is their molecular properties. A high LogKow value indicates an increased bioaccumulation potential according to OECD criteria. Lipophilic compounds are more retained in adipose tissues, bind strongly to sediment and soil colloids, and can persist in biological systems for longer periods of time (Puckowski et al., 2016). In contrast, hydrophilic compounds have higher mobility and can be transported over great distances. The environmental half-lives of pharmaceuticals range widely; While ibuprofen can degrade rapidly under most conditions, it can behave much more stably than expected, with half-lives of up to 413 days in certain photolytic environments (Rastogi et al., 2021). Compounds such as diclofenac and carbamazepine, on the other hand, have been identified as pseudopersicent pollutants due to their high environmental resilience.

Pharmaceutical conversion products are particularly important in environmental assessment. Some of these metabolites, which are formed as a result of photolysis, oxidation, hydrolysis and microbial processes, may have higher solubility, longer persistence or stronger toxic effect than the parent compounds. Moreover, Puckowski et al. (2016) highlight that pharmaceutical metabolites can accumulate not only in the aquatic environment but also in the tissues of organisms, that is, they are an essential component of bioaccumulation processes.

Certain groups of pharmaceuticals exhibit different behavioral patterns in the environment. Antibiotics are widely detected in both aquatic and terrestrial environments due to their wide range of uses and the stability of their metabolites; Tetracyclines tend to have strong adsorption in soil, while sulfonamides have higher mobility. NSAIDs are considered among the priority pollutants due to their high volume of use and toxic metabolites; While diclofenac exerts various toxic effects on sediment and aquatic organisms, ibuprofen may remain more stable than expected in certain conditions (Rastogi et al., 2021).. Hormones, on the other hand, exert endocrine disrupting effects even at very low concentrations and are easily deposited in sediments and biotic tissues due to their lipophilic nature (Maculewicz et al., 2022; Puckowski et al., 2016).

All these data show that the environmental fate of pharmaceuticals is shaped not only by chemical properties, but also by the physical and biological properties of the environmental matrices to which they are exposed. When processes such as persistence, transport, formation of transformation products, and bioaccumulation are considered together, it is clear that pharmaceuticals can have long-term and complex effects on ecosystem components. For this reason, it is critical to consider not only the main compounds but also metabolites and transformation products in an integrated manner in environmental risk assessments.

MULTITROPHIC ECOTOXICOLOGICAL EFFECTS OF PHARMACEUTICAL POLLUTION

Pharmaceutical and personal care products (PPCPs) are widely detected in freshwater and marine ecosystems at nanogram–microgram per liter levels and are capable of eliciting pronounced biological responses in environmental organisms even at low concentrations due to their inherent biological activity. In a large-scale global survey covering 104 countries, it was reported that at least one pharmaceutical compound exceeded ecological safety thresholds or posed a risk for antimicrobial resistance in approximately one-quarter of the

investigated rivers (Wilkinson et al., 2022). Species sensitivity distribution analyses for marine organisms further indicate that current regulatory frameworks are insufficient to adequately assess the environmental risks of these compounds, particularly because hormones and neuropsychiatric drugs exert chronic effects at extremely low concentrations (Guo et al., 2025). Unlike many classical chemical pollutants, pharmaceuticals may therefore induce "therapeutic-like" effects in non-target organisms owing to the evolutionary conservation of the biological pathways they target.

Fish represent one of the best-documented organism groups for evaluating the ecotoxicological effects of pharmaceutical pollution. A systematic mapping study focusing on behavioral endpoints demonstrated that pharmaceuticals affect a wide range of traits, including locomotor activity, risk-taking behavior, predator avoidance, and reproductive behavior in fish (Martin et al., 2025). These behavioral endpoints are considered more sensitive than conventional toxicity indicators such as growth and mortality and thus serve as early-warning signals of sublethal contaminant effects. In particular, antidepressants such as selective serotonin reuptake inhibitors (SSRIs) act on neurotransmitter systems that are highly conserved among vertebrates, resulting in therapeutic-like behavioral alterations in fish, including increased shyness, reduced activity, and impaired social interactions. In a multigenerational fluoxetine exposure experiment conducted by Aich et al. (2025), guppy populations exhibited reduced activity levels, increased shelter use, impaired body condition, altered gonopodium height, decreased sperm motility, and a significant reduction in behavioral plasticity. Such changes affect not only individual fitness but also the adaptive capacity of populations, thereby increasing the fragility of population dynamics over the long term. From an endocrine disruption perspective, hormone derivatives such as estradiol have been shown to induce vitellogenin production, cause gonadal feminization, and lead to severe losses in reproductive success, with population-level effects observed up to population collapse in experimental mesocosms (Guo et al., 2025). These findings indicate that pharmaceuticals threaten the sustainability of fish populations through both neurophysiological and hormonal mechanisms.

Phytoplankton and algal communities are likewise highly sensitive to pharmaceutical contamination, and disturbances at this trophic level can generate cascading effects throughout aquatic ecosystems. Chia et al. (2021) demonstrated that antibiotics reduce photosynthetic pigment content in phytoplankton, suppress chloroplast ribosomes, damage Photosystem II, and disrupt the electron transport chain. These impairments lead to fundamental physiological deteriorations, including decreased photosynthetic efficiency, reduced growth rates, and lower biomass production. Pharmaceuticals have also been shown to alter phytoplankton community composition owing to species-specific differences in susceptibility; sensitive taxa decline while tolerant species proliferate, thereby reshaping the basic structure of the trophic

web. Moreover, antibiotics have been reported to disrupt colony (coenobial) structures in certain green algal species, affecting ecological traits such as light utilization, defensive capacity, and resistance to sedimentation. Such phytoplankton-level alterations can indirectly impact zooplankton and fish populations through reductions in primary production and oxygen generation, deterioration of water clarity, and upward propagation of effects through the food web.

The impacts on microbial communities are primarily associated with antibiotic contamination and are directly linked to the disruption of essential ecosystem functions. In a study conducted by Ullah (2025) in Dapeng Bay, microbial alpha diversity was shown to decline significantly at stations with high antibiotic loads, while resistant taxa such as Umbonibacter became dominant. In the same areas, critical resistance genes, including sul1, tetA, and intI1, were found at elevated concentrations, accompanied by enhanced horizontal gene transfer. These changes represent not only a restructuring of microbial community composition but also a weakening of key ecosystem processes such as organic matter mineralization and nitrogen cycling. Furthermore, the destabilization of microbial food webs and the accumulation of resistance genes in aquatic environments increase the likelihood of their transfer to higher trophic levels and ultimately to humans. Losses in microbial diversity also reduce ecosystem resilience, thereby increasing vulnerability to environmental stressors and pollution.

Taken together, these findings clearly demonstrate that pharmaceutical pollution exerts multidimensional ecological consequences across different trophic levels. Behavioral disturbances, impaired reproductive success, disruptions in photosynthetic processes, shifts in community composition, and the dissemination of resistance genes collectively threaten biodiversity, species interactions, and essential ecosystem services in aquatic environments. Moreover, the ability of pharmaceuticals to induce strong pharmacological effects even at very low concentrations underscores the necessity of addressing chronic and sublethal responses in ecotoxicological assessments. Accordingly, behavioral endpoints, community ecology approaches, microbial resistance profiling, and long-term multigenerational studies should be integrated into environmental risk assessment frameworks. In parallel, the development of pharmaceutical-specific, science-based water quality guidelines is particularly critical for both freshwater and marine ecosystems. Only through the adoption of such a holistic approach can the ecological and public health risks associated with pharmaceutical pollution be effectively mitigated.

INDIRECT EFFECTS ON HUMAN HEALTH

The persistence of pharmaceutical residues in environmental compartments triggers toxicological mechanisms that affect human health not only through direct exposure but also through more complex indirect and cumulative risk pathways. The incomplete removal of pharmaceutical active compounds in wastewater treatment plants, intense point-source discharges from pharmaceutical manufacturing facilities, improper drug disposal, the use of antibiotics in animal production, and the application of biosolid-based fertilizers continuously introduce these pollutants into the environmental cycle (Ashfield et al., 2025; Lunghi et al., 2025). As a consequence, humans are exposed to low but persistent concentrations through water and food chains. Although such exposures rarely result in acute toxicity, they may gradually contribute to the development of multidimensional health disorders over the long term.

Human exposure to pharmaceuticals via water and food is predominantly chronic, insidious, and cumulative in nature. Persistent antibiotics, NSAIDs, antidepressants, and endocrine-disrupting hormones are continuously detected at nanogram–microgram per liter levels in drinking water, agricultural products, fish, and other animal-derived food items (Samal et al., 2022; Zampelli & Verna, 2025). Because the metabolites of these compounds are often more stable than their parent substances, they may persist longer in the environment after undergoing purification and transformation processes. This creates a background "chemical exposure" capable of interfering with fundamental biological mechanisms in the human body, including cellular stress responses, activation of inflammatory pathways, oxidative damage, and disruption of hormonal homeostasis.

One of the most critical indirect consequences of pharmaceutical pollution for human health is the acceleration of antibiotic resistance through environmental selection. It is well established that 30–90% of administered antibiotics are excreted unmetabolized through urine and feces and cannot be completely removed during wastewater treatment (Zampelli & Verna, 2025). This leads to the continuous presence of low concentrations of antibiotic residues in surface waters, sewage sludge, and agricultural soils. The selective pressure exerted by environmental antibiotic residues promotes the spread of resistance genes within the environmental microbiota via horizontal gene transfer mechanisms such as conjugation, transformation, and transduction (Lunghi et al., 2025). Consequently, even natural ecosystems unrelated to clinical settings become reservoirs of resistance genes, which can be transferred to humans through water, food, or environmental contact. This process contributes to increasing treatment failures, harder-to-control infections, and the growing need for more potent or toxic antimicrobial agents, ultimately driving a major global public health crisis.

Endocrine-disrupting pharmaceuticals represent another significant indirect threat to human health. Synthetic estrogens such as 17α-

ethinylestradiol induce marked effects including feminization and reproductive disorders in fish, and it is widely anticipated that similar disruptions may occur in humans (Zampelli & Verna, 2025). The human health risk becomes particularly pronounced under conditions of low-dose but long-term exposure. Mounting biological evidence indicates that these compounds can interfere with thyroid hormone metabolism, reproductive hormone regulation, and stress hormone signaling. Chronic exposure to estrogenic substances through water and food has been associated with reduced male fertility, decreased sperm quality, menstrual cycle irregularities, and an increased risk of hormone-related cancers such as breast and testicular cancer (Lunghi et al., 2025). The vulnerability of the endocrine system is especially pronounced during fetal development and childhood, when even very low concentrations may disrupt critical developmental processes.

Neuroactive pharmaceutical residues, which are being detected with increasing frequency in environmental matrices, also pose potential long-term risks to the human nervous system. Antidepressants, antipsychotics, and antiepileptic drugs are known to induce behavioral alterations in aquatic organisms, and comparable mechanisms are predicted to operate in humans (Zampelli & Verna, 2025). Chronic microdose exposure via drinking water and food may lead to long-term dysregulation of neurotransmitter systems such as serotonin and dopamine, contributing to mood disturbances, sleep–wake cycle disruptions, and impairments in cognitive performance. Exposure during sensitive developmental windows, particularly the prenatal period and early childhood, is thought to pose an even greater risk to neurodevelopmental processes.

Chronic toxicity risks are further complicated by the synergistic effects of pharmaceutical mixtures present in the environment. Humans are simultaneously exposed to multiple pharmaceutical active substances rather than to individual compounds in isolation. These mixture effects may trigger pathophysiological processes including oxidative stress, DNA damage, mitochondrial dysfunction, immune modulation, and chronic inflammation (Samal et al., 2022). For example, the co-occurrence of low-dose NSAIDs with antibiotics can enhance cellular oxidative burden and promote chronic inflammatory responses. Likewise, interactions between antidepressants and hormone residues may exert combined stress on the neuroendocrine system, contributing to sleep disturbances, anxiety-related tendencies, and metabolic dysregulation.

The clinical implications of these indirect mechanisms are becoming increasingly apparent. The acceleration of antibiotic resistance leads to higher failure rates in the treatment of infectious diseases, imposing a substantial burden on both individuals and healthcare systems. Hormonal imbalances linked to endocrine-disrupting pharmaceuticals are associated with rising incidences of obesity, infertility, thyroid dysfunction, and hormone-related cancers. Chronic exposure to neuroactive compounds is believed to contribute

to mood disorders, attention deficits, and declines in cognitive performance. In addition, antibiotic residues that alter gut microbiota composition may facilitate immune system dysregulation, chronic inflammation, and an elevated risk of metabolic syndrome (Lunghi et al., 2025).

In conclusion, pharmaceutical pollution has emerged as an invisible yet powerful public health challenge in modern societies. Continuous exposure to very low concentrations of complex pharmaceutical mixtures through water and food contributes to a wide range of adverse outcomes, including the spread of antibiotic resistance, endocrine disruption, metabolic disorders, and neurological effects. Therefore, reducing the environmental burden of pharmaceuticals is critical not only for ecosystem protection but also for safeguarding public health, and the development of sustainable, science-based management strategies has become imperative.

MONITORING, TREATMENT, AND MANAGEMENT STRATEGIES

Pharmaceutical pollution has become a central item on the global environmental agenda as an inevitable by-product of modern healthcare systems. Conventional wastewater treatment plants are not specifically designed to completely remove pharmaceutical compounds, resulting in the frequent detection of antibiotics, analgesics, antidepressants, hormones, and other pharmaceuticals in surface waters, drinking water, and sediments. Consequently, the management of pharmaceutical pollution requires multidimensional strategies supported by advanced analytical monitoring, innovative treatment technologies, source reduction, pharmacy-based take-back programs, and strong regulatory frameworks. The persistence of pharmaceuticals in the environment is therefore not merely a technical issue but a complex environmental challenge with direct implications for public health, ecological stability, and sustainable development. Population growth, aging societies, expanding access to medical treatment, and the global rise in drug consumption further indicate that pharmaceutical contamination will become an even more critical environmental problem in the future.

In recent years, substantial progress has been achieved in analytical techniques for monitoring pharmaceuticals in environmental matrices. High-resolution mass spectrometry methods such as LC–MS/MS, UHPLC–MS/MS, and DI-LC–MS/MS enable sensitive multi-compound detection, even in highly complex matrices. Studies investigating pharmaceutical behavior in wastewater demonstrate that DI-LC–MS/MS shortens sample preparation time while allowing accurate quantification at trace

concentrations (Domínguez-García et al., 2025). These advances have made it possible to include biotransformation products, metabolites, and toxic by-products in routine analyses, thereby enabling a more holistic environmental risk assessment. In addition, the rapid scanning capacity of these techniques facilitates the generation of large datasets, supporting the development of artificial intelligence and machine-learning-based pollution modeling.

Monitoring studies of hospital wastewater further indicate that UHPLC–MS/MS provides very high accuracy and reproducibility for antibiotic and hormone analysis. Hospitals represent a major point-source category of pharmaceutical contamination, as their effluents may contain high concentrations of antibiotics, antineoplastic agents, and contrast media. Monitoring these sources is therefore essential for evaluating treatment plant performance and quantifying human-derived pollution loads. Recent high-resolution monitoring of hospital effluents using UHPLC–MS/MS demonstrated that although compounds such as paracetamol and caffeine were almost completely removed, several pharmaceuticals including carbamazepine, diclofenac, tramadol and sulfamethoxazole exhibited low or even negative removal efficiencies, highlighting the limitations of conventional wastewater treatment systems (Molnarova et al., 2024). These advances in analytical monitoring are also critical for regulatory authorities, since the establishment of environmental threshold values, risk assessments, and regulatory enforcement strongly depend on precise and reliable measurements.

A consistent finding of monitoring studies is that conventional treatment processes are unable to fully remove most pharmaceutical compounds. For this reason, advanced purification technologies play a key role in mitigating pharmaceutical contamination. Ozonation and advanced oxidation processes (AOPs) achieve high degradation efficiencies for a wide range of pharmaceuticals via hydroxyl radical production. Studies on ozonation-based systems indicate particularly high removal efficiencies for recalcitrant compounds such as diclofenac, carbamazepine and sulfamethoxazole, with reported removals frequently exceeding 90% under optimized conditions (Angeles et al., 2020; Mansımlı and Ateş, 2025, Shah et al., 2025). These observations highlight that chemical stability, polarity, and molecular structure strongly influence treatment performance. However, several studies also report that the toxicity of transformation products formed during oxidation may in some cases exceed that of the parent compounds, emphasizing the importance of evaluating AOPs alongside comprehensive toxicity monitoring.

Despite their effectiveness, AOPs entail several limitations including high energy demand, by-product formation, and the need for careful process optimization at large scale (Kanakaraju et al., 2025). When cost-effectiveness, chemical consumption, and operational stability are considered, AOPs alone may not always represent a fully sustainable solution. However, their

integration with membrane-based technologies significantly enhances overall removal performance.

Membrane processes such as membrane bioreactors (MBR), nanofiltration (NF), and reverse osmosis (RO) provide efficient options for the physical separation and biodegradation of pharmaceutical compounds. These systems exhibit particularly high retention efficiencies for small and polar pharmaceuticals and may allow near-complete removal of pesticides, antibiotics, and analgesics at treatment plant effluents (Rostam et al., 2020). Beyond their high removal capacity, membrane systems contribute to ecological risk reduction by limiting the discharge of toxic micropollutants into receiving waters. Nonetheless, membrane fouling, high energy requirements, and maintenance costs remain important technical challenges.

Hybrid AOP–membrane systems offer additional advantages by achieving higher mineralization rates and lower toxic by-product formation compared with individual treatment processes (Silva et al., 2025). By combining chemical oxidation with physical separation, these systems demonstrate superior performance in the treatment of persistent pharmaceuticals. Such technological advances are particularly important for water reuse and sustainable water-management strategies. Increasing water scarcity under climate-change conditions further elevates the necessity of ensuring that reclaimed water is free from pharmaceutical residues.

The need for advanced treatment technologies also reflects the high volume of pharmaceutical waste generated at the source. Improper disposal of unused or expired medicines represents a major contributor to pharmaceutical contamination. Surveys indicate that a large proportion of the population still discards medicines in household garbage, sinks, or toilets. A study conducted in community pharmacies in Zambia reported that most pharmacists observed incorrect disposal practices among patients and emphasized the urgent need for a national medicine take-back program (Kampamba et al., 2025). These findings suggest that public awareness of the environmental consequences of pharmaceutical waste remains insufficient and that the success of take-back programs is strongly linked to accessibility and awareness. A recent systematic review further confirms that well-designed unused-medicine take-back programmes substantially reduce improper household disposal and significantly increase the safe collection rates of pharmaceutical waste, particularly when supported by legislation and accessible return infrastructures (Wang et al., 2024b).

This situation is not unique to Zambia, as similar behavioral patterns have been documented in many countries. Studies evaluating “Drug Take-Back Day” programs demonstrate that such initiatives significantly increase public awareness and enable the safe collection of substantial quantities of pharmaceutical waste (Pangilinan et al., 2025). These programs are therefore widely regarded as effective socio-economic instruments that promote public participation in pharmaceutical waste management.

Internationally, the role of pharmacists in pharmaceutical pollution prevention is receiving increasing recognition. The sustainability report of the International Pharmaceutical Federation (FIP, 2023) highlights that pharmacists should educate patients about correct disposal practices, provide counseling to reduce unnecessary drug use, and actively participate in medicine take-back programs. Pharmacists are also strategically positioned to minimize unused drug volumes through prescription optimization and patient follow-up. This emerging role reflects a growing integration of pharmaceutical practice into environmental sustainability frameworks.

A key pillar of holistic pharmaceutical waste management is the establishment of robust policy and regulatory frameworks. The European Union's Pharmaceuticals in the Environment (PiE) strategy constitutes one of the most comprehensive global regulatory approaches for controlling pharmaceutical contamination. The 2019 EU Strategic Approach and the 2025 Working Group Reports propose a wide range of measures including environmentally responsible prescribing, training of healthcare professionals, environmental-risk assessment (ERA) of pharmaceuticals, packaging optimization, strengthening of national take-back networks, and extended producer responsibility (EPR). These regulations aim to address pharmaceutical pollution along the entire life cycle, from production to post-consumption stages (EC, 2025). Recent stakeholder-based evaluations of the EU regulatory system further demonstrate that the mandatory Environmental Risk Assessment (ERA) requirement in the authorization of medicinal products plays a central role in integrating environmental protection into pharmaceutical governance and is expected to gain even greater importance in future regulatory revisions (Zinken et al., 2024).

The OECD report on the management of household pharmaceutical waste (OECD, 2024) further provides a valuable policy framework by recommending financing mechanisms, standardized national collection systems, public information strategies, and shared producer responsibility. Although pharmaceutical waste management in Türkiye is currently regulated mainly under medical and hazardous-waste frameworks, increasing alignment with EU and OECD standards is anticipated. In this context, the training of healthcare professionals and the strengthening of public awareness are essential for the establishment of sustainable pharmaceutical waste-management practices in Türkiye.

Taken together, these findings demonstrate that pharmaceutical pollution management is not solely a scientific or technical challenge but also requires broad social transformation, strong public-health policies, and sustainability-oriented regulatory frameworks. Advanced monitoring capacity, high-performance treatment technologies, the active involvement of pharmacists in public awareness, and effective integration of international policy frameworks represent interdependent components of a comprehensive solution. The mitigation of pharmaceutical pollution can therefore only be

achieved through a multi-layered, interdisciplinary, and sustainable management approach.

FUTURE PERSPECTIVES AND HOLISTIC APPROACHES

Effective future management of pharmaceutical pollution requires moving beyond approaches limited solely to wastewater treatment technologies. Current trends clearly indicate that the entire life cycle of pharmaceuticals—from design and production to clinical use and final disposal—must be addressed together with their environmental impacts. The conceptual foundation of this holistic framework is provided by the One Health and Ecopharmacovigilance (EPV) approaches (Velo & Moretti, 2010; Trifirò & Crisafulli, 2022).

From a future-oriented perspective, Ecopharmacovigilance is no longer confined to environmental monitoring alone; rather, it is evolving into a strategic governance and policy tool that guides sustainable prescribing practices, supports pharmaceutical waste management, strengthens environmental risk assessment during drug authorization processes, and enables the establishment of environmental adverse-effect reporting systems (Singh, 2023; Kumari et al., 2023). In this way, drug safety is transformed into an integrated safety domain that is evaluated not only at the level of individual patient health but also at the ecosystem scale.

One of the key technological drivers of this transformation is the development of artificial intelligence (AI)-based environmental risk monitoring systems. AI is increasingly applied to the detection of pharmaceutical residues in environmental compartments, the interpretation of large monitoring datasets, the establishment of early-warning systems, and the generation of predictive risk models (Trifirò & Crisafulli, 2022; Kavianpour et al., 2024). Through these tools, pharmaceutical contamination can be addressed not solely through retrospective measurements, but within a predictive and proactive risk-management framework.

In parallel with digital monitoring and prediction systems, source-oriented prevention strategies represent one of the most critical pillars of future action. In this context, Green Pharmacy and Green Chemistry approaches emerge as fundamental tools for reducing the environmental footprint of both pharmaceutical production and pharmaceutical practice (Srivastava et al., 2022; Andersen, 2025). Green Chemistry promotes environmentally benign synthesis routes and low-waste manufacturing processes, whereas Green Pharmacy aims to implement sustainable practices

along the entire pharmaceutical value chain, from supply management to pharmaceutical waste handling.

Another core element of this holistic transformation is the restructuring of clinical practices and prescribing behaviors along sustainability principles. Reducing unnecessary drug use, prioritizing active ingredients with lower environmental persistence, and ensuring the safe collection of unused medicines play critical roles in mitigating the environmental pharmaceutical burden (Trifirò & Crisafulli, 2022; Andersen, 2025). Within this framework, clinical pharmacists emerge as key healthcare actors who not only ensure medication safety and therapeutic effectiveness but also actively contribute to environmental sustainability.

Overall, the future management of pharmaceutical pollution is shaped by a multi-layered governance model in which EPV-based monitoring systems, AI-supported environmental risk assessments, Green Chemistry in production, Green Pharmacy in practice, and sustainable prescribing policies operate in an integrated manner. This holistic approach is an unavoidable requirement not only for environmental protection but also for the long-term safeguarding of public health and the sustainability of healthcare systems.

CONCLUSION

The findings presented in this chapter demonstrate that pharmaceutical pollution is no longer solely an environmental contamination issue, but has evolved into a multidimensional global risk that threatens ecosystem integrity, biodiversity, and indirectly human health. The inability of wastewater treatment plants to completely remove pharmaceutical compounds, point-source discharges from pharmaceutical manufacturing and healthcare facilities, improper drug disposal practices, and the intensive use of antibiotics and non-steroidal anti-inflammatory drugs (NSAIDs) collectively generate a persistent chemical pressure on aquatic and terrestrial ecosystems. This pressure drives a wide range of ecotoxicological consequences, including the spread of antibiotic resistance, endocrine-disrupting effects, behavioral and reproductive impairments, the weakening of microbial cycles, and the disruption of trophic networks. Indirect exposure through water and the food chain further creates a chronic health-risk profile that cannot be overlooked, even at very low concentrations.

Accordingly, the management of pharmaceutical pollution cannot be confined to technical solutions based solely on advanced treatment technologies. Instead, Ecopharmacovigilance (EPV), integrated with the One

Health approach, requires a comprehensive governance model that incorporates artificial intelligence–supported monitoring systems, Green Chemistry and Green Pharmacy practices, sustainable prescribing policies, and effective pharmaceutical waste management strategies. International regulatory frameworks, the expanding role of pharmacists, community-based take-back programs, and advanced analytical monitoring infrastructures constitute the fundamental building blocks of this integrated approach. From a future perspective, the successful reduction of pharmaceutical pollution will depend on the consistent implementation of interdisciplinary, data-driven, and preventive policies that simultaneously prioritize environmental protection, public health, and sustainable development goals. Without the adoption of such an integrated strategy, the ecological and societal burden of pharmaceutical pollution is likely to continue to increase.

REFERENCES

Aich, U., Polverino, G., Yazdan Parast, F., Melo, G.C., Tan, H., Howells, J., Nosrati, R., Wong, B.B.M. (2025). Long-term effects of widespread pharmaceutical pollution on trade-offs between behavioural, life-history and reproductive traits in fish. *The Journal of Animal Ecology,* 94(3), 340–355. https://doi.org/10.1111/1365-2656.14152.

Anbarasu, K., Yuvarajan, D. (2025). Pharmaceutical Pollution in Soil and Water and Its Impacts On Human Health and the Environment. *Water Air Soil Pollut,* 236, 877. https://doi.org/10.1007/s11270-025-08552-4.

Andersen M (2025). Green pharmacy practices: A Sustainability-Based Model for Hospital Pharmacy Operations. *J Appl Pharm.,* 17, 466. https://doi.org/10.35248/1920-4159.25.17.466.

Angeles, L.F., , Mullen, R. A., Huang, I.J., Wilson, C., Khunjar, W. Sirotkin, H. I. McElroy, A.E., Aga, D.S. (2020). Assessing pharmaceutical removal and reduction in toxicity provided by advanced wastewater treatment systems. *Environ. Sci.: Water Res. Technol.,* 6, 62-77 https://doi.org/10.1039/C9EW00559E †

Ashfield, N., Li, J., Bouzas-Monroy, A., Boxall, A.B.A. (2025). Silent side effects: Pharmaceuticals as contaminants of emerging concern. *Annual Review of Environment and Resources,* 50, 273–301. https://doi.org/10.1146/annurev-environ-111523-101837.

Chia, M.A., Lorenzi, A. S., Ameh, I., Dauda, S., Cordeiro-Araújo, M.K., Agee, J.T., Okpanachi, I.Y., Adesalu, A.T. (2021). Susceptibility of phytoplankton to the increasing presence of active pharmaceutical ingredients (APIs) in the aquatic environment: A review. *Aquatic Toxicology* (Amsterdam, Netherlands), 234, 105809. https://doi.org/10.1016/j.aquatox.2021.105809.

Domínguez-García, P., Tchantchane, C., Artigues, M. et al (2025). Application of direct injection LC–MS/MS for pharmaceutical monitoring in wastewater and reclaimed water from the 14 wastewater treatment plants in Costa Brava (Catalonia, Spain). *Appl Water Sci,* 15, 287. https://doi.org/10.1007/s13201-025-02615-x.

EC (European Commission). (2025). Pharmaceuticals in the environment: Implement the EU strategic approach. European Commission HEALTH AND FOOD SAFETY DIRECTORATE-GENERAL. https://health.ec.europa.eu

Elizalde-Velázquez, A., Gómez-Oliván, L.M., Islas-Flores, H. (2019). Introduction. In L.M. Gómez-Oliván (Ed.), *Ecopharmacovigilance: Multidisciplinary Approaches to Environmental Safety of Medicines* (pp. 1–12). Springer, Cham. https://doi.org/10.1007/698_2017_147.

FIP (International Pharmaceutical Federation) (2023). *Environmental sustainability within pharmacy*. FIP. https://www.fip.org.

Guo, X., Wang, J., Wang, Y.Y.L., Nguyen, T.D., Hua, F., Liu, W., Wang, Z. (2025). Derivation of interim guideline values for pharmaceuticals and personal care products in marine environment. *Marine pollution bulletin,* 219, 118325. https://doi.org/10.1016/j.marpolbul.2025.118325.

He, Z., Cheng, X., Kyzas, G.Z., Fu, J. (2016). Pharmaceuticals pollution of aquaculture and its management in China. *Journal of Molecular Liquids*, 223, 781–789. https://doi.org/10.1016/j.molliq.2016.09.005.

Kampamba, M., Chawangwana, M., Mwanza, B. et al. (2025). Assessment of pharmacists' perspectives on state-run drug take-back programs and disposal practices for unused and expired medications in Zambian community pharmacies. *Discov Public Health,* 22, 522. https://doi.org/10.1186/s12982-025-00941-8.

Kanakaraju, D., Glass, B. D., Goh, P. S. (2025). Advanced oxidation process-mediated removal of pharmaceuticals from water: a review of recent advances. *Environmental science and pollution research international*, 32(24), 14316–14350. https://doi.org/10.1007/s11356-025-36547-5.

Kavianpour, B, Piadeh, F, Gheibi, M, Ardakanian, A, Behzadian, K, Campos, LC. (2024). Applications of artificial intelligence for chemical analysis and monitoring of pharmaceutical and personal care products in water and wastewater: A review. *Chemosphere.* 368, 143692. https://doi.org/10.1016/j.chemosphere.2024.143692.

Klein, E.Y., Impalli, I., Poleon, S., Denoe, l. P., Cipriano, M., Van Boeckel, T.P., Pecetta, S., Bloom, D.E., Nandi, A. (2024). Global trends in antibiotic consumption during 2016-2023 and future projections through 2030. *Proc Natl Acad Sci USA,* 3,121(49), e2411919121. https://doi.org/10.1073/pnas.2411919121.

Kumari, M., Singh, H., Kashyap, R., Singh, V., Singhal, M., Sethiya, N. K., Kumar, B. (2023). Ecopharmacovigilance: An emerging concern for aquatic and terrestrial life. *World Journal of Pharmacy and Pharmaceutical Sciences,* 12(6), 373–390. https://doi.org/10.20959/wjpps20236-24844.

Lunghi, C., Osimani, A., Clementi, F., Aquilanti, L. (2025). Call to action: Pharmaceutical residues in the environment: threats to ecosystems and human health. *Drug Safety,* 48, 315–320. https://doi.org/10.1007/s40264-024-01497-3.

Maculewicz, J., Kowalska, D., Świacka, K., Toński, M., Stepnowski, P., Białk-Bielińska, A., Dołżonek, J. (2022). Transformation products of pharmaceuticals in the environment: Their fate, (eco)toxicity and bioaccumulation potential. *The Science of the total environment*, 802, 149916. https://doi.org/10.1016/j.scitotenv.2021.149916.

Mansımlı, M., & Ateş, H. (2025). Removal of pharmaceutically active compounds from hospital wastewater by ozonation pretreatment. *Environmental Research and Technology,* 8(3), 571-580. https://doi.org/10.35208/ert.1511061.

Martin JM, Michelangeli M, Bertram MG, Blanchfield PJ, Brand JA, Brodin T, Brooks BW, Cerveny D, Fergusson KN, Lagisz M, Lovin LM, Ligocki IY, Nakagawa S, Ozeki S, Sandoval-Herrera N, Scarlett KR, Sundin J, Tan H, Thoré ESJ, Wong BBM, McCallum ES. (2025). Evidence of the impacts of pharmaceuticals on aquatic animal behaviour (EIPAAB): A systematic map and open access database. *Environ Evid.,* 20;14(1), 4. doi: 10.1186/s13750-025-00357-6.

Molnarova, L., Halesova, T., Tomesova, D., Vaclavikova, M., Bosakova, Z. (2024). Monitoring pharmaceuticals and personal care products in healthcare effluent wastewater samples and the effectiveness of drug removal in wastewater treatment plants using the UHPLC-MS/MS method. *Molecules* (Basel, Switzerland), 29(7), 1480. https://doi.org/10.3390/molecules29071480.

Moreno Ríos, A.L., Gutierrez-Suarez, K., Carmona, Z., Ramos, C.G., Silva-Oliveira, L.F. (2022). Pharmaceuticals as emerging pollutants: Case naproxen an overview. *Chemosphere,* 291, 1, 132822. https://doi.org/10.1016/j.chemosphere.2021.132822.

Naseem, A., Alneghery, L.M., Al-Zharani, M., Nasr, F.A., Jawad, S.S., Umer, M., Sayyed, R., Ilyas, N. (2025). An insight into the impacts of pharmaceutical pollutants on the ecosystem and the potential role of bioremediation in mitigating pharmaceutical pollutant. *International Journal of Pharmaceutics,* 680, 125791. https://doi.org/10.1016/j.ijpharm.2025.125791.

OECD (Organisation for Economic Co-operation and Development). (2024). *Management of pharmaceutical household waste*. OECD Publishing. https://www.oecd.org

OECD (Organisation for Economic Co-operation and Development). (2019), *Pharmaceutical Residues in Freshwater: Hazards and Policy Responses*, OECD Studies on Water, OECD Publishing, Paris

Pandey, S., Jaiswal, R., Ahmad, S., Ali, A., Jaiswal, R., Yadav, R., Yadav, R., Ahsan, R.,Dwivedi, T., Sonali, S. (2025). Pharmaceutical waste management using functional materials for environmental remediation. *Sustainable Chemistry One World*, 6, 100068. https://doi.org/10.1016/j.scowo.2025.100068.

Pangilinan, K., McDaniel, J., Nadpara, P., Goode, J.V.R. (2025). Impact of a Drug Take Back Day event at a large chain community pharmacy. *JAPhA Practice Innovations*, 2, 3, 100042. https://doi.org/10.1016/j.japhpi.2025.100042.

Perez-Lucas, G., Navarro, S. (2024). How pharmaceutical residues occur, behave, and affect the soil environment. *Journal of Xenobiotics,* 14(4), 1343–1377. https://doi.org/10.3390/jox14040076.

Puckowski, A., Mioduszewska, K., Łukaszewicz, P., Borecka, M., Caban, M., Maszkowska, J., Stepnowski, P. (2016). Bioaccumulation and analytics of pharmaceutical residues in the environment: A review. *Journal of*

pharmaceutical and biomedical analysis, 127, 232–255. https://doi.org/10.1016/j.jpba.2016.02.049.

Rastogi, A., Tiwari, M.K., Ghangrekar, M.M. (2021). A review on environmental occurrence, toxicity and microbial degradation of non-steroidal anti-inflammatory drugs (NSAIDs). *Journal of Environmental Management,* 300, 113694. https://doi.org/10.1016/j.jenvman.2021.113694.

Shah, R. Shah, S.I.A., Anwar, M., Ahmad, W., Khan, D., Ullah, I., Khan, J.A., Han, C., Shah, N.S. (2025), Pharmaceuticals and personal care products: sources, occurrences, toxicities, and degradation by advanced oxidation and reduction processes- a review. *Separation and Purification Technology*, 135850, https://doi.org/10.1016/j.seppur.2025.135850.

Rostam, A.B., Taghizadeh, M. (2020), Advanced oxidation processes integrated by membrane reactors and bioreactors for various wastewater treatments: A critical review. *Journal of Environmental Chemical Engineering,* 8, 6, 104566. https://doi.org/10.1016/j.jece.2020.104566.

Samal, K., Mahapatra, S. Ali, M.H. (2022). Pharmaceutical wastewater as Emerging Contaminants (EC): Treatment technologies, impact on environment and human health. *Energy Nexus,* 6, 100076. https://doi.org/10.1016/j.nexus.2022.100076.

Schlander, M., Hernandez-Villafuerte, K., Cheng, C.Y., Mestre-Ferrandiz, J., Baumann, M. (2021). Pharmaceutical innovation and global spending on medicines. *PNAS,* 118(25), e2016968118. https://doi.org/10.1073/pnas.2016968118.

Singh, T (2023). The significance of ecopharmacovigilance in the current environmental situations. *IJPPR Human.,* Vol. 26 (2): 1-7.

Srivastava, P., Wal, A., Chaturvedi, V., Wal, P., Srivastava, A., Yadav, A., Sachan, P.P. (2022). Sustainability in pharmaceuticals: An in-depth look at the green pharmacy principles. *Journal of Population Therapeutics and Clinical Pharmacology,* 29, 1304–1325. https://doi.org/10.53555/jptcp.v29i04.4373.

Stacciarini, J.H.S. (2024). The global pharmaceutical sector: Numbers and dynamics. *Caminhos de Geografia,* 25, 101, 240-251. https://doi.org/10.5281/zenodo.13891633.

Trifirò, G., Crisafulli, S. (2022). A new era of pharmacovigilance: Future challenges and opportunities. *Frontiers in Drug Safety and Regulation,* 2, 866898. https://doi.org/10.3389/fdsfr.2022.866898.

Ullah, K. (2025). Understanding the effects of antibiotic contamination on marine ecosystems: Microbial diversity and resistance gene dynamics. *Regional Studies in Marine Science,* 90, 104428. https://doi.org/10.1016/j.rsma.2025.104428.

Vatovec, C., Kolodinsky, J., Callas, P., Hart, C., Gallagher, K. (2021). Pharmaceutical pollution sources and solutions: Survey of human and veterinary medication purchasing, use, and disposal. *Journal of Environmental Management,* 285, 112106. https://doi.org/10.1016/j.jenvman.2021.112106.

Velo, G., Moretti, U. (2010). Ecopharmacovigilance for better health. *Drug-Safety,* 33, 963–968. https://doi.org/10.2165/11539380-000000000-00000.

Wang, F., Xiang, L., Leung, K.S.Y., Elsner, M., Zhang, Y., Guo, Y., … Tiedje, J.M. (2024a). Emerging contaminants: A One Health perspective. *The Innovation,* 5(4), 100612. https://doi.org/10.1016/j.xinn.2024.100612.

Wang, L.S., Aziz, Z., Wang, E.S., Chik, Z. (2024b). Unused medicine take-back programmes: a systematic review. *Journal of pharmaceutical policy and practice,* 17(1), 2395535. https://doi.org/10.1080/20523211.2024.2395535.

Wilkinson, J. L., Boxall, A.B.A., Kolpin, D.W., Leung, K.M.Y., Lai, R.W.S., et. al. (2022). Pharmaceutical pollution of the world's rivers. *Proceedings of the National Academy of Sciences of the United States of America*, 119(8), e2113947119, https://doi.org/10.1073/pnas.2113947119.

Zampelli, S., Verna, R. (2025). Environmental Pollution from Pharmaceuticals. *Life* (Basel, Switzerland), 15(9), 1341. https://doi.org/10.3390/life15091341.

Zhang, R., Du, J., Dong, X., Huang, Y., Xie, H., Chen, J., Li, X., Kadokami, K. (2021). Occurrence and ecological risks of 156 pharmaceuticals and 296 pesticides in seawater from mariculture areas of Northeast China. *Science of the Total Environment,* 792, 148375. https://doi.org/10.1016/j.scitotenv.2021.148375.

Zinken, J.F., Pasmooij, A.M.G., Ederveen, A.G.H., Hoekman, J., Bloem, L.T. (2024). Environmental risk assessment in the EU regulation of medicines for human use: An analysis of stakeholder perspectives on its current and future role. *Drug Discovery Today*, 29(12), 104213. https://doi.org/10.1016/j.drudis.2024.104213.

Evaluating the Thermal Comfort Conditions in Traditional Mud Buildings of South-Eastern Nigeria

Akubue Jideofor ANSELM

Department Architecture, Faculty of Environmental Sciences, Baze University Nigeria
(akjideofor@yahoo.com)

ABSTRACT

The issue of occupant thermal comfort is paramount in the discussion of sustainable housing. In the case of Sub-Saharan Africa where majority of the population live in Low-cost houses, identifying the best solutions to indoor thermal comfort is of significant value. This study was conducted in order to gather information on the thermal and climatic conditions as well as the thermal preferences of occupants of traditional mud houses in the South-Eastern part of Nigeria. The data collected through physical measurement and surveys were subjected to descriptive analysis. Graphs and normal distribution tables were used to determine the comfort temperature, thermal sensation and thermal preference of the occupants. The result of the study showed that thermal comfort has significant value as well as indicated that more research is required in some of the aspects of the study, especially in the area of design, performance of traditional building materials and other variables that determine thermal comfort in the traditional mud residential buildings in the study area.

Keywords – Thermal Comfort, Traditional Buildings, Mud Houses, Thermal Sensation, Thermal Preferences, South-Eastern Nigeria.

INTRODUCTION

Thermal comfort is defined by experts as "that condition of mind which expresses satisfaction with the thermal environment" American Society of Heating, Refrigerating and Air Conditioning Engineers [1]. Thermal comfort depends on the mean radiant temperature and air temperature. Overheating would be experienced when either or both of these rise to an unacceptable level [2].

Human body is often exposed to a series of heat production, both from internal metabolism of company and that of surrounding elements. As these happen, the human body tries to maintain a thermal balance with its environment [3]. These heat exchanges through the ambient environment can be controlled so that the occupants of building will be thermally comfortable through architectural designs options aimed at providing optimal conditions for livability [4]. Parameters that are effective in thermal comfort analyses are often categorized into personal and environmental parameters. While, air temperature, relative humidity, air velocity and radiant temperature are considered as environmental parameters; factors considered as personal parameters consists of the human metabolic activity level and clothing [5]. Another major parameters for thermal comfort studies is the element of sensation and perception. Different Models exist for the study and analysis of thermal comfort amongst which is thermoneutrality [6]. One of the mosrt common models is the PMV-model, which suggests

that the most comfortable thermal sensation is neutral. However, other than thermal sensations, there exists other measures associated with perception of the thermal environment, which includes thermal satisfaction, thermal comfort, thermal acceptability and thermal preferences. These conditions are believed to impact on the appropriateness of any thermal condition [7]. It is believed that thermoneutrality does not essentially correspond to the preferred thermal sensation. A typical instance, may refer to a condition that is very warm outside, whereas people prefer somewhat cool conditions over thermoneutrality or slightly warm feeling. On the other hand, when it is very cold outside, people often prefer slightly warm conditions instead, thus thermoneutrality is not always often the ideal preference [8], [9]. As stated earlier, the human body maintains thermal balance with the environment through varieties of autonomous activities. Other than these activities, people often employ strategies to assist in adapting to indoor or outdoor thermal conditions Study has it that people in hot climate zones prefer higher indoor temperatures than people living in the colder climatic regions. This contrasts with the fundamental assumptions for comfort standards based on PMV-model [10]. Hence, the element of adaptation becomes a gradual reduction of the human reactions to repeated environmental stimulation, which can be behavioral with regards to clothing, physiological with regards to acclimatization or psychological with regards to expectations [11], [12].

It is common knowledge that thermal discomfort is a challenge to poor regions like the tropical regions of Sub-Saharan Africa due to heat and humidity stress, as buildings in this region are constantly exposed to solar radiation daily [13].

Study shows that a large number of thermal comfort studies have been conducted in buildings for all types of climate. Most of these were carried out in tropical, sub-tropical and temperate climate zones [14]. Indication suggests that thermal comfort in traditional mud residential buildings are way higher than those of conventional ones. This present study is mainly concerned with traditional mud residential buildings as characterized by mud walls and thatch roofs used to fully envelope the indoor space without considerations for windows [15].

BACKGROUND OF THE STUDY

The South-Eastern Nigerian regions comprises of the mainly Igbo tribe, who occupy the area of six geopolitical states in Nigeria [16]. It lies within the warm-humid zone of the tropics and characterized by high temperature and high humidity. Dominating winds in this region are South-West trade

winds from March to October and North-East trade winds from November to March. The wind direction at various times of the year plays a domineering role in determining the climate of the country [17].

The sun is one of the most abundant resources in the South East zone of Nigeria which is within the warm humid zone of the tropics. The zone is characterized by high temperature and high humidity. The effect of these therefore results in thermal discomfort associated with the zone. Contemporary and traditional buildings in the zone are constantly exposed to solar radiation every day. Despite this, a casual observation reveals that there is thermal comfort in the indoor space of the traditional mud residential buildings, even during the daytime when solar radiation is much felt at the outdoor of the buildings. Probably, the thermal sensation expressed by the observer might be as a result of the design of such buildings. A review of available literature reveals that there is no verification to validate the opinion surrounding the comfort conditions in mud houses, as most of these speculations are devoid of any empirical analyses. Owing to this, there is need for empirical study on effectiveness of thermal comfort in these traditional mud residential buildings in the South East zone of Nigeria. This study therefore aims to investigate the effectiveness of thermal comfort in the traditional residential buildings in the South East Zone of Nigeria, with a view to contributing towards measures to improve the thermal comfort in the indoor space of the buildings, in the case the existing ones are found inadequate.

A. Description of study site

The study in this work was conducted around the Enugu region which is located in the Southeastern part of Nigeria with geo-location as 6.26° North and 7.29° East. The predominant human activity in the study area is characterized as farming community. The regional climate is classified as a tropical rain forest with a derived savannah. Set on an altitude of 304.7m above sea level, its climate is characteristically humid and this humidity is oppressive between March and November as indicated in figure 1 [18]. The average maximum temperature is 34.9°C, with average low of 22.3°C the annual mean temperature is 26.7°C. Typical with the rest of West Africa, the rainy and dry seasons are the only weather periods in this location. The hottest months of the year are February to April, a period in which isolation increases with the apparent north ward movement of the sun and the cloud cover is sparse. Temperatures for February-April are typically above 27°C, while those of September, October and November are typically above 26°C. The coolest month is usually August, middle of rainy seasons, high cloud cover, temperatures are typically around 24°C. Annual temperature in south east are fairly high generally around 26°C.The average hourly wind speed experiences significant seasonal variation over the course of the year. The windier part of the year lasts for 7 months, between March to September,

with average wind speed of 5.5m/s. The windiest month of the year is August, with an average wind speed of 7.1m/s. The wind direction comes relatively from the west for 10 months (monsoon season), from January to November, with a peak percentage of 75% in July. Additionally the wind direction comes from the north for 2 months (Harmattan season) through November to January, with a peak percentage of 41% in January. The surroundings of the measurement sites can be characterized as typical rural land-use, with residential housing. The ground in the study zone is firm with abundance of building earth for traditional earthen architecture. Forest products provide materials for construction ranging from post (to hold the roof) to sticks for wattle-and-daub construction, and thatch and raffia mats for covering the roof. Conversely, with the rise of modern agriculture, the South east area shed most of its thick forests giving way to open cultivated lands. Also, the growth of towns and modern institutions has equally contributed to a transformation of the vegetation and physical features of the area.

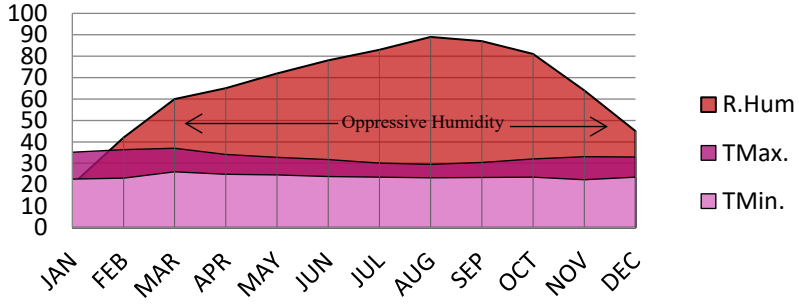


Figure 1. Average climatic data of Enugu city

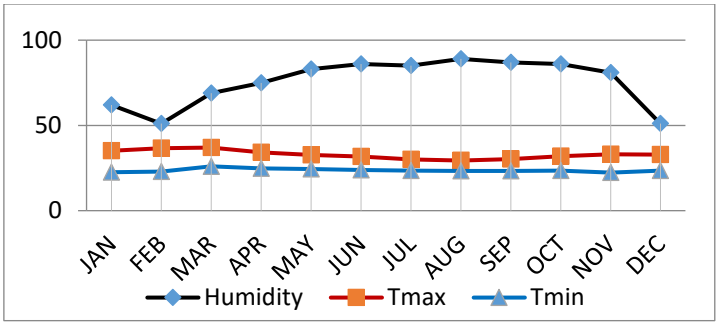


Figure 2. Average monthly temperature and relative humidity of Enugu city during the year of study

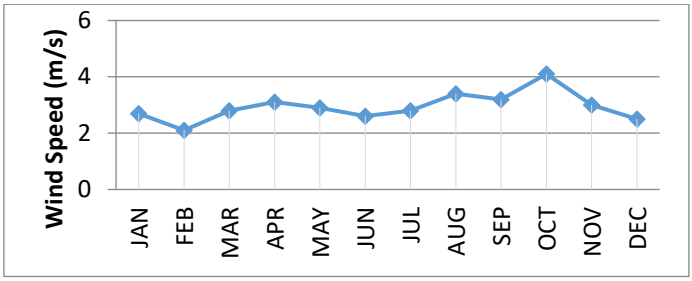


Figure 3. Average monthly wind speed of Enugu city during the year of study

STUDY METHODOLOGY

The authors conducted field assessments to gather information on the thermal and climatic conditions as well as the thermal preferences of occupants. The data gathered were supplemented with the experiment process to obtain the result.

All the climate data and other information were analyzed using graphical analysis mode (mathematical graphs and charts) to produce the detailed report for analyzing the effectiveness of thermal comfort in this study.

The following instruments were employed for the experimental studies as shown in Figure 4:

a. Digital thermometer (for air temperature)

b. Thermo-anemometer (for air movement/ or air velocity) indoor and outdoor.

c. Electronic sensor connected to data logger (for relative humidity)

d. A digital still camera (Sony cyber shot) was used to take photographs of some of the selected Igbo traditional residential buildings in the study area.

e. A laptop (computer) was used for Computational Fluid Dynamics (CFD) simulation. It is useful in predicting the indoor thermal environment. The laptop was also used for collation of data generated during the study.

The data collected from the study were subjected to descriptive analysis by describing them in frequency tabulation. Graphs and normal distribution tables were used to determine the comfort temperature, thermal sensation and thermal preference. The input comprised the data collected from the primary and secondary sources. These were data collected through physical measurement and surveys. The information obtained was inputted graphically and in tables. The output from data analysis was utilized in the

research. The data from the secondary source and the results from the analysis of primary source were applied to determine the thermal comfort. The result of the application was presented and showed that thermal comfort has value. It also showed that more research is required in some of the aspects of the transposition, especially in the area of design, performance of traditional building materials and other variables that determine thermal comfort in the traditional residential buildings in the study area. The data output formed the basis of subsequent studies and conclusions reached by the study. Data output involved building typologies, graphs and analysis of the results.

Raytek, (Raynger ST) Infra-red Sensor temperature monitoring equipment.	Temperature-Velocity-Humidity monitoring equipment

Figure 4. Climate and thermal Data equipment used in the field experiment measurements

INVESTIGATIONS AND FIELD MEASUREMENTS

A. Experimental Measurement

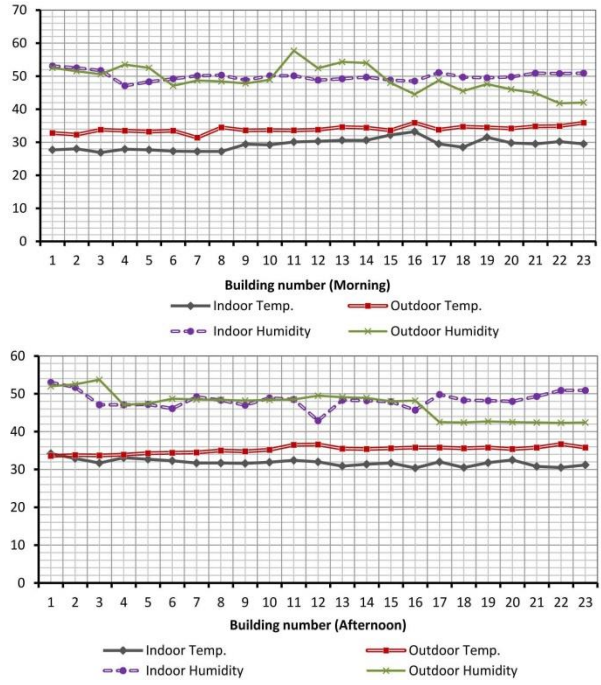
The field survey recorded the environmental and personal parameters which are important factors affecting thermal comfort. The relatively low mean clothing values of 0.27 clo and 0.34 clo reflect the casual dress codes in their homes. The typical male attire is short pants or and t-shirt, while the common female attire consists of a wrapper and blouse. In some occasions, the male respondents were observed to wear only pants with upper body bared during the hot afternoon survey in their homes and sweaters during the

early mornings and late evenings. Relatively equal numbers of male and female respondents participated in the survey. In terms of age distribution, the biggest percentage of respondents (about 50%) was represented by young adult age group (between 21 and 30 years old) and the smallest percentage was from oldest age group (more than 50 years old). The average height and weight of the respondents indicated that they are within the typical (normal) African human body proportion.

B. Investigation and Field measurements

The survey was conducted at the traditional residential buildings cluster of Ugwogo -Nike (figure 5). The Figure 6 shows the graphs representing the information gathered from the survey both in the morning and afternoon periods.

Figure 5. Images of the typical traditional mud houses in the study area



(a) Morning period (b) Afternoon period

Figure 6. Graphs showing the information from survey (morning and afternoon periods)

Using Humphrey's comfort temperature equation for deriving the global expected comfort temperature stated as follows:

Tc = 0.534To + 12.9 ……..…… (equation 1)

Where Tc is the comfort temperature and To is the mean outdoor temperature, the derived comfort temperature for this location is presented in Figure 7.

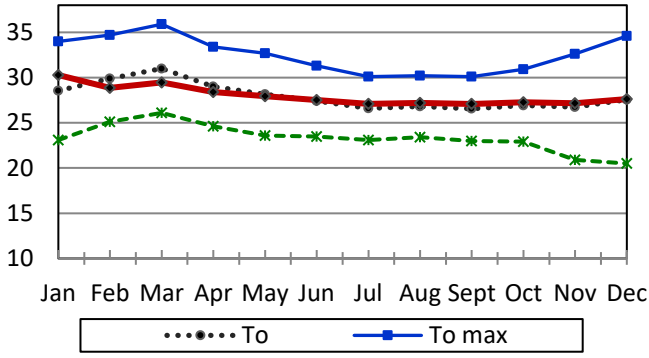


Figure 7. Average comfort temperature Tc indicated in red, as compared to mean outdoor temperature to

C. Summary of Field Investigation

The following were observations from the field investigation:

a. The mean temperatures recorded indoors in the morning hours as shown in Table 1, ranged from 26.9 ºC to 33.20 ºC, while that of afternoon (indoors) ranged from 30.4 ºC to 34.2ºC. The mean temperature outdoor for morning hours ranged from 31.40 ºC to 33.96ºC, while that of the afternoon periods ranged from 33.6 ºC to 36.70 ºC.

b. The mean relative humidity (RH) recorded indoors in the morning hours ranged from 47.10% to 53.1%. On the other hand, outdoor morning hours relative humidity measurements ranged from 41.80% to 57.70%. The afternoon hours recorded 42.30% to 53.70%. This is due to the dry season when the moisture content of the atmosphere is less because the sky is bright with sunshine.

c. The mean air velocity (m/s) recorded indoors in the morning hours were 0.0m/s. In the afternoon, the same 0.0m/s was recorded. The recorded low air velocities for the dry season indoors were due to the fact that all measurements were taken inside the typical traditional residential buildings in the study area. Secondly, the miniature windows used in the huts where considered to be responsible for lack of air flow and ventilation. At the outdoor in the morning hours as shown in Table 5, the air velocities ranged from 1.08 to 1.80m/s while in the afternoon periods, the air velocities ranged from 1.17 to 1.80m/s.

Table 1. Summary of climatic data as measured

	Morning period (Indoor)	**Afternoon Period (Indoor)**	**Morning period (Outdoor)**	**Afternoon Period (Outdoor)**
Mean air temperature (°C) (Mean, Standard deviation) (Minimum, maximum)	(29.30, 1.69) (26.90, 33.20)	(31.82, 0.91) (30.40, 34.20)	(33.96, 1.02) (31.40, 35.9)	(35.20, 0.92) (33.60, 36.70)
Mean relative humidity (%) (Mean, Standard deviation) (Minimum, maximum)	(49.96, 1.38) (47.10, 53.10)	(48.37, 2.11) (42.90, 53.00)	(49.08, 4.05) (41.80, 57.70)	(47.14, 3.53) (42.30, 53.70)
Mean air velocity (m/s) (Mean, Standard deviation) (Minimum, maximum)	(0.04, 0.14) (0.0, 0.50)	(0.0, 0.0) (0.0, 0.0)	(1.08, 0.56) (0.0, 1.80)	(1.17, 0.56) (0.0, 1.80)

D. Thermal Comfort Measurements

This part of the survey was conducted simultaneously with the field measurements. It involved a survey conducted through questions directed to the occupants of the houses. The questions were intended to identify Thermal Comfort elements based on human body reactions to the environment.

The questions were made up of the following:

a. ASHRAE 7-point scales for thermal sensation vote (TSV). These involve subjective thermal sensation vote in the following manner: (-3) cold, (-2) cool,(-1) slightly cool, (0) neutral, (1)slightly warm, (2) warm, (3) hot.
b. McIntyre (preference) 3-point scales, such as (-2) much cool, (-1) bit cool, (0) no change, (1) bit warm, (2) much warmer.

The data generated from the experiments conducted in this location were put through a Thermal Comfort evaluation system using standard evaluation parameters in order to significantly identify the levels of comfort and its effects within the selected study regions. The ASHRAE 7-Point scale (Table 2) and the Thermal Comfort Scale (Table 3) were adopted for the thermal

comfort assessment as analysis in order to determine the Thermal Sensation factor (TS) and the Thermal Preference factor (TP).

Table 2. ASHREA 7 point scales

Cold	Cool	Slightly cool	Neutral	Slightly warm	Warm	Hot
-3	-2	-1	0	1	2	3

Table 3. Thermal comfort Scale (Preference scale)

Cooler	Slightly cooler	Neutral (No change)	Slightly warmer	Warmer
-2	-1	0	1	2

ANALYSIS AND DISCUSSION OF RESULTS

The study further applied the thermal feeling measuring scale ASHRAE 7-Point scales and Thermal Comfort Scale for the measurements of the feelings of occupants. The comfort temperature (Tc) using M. Humphrey's equations were calculated. The summary of thermal sensation and thermal preferences is shown in Table 4. This demonstrated the general conditions of the occupants in terms of their acceptability of the indoor conditions. The conditions gave a mean regional comfort temperature value of 29.0 °C with mean indoor temperatures of 29.3 °C (morning) and 31.82 °C (afternoon). Relative humidity values were 49.96% for the morning and 48.37% for the afternoon period. Actual Mean Vote (AMV) of the occupants for the surveyed traditional buildings in the morning and afternoon periods indicated that none of the respondents felt neutral (0). On the other hand, the TSV (thermal Sensation vote) for the morning period recorded a mean Sensation value of -2.4 with SD 1.2 (indicating a cool condition) and 1.0 with SD 0 for the afternoon period (indicating a warm condition). The mean TPV (thermal Preference vote) for the morning gave a 0.7 value with SD 0.6 (indicating the preference for slightly cool condition) which accounted for 37.8% of the vote ,while the afternoon period gave a mean preference value of -1.1 with SD 0 (preference for slightly cooler conditions) with 0%.

Table 4. Summary of thermal sensation and thermal preferences

Time	Sample size	Thermal sensation (-3 to +3) TSV AMV			Thermal preference (-2 to +2) TPV			Tg (when voting)		Tgnv (voting neutral)	
		Mean	SD	Frequency of neutral %	mean	SD	Frequency of no change%	mean	SD	mean	SD
Morning	37	-2.4	1.2	Nil	0.7	0.6	37.8	29.2	2.9	nil	Nil
Afternoon	nil	Nil 1.0	nil	Nil	Nil -1.1	nil	Nil	nil	nil	nil	Nil

CONCLUSION

In this study, the indoor conditions of mud traditional buildings were studied within the South-Eastern part of Nigeria. The study reveals that the traditional mud buildings were at least more adaptable to warm and cold conditions. The traditional building materials are actually more suitable and adaptable to warm humid climate. The occupants of these traditional mud buildings in the study area voted in a unique way between thermal sensation and thermal preference agreeing with the common logical assumption of thermal comfort standard which assumed that under neutral conditions the occupants should prefer to feel neither cooler nor warmer meaning that their conditions were neutral in overall preference. However to further improve the comfort conditions of occupants, the miniature window openings observed in the buildings need to be modified in other to further improve the indoor natural ventilation conditions.

REFERENCES

[1] Goodchild B. (1974). Class differences in environmental perception: an exploratory study. Urban Studies, 11: 157-169.

[2] ASHRAE. ANSI/ASHRAE Standard 55-2020 (2020). Thermal Environmental Conditions for Human Occupancy. Atlanta: ASHRAE

[3] Heerwagen D. (2004). Passive and Active Environmental Controls: Informing the Schematic Designing of Buildings. The McGraw Hill Co. Inc; New York.

[4] Olgyay, V. (1958). Design with Climate. Princeton University Press, Princeton, N.J.

[5] Özdamar S. M., Umaroğulları F. (2018). Thermal Comfort and Indoor Air Quality. International journal of scientific research and innovative technology, 5: 90-109.

[6] Hoof J., Mazej M., Hensen J. (2010). Thermal comfort: Research and practice. Frontiers in Bioscience. 15: 765-788.

[7] Humphreys M. A., Hancock M. (2007). Do people like to feel "neutral"?: Exploring the variation of the desired thermal sensation on the ASHRAE scale. Energy and Buildings, 39(7):867-874

[8] Butala V., Muhič S. (2007). Perception of air quality and the thermal environment in offices. Indoor Built Environ 16(4): 302-310

[9] Fountain M. E., Brager G. S., de Dear R. J. (1996). Expectations of indoor climate control. Energy Buildings, 24(3): 179-182

[10] de Dear R. J., Brager G. S., Cooper D. (1997). Developing an adaptive model of thermal comfort and preference, Final Report ASHRAE RP-884. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers.

[11] de Dear R. J., Brager G. S. (2002) .Thermal comfort in naturally ventilated buildings, revisions to ASHRAE Standard 55. Energy and Buildings, 34(6): 549-561

[12] de Dear R. J., Brager G. S. (1998). Developing an adaptive model of thermal comfort and preference. ASHRAE Trans 104(1A): 145-167

[13] Olufowobi M. B. (1988). Design Impute for Rural Housing in Nigeria, (NIA) Nigeria Institute of Architects, Journal, 4: (1) 23-25

[14] Ajibola K. (2001) Design for Comfort in Nigeria: A Bioclimatic Approach. Renewable Energy, 23: 57-76.

[15] Al-ajim F., Loveday D.L. (2010). Indoor thermal conditions and thermal comfort in air-conditioned domestic buildings in the dry-dessert climate of Kuwait. Building and Environment, 45:704-710

[16] Dmochowoski: Z. R. (1990). An Introduction to Nigerian Traditional Architecture Vol. 3, Ethnographical Ltd. London, in Association with the National Commission for Museums and Monuments. Lagos

[17] Aniakor C. C. (2002) Igbo Architecture: A survey of the Igbo Nation. Africana Publishers Ltd. Onitsha, Nigeria.

[18] Anyadike, R. N. C. (2002) Climate and Vegetation: A Survey of the Igbo Nation. Africana First Publishers Ltd. Onitsha, 73-82.

Investigation of drilling parameters of CNT and graphene doped hybrid carbon-aramid fibre reinforced composite plates

Serkan ERDEM[1]

1- Assoc. Prof..; Fırat Üniversitesi Mühendislik Fakültesi Makine Mühendisliği Bölümü
serdem@firat.edu.tr ORCID No: 0000-0002-2504-0135

ABSTRACT

In this study, the puncture behavior of composite plates produced by vacuum infusion method with the help of epoxy resin from 4-layer carbon-aramid hybrid fabric was investigated at 500-1000-2000 rpm rotation speeds and 50-100-200 mm.min-1 feed rates. In addition, by adding wt.1-2% nano-sized carbon nano tube (CNT) and graphene into the epoxy resin, its effect on the drilling performance was investigated. The fuzzing, burr and crack behavior of the plates as a result of drilling were investigated for the front and back surfaces of the plate. As a result, visual fuzzing in the carbon-intensity regions was much less than in the aramid-intensity region. For the front and back surfaces, the cutting quality decreased as the feed rate increased at 500 rpm cutting speed, and the addition of CNT to the fabric increased the cutting quality. The effect of the variation of rotation and feed rate on hole quality could not be obtained in a certain trend. Similarly, the addition of CNT and graphene to the composite did not show a single trend. Therefore, the designer should choose the parameters that are suitable for him and make his design that way.

Keywords – CNT, Graphene, Drilling, Fuzzing, Burr.

INTRODUCTION

The whole mechanism is formed by the joint of the elements that make up the structure and the tools. This joint is made chemically, thermally and mechanically. Holes are drilled in the parts so that the mechanical joint can be made. Parts can be joined by putting in screws and rivets through these holes. The effect of parameter changes in drilling holes in isotropic materials is known. However, examining the behavior of fiber-reinforced composite materials during drilling is a current issue. When the studies on the drilling of fiber reinforced composites are examined: Geier et al. found a method to predict the burrs formed during drilling in carbon fiber reinforced polymer (CFRP's). First, digital images of the CFRP plate were taken under different lighting. Next, the fiber orientation of the chopped fiber group in each image was determined. These images were associated with each other. Finally, areas with burrs were predicted by the local characteristics of the chopped fibers. With the proposed method, burr prediction in chopped CFRPs could be made at a rate of 64–97% (Geier et al., 2022). Nagaraj et al. investigated the dry drilling machining performance under minimum lubrication and hybrid conditions. The effect of the change of thrust power, torque and feed on the drilling parameters was evaluated in terms of surface roughness and delamination damage. The results showed that drilling the CFRP composite material using coolant/lubricant gave better results than dry drilling (Nagaraj et al., 2022). Mathew and Shunmugesh optimum values of

drill diameter, spindle speed rotation and drill feed rate were obtained using Taguchi's L27 for light and gray analysis in MATLAB. From the analyzes made via MATLAB, it has been observed that the feed rate is the most effective parameter in drilling holes (Mathew et al., 2023). Benyettou et al. investigated the drilling process of bio composites reinforced with palm fibers (CDPF) at different feed rate and rotation speeds. Drilling performance was evaluated in terms of circularity of the hole, delamination and surface quality. The results showed that the hole quality increased with the increase of feed rate and decreased with the increase of spindle speed (Benyettou et al., 2022). Chrispin Das et al. produced pineapple leaf reinforced phenolic resin composite by hand laying and examined its mechanical properties. The effect of drilling time on this composite was analyzed with the Taguchi technique. The most ideal spindle speed feed rate to obtain minimum drilling time was obtained as 1000 rpm and 0.15 mm.min-1 (Chrispin Das et al., 2023). Ahuja et al. investigated the effect of hole drilling on delamination in composite. There will be 8 layers of fiber; 8 jute, 6 jute and 2 silk, 4 jute and 4 silk, 2 jute and 6 silk, 8 silk were used. The produced composites were drilled at different drilling speeds. Lower delamination was obtained in hybrid composites than in pure jute and pure silk (Ahuja et al., 2023). Gupta et al. examined the delamination that occurred during drilling in the Pultruded glass fiber reinforced polymer composite (PGFRP) with hybrid filler composites produced in terms of drill tip angle, spindle speed and feed. They used Taguchi method for optimization. In terms of delamination, 3000 rpm spindle speed, 200 mm.min-1 feed rate and 1000 lip angle of twist drill were found to be optimum (Gupta et al., 2022). Bhat et al. examined the effect of sea water aging, spindle speed, feed and section thickness in composites in terms of peeling and delamination. The most damage was obtained in samples immersed in seawater for 60 days (Bhat et al., 2022). Singh et al. experimentally observed laser drilling of 1.2 mm thick Kevlar-29 composite plate in terms of hole circularity and taper. Optimum parameters were obtained with the Taguchi method by measuring the circularity and taper of each hole opened. It was determined that the laser lamp current was the most important parameter for circularity and air pressure was the most important parameter for taper (Pratap Singh et al., 2022). Bharath et al. carried out drilling tests to evaluate the thrust force and delamination of hybrid composites using the Taguchi method and evaluated the results in detail (Bharath et al., 2021). Shard et al. compared normal rotational drilling of composites with ultrasonic drilling in terms of cutting force, temperature, torque, surface roughness, metal removal rate and micro cracks. When evaluated in terms of smooth surface quality, ultrasonic drilling provided a quality surface (Shard et al., 2023). Zhang et al. conventional and ultrasonic drilling were compared in terms of the effect of change of spindle and feed on hole quality.The results showed that ultrasonic hole drilling improved

hole quality and reduced delamination and burr formation (ZHANG et al., 2023). Mustari et al. examined the effect of glass fiber content in composites on the surface quality during drilling, parameters such as delamination and taper angle. The results showed that a lower glass fiber ratio gave better results (Mustari et al., 2020). Ramesh et al. investigated the effects of drill rotation, feed speed and drill coating material parameter changes on the surface quality of nano granite powder added glass fiber composites. They found that the effect of drill coating material on surface quality is the most important parameter (Ramesh et al., 2022). Saravanakumar and Aravindanath Reddy examined the thrust forces and surface roughness during drilling of composites produced from epoxy resin and Snake Grass Fibers by hand lay-up method. Optimum thrust force was achieved at 1800 rpm spindle speed, 10 mm.min-1 feed and 4 mm drill diameter (Saravanakumar et al., 2022). Maleki et al. applied three different non-destructive methods to examine the delaminations that occur when drilling natural fiber reinforced composites. Of the methods used, the flatbed scanner is the most ideal because it is accessible, economical and efficient (Maleki et al., 2022). Bhadra and Ranjan Dhar examined the roundness error, delamination and hole quality during drilling of Sisal and Coir fiber reinforced CNT-based nanocomposites produced by hand lay-up technique. The best hole quality, surface roughness and low delamination were obtained in the nanocomposite produced hybridly from Sisal and Coir fiber (Bhadra et al., 2022).

In this study, composite plate was produced from carbon-aramid hybrid fabric with epoxy resin by vacuum infusion method. In addition, composite plate production was made by adding wt.1-2% nano graphene and CNT to the resin. Holes were drilled in the produced composite plates with a Ø3 mm HSS drill bit at a rotation speed of 500-1000-2000 rpm and a feed rate of 50-100-200 mm.min-1 on a CNC milling machine and the damage status of these holes was examined.

EXPERIMENTAL APPROACH

In this study, 4 layers of carbon-aramid hybrid fabric at 0o orientation angle, undoped, wt.1-2% CNT and graphene doped composite plates were produced. Then, these produced samples were drilled with HSS drill bits at different rotation and feed rates on the CNC milling machine. Thus, the effect of adding nanoparticles to the composite plates on the hole drilling was observed.

In the study, composite plate production was carried out using vacuum infusion method. Carbon-aramid 200 g.m-2 hybrid fabric was used as reinforcement element. F1564 epoxy resin and the appropriate F3487 hardener set were used as matrix. Hybrid fabrics and epoxy set were supplied by Kompozitsan, IZMIR, Turkey. The matrix set consisting of epoxy resin+hardener and carbon-aramid hybrid fabric used are given in Figure 1.

Figure 1: Hybrid carbon-aramid fabric and epoxy resin+hardener set used in composite plate production.

In addition, since the effect of nanoparticle addition was desired to be observed in the study, wt.1-2% CNT and graphene were poured into the matrix by hand, and then the doped sample was produced by exposing it to vacuum process. The CNT and graphene properties used are given in Table 1. Production phases were carried out by following the practices of Erdem et al. (Erdem et al., 2023).

Table 1: Properties of nano-sized CNT and graphene added into the resin.

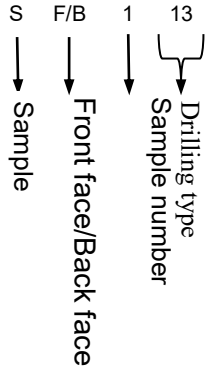
Name of the product	**CARBON NANO TUBE (MWCNT)**	**GRAPHENE**
Purity	> 95%	> 99.9%
Color	Black	Grey
Outer diameter	30-50 nm	7 nm
Inner diameter	5-10 nm	-
Length	10-25 nm	-
Production method	Chemical evaporation	-
Specific surface area	50 $m^2 \cdot g^{-1}$	170 $m^2.g^{-1}$

Five different samples were produced and numbered.
Sample 1 (S1): Non-additive
Sample 2 (S2): %1 Grafen additive
Sample 3 (S3): %2 Grafen additive

Sample 4 (S4): %1 CNT additive
Sample 5 (S5): %2 CNT additive

In order to observe the parameter changes in the produced samples, the samples were coded.

Sample coding:



Holes were drilled in the produced composite plates with a Ø3 mm HSS drill bit on a Microcut brand CNC milling machine at 500-1000-2000 rpm rotation speeds and 50-100-200 mm.min-1 feed rates (Figure 2). The rotation speed and feed rate applied to the test specimens while drilling holes on the CNC milling machine are given in Table 2.

Table 2: Drilling types of composite plates.

Drilling type	**Rotatıon speed (rpm)**	**Feedrate (mm.min^{-1})**
1.1	500	50
1.2	500	100
1.3	500	200
2.1	1000	50
2.2	1000	100
2.3	1000	200
3.1	2000	50
3.2	2000	100
3.3	2000	200

Figure 2: Microcut CNC milling machine and Ø3 mm diameter drill bit for drilling composite plates.

Figure 2: Microcut CNC milling machine and Ø3 mm diameter drill bit for drilling composite plates.

Holes were drilled with a Ø3 mm HSS drill bit at a rotation speed of 500-1000-2000 rpm and a feed rate of 50-100-200 mm.min-1 on the samples taken from the plates produced by the vacuum infusion method. The images of the samples were examined under an optical microscope, the surface where the drill enters is the front surface and the surface where the drill exits is the back surface and damage analysis was made by measuring the damaged area with the help of caliper. In the form of a table, two circles are drawn in yellow, intertwined. While the inner circle is the drill diameter of Ø3 mm, the outer circle represents the damaged area obtained differently in each sample.

RESULTS AND DISCUSSION

In this study, 4 layers of carbon-aramid hybrid fabric at 0o orientation angle, non-additive, wt.1-2% CNT and Graphene additived composite plates were produced. Then, the effect of adding nanoparticles to the composite plates on the drilling of holes was observed by drilling holes with an HSS drill bit at different rotation and feed rates on the CNC milling machine. Fuzzing, burrs and crack damage occur in drilled samples (Figure 3). Therefore, the damages caused by the change of drilling parameters were investigated.

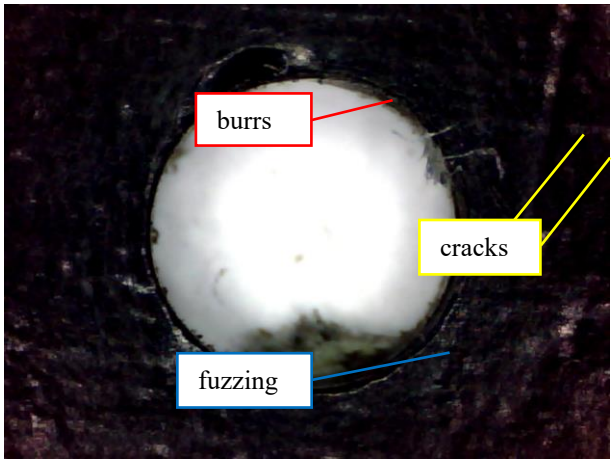


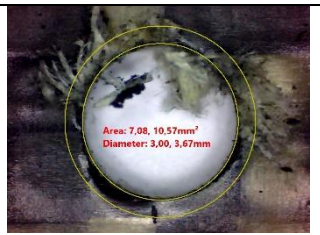
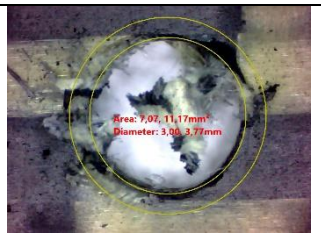
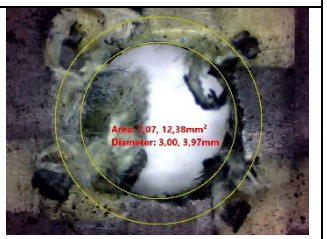
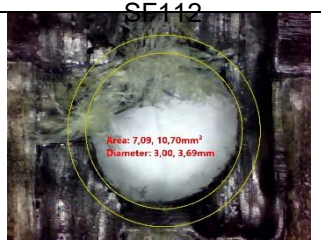
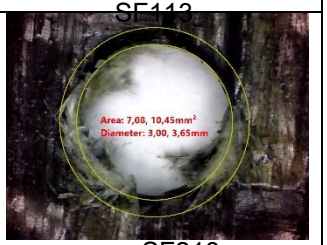
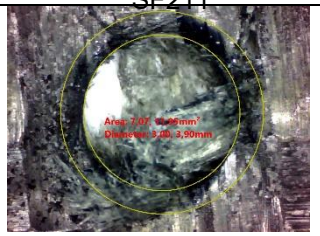
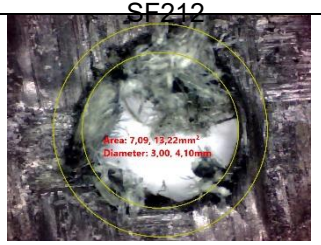
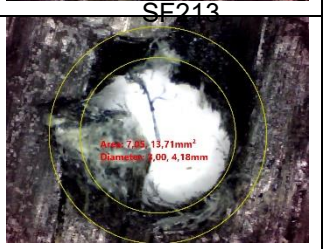
Figure 3: Hole damage occurring in the material when drilling composite plates.

Effect of feedrate at 500 rpm rotational speed:

Holes were drilled with a Ø3 mm HSS drill bit at 500 rpm rotation speed and 50-100-200 mm.min-1 feed rates on the samples taken from the produced plates. The image formed by the drilled holes on the front surface of the sample is given in Table 3, and the image formed on the back surface of the sample is given in Table 4.

The composite plate produced is made of carbon-aramid hybrid fabric. Therefore, when Table 3 and Table 4 are examined, the regions where the drilled holes coincide with carbon fibre (black colored) in some places and aramid fibre (yellow colored) in other places (S3). Visual fuzzing in the carbon-dense regions was much less than in the aramid-intensity region (S5). The cutting quality decreased as the feed rate increased at 500 rpm cutting speed for the front surface and the back surface, and the addition of CNT to the fabric increased the cutting quality (Table 3-Table 4).

Table 3: Hole damages occurred which in the front surface at 500 rpm rotation speed, 50-100-200 mm.min^{-1} feed rate.

500 rpm	Feedrate (mm.min^{-1})		
Sample number	50	100	200
S1	 SF111	 SF112	 SF113
S2	SF211	 SF212	 SF213
S3	 SF311	 SF312	 SF313

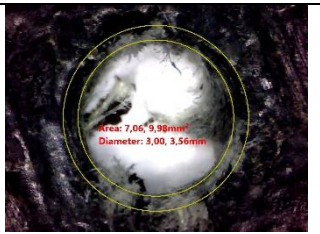
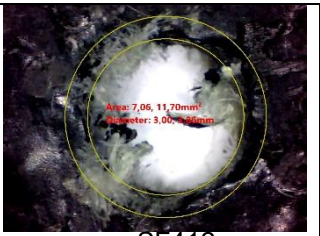
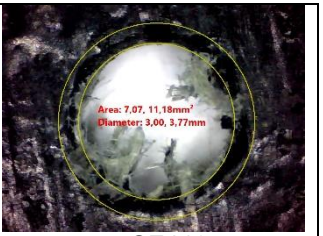
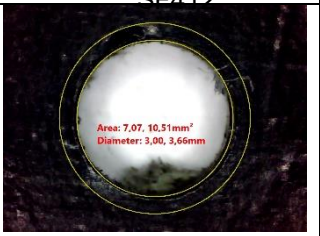
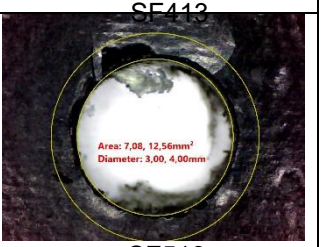
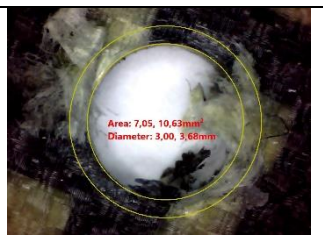
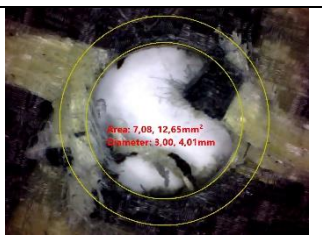
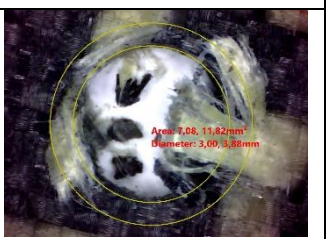
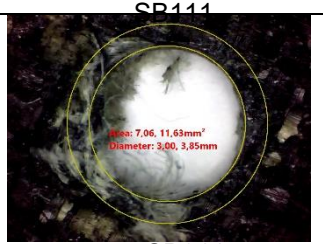
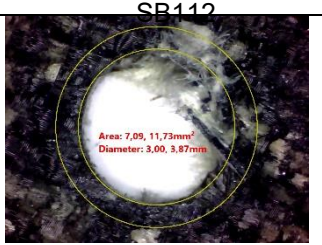
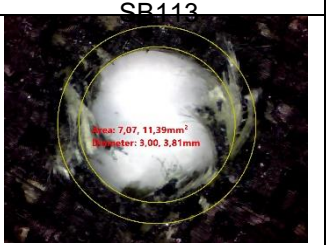
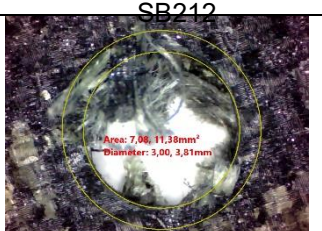
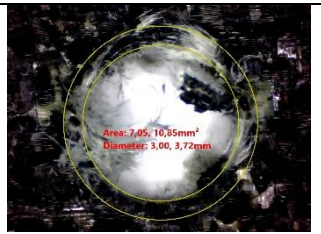
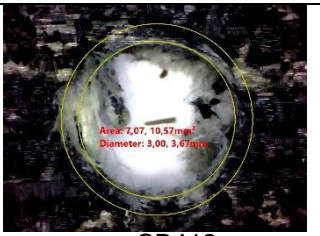
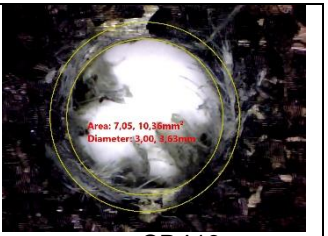
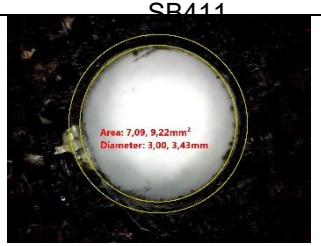
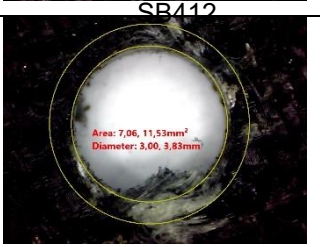
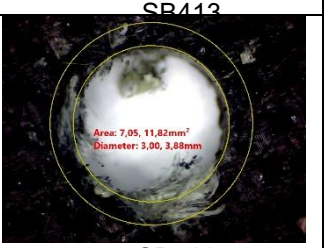
S4	 SF411	 SF412	 SF413
S5	SF511	 SF512	 SF513

Table 4: Hole damages occurred which in the back surface at 500 rpm rotation speed, 50-100-200 mm.min^{-1} feed rate.

500 rpm	Feedrate (mm/d)		
Sample number	50	100	200
S1	 SB111	 SB112	 SB113
S2	 SB211	 SB212	 SB213
S3	SB311	 SB312	SB313

S4	 SB411	 SB412	 SB413
S5	 SB511	 SB512	 SB513

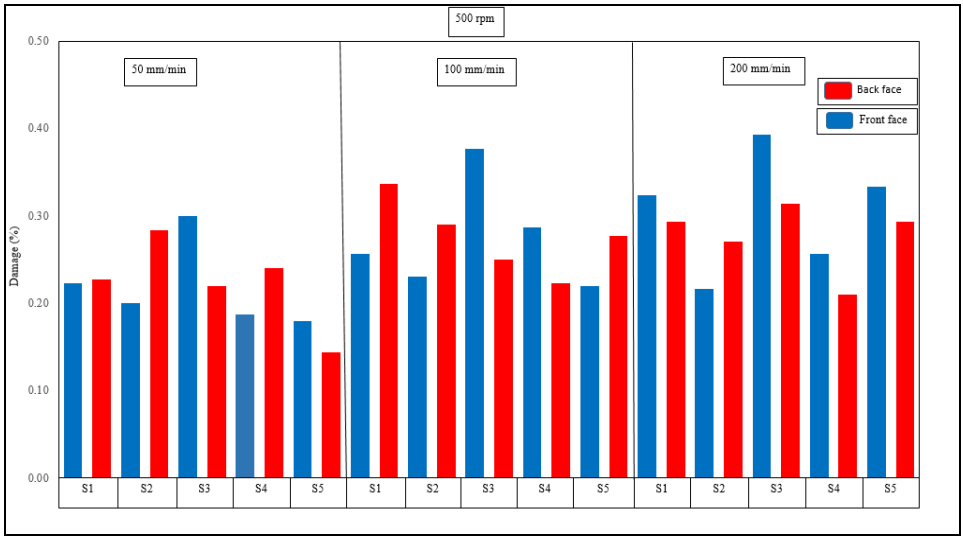


Figure 4: Hole damage rates occurred which in the front surface and back surface at 500 rpm rotation speed, 50-100-200 mm.min^{-1} feed rate.

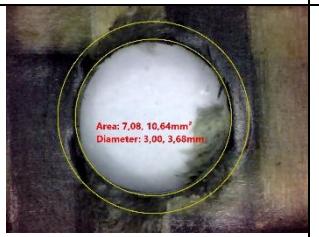
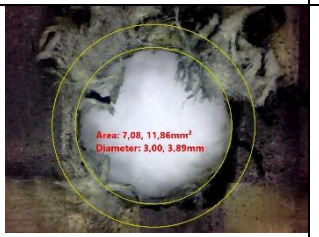
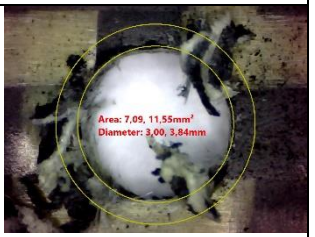
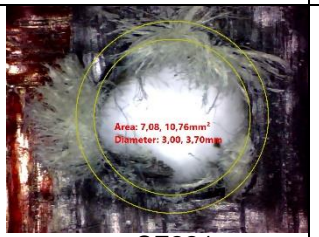
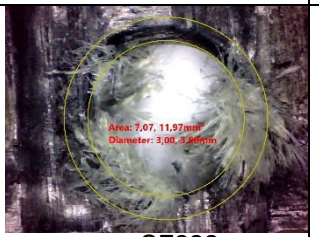
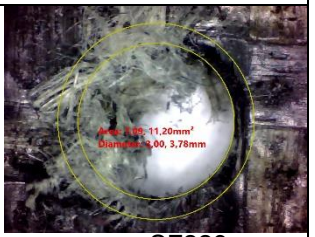
When the damage rate in Figure 4 is examined for the front surface where the drill enters; For the non-additive and wt.2% CNT and graphene additived samples, the feed rate increased from 50 mm.min-1 to 200 mm.min-1, while the damage rates increased in each. However, in samples with wt.1% CNT and graphene; While the damage rate increased at 100 mm.min-1 feed rate, there was no significant change in 200 mm.min-1 feed rate. When the damage rate is examined for the rear surface; While it increased with the feed rate in wt.2% CNT and graphene doped samples, it did not show a significant trend in other samples. Increasing the nano graphene particle ratio for each feed rate increased the damage rate. However, the addition of CNT decreased the damage rate for 50-100 mm.min-1 feed rate and increased it for 200 mm.min-1 feed rate.

Effect of feedrate at 1000 rpm rotational speed:

Holes were drilled with Ø3 mm HSS drill bit at 1000 rpm rotation speed and 50-100-200 mm.min-1 feed rates. The image formed on the front surface of the sample by the drilled holes is given in Table 5, and the image formed on the back surface of the sample is given in Table 6.

When these holes drilled into the test samples at different rotational and advance speeds are examined; At the points where the hole coincides with the aramid part of the fabric (S2), fuzzing was obtained much more than at the points where it coincides with the carbon fibre part (S5). This can occur even within the same hole (SF122).

Table 5: Hole damages occurred which in the front surface at 1000 rpm rotation speed, 50-100-200 mm.min^{-1} feed rate.

1000 rpm	Feedrate (mm.min^{-1})		
Sample number	50	100	200
S1	 SF121	 SF122	 SF123
S2	 SF221	 SF222	 SF223

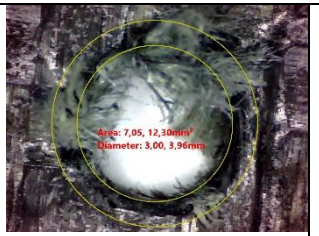
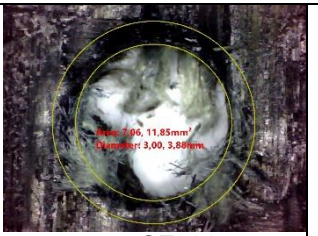
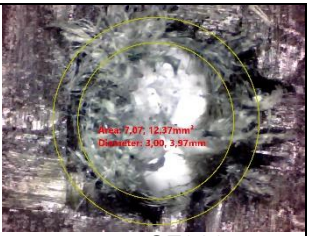
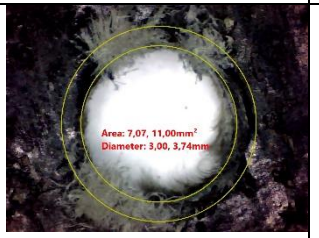
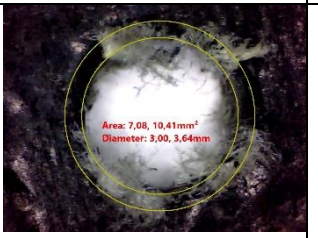
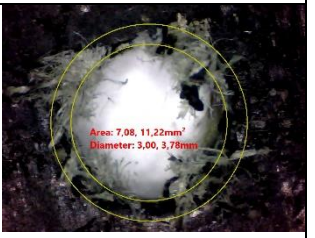
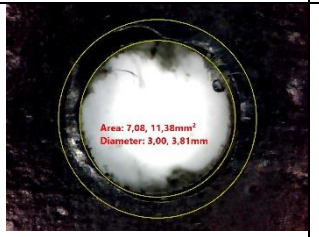
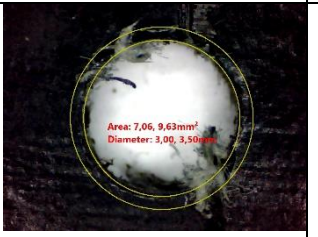
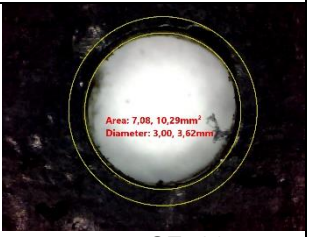
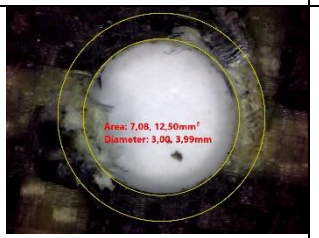
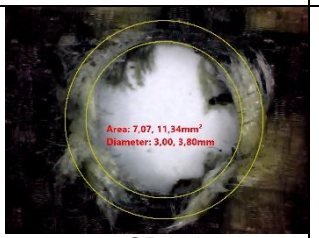
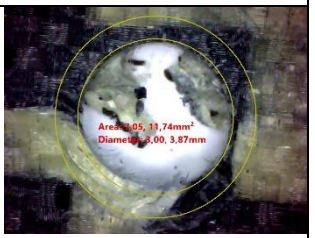
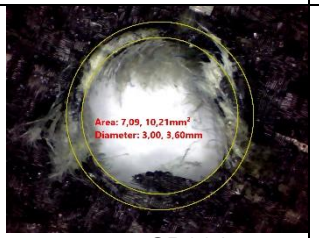
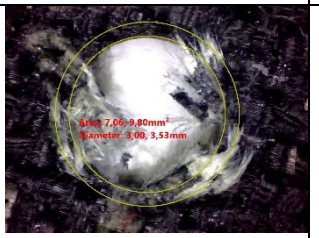
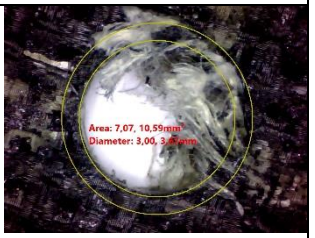
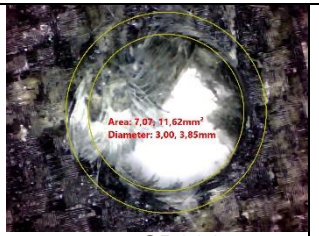
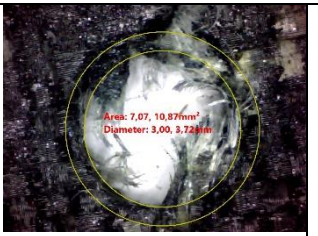
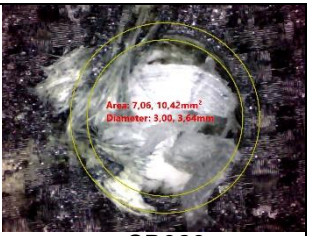
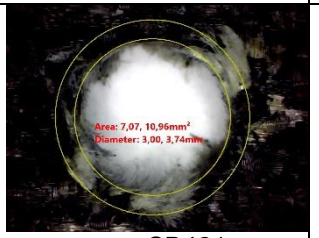
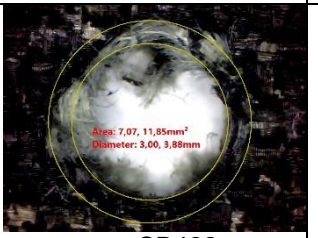
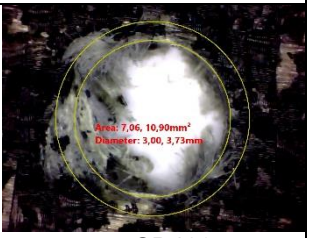
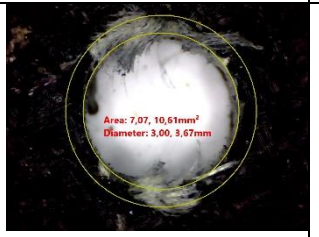
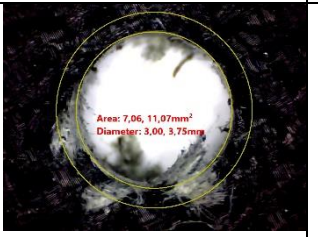
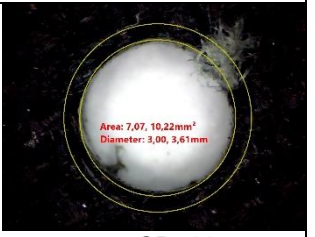
S3	 SF321	 SF322	 SF323
S4	 SF421	 SF422	 SF423
S5	 SF521	 SF522	 SF523

Table 6: Hole damages occurred which in the back surface at 1000 rpm rotation speed, 50-100-200 mm.min^{-1} feed rate.

1000 rpm	Feedrate (mm.min^{-1})		
Sample number	50	100	200
S1	 SB121	 SB122	 SB123
S2	 SB221	 SB222	 SB223

S3	 SB321	 SB322	 SB323
S4	 SB421	 SB422	 SB423
S5	 SB521	 SB522	 SB523

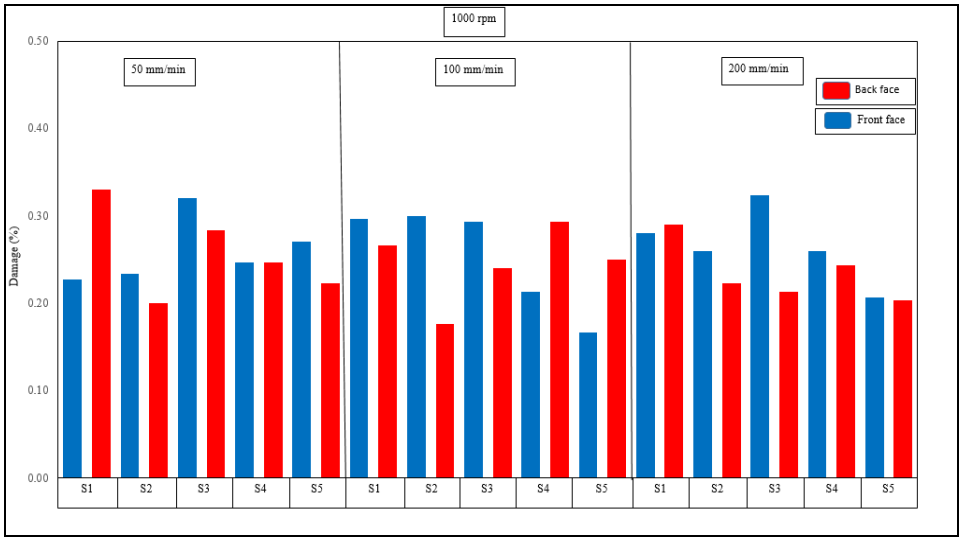


Figure 5: Hole damage rates occurred which in the front surface and back surface at 1000 rpm rotation speed, 50-100-200 mm.min^{-1} feed rate.

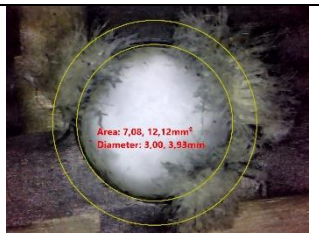
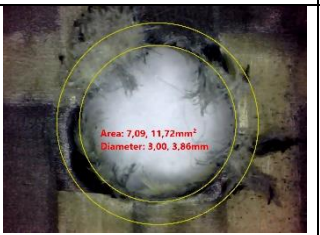
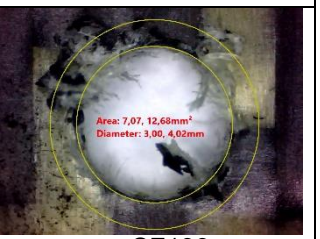
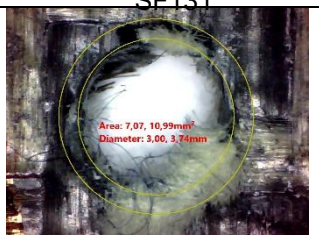
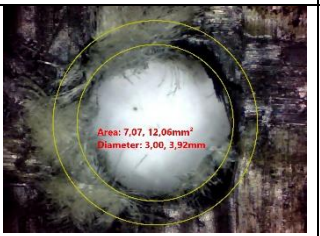
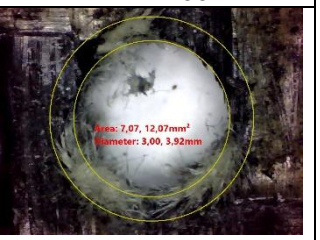
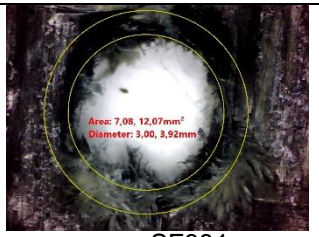
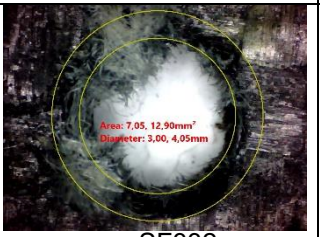
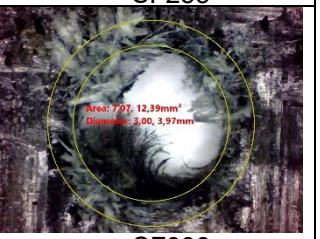
In Figure 5, wt.1% and wt.2% CNT and graphene additives increased the damage rate when the front face damage was taken into account for 1000 rpm rotation and 50 mm.min-1 feed rate. For 100 mm.min-1 feed rate, 1-2% graphene additive did not significantly affect the damage rate, while CNT additive reduced the damage rate linearly. Linear reduction of damage by CNT additive was achieved within 200 mm.min-1 feed. While wt.1% graphene additive decreased the damage rate at 200 mm.min-1 feed rate, the wt.2% additive increased the damage rate. However, wt.1-2% CNT contribution for the same progression significantly reduced the damage rate.

Effect of feedrate at 2000 rpm rotational speed:

The front surface view of the holes drilled at 2000 rpm speed with Ø3 mm HSS drill bit at 50-100-200 mm.min-1 feed speed is given in Table 7 and the back surface image is given in Table 8.

Due to the characteristic of the aramid fabric, the fibers do not break during drilling and fuzzing occurs. When the images are examined, the fuzzing on the front surface is higher in that sample, on the back surface (S2).

Table 7: Hole damages occurred which in the front surface at 2000 rpm rotation speed, 50-100-200 mm.min^{-1} feed rate.

2000 rpm	Feedrate (mm/d)		
Sample number	50	100	200
S1	 SF131	 SF132	 SF133
S2	 SF231	 SF232	 SF233
S3	 SF331	 SF332	 SF333

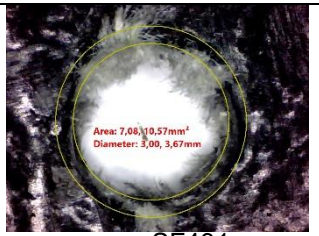
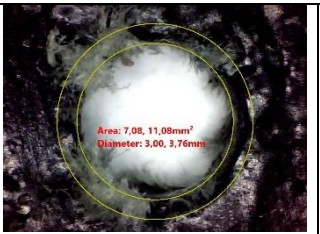
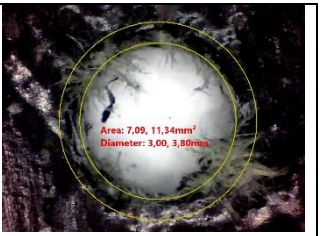
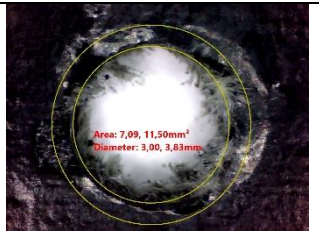
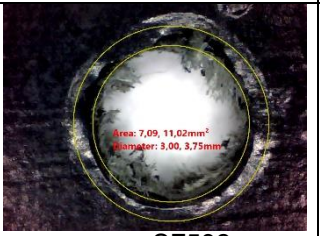
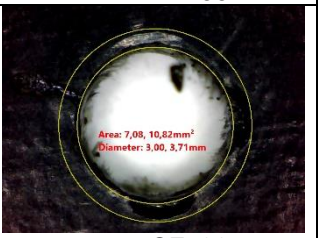
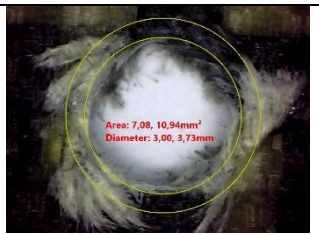
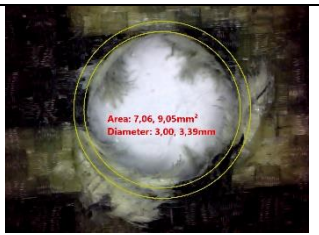
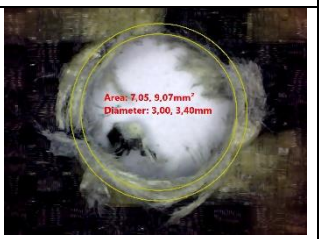
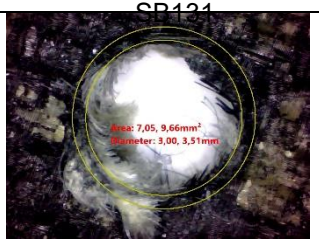
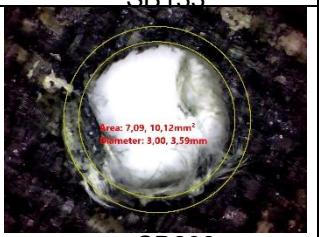
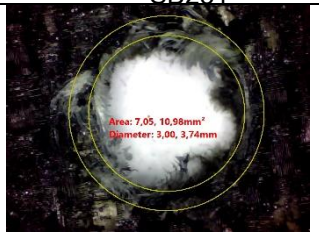
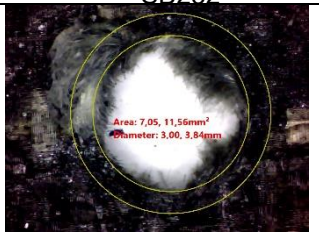
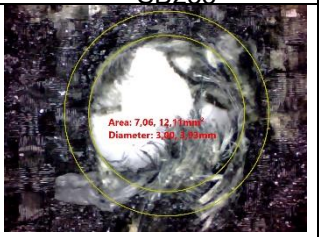
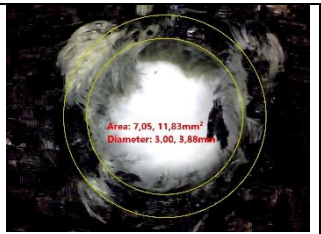
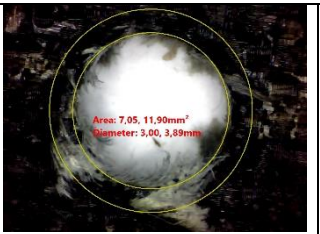
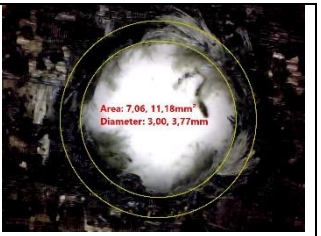
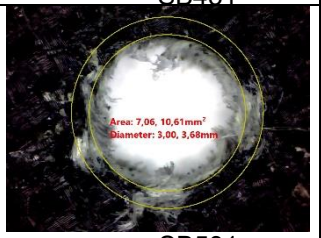
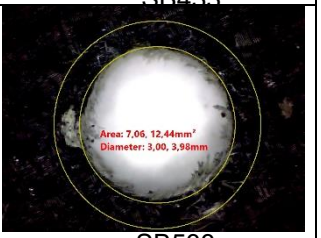
S4	 SF431	 SF432	 SF433
S5	 SF531	 SF532	 SF533

Table 8: Hole damages occurred which in the back surface at 2000 rpm rotation speed, 50-100-200 mm.min^{-1} feed rate.

2000 rpm	Feedrate (mm/d)		
Sample number	50	100	200
S1	 SB131	 SB132	 SB133
S2	 SB231	SB232	 SB233
S3	 SB331	 SB332	 SB333

S4	 SB431	 SB432	 SB433
S5	 SB531	SB532	 SB533

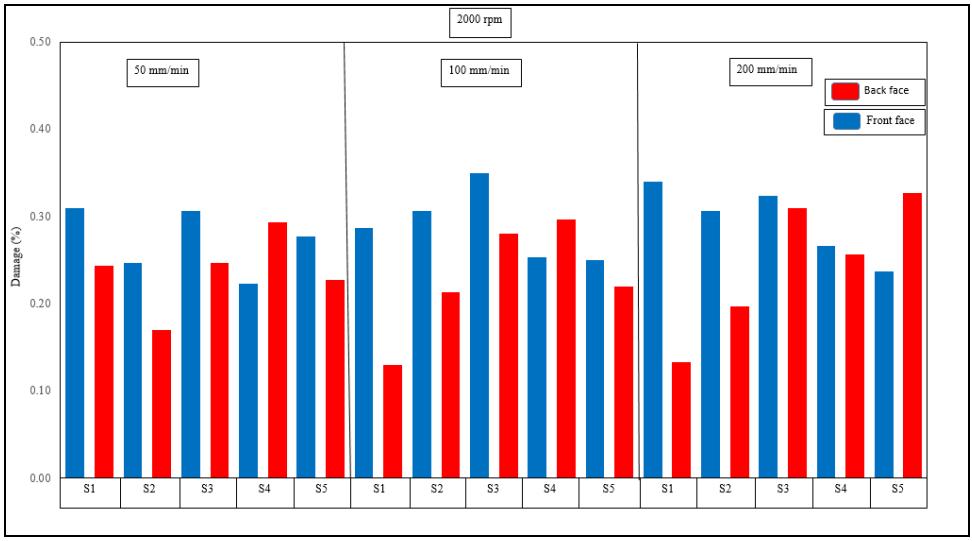


Figure 6: Hole damage rates occurred which in the front surface and back surface at 2000 rpm rotation speed, 50-100-200 mm.min^{-1} feed rate.

When Figure 6 is examined, the addition of wt.2% graphene particles at 2000 rpm at every 50-100-200 mm.min-1 feed rate increased the damage rate compared to wt.1%. The same result is valid for CNT for 500 rpm. However, adding wt.2% CNT for 100-200 mm/min feed reduced the damage rate compared to wt.1%.

CONCLUSIONS

In order for vehicles and structures to come together and form mechanisms, they must be jointed to each other. One of the most basic of these joints is mechanical joints. In order to joint the parts mechanically, holes must be drilled on them. The effects of parameters are known when drilling holes in isotropic materials.

The study is based on the determination of the damage rate by examining the cracks formed around the hole after drilling under an optical microscope. No significant trend effect of drilling parameters was observed in fiber-reinforced composite plates. This can be explained by the fact that

the drill coincides with the fiber or matrix during drilling. This suggests that it also prevents the observation of the effect of the nanoparticle additive.

- ✓ Cracking, fuzzing, burrs and delamination faults are common defects in drilling holes in fiber-reinforced composites and need to be controlled, which affects hole quality. Controlling and reducing these errors will increase the life of holed samples for the next stage of use.
- ✓ The most fuzzing damage was obtained in aramid fiber reinforced composites. In cases where fuzzing damage is not desired, the use of carbon fiber reinforced composite can be recommended.
- ✓ The addition of CNT to the fiber element reduced fuzzing damage, while the addition of graphene did not make any significant changes.
- ✓ Drilling holes is difficult because fiber-reinforced composite materials are not homogeneous and anisotropic. Therefore, preliminary experiments and optimization are often required before mass production.
- ✓ The lower cutting feed rate at each cutting speed increased the cutting quality.
- ✓ The effect of variation of rotational speed and feed on hole quality could not be obtained in a certain trend. Similarly, the addition of CNT and graphene to the composite did not show a single trend. Therefore, the designer should choose the parameters that are suitable for him and make his design that way.

REFERENCES

Ahuja, B., Johri, N., Chandra Kandpal, B., Kumar Singh, L., & Parkash Singh, P. (2023). Drilling process parameter optimization of natural fibre reinforced polymer matrix composites. *Materials Today: Proceedings*, *xxxx*. doi: 10.1016/j.matpr.2023.02.364

Benyettou, R., Amroune, S., Slamani, M., Seki, Y., Dufresne, A., Jawaid, M., & Alamery, S. (2022). Assessment of induced delamination drilling of natural fiber reinforced composites: A statistical analysis. *Journal of Materials Research and Technology*, *21*, 131–152. doi: 10.1016/j.jmrt.2022.08.161

Bhadra, D., & Ranjan Dhar, N. (2022). Study of the delamination factor and taper angle in drilling of natural fiber reinforced epoxy nanocomposite materials.

Materials Today: Proceedings, *60*, 686–693. doi: 10.1016/j.matpr.2022.02.318

Bharath, K. N., Roopa, D., Indran, S., Basavarajappa, S., Sanjay, M. R., & Siengchin, S. (2021). Influence of the stacking sequence and coconut husk micro fillers on the drilling parameters of coconut leaf sheath/glass/jute fiber hybrid phenol formaldehyde composites. *Materials Today: Proceedings*, *52*, 2427–2431. doi: 10.1016/j.matpr.2021.10.422

Bhat, R., Mohan, N., & Sharma, S. (2022). Effect of seawater immersion on the drill induced delamination failure in glass fiber reinforced polymer composites – Experimental and statistical analysis. *Engineering Failure Analysis*, *142*(September). doi: 10.1016/j.engfailanal.2022.106803

Chrispin Das, M., Prathap Singh, S., Rangarajan, V., & Prabhuram, T. (2023). Characterization of pineapple leaf fiber, areca fiber and egg shell powder reinforced phenolic resin composites and finding optimal parameters for sustainable machining. *Materials Today: Proceedings*, *xxxx*. doi: 10.1016/j.matpr.2023.04.023

Erdem, S., Gur, M., & Kaman, M. O. (2023). Nanoparticle effects on post-buckling behaviour of patched hybrid composites. *Materialpruefung/Materials Testing*, *65*(1), 111–123. doi: 10.1515/mt-2022-0223

Geier, N., Póka, G., Jacsó, Á., & Pereszlai, C. (2022). A method to predict drilling-induced burr occurrence in chopped carbon fibre reinforced polymer (CFRP) composites based on digital image processing. *Composites Part B: Engineering*, *242*(December 2021). doi: 10.1016/j.compositesb.2022.110054

Gupta, A., Vaishya, R., Kumar, R., Khan, K. L. A., Chhabra, S., Singh Verma, A., & Bharadwaj, A. (2022). Effect of drilling process parameters on delamination factor in drilling of pultruded glass fiber reinforced polymer composite. *Materials Today: Proceedings*, *64*, 1290–1294. doi: 10.1016/j.matpr.2022.04.110

Maleki, H. R., Abazadeh, B., Arao, Y., & Kubouchi, M. (2022). Selection of an appropriate non-destructive testing method for evaluating drilling-induced delamination in natural fiber composites. *NDT and E International*, *126*(September 2021). doi: 10.1016/j.ndteint.2021.102567

Mathew, S., & Shunmugesh, K. (2023). Application of grey-fuzzy logic for the optimization of drilling parameters for coir fibre reinforced composite. *Materials Today: Proceedings*, *72*, 2082–2088. doi: 10.1016/j.matpr.2022.08.144

Mustari, A., Chakma, P., Sobhan, R., & Dhar, N. R. (2020). Investigation on roundness and taper of holes in drilling GFRP composites with variable weight percentages of glass fiber. *Materials Today: Proceedings*, *38*, 2578–2583. doi: 10.1016/j.matpr.2020.07.685

Nagaraj, A., Uysal, A., Gururaja, S., & Jawahir, I. S. (2022). Analysis of surface integrity in drilling carbon fiber reinforced polymer composite material under various cooling/lubricating conditions. *Journal of Manufacturing Processes*, *82*(July), 124–137. doi: 10.1016/j.jmapro.2022.07.065

Pratap Singh, K., Bahl, A., Norkey, G., & Dutt Gautam, G. (2022). Experimental investigation and parametric optimization of the hole-circularity and taper angle during laser drilling kevlar-29 fiber composite. *Materials Today: Proceedings*, *56*, 3325–3329. doi: 10.1016/j.matpr.2021.10.155

Ramesh, B., Sathish Kumar, S., Elsheikh, A. H., Mayakannan, S., Sivakumar, K., &

Duraithilagar, S. (2022). Optimization and experimental analysis of drilling process parameters in radial drilling machine for glass fiber/nano granite particle reinforced epoxy composites. *Materials Today: Proceedings*, *62*, 835–840. doi: 10.1016/j.matpr.2022.04.042

Saravanakumar, A., & Aravindanath Reddy, S. (2022). Optimization of process parameter in drilling of snake grass fiber reinforced composites. *Materials Today: Proceedings*, *62*, 5460–5466. doi: 10.1016/j.matpr.2022.04.144

Shard, A., Agarwal, R., Gupta, V., & Garg, M. P. (2023). Influence of ultrasonic vibrations during drilling of carbon-fiber-reinforced polyetherimide composites. *Applied Acoustics*, *202*. doi: 10.1016/j.apacoust.2022.109163

ZHANG, S., WANG, W., XIONG, Y., HUANG, B., & JIANG, R. (2023). Defect suppression mechanism of ultrasonic vibration-assisted drilling carbon fiber/bismaleimide composite. *Chinese Journal of Aeronautics*, *36*(8), 366–380. doi: 10.1016/j.cja.2022.10.008

Large Language Models: Foundations, Architectures and Applications

Tülay YILDIRIM

Doç. Dr.; Karabük Üniversitesi, Eskipazar Meslek Yüksek Okulu, tulayturan@karabuk.edu.tr, ORCIH No: 0000-0001-5933-9752.

ABSTRACT

Large Language Models (LLMs) have quickly emerged as one of artificial intelligence's most revolutionary technologies, radically altering how machines comprehend, produce, and engage with natural language. LLMs can now execute a wide range of challenging linguistic tasks with previously unheard-of fluency and contextual awareness thanks to recent advances in transformer-based architectures, large-scale self-supervised training, and alignment techniques. It is becoming more and more crucial to have a clear and organized understanding of how these models are created, trained, and used as their use grows in the public, commercial, and scientific spheres. The primary objective of this study is to present a comprehensive yet accessible overview of Large Language Models, bringing together key theoretical foundations and practical insights within a unified framework. The study summarizes the key phases of the training pipeline, such as pre-training, fine-tuning, and alignment with human feedback. It also examines the historical development of language modeling and emphasizes the architectural principles that underpin contemporary LLMs, such as layered transformer structures, attention mechanisms, tokenization, and embedding representations. Furthermore, the work addresses issues of computational cost, bias, sustainability, and responsible use while analyzing the benefits and difficulties of LLM deployment across various application areas. To determine which models are most appropriate for particular tasks and goals, well-known model families like GPT, BERT, and T5 are studied in detail. Lastly, a forward-looking perspective is provided by outlining new research directions, open model ecosystems, and changing regulatory frameworks. All things considered, this study provides readers with a logical road map for comprehending the present situation as well as the future course of LLMs.

Keywords – Large Language Model (LLM), Transformer-Based Models, GPT, BERT, T5.

INTRODUCTION

Large Language Models (LLMs) have become one of the most significant technologies influencing both academic research and practical applications in the rapidly evolving field of artificial intelligence. These models have proven to be remarkably adept at comprehending and producing natural language, allowing them to perform tasks that were previously believed to be unique to humans. These include translating between languages, creating cohesive narratives, reasoning through complex texts, and participating in interactive discourse.

These models are becoming more and more important due to developments in deep learning, the availability of large amounts of textual

data, and the creation of effective training paradigms. LLMs acquire a comprehensive grasp of linguistic structure, semantics, and contextual relationships by utilizing sizable datasets and self-supervised learning techniques. As a result, they are now crucial parts of many different systems, such as scientific analysis procedures, content creation platforms, speech-based AI systems, and decision support tools.

LLMs have created new concerns about scalability, efficiency, interpretability, and responsible use in addition to their technical accomplishments. Although a larger model improves performance significantly, there are now more worries about computational cost, energy consumption, bias, and regulatory compliance. As a result, in addition to creating larger models, current research concentrates on creating more effective structures, creating transparent and open ecosystems, and putting strong governance frameworks in place.

This study is structured to provide a systematic overview of large language models (LLMs) from both theoretical and practical perspectives. First, the conceptual foundations and historical development of LLMs are addressed to provide a backdrop for modern language modeling. Then, the architectural design of LLMs is examined in detail, with particular emphasis on transformer layers, attention mechanisms, and representation learning. Next, the training process of LLMs, encompassing data preparation, pre-training, fine-tuning, and alignment strategies, is summarized. Following this, the advantages and challenges associated with large-scale language models are analyzed, highlighting their impact in various application areas and their technical and ethical limitations. The study also examines prominent LLM families and use cases, addressing the question of which model is best suited for which purpose. Finally, future directions in LLM research are discussed, focusing on the development of more efficient architectures, the rise of open models, and the increasing importance of regulatory and governance frameworks in shaping responsible and sustainable deployment.

THE LARGE LANGUAGE MODEL CONCEPT AND ITS HISTORICAL DEVELOPMENT

Nowaday, the first technology that comes to mind in the field of artificial intelligence is Large Language Models (LLM). These models can write, summarize, answer questions, make code, do scientific research, come up with design ideas, and talk to people in a natural way.

Large Language Models (LLMs) are language models that use deep learning and are trained on huge sets of text. Their main job is to guess what the next word in a sequence will be based on probability. This lineage has changed over time from n-gram and statistical language models to word embeddings (Word2Vec, GloVe), and finally to modern LLMs with the transformer architecture (Attention Is All You Need) in 2017 (Vaswani et al., 2017). Models like the GPT series, BERT family, and others, with their self-supervised pre-training and fine-tuning approach, have made the LLM concept widespread and visible since the 2020s (Wang et al., 2025, Zhao et al., 2025).

Mathematically, a language model learns the probability distribution P (w_t | w_1, w_2, ..., w_{t-1}); that is, it calculates the probability of the next word by looking at the previous words. The basic principle of this modelling is quite simple: to predict what will follow a given word. For example:

Input**:** "The sun rises in the east..."

Model Prediction**:** "rises" (high probability), "sets" (low probability).

This simple task reveals surprising capabilities when applied on a sufficiently large scale.

There are three main reasons why LLMs are called "big":

(i) Number of Parameters**:** Modern LLMs contain billions, even trillions, of parameters. GPT-3 has 175 billion, and GPT-4 has an estimated 1.8 trillion parameters.

(ii) Training Data: These models are trained on terabytes of text data that make up a significant portion of the internet books, articles, websites, code repositories.

(iii) Computational Power: Training an LLM requires months of computation with thousands of GPUs and costs millions of dollars.

Theoretical Foundations of Large Language Models

Probabilistic language modeling is the basis for large language models (LLMs). The goal is to find out how likely it is that a certain sequence of words will happen. At each step, the model answers the question, "Which word is most likely to come next?". Mathematically, this is done using the chain rule:

$$P(w_1, \ldots, w_n) = \prod_{i=1}^{n} P(w_i \mid w_1, \ldots, w_{i-1}).$$

In the past, early language models used the n-gram method. Only the words that came before it affected the probability of each word. After that, they were estimated using frequency counts and smoothing methods like Kneser–Ney. (Chen et al., 1999). While these methods were straightforward

and explicable, they were insufficient for modeling extensive, complex languages because they could not leverage long contexts and faced data sparsity challenges. (Swaileh, 2017).

Language models based on neural networks have been around since the 2000s. Words were represented by high-dimensional "embedding" vectors and learned through multilayer networks. RNN, LSTM, and GRU are examples of structures that used hidden states to carry contextual information over time. This made the data less sparse and made it possible to find semantic similarities. But gradient extinction and sequential computation in long sequences made it hard to scale.

The fast progress in this field reached its peak in 2018 with the release of "Attention Is All You Need," a groundbreaking study by eight researchers at Google that is now considered a landmark in the machine learning literature (Vaswani et al., 2017). This work introduced the **Transformer** architecture, based entirely on attention mechanisms, as an alternative to the recurrent and convolutional networks long used in language modelling and natural language processing. This new kind of architecture has made it possible to better understand and generate language by better capturing long-range dependencies in text, allowing for parallel computation, and learning with high accuracy on large-scale data. The Transformer method, which is the basis for large language models like the GPT series, BERT, and their offshoots, has been very important in the rise and widespread use of modern LLMs (Wang et al., 2025, Ji et al., 2025).

LLM ARCHITECTURE: LAYERS, ATTENTION MECHANISM AND REPRESENTATIONS

Large Language Models (LLMs) are built on Transformer architecture, which can overcome the limitations of models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks. Transformers process input sequences in parallel, unlike previous models, enabling effective training for lengthy texts.

Layer Structure

Modern LLMs include dozens or even hundreds of transformer layers that repeat. Every layer gets the representation from the layer before it, works on it, and sends it on to the next layer. A typical transformer layer has a multi-headed self-attention sublayer, a position-based feedforward

network, layer normalization, and residual links that all function together (Beggag et al., 2025).

This layering allows upper layers to capture more complex linguistic phenomena, while lower layers focus more on syntax and local patterns. As the number of parameters and layer depth increase, the model's expressive power and the variety of patterns it can learn also increase.

Attention Mechanism

Attention is a weighting mechanism that figures out how each token in a series relates to the others. It is at the center of the Transformer design. We can mention two kinds of attention mechanisms:

In self-attention, each token figures out a weighted average of all the other tokens in the input sequence. These weights show how relevant one token is to another in a certain context. Because of this, each token's representation becomes context-aware, and changes based on the words around it (Gurgurov et al., 2024).

Large language models utilize multi-head attention, enabling the model to focus on various aspects of the input simultaneously. Each attention head targets specific linguistic features, including syntactic structure, semantic roles, and contextual emphasis. The outputs of these heads are subsequently integrated to create a more sophisticated and comprehensive representation (Voita et al., 2019).

The Concept of Tokens

Large language models (LLMs) analyze text not by individual words, but by smaller components known as tokens. A token may represent an entire word, a fragment of a word, or a single character, depending upon the tokenizer employed by the model. The model efficiently represents common words and processes rare or new words by decomposing them into component parts.

Upon user input, the text is initially decomposed into tokens, and all model calculations are executed based on this token sequence. The maximum context length denotes the highest number of tokens that can be processed simultaneously.

The Concept of Embedding

Tokens are digital identifiers; for the model to do significant computations with them, each token must be represented by a fixed-dimensional vector, a process known as embedding. The embedding layer associates each token identity with a high-dimensional vector, positioning functionally comparable tokens in proximity throughout this vector space.

Language models usually include two predominant forms of embeddings: token embeddings and positional embeddings. Token embedding applies to the identification of the specific word or subunit,

whereas positional embedding deals to its location within the sequence. The final representation is derived from these components and serves as input to the transformer layers (Bingi et al. 2023, Zhang et al., 2024).

EDUCATION PROCESS

The LLM education process consists of several basic stages:

Data Collection and Preparation: The first step is to collect a very large and diverse text dataset for the model to learn from. Pages on the internet, books, news, and conversations are usually considered resources. Unwanted content such as spam and personal data is extracted from this data, formats are combined, and the text is converted into a format that the model can understand (e.g., token sequences).

Pre-training: During pre-training, the model acquires the ability to predict missing tokens within extensive text databases, primarily through unsupervised or semi-supervised methods. The objective is to develop an internal representation that encapsulates the statistical structure, grammar, and overarching world knowledge of the language. This phase involves the highest cost and requires the most time.

Fine-tuning and task adaptation: Following pre-training, smaller, target-specific datasets are used to refine the model. For instance, question-and-answer formats, dialogues, or labeled data relevant to specific fields such as law or medicine are utilized. This guarantees the continuation of general language skills while generating output in a more uniform format and with increased accuracy in the specified tasks.

Human Feedback and Alignment: Present LLMs use reinforcement learning (e.g., RLHF) to incorporate human ratings and rankings of the model's answers. During this phase of alignment, we want to make sure the model follows user expectations for safe, harmless behavior and not just gives "close to correct" answers.

ADVANTAGES AND CHALLENGES

Advantages

LLMs learn the small connections and patterns between words by reading a lot of text. This lets them write texts that sound like human language and are fluent and coherent. They can write content that fits the situation in many different styles, like long descriptions, summaries, anecdotes, technical writings, or formal letters.

They can follow the context by examining previous parts of the text, appropriately connect pronouns, and structure their responses accordingly while keeping conceptual coherence. Their enhanced knowledge of user intent enables them to both interpret questions accurately and deliver more useful and focused answers.

Once LLMs have learned the basics of language, they can be used in a variety of disciplines, such as health, finance, law, and education, with more data. They are general-purpose tools that can learn more about a given field, which makes them versatile. You can instantly adjust to completely new situations by only altering the data and settings, as long as you use the same scheme.

By doing so, companies may make solutions that can be changed to fit their needs, all based on one LLM-based core, instead of having separate systems for each department. This kind of strategy lowers maintenance costs and speeds up the implementation of the solution as new needs come up. It has a wide range of applications:

Scientific research and knowledge discovery: It makes it easier to discover researchers by doing things like summarizing scientific articles, finding comparable studies in the literature, and pulling out the most important results from long reports. It might help come up with first ideas for patterns, hypotheses, or research questions by putting together text data from multiple sources.

Health: In healthcare, LLMs can alleviate the documentation stress for professionals by efficiently summarizing texts, including medical notes, lab result reports, and patient histories. They can emphasize critical insights from unorganized content, contributing to clinical decision support systems.

They may also be utilized to produce instructive texts for patient communication, appointment reminders, and clear descriptions of the treatment procedure. Well-structured and monitored procedures enable healthcare practitioners to concentrate on more intricate circumstances by supplying preliminary responses to patients' inquiries.

Human Resources and Office Automation: From an HR point of view, it makes faster resume screening, basic question generation, job ad text writing, and performance feedback draft creation. It can reply to emails, take notes during meetings, and write reports automatically, which lets employees spend their time on more important things.

Benefits in Finance and Business: In finance, LLMs can help analysts and managers get a rapid "pre-reading" by summarizing long reports, market analysis, and contract texts. By looking at how customers interact with them and what they say, they may learn about how people see risk, how happy they are, and where problems keep coming up.

Law, Public and Policy: In the legal area, it saves a lot of time on things like writing agreements, summarizing long legal documents, and finding the right article faster. In public institutions, it can help cut down on paperwork by helping to organize and summarize requests, applications, and minutes, and by making policy texts easier to understand for different groups of people.

Marketing, Media, and Creative Industries: It speeds up the process of making marketing materials and content by coming up with texts, slogans, campaign ideas, and versions that work on numerous platforms. In the media and creative industries, it helps teams come up with ideas and plans by writing draft screenplays, news articles, product descriptions, and social media posts.

Challenges

Computing and Energy Requirements: Large language models need a lot of computational power, storage space, and energy to run both while they are being trained and when they are being used. Costs go up quickly, especially when processing big documents of text. This is because memory and processing burdens go up, which could make it harder for smaller firms to get access.

This use of a lot of resources also has a big carbon footprint, which is bad for the environment. To keep the LLM ecosystem going, it is now necessary to work on better hardware, compression methods, and more energy-efficient architectures.

Ethical Concerns and Misinformation: LLMs can learn and reproduce biases and stereotypes from the datasets they are trained on. This means that they could provide outputs that are unjust, discriminating, or exclusionary toward some groups. So, finding and getting rid of bias should be a constant part of choosing data, cleaning it up, and checking it after the fact.

Also, models might make content that is very convincing yet wrong or made up; this can help spread false information, fake news, phishing emails, and other harmful materials. Many nations have various restrictions about what information and expression are allowed, which makes it hard for businesses to manage these platforms, both legally and technically. This also makes it hard for enterprises to create trust with users.

Update and Version Management: LLMs need to be up to current because the world and knowledge change quickly. But retraining or retuning big models costs a lot of money. This always makes people wonder things

like "How latest is the model?" and "What date does it have information for?"

Also, when a new version comes out, it might be hard to keep things consistent with the old version, make backward integrations, and get used to new ways of doing things. Every update to a release needs to be retested and checked for quality, security, and the risk of unexpected behavior.

User Dependence and Skill Erosion: Because LLMs make writing, summarizing, coding, and problem-solving easier, people may not want to learn these skills on their own as much. In the long term, this risks undermining core research, critical thinking, and expression skills, especially for students and emerging professionals.

Also, if companies rely too much on the model for their procedures, a model failure, price rise, or access restriction might mess up the whole workflow. The response to the question "Can we survive without LLMs?" is starting to get weaker from a strategic point of view.

LARGE LANGUAGE MODELS: WHICH MODEL? WHICH PURPOSE?

The organization and purposes of their training are different for each Large Language Model. These distinctions let you figure out which model works best for each part. This part gives a brief overview of the most important model families and the types of problems they can help with.

GPT Series

OpenAI invented GPT, a family of LLMs that takes free text generation, discussion, and code output to new heights by being able to guess what the next token will be. Introduced in 2018, the first GPT model developed the essential training paradigm for LLMs by systematizing a pre-training and task-oriented fine-tuning technique using large-scale data. The next version, GPT-2, was better at modeling context since it had more parameters and data. GPT-3, in turn, considerably boosted its power to execute complicated instructions and multi-step operations.

GPT-4, which came out in 2023, made contextual awareness much deeper, answers even more consistent, and reactions to prompts even more subtle. The GPT series is known for its great performance, especially when it comes to making text, conversation systems, and creative material. Every successive iteration adds to what the last one could do, pushing the limits of what generative AI can do.

BERT and its Variations: Understanding and Representation-Oriented Models

BERT, which was made by Google, is different from GPT models in that it "understands text rather than generating it." This is because its encoder structures handle text in both directions, with context coming from both the left and the right. This enables it to excel in tasks including sentiment analysis, topic categorization, name and entity recognition, retrieval of passages for questions and answers, and search ranking.

Enhanced versions of BERT, like RoBERTa, make the BERT architecture even better by using bigger datasets and better pre-training methods. For businesses, it typically makes sense to use a BERT-based model for activities like classifying documents, moderating content, enhancing search quality, and automatically sorting support queries. This is because it strikes a good compromise between performance and cost.

T5 and The Text-To-Text Approach: Multitasking Transformations

Text-to-text and T5 models both have a consistent framework that says, "input is text, output is text." This means that the same model may be used to do a lot of various things, such translating, summarizing, making headlines, answering questions, and writing comments. The only thing that changes is the task name.

This method works best for businesses who wish to use one general-purpose model to do a lot of different text processing jobs. It is a good solution for teams who want to construct a "single core model adaptable to many tasks" since it makes training and fine-tuning easier, transferring information between tasks easier, and maintenance costs lower.

Conversation Agents and Chatbots

Conversational systems are built on large language models, which make it possible for people to talk to each other in a way that is similar to how people talk to each other. These models can not only answer individual queries, but they can also keep multi-round and complex conversations going by understanding user input in context.

Because of these qualities, chatbots and virtual assistants can help with a lot of things around the clock, especially customer service. This makes response times shorter and makes users happier.

Language Translation and Sentiment Analysis

When trained on multilingual data, LLMs can generate translations between languages that are quite accurate and fit the context. Compared to traditional methods, their capacity to take into account complex linguistic aspects including phrases, cultural expressions, and language details greatly increases the quality of translations.

Also, LLMs do a good job of sentiment analysis by looking at the emotional tone and approach of texts. These models are useful for analyzing social media and evaluating user comments since they can pick up on latent meanings like irony and implication.

Which model, which purpose?

- Free text, dialogue, code generation, and creative tasks: GPT, instruction-oriented models.
- Sentiment analysis, classification, information extraction, and search: Bidirectional encoder-based models such as BERT and RoBERTa.
- Summarizing, translating, and multitasking text transformations: T5 and text-to-text designed models.
- Multilingual systems and global products: Multilingual pre-trained GPT or T5.

THE FUTURE OF LARGE LANGUAGE MODELS

More Efficient and Specialized Architectures

A lot of LLM research will be on structures that can do the same thing with fewer parameters and less energy in the near future. In addition to big, single-piece models, smaller, modular, and specialized models that are optimized for certain tasks (such mixture-of-experts and task-specific sub-models) are also likely to become common. These changes will make it possible for smaller businesses as well as big IT corporations to employ LLMs.

Open models and strengthening ecosystems

More and more research groups and businesses are putting out open or partially open models that come with pre-trained weights and training scripts. This tendency is very important for both academic and local adaptation, as well as for auditability. Researchers and developers can adjust their own data without having to rely on closed systems.

There is a lot of activity around open models, including datasets, assessment metrics, security checks, and toolchains. This allows for the spread of specialized LLM solutions, especially for certain languages, industries, or rules. In other words, it leads to a more decentralized wave of innovation instead of a "centralized" one.

Regulatory frameworks and governance

As LLMs acquire increasingly prevalent, it is predicted that rules around data privacy, copyright, bias, and security will become clearer both in the US and around the world. For enterprises, this means that the design should focus on not only technological performance but also legal compliance, auditability, and reportability.

In the near future, genuine LLM initiatives may include things like model cards, risk classifications, explainability reports, and limits on how they can be used. So, the question "which model to use?" will no longer be just a technological one; it will also involve legal, ethical, and corporate responsibility issues.

RESULTS AND DISCUSSION

This study has offered a detailed review of Large Language Models by evaluating their theoretical foundations, architectural principles, training procedures, application fields, and upcoming difficulties. The analysis shows how improvements in attention mechanisms, representation learning, and large-scale self-supervised training have made LLMs able to do more and more complicated language tasks with high accuracy and flexibility. This is done by following the development of early statistical language models to modern transformer-based systems. A review of well-known model families shows that the choices made about the architecture and the goals of training directly affect how well a model works for a given job. This shows how important it is to choose the right model for real-world use.

In the future, more research focused at making models easier to comprehend, more efficient, and more scalable is likely to make LLMs even more powerful. As more people use it, businesses will probably gain from automation that works better, more personalized experiences, and better ways to make decisions. At the same time, continuous work to fix bias and reliability—through debiasing algorithms, balanced and synthetic data production, and explainability tools—will be very important for making systems that people can trust. At the same time, the development of increasingly advanced multimodal models that can interpret text, images, audio, and video data together leads to a future in which language models are a key part of more natural and richer interactions between people and AI. These changes all point to the idea that the future influence of LLMs will depend not just on how well they work, but also on how well they are designed to be fair, open, and long-lasting.

REFERENCES

Beggag, A., and Saad, Y., (2025). Deep learning, transformers and graph neural networks: a linear algebra perspective. 100, 2095-2134.

Bingi, Y., and Yin, Y., (2023). An analysis of embeddinf layers and similarity scores using siamese neural networks. Arxiv: 2401-00582.

Chen, S. F., and Goodman, J., (1999). An empirical study of smoothing techniques for language modeling. Computer speech and language, 13(4), 359-394.

Gurgurov, D., Baumel, T., and Anikina, T., (2024). Multilingual large language models and curse of multilinguality. Arxiv: 2406.10602.

Ji, W., Yuan, W., Getzen, E., Cho, K., Jordan, M. I., Mei, S., Wston, J. E., Su, W. J., Xu, J., and Zhang, L., (2025). An overview of large language models for statisticians. Arxiv: 2502. 17814.

Swaileh, W., (2017). Language modelling for handwriting recognition, Normandie Universite. NNT: 2017NORMR024.

Wang, Z., Chu, Z., Doan, T. V., Ni, Shiwen, Yang, M., and Zhang, W., (2025). History, development, and principles of large language models: an introductory survey. AI and Ethics, 5, 1955-1971.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Fomez, A. N., Kaise, L., and Polosukhin, I., (2017). Attention is all you need. 31st Conference on neural information processing system, Long Beach, CA, USA.

Voita, E., Talbot, D., Moiseev, F., Sennrich, R., and Titov, I., (2019). Analyzing Multi-head self-attention: specialized heads do the heavy lifting, the rest can be pruned. Procedinf og the 57th annual meeting of the association for computational linguistics. Florence, Italy, 5707-5808.

Zhang, C., Peng, B., Sun, X., Niu, Q., Liu, J., Chen, K., Li, M., Feng, P., Bi, Z., Liu, M., Zhang, Y., Song, X., Fei, C., Yin, C. H., Yan, L. K., He, H., and Wang, T., (2024). From word vectors to multimodelal embeddings: techniques, applications, and future directions for large language models. Arxiv: 2411.05036.

Zhao, W. X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., Min, Y., Zhang, B., Zhang, J., Ding, Z., Du, Y., Yang, C., Chen, Y., Chen, Z., Jiang, J., Ren, R., Li, Y., Tang, X., Liu, Z., Liu, P., Nie, J., and Wen, J., (2025). A survey of Large language models. Arxiv, 16, 2303.18223.

Energy and Exergy Analysis of Coal Combustion

M. Özgün KORUKÇU[1]

1- Prof. Dr.; University of Bursa Uludağ, Faculty of Engineering, Department of Mechanical Engineering. ozkorukcu@uludag.edu.tr ORCID No: 0000-0002-4761-4304

ABSTRACT

This chapter presents a comprehensive computational framework for analyzing coal combustion using a Python-based graphical user interface (GUI) that integrates energy, exergy, and carbon-footprint assessments. The developed tool performs detailed thermochemical calculations based on the ultimate analysis of coal, including product species determination, stoichiometric oxygen demand, adiabatic flame temperature, and the distribution of CO_2, CO, H_2O, SO_2, and residual oxygen. In addition to the combustion equilibrium calculations, the GUI evaluates flue-gas cooling behavior, sensible energy transfer, entropy generation, and exergy destruction during the cooling process from the adiabatic flame temperature to the exhaust conditions.

A key contribution of the study is the incorporation of carbon-intensity metrics—including CO_2 emissions per unit fuel mass and per unit thermal energy (kg CO_2/kWh$_{th}$)—enabling an integrated thermodynamic and environmental evaluation of coal combustion. The program also performs parametric analysis across a wide range of equivalence ratios (0.5–1.5), allowing users to investigate how lean and rich combustion conditions influence flame temperature, heat transfer, irreversibilities, and carbon emissions. Results demonstrate that variations in equivalence ratio significantly affect both the thermodynamic performance and the environmental footprint of combustion systems.

By combining first-principles thermochemical modeling with an intuitive user interface, the proposed tool provides a transparent, flexible, and open-source alternative to commercial simulation packages. It serves as an effective platform for engineering analysis, optimization studies, and educational purposes, offering valuable insights into the linked behavior of energy efficiency, exergy losses, and carbon emissions in solid-fuel combustion processes.

Keywords – Coal Combustion, Adiabatic Flame Temperature, Exergy Destruction, Carbon Footprint, Thermochemical Modeling.

INTRODUCTION

The increasing global demand for energy, coupled with the urgent need to reduce greenhouse gas emissions, has elevated the importance of advanced combustion analysis in modern power and process industries. Coal, despite the rapid penetration of renewable energy sources, remains one of the most widely used primary fuels for electricity and heat generation. Understanding the thermochemical behavior of coal during combustion—particularly the distribution of products, the adiabatic flame temperature, the heat transfer characteristics, and the associated exergy losses—continues to

be essential for improving system efficiency and reducing environmental impact.

Luo et al. (Luo et al., 2021) presented a comprehensive exergy-based assessment of a coal-fired Allam cycle aimed at identifying the principal thermodynamic inefficiencies within the system. The findings demonstrate that the cycle achieves an overall exergetic efficiency of approximately 40.6%, while maintaining near-complete CO_2 capture, underscoring its potential as a low-emission coal utilization technology. The exergy analysis reveals that the gasifier (29.1%) and combustion chamber (17.4%) constitute the dominant sources of irreversibility due to the inherent inefficiencies associated with chemical reactions. Sensitivity analyses further show that the power consumption of the air separation unit and turbine inlet and outlet pressures exert a significant influence on the overall exergetic efficiency. A comparative evaluation with a gas-fired Allam cycle indicates distinct differences in the distribution of exergy destruction, particularly a comparatively lower contribution from the combustor in the coal-fired configuration.

Bhunia et al. (Bhunia et al., 2017) investigated the oxy-fuel combustion behavior of coarse sub-bituminous coal char through a combination of detailed experiments and a fully transient, one-dimensional, non-isothermal combustion model. Experiments conducted at 1100 K with varying oxygen concentrations (42–60 percent) in both O_2–N_2 and O_2–CO_2 environments demonstrate significantly lower peak particle temperatures and slower combustion rates in CO_2-rich atmospheres due to reduced oxygen diffusivity and the endothermic CO_2-gasification reaction. The developed model incorporates multicomponent mass transfer (via the Maxwell–Stefan formulation), heat transfer, and heterogeneous and homogeneous reaction kinetics, and shows close agreement with experimental center-temperature, conversion, and carbon-consumption measurements across the tested conditions. Simulation results reveal that particle size, reactor temperature, and oxygen concentration strongly influence burnout time and reaction rates, with larger particles exhibiting higher peak temperatures and shorter burnout times as a result of reduced heat loss per unit mass. The study provided a reliable predictive framework for understanding char combustion mechanisms under oxy-fuel conditions and offers guidance for the design and optimization of combustion systems.

Aydın and Kopaç (Aydın & Kopaç, 2025) presented ChemEqT, a unified computational tool designed to predict chemical equilibrium compositions and adiabatic flame temperatures in multicomponent fuel–air systems. The model integrates both Gibbs free energy minimization and the element potential method, enabling flexible and accurate combustion calculations across a wide range of species and thermochemical conditions. Validation using propane-air and methane-air mixtures, as well as ten additional fuel–air systems, demonstrates exceptionally low deviations—

typically below 1%—from established literature values for both species distributions and adiabatic flame temperatures. The tool leverages NASA's 7- and 9-coefficient polynomial databases to compute thermodynamic properties with high fidelity, ensuring robust performance even for large reaction mechanisms. Overall, the study establishes ChemEqT as an efficient, modular, and high-accuracy platform suitable for combustion modeling, parametric studies, and broader energy system simulations.

Kurose et al. (Kurose, 2009) presented a comprehensive review of recent advances in the numerical simulation of pulverized coal combustion, focusing on Reynolds-Averaged Navier–Stokes (RANS) and Large-Eddy Simulation (LES) approaches. The authors describe how detailed computational fluid dynamics (CFD) models, validated using single- and multi-burner furnace experiments, can capture complex interacting phenomena such as devolatilization, char oxidation, turbulent mixing, NOx formation, and swirling recirculation zones. Their results show that LES provides markedly superior predictions of unsteady flow structures and recirculation behavior compared to RANS, particularly in regions of strong shear and swirl, as illustrated in velocity and species-distribution profiles throughout the furnace. Parametric studies further demonstrate the influence of coal properties—including fuel ratio, ash content, and moisture—on burnout, NOx emissions, and temperature fields. The study concludes that LES is a promising tool for future furnace and burner design, though challenges remain in turbulence–chemistry interaction modeling and the computational cost of large-scale simulations.

Korukçu (Korukcu, 2024) presented a graphical user interface (GUI) designed to compute adiabatic flame temperature, combustion equilibrium products, and exergy destruction for a wide range of alkanes and alcohols. The software incorporates thermochemical property databases and chemical equilibrium formulations to calculate combustion behavior using both dry and moist air. Validation results show strong agreement with reference data, demonstrating the high accuracy of the program. The study additionally examines the influence of relative humidity on adiabatic flame temperature and exergy destruction, revealing a linear decrease in flame temperature and a quadratic increase in exergy destruction with increasing humidity. This tool offers a practical and accurate platform for academic instruction, research studies, and preliminary combustion system design.

Goreliko et al. (Goreliko et al., 2025) presented an experimental investigation of pulverized coal flame combustion in a 50 kW two-stage tangential vortex burner. The study compares co-swirl and counter-swirl configurations by measuring temperature distributions, gas composition, and detailed flow fields using two-component laser Doppler anemometry, with the particle velocity. Results indicate that the counter-swirl configuration enhances mixing, increases centerline temperatures beyond 1000 °C, and yields higher CO_2 and lower O_2 concentrations at the burner outlet. Axial and

tangential velocity measurements demonstrate more uniform velocity profiles and disappearance of the propane–butane jet influence along the chamber length under counter-swirl conditions. The authors conclude that counter-swirl operation provides more stable ignition and more efficient combustion of pulverized coal fuel.

Cantera is an online tool capable of estimating the adiabatic flame temperature (AFT) using inputs such as pressure, initial temperature, fuel composition, and equivalence ratio (*Cantera*). Pratt's EQLBRM program (Pratt), written in Fortran IV, provides a text-based computational environment modeled after early DOS/UNIX systems.

Yazar and Demir (Yazar & Demir, 2022) introduced a Java-based graphical user interface (GUI) designed to evaluate combustion behavior, emissions, and engine performance for internal combustion engines, enabling analysis of in-cylinder pressure, temperature, gas composition ratios, and performance parameters under various excess-air conditions ($\phi > 1$).

Jakoubek and Hanzal (Jakoubek & Hanzal, 1990) developed online control algorithms aimed at regulating flue gas emissions in power plants, integrating real-time communication with operational systems to maintain acceptable emission levels while sustaining energy production.

Yıldız (Yıldız, 2023) applied a chemical-equilibrium combustion model to investigate the influence of hydrogen enrichment on biogases with varying CO_2 concentrations, concluding that hydrogen addition leads to an increase in adiabatic flame temperature.

Cui et al. (Cui et al., 2024) conducted an experimental investigation into the combustion behavior of coal–ammonia gas–solid fuel jets using a flat flame burner, focusing on ignition characteristics and gas-phase species evolution. The study demonstrates that adding ammonia significantly modifies the preheating environment, resulting in altered ignition delay times that depend strongly on pulverized coal concentration. Measurements of O_2, CO, and NOx along the burner height reveal that ammonia addition promotes oxygen competition, enhances CO formation at low coal concentrations, and increases NO production through accelerated fuel-N conversion. The authors also show that lower coal concentrations favor the formation of NO_2 and N_2O in the early combustion stages, although these species decline rapidly as temperature increases. Overall, the results highlight the complex thermal and chemical interactions in coal–ammonia co-combustion and identify an optimal coal concentration of approximately 0.41 kgc/kga for stable and low-nitrogen operation.

Özer (Özer, 2019) investigated the combustion behavior of lignites from different regions of Turkey and Germany using both experimental and numerical methods. The study employs thermogravimetric analysis (TGA) under low heating rates and drop tube furnace (DTF) experiments under high heating rates to quantify ignition characteristics, volatile release, and char reactivity of the fuels. Results show that volatile-to-fixed-carbon ratios

significantly influence ignition times, while ash content inversely affects maximum reaction rates. High-heating-rate DTF measurements confirm these trends, revealing that Rhenish lignite exhibits the highest reactivity among the tested fuels. Numerical modeling using CFD and the PoliMi multi-step kinetic mechanism shows good agreement with experimental volatile yields and char oxidation behavior, providing a validated framework for simulating single-particle coal combustion.

Kuznetsov et al. (Kuznetsov et al., 2024) conducted a comprehensive computational analysis of pulverized coal combustion under oxygen-enriched conditions in a vortex burner. Using a URANS-based CFD framework with the k–ω SST turbulence model, discrete ordinates radiation modeling, and CPD devolatilization kinetics, the study examines how varying oxygen concentrations influence flow structure, ignition, flame characteristics, and burnout behavior. The results show that increasing oxygen concentration decreases the velocity of the secondary-blast jet and weakens the reverse-flow zone, which in turn elongates the flame and increases unburned carbon concentrations downstream. Numerical predictions demonstrate good agreement with experimental data, confirming the accuracy of the proposed modeling approach. The authors conclude that although higher oxygen levels reduce ignition delay, burner aerodynamics may offset this benefit by promoting longer flames and lower combustion completeness.

Kim et al. (Kim et al., 2024) experimentally investigate oxy-fuel combustion behavior and emission characteristics in a 10 kWth pressurized fluidized bed combustor across operating pressures from 3 to 8 bar(g). Their results show that CO_2 concentrations in the flue gas consistently exceed 90% under all oxy-fuel conditions, indicating strong potential for efficient carbon capture. Increasing pressure reduces unburned carbon and CO levels in the fly ash, thereby improving combustion efficiency, which reaches 98.3% at 8 bar(g). The study also demonstrates that NO and N_2O emissions decrease with pressure, while NO_2 becomes a larger fraction of total NOx, and that SO_2 emissions decrease under oxy-fuel conditions due to enhanced direct desulfurization. Biomass cofiring further lowers CO, NOx, and SO_2 emissions, suggesting that pressurized oxy-fuel FBC offers both environmental and performance advantages for mixed coal–biomass combustion systems.

Ozsari et al. (Ozsari et al., 2021) present a detailed thermodynamic comparison of oxy-fuel combustion and conventional air combustion using a validated equilibrium combustion model. The study examines the effects of oxygen concentration on adiabatic flame temperature, product species, entropy generation, and NOx formation, employing natural gases from four major global sources. Their results show that oxy-combustion significantly reduces NOx emissions—up to forty-fold—while also lowering entropy generation compared to air combustion, thus offering higher exergy potential. Validation against NASA CEA and GASEQ demonstrates strong agreement

in flame temperatures, enthalpy, and major species fractions. The authors conclude that oxy-combustion provides important environmental and thermodynamic advantages, though performance varies with fuel origin.

Conventional combustion engineering relies heavily on manual stoichiometric calculations, empirical correlations, or dedicated commercial software packages. However, these approaches often lack transparency, require extensive user expertise, or are not sufficiently flexible for parametric studies. Modern engineering practice increasingly demands computational tools that are both physically rigorous and accessible to engineers, researchers, and students. Graphical User Interface (GUI)–based tools offer a unique advantage by integrating detailed thermochemical modeling with intuitive user interaction, enabling real-time exploration of the effects of fuel composition, equivalence ratio, and operating conditions.

This chapter presents a comprehensive combustion analysis framework based on a Python-powered GUI capable of performing energy, exergy, and carbon-footprint calculations for solid fuels. The tool computes product species, adiabatic flame temperature, flue-gas cooling energy, entropy generation, and exergy destruction, while also determining CO_2 emissions and CO_2-equivalent metrics per unit fuel mass and per unit useful heat output. By enabling systematic variation of the equivalence ratio between lean and rich combustion regimes, the GUI facilitates detailed parametric investigations that reveal the thermodynamic and environmental performance of coal combustion.

The chapter aims to bridge the gap between fundamental combustion theory and practical engineering analysis. It emphasizes transparent thermochemical calculations based on first principles, while demonstrating how Python can be used to build reproducible, open-source tools for energy systems modeling. The presented methodology serves not only as a practical engineering instrument but also as an educational platform for understanding the interconnected roles of thermodynamics, exergy analysis, and emissions assessment in fossil-fuel combustion.

THEORY OF COMBUSTION

Coal does not have a fixed chemical formula like gaseous or liquid hydrocarbon fuels; instead, its composition is defined through its ultimate analysis, specifying the mass fractions of carbon, hydrogen, oxygen, nitrogen, sulfur, and ash. During combustion, the reactive elements in coal—primarily carbon and hydrogen—undergo oxidation, releasing heat energy to the

surroundings. The oxidizer is predominantly atmospheric air, which is typically modeled as a mixture of O_2 and 3.76 kmol of N_2. Because coal composition varies with source and rank, its combustion reaction must be written using elemental balances derived from the ultimate analysis rather than a single molecular formula. Stoichiometric combustion (ϕ = 1), fuel-rich combustion ($\phi > 1$), and excess-air combustion ($\phi < 1$) can then be expressed analogously to hydrocarbon systems by incorporating the appropriate oxygen demand and product distributions calculated from these elemental ratios.

$$a\,C + b\,H_2 + c\,O_2 + d\,N_2 + e\,S + \emptyset(O_2 + 3.76\,N_2) \rightarrow x\,CO_2 + y\,H_2O + z\,SO_2 + w\,N_2 \quad \emptyset = 1 \tag{1}$$

$$a\,C + b\,H_2 + c\,O_2 + d\,N_2 + e\,S + \emptyset(O_2 + 3.76\,N_2) \rightarrow x\,CO_2 + k\,CO + y\,H_2O + z\,SO_2 + w\,N_2 \quad \emptyset > 1 \tag{2}$$

$$a\,C + b\,H_2 + c\,O_2 + d\,N_2 + e\,S + \emptyset(O_2 + 3.76\,N_2) \rightarrow x\,CO_2 + y\,H_2O + z\,SO_2 + l\,O_2 + w\,N_2 \quad \emptyset < 1 \tag{3}$$

The air–fuel ratio (*AF*) represents the ratio between the mass of supplied air (m_{air}) and the mass of the fuel (m_{fuel}), as shown below.

$$AF = \frac{m_{air}}{m_{fuel}} \tag{4}$$

The overall enthalpy difference between the reactants and products allows the calculation of the heat released from the combustion process (Q_{out}) when no work interactions occur. Equation 5 accounts for several thermodynamic parameters, including the number of moles (*N*), the standard enthalpy of formation ($\bar{h}^{\circ}_{for}$), the enthalpy of each species at the specified temperature ($\bar{h}_T$), and the reference enthalpy at 1 atm and 25 °C ($\bar{h}^{\circ}_{ref}$).

$$Q_{out} = \underbrace{\sum N_R \left(\bar{h}^{\circ}_{for} + \bar{h}_T - \bar{h}^{\circ}_{ref}\right)_R}_{\substack{Energy\ in\ by\ mass \\ per\ mole\ of\ fuel}} - \underbrace{\sum N_P \left(\bar{h}^{\circ}_{for} + \bar{h}_T - \bar{h}^{\circ}_{ref}\right)_P}_{\substack{Energy\ out\ by\ mass \\ per\ mole\ of\ fuel}} \tag{5}$$

$$Q = W = 0 \tag{6}$$

For an adiabatic combustion process, the total enthalpy of the reactants (H_R) must equal the total enthalpy of the products (H_P), as expressed in Equations 7–8.

$$H_R = H_P \tag{7}$$

$$\sum N_R \left(\bar{h}^{\circ}_{for} + \bar{h}_T - \bar{h}^{\circ}_{ref}\right)_R = \sum N_P \left(\bar{h}^{\circ}_{for} + \bar{h}_T - \bar{h}^{\circ}_{ref}\right)_P \quad (8)$$

The entropy change of the combustion can be found as

$$\underbrace{S_{in} - S_{out}}_{\substack{Net\ entropy \\ transfer\ by \\ heat\ and\ mass}} + \underbrace{S_{gen}}_{\substack{Entropy \\ generation}} = \underbrace{\Delta S_{system}}_{\substack{Change\ in \\ entropy}} \quad (9)$$

The entropy generation of the system is

$$S_{gen} = S_P - S_R + \sum \frac{Q_{out}}{T_{ambient}} \quad (10)$$

In an adiabatic combustion reaction

$$S_{gen,adiabatic} = S_P - S_R \geq 0 \quad (11)$$

The entropy of each component can be calculated with

$$\bar{s}_n(T, P_n) = \bar{s}^{\circ}_n(T, P_0) - R_u ln \frac{y_n P_m}{P_{ambient}} \quad (12)$$

Where Ru is the universal gas constant and 8.3144 7 kJ/kmolK.

The total entropy generation is calculated as

$$S_{gen} = \sum N_P \bar{s}_P - \sum N_R \bar{s}_R + \sum \frac{Q_{out}}{T_{ambient}} \quad (13)$$

Exergy destruction can be found with

$$X_{destroyed} = T_{ambient}\, S_{gen} \quad (14)$$

The aforementioned equations are from Cengel and Boles (Cengel & Boles, 2011)

The carbon footprint of coal combustion can be quantified by evaluating the mass of carbon dioxide and carbon monoxide generated during oxidation and expressing the emissions relative to the mass or energy content of the fuel. For solid fuels such as coal, the total CO_2 produced is directly

related to the carbon content identified in the ultimate analysis. The CO_2 generated is calculated using the elemental carbon balance:

$$m_{CO_2} = m_{C\ fuel}\ \frac{M_{CO_2}}{M_C} \tag{15}$$

where M_{CO_2}/M_C describes the stoichiometric conversion of carbon to CO_2 (Turns, 2012; Haywood, 1991). In cases where CO is present in the flue gas, its eventual oxidation to CO_2 in the atmosphere is accounted for using a CO_2-equivalent contribution:

$$m_{CO_2\ eq} = m_{CO_2} + m_{CO}\frac{M_{CO_2}}{M_C} \tag{16}$$

which reflects the full oxidation pathway of CO as documented in combustion chemistry literature (Glassman et al., 2014) and climate models (Kikstra et al., 2022).

The carbon intensity of the fuel may be expressed per unit mass of fuel burned:

$$CI_{kg/kg} = \frac{m_{CO_2}}{m_{fuel}} \tag{17}$$

following emission-factor methodology applied by the U.S. EPA (*AP-42*). Alternatively, the emissions can be normalized per unit of useful energy released during combustion:

$$CI_{kg/kWh} = \frac{m_{CO_2}3600}{Q_{out}} \tag{17}$$

which links the mass of CO_2 to the actual heat release obtained from thermodynamic analysis (Cengel & Boles, 2011). For complete climate-impact assessment, a general carbon-footprint formulation may be applied:

$$Carbon\ Footprint = \sum_i m_i\ GWP_i \tag{18}$$

where m_i represents the emitted mass of species i, and GWP_i is its global warming potential (IPCC, 2021). For coal combustion, the dominant species is CO_2, with CO included through Equation (16) due to its ultimate conversion in the atmosphere.

RESULTS

In this section, combustion analyses of different lignites that are extracted in Turkey are compared. Table 1 shows the ultimate analysis of lignites from four sites in Turkey.

Table 1: The ultimate analysis of lignites from four sites in Turkey (Korkmaz, 2022)

Site	C (%)	H_2 (%)	N_2 (%)	S (%)	O_2 (%)
Çayırlı	69.26	3.98	1.36	1.42	23.98
Ermenek	39.49	5.32	0.90	3.94	50.35
Ilgın	40.45	2.96	0.68	4.52	51.39
Sorgun	58.70	3.54	0.87	1.98	34.91

Figure 1 shows the combustion calculation results for Çayırlı site. As you see from Figure 1, the equivalence ratio is taken as 1. The air and coal inlet temperature and ambient temperature values are taken as 25 °C. Flue gas outlet temperature is taken as 800 °C. According to the results, the air-fuel ratio is found as 30.437 kg air/kg coal. The adiabatic flame temperature is found as 2516.3 °C. Released heat to ambient is 1901245.2 kJ/100 kg coal. Entropy generation is 5318.64 kJ/k and exergy destruction is 1585751.4 kJ. CO_2-based carbon footprint results reveal that 2.5378 kg/kg of coal is produced, and the carbon intensity is 0.331 kg/kWh_{th}.

Figure 2 shows the combustion equation of Çayırlı site for the equivalence ratio f 1. As seen in Figure 2, the user can see the entire combustion reaction when the "Show LaTex Equation" button is clicked.

Figure 3 shows the variation of adiabatic flame temperature with equivalence ratio. The highest adiabatic flame temperature value is obtained for Ilgın site at an equivalence ratio of 1 with 3140.2 °C, following Ermenek site with 3013.6 °C, Sorgun site with 2670.7 °C, and Çayırlı site with 2516.3 °C.

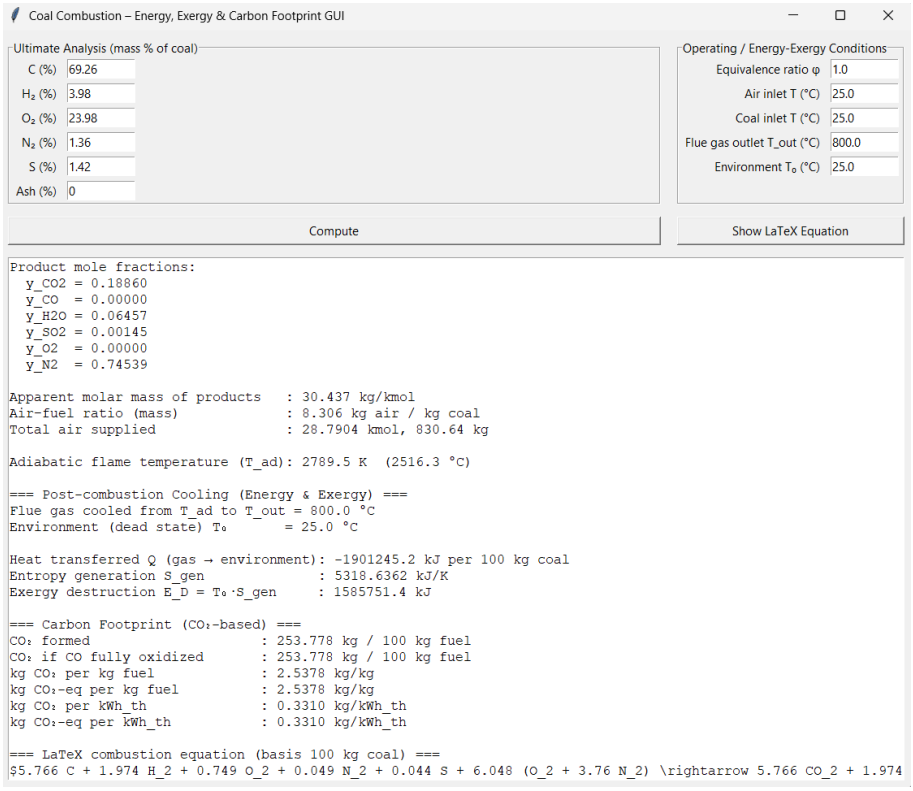


Figure 1: Combustion results for Çayırlı site.

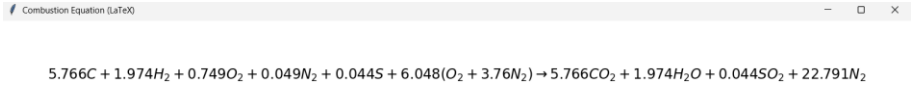


Figure 2: Combustion reaction for Çayırlı site.

Figure 4 shows the variation of heat transfer with equivalence ratio. The highest heat transfer value is obtained for Çayırlı site at an equivalence ratio of 1 with 1901245.2 kJ/100kg, following Sorgun site with 1673057.0 kJ/100kg, Ermenek site with 1457988.2 kJ/100kg, and Ilgın site with 1293733.6 kJ/100 kg.

Figure 5 shows the variation of exergy destruction with equivalence ratio. The highest exergy destruction value is obtained for Çayırlı site at an equivalence ratio of 1 with 1585751.4 kJ/100kg, following Sorgun site with 1403977.6 kJ/100kg, Ermenek site with 1238185.1 kJ/100kg, and Ilgın site with 1103014.3 kJ/100kg.

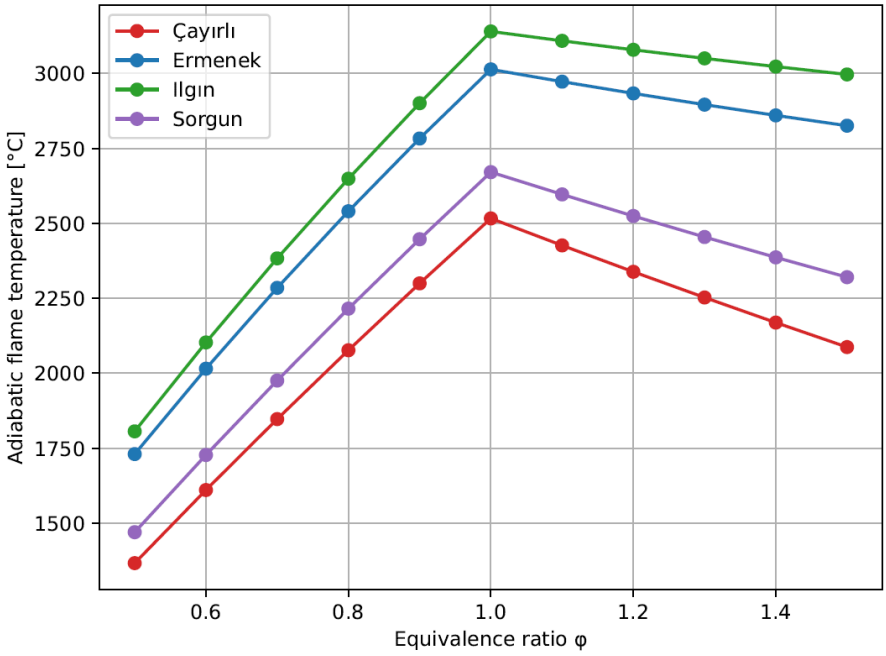


Figure 3: Variation of adiabatic flame temperature with equivalence ratio for sites.

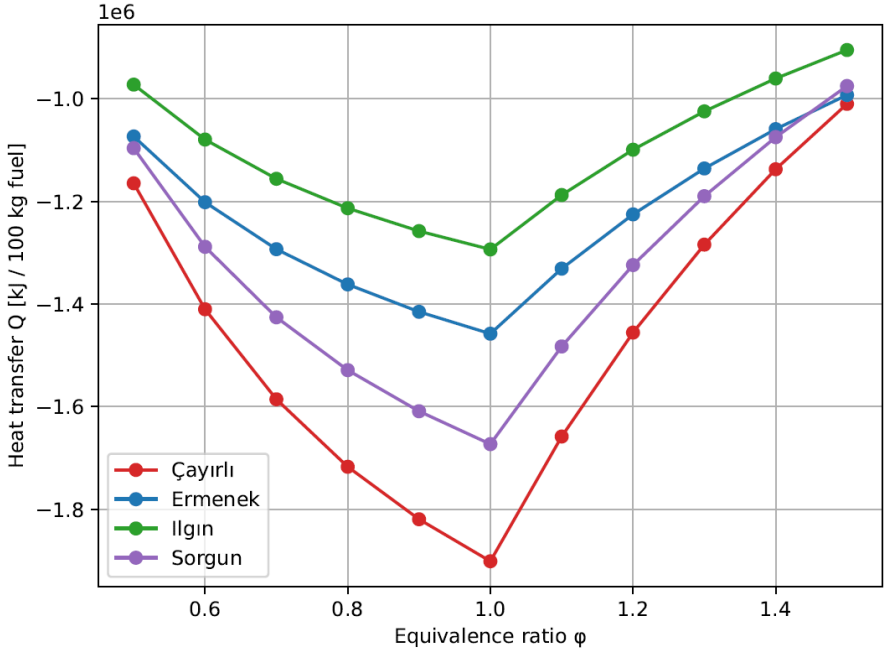


Figure 4: Variation of heat transfer with equivalence ratio for sites.

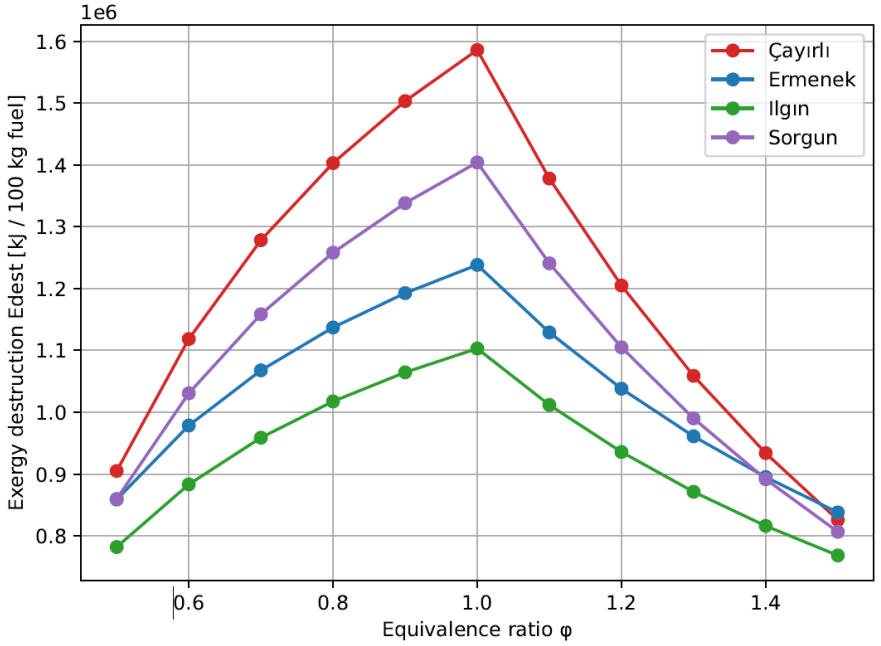


Figure 5: Variation of exergy destruction with equivalence ratio for sites.

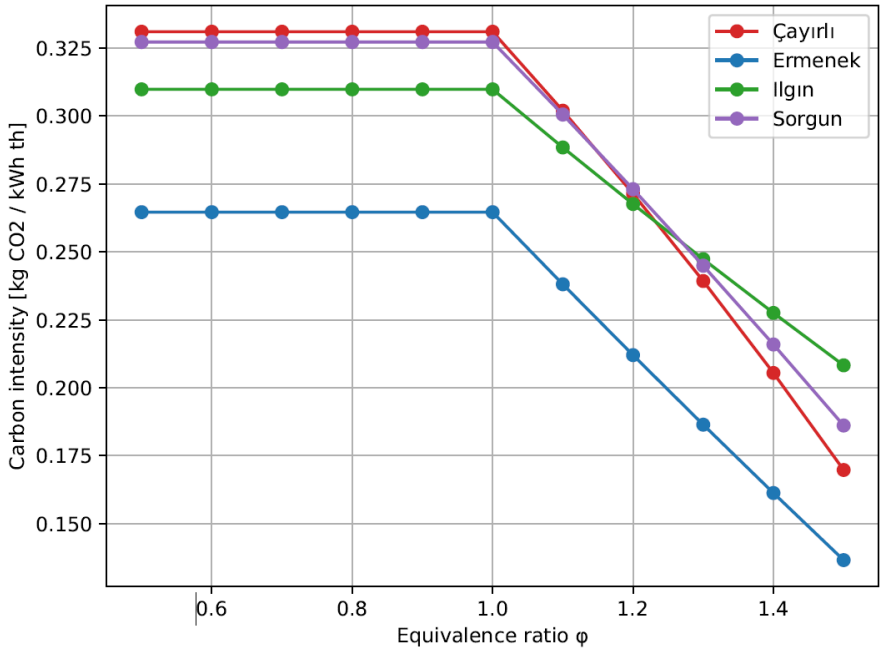


Figure 6: Variation of carbon intensity with equivalence ratio for sites.

Figure 6 shows the variation of carbon intensity with equivalence ratio. The highest carbon intensity value is obtained for Çayırlı site at an equivalence ratio of 1 with 0.3310 kg/kWh$_{th}$, following Sorgun site with 0.3272 kg/kWh$_{th}$, Ilgın site with 0.3098 kg/kWh$_{th}$, and Ermenek site with 0.2646 kg/kWh$_{th}$.

CONCLUSION

Energy, exergy, and carbon footprint analysis of coal combustion is investigated. For combustion calculations, a graphical user interface (GUI) is designed with Python and PyQt5. This user-friendly GUI is capable of versatile calculations.

The ultimate analysis of coal by mass can be given as input, and the program calculates adiabatic flame temperature, heat released to the ambient, entropy generation, exergy destruction, and carbon footprint. GUI also shows the whole combustion equation.

According to the calculations, The highest adiabatic flame temperature value is obtained for Ilgın site at an equivalence ratio of 1 with 3140.2 °C, following Ermenek site with 3013.6 °C, Sorgun site with 2670.7 °C, and Çayırlı site with 2516.3 °C.

The highest heat transfer value is obtained for Çayırlı site at an equivalence ratio of 1 with 1901245.2 kJ/100kg, following Sorgun site with 1673057.0 kJ/100kg, Ermenek site with 1457988.2 kJ/100kg, and Ilgın site with 1293733.6 kJ/100 kg.

The highest exergy destruction value is obtained for Çayırlı site at an equivalence ratio of 1 with 1585751.4 kJ/100kg, following Sorgun site with 1403977.6 kJ/100kg, Ermenek site with 1238185.1 kJ/100kg, and Ilgın site with 1103014.3 kJ/100kg.

The highest carbon intensity value is obtained for Çayırlı site at an equivalence ratio of 1 with 0.3310 kg/kWh$_{th}$, following Sorgun site with 0.3272 kg/kWh$_{th}$, Ilgın site with 0.3098 kg/kWh$_{th}$, and Ermenek site with 0.2646 kg/kWh$_{th}$.

Finally, the integration of energy, exergy, and emissions analyses into a single computational platform demonstrates the pedagogical and practical value of the proposed GUI. By allowing users to interactively modify coal composition, inlet temperatures, and equivalence ratio, the tool supports both conceptual understanding and early-stage engineering evaluations. Such tools are increasingly important as coal-based systems transition toward higher-efficiency and lower-emission operation, including co-firing, oxy-fuel

strategies, and carbon-capture applications. Overall, the results presented in this chapter provide a solid framework for analyzing coal combustion systems and can serve as a foundation for future extensions involving detailed kinetics, multi-component fuels, and advanced environmentally conscious combustion strategies.

REFERENCES

AP-42: Compilation of Air Emissions Factors from Stationary Sources. (n.d.). https://www.epa.gov/air-emissions-factors-and-quantification/ap-42-compilation-air-emission-factors

Aydın, N. Ö., & Kopaç, M. (2025). Isı Bilimi ve Tekniği Dergisi ChemEqT : A Unified Computational Tool for Predicting Equilibrium Composition and Adiabatic Flame Temperature in Multicomponent Fuel – Air Systems. *Journal of Thermal Science and Technology*, *45*(2), 325–334. https://doi.org/10.47480/isibted.1716739

Bhunia, S., Sadhukhan, A. K., & Gupta, P. (2017). Modelling and experimental studies on oxy-fuel combustion of coarse size coal char. *Fuel Processing Technology*, *158*, 73–84. https://doi.org/10.1016/j.fuproc.2016.11.015

Cantera.http://elearning.cerfacs.fr/combustion/tools/adiabaticflametemperature/index.php

Cengel, Y., & Boles, M. A. (2011). *Thermodynamics An Engineering Approach 5th. Edition* (5th ed.). Mc-Graw Hill.

Cui, M. S., Niu, F., Ji, R. S., Duan, L., & Zhang, X. (2024). *Experimental Study on Flame Chemical Composition of Coal and Ammonia Gas − Solid Jet in Flat Flame Burner*. https://doi.org/10.1021/acsomega.3c09231

Glassman, I., Yetter, R. A., & Glumac, N. G. (2014). *Combustion* (5th Editio). Academic Press.

Goreliko, E. Y., Litvinov, I. V, Butakov, E. B., & Shtork, S. I. (2025). *Experimental Study of Flame Combustion of Pulverized Coal Fuel in a Two-Stage Vortex Burner*. *34*(1). https://doi.org/10.1134/S1810232825010011

Haywood, R. W. (1991). Analysis of Engineering Cycle 4th. Edition. *Pergamon Press*.

Jakoubek, P., & Hanzal, L. (1990). Matlab As a Tool for Fossil Fuel Power Plant Combustion Control. *Communication*.

Kikstra, J. S., Nicholls, Z. R. J., Smith, C. J., Lewis, J., & Lamboll, R. D. (2022). *The IPCC Sixth Assessment Report WGIII climate assessment of mitigation pathways : from emissions to global temperatures*. 9075–9109.

Kim, D. W., Lee, J. M., Lee, G. H., & Park, K. Il. (2024). *Experimental Study of Oxy-fuel Combustion and Emission Characteristics using a 10 kWth Pressurized Fluidized Bed Combustor*. 0–22.

Korkmaz, A. A. (2022). Farklı Linyitlerin Kısa ve Elementel Analiz Verilerine Dayanarak Üst Isıl Değerlerinin Hesaplanması. *Konya Journal of Engineering*

Sciences, *8055*, 49–60. https://doi.org/10.36306/konjes.869637

Korukcu, M. Ö. (2024). A Graphical User Interface for Calculating Exergy Destruction for Combustion Reactions. *Processes*, *12*(294).

Kurose, R. (2009). *Numerical Simulations of Pulverized Coal Combustion*. *27*(27), 144–156.

Kuznetsov, V. A., Bozheeva, D. M., Dekterev, A. A., & Minakov, A. V. (2024). *Computational study of the processes of pulverized coal oxygen combustion in a vortex burner *. 31*(3), 523–530. https://doi.org/10.1134/S0869864324030156

Luo, J., Emelogu, O., Morosuk, T., & Tsatsaronis, G. (2021). Exergy-based investigation of a coal- fi red allam cycle. *Energy*, *218*, 119471. https://doi.org/10.1016/j.energy.2020.119471

Özer, B. (2019). *A Study on Coal Combustion: Experiments and Modelling*. *December*.

Ozsari, I., Ust, Y., & Kayadelen, H. K. (2021). Comparative Energy and Emission Analysis of Oxy-Combustion and Conventional Air Combustion. *Arabian Journal for Science and Engineering*, *46*(3), 2477–2492. https://doi.org/10.1007/s13369-020-05130-0

Pratt, D. T. h*ttps://web.archive.org/web/20150521041644/http://astronautics.usc.edu/utility/flame_temperature.php*.

Yazar, O., & Demir, B. (2022). Development Of the Software Program That Can Account the Combustion, Emission and Engine Performance Values of Internal Combustion Engines. *European Journal of Technic*, *12*(2), 129–138. https://doi.org/10.36222/ejt.1147020

Yıldız, M. (2023). Chemical equilibrium based combustion model to evaluate the effects of H2 addition to biogases with different CO2 contents. *International Journal of Hydrogen Energy*, *52*, 1334–1344. https://doi.org/10.1016/j.ijhydene.2023.06.077

TCDC: Thermal Circuit Designer and Calculator

M. Özgün KORUKÇU[1]

1- Prof. Dr.; University of Bursa Uludağ, Faculty of Engineering, Department of Mechanical Engineering. ozkorukcu@uludag.edu.tr ORCID No: 0000-0002-4761-4304

ABSTRACT

This chapter introduces the *TCDC Thermal Circuit Designer and Calculator tool*, developed for the graphical modeling and analysis of conduction and convection-based thermal resistance networks. The tool employs the electrical circuit analogy to perform node-based analysis of thermal systems and allows users to construct complex thermal circuits through a drag-and-drop interface. Various thermal elements—including planar walls, cylindrical and spherical conduction resistances, convection resistances, and fixed-temperature sources (T_h, T_c) can be defined parametrically; the software automatically determines the circuit topology, calculates the equivalent thermal resistance (R_{eq}), node temperatures, and the total heat transfer rate (Q).

A key capability of TCDC is the Q+ inverse-problem solver, which computes the value of a single parameter required to achieve a target heat transfer. Through this feature, properties such as material thermal conductivity, wall thickness, convection coefficient, or source temperatures can be determined from a desired Q value. The tool also incorporates advanced interaction features such as automatic node labeling, temperature tooltip updates, undo support, rotation, duplication, and port visibility toggling.

This chapter details the software architecture, solution algorithms, GUI components, and example engineering applications, highlighting the contribution of the thermal-circuit approach to education, research, and design workflows.

Keywords – Thermal Circuit, Heat Transfer, Heat Conduction, Heat Convection, Thermal Resistance.

INTRODUCTION

Thermal analysis plays a central role in many branches of engineering, including energy systems, electronics cooling, building physics, materials engineering, and industrial process design. Engineers routinely encounter problems where accurate prediction of temperature distributions and heat-transfer rates is essential for ensuring system performance, durability, and safety. While numerical approaches such as finite element or finite volume methods offer high precision, they often require significant computational resources, specialized software, and expert knowledge. For preliminary design, rapid assessments, conceptual understanding, and educational purposes, such heavy methods may not always be ideal.

For these reasons, thermal circuit modeling has emerged as an effective intermediate approach that balances physical accuracy with computational efficiency. In this method, conduction, convection, and sometimes radiation

pathways are represented as resistances, while temperature points are represented as nodes. This formulation directly mirrors electrical circuit theory—an analogy that has long served as a powerful tool for understanding and simplifying thermal systems. By applying nodal analysis, engineers can estimate heat-flow rates and temperature distributions quickly, even in relatively complex systems. The intuitive nature of the electrical analogy also makes thermal circuit modeling highly accessible to students and practitioners with diverse backgrounds.

Dang et al. (Dang et al., 2024), developed a detailed thermal resistance network model for flip chip ball grid array (FCBGA) packages by combining one-dimensional conduction resistances with spreading resistances obtained from COMSOL simulations. Their three-dimensional package model enables calculation of individual thermal resistance components along multiple heat-dissipation paths. The approach predicts junction temperatures with high accuracy, showing only 0.24% deviation under varying power loads. The study also highlights major thermal bottlenecks—particularly spreading resistance at the heat sink floor and non-uniform heat flux at the die—providing practical guidance for thermal optimization.

Duan et al. (Duan et al., 2022) proposed an efficient transient thermal model for multilayer building walls using an resistance and capacitance (RC) thermal network formulation. Their analytical approach is validated with in-situ wall temperature measurements, showing good agreement with errors generally under 6%. The study convincingly demonstrates that the simplified circuit model captures attenuation and phase-lag behavior while significantly reducing computational effort. Overall, the article offers a practical and reliable alternative to more complex transient heat-transfer methods.

Hao et al. (Hao et al., 2014) developed a simplified thermal resistance network to model heat transfer and pressure drop in serpentine-channel liquid-cooled heat sinks. The approach breaks the heat sink into conduction and convection resistance elements and incorporates a flow-network representation for channel temperature rise. Validation against full 3D CFD simulations shows very good agreement, while reducing computation time dramatically. The authors also examine Reynolds number and channel aspect ratio, concluding that Reynolds number has the strongest influence on both cooling performance and hydraulic losses. Overall, the study demonstrates that thermal-network modeling offers an efficient and accurate tool for preliminary heat-sink design.

Huang (Huang, 2007) presents a rigorous compact thermal modeling framework, HotSpot, which formulates chip and package heat-transfer behavior as a parameterized lumped R–C network derived from stacked-layer packaging geometries. The methodology systematically incorporates lateral heat spreading, transient thermal capacitances, and interconnect self-heating, thereby addressing key deficiencies in earlier compact models that relied on oversimplified package assumptions or coarse spatial granularity. Validation

against finite-element simulations, thermal test chips, and FPGA-based prototypes confirms that HotSpot achieves high fidelity while maintaining computational efficiency suitable for early design-stage analysis. As a result, the work provides a scientifically robust foundation for temperature-aware very large scale integration (VLSI design, including leakage–temperature coupling, dynamic thermal management, and reliability assessment.

Kirchgässner et al. (Kirchgässner et al., 2023) introduce the Thermal Neural Network (TNN), a hybrid modeling framework that embeds heat-transfer principles from lumped-parameter thermal networks (LPTN into a state-space machine-learning architecture. The study shows that TNNs achieve higher temperature-estimation accuracy than both conventional LPTNs and black-box neural models, while maintaining physical interpretability and requiring minimal expert or geometric input. Extensive experiments on a permanent-magnet synchronous motor dataset demonstrate that small TNNs (≈64 parameters) can outperform more complex models with significantly lower computational cost. Overall, the work provides a scientifically grounded and computationally efficient approach for real-time thermal monitoring in power-electronic and electromechanical systems.

Mannapperuma et al. (Mannapperuma et al., 2025) presented a comprehensive electro-thermal modeling framework for an electric vehicle (EV) lithium-ion battery pack by integrating a first-order equivalent circuit model with a four-zone lumped-parameter thermal network. The model is distinguished by its exclusive use of vehicle-level CAN data—rather than laboratory measurements—for identifying both electrical and thermal parameters, enabling practical deployment under real driving conditions. Validation across 31 datasets (summarized in Table 3 on page 15) demonstrates strong accuracy, with terminal-voltage RMSE below 2% and battery-temperature errors under 2 °C for most cases. Overall, the study provides a cost-effective and experimentally grounded method for predicting pack-level thermal behavior, highlighting its value for real-time battery management and EV thermal-control strategies.

Merrikh (Merrikh, 2015) develops a compact thermal modeling framework for predicting the transient surface temperature of passively cooled, fanless electronic devices using Foster-type RC thermal networks. The study carefully benchmarks the RC models against full 3D transient CFD simulations, demonstrating that a two-ladder RC network can reproduce hot-spot temperature evolution with errors generally below 5%, while a single-ladder model performs poorly. A key finding is that thermal resistance and time-constant parameters are strongly power-dependent due to nonlinear natural-convection and radiation effects, necessitating a variable-coefficient formulation. Overall, the work provides an efficient and accurate alternative to CFD for real-time thermal prediction in thin, passively cooled mobile devices.

Uhlmann et al. (Uhlmann et al., 2020) developed an energy-self-sufficient thermoelectric self-cooling (TSC) system for linear direct drives and model its thermal–electrical behavior using an extended thermal resistance network implemented in MATLAB/Simulink. The model integrates conduction, convection, and thermoelectric effects (including Seebeck, Peltier, and Joule contributions) and is validated through a detailed test-bench experiment, showing high prediction accuracy for both steady-state and transient thermal loads. As demonstrated in the results on pages 49–54, the system adaptively regulates coolant flow without closed-loop control, maintaining stable temperature differences and generating sufficient electrical power to operate the cooling components autonomously. Overall, the study provides a robust modeling framework and experimental evidence supporting TSC systems as an efficient and energy-saving alternative to conventional machine-tool cooling architectures.

Xu and Nikovski (Xu & Nikovski, 2016) proposed a grey-box building thermal model that couples a third-order thermal circuit with an enhanced Building Research Establishment (BRE) admittance humidity model to simultaneously predict indoor temperature and humidity with a 5-minute resolution. The model accounts for key factors such as ventilation, human activity, moisture absorption/desorption, condensation, and HVAC operation, while maintaining computational efficiency through linearization and a state-space formulation. Laboratory and field tests demonstrate that the enhanced model significantly improves humidity-prediction accuracy—reducing mean absolute percentage error from ~7% to below 2%—while retaining reliable temperature predictions. Overall, the study delivers a practical and accurate short-term indoor climate prediction framework suitable for comfort control and HVAC optimization.

Despite its advantages, traditional thermal circuit modeling has often been underutilized due to the lack of user-friendly tools. Many existing software environments require manual equation setup, lack interactivity, or do not support modern features such as automatic node recognition or mixed conduction–convection networks. In educational settings, especially, GUI-based tools significantly enhance learning by enabling students to visualize thermal paths, experiment with parameters, and immediately observe the effect of design choices. In research and design, interactive interfaces accelerate prototyping and facilitate communication of ideas through clear and adaptable visual diagrams.

Motivated by these needs, the TCDC Thermal Circuit Designer and Calculator was developed as a modern, extensible, and interactive environment for creating and solving thermal resistance networks. TCDC combines a drag-and-drop graphical interface with an automated computational core capable of analyzing arbitrary topologies, labeling nodes, and dynamically updating temperature distributions. It supports building multi-element conduction walls, convection interfaces, cylindrical and

spherical resistances, and temperature sources, all within a unified interface. Furthermore, the tool introduces the Q+ inverse-design module, which enables users to determine the value of a single parameter required to achieve a specified heat-transfer rate—an essential capability for material selection, insulation design, and system optimization.

This chapter aims to provide a comprehensive overview of the theoretical foundations and practical capabilities of TCDC. It covers the principles of thermal circuit modeling, the mathematical basis of nodal analysis, the architecture of the software, and the functionalities of the graphical user interface. In addition, several example applications are presented to demonstrate how TCDC can be used effectively in engineering analysis, classroom instruction, and early-stage design workflows.

THEORY OF THERMAL CIRCUIT

The thermal circuit method utilizes the well-known analogy between heat transfer and electrical networks, representing temperature as a potential and heat flow as an analogous electrical current. In this framework, individual heat-transfer mechanisms are modeled as thermal resistances. Series thermal resistances arise when heat flows sequentially through multiple layers or components, with the total resistance obtained by summation (Kılıç & Yiğit, 2024).

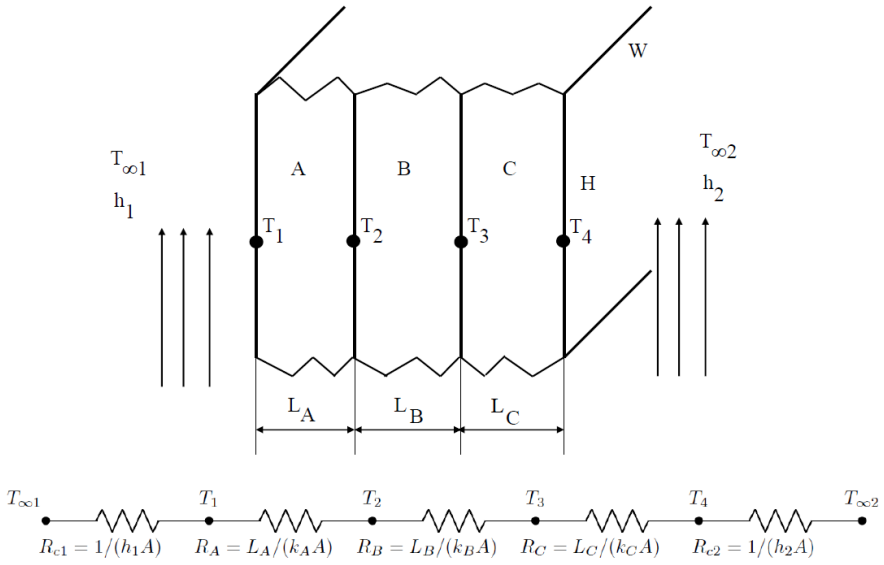


Figure 1: Thermal network for composite wall

Consider a composite wall of H height and W width with fluid flow on both sides that as shown in Figure 1. Ambient temperatures are $T_{\infty 1}$, $T_{\infty 2}$ and heat transfer coefficient values on the wall surfaces are h_1 and h_2, respectively. The wall consists of A, B, and C layers. The lengths of the layers are L_A, L_B, and L_C. The thermal conductivity values of the layers are k_A, k_B, and k_C. The thermal resistances of heat convection and heat conduction can be found with appropriate formulas that are presented in Table 1. The equivalent resistance of the system is the sum of all resistances in the thermal network.

$$Q = \frac{\Delta T}{\sum R} = \frac{T_{\infty 1} - T_{\infty 2}}{R_{c1} + R_A + R_B + R_C + R_{c2}}$$

Since the heat transfer rate through the resistances is constant. Nodal temperatures can be found as:

$$Q = \frac{\Delta T}{\sum R} = \frac{T_{\infty 1} - T_1}{R_{c1}} = \frac{T_1 - T_2}{R_A} = \frac{T_2 - T_3}{R_B} = \frac{T_3 - T_4}{R_C}$$

Or

$$Q = \frac{\Delta T}{\sum R} = \frac{T_{\infty 1} - T_2}{R_{c1} + R_A} = \frac{T_2 - T_4}{R_B + R_c} \ldots$$

Table 1: Conduction and convection resistance formulas for planar wall, cylinder, and spherical systems.

Geometry	Conduction Resistance	Convection Resistance	Heat Transfer Surface Area
Planar Wall	$R = \frac{L}{kA}$	$R = \frac{1}{hA}$	$A = HW$
Cylinder	$R = \frac{\ln(r_2/r_1)}{2\pi kL}$	$R = \frac{1}{hA}$	$A = 2\pi rL$
Sphere	$R = \frac{(r_2 - r_1)}{4\pi k r_1 r_2}$	$R = \frac{1}{hA}$	$A = 4\pi r^2$

Parallel thermal resistances appear when multiple paths for heat flow exist simultaneously, in which case the inverse of the equivalent resistance is the sum of the inverses of each branch. This abstraction allows engineers to reduce complex, multi-layer thermal systems into manageable network structures.

$$\frac{1}{R_{eq}} = \sum_{1}^{i} \frac{1}{R_i}$$

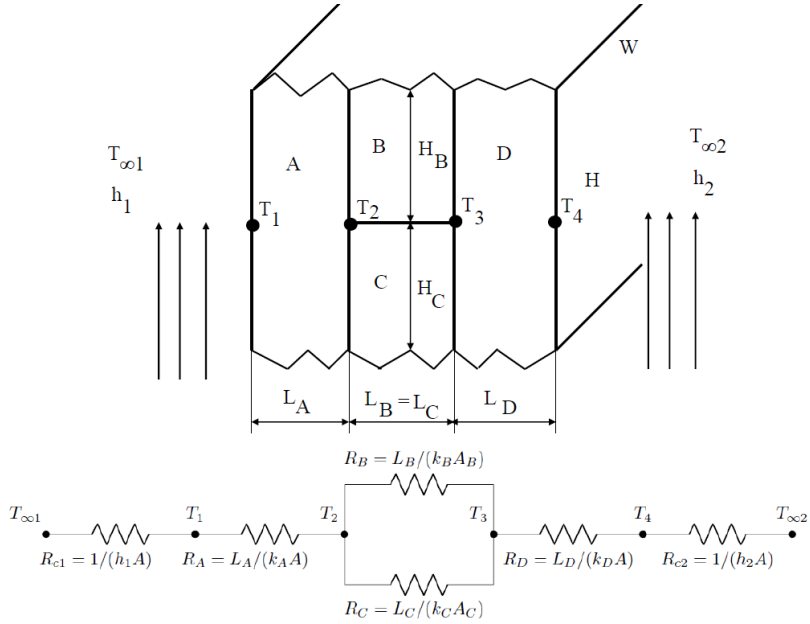


Figure 2: Thermal network for composite wall with parallel resistances

In Figure 2, a composite wall containing two layers placed on top of each other is shown in the middle of the wall. In this problem, the equivalent resistance of the B and C layers should be:

$$\frac{1}{R_{eq}} = \frac{1}{R_B} + \frac{1}{R_C}$$

Where:

$$A_B = H_B W \text{ and } A_C = H_C W$$

The total thermal resistance of the system would be:

$$\sum R = R_{c1} + R_A + R_{eq} + R_D + R_{c2}$$

INTRODUCTION OF TCDC

In this section, the layout of TCDC GUI will be presented. TCDC is developed with Python and PyQt5. In Figure 3, the main screen of the program can be seen. On the left panel, Thot, Tcold temperatures, and resistances for wall, cylinder, sphere, and convection can be seen. Over the canvas, select, wire, heat transfer calculation button, set target button, Q+ Solve button, and hide ports button can be seen. Desired resistances can be dragged and dropped into the canvas, wire button is used to connect resistances. After building the thermal circuit, to calculate heat transfer "Calculate heat transfer (Q)" button should be pressed. The main screen of the software can be seen in Figure 3.

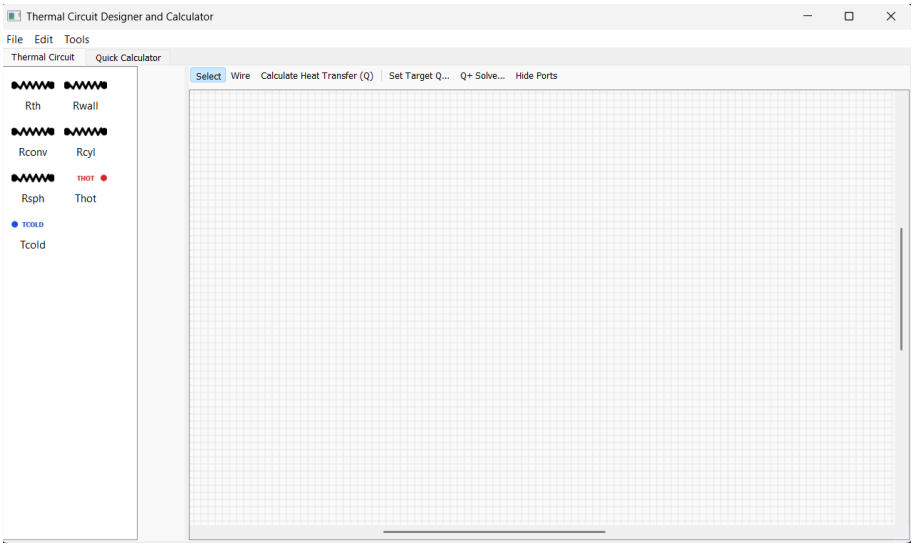


Figure 3: The main screen of TCDC

There is also a Quick Calculator button that allows users to calculate all parameters in heat transfer equation, like temperature on the left or right side of resistances, heat transfer, and resistance values. Figure 4 shows the Quick Calculator button features.

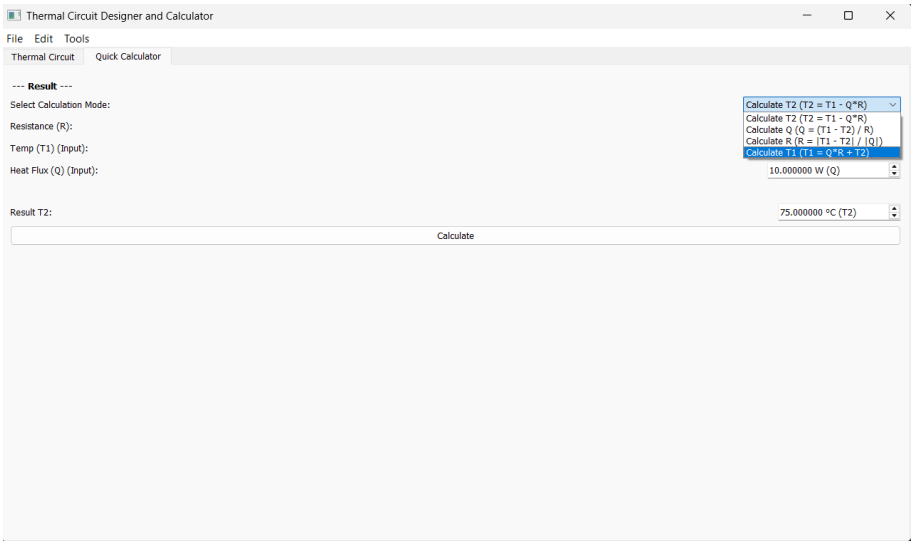


Figure 4: Quick calculator button features

The “File” button is used for opening a new project, saving the existing project, or loading a project that was saved before. The user can also export the thermal circuit in PNG or JPEG format. The “Edit” button is used

to copy and paste the resistances. The “Tools” button has the same features as the buttons that are placed on the canvas. Figure 5 shows the File (a) and Tools (b) buttons of the software.

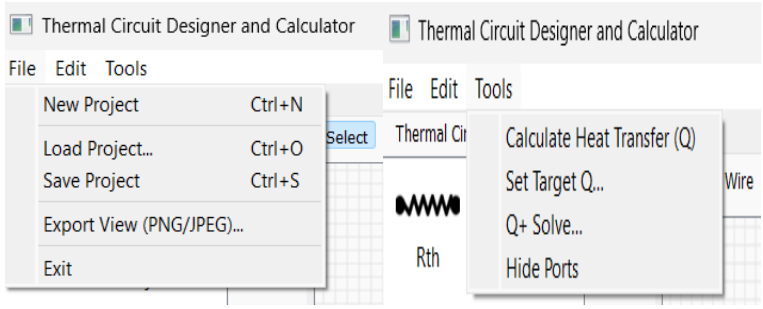


Figure 5: The File (a) and Tools (b) buttons.

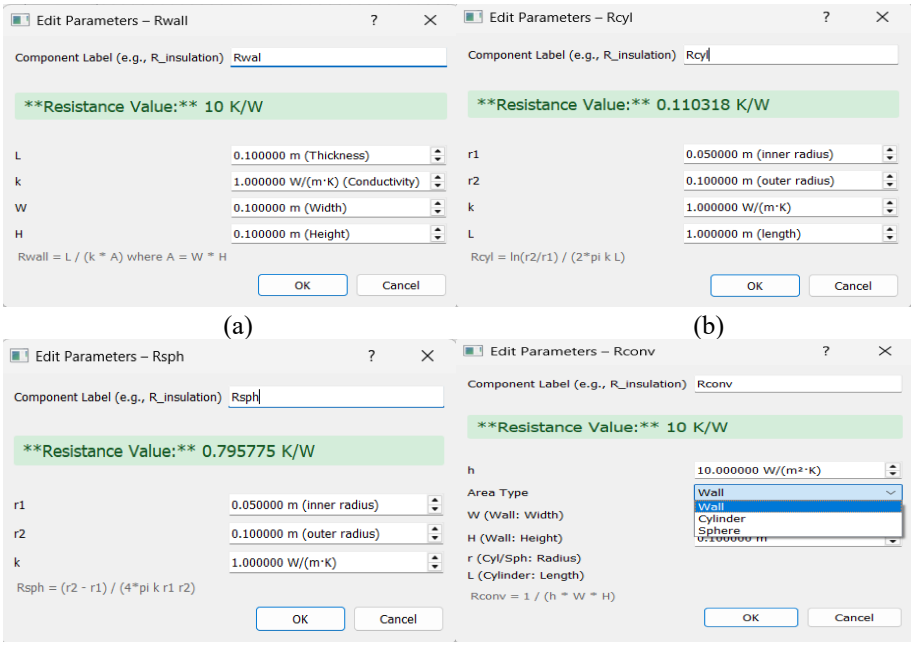


Figure 6: Calculation of thermal resistances (a) Wall, (b) Cylinder, (c) Sphere, and (d) Convection resistance

Once the desired resistance is dragged and dropped into the canvas, parameters can be assigned to calculate the resistance values. Figure 6 shows the calculation of thermal resistances for walls, cylinders, spheres, and convection. For convection resistance, there is also a small combobox that changes the formulas for the desired geometry.

To find any parameter for a different heat transfer value, the user first assigns a new heat transfer value with the “Set Target Q” button and then chooses the desired resistance and parameter in the chosen resistance. After setting the target heat transfer value, the “Q+ Solve” button should be used. The new parameter, equivalent resistance, and heat transfer value will be prompted on the screen.

APPLICATIONS

In this section, applications of TCDC will be presented.

The first example is in cylindrical coordinates and is as follows:

A cast-iron pipe (k_{pipe}=52 W/mK) with an inner diameter of 12 cm has a wall thickness of 2 cm and a length of 10 m. The pipe is insulated with 1.5 cm of glass wool ($k_{glass\ wool}$=0.04 W/mK) on the outside. Steam at 80 °C flows inside the pipe, and the internal convective heat transfer coefficient is 50 W/m²·K. The ambient temperature outside the pipe is 15 °C, and the interface temperature between the pipe and the glass wool is 75.6437 °C. Determine the convective heat transfer coefficient of the outside environment. The problem is sketched in Figure 7.

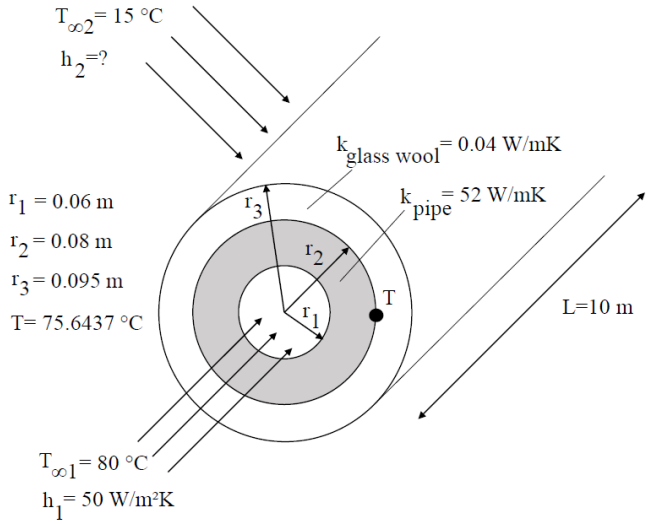


Figure 7: An insulated pipe with convective surface conditions

The problem will be solved in two steps. In the first step, heat transfer will be found with the given interface temperature, and then, by using the same heat transfer value, the outer heat transfer coefficient will be found. Figure 8 shows the first step, and Figure 9 shows the results of the first step.

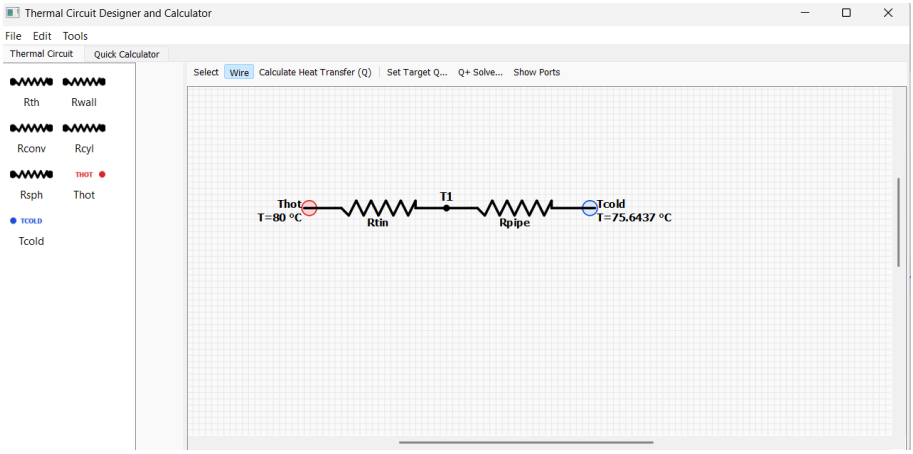


Figure 8: The first step of the problem

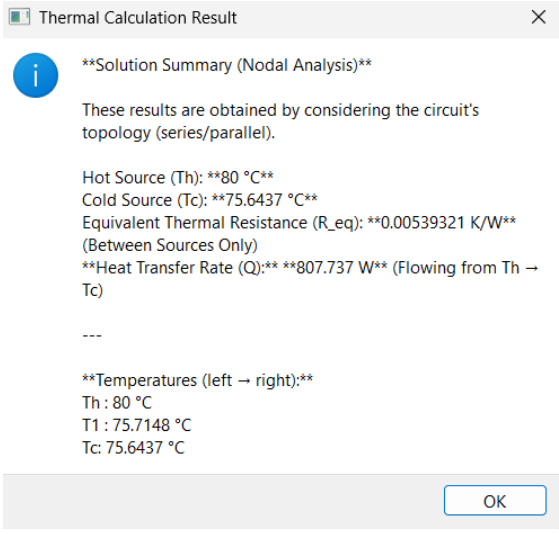


Figure 9: The results of the first step

As seen from Figure 9, the heat transfer value of the problem is 807.737 W.

In the second step, two more resistances should be added to the system, and then Rout should be clicked and checked as “Solve this R from Target Q” and adjust the parameter as the heat transfer coefficient, h. After setting the target heat transfer, the user should click the “Q+ Solve” button. Figure 10

shows the thermal circuit and the adjustment of the resistance parameter. Figure 11 shows how to set the target heat transfer. Figure 12 shows the solution of the problem for the heat transfer coefficient (a) and the nodal temperature values on the thermal circuit (b).

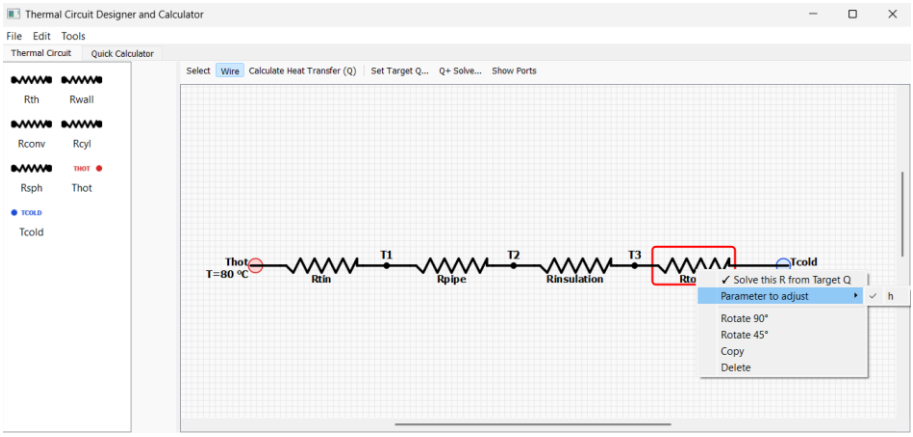


Figure 10: Thermal circuit of the second part of the problem and the adjustment of the resistance parameter

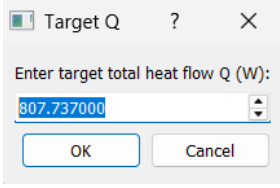
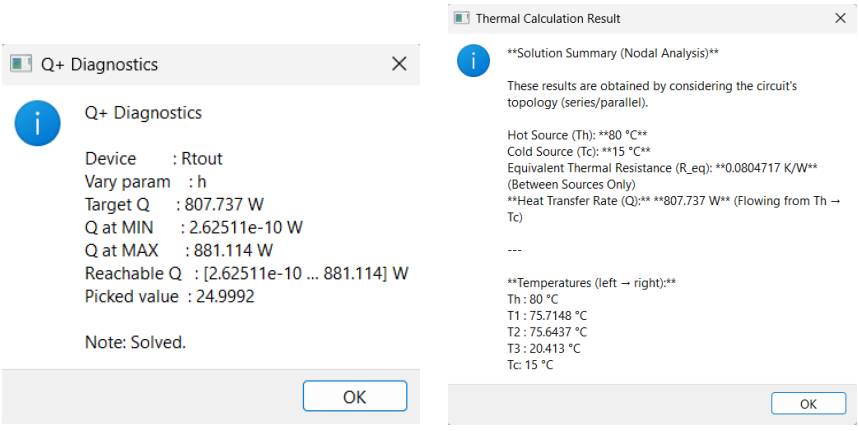


Figure 11: Setting the target temperature



(a) (b)

Figure 12: The results of the first problem (a), and the nodal temperature values of the thermal circuit in the second step (b)

As it is seen from Figure 12(a), the user can not set an arbitrary heat transfer when setting a new target. The software gives maximum and minimum target heat transfer values according to the equivalent resistance value of the thermal circuit. According to the results outer heat transfer coefficient should be 25 W/m²K.

The second problem contains parallel resistances in the thermal circuit, and the thermal conductivity of one of the layers that comprises the parallel resistance is asked for.

The components of a composite wall are shown in Figure 13. Given that the interface temperature between layers B and C and layer D is measured as 83.74132 °C, determine the thermal conductivity of layer C.

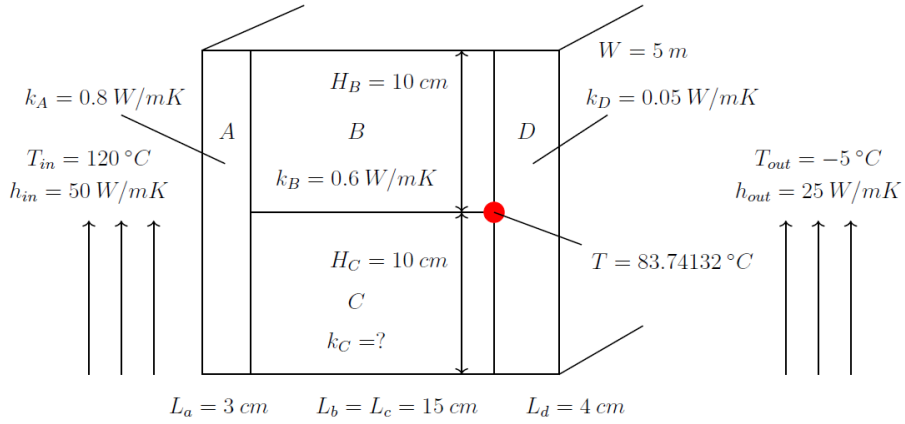


Figure 13: Sketch of the second problem

This problem will also be solved with two steps. In the first step, heat transfer will be found by using the given interface temperature with using conduction resistance of layer D and the outer convection resistances. Figure 14 shows the thermal circuit of the first part, and the heat transfer is found as 105.644 W.

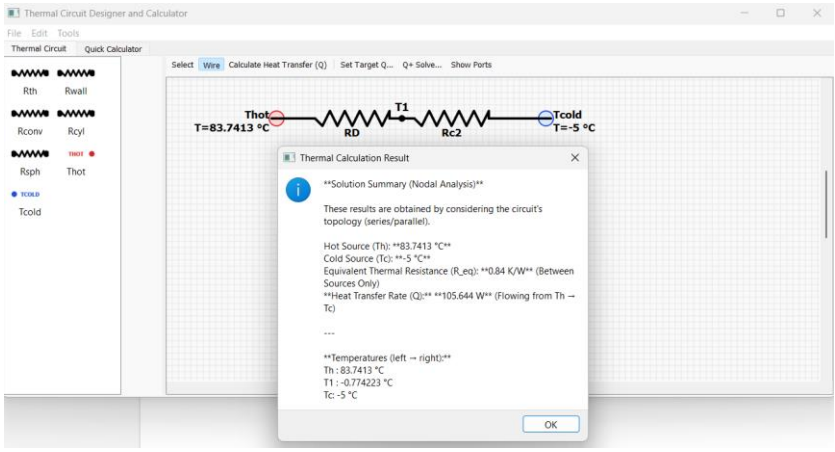


Figure 14: Thermal circuit of the first part of the problem and the heat transfer calculation results

In the second part, the resistances on the left side of layer D should be added by dragging and dropping and connecting them to the first part. Figure 15 shows the thermal circuit of the second part of the problem and the adjustment of the thermal conductivity value of resistance Rc.

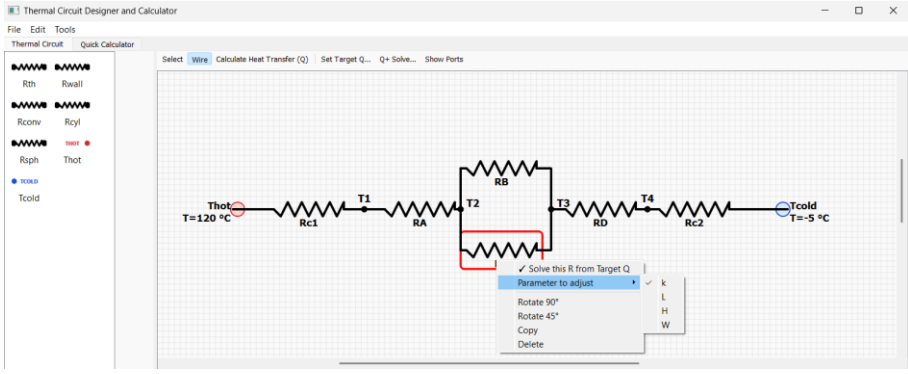


Figure 15: Thermal circuit of the second part of the problem and adjustment of the thermal conductivity

After setting the target Q as 105.644 W and clicking the “Q+ Solve” button, the thermal conductivity of layer C can be found. Figure 16 shows how to set the target heat transfer for this problem. Figure 17 shows the results of the problem (a) and the nodal temperature values (b). As seen from Figure 17-a, the varying parameter of the chosen resistance is written in the results. According to the results, the thermal conductivity of layer C should be 0.45 W/mK.

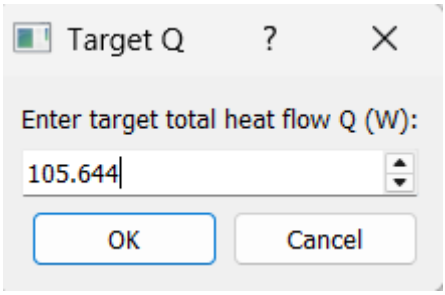
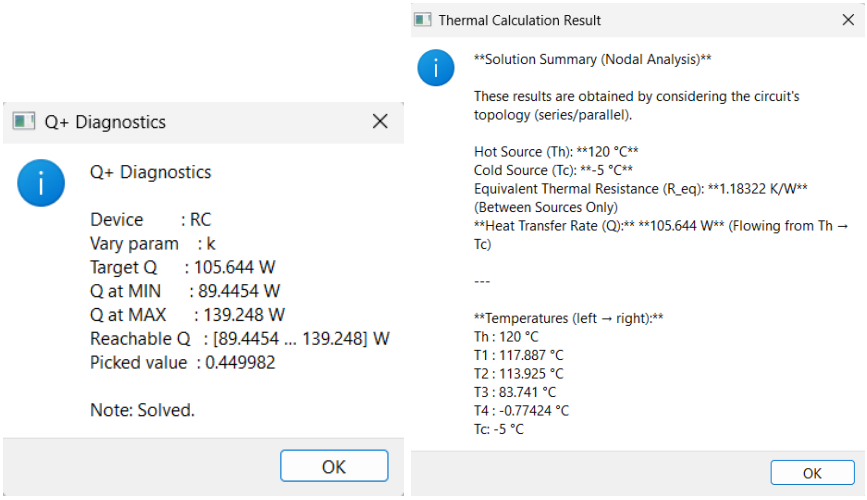


Figure 16: Setting the target temperature for the second problem



(a) (b)

Figure 17: The results of the second problem (a), and the nodal temperature values of the thermal circuit in the second step (b)

CONCLUSION

A user-friendly and easy to learn graphical user interface (GUI) software for one-dimensional steady-state heat transfer of conduction and convection by using the thermal circuit concept has been developed using Python and PyQt5. The drag-and-drop feature of the software allows users to build their own thermal circuit. Heat transfer of composite walls, insulated pipes, or spherical systems can be found with serial and parallel resistances.

Another unique feature of the developed software is the “Q+ Solve” button that allows users to find any parameter in the thermal resistances by setting the target heat transfer value.

TCDC is fast and can be an alternative tool to complex FEM or CFD solvers.

TCDC can be used for heat transfer courses, rapid prototyping of thermal systems, where circuit-based systems are needed.

REFERENCES

Dang, H., Yang, L. U., Yanzheng, D. U., Xiu, Z., Qian, Z., & Weigang, M. A. (2024). *A Detailed Thermal Resistance Network Analysis of FCBGA Package*. *33*(1), 18–28.

Duan, J., Li, N., Peng, J., & Wang, C. (2022). Full-response model of transient heat transfer of building walls using thermoelectric analogy method. *Journal of Building Engineering*, *46*(November 2021), 103717. https://doi.org/10.1016/j.jobe.2021.103717

Hao, X., Li, X., Peng, B., Zhang, M., & Zhu, Y. (2014). *Thermal resistance network model for heat sinks with serpentine channels*. *August 2013*, 298–308. https://doi.org/10.1002/jnm

Huang, W. (2007). *HotSpot — A Chip and Package Compact Thermal Modeling Methodology for VLSI Design*. *January*.

Kirchgässner, W., Wallscheid, O., & Böcker, J. (2023). Engineering Applications of Artificial Intelligence Thermal neural networks : Lumped-parameter thermal modeling with state-space machine learning. *Engineering Applications of Artificial Intelligence*, *117*(October 2022), 105537. https://doi.org/10.1016/j.engappai.2022.105537

Kılıç, M., & Yiğit, A. (2024). *Isı Transferi*. Dora Yayıncılık. Bursa.

Mannapperuma, V., Gaddala, L. C., Zheng, R., Kim, D., Kim, Y., Ullal, A., Zhu, S., & Ha, K. P. (2025). *Electro-Thermal Modeling and Parameter Identification of an EV Battery Pack Using Drive Cycle Data*. 1–24.

Merrikh, A. A. (2015). Compact thermal modeling methodology for predicting skin temperature of passively cooled devices. *Applied Thermal Engineering*, *85*, 287–296. https://doi.org/10.1016/j.applthermaleng.2015.04.007

Uhlmann, E., Salein, S., Polte, M., & Triebel, F. (2020). MODELLING OF A THERMOELECTRIC SELF-COOLING SYSTEM BASED ON THERMAL RESISTANCE NETWORKS FOR LINEAR DIRECT DRIVES IN MACHINE TOOLS. *Journal of Machine Engineering*, *20*(1), 43–57.

Xu, J., & Nikovski, D. (2016). *A Humidity Integrated Building Thermal Model*. 1492–1499.

A Simple Deep Learning Application for Apple Disease Classification

Nadir İLTEN[1]

1- Prof. Dr.; Balikesir University Faculty of Engineering Department of Mechanical Engineering. nilten@balikesir.edu.tr ORCID No: 0000-0003-4009-5078

ABSTRACT

Apple diseases have a substantial effect on both yield and fruit quality, highlighting the importance of early and precise detection to support sustainable agricultural practices. Recent advances in deep learning have demonstrated strong potential for automated plant disease detection using image-based analysis. This study presents a simple and effective deep learning application for apple disease classification implemented in the MATLAB environment. The proposed approach utilizes a GoogLeNet structure trained on labeled apple leaf images to classify common apple diseases, including apple scab, apple blotch, and apple rot, along with healthy samples. MATLAB's *Deep Learning Toolbox* is employed for data preprocessing, network design, training, and performance evaluation, providing a user-friendly and reproducible workflow. The experimental findings show that the proposed system attains acceptable classification accuracy with low computational demands, making it well suited for educational use and for practical implementation in resource-constrained environments. This work highlights the effectiveness of MATLAB as a rapid prototyping platform for developing deep learning-based agricultural applications and provides a foundation for more advanced plant disease diagnosis systems.

Keywords – Apple Disease, Machine Learning, Deep Learning, Classification, MATLAB.

INTRODUCTION

Apple (Malus domestica) production is significantly influenced by several major diseases that affect fruit quality, tree health, and overall orchard productivity. Among these, apple blotch, apple rot, and apple scab are three of the most common and economically important disorders encountered in apple-growing regions worldwide. Apple blotch (Wöhner & Emeriewen, 2019), typically caused by Marssonina coronaria, leads to dark leaf lesions and premature defoliation, weakening trees and reducing yield potential. Apple rot (Henriquez, 2005), a general term encompassing several fungal fruit rots such as black rot, bitter rot, and white rot, results in fruit decay both pre- and post-harvest, contributing to major economic losses. Apple scab (MacHardy, 1996), caused by Venturia inaequalis, remains one of the most persistent challenges in apple orchards, producing characteristic olive-brown leaf and fruit lesions that can deform fruit and significantly decrease market value. Effective management of these diseases requires understanding their infection cycles, environmental conditions favoring their development, and integrated control strategies involving resistant cultivars, cultural practices, and targeted fungicide applications. Because these

pathogens can survive year-round in plant debris or overwintering structures, orchard sanitation and seasonal monitoring are essential components of disease prevention. By recognizing the symptoms and epidemiology of apple blotch, rot, and scab, growers can improve disease prediction and implement timely interventions to maintain healthy orchards and ensure high-quality fruit production.

Machine learning is a fundamental branch of artificial intelligence concerned with creating computational systems that enhance their performance on specific tasks by learning from experience rather than relying on explicit programming. At its core, it focuses on developing algorithms capable of recognizing patterns, generating predictions, and extracting meaningful information from data. These algorithms construct mathematical models that capture underlying relationships within datasets, allowing them to generalize from known examples to previously unseen cases. The field of machine learning includes several learning paradigms, such as supervised, unsupervised, semi-supervised, and reinforcement learning. In supervised learning, models are trained on labeled data to address problems like classification and regression, whereas unsupervised learning aims to uncover latent structures in unlabeled data through techniques such as clustering and dimensionality reduction. Reinforcement learning emphasizes sequential decision-making, enabling agents to learn optimal behaviors by interacting with an environment and receiving feedback in the form of rewards. Driven by progress in computing hardware, the availability of large-scale data, and improvements in algorithmic methodologies, machine learning has become a cornerstone of modern technology. It plays a critical role in a wide range of applications, including computer vision, natural language processing, healthcare analysis, robotics, finance, and autonomous systems. Despite this progress, machine learning continues to face challenges, including issues of model interpretability, bias and fairness, data quality, and the need for robust generalization. Addressing these limitations remains a key focus of ongoing research, shaping the development of next-generation learning algorithms (Alpaydin, 2021; Paluszek & Thomas, 2016, 2020).

Deep learning is a specialized subfield of machine learning that has become a central approach for developing computational models capable of learning complex, nonlinear representations from large-scale datasets. Built upon artificial neural networks with multiple hidden layers, deep learning architectures employ hierarchical feature learning to automatically convert raw inputs into progressively more abstract and discriminative representations. This layered learning strategy reduces the reliance on manual feature design and allows models to capture the intricate structures present in high-dimensional data. The renewed prominence of deep learning

has been driven by three major factors: the availability of large, well-annotated datasets; advances in optimization techniques such as stochastic gradient descent and its variants; and rapid growth in computational power, particularly through GPU- and TPU-based acceleration. Collectively, these advances have made it possible to efficiently train deep, parameter-rich models (such as convolutional neural networks, recurrent neural networks, transformers, and generative architectures) that achieve state-of-the-art results across diverse applications. Deep learning has demonstrated exceptional effectiveness in domains characterized by complex statistical dependencies, including computer vision, natural language processing, speech recognition, biomedical signal analysis, and autonomous systems. Its success in these areas has not only transformed academic research but also driven significant industrial innovation. Despite these advancements, challenges remain, including issues related to model interpretability, robustness, data efficiency, and the computational cost associated with large-scale model training. Addressing these challenges forms the basis of ongoing research and continues to shape the future trajectory of deep learning methodologies (Kaur & Mohta, 2019; LeCun, Bengio, & Hinton, 2015; Paluszek & Thomas, 2020).

MATLAB (Matrix Laboratory) is a high-level numerical computing environment and programming platform widely used in engineering, scientific research, and industrial applications. Developed by MathWorks, MATLAB is designed to facilitate matrix-based computation, data analysis, algorithm development, modeling, simulation, and visualization within a unified and interactive framework. Its intuitive syntax and extensive built-in functions allow users to efficiently implement complex mathematical operations while maintaining clarity and readability in code development. One of MATLAB's key strengths lies in its rich ecosystem of domain-specific toolboxes, which extend its core capabilities to specialized areas such as signal processing, control systems, image processing, communications, robotics, and machine learning. These toolboxes provide pre-optimized algorithms, visualization tools, and workflow support that significantly reduce development time and enhance reproducibility. As a result, MATLAB has become a standard platform in both academia and industry for rapid prototyping and algorithm validation. In addition to numerical computation, MATLAB offers powerful modeling and simulation capabilities through Simulink, a graphical environment for multi-domain system design. Simulink enables users to model dynamic systems using block diagrams, perform simulations, and generate executable code for real-time and embedded hardware. This tight integration between MATLAB and Simulink supports a seamless transition from theoretical development to practical implementation. With built-in support for parallel computing, GPU acceleration, and code generation, MATLAB is well suited for handling

computationally intensive tasks and deploying algorithms to embedded systems, desktop applications, and cloud platforms. These features, combined with its ease of use and strong community support, make MATLAB a versatile and widely adopted tool for research, education, and industrial innovation (Beucher & Weeks, 2016; MathWorks, 2025).

Deep learning has become an essential methodology for building data-driven models capable of learning complex, hierarchical representations from large datasets. As the field continues to expand across domains such as computer vision, signal processing, natural language understanding, and autonomous systems, the need for accessible, reliable, and efficient development environments has grown correspondingly. MATLAB offers a cohesive and intuitive environment for the development, training, and deployment of deep learning models, supported by its strong numerical computing foundation and a wide range of specialized toolboxes. In particular, the *Deep Learning Toolbox* provides a flexible and comprehensive framework for building neural network architectures (including convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory networks (LSTMs), and transformer-based models) while accommodating both high-level workflows and detailed low-level customization. Its built-in functions simplify data preprocessing, model visualization, hyperparameter tuning, and performance evaluation. Furthermore, MATLAB facilitates GPU acceleration, automatic differentiation, and interoperability with external deep learning libraries through ONNX and Python integration, enabling researchers to combine MATLAB's productivity with the flexibility of broader deep learning ecosystems. In engineering, science, and industry, MATLAB's deep learning environment has proven effective for rapid prototyping, algorithm validation, and deployment to embedded systems, GPUs, FPGAs, and cloud services via tools such as MATLAB Coder, GPU Coder, and Simulink. This seamless transition from model design to real-world implementation makes MATLAB a valuable platform for deep learning research and application development. As deep learning continues to evolve, MATLAB remains a powerful tool for exploring advanced architectures, optimizing workflows, and accelerating innovation across a wide range of computational disciplines (Kim, 2017; Paluszek & Thomas, 2020).

This paper is organized as follows: The deep learning GoogLeNet structure and apple diseases are explained in Materials and Method section. The modeling process and the obtained results are presented in Results and Discussion section.

MATERIALS AND METHOD

In this study, a deep learning model was developed to identify three different apple diseases (apple blotch, apple rot and apple scab) and the healthy state of the apple. Sample images for the healthy state and the three different apple diseases are presented in Figure 1.

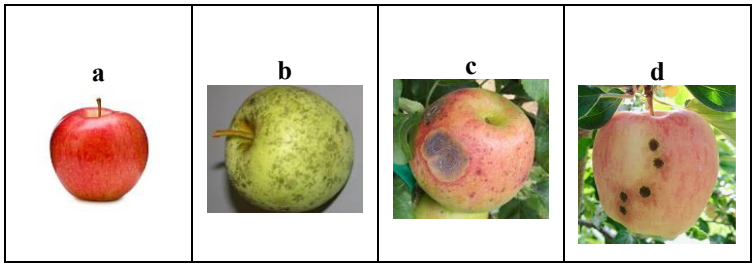


Figure 1: Apple disease: a) Healthy apple, b) Apple blotch, c) Apple rot, d) Apple scab.

A dataset consisting of 10 images was prepared for each condition. The prepared dataset is given in Figures 2, 3, 4, and 5. Although the limited dataset will negatively affect model accuracy, it is believed to be sufficient for a simple modeling study.

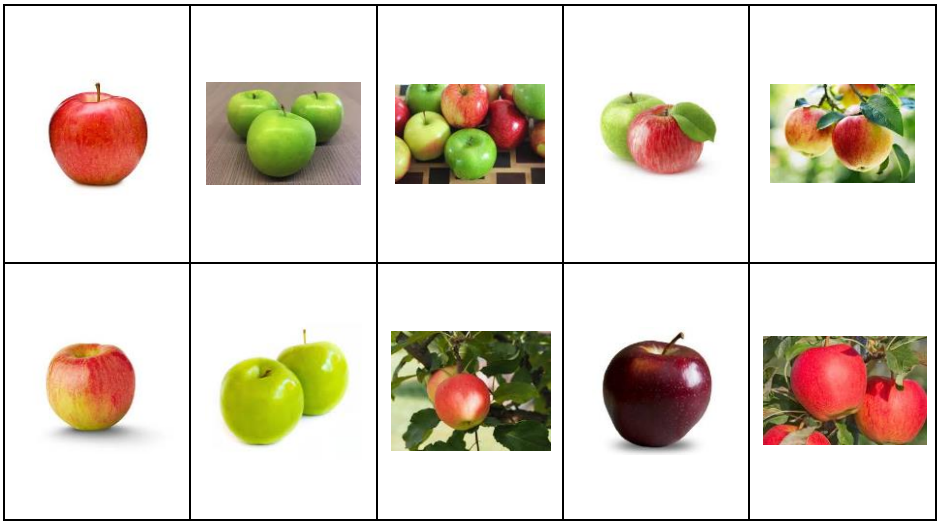


Figure 2: Dataset for healthy apple.

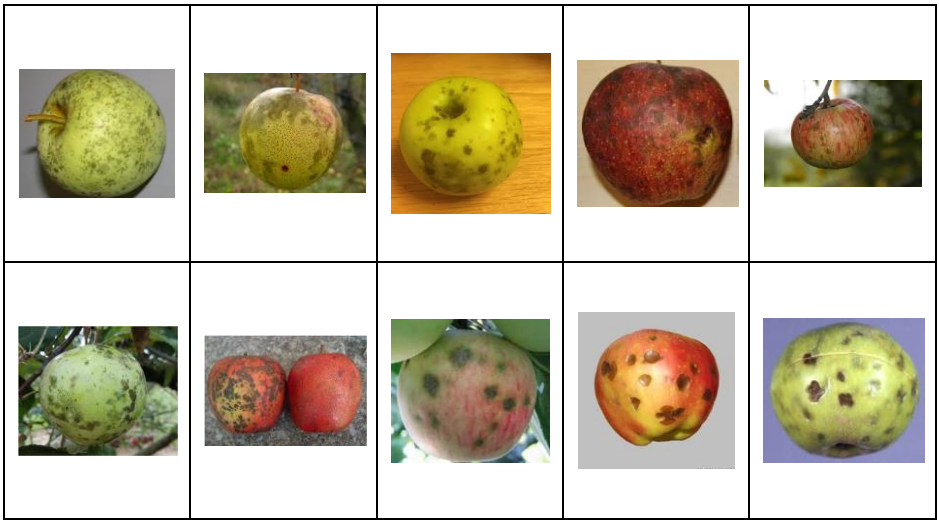


Figure 3: Dataset for apple blotch.

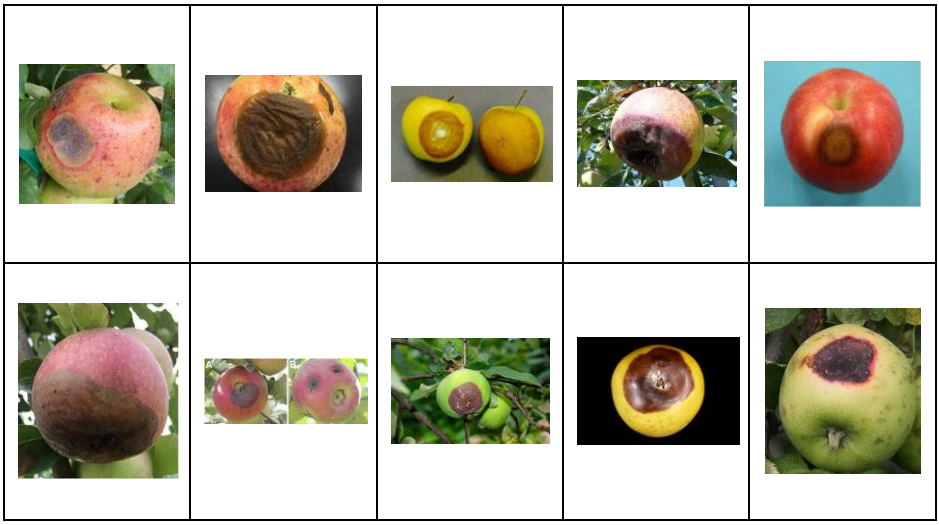


Figure 4: Dataset for apple rot.

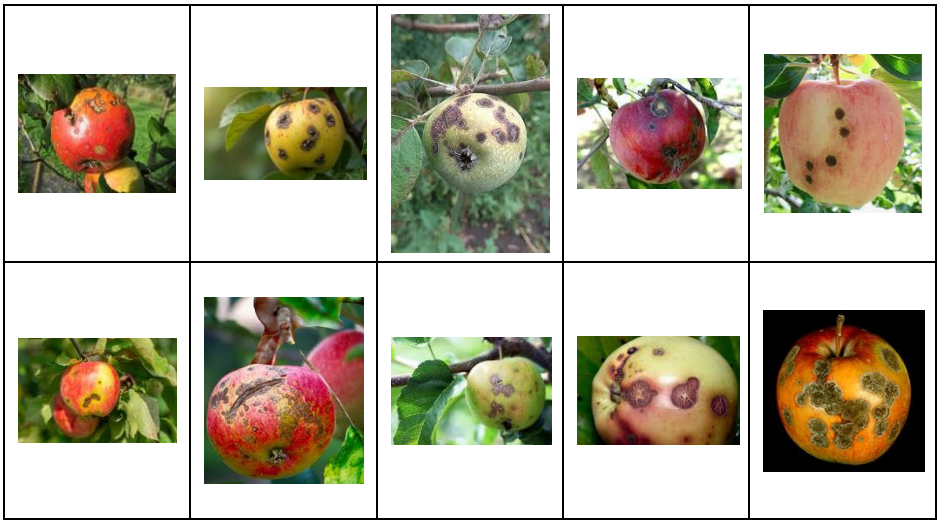


Figure 5: Dataset for apple scab.

Deep convolutional neural networks (CNNs) have achieved remarkable success in computer vision tasks such as image classification, object detection, and scene understanding. However, early CNN architectures faced significant challenges related to computational cost, memory usage, and overfitting as network depth increased. To address these limitations, the GoogLeNet architecture, also known as Inception v1, was introduced by Szegedy et al. in 2014 as the winning model of the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (Szegedy et al., 2015). GoogLeNet introduced a novel architectural concept called the Inception module, which enables the network to efficiently capture multi-scale visual features within a single layer. Instead of relying on a fixed convolutional filter size, the Inception module performs parallel convolutions using multiple kernel sizes (such as 1×1, 3×3, and 5×5), along with max pooling, and concatenates their outputs. This design allows the network to learn both local and global features while maintaining computational efficiency. A key innovation of GoogLeNet is the extensive use of 1×1 convolutions for dimensionality reduction. These layers significantly decrease the number of parameters and computational load, making it possible to construct a much deeper network without a proportional increase in complexity. As a result, GoogLeNet contains 22 learnable layers while having far fewer parameters compared to earlier architectures such as AlexNet and VGGNet. In addition to its efficient architecture, GoogLeNet incorporates auxiliary classifiers during training to mitigate the vanishing gradient problem and improve convergence. These auxiliary outputs act as regularizers, enhancing the network's generalization capability. Owing to these design choices, GoogLeNet achieves high classification accuracy while remaining suitable

for deployment in environments with limited computational resources. Due to its balance between accuracy and efficiency, GoogLeNet has been widely adopted in various applications, including medical image analysis (Subba & Sunaniya, 2025), plant disease classification (Yang et al., 2023), and embedded vision systems. Its compatibility with modern deep learning frameworks, such as MATLAB's *Deep Learning Toolbox*, further facilitates rapid development and experimentation. Consequently, GoogLeNet remains a foundational architecture that has influenced the design of many subsequent deep learning models (Al-Qizwini, Barjasteh, Al-Qassab, & Radha, 2017; Anand, Shanthi, Nithish, & Lakshman, 2019). Block diagram of the GoogLeNet is presented in Figure 6.

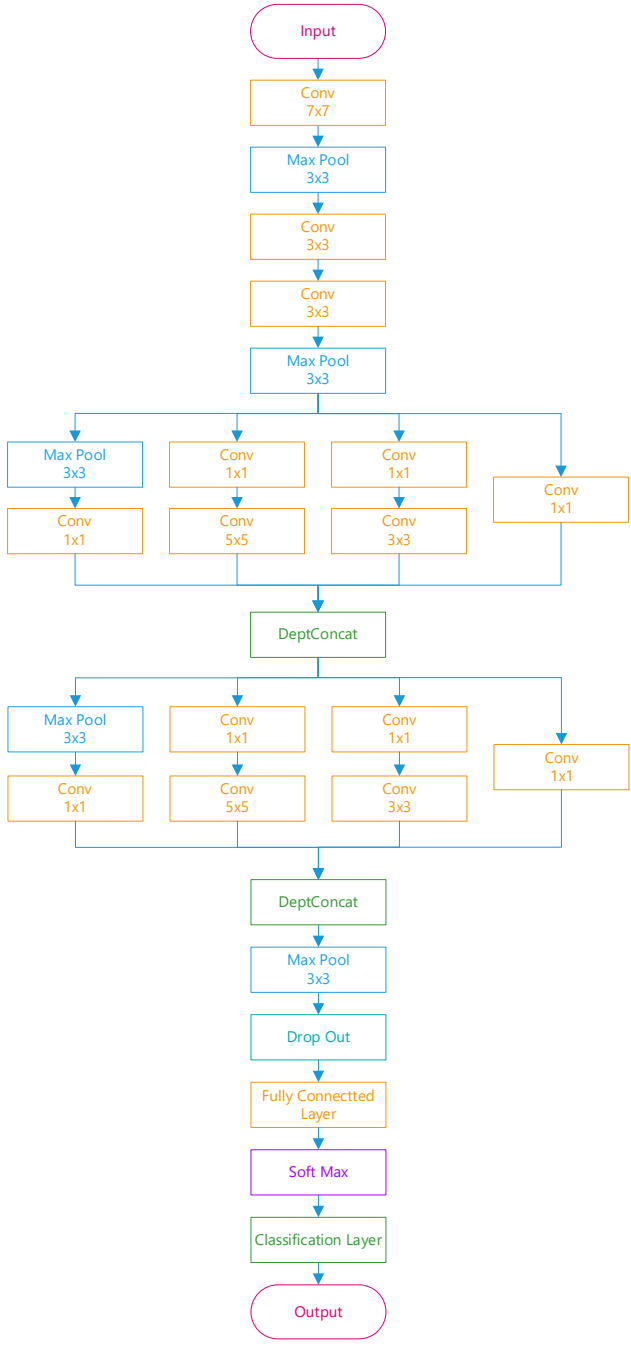


Figure 6: Block diagram of the GoogLeNet.

MATLAB's *Deep Learning Toolbox* was used to classify apple diseases with GoogLeNet. The options set for model training are presented in Table 1.

Table 1: Training options.

Method	SGDM
Momentum	0.9000
Initial Learn Rate	3.0000e-04
Max Epochs	100
Learn Rate Schedule	'none'
Learn Rate Drop Factor	0.1000
Learn Rate Drop Period	10
Mini Batch Size	5
Shuffle	'every-epoch'
Checkpoint Frequency	1
Checkpoint Frequency Unit	'epoch'
Sequence Length	'longest'
Preprocessing Environment	'serial'
L2 Regularization	1.0000e-04
Gradient Threshold Method	'l2norm'
Gradient Threshold	Inf
Verbose	0
Verbose Frequency	50
Validation Data	[1×1 augmentedImageDatastore]
Validation Frequency	8
Validation Patience	Inf
Objective Metric Name	'loss'
Checkpoint Path	''
Execution Environment	'auto'
Output Fcn	[ ]
Metrics	[ ]
Plots	'training-progress'
Sequence Padding Value	0
Sequence Padding Direction	'right'
Input Data Formats	"auto"
Target Data Formats	"auto"
Reset Input Normalization	1
Batch Normalization Statistics	'auto'
Output Network	'auto'
Acceleration	"auto"

RESULTS AND DISCUSSION

For the deep learning training process, the dataset was split into two subsets, with 70% allocated for training and the remaining 30% used for validation. Training with the *Deep Learning Toolbox* in MATLAB took 3 minutes and 50 seconds. A validation accuracy of 75.00% was achieved as a result of the training. The training process is presented in Figure 7.

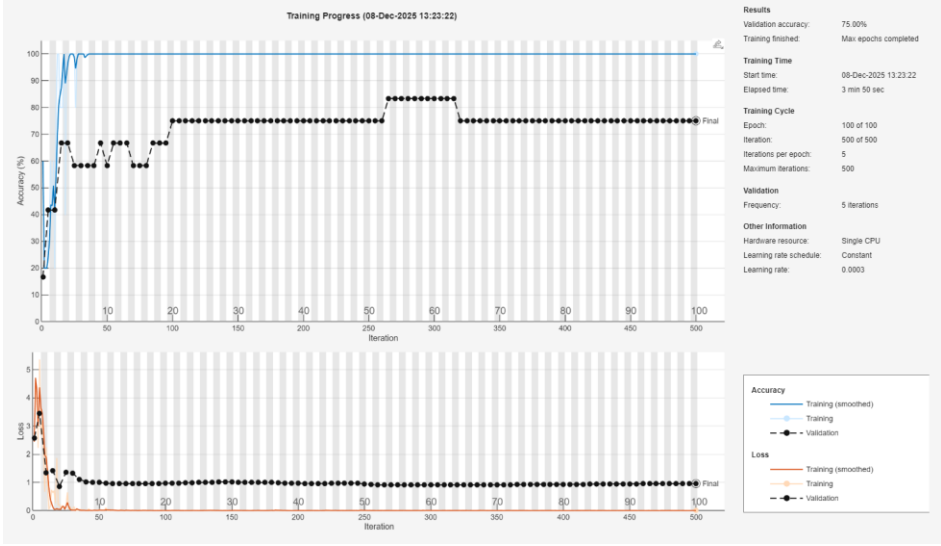


Figure 7: Training progress.

This study presented a simple yet effective deep learning–based approach for apple disease classification using the GoogLeNet architecture implemented in the MATLAB environment. By leveraging a convolutional neural network and image-based analysis, the proposed system successfully classified common apple diseases, including apple blotch, apple rot, and apple scab, alongside healthy samples. The use of MATLAB's *Deep Learning Toolbox* enabled an efficient and reproducible workflow for data preprocessing, network training, and performance evaluation. The experimental results demonstrate that GoogLeNet provides a strong balance between classification accuracy and computational efficiency, making it well suited for agricultural image analysis tasks. The use of transfer learning and a relatively lightweight architecture allowed the model to achieve satisfactory performance even with a limited dataset, highlighting its suitability for practical and educational applications. Furthermore, the simplicity of the proposed framework makes it adaptable to resource-constrained systems and real-world deployment scenarios.

REFERENCES

Al-Qizwini, M., Barjasteh, I., Al-Qassab, H., & Radha, H. (2017). Deep learning algorithm for autonomous driving using googlenet. *2017 IEEE Intelligent Vehicles Symposium (IV)*, 89–96. IEEE.

Alpaydin, E. (2021). *Machine learning*. MIT press.

Anand, R., Shanthi, T., Nithish, M. S., & Lakshman, S. (2019). Face recognition and classification using GoogleNET architecture. In *Soft Computing for Problem Solving: SocProS 2018, Volume 1* (pp. 261–269). Springer.

Beucher, O., & Weeks, M. (2016). *Introduction to MATLAB & SIMULINK: A Project Approach*. Infinity Science Press.

Henriquez, J. L. (2005). First report of apple rot caused by Neofabraea alba in Chile. *Plant Disease*, *89*(12), 1360.

Kaur, M., & Mohta, A. (2019). A review of deep learning with recurrent neural network. *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 460–465. IEEE.

Kim, P. (2017). Matlab deep learning. *With Machine Learning, Neural Networks and Artificial Intelligence*, *130*(21), 151.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

MacHardy, W. E. (1996). *Apple scab: biology, epidemiology, and management.*

MathWorks. (2025). *MATLAB documentation*. The MathWorks.

Paluszek, M., & Thomas, S. (2016). *MATLAB machine learning*. Apress.

Paluszek, M., & Thomas, S. (2020). MATLAB machine learning toolboxes. In *Practical MATLAB deep learning: A project-based approach* (pp. 25–41). Springer.

Subba, A. B., & Sunaniya, A. K. (2025). Computationally optimized brain tumor classification using attention based GoogLeNet-style CNN. *Expert Systems with Applications*, *260*, 125443.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., … Rabinovich, A. (2015). Going deeper with convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1–9.

Wöhner, T., & Emeriewen, O. F. (2019). Apple blotch disease (Marssonina coronaria (Ellis & Davis) Davis)–review and research prospects. *European Journal of Plant Pathology*, *153*(3), 657–669.

Yang, L., Yu, X., Zhang, S., Long, H., Zhang, H., Xu, S., & Liao, Y. (2023). GoogLeNet based on residual network and attention mechanism identification of rice leaf diseases. *Computers and Electronics in Agriculture*, *204*, 107543.

Investigation of The Effect of Shot Peening Process on The Mechanical and Surface Properties of Aluminum Materials

Ahmet DEMİRER

Sakarya Sakarya University of Applied Sciences Faculty of Technology
Department of Mechanical Engineering, ademirer@subu.edu.tr, ORCID: 0000-0003-1252-9203

INTRODUCTION

Shot peening is a cold working process used to improve the surface and subsurface properties of metallic materials. It is applied to increase the fatigue life of materials [1]. As a result of the repeated impacts of small-diameter shots projected at high velocities onto the material surface, a plastically deformed layer is formed [2]. This phenomenon is observed as grain refinement in the material microstructure, an increase in tensile strength, the formation of residual stresses on the surface, and an increase in hardness in the surface and subsurface regions. In the shot peening process, kinetic energy is transferred to the material through the shots. The Almen test was developed to measure this energy. Almen intensity varies depending on shot diameter, shot type, peening pressure, peening time, and surface coverage ratio [3]. As the Almen intensity increases, the thickness of the plastically deformed layer also increases [4]. Shot peening is an inexpensive and accessible method widely used in the automotive and aerospace industries.

Shot peening is a cold forming method created by controlled, repeated impacts of small shots onto the surface of a material. Each shot creates a small dimple on the peened material surface, acting like a hammer (Figure 1) [5].

Shot peening is used in the automotive and aerospace industries to increase surface hardness, reduce surface and subsurface grain size, prevent crack initiation, induce compressive residual stresses, improve microstructural properties, enhance fatigue strength, and eliminate internal porosity [1,6,7]. In the aerospace industry, shot peening is applied to engines and fuselage blocks, landing gear, propellers, and wings. In the automotive industry, it is frequently used for camshafts, crankshafts, cylinder blocks, gears, and springs operating under load [8].

Shot peening induces compressive residual stresses in the material. The surface and the region immediately beneath the surface affected by shot peening are subjected to compressive residual stresses. In contrast, tensile stresses are formed in regions farther from the surface that undergo elastic deformation. This effect increases the fatigue strength of the material [6]. As a result of the shot peening process, improvements occur in the mechanical properties of the material [9]. Shot peening is an economical and reliable process. The repeated impacts of shots on the surface increase material hardness due to changes occurring on and beneath the surface. Figure 2 illustrates the changes occurring in the peened material after shot peening.

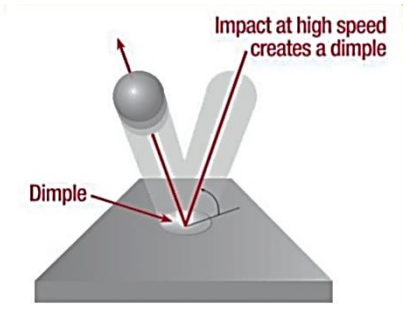


Figure 1. High-velocity impact effect of a shot during the shot peening process [5]

The most commonly used materials in shot peening applications are summarized in Table 1.

Table 1. Main Material Types Subjected to Shot Peening and Associated Improvements

Material Type	Main Purpose / Achieved Improvement	Typical Application Areas	References
Steels (High- and Low-Carbon, Alloy Steels)	Increasing fatigue life and improving surface hardness.	Gears, shafts, crankshafts, turbocharger components, landing gear parts.	[4,10,11]
Aluminum Alloys (2xxx, 7xxx, 6xxx series)	Increasing resistance to fatigue and stress corrosion cracking (SCC).	Aerospace structural components, automotive engine parts, wheels.	[12,13]
Titanium Alloys (e.g., Ti-6Al-4V)	Increasing high-temperature fatigue (High-Cycle Fatigue) and wear resistance.	Biomedical implants, aerospace engine parts (blades), high-performance components.	[6,14,15]
Nickel-Based Superalloys (Inconel, Waspaloy)	Enhancing high-temperature fatigue resistance and optimizing contact fatigue resistance.	Gas turbine blades and disks, nuclear reactor components.	[16]
Composite Materials (Metal Matrix Composites)	Reducing interfacial stresses between the matrix and reinforcement.	Specialized applications.	[17]

Various parameters in the shot peening process directly affect the properties of the peened material. These parameters include shot diameter, shot type, peening pressure, peening angle, surface coverage ratio, peening time, nozzle diameter, and nozzle distance. Almen intensity is used to measure the kinetic energy transferred by the shots during shot peening.

Shot peening parameters vary depending on the applied process and the component. Incorrect selection of these parameters may cause excessive surface erosion. Shot peening parameters can be classified as shot characteristics and operational parameters. Shot characteristics include shot type, size, and hardness, while operational parameters include Almen intensity, surface coverage ratio, peening pressure, and peening time [3].

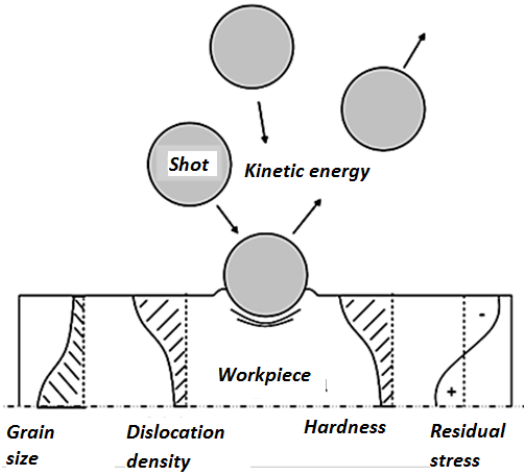


Figure 2. Changes occurring in the material after shot peening [6]

Aluminum materials are among the most widely used engineering materials after steel due to their high corrosion resistance, high specific strength, ease of formability, and lightweight characteristics. Compared to steel, aluminum is three times lighter and is preferred in many fields such as defense industry applications where energy efficiency is critical, as well as trains, aircraft, ships, and automobiles [18]. In recent years, increasing speed and maneuverability in military vehicles has become an important issue. Improving the mechanical properties of aluminum and aluminum alloys to replace steel materials has been a topic of extensive research. The 6xxx series aluminum alloys contain magnesium (Mg) and silicon (Si) and exhibit high weldability, which reduces manufacturing costs [19].

By taking advantage of the excellent machinability of aluminum materials, maximum performance can be achieved through shot peening, which is an economical and repeatable process.

When shot peening studies conducted on aluminum alloys are examined in the literature, Chen et al. investigated AA6061-T6 material peened at different pressures. They observed that as the pressure increased, surface roughness, plastic deformation layer thickness, and corrosion resistance increased [12]. Alkan [9], in his doctoral dissertation on AA1040 and AA2024 aluminum alloys, performed shot peening using S230 and S460 shots and reported an increase in yield strength while noting a decrease in

elongation at fracture. Sathyajith applied laser peening to AA6061-T6 aluminum alloy and determined that the plastic deformation on the surface increased with shot density and that microhardness increased by 1.2 mm in depth [15]. Markovina et al. [20] investigated the effects of nozzle angle and nozzle distance, which are among shot peening parameters, on AA2024-T3 and stated that incorrect selection of shot peening parameters could lead to excessive surface erosion. Patil et al., in their review study, addressed research examining the effects of shot peening on the mechanical properties of aluminum alloys and reported that shot peening is directly related to tensile strength, fatigue life, and corrosion resistance [17]. Hatamleh et al. applied laser peening and shot peening to specimens joined by friction stir welding and examined their mechanical properties. The improvement in mechanical properties was more pronounced in laser peening compared to shot peening, and this improvement was attributed to strain hardening [13]. Numerous studies on different aluminum alloys and various peening techniques are still ongoing.

SHOT MATERIALS AND THEIR PROPERTIES

The most effective parameter in shot peening is the shot itself. Selecting an appropriate shot type and size suitable for the surface of the material to be peened is critically important. Commonly used shot materials include steel, stainless steel, glass, ceramic, aluminum, and cut wire shots [1]. Shot materials are presented in Table 2.

Table 2. Shot Materials [21]

Shot Materials		
Ferrous	Steel Shot, Rounded Cut Wire Shot, Stainless Steel Shot	
Non-Ferrous	Metallic	Aluminum, Zinc, Copper, Bronze
	Non-metallic	Glass, Spherical Ceramic Shot

Steel shots are the most widely used due to their high density and cost effectiveness. Cast Steel Shot is preferred for general aerospace and automotive components requiring medium to high peening intensity in broad and cost-effective applications. Cut Wire Shot is used for components

requiring high hardness and intensive peening, such as springs, gears, and fasteners. Stainless Steel Shot is used for sensitive components, such as titanium, aluminum, or stainless steel parts, to eliminate or reduce corrosion risk. Cast Iron Shot is used on very hard surfaces requiring wear resistance but tends to fracture quickly. High Manganese Steel Shot is used to increase toughness and impact resistance in components with high hardness and impact resistance requirements. For sensitive surfaces, Glass Beads are preferred; they provide good results in surface finishing and cosmetic improvements for light alloys (Al, Mg) and thin-walled components. Ceramic Shot is used for titanium alloys and nickel-based superalloys where chemical purity and smooth surface finish are critical. Additionally, Aluminum Shot is used for very delicate, low-energy peening and surface cleaning, while Plastic/Polymer Shot is preferred for removing contaminants or creating a very light peening effect without causing surface damage [10,11].

In addition to shot material, other important parameters include shot diameter, shot type, and hardness, all of which directly affect peening intensity. The shot size must be selected to be the maximum size that does not damage the peened component. As shot diameter increases, peening intensity increases when other parameters are kept constant. Accordingly, the thickness of the plastically deformed region in the peened component also increases [1].

Shot sizes are commonly classified according to the SAE standard. For example, if a shot is defined as “S230,” the letter “S” indicates that the shot is made of steel, and the value “230” represents the shot diameter in inches multiplied by 10,000 [22]. Figure 3 shows shots according to the SAE standard.

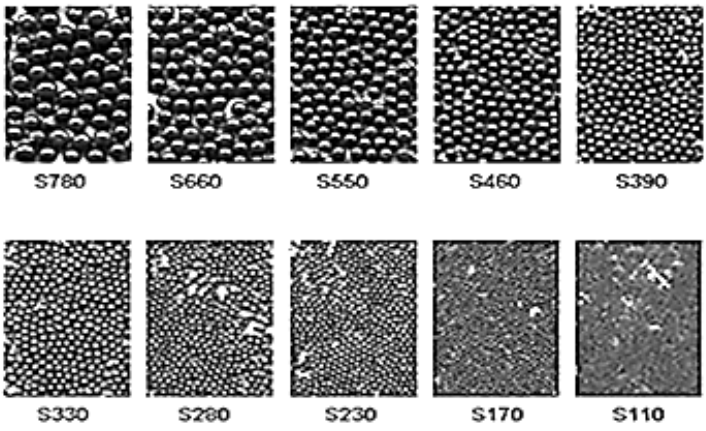


Figure 3. Shots according to the SAE standard [23]

Shot size affects the surface quality of the peened component. Surface roughness is determined by the impact energy of the shots and the resulting dimples. An increase in shot diameter increases dimple size [1]. Figure 4 illustrates the change in surface roughness with increasing shot diameter. Surface roughness does not always increase with increasing shot diameter. As shown in Figure 4, the best Ra value (average roughness) for the material with VSD 365 was obtained using a shot diameter of 0.6 mm.

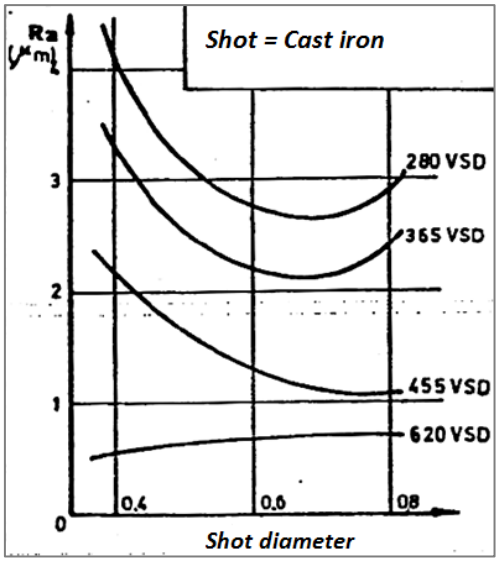


Figure 4. Effect of shot diameter on surface quality in cast iron [1]

In the literature, the most frequently investigated and compared shot types are generally Stainless Steel Shot (for corrosion and contamination control) and Cast Steel Shot (for general mechanical property improvement). Studies typically focus on the effects of shot hardness, size, and density on compressive residual stress depth and surface hardness.

ALMEN INTENSITY

Almen intensity is an inspection method frequently used prior to the peening process. The Almen test was developed by John Almen in 1945. It is used to measure the kinetic energy transferred by the shots projected during the shot peening process. It enables the assessment of the adequacy of process parameters.

Standard strips are used for measuring Almen intensity. The Almen strip is fixed onto a holder (Figure 5a), and the peening process is carried out. As a result of shot peening, convex bending occurs in the strip. Almen intensity is defined as the arc height of the curvature formed on the deformed strip and is determined using an Almen gauge (Figure 5b). To measure the Almen intensity of the shot peening process, peening is repeated for different durations [9].

Figure 5. a) Almen strips and Almen holder b) Almen measuring device

Standard strip types are used in the measurement of Almen intensity (N: light peening intensity, A: medium peening intensity, C: high peening intensity). The properties of Almen strips are given in Table 3. Almen intensity is obtained by multiplying the arc height formed on the Almen strip as a result of the peening process by 1000 in inch units [24].

Table 3. Properties of Almen strips [25]

Strip types	Standard	Flatness	Hardness	Thickness
C	AMS-S 13165	+/- 0,038 mm	44-50 HRC	2.39 mm
A	SAE J442	+/- 0,025 mm	44-50 HRC	1.29 mm
N	SAE AMS 2432	+/- 0,013 mm	45-58 HRC	0.76 mm

SURFACE COVERAGE RATIO

The ratio of the surface area affected by the shots to the total surface area during shot peening is defined as the surface coverage ratio. Surface coverage ratio is a parameter frequently used in shot peening and is important in terms of homogeneity and process reliability. There is no universally accepted definition or standard method for determining surface coverage ratio. Surface coverage ratio is directly proportional to peening time. Practically, when the peening time is doubled and the increase in arc height on the strip is less than 10%, it is accepted that 100% surface coverage has been achieved [26]. Figure 6 shows the relationship between peening time and arc height.

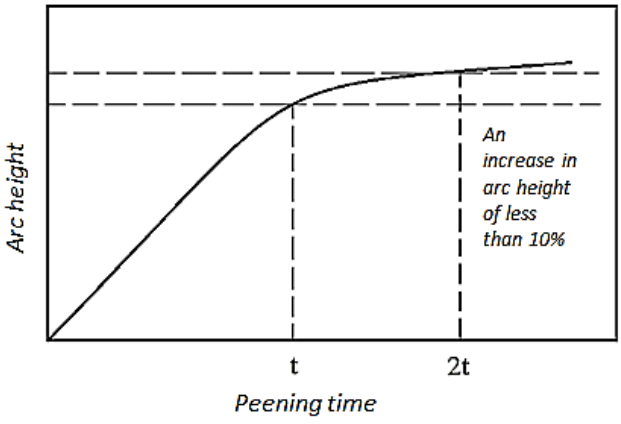


Figure 6. Determination of surface coverage ratio from the peening time–arc height graph [21]

Surface coverage ratio can also be examined visually. Figure 7 presents example images of fully covered and partially covered surfaces. In some cases, surface coverage ratios greater than 100% may be required. In such cases, the peening process is carried out by doubling the time required to reach 100% surface coverage under the specified process parameters [1].

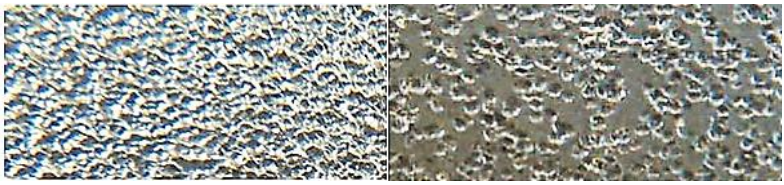


Figure 7. a) Fully covered surface by shot peening b) Partially covered surface images [5]

MICRO- AND MACROSTRUCTURAL PROPERTIES

Shot peening creates controlled plastic deformation in the surface layer, resulting in a fine-grained surface microstructure, increased surface hardness, and generally compressive residual stresses at the surface. These changes directly affect component performance, particularly fatigue, creep, and wear behavior. Grain refinement at the surface and high dislocation density can lead to the formation of subsurface nanostructures in certain materials. These structures result in both increased hardness and localized ductility changes; therefore, component behavior depends on the material–process combination [27].

When the mechanical effects of shot peening are examined, improvements in fatigue and wear performance have been reported through increased surface hardness, grain refinement, and compressive residual stresses retained in the surface layer. This increase in hardness raises the fatigue limit in most metals and alloys and extends the service life of components sensitive to fretting. Advanced peening methods such as USP and LSP can provide deeper compressive residual stress profiles and higher hardness increases compared to conventional shot peening. However, some studies have also reported adverse changes in surface corrosion resistance for certain alloys [27,28].

Compressive residual stresses at the surface reduce the tendency for crack initiation, resulting in significant improvements in fatigue life. However, if peening parameters (impact velocity, shot diameter, impact density, energy) are not appropriately selected, the expected benefits may be reduced due to excessive roughness or surface microcracks. Studies have shown that parameter optimization is critical for fatigue performance [29].

The peening process generally increases surface roughness; however, the compressive stresses induced often provide a net benefit. If surface roughness is reduced through post-process superfinishing while retaining compressive residual stresses, optimal results can be achieved [30].

Many studies have reported improvements in wear resistance through controlled peening, primarily due to increased surface hardness and modified surface topography. However, in some cases—such as peening with excessively hard particles or parameter selections that leave microcracks on the surface—friction coefficients may increase or localized material removal may occur [31].

Kocaman and Demirer [ ] performed microhardness measurements at six equally spaced points along the thickness from the lateral cross-sections of AA6061-T6 and St32 material specimens and found that hardness gradually decreased from the surface toward the center in all samples. An increase in hardness was observed with increasing peening pressure and shot diameter [32].

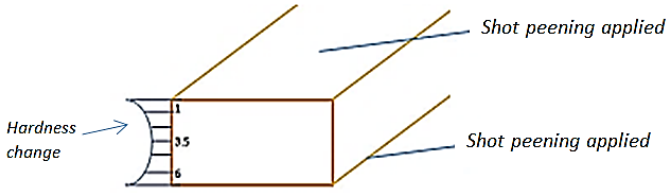


Figure 8. Schematic illustration of microhardness measurement [32]

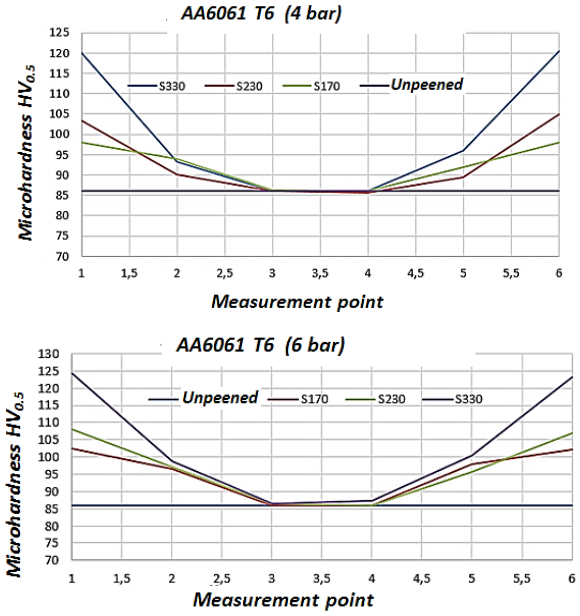


Figure 9. Microhardness results of AA6061-T6 material after shot peening (4–6 bar) [32]

The mechanical and corrosion properties of AA5754-H111 aluminum alloy were investigated after shot peening. Corrosion test results showed that partial corrosion damage occurred in specimens peened with S110 (0.3 mm) shot at 15 Almen intensity, whereas in specimens peened with S170 (0.4 mm) shot at 15 Almen intensity, both an increase in tensile strength and no observable corrosion damage were detected [33].

RESULTS AND NEW SHOT PEENING TECHNIQUES

In recent years, alongside conventional shot peening methods, extensive research has focused on ultrasonic shot peening (USP), laser shock peening (LSP), and various hybrid methods [29]. Methods such as LSP and USP can produce deeper and more homogeneous compressive residual stress profiles compared to conventional shot peening, providing performance advantages for critical components. These effects have been validated through X-ray

diffraction (XRD), microhardness, and optical/SEM cross-sectional analyses [34].

Ultrasonic and impulse applications: USP and impulse peening methods can provide more pronounced strengthening by enabling surface nanostructuring and deeper residual stresses [34].

1. Ultrasonic Peening (USP)

Ultrasonic peening is a mechanical surface enhancement method developed as an alternative to conventional shot peening, aiming to induce compressive residual stresses on the surface. Both LSP and USP methods are based on the principle of continuously impacting the material surface with a group of impact elements (peening pins or shots) using high-frequency vibrations (generally above 20 kHz, i.e., within the ultrasonic range) [35–38].
The vibrations are generated by a transducer and transmitted to the peening elements through a sonotrode (vibration horn).
These high-frequency impacts cause plastic deformation at the surface and consequently generate compressive residual stresses.
The method is generally used to prevent fatigue crack initiation at the weld toe of welded joints.

Table 4. Reasons for Use and Application Areas of Ultrasonic Peening (USP)

	Why It Is Used
Weld Seams	To reduce stress concentration and surface roughness (crack initiation sites), especially at the toes of steel and aluminum welds.
Large Structures	Critical weld regions of bridges, ships, offshore oil platforms, and large machinery.
Special Alloys	To improve the fatigue strength of high-performance materials such as titanium and nickel-based superalloys.

Table 5. Advantages and Disadvantages of the USP Method

No	Advantages	Disadvantages
1	Depth and Effectiveness: Can induce deeper and more effective compressive residual stresses than conventional peening (ideal for weld seams).	Noise and Vibration: Due to high-frequency vibrations, it can be noisy and uncomfortable for the operator.
2	Portability: Equipment can be easily transported and applied on-site or on large structures.	Heating: Prolonged operation may cause heating problems due to vibration energy.
3	Surface Roughness: Has the potential to reduce surface roughness.	Large Areas: Not practical for processing very large surfaces due to low processing speed.

2. Impulse / Accelerated Impulse Peening Method

As implied by its name, this method is based on the principle of applying short-duration, high-energy, and controlled impacts to the surface. These systems generally operate using pneumatic (air pressure) or electromagnetic principles, accelerating a single shot or pin to strike the surface at high velocity [35–38].
Repeated impacts create localized and intense plastic deformation on the surface, thereby inducing compressive residual stresses.
The energy level is higher than that of conventional shot peening but lower than that of Laser Shock Peening (LSP).
The system is particularly used in localized areas where precise control and high energy density are required.

Table 6. Reasons for Use and Application Areas of Impulse / Accelerated Impulse Peening

Application Area	Why It Is Used
Local Improvement	Small areas with critical stress concentrations (holes, corner radii, special joints).
Precision Control	Applications requiring highly precise control of energy and impact frequency.
Repairs	In-situ repair and strengthening of components that have experienced fatigue damage.

Table 7. Advantages and Disadvantages of Impulse / Accelerated Impulse Peening

No	Advantages	Disadvantages
1	High Energy: Provides high energy density in a single impact, inducing effective residual stresses.	Localization: Since the peening effect is highly localized, it is not efficient for large areas.
2	Precision Control: Pneumatic/electromagnetic impact frequency and energy can be adjusted.	Equipment Complexity: System setup and control are more complex than conventional shot peening.
3	Portability: Similar to USP, portable devices are available for field applications.	Repeatability: Maintaining high precision in energy and impact location between successive impacts can be challenging.

3. Laser Shock Peening (LSP)

Laser Shock Peening is an advanced surface enhancement method aimed at generating deep and permanent compressive residual stresses on metal surfaces using high-energy laser pulses. This method is considered an evolution of conventional shot peening and provides significantly greater stress depth and controllability [39–42].

LSP operates by applying extremely short-duration and high-energy laser pulses to the surface. The process steps are as follows:

Protective Coating: The metal surface to be treated is typically covered with an absorptive coating such as paint or adhesive tape to ensure efficient absorption of laser energy by the surface.

Confinement Medium: A shock confinement medium, such as water or transparent glass, is placed over the coating.

Laser Pulse: A high-energy laser pulse strikes the absorptive coating, which instantaneously vaporizes and transforms into plasma.

Shock Wave: The plasma rapidly expands against the confinement medium, generating a high-pressure shock wave that propagates into the metal.

Plastic Deformation: The shock wave exceeds the yield strength of the material, creating plastic deformation beneath the surface.

Compressive Residual Stress: This subsurface deformation leads to the formation of deep and permanent compressive residual stresses at the surface, preventing fatigue crack initiation.

The primary purpose of LSP is to increase the fatigue life and durability of materials.

Reasons for Use:

Deep Residual Stress: Compared to conventional shot peening, LSP can generate compressive residual stresses 5 to 10 times deeper (typically 1–2 mm), more effectively preventing fatigue crack initiation originating from subsurface defects.

Table 8. Main Application Areas of the Laser Shock Peening (LSP) Method [39–42]

Application Area	Application Example	Improved Property
Aerospace	Gas turbine engine blades, landing gear, titanium and nickel-based superalloy components.	High-cycle fatigue (HCF) life, stress corrosion cracking (SCC) resistance.
Automotive	Crankshafts, transmission gears, critical Aluminum engine components.	Mechanical fatigue life, impact resistance.
Biomedical	Titanium implants, prostheses.	Biocompatible surface and wear resistance.
Nuclear	Pressure vessel components, pipe joints.	Corrosion fatigue and SCC resistance.

Superior Surface Finish: As a non-contact process, LSP does not cause significant surface roughness or deformation, unlike conventional shot peening.
Precision Control: Laser parameters (energy, pulse duration, overlap) can be precisely controlled to generate residual stresses at desired depths in specific regions [39–42].

Table 9. Advantages and Disadvantages of the LSP Method [39–42]

No	Advantages	Disadvantages
1	Depth: Achieves compressive residual stresses at much greater depths than conventional peening methods.	Cost: Equipment (high-energy laser systems) and operating costs are very high.
2	Surface Finish: Does not cause significant surface roughness deterioration (non-contact process).	Productivity: Processing speed may be slower than conventional shot peening for large areas.
3	Control: Process parameters can be adjusted with high precision and repeatability.	Application Difficulty: Adapting to complex geometries can be challenging and requires experienced process control.
4	Corrosion Resistance: Significantly improves stress corrosion cracking (SCC) resistance by preventing surface crack initiation.	Absorptive Coating: Requires application of a protective/absorptive coating before processing and subsequent cleaning.

Recent studies in the literature have focused on optimizing peening parameters (shot size, velocity/energy, coverage density, stand-off distance, and frequency for USP). Process optimization is commonly performed using design of experiments (DoE), response surface methodology (RSM), or finite element (FE) dynamic simulations, which help balance desired hardness, residual stress, and surface roughness targets [43].

The literature of the last five years indicates that shot peening-type processes significantly improve mechanical and surface performance when applied appropriately. However, optimization, material-specific effects, and post-peening corrosion or corrosive environment behavior remain active research topics. Further long-term fatigue data, industrial-scale experiments, and systematic studies on combined peening + heat treatment/coating processes are recommended [29].

REFERENCES

1. Varol, R. (1990). *2024 alüminyum alaşımının yorulma ömrü üzerine farklı bilyeli dövme işlem parametrelerinin etkisi* (*Effects of different shot peening process parameters on the fatigue life of 2024 aluminum alloy*). Doctoral dissertation, Selçuk University, Graduate School of Natural Sciences, Thesis No. 17040.
2. Alkan, Z., Varol, R., & Server, R. (2016). Bilyeli dövme uygulanmış AA1050 alüminyum malzemenin metalografik incelenmesi (*Metallographic investigation of shot-peened AA1050 aluminum material*). *Süleyman Demirel University Journal of Natural and Applied Sciences*, 20(5), 524–529. https://doi.org/10.19113/sdufbed.12415
3. Avcu, E. (2018). Bilyalı dövme parametrelerinin AA7075-T6 alüminyum alaşımının yüzey altı özelliklerine etkileri (*Effects of shot peening parameters on subsurface properties of AA7075-T6 aluminum alloy*). *Gazi University Journal of Science, Part C*, 6(4), 741–752.
4. Ünal, O., & Varol, R. (2014). Almen intensity effect on microstructure and mechanical properties of low carbon steel subjected to severe shot peening. *Applied Surface Science*, 290, 40–47. https://doi.org/10.1016/j.apsusc.2013.10.184
5. Metal Improvement Company. (2005). *Shot peening applications*. https://www.cwst.co.uk/wp-content/uploads/2015/08/MIC_Green_Book_9th_Edition.pdf
6. Avcu, Y. Y. (2019). *Ti6Al4V titanyum alaşımın bilyalı dövme sonrası aşınma davranışının incelenmesi* (*Investigation of wear behavior of Ti6Al4V titanium*

alloy after shot peening). Doctoral dissertation, Kocaeli University, Graduate School of Natural Sciences.
7. Önen, B., Yıldıran, Y., & Çınar, A. (2015). Investigation of the effects of erosion test parameters on particle impingement velocity using CFD analysis. *International Advances in Applied Physics and Material Science Congress*, 127(4), 1225–1229.
8. Yalçınkaya, A. (2019). *Bilyalı dövme prosesi uygulanan tank palet pimlerinin yorulma analizi* (*Fatigue analysis of tank track pins subjected to shot peening process*). Master's thesis, National Defence University, Alparslan Defence Sciences Institute.
9. Alkan, Z. (2014). *Alüminyum alaşımlarında bilyalı dövmenin fiziksel, mekanik ve yüzey özelliklerine etkisi* (*Effects of shot peening on physical, mechanical and surface properties of aluminum alloys*). Doctoral dissertation, Süleyman Demirel University, Graduate School of Natural Sciences.
10. Guagliano, M., & Boniardi, M. (2018). Shot peening: Applications and effects on materials. In S. Smith & J. Doe (Eds.), *Surface engineering and material fatigue* (pp. 123–145). Elsevier. https://doi.org/10.1016/B978-0-12-345678-9.00007-X
11. López-Crespo, P., & Infante, V. (2018). A review of shot peening effect on fatigue performance in metallic alloys. *Fatigue & Fracture of Engineering Materials & Structures*, 41(8), 1618–1634. https://doi.org/10.1111/ffe.12781
12. Chen, B., Huang, B., Liu, H., Li, X., Ni, M., & Lu, C. (2014). Surface nanocrystallization induced by shot peening and its effect on corrosion resistance of 6061 aluminum alloy. *Journal of Materials Research*, 29(24), 3002–3010. https://doi.org/10.1557/jmr.2014.323
13. Hatamleh, O., Mishra, R. S., & Oliveras, O. (2009). Peening effects on mechanical properties in friction stir welded AA2195 at elevated and cryogenic temperatures. *Materials & Design*, 30(8), 3165–3173. https://doi.org/10.1016/j.matdes.2008.11.010
14. Liu, R., Yuan, S., Lin, N., Zeng, Q., Wang, Z., Wu, Y. (2021). Application of ultrasonic nanocrystal surface modification (UNSM) technique for surface strengthening of titanium and titanium alloys: A mini review. *Journal of Materials Research and Technology*, 11, 351–377.
15. Sathyajith, S., & Kalainathan, S. (2012). Effect of laser shot peening on precipitation hardened aluminum alloy 6061-T6 using low energy laser. *Optics and Lasers in Engineering*, 50, 345–348.
16. Chen, Z. (2017). *Surface integrity and fatigue life improvement of nickel-based superalloys through shot peening*. Doctoral dissertation, Linköping University, Linköping Studies in Science and Technology, Dissertation No. 1821. https://www.diva-portal.org/smash/get/diva2:1068050/FULLTEXT01.pdf
17. Patil, S. S., K. S., S. P. S., & S. A. S. (2018). A review on influence of various technological processes on mechanical properties of aluminum alloys. *IOP Conference Series: Materials Science and Engineering*, 376, 012043. https://doi.org/10.1088/1757-899X/376/1/012043
18. Başer, T. A. (2012). Alüminyum alaşımları ve otomotiv endüstrisinde kullanımı (*Aluminum alloys and their use in the automotive industry*). *Mühendis ve Makina*, 53(635), 51–58.

19. Coşkun, K. (2002). *Hadde alüminyum alaşımlarının uçaklarda kullanılması* (*Use of rolled aluminum alloys in aircraft*). Master's thesis, Istanbul Technical University, Graduate School of Natural Sciences.
20. Markovina, R., Blagojević, B., & Ban, D. (2008). Investigation of influential parameters on shot peening of aluminum alloys. *12th International Research/Expert Conference*, Istanbul.
21. Ünal, O. (2011). *Bilyalı dövme işleminin tane boyutuna etkisinin deneysel olarak incelenmesi* (*Experimental investigation of the effect of shot peening on grain size*). Master's thesis, Bartın University, Graduate School of Natural Sciences.
22. Ay, İ., & Sakin, R. (1988). Kumlama ile parça temizleme ve bilyalı dövme prosesi (*Part cleaning by sandblasting and shot peening process*). *Makina & Metal Teknolojisi*, 68–72.
23. BDM Bilginoğlu Döküm. (n.d.). *Çelik granül bilya* (*Steel granule shot*). https://www.bdmbilginogludokum.com.tr/urunler/celik-granul-bilya.png
24. Harbelioğlu, E. (2018). *Bilyalı dövmenin AA7075-T6 alaşımının yüzey özelliklerine ve erozif aşınma davranışına etkisi* (*Effect of shot peening on surface properties and erosive wear behavior of AA7075-T6 alloy*). Master's thesis, Kocaeli University, Graduate School of Natural Sciences.
25. Bek, N. (2009). Peening uygulamaları (*Peening applications*). *Yüzey İşlem Kumlama Dergisi*, 7, 26–28.
26. Nguyen, V. B., Poh, H. J., & Zhang, Y.-W. (2014). Predicting shot peening coverage using multiphase computational fluid dynamics simulations. *Powder Technology*, 256, 100–112.
27. Wen, Y., Wu, Y., Hua, L., Xie, L., Wang, L., Zhang, L. C., & Lu, W. (2021). Effects of shot peening on microstructure evolution and mechanical properties of surface nanocrystal layer on titanium matrix composite. *Materials & Design*, 206, 109760.
28. Walczak, M., & Szala, M. (2021). Effect of shot peening on the surface properties, corrosion and wear performance of 17-4PH steel produced by DMLS additive manufacturing. *Archives of Civil and Mechanical Engineering*, 21(4), 157. https://doi.org/10.1007/s43452-021-00306-3
29. Świetlicki, A., Szala, M., & Walczak, M. (2022). Effects of shot peening and cavitation peening on properties of surface layer of metallic materials—A short review. *Materials*, 15(7), 2476. https://doi.org/10.3390/ma15072476
30. Jia, L., & Zhang, X. (2021). Effects of conventional and severe shot peening on residual stress and fatigue strength of steel AISI 1060 and residual stress relaxation due to fatigue loading: Experimental and numerical simulation. *Metals and Materials International*, 27, 2575–2591.
31. Jia, Y., Liu, X., Wang, H., Zhang, L., & Chen, Q. (2024). Influence of shot-peening treatment on wear resistance of medium manganese steel. *Journal of Materials Research and Technology*, 31, 3703–3711. https://doi.org/10.1016/j.jmrt.2024.07.101
32. Kocaman, A. G., & Demirer, A. (2021). Investigation of effect of shot peening on the mechanical properties of AA6061-T6 aluminum alloy. *Afyon Kocatepe University International Journal of Engineering Technology and Applied Sciences*, 4(2), 49–58. https://doi.org/10.53448/akuumubd.959091
33. Bozkurt, M., & Demirer, A. (2023). AA5754-H111 alüminyum alaşımının bilyeli dövme prosesi mekanik özelliklerin incelenmesi ve Cr6+ kaplama

uygulamasının korozyona bir bakış (*Investigation of mechanical properties of AA5754-H111 aluminum alloy subjected to shot peening process and an overview of corrosion behavior after Cr6+ coating*). *Afyon Kocatepe University International Journal of Engineering Technology and Applied Sciences*, 6(2). https://doi.org/10.53448/akuumubd.1295992

34. Yue, X., Hu, S., Wang, X., Liu, Y., Yin, F., & Hua, L. (2022). Understanding the nanostructure evolution and the mechanical strengthening of the M50 bearing steel during ultrasonic shot peening. *Materials Science and Engineering: A*, 836, 142721. https://doi.org/10.1016/j.msea.2022.142721
35. Marines, B., & Shiozawa, K. (2018). A comparison of the fatigue life improvement in welding joints by ultrasonic peening and laser peening. *Welding in the World*, 62(3), 565–577. https://doi.org/10.1007/s40194-018-0557-6
36. Statnikov, A. P., & Shcheglov, V. (2020). Ultrasonic impact treatment (UIT): Theory and applications in fatigue life improvement of welded structures. *Materials Science and Engineering: A*, 790, 139726. https://doi.org/10.1016/j.msea.2020.139726
37. López-Crespo, P., Infante, V., Mikkola, H. E., & Martínez-Berridi, P. (2018). Fatigue life improvement of structural details treated by ultrasonic peening and other mechanical surface treatments: A review. *Engineering Fracture Mechanics*, 194, 148–172. https://doi.org/10.1016/j.engfracmech.2018.03.013
38. Hassan, H. E., & Ahmed, M. S. (2020). Influence of mechanical treatment methods (shot peening, ultrasonic peening and accelerated impulse peening) on the fatigue performance of welded joints. *Journal of Materials Research and Technology*, 9(3), 5202–5217. https://doi.org/10.1016/j.jmrt.2020.03.031
39. Yin, Y., Chen, Z., & Chen, G. (2020). Laser shock peening: A review of its processing mechanism, parameter optimization, and fatigue performance improvement. *Optics & Laser Technology*, 131, 106404. https://doi.org/10.1016/j.optlastec.2020.106404
40. Lu, J., Luo, K., Deng, M., Li, C., Ren, X., & Zhang, Y. (2021). Review on the residual stress, microstructural evolution and fatigue property enhancement of metallic materials by laser shock peening. *Optics and Lasers in Engineering*, 147, 106720 https://doi.org/10.1016/j.optlaseng.2021.106720.
41. Hatamleh, O., & Oliveras, O. (2018). Laser peening of metallic alloys: Fundamentals, applications and effects on mechanical properties. In S. Smith & J. Doe (Eds.), *Advanced surface engineering techniques* (pp. 123–145). Springer.
42. Sathish, T., & Krishna, V. S. (2019). A review on influence of laser shock peening on the mechanical properties of materials. *Journal of Manufacturing Processes*, 47, 337–356. https://doi.org/10.1016/j.jmapro.2019.09.006
43. Jantepa, N., Siripath, N., & Suranuntchai, S. (2024). Predictive modeling and optimization of hot forging of AISI 1045. *Metals*, 14(10), 1198. https://doi.org/10.3390/met14101198

Scientific Advances in Sustainable Carbon Sources and Adsorption Technologies: A Bibliometric Analysis

Hatice ERDEM[1]

1- Asst. Prof.; Muş Alparslan University, Faculty of Engineering and Architecture, Deparment of Food Engineering, h.erdem@alparslan.edu.tr ORCID No: 0000-0002-7666-8301

ABSTRACT

The development of sustainable material solutions for energy and environmental applications is now necessary due to increased environmental pressures brought on by rapid industrialization, population growth, and rising energy demand. In this regard, the development of effective and low-impact adsorption technologies now depends on the use of carbon-based adsorbents made from sustainable carbon sources. In this study, a comprehensive bibliometric analysis was conducted using the bibliometrix R-Tool software to examine 679 publications on adsorption integrated with sustainable carbon sources indexed in the Web of Science database between 1980 and 2026. This analysis provides new fundamental insights into the field and identifies potential research directions for future studies, contributing to a deeper and more systematic understanding of the role of sustainable carbon sources in adsorption research. The analysis results show that research activity in this field remained limited prior to 2010, began to emerge between 2010 and 2015, and has experienced rapid growth since 2015. The majority of publications are concentrated in the disciplines of Chemical Engineering, Chemistry, and Materials Science, indicating a strong materials- and process-oriented research focus. The results also reveal a high concentration of publications in a small number of leading journals, with the *Chemical Engineering Journal* serving as the primary knowledge hub, followed by *Separation and Purification Technology* and the *Journal of Environmental Chemical Engineering*. Furthermore, the findings highlight a clear dominance of China-based authors and institutions, reflecting the country's strong research capacity and sustained investments in advanced materials and environmental applications.

Keywords – Adsorption; Bibliometric Analysis; Biblioshiny; Carbon Sources; Sustainability.

INTRODUCTION

Population growth, accelerated urbanization, and rising living standards over the past half-century have led to a sharp increase in global energy consumption and associated environmental burdens, a trend expected to continue in the coming years. To meet the increasing energy demand, the development of efficient and clean energy sources and energy carriers such as bioenergy, lithium-ion batteries, and supercapacitors has become a critical necessity. In parallel, intensive industrialization and prolonged agricultural activities have caused a serious deterioration in environmental quality; effectively controlling and removing environmental pollutants has emerged as an urgent priority. These multifaceted energy and environmental problems place the development of high-performance materials and technology

systems with a low carbon footprint at the heart of the solution (Zhou et al., 2019).

In this context, carbon-based materials derived from renewable biomass or waste sources, such as carbon nanotubes, graphene, activated carbons, and carbon aerogels, are considered prominent material platforms in energy and environmental applications due to their low energy requirements, low-emission synthesis processes, and superior physicochemical properties (Sun et al., 2025). These materials offer solutions aligned with sustainable development goals by demonstrating higher performance compared to conventionally synthesized carbon counterparts in energy storage, catalysis, and environmental applications (Varghese, Ralph, & Kuzhiumparambil, 2025).

Adsorption systems are directly related to the United Nations Sustainable Development Goals, such as supporting sustainable economic growth, developing resilient infrastructure, making cities more sustainable, combating climate change, and ensuring access to affordable energy. Applications such as air purification and dehumidification, industrial pollutant removal, thermal energy storage, and adsorption cooling systems clearly demonstrate that adsorption offers holistic solutions that can increase energy efficiency while reducing environmental impacts (Shitole, Tiwari, Sinhmar, & Bellos, 2025). In this context, carbon-based adsorbents derived from sustainable carbon sources have attracted increasing research attention due to their advanced porous structures, wide range of biomass- and waste-derived precursors, tunable surface chemistry, and high adsorption capacities toward various pollutants. Their performance advantages in air and water treatment, carbon capture, and energy-related adsorption processes have contributed to a rapid quantitative and thematic expansion of research in carbon-based adsorption technologies (Ahmed, Hameed, & Khan, 2024; Reza et al., 2023). However, this rapid and multifaceted growth in the literature makes it difficult to holistically assess research trends, prominent themes, leading countries and institutions, as well as interdisciplinary collaboration networks. Therefore, bibliometric analysis stands out as a powerful and necessary tool for systematically mapping the body of knowledge in the field of sustainable carbon sources and adsorption, identifying existing research gaps, and rationally outlining future research directions.

Bibliometric analysis is a widely recognized and robust approach for systematically examining large bodies of scientific literature from both quantitative and qualitative perspectives (Baarimah et al., 2024). By integrating principles from statistics, mathematics, and visualization techniques, this method enables the structured analysis of publication characteristics, citation patterns, collaboration networks, and keyword relationships, thereby revealing the intellectual structure and internal linkages of a research field (Baarimah et al., 2024; Hassan & Duarte, 2024;

Xu, Li, Zhang, & Zhang, 2024). A comprehensive bibliometric assessment provides detailed descriptive information on research outputs, leading authors, institutions, countries, and thematic trends, allowing the developmental trajectory and evolving directions of a given domain to be clearly identified (Xu et al., 2024). One of the key strengths of bibliometric analysis lies in its ability to transform complex and abstract literature into an organized, transparent, and analytically manageable framework, making it particularly suitable for evaluating rapidly expanding research areas (Hassan & Duarte, 2024). In the context of the rapidly expanding literature on sustainable carbon-based adsorption technologies, bibliometric analysis offers a robust framework for identifying influential contributions and emerging research trends.

The aim of this book chapter is to systematically examine scientific publications on sustainable carbon sources and adsorption applications of carbon-based adsorbents using bibliometric analysis. In this context, publications obtained from selected scientific databases were analyzed in terms of publication performance by year, leading journals, authors, institutions and countries, citation structures, collaboration networks, and keyword-based research trends. The study does not aim to evaluate the information content in the literature; rather, it aims to reveal research dynamics, thematic concentrations, and the structural development of the field through quantitative indicators. In this respect, the presented bibliometric framework aims to objectively reflect the current state of sustainable carbon-based adsorption research and provide a strategic perspective for future studies.

METHODOLOGY

In this study, a bibliometric analysis method was used to quantitatively evaluate the structural characteristics and research trends of scientific production in the field of sustainable carbon sources and adsorption. Literature data were obtained from the Web of Science (WoS) Core Collection database due to its provision of comprehensive and standardized bibliographic records. The WoS database provides multidimensional metadata necessary for bibliometric analysis, such as publication title, author information, institutional affiliations, country distributions, publication years, citation data, and keywords. The data processing, analysis, and visualization stages were performed using the

Bibliometrix package developed with the R programming language and its web-based interface Biblioshiny, which allows for the examination of bibliometric performance, conceptual structure, and collaboration networks (Aria & Cuccurullo, 2017). Additional graphical analyses and tabulation processes were performed using Microsoft Excel software to validate and support the results. A comprehensive literature review was conducted on December 2, 2025. The WoS Core Collection database was used as the bibliographic data source for the study. The search strategy was structured to focus on sustainable carbon-based adsorbents, and the following keywords were used in the query: (*"waste-derived carbon"* or *"biomass-derived carbon"* or *"lignin-based carbon"* or *"agricultural waste carbon"* or *"carbonaceous adsorbent"* or *"sustainable carbon materials"* or *"green carbon materials"*) AND "*adsorption*". As a result of this search strategy, a total of 679 scientific publications were identified for inclusion in the analysis.

RESULTS AND DISCUSSION

A bibliometric analysis of publications on the use of sustainable carbon sources in adsorption technologies within the WoS database comprehensively reveals scientific development over a 46-year period from 1980 to 2026 (Table 1). The dataset consists of a total of 679 documents from 254 sources. These documents include 592 research articles, 48 review articles, 16 early access articles, 14 conference proceedings, 8 conference papers, and 1 data paper. The calculated annual growth rate of 4.62% indicates a steady expansion of the field. The average age of the documents is 4.87 years, suggesting that the studies refer to a relatively recent literature. The average number of citations per document is 31.78. Furthermore, the dataset contains a total of 34,556 references. These indicators reveal that scientific output in the field of sustainable carbon sources and adsorption has increased steadily over time and has a high citation impact.

The analyzed dataset contains 1455 Keywords Plus terms and 1735 author keywords. Of the studies with contributions from a total of 3255 authors, only 5 produced single-authored documents, resulting in a total of 6 single-authored documents. The average number of co-authors per document in the dataset is 5.96, and the international co-authorship rate is 24.89%. These indicators reveal that research on sustainable carbon sources and

adsorption is developing through intensive collaboration and multi-author production dynamics.

Table 1. Main information about the data obtained from the Web of Science.

Description	Results
Main Information About Data	
Timespan	1980:2026
Sources (Journals, Books, etc)	254
Documents	679
Annual Growth Rate %	4,62
Document Average Age	4,87
Average citations per doc	31,78
References	34556
Document Contents	
Keywords Plus (ID)	1455
Author's Keywords (DE)	1735
Authors	
Authors	3255
Authors of single-authored docs	5
Authors Collaboration	
Single-authored docs	6
Co-Authors per Doc	5,96
International co-authorships %	24,89
Document Types	
article	592
article; data paper	1
article; early access	16
article; proceedings paper	8
proceedings paper	14
review	48

Publication Trends and Subject Areas

The trend in scientific publications on sustainable carbon sources and adsorption technologies between 1980 and 2026 has been analyzed. Figure 1 shows that publications on sustainable carbon sources and adsorption technologies during the 1980–2026 period exhibit a clear upward trend over the years. Throughout the 1980s and 1990s, the number of publications remained quite low and followed a flat trajectory; this indicates that the field was still in its early stages of development and that only a limited number of studies were produced. Although a slight increase in the number of publications was observed from the early 2000s onwards, the real momentum began after 2010. This period has seen the field become more visible, driven by both the strengthening of the concept of sustainability and the growing interest in waste-based carbon materials. Publication numbers have shown a sharp growth trend since 2015. The graph shows a particularly pronounced vertical rise after 2018. Publication numbers peaked in the 2022–2024 period, with production much more intense than in previous

years. This situation indicates that the field has matured, is being intensively studied by the scientific community, and that the global research trend is continuing strongly. The field was low-volume before 2010, emerging between 2010 and 2015, and has rapidly grown since 2015. The vertical rise in recent years clearly shows that interest in sustainable carbon sources will continue to increase in the long term.

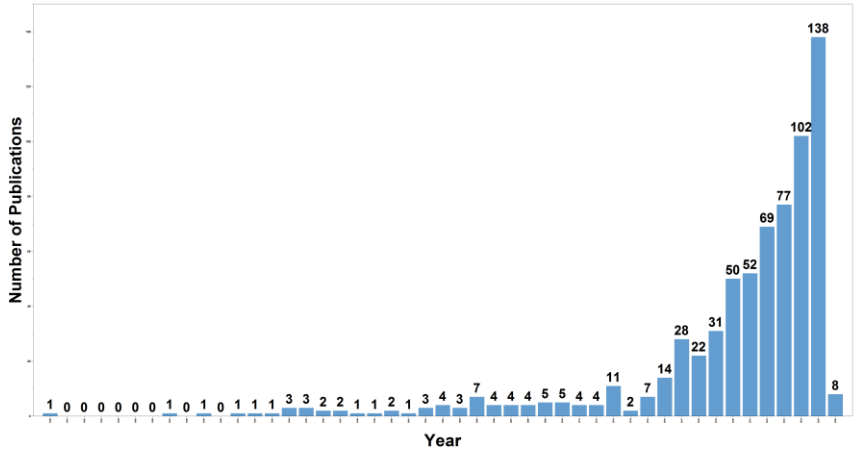


Figure 1: Annual Scientific Publication Trend (1980-2026)

According to the subject area classification presented in Figure 2, the majority of studies are concentrated in the disciplines of *Chemical Engineering*, *Chemistry*, and *Materials Science*, indicating that the use of sustainable carbon sources in adsorption processes is primarily shaped around understanding chemical processes, material synthesis, and mechanisms. Significant proportions in *Environmental Science* and *Energy* reveal that research is strongly related to environmental remediation applications and themes of energy efficiency and sustainability. Studies in other fields, represented at lower rates, support the multidisciplinary nature of the subject, pointing to a broad scientific framework developing at the intersection of materials science, environmental engineering, chemistry, and energy technologies.

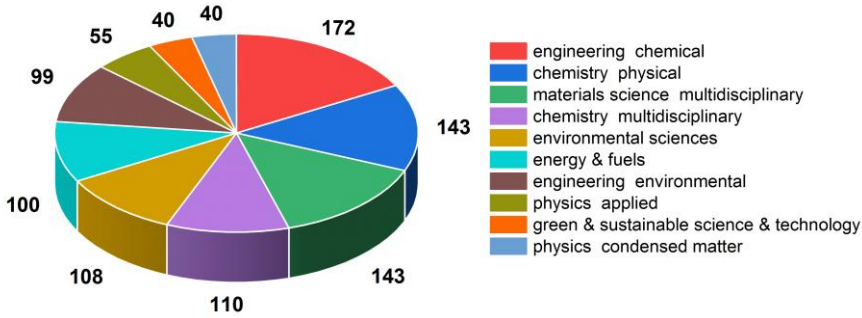


Figure 2: Top 10 Subject Areas of Publications

Most Relevant Sources

This bibliometric study analyzed a total of 254 sources, and the 25 sources with the highest number of publications are presented graphically in Figure 3. The analysis results clearly show that publications are concentrated in a few specific journals. *The Chemical Engineering Journal* stands out with the highest number of publications, revealing its position as the core knowledge production center in the field. It is followed by *Separation and Purification Technology* and *Journal of Environmental Chemical Engineering*; these three journals together form the core publishing platforms that guide the field. The concentration of publication volume in the first few sources indicates that researchers mostly prefer these high-impact channels to make their work visible. Journals following this core group exhibit a more balanced and widespread distribution of publication numbers. This indicates that the field is not dependent on a single journal, but that scientific influence is strongly clustered around the core journals.

Figure 4 presents the results of the Bradford's law analysis. The Bradford analysis shows that the literature is concentrated in a limited number of core journals. The graph shows that the number of publications is concentrated in a few highly productive sources (Zone 1, 18 sources), such as the *Chemical Engineering Journal*, *Separation and Purification Technology*, and *Journal of Environmental Chemical Engineering*, while the remaining sources form a broad middle band (Zone 2, 49 sources) and a long tail (Zone 3, 188 sources) with a gradually decreasing publication volume. This distribution clearly confirms that the field's core knowledge production is driven by a small number of core sources, while other journals expand the field's scope with more dispersed and lower-density contributions.

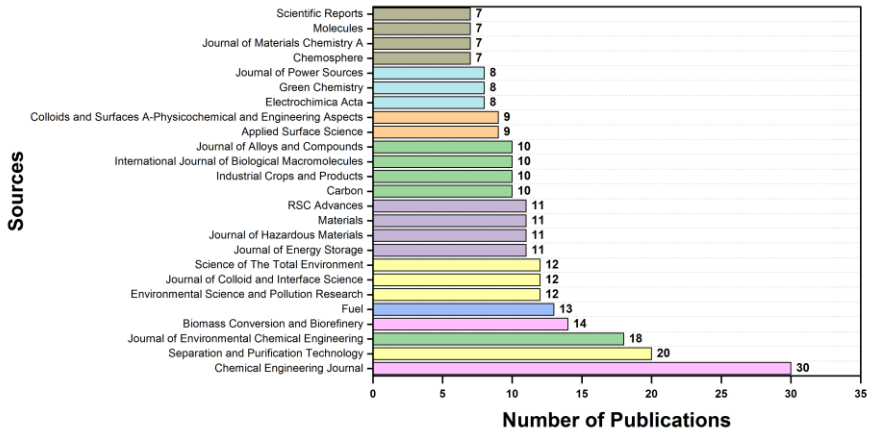


Figure 3: Top 25 Most Relevant Sources

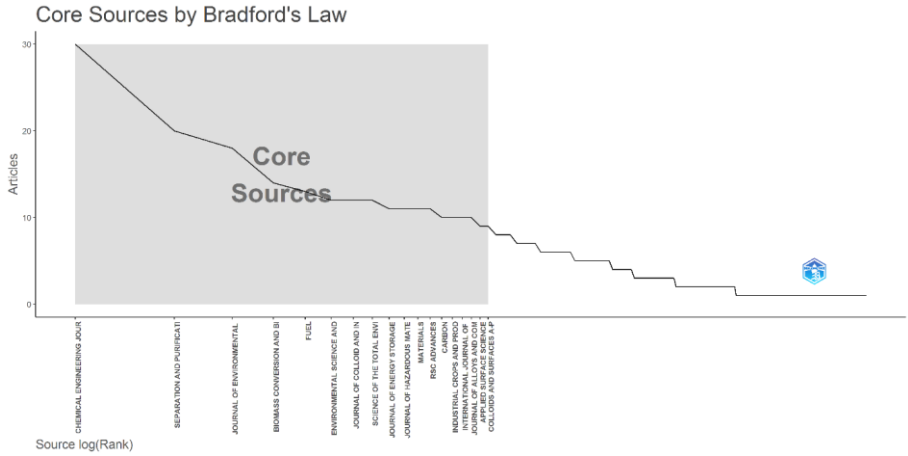


Figure 4: Core Sources According to Bradford's Law with A Three-Zone Model

Most Productive Authors

Figures 5 and 6 show the analysis of the top 10 most productive authors from 1980 to 2026. The analysis results reveal that the most prominent author is Wang Y., who has 16 publications in the WoS database since 2010. This author is followed in order of publication count by Bhatnagar A. (9), Jain A.K. (9), Li J. (9), Li X. (8), Wang J. (8), Chen Y. (7), Gupta V.K. (7), Kim H. (7), and Suhas (7).

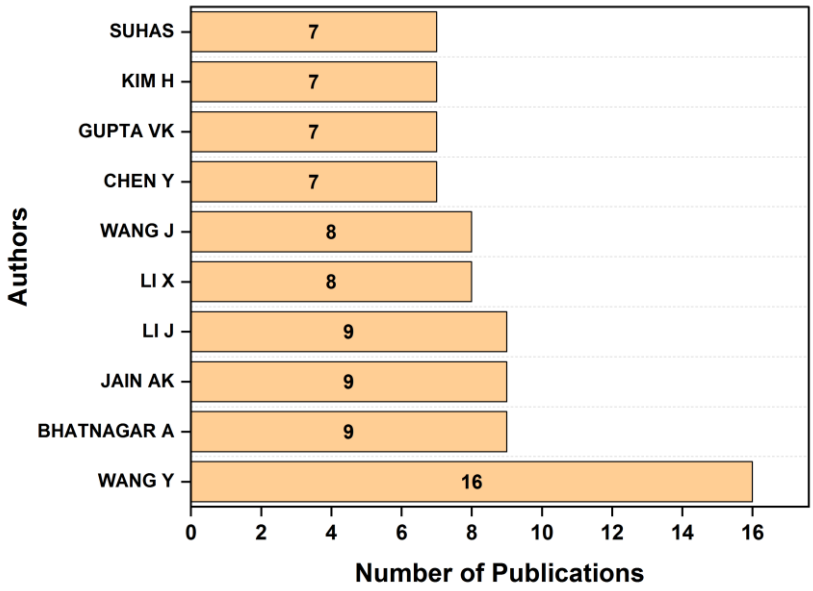


Figure 5: Top 10 Most Productive Authors

Figure 6 clearly illustrates the time-dependent productivity dynamics of prominent authors in the literature on sustainable carbon sources and adsorption. Authors such as Gupta VK (2000-2014), Jain AK (2002-2006), Bhatnagar A (2002-2007), and Suhas (2002-2006) shaped the fundamental methodological and applied framework of the field through their intensive publishing activities in the early stages. This productivity, which intensified particularly in the early 2000s, represents the first maturation phase in which the conceptual foundations of sustainable carbon-based adsorption research were laid. In contrast, authors such as Wang Y (2010-2026), Li X (2017-2025), Kim H (2018-2024), Chen Y (2019-2024), Wang J (2019-2024), and Li J (2022-2025) stand out with their increased publication numbers and higher annual citation impacts, particularly after 2015. This indicates that the research focus has shifted in recent years toward advanced functional carbon materials, performance optimization, and sustainability-based applications. Overall, the graph reveals that the field was shaped by a few pioneering researchers in its early stages, but in recent years, new and consistently productive authors have begun to dominate the literature, and the research momentum has increased significantly.

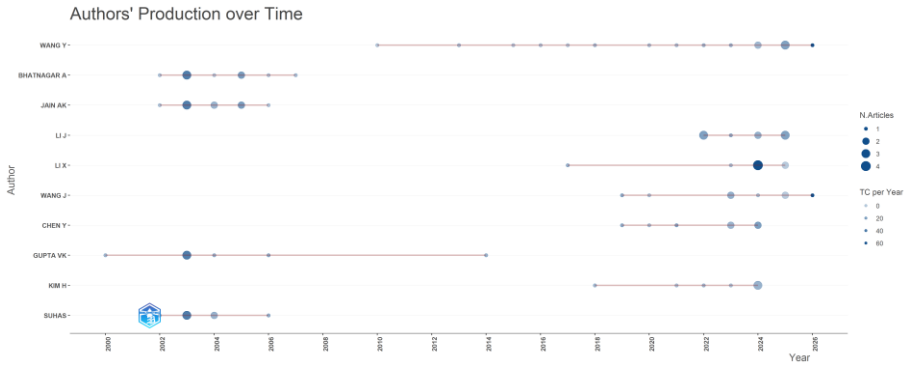


Figure 6: Authors' Production Over Time

Most Relevant Affiliations

Table 2 presents the data on the top 10 most productive institutions (affiliations) obtained from the bibliometric analysis. The data show that publication output is clearly concentrated around a few large research institutions. *Chinese Academy of Sciences* has by far the highest productivity (56 publications), reflecting the institution's leadership in the field, strong research funding, and extensive academic network. *Indian Institute of Technology System (IIT System)* (29 publications) and *South China University of Technology* (21 publications), which follow closely behind, demonstrate that Asia is the dominant research center in this field regionally. The total output of these three institutions represents a large portion of the top 10 institutions, indicating that the literature is strongly Asia-centric both geographically and institutionally.

The remaining institutions on the list, such as *Jiangsu University*, *King Fahd University of Petroleum and Minerals*, *Nanjing Forestry University*, *IIT Roorkee*, *Tianjin University*, *Guangxi University*, and *Qilu University of Technology*, demonstrate more balanced yet consistent productivity. This distribution shows that research activities are concentrated in China as well as in leading universities in India and the Middle East; therefore, the field's scientific contributions are shaped around a few strong national research systems. Overall, institution-based productivity data reveal that the research ecosystem is geographically weighted toward Asia and that universities with strong research infrastructure are the primary drivers of the literature.

Table 2. Top 10 most productive affiliations

Affiliation	Articles
Chinese Academy Of Sciences	56
Indian Institute of Technology System (IIT System)	29
South China University Of Technology	21
Jiangsu University	19
King Fahd University of Petroleum and Minerals	16
Nanjing Forestry University	15
Indian Institute of Technology (IIT) - Roorkee	14
Tianjin University	14
Guangxi University	13
Qilu University of Technology	12

Countries' Scientific Production

The country data reveals a very pronounced geographical concentration in the production of publications on sustainable carbon sources and adsorption (Figure 7). China's clear lead with 920 publications indicates that research capacity in this field has become largely China-centric and that the country's R&D investments in both material development and environmental applications are yielding strong results. The high productivity of other Asian countries such as India (171) and South Korea (89) confirms that the region has become a global center of gravity for research on sustainable carbon and adsorption in recent years. Production from countries outside Asia appears more limited and the distribution more balanced. Although countries such as the US (55), Japan (42), Saudi Arabia (44), Vietnam (37), Spain (36), Brazil (35), and the UK (35) make significant contributions, their growth rates remain lower than China's. This situation shows that, despite the field receiving widespread global interest, high-volume scientific production is concentrated in a few strong national research systems. The general trend indicates that sustainable carbon-based adsorption studies are particularly driven by Asian countries and that future scientific developments will continue to emerge largely from this region.

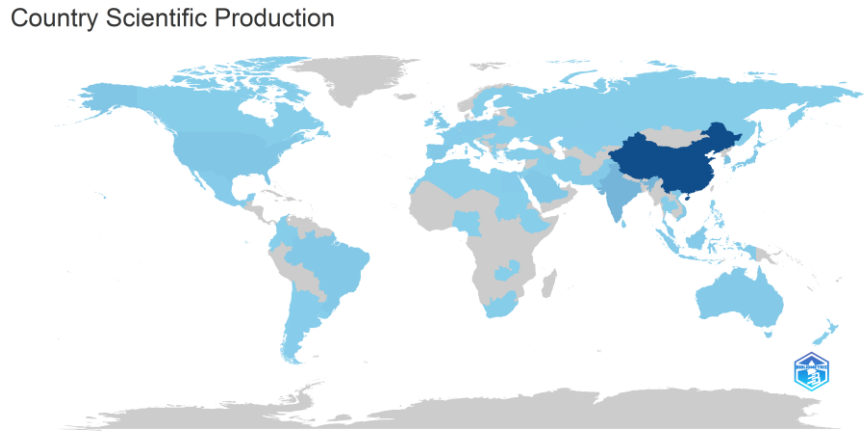


Figure 7: Map of Most Productive Countries

Most Global Cited Documents

The citation distribution shows that the information structure of sustainable carbon sources and adsorption areas exhibits both temporal continuity and thematic transformation (Table 3) The study published by Titirici (2010, *Chemical Society Reviews*), which received 1036 citations, stands out as the most influential publication in the field; it has become a reference point in the literature by establishing a conceptual framework for bio-carbon and sustainable carbon-based materials. The subsequent studies by Bi (2019, *Journal of Materials Chemistry* A) (803 citations) and Zhao (2016, *Journal of Power Sources*) (517 citations) demonstrate that adsorption research has integrated with energy storage and advanced material applications, extending beyond environmental remediation.

Table 3. Top 25 Most Global Cited Documents

Paper	DOI	Total Citations
Titirici MM, 2010, Chem Soc Rev	10.1039/b819318p	1036
Bi ZH, 2019, J Mater Chem A	10.1039/c9ta04436a	803
Zhao YQ, 2016, J Power Sources	10.1016/j.jpowsour.2016.01.020	517
Jain AK, 2003, J Hazard Mater	10.1016/S0304-3894(03)00146-8	453
Bhatnagar A, 2005, J Colloid Interf Sci	10.1016/j.jcis.2004.08.076	448
Jain AK, 2003, Sep Sci Technol	10.1081/SS-120016585	378
Zhu Ye, 2017, Adv Energy Mater	10.1002/aenm.201701222	328
Saini S, 2021, J Energy Storage	10.1016/j.est.2021.102646	328
Liao P, 2013, Chem Eng J	10.1016/j.cej.2013.04.118	302
Srivastava SK, 1989, Water Res	10.1016/0043-1354(89)90160-7	265
Yu LH, 2012, Langmuir	10.1021/la3024277	249
Gupta VK, 2006, J Colloid Interf Sci	10.1016/j.jcis.2006.02.017	248
Jain AK, 2004, Environ Sci Technol	10.1021/es034412u	245
Reguyal F, 2017, J Hazard Mater	10.1016/j.jhazmat.2016.10.006	234
Karaman C, 2022, Chemosphere	10.1016/j.chemosphere.2021.133346	211
Manasa P, 2022, J Energy Storage	10.1016/j.est.2022.105290	208
Gupta VK, 2000, Water Res	10.1016/S0043-1354(99)00322-X	188
Hoang AT, 2022, Chemosphere	10.1016/j.chemosphere.2021.131959	181
Sun ZQ, 2020, Chem Eng J	10.1016/j.cej.2019.122510	178
Zhang YP, 2005, Water Res	10.1016/j.watres.2005.07.019	168
Chen MF, 2017, Chemsuschem	10.1002/cssc.201700050	160
Hassan MM, 2021, Chemosphere	10.1016/j.chemosphere.2020.129087	160
Ajuria J, 2017, J Power Sources	10.1016/j.jpowsour.2017.04.107	152
Khalafallah D, 2021, Renew Energ	10.1016/j.renene.2021.01.077	151
Gao YF, 2020, Carbon	10.1016/j.carbon.2020.06.063	151

Earlier highly cited publications laid the methodological and experimental foundation for the field. Jain's 2003–2004 studies published in the *Journal of Hazardous Materials*, *Separation Science and Technology*, and *Environmental Science & Technology*, along with articles by Bhatnagar (2005) and Gupta (2006), are still heavily referenced in adsorption kinetics, isotherm models, and aqueous environment applications. Srivastava's (1989) early work continues to receive high citations despite the passage of time, demonstrating the theoretical continuity of the field. The fact that recent studies such as Zhu (2017) and Saini (2021) have reached hundreds of citations in a short time clearly shows that sustainable carbon-based

adsorbents and energy-related adsorption systems have taken a central place in the current research agenda.

Most Frequent Keywords

Figure 8 illustrates the most frequent keywords identified in the analysis. The prominent use of terms like *"activated carbon"* and *"adsorption"* indicates that activated carbon-based adsorbents are central to the research. The frequent use of terms such as *"removal", "water","aqueous solution,* and *"waste-water"* reveals that the research largely focuses on pollutant removal applications in water and wastewater treatment. However, the prominence of the keyword *"performance"* suggests that the studies focus not only on adsorption mechanisms but also on performance indicators of the developed carbon-based materials, such as adsorption capacity, kinetic behavior, selectivity, and reusability.

The association of terms such as "*biomass"*, "*biochar*", *"porous carbon"*, and *"waste"* indicates a strong research trend towards converting sustainable carbon sources into high-performance adsorbents. Furthermore, the striking frequency of carbon-based advanced material terms like "biochar", "porous carbon", "graphene oxide", and "nanoparticles" suggests a diversification of sustainable carbon sources and increasing interest in nanostructured materials. Keywords focusing on processes and mechanisms, such as "*kinetics*", "*equilibrium*", *"surface-area*", and "*pyrolysis*", indicate that the fundamental principles of adsorption behavior and the synthesis methods of carbon materials have been extensively studied in the literature. Overall, this distribution demonstrates that sustainable carbon sources and adsorption technologies constitute a multidisciplinary research field situated at the intersection of environmental remediation, materials science, and energy applications.

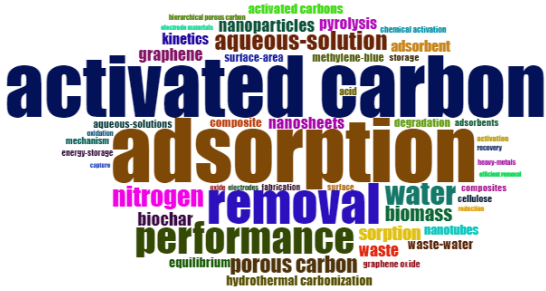


Figure 8: World Cloud of the Most Frequent "KeyWords Plus" Terms

Thematic Analysis

The thematic map examines global research on the sustainable carbon sources and adsorption technologies, and classifies research topics into four quadrants based on their level of development. Each quadrant represents a different type of research focus. The first quadrant includes motor themes, which are well-developed and widely studied topics that have attracted sustained attention from researchers over time. The second quadrant contains niche themes, which are well-structured but less central to the field. These topics are likely to receive more attention in the future and point to new research opportunities. The third quadrant represents emerging or declining themes, consisting of topics with limited development and low visibility. These themes either reflect new areas that have not yet matured or older topics that have gradually lost importance. The fourth quadrant includes basic themes, which are closely connected to the field but not deeply developed. Despite this, these fundamental topics continue to play an important role in carbon neutrality research (Tan, Yang, & Zhou, 2024).

Examining the thematic map in Figure 9, it can be seen that the literature on sustainable carbon sources and adsorption is structurally organized around basic-core-niche themes. The placement of the keywords *"adsorption"*, *"activated carbon",* and "*biochar*" in the Basic Themes region (high centrality, low density) indicates that these concepts form the foundation of the field and are associated with numerous studies, but are still open to development within themselves. This clearly shows that adsorption and carbon-based adsorbents constitute an indispensable backbone in the literature. The positioning of the clusters "*Performance*", "*nitrogen*" and *"porous carbon"* in the Motor Themes region shows that these themes have both high centrality and high density, meaning they actively drive the current research direction of the field. This result confirms that studies are increasingly focused on surface functionalization (especially nitrogen doping), pore structure engineering, and performance-oriented evaluations, as also seen in the Keyword Plus word cloud. In contrast, the lower centrality of terms such as "*water*", "*aqueous solution*" and "*removal*" indicates that these concepts have become established and methodologically mature themes and provide a supportive context for new research.

Overall, the thematic map shows that the field is evolving toward the structural design and performance optimization of sustainable carbon-based adsorbents, while still maintaining its roots in conventional water treatment and adsorption applications. This framework clearly demonstrates that research on sustainable carbon sources and adsorption is supported by a strong theoretical foundation and is simultaneously advancing dynamically in terms of advanced materials and functional performance.

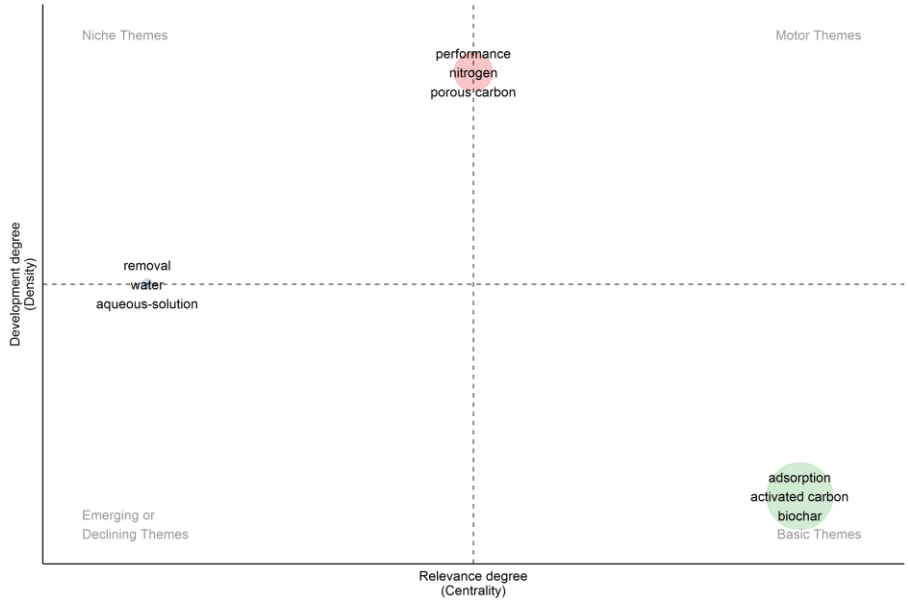


Figure 9: Thematic Map Generated Using "KeyWords Plus" Terms

Co-occurence Network

The keyword co-occurrence network shown in Figure 10 clearly reveals the conceptual structure of the literature on sustainable carbon sources and adsorption, as well as the relationships among its sub-themes. The high number of connections and the large node sizes of *"activated carbon"* and "*adsorption*" at the center of the network indicate that these concepts form the core of the literature and play a key role in linking different research themes. This central position is consistent with the identification of these topics as "basic themes" in the previous Thematic map and Keyword Plus analyses.

Two distinct clusters are evident in the network structure. The red cluster includes terms related to water treatment, pollutant removal, and conventional adsorption applications, while the blue cluster represents advanced material-oriented concepts such as porous carbon structures, surface modification, and performance enhancement. The strong connections between these two clusters indicate that the field is not only application-driven but also increasingly integrated with material design and performance optimization. Overall, the co-occurrence network clearly demonstrates that research on sustainable carbon-based adsorption is organized around a strong core and is evolving from traditional applications toward functional, high-performance carbon materials.

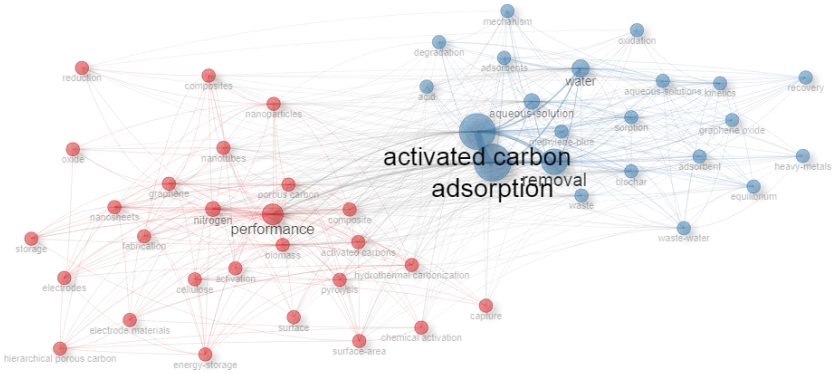


Figure 10: Co-occurrence Network of "KeyWords Plus" Terms

Collaboration Network

The international collaboration network for research on sustainable carbon sources and adsorption is depicted in Figure 11. This cross-country collaboration graph clearly reveals the global interaction structure of research in the field of sustainable carbon sources and adsorption. China, at the center of the network, stands out as the most dominant actor in terms of node size and connection density; it forms the main collaborative hub of the literature by establishing strong joint publication relationships with numerous countries. This, consistent with previously observed high publication output data, shows that China leads the field both quantitatively and structurally. In the graph, the US, South Korea, India, Japan, and European countries clustered around China represent secondary but strategic partners in the field. The intensity of connections between these countries indicates that research is not confined to a single geography, but rather that the flow of information is strengthened through international collaborations. In contrast, countries in a more environmentally oriented position with fewer connections contribute more limitedly to the literature or participate in the system through specific bilateral collaborations. Overall, the network clearly reveals that sustainable carbon-based adsorption research has a China-centric but globally integrated collaborative structure.

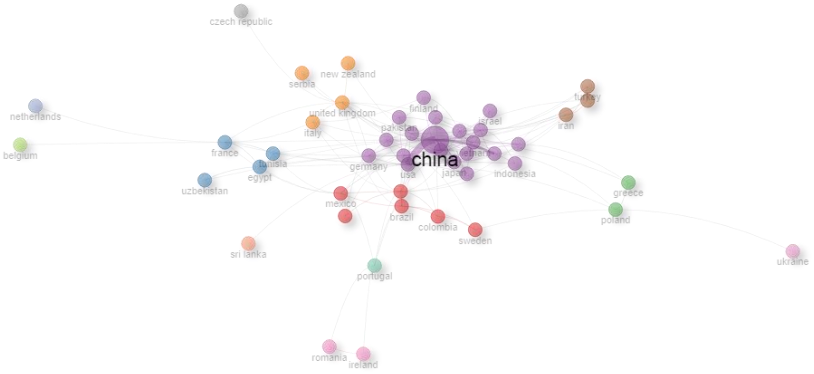


Figure 11: Country Collaboration Network

CONCLUSIONS

This study provides a thorough bibliometric analysis of scientific literature on sustainable carbon sources and adsorption published in the Web of Science database between 1980 and 2026. Analysis of a total of 679 publications clearly reveals that the field progressed at a limited volume before 2010, entered a development phase between 2010 and 2015, and gained rapid growth momentum after 2015. According to subject area classification, studies are predominantly concentrated in the disciplines of Chemical Engineering, Chemistry, and Materials Science, followed by Environmental Science and Energy. This distribution demonstrates that the use of sustainable carbon sources in adsorption processes is primarily shaped by studies based on chemical processes, material synthesis, and mechanisms; and also shows a strong connection to environmental remediation, energy efficiency, and sustainability themes.

When evaluated in terms of geographical distribution, research activities were found to be particularly concentrated in China, with leading universities in India and the Middle East also making significant contributions to the field. China's clear leadership with 920 publications demonstrates that global research capacity in sustainable carbon-based adsorption research has become largely China-centric. Keyword analyses reveal the clear dominance of the terms “activated carbon” and ‘adsorption’ and the high frequency of concepts such as “removal,” “water,” “aqueous

solution," and "wastewater," indicating that research is particularly focused on contaminant removal in water and wastewater treatment. Thematic map results show that adsorption and carbon-based adsorbents form the fundamental backbone of the literature, with a growing focus on surface functionalisation (particularly nitrogen doping), pore structure engineering and performance-oriented evaluations. Co-occurrence networks reveal that sustainable carbon-based adsorption studies are organised around a strong core structure, indicating a clear paradigm shift from classical applications to functional, high-performance carbon materials.

China is the dominant actor in terms of both node size and connection density, according to analysis of international collaboration networks; as a result of its strong co-publication relationships with numerous nations, it serves as the primary hub for collaboration in the literature. As secondary but strategic partners in the field, the United States, South Korea, India, Japan, and European nations centered around China stand out. The close ties between these nations show that adsorption research and sustainable carbon resources are not limited to a single region, but rather are a part of a global knowledge production and sharing structure that is reinforced by international cooperation.

REFERENCES

Ahmed, Muthanna J., Hameed, Bassim H., & Khan, Moonis Ali. (2024). Recent progress on carbonaceous materials-based adsorbents derived from cigarette wastes for sustainable remediation of aquatic pollutants: A review. *Journal of Analytical and Applied Pyrolysis, 183*, 106779. doi:https://doi.org/10.1016/j.jaap.2024.106779

Aria, Massimo, & Cuccurullo, Corrado. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959-975. doi:https://doi.org/10.1016/j.joi.2017.08.007

Baarimah, Abdullah O., Bazel, Mahmood A., Alaloul, Wesam Salah, Alazaiza, Motasem Y. D., Al-Zghoul, Tharaa M., Almuhaya, Basheer, . . . Mushtaha, Ahmed W. (2024). Artificial intelligence in wastewater treatment: Research trends and future perspectives through bibliometric analysis. *Case Studies in Chemical and Environmental Engineering, 10*, 100926. doi:https://doi.org/10.1016/j.cscee.2024.100926

Hassan, Waseem, & Duarte, Antonia Eliene. (2024). Bibliometric analysis: A few suggestions. *Current Problems in Cardiology, 49*(8), 102640. doi:https://doi.org/10.1016/j.cpcardiol.2024.102640

Reza, Md S., Afroze, Shammya, Kuterbekov, Kairat, Kabyshev, Asset, Zh. Bekmyrza, Kenzhebatyr, Haque, Md N., . . . Azad, Abul K. (2023).

Advanced Applications of Carbonaceous Materials in Sustainable Water Treatment, Energy Storage, and CO2 Capture: A Comprehensive Review. *Sustainability, 15*(11), 8815. doi:10.3390/su15118815

Shitole, Abhishek Madhav, Tiwari, Sumit, Sinhmar, Harender, & Bellos, Evangelos. (2025). Emerging trends in sustainable adsorption systems: Materials, optimization, and applications in carbon capture, cooling, storage, and atmospheric water harvesting. *Renewable and Sustainable Energy Reviews, 221*, 115899. doi:https://doi.org/10.1016/j.rser.2025.115899

Sun, Zhenli, Liao, Yun, Zhang, Yiyan, Sun, Shurui, Kan, Qihui, Wu, Zhiyao, . . . Wang, Xiangke. (2025). Sustainable carbon materials in environmental and energy applications. *Sustainable Carbon Materials, 1*(1). doi:10.48130/scm-0025-0002

Tan, Fenglong, Yang, Jun, & Zhou, Cheng. (2024). Historical review and synthesis of global carbon neutrality research: A bibliometric analysis based on R-tool. *Journal of Cleaner Production, 449*, 141574. doi:https://doi.org/10.1016/j.jclepro.2024.141574

Varghese, Shintu, Ralph, Peter, & Kuzhiumparambil, Unnikrishnan. (2025). Sustainable carbon-rich materials from algae and biomass: Synthesis, characterisation, and applications. *Results in Engineering, 27*, 106439. doi:https://doi.org/10.1016/j.rineng.2025.106439

Xu, Jiahui, Li, Yanzi, Zhang, Meiping, & Zhang, Shuhan. (2024). Sustainable agriculture in the digital era: Past, present, and future trends by bibliometric analysis. *Heliyon, 10*(14). doi:10.1016/j.heliyon.2024.e34612

Zhou, Shaofeng, Zhou, Lihua, Zhang, Yaping, Sun, Jian, Wen, Junlin, & Yuan, Yong. (2019). Upgrading earth-abundant biomass into three-dimensional carbon materials for energy and environmental applications. *Journal of Materials Chemistry A, 7*(9), 4217-4229. doi:10.1039/C8TA12159A

Noise Management for Sustainable Urban Living

Melike YALILI KILIC[1]
Semanur OZTURK[2]

1- Prof. Dr.; Bursa Uludag University Faculty of Engineering Environmental Engineering Department. myalili@uludag.edu.tr ORCID No: 0000-0001-7050-6742
2- MSc Student; Bursa Uludag University Faculty of Engineering Environmental Engineering Department. s.ozturrkk@gmail.com ORCID No: 0009-0009-6744-4857

ABSTRACT

Noise pollution has emerged as a significant environmental problem, particularly under conditions of rapid urbanization, adversely affecting the quality of urban living and exerting multidimensional impacts on human health and ecosystems. Transportation systems, industrial activities, and intensive land-use patterns contribute to increased levels of noise exposure, rendering this issue increasingly critical in relation to sustainable urban development goals. In this study, noise pollution in urban areas is examined within the framework of sustainable urban living, and its main sources and effects at the urban scale are evaluated. The effects of noise pollution on human health, ecosystems, and social well-being are analyzed based on findings from the existing literature. In addition, current policies and approaches to noise management implemented by the World Health Organization, the European Union, and Turkey are reviewed. The results indicate that the development of holistic and preventive approaches to noise management is essential for promoting sustainable and livable urban environments.

Keywords – Noise Pollution, Sustainable Urban Living, Urban Quality Of Life, Public Health, Noise Management.

INTRODUCTION

Cities in which individuals with diverse lifestyles coexist are currently undergoing a significant transformation driven by rapid population growth, technological advancements, and the intensification of industrial activities. This process not only reshapes the physical structure of urban spaces but also leads to a marked increase in environmental pressures, including noise pollution. Rapid and often unplanned urban development contributes to the expansion of uncontrolled construction and the progressive reduction of green spaces. In parallel, the continuous rise in the number of motor vehicles, together with the limited capacity of existing transportation infrastructure to accommodate this demand, places substantial pressure on urban transport systems. As a result, transportation-related noise has emerged as a major environmental problem, particularly in metropolitan areas (Cansaran, 2019; Onay, 2021; Sava et al., 2025).

Changes in the social, cultural, and economic structures of urban populations contribute to the emergence of various environmental challenges in cities. According to data from the Turkish Statistical Institute (TÜİK, 2024), 93.4% of the country's population resides in urban areas, indicating that risks associated with environmental pressures affect a large proportion of society. Environmental pressures arising particularly from rapid urbanization are increasingly emphasized as requiring a holistic approach

within the context of urban quality of life and sustainability (Türkoğlu, 2015). In this regard, noise pollution, together with increasing traffic density and industrial activities, has become one of the major environmental problems threatening the sustainability of urban ecosystems (Yerli and Demir, 2016). The cumulative impact of these pressures necessitates the development of comprehensive policies and the implementation of effective measures to safeguard healthy, livable, and resilient cities. Accordingly, noise management has emerged as a key policy area in efforts to promote sustainable urban living

These multifaceted urbanization challenges make the adoption of sustainable urban living strategies inevitable in order to safeguard quality of life and enhance the resilience of cities to future risks. Sustainable urbanization is defined as the guidance of urban development and transformation processes in accordance with the principles of sustainable development, requiring a holistic consideration of economic, social, and environmental dimensions (Tan et al., 2016). Under these conditions, reforms can contribute more effectively to sustainability objectives when they address not only the reduction of general environmental pressures but also the monitoring and control of urban noise levels. In this respect, the monitoring, mapping, and management of noise through action plans under the European Union's Environmental Noise Directive (2002/49/EC) have become an increasingly visible component of sustainable urban policies (Roy, 2009). In Turkey, strategic noise maps have been prepared, particularly by metropolitan municipalities, in line with the European Union harmonization process; however, the scope and effectiveness of these practices vary considerably depending on the institutional and technical capacities of local governments. The successful implementation of this approach requires the active involvement of not only public authorities but also civil society organizations and other stakeholders in decision-making and planning processes.

This study examines the noise problem affecting urban quality of life from a holistic perspective by evaluating the main sources and impacts of noise in urban areas. In addition, existing noise management approaches are reviewed, and recommendations for improvement are presented within the context of sustainable urban living.

NOISE POLLUTION AND ITS EFFECTS

Major Sources of Noise Pollution

Noise pollution is an environmental pollutant that emerges as a result of urbanization and human activities and poses a serious threat to public health when not adequately controlled. Major noise sources, including transportation, industrial activities, and urban land-use patterns (Figure 1), are associated not only with increased annoyance but also with sleep disturbances, cardiovascular diseases, and reduced quality of life (Basner et al., 2014; Münzel et al., 2014; MedlinePlus, 2025). In this context, the assessment of exposure to environmental noise requires consideration of not only noise levels but also the size of the exposed population and the duration of exposure to these sources (Faulkner and Murphy, 2022; Houthuijs et al., 2018).

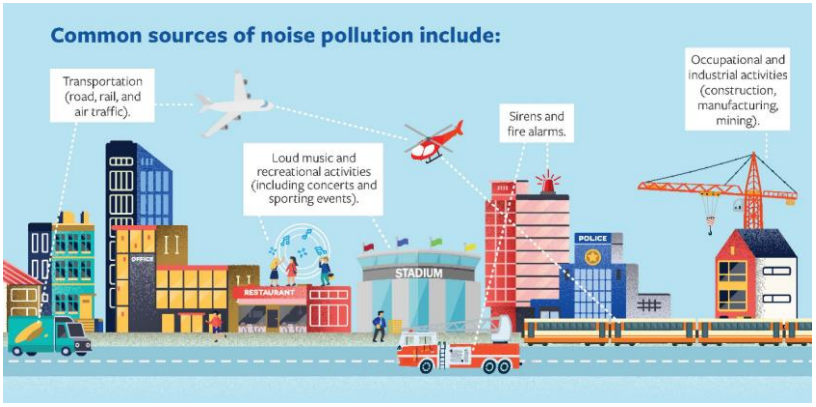


Figure 1: Major Sources of Noise Pollution in Urban Environments

Transportation systems represent the most widespread source of environmental noise exposure in urban areas. Road traffic generates continuous sound resulting from engine and exhaust emissions as well as tire–road surface interactions, and it is particularly influential throughout much of the day in residential areas located adjacent to major arterial roads. Modeling and noise mapping studies conducted in various European cities have demonstrated that road traffic noise affects significantly larger population groups during both daytime and nighttime periods compared to other sources such as railways, aviation, and industrial activities (Houthuijs et al., 2018; Faulkner and Murphy, 2022). Long-term exposure to road traffic noise has been shown to be associated with cardiovascular diseases (Cai et al., 2017; Yucel, 2017). Although railway and aircraft noise have a more limited spatial extent, they are identified as important sources due to their potential to cause sudden noise peaks and increased annoyance, particularly during nighttime hours (Basner et al., 2014).

Industrial facilities also constitute a significant source of noise for both workers and surrounding communities. Heavy machinery, compressors, pumps, and ventilation systems used in industrial processes can lead to various health problems among exposed employees, particularly hearing loss, while simultaneously increasing levels of annoyance and stress in residential areas located near industrial sites (Doğan and Aslan Çataltepe, 2018; Stanovská et al., 2024).

Field studies conducted in areas surrounding industrial zones have reported that residents exposed to industrial noise may simultaneously experience noise annoyance, symptoms of psychological stress, and increases in physiological stress indicators such as urinary cortisol levels (Onmek et al., 2024). Moreover, studies that have jointly examined noise originating from industrial facilities and road traffic indicate that the combined exposure to these sources can lead to higher perceived annoyance levels compared to exposure to single noise sources alone (Pierrette et al., 2012).

Urbanization patterns and land-use intensity also exert a decisive influence on urban noise levels. In a study conducted in Tehran, noise levels measured across different land-use types—including roads, residential areas, commercial zones, industrial areas, and green spaces—were compared. The results indicated that noise levels were statistically significantly higher in areas with high road and building density, whereas an inverse relationship was identified between the proportion of green spaces and noise levels (Naji et al., 2020). Similarly, a field study carried out in the city of Düzce demonstrated that noise levels measured in areas dominated by residential, commercial, and industrial land uses were higher than those observed in areas characterized by green spaces and low-density settlements (Yerli and Demir, 2016).

These findings highlight that noise pollution is influenced not only by the type of source but also by the spatial organization of the city, building density, and green space planning. They further demonstrate that urban planning decisions can play a critical role in reducing long-term noise exposure.

Effects of Noise Pollution on Human Health, Ecosystems, and Social Well-Being

Long-term exposure to environmental noise is recognized as a significant environmental risk factor that affects human health at both physiological and psychosocial levels. Numerous studies have demonstrated that prolonged exposure to various noise sources, particularly road traffic noise, is associated with cardiovascular outcomes such as hypertension, coronary heart disease, and stroke (Babisch, 2014; Héritier et al., 2017). These health effects are not limited to hearing loss; rather, they generate a broader physiological burden through mechanisms including stress

activation, disruption of hormonal balance, and impacts on the autonomic nervous system.

One of the earliest and most widespread effects of noise on human health is sleep disturbance. During nighttime hours, low-intensity but persistent environmental noise can delay sleep onset, increase the frequency of nocturnal awakenings, and ultimately lead to daytime sleepiness and reduced cognitive performance (Halperin, 2014). Studies focusing on children have shown that exposure to environmental noise extends beyond subjective annoyance, resulting in elevated stress hormone levels, behavioral problems, and adverse effects on reading ability, memory, and academic performance, particularly in noisy school environments (Stansfeld and Clark, 2015). For this reason, noise control should be regarded as a fundamental public health priority not only for adult health but also for children's cognitive development and educational attainment.

The effects of noise pollution are not limited to human health but also extend to species inhabiting terrestrial and aquatic ecosystems. A study examining the impacts of anthropogenic noise on animals demonstrated significant changes in communication, movement, foraging behavior, and physiological stress responses across a wide range of species, including amphibians, birds, fish, and mammals (Kunc and Schmidt, 2019). Another study further reported that chronic exposure to noise levels exceeding 40 dB(A) leads to behavioral alterations and disruptions in ecological functions in many species, thereby affecting ecosystem services such as predator–prey interactions, pollination, and seed dispersal (Shannon et al., 2016). These findings indicate that noise constitutes an important environmental pressure not only for individual organisms but also for community structures and ecosystem dynamics.

Noise also plays a decisive role in individuals' quality of life, psychosocial well-being, and satisfaction with the urban environment. Studies conducted in urban settings have shown that higher levels of road traffic noise are associated with increased annoyance, stress-related symptoms, and, in some cases, sleep disturbances. In contrast, access to quiet or relatively calm areas in the vicinity of residential environments can have a protective effect on quality of life (Öhrström et al., 2006). A large-scale study involving populations living near airports further demonstrated that increasing levels of aircraft noise–related annoyance are linked to significant deterioration in both environmental quality of life and health-related quality of life indicators. Moreover, individual characteristics such as noise sensitivity were found to strengthen this relationship (Schreckenberg et al., 2010). For this reason, noise management should not focus solely on the technical reduction of sound levels but should also emphasize the preservation of quiet areas, the consideration of sensitive groups in land-use and transportation planning, and the comprehensive promotion of social well-being.

NOISE MANAGEMENT APPROACHES

Effective management of noise pollution at the urban scale requires not only measurement and monitoring practices but also the integration of planning and control strategies within a holistic framework. Therefore, noise control strategies should be addressed through a systematic approach. In the literature, noise management practices are commonly classified into three main approaches. The first approach focuses on controlling noise at its source; the second aims to reduce noise levels along the transmission path between the source and the receiver. The third approach is based on protecting individuals exposed to noise and includes control measures applied at the receiver level (ISO 11690-1, 1996).

Traditional noise control methods are generally discussed in the literature under acoustic insulation practices, noise barriers, and planning-based approaches. These methods encompass structural and spatial measures designed to limit noise propagation and reduce exposure levels. Particularly when incorporated during the planning and design stages of urban and built environments, such approaches are considered fundamental components of effective noise management (Vér and Beranek, 2006; ISO 11690-1, 1996; Şahin, 2003).

Measures for Noise Control

Measures for noise control encompass technical and managerial practices aimed at regulating disturbing sounds occurring in urban and built environments. These measures can be implemented through various strategies, including the reduction of noise levels, the limitation of exposure duration, and the mitigation of perceived annoyance. Noise control is therefore regarded as a holistic approach that seeks to establish an acceptable acoustic environment in alignment with technical feasibility and economic conditions (Tufaner, 2010).

In determining appropriate noise control measures, not only the level of noise exposure but also the duration of exposure, the temporal characteristics of noise, and the sensitivity of the affected population should be considered. Consequently, noise management should be viewed as a multidimensional field that extends beyond assessments based solely on technical threshold values and should be addressed in conjunction with health-based risk assessments and urban planning decisions.

New Technologies and Smart City Applications

The limitations of traditional noise control methods have made the integration of new technologies and smart city applications into noise management necessary, particularly in areas with complex urban structures and dense transportation networks. In recent years, advances in acoustic sensor networks, geographic information systems (GIS), and real-time data

collection technologies have enabled continuous monitoring of urban noise levels and the production of noise maps that reflect spatial and temporal variations. These approaches allow for the dynamic identification of noise sources and the development of targeted intervention strategies (Maisonneuve et al., 2009; Murphy and King, 2022). A schematic representation of IoT sensors, fog computing, and cloud-based data processing architectures employed in smart city applications is presented in Figure 2.

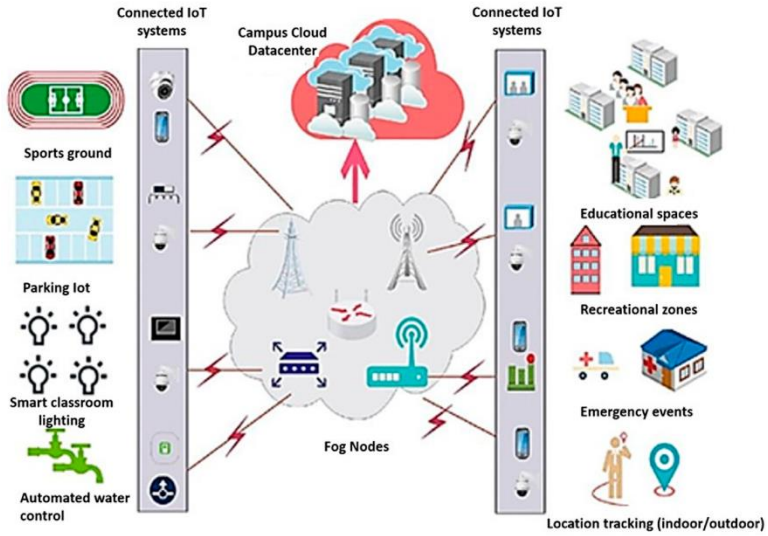


Figure 2: Schematic Representation of Sensor-Based Noise Monitoring Systems in Smart City Applications (İşler, 2025)

Applications developed within the smart city framework are not limited to measurement and monitoring activities; rather, they are integrated with decision-support systems and directly contribute to the planning and management processes of local authorities. The combined use of sensor-based noise monitoring systems with big data analytics and numerical modeling tools enables noise-related problems to be managed in a more flexible, predictive, and effective manner. In this respect, new technologies and smart city applications transform noise management from a purely technical control tool into a comprehensive management approach that supports sustainable urban planning and enhances quality of life (Bibri and Krogstie, 2017).

Contribution of Noise Management to Environmental, Social, and Economic Sustainability

Noise management is regarded as a multidimensional field of urban environmental governance that is directly linked to the environmental,

social, and economic dimensions of sustainability. From an environmental perspective, noise reduction contributes to mitigating environmental pressures through its positive effects on ecosystems, wildlife, and overall urban environmental quality. In particular, controlling continuous noise generated by transportation systems and industrial activities is considered an essential component of sustainable urban development goals (World Health Organization, 2018).

From the perspective of social sustainability, noise management plays a critical role in improving quality of life, preserving sleep patterns, and enhancing public health. The literature emphasizes that long-term exposure to noise is associated with increased stress, reduced cognitive performance, and elevated cardiovascular disease risks; conversely, effective noise policies are shown to mitigate these adverse effects and support overall social well-being (Stansfeld and Matheson, 2003; Basner et al., 2014).

From an economic sustainability perspective, controlling noise pollution provides indirect economic benefits by reducing healthcare costs, increasing labor productivity, and improving the overall livability of urban areas. Reports published by the European Environment Agency indicate that the health-related costs associated with environmental noise amount to billions of euros across Europe; in contrast, effective noise management policies are shown to generate substantial economic benefits in the long term (European Environment Agency, 2020).

NOISE MANAGEMENT POLICIES AND APPROACHES

World Health Organization Approach to Noise Management

The World Health Organization (WHO) addresses environmental noise as a major public health issue at the global level and evaluates the health impacts of noise exposure through evidence-based guidelines. The Organization's principal policy document in this field is the Environmental Noise Guidelines for the European Region. This guideline defines health-based threshold values aimed at reducing exposure to environmental noise and provides a scientific reference framework for policymakers at multiple governance levels (World Health Organization, 2018).

Within the WHO framework, noise management is not limited to environmental measurement and technical control practices; rather, it is evaluated through the short- and long-term health effects of noise exposure. In the guideline, major noise sources—including road traffic, railway traffic,

air transport, and wind turbines—are addressed separately, and prolonged exposure to these sources is emphasized as being associated with cardiovascular diseases, sleep disturbances, noise annoyance, and cognitive effects (World Health Organization, 2018). The recommended threshold values are based on systematic evaluations of available epidemiological and experimental evidence.

One of the key features of the WHO guideline is the classification of noise-related recommendations according to levels of evidence. Taking into account the strength of evidence regarding health effects, strong recommendations are provided for certain noise sources, while conditional recommendations are proposed for others. This approach enables policymakers to make flexible yet health-oriented decisions by considering local conditions and existing scientific uncertainties (World Health Organization, 2018). Consequently, noise management is addressed within a risk-based and preventive public health framework rather than relying solely on rigid limit values.

The WHO also emphasizes the importance of the Health Impact Assessment (HIA) approach in the evaluation and management of environmental noise. This approach aims to anticipate the potential effects of noise exposure on population health and to systematically assess the health consequences of policy or planning decisions (World Health Organization, 2018). From this perspective, the WHO's approach to noise management positions environmental noise not merely as a technical environmental issue but as a comprehensive policy domain focused on the protection of public health.

Overall, the WHO approach to environmental noise management is based on the principles of establishing scientifically grounded, evidence-based threshold values, protecting vulnerable groups, and reducing long-term health risks. This framework serves as a complementary guide for noise policies developed at national and local levels and promotes the consideration of noise control from a health-oriented perspective (World Health Organization, 2018).

Noise Management Approaches in the European Union

In the European Union, environmental noise management is addressed as a comprehensive policy domain aimed at protecting human health and enhancing overall quality of life. The primary legal framework in this field was established through the Environmental Noise Directive (2002/49/EC), adopted by the European Parliament and the Council. The main objective of the Directive is to assess exposure to environmental noise, prevent adverse effects of noise on human health, and develop a common approach to assessment and reporting among Member States (European Parliament and Council, 2002).

Under the provisions of the Directive, Member States are required to prepare strategic noise maps for major urban agglomerations as well as for major roads, railways, and airports. These maps provide a quantitative representation of population exposure to noise and enable the identification of areas at risk. The use of the L_den and L_night indicators is mandated in noise assessment in order to ensure a comparable evaluation of noise exposure during daytime, evening, and nighttime periods (European Parliament and Council, 2002).

Another key component of the Directive is the action plans prepared on the basis of strategic noise maps. These action plans include priorities such as reducing noise levels, protecting vulnerable groups, and preserving quiet areas. In this regard, environmental noise is addressed not merely as a technical environmental issue but as a risk factor directly linked to public health (European Parliament and Council, 2002). Reports published by the European Environment Agency indicate that transport-related noise continues to pose a significant health risk for millions of people across Europe and that noise management remains a persistent item on the policy agenda (European Environment Agency, 2020).

The European Union also places particular emphasis on the harmonization of noise assessment methods. In this context, the CNOSSOS-EU (Common Noise Assessment Methods in Europe) framework was developed to provide a common methodology for the calculation of environmental noise and to enhance data comparability among Member States. By defining shared technical principles for the assessment of road traffic, railway, aircraft, and industrial noise sources, the CNOSSOS-EU method strengthens the scientific consistency and comparability of strategic noise mapping across Europe (European Commission, 2015).

Overall, the European Union's approach to environmental noise management offers a policy framework that goes beyond measurement, mapping, and reporting by integrating long-term planning, health-based threshold values, and the protection of quality of life at the urban scale. This framework enables the systematic development of preventive and mitigation-oriented practices addressing environmental noise across Member States (European Parliament and Council, 2002; European Environment Agency, 2020).

Noise Management Approaches in Turkey

In Turkey, environmental noise management is largely based on legal and institutional arrangements developed in alignment with European Union environmental legislation. The fundamental framework in this field has been shaped in line with the Environmental Noise Directive (2002/49/EC) adopted by the European Parliament and the Council, and national regulations on the assessment and management of environmental noise have been established within the scope of Turkey's European Union

harmonization process (European Parliament and Council, 2002). In this context, the principal national legal instrument governing environmental noise in Turkey, the Regulation on the Assessment and Management of Environmental Noise, entered into force following its publication in the Official Gazette on 28 May 2010 (Regulation on the Assessment and Management of Environmental Noise, 2010).

The Regulation mandates the preparation of strategic noise maps and the development of action plans based on these maps in order to identify and reduce the adverse effects of environmental noise on human health and quality of life. Within the scope of the Regulation, strategic noise maps are prepared for urban agglomerations with populations exceeding 100,000, as well as for areas surrounding major roads, railways, and airports. Through these maps, the size of the population exposed to noise and the associated risk levels can be systematically determined (Regulation on the Assessment and Management of Environmental Noise, 2010). Subsequently, the action plans developed include measures aimed at reducing noise levels, preserving quiet areas, and integrating noise considerations into urban planning processes.

In Turkey, the approach to noise management is not limited to the identification of environmental noise sources but also encompasses enforcement and sanction mechanisms aimed at implementation. The Regulation defines the distribution of responsibilities between central authorities and local administrations, assigning municipalities and relevant public institutions responsibility for the preparation of noise maps, the implementation of action plans, and monitoring and inspection processes (Ministry of Environment and Urbanization, 2011). In addition, noise limit values related to construction activities, entertainment venues, and workplaces are regulated through relevant national legislation and technical guidelines (Ministry of Labor and Social Security, 2013).

This holistic approach demonstrates that noise management in Turkey is not confined solely to technical measurement and reporting processes, but is addressed as a multidimensional environmental policy domain aimed at protecting public health, enhancing quality of life, and promoting sustainable urban development. Turkey’s current legislative framework provides a structure that is largely aligned with European Union standards in environmental noise management and offers policy instruments that are applicable at the local level (European Parliament and Council, 2002; Regulation on the Assessment and Management of Environmental Noise, 2010).

CONCLUSION

In this study, noise pollution has been examined within the context of urban living, and its underlying causes and impacts have been evaluated. Transportation systems, industrial activities, and intensive land-use patterns have been identified as key factors contributing to increasing noise levels. This situation results in continuous noise exposure, particularly in densely populated urban areas.

An evaluation of the effects of noise pollution on human health demonstrates that the problem is not limited to hearing-related impacts alone. Long-term exposure to environmental noise may lead to adverse outcomes affecting cardiovascular health, sleep patterns, and overall quality of life. Moreover, the impacts of noise on ecosystems manifest through changes in species behavior and disruptions to ecological processes, thereby negatively affecting natural balance. These findings highlight that noise pollution is a multifaceted issue with both environmental and social dimensions.

An assessment of noise management policies and practices indicates that the approaches developed by the World Health Organization and the European Union provide important guidance for integrating noise monitoring into planning and decision-making processes. In Turkey, despite the existence of a comprehensive legal framework, the effectiveness of implementation varies depending on the technical capacity and institutional performance of local governments. This situation demonstrates that noise management should be supported not only by legal regulations but also by effective, implementation-oriented mechanisms.

Accordingly, measures aimed at reducing noise pollution should be addressed in conjunction with transportation planning, land-use policies, and building development decisions. The accurate identification of noise sources, the protection of sensitive and quiet areas, and the development of locally applicable solutions can contribute significantly to improving urban quality of life. Ultimately, the management of noise pollution should be regarded as one of the fundamental components in the pursuit of healthier, more livable, and sustainable cities.

REFERENCES

Babisch, W., Wölke, G., Joachim Heinrich, J., and Straff, W. (2014). Road traffic noise and hypertension—accounting for the location of rooms. *Environmental Research,* 133, 380-387. https://doi.org/10.1016/j.envres.2014.05.007

Basner, M., Babisch, W., Davis, A., Brink, M., Clark, C., Janssen, S., and Stansfeld, S. (2014). Auditory and non-auditory effects of noise on health. *The Lancet*, 383(12), 1325-1332. http://dx.doi.org/10.1016/S0140-6736(13)61613-X

Bibri, S.E., and Krogstie, J. (2017). Smart sustainable cities of the future: An extensive interdisciplinary literature review. *Sustainable Cities and Society*, 31, 183-212. https://doi.org/10.1016/j.scs.2017.02.016

Cai, Y., Hansell, A.L., Blangiard, M., Burton, P.R., BioSHaRE, de Hoogh, K., Doiron, D., Fortier, I., Gulliver, J., Hveem, K., Mbatchou, S., Morley, D.W., Stolk, R.P., Zijlema, W.L., Elliott, P., and Hodgson, S. (2017). Long-term exposure to road traffic noise, ambient air pollution, and cardiovascular risk factors in the HUNT and lifelines cohorts. *European Heart Journal*, 38, 2290-2296.

Cansaran, D., (2019). A study on the determination of noise pollution level: Amasya Sample. *Ankara University Faculty of Political Science Journal*, 74(1), 89-108.https://doi.org/10.33630/ausbf.523208

Doğan, H., and Çataltepe, Ö.A. (2018). The Effect of the Noise on Human Health, *Journal of Health and Sport Sciences (JHSS)*, 1(1-2-3), 29-38.

European Parliament and Council. (2002). Directive 2002/49/EC of the European Parliament and of the Council relating to the assessment and management of environmental noise. Official Journal of the European Communities, L189, 12-25.

European Commission. (2015). Commission Directive (EU) 2015/996 establishing common noise assessment methods according to Directive 2002/49/EC of the European Parliament and of the Council (CNOSSOS-EU). Official Journal of the European Union, L168, 1-823.

European Environment Agency. (2020). Environmental noise in Europe – 2020. EEA Report No 22/2020. Luxembourg: Publications Office of the European Union.

Faulkner, J.-P., and Murphy, E. (2022). Road traffic noise modelling and population exposure estimation using CNOSSOS-EU: Insights from Ireland. *Applied Acoustics*, 192, 108692. https://doi.org/10.1016/j.apacoust.2022.108692

Halperin, D. (2014). Environmental noise and sleep disturbances: A threat to health?. *Sleep Science*, 7(4), 209-212. https://doi.org/10.1016/j.slsci.2014.11.003

Héritier, H., Vienneau, D., Foraster, M., Eze, I.C., Schaffner, E., Thiesse, L., Rudzik, F., Habermacher, M., Köpfli, M., Pieren, R., Brink, M., Cajochen, C., Wunderli, J.M., Probst-Hensch, N., and Röösli, M. (2017). Transportation noise exposure and cardiovascular mortality: A nationwide cohort study from Switzerland. *European Journal of Epidemiology*, 32(4), 307-315.

Houthuijs, D., Fons, J., de la Maza, M.S., Swart, W., and van Beek, A. (2018). Method to estimate the road traffic noise exposure distribution in Europe. ETC/ACM Technical Paper 2017/10. European Topic Centre on Air Pollution and Climate Change Mitigation.

ISO 11690-1 (1996). Acoustics — Recommended practice for the design of low-noise workplaces containing machinery — Part 1: Noise control strategies. International Organization for Standardization, Geneva.

İşler, B. (2025). Urban sound recognition in smart cities using an IoT–fog computing framework and deep learning models: a performance comparison. *Applied Sciences*, 15, 1201 https://doi.org/10.3390/app15031201.

Kunc, H.P., and Schmidt, R. (2019). The effects of anthropogenic noise on animals: A meta-analysis. Biology Letters, 15(11), 20190649. https://doi.org/10.1098/rsbl.2019.0649

Maisonneuve, N., Stevens, M., Niessen, M.E., Hanappe, P., and Steels, L. (2009). Citizen noise pollution monitoring. The proceedings of the 10 th International Digital Government Research Conference, 96-103.

MedlinePlus. (2025). Noise Pollution. U.S. National Library of Medicine. Available at: https://magazine.medlineplus.gov/pdf/Noise_pollution.pdf (accessed 15.12.2025).

Ministry of Environment and Urbanization. (2011). Guidelines for the Preparation of Strategic Noise Maps and Action Plans. Ankara: Republic of Turkey Ministry of Environment and Urbanization.

Ministry of Labor and Social Security. (2013). Regulation on the Protection of Employees from Noise-Related Risks. Ankara: Republic of Turkey Ministry of Labor and Social Security.

Murphy, E., and King, E.A. (2022). Environmental noise pollution: Noise mapping, public health, and policy. Elsevier, ISBN 978-0-12-820100-8

Münzel, T., Gori, T., Babisch, W., and Basner, M. (2014). Cardiovascular effects of environmental noise exposure. *European Heart Journal*, 35(13), 829-836. https://doi.org/10.1093/eurheartj/ehu030

Naji, F.K., Djahed, B., Mahvi, A.H., Nowrouz, P., Nazmara, S., and Kardooni, H. (2020). Urban land use effect analysis on the level of noise pollution using satellite and GIS technologies: A case study in Tehran city. *Global NEST Journal*, 22(3), 361-368. https://doi.org/10.30955/gnj.003337

Onay, B. (2021). Landscape architecture approaches to prevent noisepollution in school gardens and surroundings: the case of Isparta. Suleyman Demirel University, Institute of Science, Department of Landscape Architecture, PhD Thesis, 181 pages, Isparta.

Onmek, N., Kengtrong Bordeerat, N., and Wangwongwatana, S. (2024). The Effects of Environmental Noise on Annoyance, Stress, and Urine Cortisol Levels Among Residents Living Near Industrial Sites in Bangkok, Thailand. *Kesmas: Jurnal Kesehatan Masyarakat Nasional*, 19(1), 18-26. DOI: 10.21109/kesmas.v19i1.7521

Öhrström, E., Skånberg, A., Svensson, H., and Gidlöf-Gunnarsson, A. (2006). Effects of road traffic noise and the benefit of access to quietness. *Journal of Sound and Vibration*, 295(1-2), 40-59. https://doi.org/10.1016/j.jsv.2005.11.034

Pierrette, M., Marquis-Favre, C., Morel, J., Rioux, L., Vallet, M., Viollon, S., and Moch, A. (2012). Noise annoyance from industrial and road traffic combined noises: A survey and a total annoyance model comparison. *Journal of Environmental Psychology*, 32(2), 178-186. doi:10.1016/j.jenvp.2012.01.006

Roy, M. (2009). Planning for sustainable urbanisation in fast growing cities: mitigation and adaptation ıssues addressed in Dhaka, Bangladesh, Habitat

International, 33(3), 276-286. https://doi.org/10.1016/j.habitatint.2008.10.022

Sava, B., Kuş Şahin, C., and Coşkun, S. (2025). An analysis of noise pollution in eğirdir district of Isparta province. *Turkish Journal of Forest Science*, 9(1), 59-74. https://doi.org/10.32328/turkjforsci.1549844

Schreckenberg, D., Meis, M., Kahl, C., Peschel, C., and Eikmann, T. (2010). Aircraft Noise and Quality of Life around Frankfurt Airport. International *Journal of Environmental Research and Public Health*, 7(9), 3382-3405. https://doi.org/10.3390/ijerph7093382

Shannon, G., McKenna, M.F., Angeloni, L.M., Crooks, K.R., Fristrup, K.M., Brown, E., Warner, K.A., Nelson, M.D., White, C., Briggs, J., McFarland, S., and Wittemyer, G. (2016). A synthesis of two decades of research documenting the effects of noise on wildlife. *Biological Reviews*, 91(4), 982-1005. https://doi.org/10.1111/brv.12207

Stanovská, M., Tomášková1, H., Šlachtová, H., Potužníková, D., and Argalášová, L. (2024). Health impact of environmental and industrial noise - A narrative review. *Med Pr Work Health Saf.*, 75(5), 425-437.

Stansfeld, S.A., and Matheson, M.P. (2003). Noise pollution: Non-auditory effects on health. *British Medical Bulletin*, 68(1), 243-257. https://doi.org/10.1093/bmb/ldg033

Stansfeld, S., and Clark, C. (2015). Health effects of noise exposure in children. *Current Environmental Health Reports*, 2(2), 171-178. DOI 10.1007/s40572-015-0044-1

Şahin, E. (2003). Noise control techniques-a case study, J. Fac. Eng. Arch. Gazi Univ. 18(4), 67-80.

Tan, Y., Xu, H., and Zhang, X. (2016). Sustainable development and quality of urban life. *Procedia – Social and Behavioral Sciences*, 202, 10-14. https://doi.org/10.1016/j.sbspro.2015.08.203

Tufaner, F. (2010). Policies on Turkey's environmental noise control an management. *Çankırı Araştırmaları Dergisi*, 5(5-6), 234-245.

Türkoğlu, H. (2015). Sustainable development and quality of urban life. https://doi.org/10.1016/j.cities.2015.03.001

TÜİK. (2024). Turkish Statistical Institute. Address-Based Population Registration System Results, Ankara.

Vér, I.L., and Beranek, L.L. (2006). Noise and Vibration Control Engineering: Principles and Applications, Wiley.

World Health Organization. (2018). Environmental Noise Guidelines for the European Region. Copenhagen: WHO Regional Office for Europe.

Yerli, O., and Demir, Z. (2016). Relation between urban land uses and noise: A case study in Duzce, Turkey. *Oxidation Communications*, 39(1-II), 732-745.

Yucel, M. Kahveci, B., and Colakkadioglu, D. (2017). Modelling the Adverse Health Effects of Road Traffic Noise: A Case Study in Adana, Bulent Angin Boulevard. *J. Int. Environmental Application & Science*, 12(4), 325-333.

Mechanical Performance and Cost Analysis of Pumice Blocks with Afsin–Elbistan Fly Ash Addition

Mustafa EKEN[1]

[1] Assoc. Prof., Kahramanmaras Istiklal University, Department of Construction Technology, Orcid: 0000-0002-7559-876X

INTRODUCTION

Global population growth, industrialization, and urbanization have rapidly increased energy demand and placed significant pressure on the built environment to improve energy efficiency (building materials consume a large portion of energy throughout lifecycle) (Usta, P., & Zengin, B. (2021). The construction sector is one of the major contributors to total energy usage and environmental impact, consuming approximately 38–40 % of global energy in building production and operation phases (Yüksek, İ. (2015). In many countries, including Turkey, energy consumption for building heating and comfort requirements remains high, emphasizing the need for efficient and sustainable materials (Usta, P., & Zengin, B. (2021).

Conventional building materials, particularly cement-based products, are environmentally disadvantageous due to high energy requirements and CO_2 emissions during production (cement manufacture contributes significantly to greenhouse gas emissions) (Juarez, R. I. C., & Finnegan, S. (2021). The environmental impact of cement production in Europe: A holistic review of existing EPDs. *Cleaner Environmental Systems*, *3*, 100053.). In this context, substituting cement with industrial by-products and waste materials has emerged as a practical approach to reduce environmental impact and lower construction costs. Fly ash, a by-product of coal combustion in thermal power plants, possesses pozzolanic properties and is widely studied as a cement replacement that can improve sustainability and durability in cementitious materials (Nsobundu, A. E. J., & Tiza, M. T. A).

Pumice blocks are widely used in the building sector due to their lightweight nature and thermal insulation properties, which can improve building energy performance and reduce structural loads (review on pumice-based concrete identifies pumice as a sustainable lightweight aggregate) (Moolchandani, K. (2025). However, the significant amount of cement used in their production increases both cost and environmental impact. Therefore, this study aims to develop a technically and economically sustainable solution by replacing cement with fly ash from the Afsin–Elbistan region in pumice block production, evaluating both mechanical performance and cost implications.

MATERIALS AND METHODS

1.Materials

In this study, Portland cement CEM I 42.5 R was utilised as the binder. The properties of CEM I 42.R are presented in Table 1. Fly ash was obtained from the Afsin–Elbistan Thermal Power Plant, and its properties are given in Tables 2–3. Pumice aggregate was obtained from the Kayseri region and has a particle size of 0–11.2 mm (see Table 3). Aggregates obtained from the Elbistan region were utilized in the mixture.

The dimensions of the produced blocks were determined as 13.5 × 39.5 × 18.5 cm. Preliminary experiments indicated that substituting 30% of the cement with fly ash yielded the optimal ratio in terms of compressive strength.

Table 1. Chemical properties of CEM I 42.5 R cement

<table>
<tr><td>Type of additive</td><td>SiO₂</td><td>Al₂O₃</td><td>Fe₂O₃</td><td>CaO</td><td>MgO</td><td>SO₃</td></tr>
<tr><td rowspan="4">CEM I 42,5 R</td><td>19,4</td><td>5,4</td><td>3,8</td><td>64,3</td><td>2,3</td><td>2,5</td></tr>
<tr><td colspan="6">Physical characteristics</td></tr>
<tr><td colspan="3" rowspan="2">Specific gravity (gr/cm³)</td><td rowspan="2">Specific surface area (cm²/gr)</td><td colspan="2">Remaining on the sieve (%)</td></tr>
<tr><td>90μ</td><td>45μ</td></tr>
<tr><td colspan="3">3,15</td><td>3480</td><td>0,1</td><td>8,2</td></tr>
</table>

Table 2. Chemical composition of Afşin-Elbistan fly ash

Component	*SiO_2*	*Al_2O_3*	*Fe_2O_3*	*CaO*	*MgO*	*Na_2O*	*K_2O*	*SO_3*	*Cr_2O_3*	*Mn_2O_3*	*Loss*
Percentage	*29.21*	*11.50*	*6.53*	*37.69*	*1.79*	*0.44*	*0.62*	*4.51*	*0.048*	*0.050*	*2.27*

Table 3. Physical properties of Afşin-Elbistan fly ash

Test	*Value*
Fineness (residue on 45 μm sieve, %)	*66.43*
Specific gravity (g/cm³)	*2.41*
Specific surface area (cm²/g)	*1868*
Initial setting time (min)	*225*
Final setting time (min)	*313*
Volume expansion (mm)	*1*
Water requirement (%)	*27.81*
Water amount (g)	*139.1*

Table 4. General chemical properties of pumice

Component	*Content (%)*
SiO_2	*60–75*
Al_2O_3	*13–15*
Fe_2O_3	*1–3*
MgO	*1.2*
CaO	*1–2*
$Na_2O + K_2O$	*7–8*

Kaynak: (Özkan ve Tuncer, 2001)

2. Production Process

Fly ash was transported from the thermal power plant to the production facility, a distance of 30 km. At the production facility, a vertical silo with a capacity of 75 tons was utilised for the storage of fly ash and to ensure continuous production. The silo is equipped with a filter and weighing system. In the production phase, cement and fly ash were initially mixed in a dry state, followed by the incorporation of pumice aggregate and water (Figure 1). The mixture was subsequently placed in moulds and subjected to standard curing conditions. Preliminary tests indicated that samples containing a 30% fly ash additive exhibited adequate compressive strength.

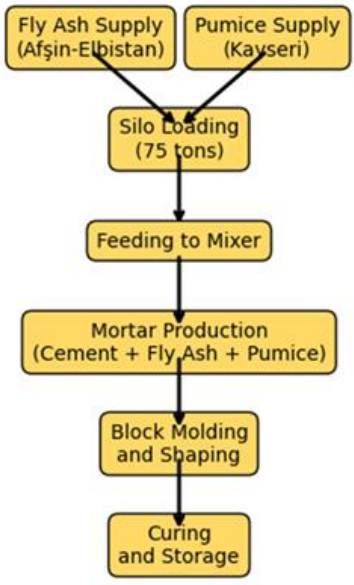


Fig. 1. Production process illustration

COST ANALYSIS

1.Material Quantities for a 200 m² Wall

In order to perform a cost analysis, the material quantities required for a 200 m² wall sample were determined. It was hypothesized that approximately 4,800 pumice blocks were utilized for this area. The material quantities calculated based on the mixing ratios are given in Figure 2.

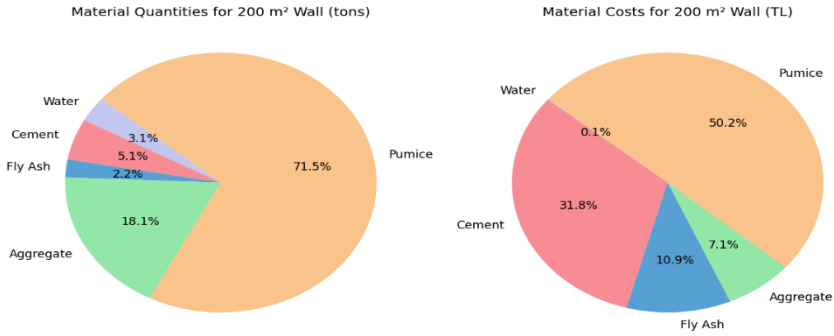


Fig. 2: 200 m² duvar için malzeme miktarları ve maliyetleri

2. Distribution of costs

The total material cost calculated for a 200m² wall was determined to be 42,590.24 TL. This value corresponds to approximately 8.87 TL per block and approximately 212.95 TL per m². As illustrated in Figure 1, the figure presents the quantity and cost distribution of materials utilized in the construction of a 200-square-metre wall. Upon examination of the graph, it is evident that pumice possesses the largest share in terms of total material quantity; however, cement and pumice emerge as the pivotal factors in determining the distribution of costs. The cost of fly ash is a minor component of the overall expenditure, and its utilization has been demonstrated to reduce cement consumption, thereby lowering the total cost of the project.

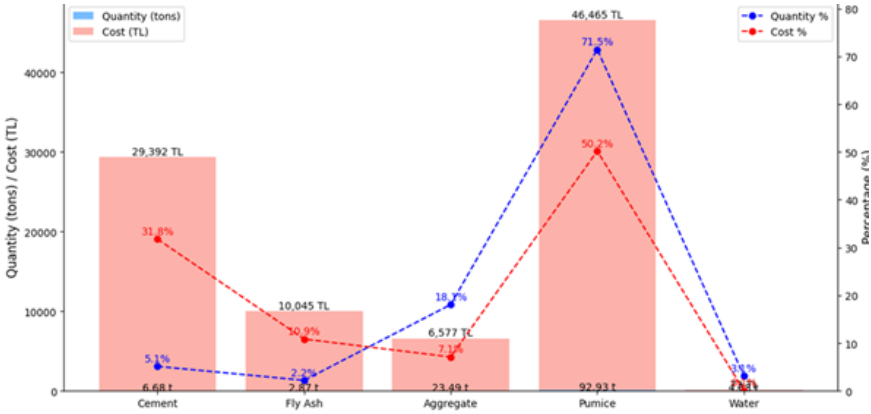


Fig. 3 Material quantities and cost for 200 m^2 wall

DISCUSSION

The findings indicate that substituting 30% of the cement with fly ash does not adversely affect the mechanical performance of pumice blocks, while concomitantly reducing cement consumption. The low unit cost of fly ash has a limited impact on the total production cost, providing an economic advantage. Moreover, the utilization of Afsin-Elbistan fly ash confers environmental benefits by means of the recycling of industrial waste.

RESULTS

The present study demonstrated that Afsin-Elbistan fly ash-modified pumice blocks are technically and economically feasible. Preliminary experiments determined a 30% fly ash substitution rate to be the optimum value, and it was found that blocks produced with this rate have sufficient compressive strength.

A cost analysis was conducted on a 200 m² wall sample, the results of which indicate that the utilization of fly ash reduces cement consumption without increasing the total cost. Furthermore, it is demonstrated that the use of fly ash contributes to the production of sustainable building materials. In conclusion, fly ash-modified pumice blocks can be regarded as an environmentally friendly, economical, and industrially viable alternative building material.

REFERENCES

Usta, P., & Zengin, B. (2021). The energy impact of building materials in residential buildings in Turkey. *Materials*, *14*(11), 2793.

Yüksek, İ. (2015). The evaluation of building materials in terms of energy efficiency. *Periodica Polytechnica Civil Engineering*, *59*(1), 45-58.

Juarez, R. I. C., & Finnegan, S. (2021). The environmental impact of cement production in Europe: A holistic review of existing EPDs. *Cleaner Environmental Systems*, *3*, 100053.

Nsobundu, A. E. J., & Tiza, M. T. A Comprehensive Evaluation of Waste-Derived Materials for Sustainable Construction Practices. *International Journal of Environmental Pollution and Environmental Modelling*, *8*(1), 26-42.

Moolchandani, K. (2025). Advancements in pumice-based concrete: A comprehensive review. *Next Materials*, *8*, 100646.

Özkan, Ş.G., Tuncer , G., 2001. An Overview of Pumice Mining, 4th Industrial Raw Materials Symposium, Izmir, 200 – 207.

Geophysical Investigations of the Seamounts in the eastern Mediterranean

Sercan KAYIN [1]
M. Nuri DOLMAZ [2]

1- Assist. Prof. Dr. : Gümüşhane University, Gümüşhane Vocational School, Department of Mining and Mine Extraction, Gümüşhane, sercankayin@gumushane.edu.tr ORCID No: 0000-0003-2279-0579
2- Prof. Dr.; Süleyman Demirel University, Engineering and Natural Sciences Faculty, Department of Geophysical Engineering, Isparta. nuridolmaz@sdu.edu.tr ORCID No: 0000-0002-9185-9835

ABSTRACT

This research presents a comprehensive geophysical investigation of the prominent seamounts in the Eastern Mediterranean, a geologically complex basin shaped by long-lived convergence between the African and Eurasian plates. Seamounts in this region serve as unique windows into mantle processes, crustal deformation, and lithospheric restructuring. Specifically focusing on the Anaximander Mountains complex (comprising the Anaximander, Anaximenes, and Anaxagoras seamounts), the Güzelyurt Seamount, and the Eratosthenes Seamount, this study utilizes an integrated approach combining magnetic anomaly mapping, bathymetric estimations, and seismotectonic analysis. Magnetic investigations were conducted based on the EMAG2-v3 global magnetic anomaly dataset with a spatial resolution of 2 arc-minutes, processed using Oasis Montaj software. Advanced techniques including Reduction to the Pole (RTP), Analytic Signal (AS), and First Vertical Derivative (FVD) were applied to reveal buried intrusive bodies, volcanic edifices, and structural boundaries.

The Eratosthenes Seamount is identified as a massive continental fragment undergoing incipient collision with Cyprus, characterized by an elliptical, elongated geometry with an N40E strike and a crustal thickness of approximately 22.49 ± 0.08 km derived from potential field data. To the north, the Güzelyurt Seamount acts as a well-expressed extensional dome within the Antalya Basin, accommodating regional deformation through active seismicity and strike-slip faulting. The Anaximander Mountains represent a critical outer-arc complex formed under a transpressional regime, featuring active fluid venting and methane emissions at the Anaxagoras peak. Seismicity analysis from 1976 to 2025, utilizing KOERI data, indicates a high density of hypocenters in the northern sector and around Cyprus, with predominantly shallow events (0-50 km) and magnitudes reaching up to M=6. These findings underscore the intense tectonic activity and ongoing subduction dynamics that define the Eastern Mediterranean's geodynamic evolution.

Keywords – Seamount, Mediterranean Sea, Magnetic, Depth, Earthquake.

INTRODUCTION

Seamounts are significant submarine topographic and volcanic structures that provide critical constraints on the tectonic and geodynamic evolution of oceanic basins. Eastern Mediterranean Sea hosts several prominent seamounts that strongly influence the region's deep-marine morphology. The most significant structures include the Eratosthenes Seamount, a large isolated carbonate platform south of Cyprus, and the Finike Seamount Group—comprising the Anaximenes, Anaxagoras, and Finike seamounts—located offshore southwestern Turkey. To the west of this group lies the Anaximander Seamount, a broad tectonically active

submarine high associated with mud volcanism, faulting, and uplift along the eastern segment of the Hellenic–Anatolian plate boundary. In the northern Levant Basin, the Latakia Ridge represents an elongated compressional structure developed under Africa–Anatolia convergence. Numerous smaller knolls and ridges across the Herodotus, Rhodes, and Levant basins further reflect the complex interplay of subsidence, accretion, and intraplate deformation that shapes the seafloor of the Eastern Mediterranean.

In the Mediterranean Sea—a geologically complex, semi-enclosed basin shaped by long-lived convergence between the African and Eurasian plates—seamounts offer unique windows into mantle processes, crustal deformation, and lithospheric restructuring. The region is characterized by a mosaic of microplates, slab rollback dynamics, back-arc extension, and lithospheric tearing (Faccenna et al., 2014; Jolivet and Faccenna, 2000). Consequently, Mediterranean seamounts reflect a diversity of formation mechanisms, including compressional uplift, transtensional faulting, and intraplate volcanism (LeBreton et al., 2020; Zitellini et al., 2009).

From a geophysical perspective, seamounts play an integral role in imaging lithospheric architecture and understanding crustal evolution. High-resolution multibeam bathymetry, magnetic anomaly mapping, and potential-field modeling have revealed intricate volcanic edifices, buried intrusive bodies, and crustal heterogeneities across the basin (Adam et al., 2015; Leandro et al., 2019). Magnetic anomalies commonly indicate zones of magmatic addition or remnant volcanic arcs, while gravity data provide insights into crustal thickness variations associated with rifting or subduction processes. Seismic reflection and refraction studies further illuminate fault geometries, sedimentary sequences, and deep crustal structure, helping to decipher the tectonic history of seamount provinces such as the Anaximander Mountains, Eratosthenes Seamount, and Vavilov Basin (Hall et al., 2014; Makris et al., 2018).

In the Eastern Mediterranean, where subduction rollback and slab segmentation have been dominant since the Miocene, seamounts serve as indicators of dynamic mantle flow and lithospheric deformation (Grevemeyer et al., 2020). Their spatial distribution often corresponds to diffuse deformation belts, transform boundaries, or ancient suture zones, highlighting the interplay between plate-boundary forces and intraplate magmatic processes. Although recent geophysical surveys have improved regional imaging, many seamount complexes remain insufficiently mapped due to limited deep-sea coverage. Continued acquisition of marine geophysical data is essential to refine models of Mediterranean tectonics, improve our understanding of lithosphere–asthenosphere interactions, and assess how seamount evolution reflects the broader geodynamic development of the basin.

Figure 1 illustrates the spatial distribution of the Anaximander–Finike, Anaxagoras, Anaximenes, Eratosthenes, and Güzelyurt seamounts in the

Eastern Mediterranean, whereas Figure 2 presents the general tectonic framework of the eastern Mediterranean region.

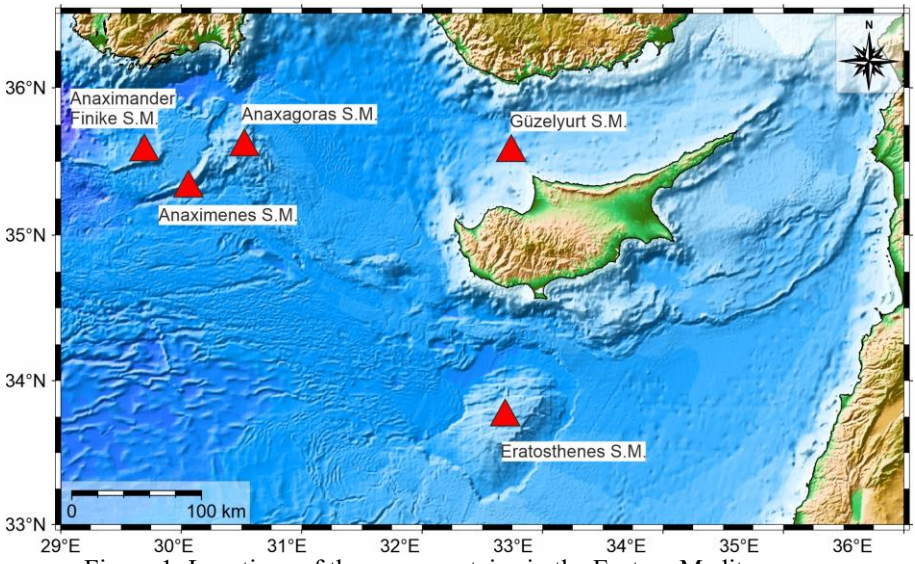


Figure 1 :Locations of the sea mountains in the Eastern Mediterranean

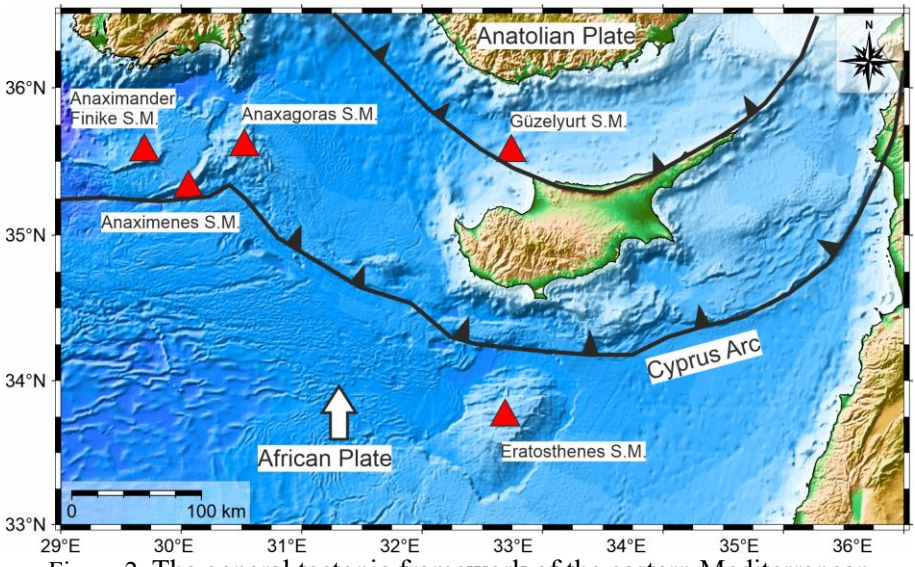


Figure 2. The general tectonic framework of the eastern Mediterranean region.

SEAMOUNTS REVIEW

The tectonic architecture of the Eastern Mediterranean is defined by the complex interaction between the African and Eurasian plates. This convergence is accommodated by a series of prominent bathymetric highs and submarine structures situated along the plate boundaries. Extensive geophysical investigations described in the literature have focused on elucidating the structural and evolutionary complexity of these features, specifically the Anaximander Mountains, the Güzelyurt Seamount, and the Eratosthenes Seamount.

The Anaximander Mountains: The Anaximander Mountains constitute an outer-arc complex comprising the Anaximander, Anaximenes, and Anaxagoras seamounts, positioned at the critical junction of the Hellenic and Cyprus arcs. ten Veen et al. (2004) provided a detailed analysis of this region utilizing a comprehensive suite of geophysical data, including multibeam bathymetry, backscatter imagery, seismic reflection profiles, and gravity/magnetic data. Their work characterized the region's neotectonic evolution, concluding that the mountains formed as a result of a protracted contractional and transpressional regime driven by the convergence of the African and Anatolian plates. The individual seamounts likely developed through Plio-Quaternary uplift and differential rotation of a major thrust fan structure (ten Veen et al., 2004).

Within this complex, *the Anaxagoras Seamount*, situated as the easternmost peak, has been a key target for multidisciplinary research focused on active margin processes. Investigations utilizing high-resolution multibeam bathymetry, deep-tow sidescan sonar, and observations from the submersible Nautile confirmed that the seamount lies within a region of complex deformation (Woodside et al., 1998; Zitter et al., 2003). Crucially, these studies identified Anaxagoras as a major site for active fluid venting. The presence of numerous cold seeps, methane emissions, and associated chemosynthetic communities along its scarps and flanks-closely linked to the regional fault architecture-highlights the active fluid escape structures characterizing this segment of the arc (Zitter et al., 2006).

Güzelyurt Seamount: East of the Anaximander complex, situated within the Antalya Basin approximately 80 km off the coast of Cyprus, lies the *Güzelyurt Seamount*. It is a prominent, well-expressed extensional dome structure formed on the southern margin of the active Cyprus Arc. Geophysical investigations, notably multi-channel seismic reflection profiles, reveal that the seamount is characterized by significant faulting and uplift that dissects the overlying sedimentary succession.

Tectonically, the formation of the *Güzelyurt Seamount* is intrinsically linked to the complex African-Eurasian plate boundary

interaction. It is believed to be a product of transtensional forces arising from the oblique convergence along the adjacent Cyprus Arc, with associated strike-slip and thrust faulting (Makris et al., 1983). The seamount exhibits active seismicity, reinforcing its role as a key structural element accommodating deformation in the transition zone between the Hellenic and Cyprus Arc systems (Woodside, 1976). Furthermore, gravity data indicates a positive free-air anomaly over the structure, suggesting a local crustal density increase or the presence of uplifted high-density material beneath it. Magnetic surveys often show associated anomalies, features commonly interpreted as evidence for the involvement of deeper basement rocks-likely continental crustal fragments-being brought closer to the seafloor during the uplift process.

Eratosthenes Seamount: To the south, the *Eratosthenes Seamount* represents one of the most prominent features in the Eastern Mediterranean, identified as a continental fragment undergoing incipient collision with Cyprus (Galindo-Zaldívar et al., 2001). Specific geophysical investigations have provided crucial insights into its deep structure. Erbek and Dolmaz (2014) utilized free-air gravity and magnetic data to generate new potential field maps of the seamount and its vicinity. Through detailed spectral analysis and various filtering techniques, including the Horizontal Gradient of the Analytic Signal (HGAS), they characterized the seamount as an elliptical, elongated structure with an approximate strike of N40°E.

Their analysis of the potential field data further estimated the crustal thickness in the region, deriving a value of approximately 22.49±0.08 km from the power spectrum of the gravity anomaly (Erbek and Dolmaz, 2014). This comprehensive study highlights the complex subsurface characteristics of the *Eratosthenes Seamount*, linking its surface morphology to deep crustal structures within the Eastern Mediterranean context. Figure 3 illustrates example of batimetrical depth estimation from the Eratosthenes Seamount. Table 1 presents the numerical locations (latitude and longitude), peak depths, and base depths of all seamounts located in the Eastern Mediterranean.

Table 1: Seamounts of the East Mediterranean (from Öztürk et al., 2015)

Seamout name	**Latitude (°)**	**Longitude (°)**	**Peak depth (m)**	**Base depth (m)**
Anaxagoras S. M.	35.67566	30.51171	920-930	1510-1520
Anaximander-Finike S. M.	35.50840	29.78160	1110-1120	2000-2010
Anaximenes S. M.	35.43070	30.16420	690-700	1500-1510
Eratosthenes S. M.	33.74444	32.73362	780-790	1920-1930
Gűzelyurt S. M.	35.67092	32.76268	1410-1420	ND

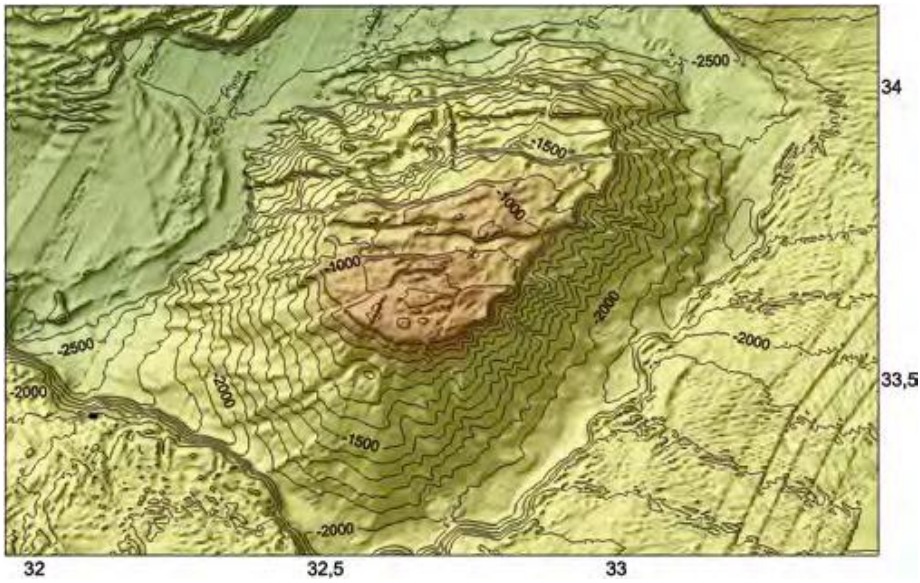


Figure 3: Batimetrical depth estimation from the Eratosthenes Seamount. Note the deeper bathymetry of 1920-1930 m, which identifies the seamount base and shallower bathymetry of 780-790 m, which identifies the seamount peak (from Öztürk et al., 2015).

SEISMICITY

The provided earthquake map clearly illustrates that the Eastern Mediterranean is a highly seismically active region, characterized by a large number of recorded earthquakes (Fig. 4). The distribution of these events is not uniform; the highest density of epicenters is predominantly observed in the northern part of the map, spanning the Aegean Sea and extending into the Turkish mainland. Furthermore, significant activity clusters around Cyprus and along the complex plate boundaries extending from the island. While the majority of the recorded events are of moderate magnitude, typically M=4, there are numerous occurrences of M=5 and several instances of M=6, underscoring the potential for major, damaging earthquakes within the region. Regarding depth, most events are relatively shallow (0-50 km), which is common for crustal seismicity along active fault zones, thereby increasing the risk of severe ground shaking. However, deeper earthquakes are also evident, particularly in the northern areas, typically associated with the subduction of the African plate beneath the Eurasian and Anatolian

plates. This intense pattern of seismicity strongly reflects the complex tectonic setting of the Eastern Mediterranean, driven by the ongoing convergence and interaction of these major lithospheric plates.

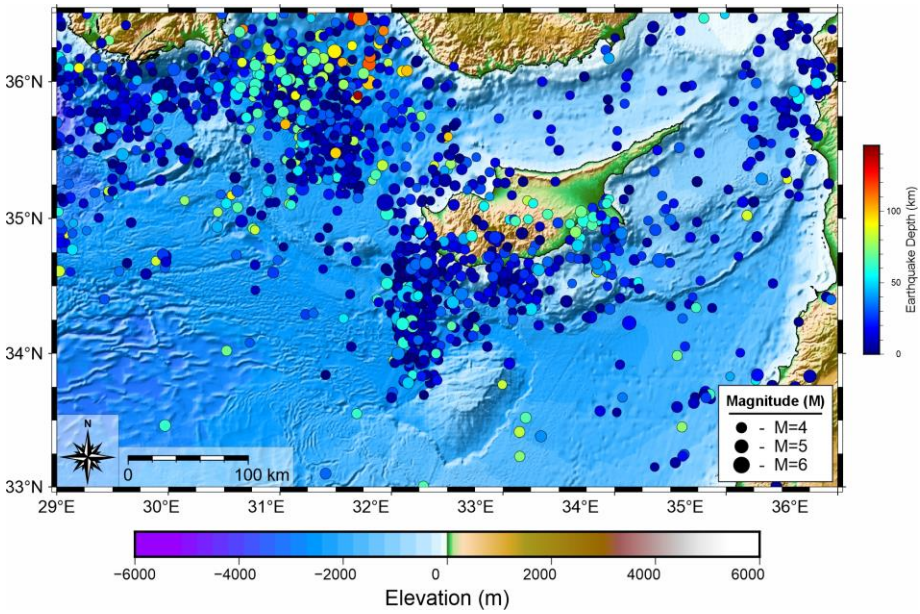


Figure 4. Combined display of the bathymetry of the Eastern Mediterranean and the spatial distribution of earthquake hypocenters and their focal depths in the region. Earthquake hypocenter data for events with M > 3.5 that occurred between 1976 and December 2025 were obtained from KOERI (Kandilli Observatory and Earthquake Research Institute). The map was produced using Generic Mapping Tools (GMT), version 6.0 (Wessel et al., 2019).

MAGNETIC PERSPECTIVE

Magnetic data analysis plays a fundamental role in imaging the lithospheric architecture and understanding the crustal evolution of the Eastern Mediterranean. The magnetic investigations presented here aim to reveal volcanic edifices, buried intrusive bodies, and crustal heterogeneities.

This research is based on the EMAG2-v3 global magnetic anomaly dataset provided by the National Centers for Environmental Information in collaboration with the National Oceanic and Atmospheric Administration (NOAA). The dataset integrates more than five decades of satellite, marine,

and airborne magnetic observations and is gridded at a spatial resolution of 2 arc-minutes at sea level (Meyer et al., 2017). Magnetic data processing and interpretation were carried out using the Oasis Montaj software package.

The Total Magnetic Intensity map (Fig. 5) represents the fundamental magnetic dataset for the study area. This map visualizes the summation of the Earth's regional field and local magnetic anomalies caused by subsurface geological structures. In the context of the Eastern Mediterranean seamounts, these anomalies typically indicate zones of magmatic addition or remnant volcanic arcs.

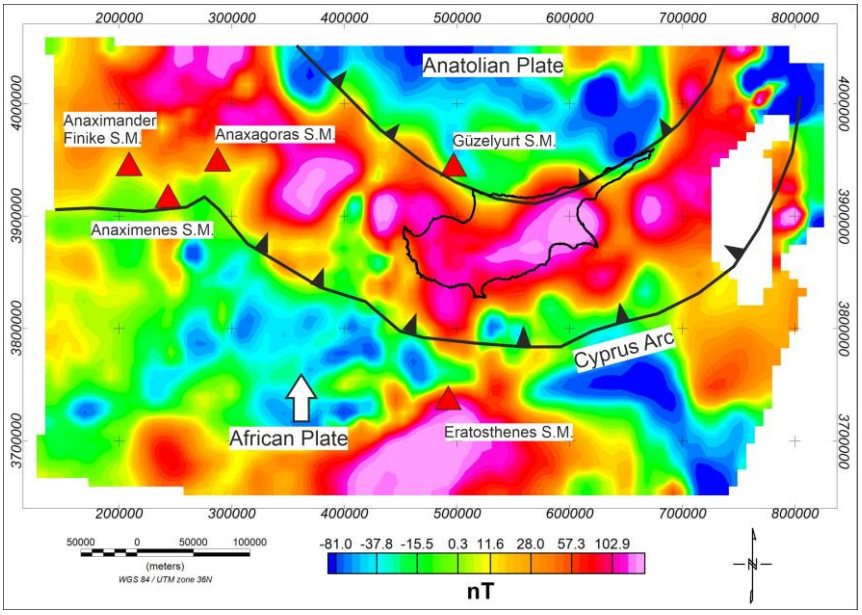


Figure 5: The Total Magnetic Intensity map of the study area

Magnetic anomalies are controlled by several factors, including the depth and magnetic susceptibility of the source bodies, as well as the inclination and declination of the Earth's magnetic field. To correct for the dipolar nature of the geomagnetic field at these latitudes and to facilitate geological interpretation - particularly in mid-latitude regions such as the Mediterranean - the Reduction to the Pole (RTP) transformation was applied to the total magnetic intensity (TMI) data (Fig. 6). This procedure mathematically transforms magnetic anomalies to their equivalent response under a vertically oriented magnetic field, effectively repositioning the anomalies above their causative sources (Baranov, 1957; MacLeod et al., 1993; Blakely, 1995; Ansari and Alamdar, 2009). The RTP correction was applied prior to spectral analysis. Figure 6 presents the RTP map of the TMI data.

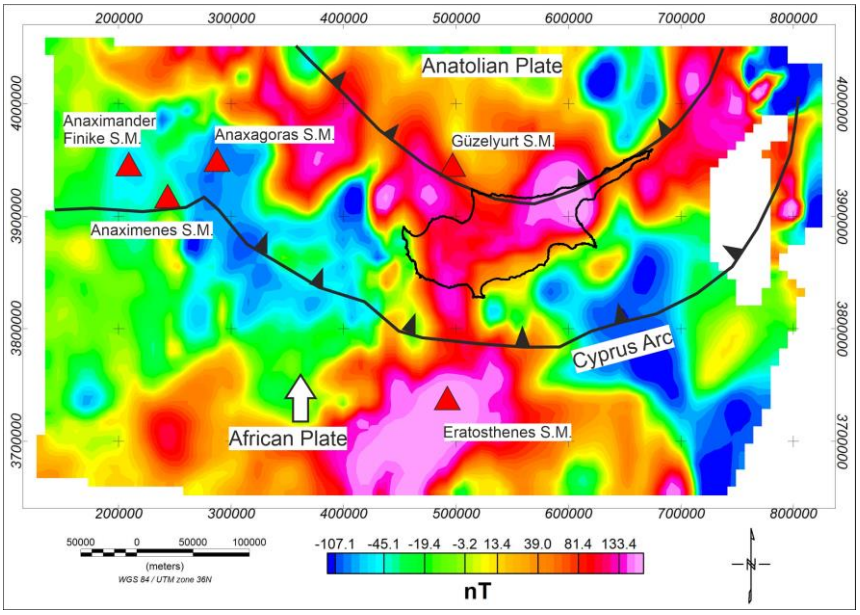


Figure 6: The Reduction to the pole map of the study area

To further delineate magnetic source bodies, the Analytic Signal (AS) method was applied, as it is independent of the direction of magnetization and is particularly effective in highlighting the edges of causative sources (Roest et al., 1992). Figure 7 illustrates the Analytic Signal map derived from the magnetic data, clearly emphasizing source boundaries regardless of magnetization direction.

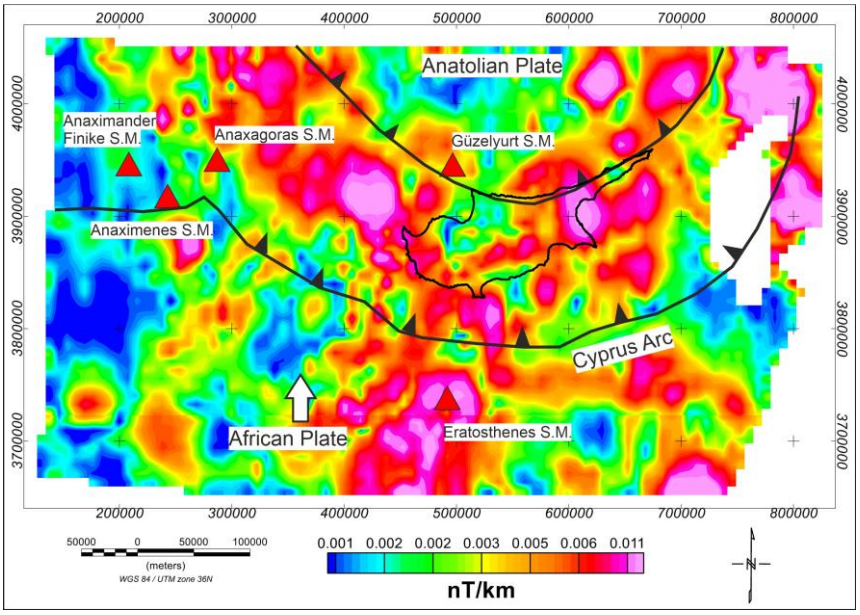


Figure 7: The Analytic Signal map of the study area

The First Vertical Derivative (FVD) was applied to enhance shallow magnetic features and sharpen anomaly boundaries, as it effectively acts as a high-pass filter by amplifying anomalies associated with near-surface sources while suppressing longer-wavelength regional components (Reeves, 2005). The resulting FVD map is shown in Figure 8.

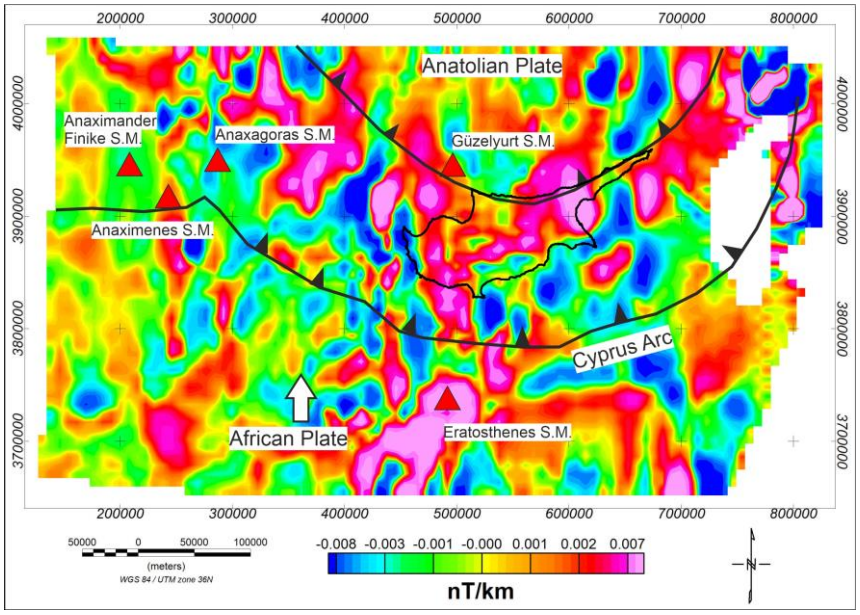


Figure 8: The first vertical derivative map of the study area

The depth to the magnetic basement was estimated using the Analytic Signal (AS) method, which allows the determination of magnetic source depths by calculating the ratio of the AS amplitude to its vertical derivative (ASI) (Reid et al., 1990). This approach produces a three-dimensional subsurface model that can be directly compared with seismicity data. Figure 9 illustrates the estimated depth to the magnetic basement derived from the spectral analysis of the potential field data.

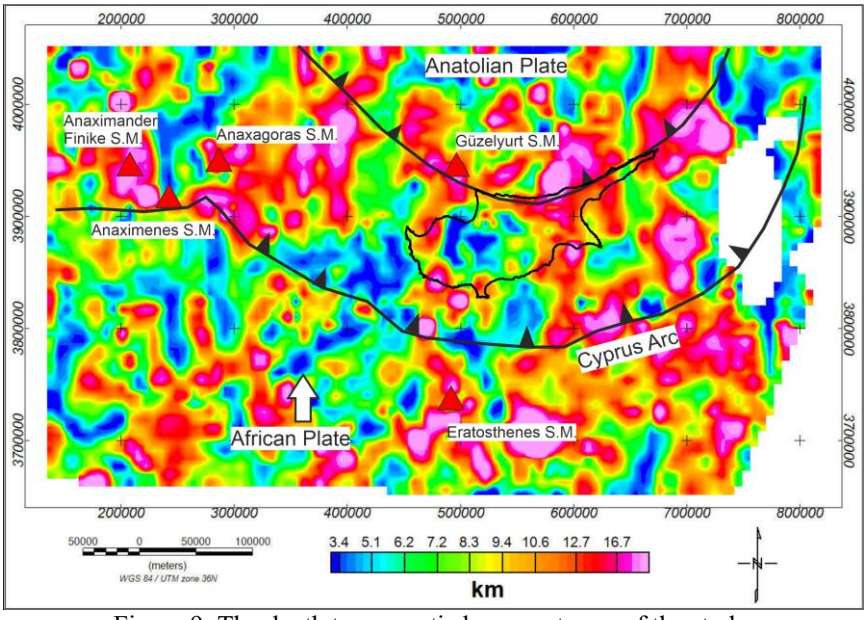


Figure 9: The depth to magnetic basement map of the study area

RESULTS AND DISCUSSION

The integrated geophysical analysis reveals that the Eastern Mediterranean seamounts are pivotal in understanding the lithospheric architecture and tectonic evolution of this complex convergent margin. The Total Magnetic Intensity (TMI) mapping highlights distinct signatures of magmatic addition and remnant volcanic arcs. By applying the Reduction to the Pole (RTP) transformation, the dipolar shift was mathematically corrected, allowing for the precise spatial correlation of magnetic anomalies with the bathymetric positions of the Anaximander, Güzelyurt, and Eratosthenes seamounts. Furthermore, the Analytic Signal (AS) method delineated the structural boundaries of these features, confirming that the

Eratosthenes Seamount possesses an elliptical geometry with an N40E strike. Shallow structural lineaments and fault geometries were further sharpened using the First Vertical Derivative (FVD), which acts as a high-pass filter to reveal the intense tectonic activity and uplifted blocks within these complexes.

The region's geodynamic complexity is further reflected in the spatial distribution of earthquake hypocenters recorded between 1976 and 2025. Seismic data indicates a high density of epicenters concentrated in the northern sector and around Cyprus, predominantly characterized by shallow crustal events occurring at depths of 0-50 km. These events, frequently ranging from M=4 to M=6, underscore the significant seismic hazard associated with the ongoing convergence and subduction of the African plate. The depth to the magnetic basement, estimated through spectral analysis, provides a three-dimensional perspective that aligns with this seismicity, reflecting the complex interplay of subsidence and accretion shaping the seafloor.

Specific structures such as the Anaximander Mountains, comprising the Anaximander, Anaximenes, and Anaxagoras seamounts, serve as a critical outer-arc complex where active margin processes like fluid venting and methane emissions are prevalent. To the east, the Güzelyurt Seamount acts as an extensional dome in the Antalya Basin that accommodates regional deformation through active seismicity and strike-slip faulting. Meanwhile, the Eratosthenes Seamount represents a massive continental fragment undergoing collision with Cyprus, with a calculated crustal thickness of approximately 22.49±0.08 km. Bathymetric data further quantifies these features, noting peak depths as shallow as 690-700 m at Anaximenes and as deep as 1410-1420 m at Güzelyurt, with base depths reaching 2000-2010 m in the Anaximander-Finike complex.

REFERENCES

Adam, C., Vidal, V., and Bonneville, A. 2015. Seafloor topography and lithospheric structure of oceanic regions. Earth and Planetary Science Letters, 429, 1–13

Ansari, A.H. and Alamdar, K. 2009. Reduction to the pole of magnetic anomalies using the analytic signal. World Appl Sci J, 7, 405-409.

Baranov, V. 1957. A new method for interpretation of aeromagnetic maps, pseudo-gravimetric anomalies. Geophys., 22, 359-363.

Blakely, R.J. 1995. Potential theory in gravity and magnetic applications. Cambridge University Press, pp 441.

Erbek, E., and Dolmaz, M.N. 2014. Geophysical Researches (Gravity and Magnetic) of the Eratosthenes Seamount in the Eastern Mediterranean Sea. Acta Geophysica, 62 (4), 762-784.

Faccenna, C., Becker, T. W., Conrad, C. P., and Husson, L. 2014. Mountain building and mantle dynamics. Earth and Planetary Science Letters, 403, 93–106.

Galindo-Zaldívar, J., Nieto, L.M., Robertson, A.H.F. 2001. Recent tectonics of Eratosthenes Seamount: an example of seamount deformation during incipient continental collision. Geo-Marine Letters, 20, 233-242.

Grevemeyer, I., Lange, D., and Moeller, S. 2020. Structure and evolution of the Eastern Mediterranean from marine seismology. Geophysical Journal International, 223(2), 1050–1071.

Hall, J. K., Tibor, G., and ben-Avraham, Z. 2014. The Eastern Mediterranean: A review of seismic and bathymetric data. Marine Geology, 354, 1–20.

Jolivet, L., and Faccenna, C. 2000. Mediterranean extension and the Africa–Eurasia collision. Tectonics, 19(6), 1095–1106.

Leandro, D., Catalão, J., Miranda, J. M., and Luís, J. F. 2019. Magnetic signatures of Mediterranean seamounts. Geophysical Research Letters, 46, 3451–3459.

LeBreton, N., Handy, M. R., and Molli, G. 2020. The geodynamic evolution of the Mediterranean region. Gondwana Research, 84, 1–22.

MacLeod, I.N., Jones, K. and Dai, T.F. 1993. 3-D analytic signal in the interpretation of total magnetic feld data at low magnetic latitudes. Explor. Geophys., 24, 679–688.

Makris, J., Ben-Avraham, Z., and Nicolich, R. 2018. Deep crustal structure of the Mediterranean Ridge and surrounding areas. Tectonophysics, 749, 90–105.

Makris, J., Ben-Avraham, Z., Behle, A., Ginzburg, A., Giese, P., Steinmetz, L., Whitmarsh, R.B., Eleftheriou, S. 1983. Seismic refraction profiles between Cyprus and Israel and their interpretation. Geophys J R Astron Soc, 75, 575-591.

Meyer, B., Chulliat, A. and Saltus, R. 2017. Derivation and error analysis of the Earth magnetic anomaly grid at 2 arc min resolution version 3 (EMAG2v3). Geochem Geophys Geosyst, 18, 4522-4537.

Öztürk, B., Rovere, M., and Würtz, M. 2015. Seamounts and Seamount-like Structures of the Eastern Mediterranean. Book chapter, edited by M. Würtz, M. Rovere, IUCN Atlas of the Mediterranean Seamounts and Seamount-like Structures, 227-275.

Reeves, C. 2005. Aeromagnetic Surveys; Principles, Practice and Interpretation. GEOSOFT, 155pp.

Reid, A.B., Allsop, J.M., Granser, H., Millett, A.J. and Somerton, I.W. 1990. Magnetic interpretation in three dimensions using Euler deconvolution. Geophysics, 55, 80-91.

Roest, W. R., Verhoef, J. and Pilkington, M. 1992. Magnetic interpretation using the 3-D analytic signal. Geophysics, 57, 116–125.

ten Veen, J. H., Woodside, J. M., Zitter, T., Dumont, J. F., Mascle, J., Volkonskaia, A. 2004. Neotectonic evolution of the Anaximander Mountains at the junction of the Hellenic and Cyprus arcs. Tectonophysics, 391, 35-65.

Wessel, P., Luis, J., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., and Tian, D. 2019. The Generic Mapping Tools Version 6, Geochemistry, Geophysics, Geosystems, 20 (11), 5556-5564.

Woodside, J. 1976. Regional vertical tectonics in the Eastern Mediterranean. Geophys J Int, 47, 493-514.

Woodside, J. M., Ivanov, M. K., Limonov, A. F. 1998. Shallow gas and gas hydrates in the Anaximander Mountains region, eastern Mediterranean Sea. In: Henriet JP, Mienert J (eds.), Gas hydrates: relevance to world margin stability and climate change, Geological Society, London, Special Publications 137: 177-193.

Zitellini, N., Gràcia, E., and Ranero, C. R. 2009. The Mediterranean Ridge: structure, evolution and seismicity. Marine Geology, 257, 1–14.

Zitter, T. A. C., Woodside, J. M., Mascle, J. 2003. The Anaximander Mountains: a clue to the tectonics of Southwest Anatolia. Geological Journal 38, 375-394.

Zitter, T., Huguen, C., ten Veen, J., Woodside, J. M., 2006. Tectonic control on mud volcanoes and fluid seeps in the Anaximander Mountains, eastern Mediterranean Sea. In: Dilek Y, Pavlides S (eds.), Postcollisional Tectonics and Magmatism in the Mediterranean Region and Asia. Marine Geology, 235(1-4), 1-25.

Transient Heat Transfer Analysis in Fuel Cell Plates Using a Nodal Network Approach

Hüseyin KAHRAMAN[*1]

[1]Mechanical Engineering Department/Faculty of Technology, Sakarya University of Applied Sciences, Türkiye
[*]huseyink@subu.edu.tr

ABSTRACT

Thermal behavior plays a critical role in the performance, durability, and efficiency of fuel cell systems. In practical fuel cell stacks, temperature distribution is often non-uniform due to heat generation from electrochemical reactions, convective heat loss, and material thermal properties. In this study, transient heat transfer in a representative fuel cell plate is numerically investigated using a one-dimensional nodal network approach. The governing energy balance equations are formulated based on conductive heat transfer within the solid domain and convective heat loss at the boundary, and the resulting system of ordinary differential equations is solved using a time-dependent numerical scheme.

The computational model discretizes the plate thickness into uniformly distributed nodes, enabling the prediction of temperature evolution as a function of both space and time. Simulations are performed for different operating durations, with particular focus on short-term (10 s) and long-term (150 s) thermal responses. The results reveal that significant temperature gradients develop during the early transient period, primarily influenced by convective cooling at the exposed surface. As the simulation time increases, thermal diffusion becomes dominant, leading to smoother temperature profiles and a gradual transition toward a quasi-steady-state condition.

The findings highlight the importance of transient thermal analysis in fuel cell components, as temperature gradients may persist over extended operating periods and influence reaction kinetics, water management, and material degradation. The proposed nodal network framework provides a computationally efficient and physically consistent tool for analyzing heat transfer in fuel cell layers and can be extended to multilayer structures and multidimensional models in future studies.

Keywords – Fuel Cell, Heat Transfer, Transient Conduction, Nodal Network Method, Thermal Management.

INTRODUCTION

In a fuel cell, temperature distribution is not always uniform, even when there is a constant mass flow rate within the channels. Uneven temperatures within the fuel cell stack are influenced by several factors, including water phase change, coolant temperature variations, air convection, water trapping, and heat generated by the catalyst layer [1].

To accurately predict temperature-dependent parameters, such as reaction rates and species transport, it is essential to understand the temperature distribution across the stack. This comprehensive understanding is crucial for optimizing fuel cell performance, ensuring longevity, and improving overall efficiency.

This paper will explore the fundamental principles of heat transfer in fuel cells, providing an overview of the mechanisms that affect temperature variation within the system. By analyzing the heat transfer processes, we aim to provide insights into how temperature impacts fuel cell operations and how efficient temperature management can enhance performance [2].

To determine the precise heat distribution across the stack, mathematical models and heat transfer equations are employed. These models account for the various heat sources and sinks within the system, allowing for accurate predictions of temperature-dependent behavior. In the following sections, we will delve into the heat transfer basics in fuel cells, including the conduction, convection, and radiation mechanisms at play.

By understanding and controlling the thermal behavior of a fuel cell, we can enhance the overall efficiency, ensure proper operation, and improve the longevity of the fuel cell stack [3].

The first step in determining the heat distribution within a fuel cell stack is to apply energy balances to the system. The total energy balance in a fuel cell is determined by the power produced, the electrochemical reactions occurring, and the heat losses within the system. Heat transfer occurs through convection between the solid surfaces and the gas streams, as well as conduction within the solid and/or porous materials. Key factors to consider when modeling heat transfer in a fuel cell include the reactants, products, and the electricity generated, as illustrated in Figure 6-1.

According to the general energy balance, the enthalpy of the reactants entering the fuel cell must equal the enthalpy of the products exiting the fuel cell, along with the heat generated by the power output and the heat lost to the surrounding environment. These fundamental heat transfer calculations provide valuable insights into predicting the temperature distribution and heat flow throughout the fuel cell stack and its components.

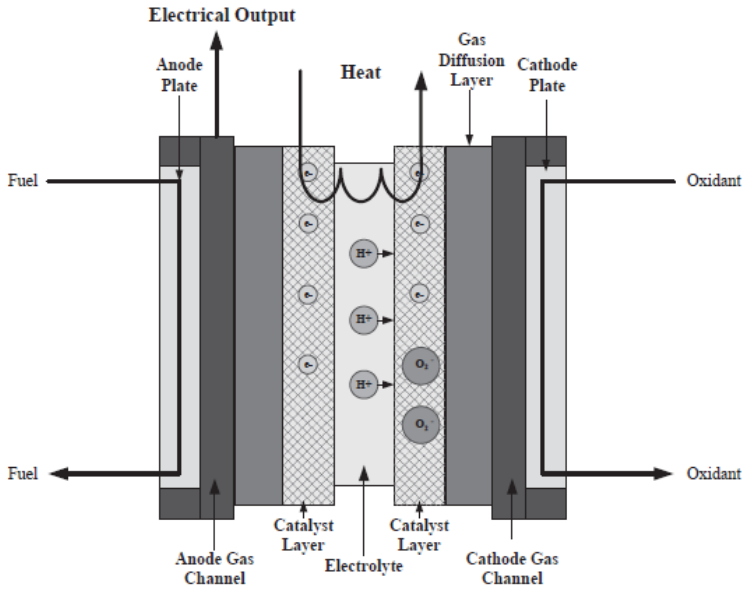


Fig. 1 Illustration of the fuel Cell stack for heat flow analysis [4]

Conduction is defined as the transfer of energy from more energetic particles to less energetic particles due to their interaction. A temperature gradient within a homogeneous material leads to energy transfer through the medium, where energy moves from the more energetic molecules to the less energetic ones. This heat transfer process can be described using rate equations [5]. The rate of heat transfer in the x-direction through a finite cross-sectional area, A, is governed by Fourier's law, which is given by:

$$q = -kA\frac{dT}{dx} \tag{1}$$

where $k$is the thermal conductivity in W/(m·K). Under steady-state conditions with linear heat transfer, the temperature gradient can be expressed as follows:

Table 1. Thermal Conductivity of Some Fuel Cell Materials

Material	**Thermal Conductivity (W/m·K) @ 300 K**
Aluminum	237
Nickel	90.5
Platinum	71.5
Titanium	22
Stainless steel 316	13
Graphite	98
Carbon cloth	1.7
Teflon	0.4

The thermal conductivity, k, is a transport property that indicates the rate at which energy is transferred through a material by the diffusion process. It depends on the atomic and molecular structure of the substance. The thermal conductivity of some fuel cell materials is provided in **Table 1** [6].

For one-dimensional, steady-state heat conduction with no heat generation, the heat flux remains constant and independent of the position x. This can be expressed as:

$$\frac{d^2T}{dx^2} = 0 \tag{3}$$

When analyzing one-dimensional heat transfer with no internal energy generation and constant properties, there is an analogy between the diffusion of heat and electrical charge. A thermal resistance can be associated with the conduction of heat, and it can be expressed in a plane wall as:

$$R_{\text{cond}} = \frac{T_1 - T_2}{q} = \frac{L}{kA} \tag{4}$$

When heat is conducted through two adjacent materials with different thermal conductivities, the third boundary condition comes from the requirement that the temperature at the interface must be the same for both materials. This can be represented as:

$$q = h_{\text{tc}} A \Delta T \tag{5}$$

Where:

- T_1 and T_2 are the temperatures at the two sides of the material,
- L is the thickness of the material,
- A is the cross-sectional area through which heat is being transferred,
- h_{tc} is the thermal contact resistance,
- ΔT is the temperature difference between the two materials.

Where $h_{\text{tc}}$is the convective heat transfer coefficient, in units of BTU/h·ft²·°F or W/m²·K, $A$is the area normal to the direction of the heat flux, in ft² or m², and $\Delta T$is the temperature difference between the solid surface and the fluid, in °F or K.

A thermal resistance can also be associated with the heat transfer by convection at a surface. This can be expressed as:

$$R_{\text{conv}} = \frac{T_1 - T_2}{q} = \frac{1}{hA} \tag{6}$$

Since the conduction and convection resistances are in series, they can be summed as follows:

$$R_{\text{tot}} = \frac{1}{\frac{1}{R_{\text{conv}}} + \frac{1}{R_{\text{cond}}}} = \frac{L}{kA} + \frac{1}{hA} \tag{7}$$

The fuel cell layers can be modeled as a "composite wall" with series thermal resistances due to the different materials. For a composite system, it is convenient to work with an overall heat transfer coefficient, U, which can be written as:

$$q = UA\Delta T \tag{8}$$

Similar to charge transport (Chapter 4), in composite systems, the temperature drop between materials can be significant, and this is referred to as thermal contact resistance. It is defined as:

$$R_{\text{tc}} = \frac{T_A - T_B}{q} \tag{9}$$

The total thermal resistance R_{th}can also be expressed as:

$$R_{\text{th}} = R_c + R_{\text{tc}} \tag{10}$$

where $R_c$is the convective thermal resistance:

$$R_c = \frac{1}{hA} \tag{11}$$

where $h$is the convective heat transfer coefficient, and $A$is the heat transfer area. $h$can also be expressed in terms of the Nusselt number (Nu) as: $h = \frac{k}{L} NuL$. Figure 2 illustrates the temperature drop due to thermal contact resistance.

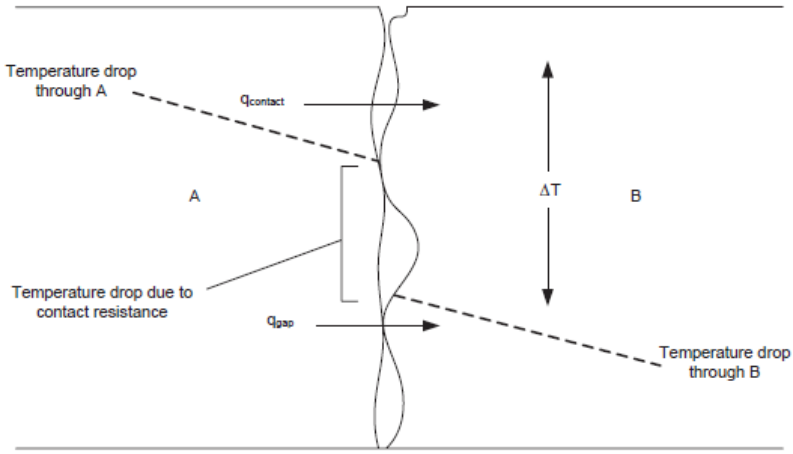


Fig. 2 Temperature drop across the interface due to thermal contact resistance [4]

MATERIALS AND METHOD

To achieve more accurate solutions for heat transfer problems, numerical techniques such as finite-difference, finite-element, and boundary-element methods are commonly employed. These methods are flexible and can be extended to handle problems in up to three dimensions, making them suitable for complex geometries and varying conditions. One of the key approaches in numerical heat transfer analysis is the nodal network solution, which enables the determination of temperature values at specific points within the system.

In a nodal network, the physical domain of interest is divided into smaller regions, each of which is represented by a discrete point called a nodal point. These points form a grid or mesh, which can be one-dimensional, two-dimensional, or three-dimensional, depending on the complexity of the problem. The temperature at each nodal point is determined by solving an energy balance for that point, taking into account heat flow and boundary conditions.

The process begins by subdividing the domain into a network of nodes, where each node represents a reference point. These nodes are interconnected, and their interactions describe the heat transfer throughout the system. By applying appropriate energy equations to each node and solving them simultaneously, the temperature distribution across the entire

domain can be obtained. This method provides a detailed, localized understanding of how heat moves within the system.

For example, in a two-dimensional heat transfer problem, the locations of the nodes in both the $x$and $y$directions are carefully chosen to represent the physical layout of the material. This is shown in Figure 3, which illustrates the placement of the nodes in a two-dimensional grid. Each node is connected to its neighboring nodes, forming a network that can be solved for the temperature at each point.

In the following section, an example of transient heat conduction in a plate using a one-dimensional nodal network will be demonstrated. This example will showcase how the nodal method can be applied to time-dependent problems, allowing for the study of temperature changes over time as heat flows through a material.

By using these numerical techniques, engineers and researchers can model and simulate heat transfer processes with high precision, providing valuable insights for designing and optimizing thermal systems, including fuel cells, electronic devices, and industrial applications.

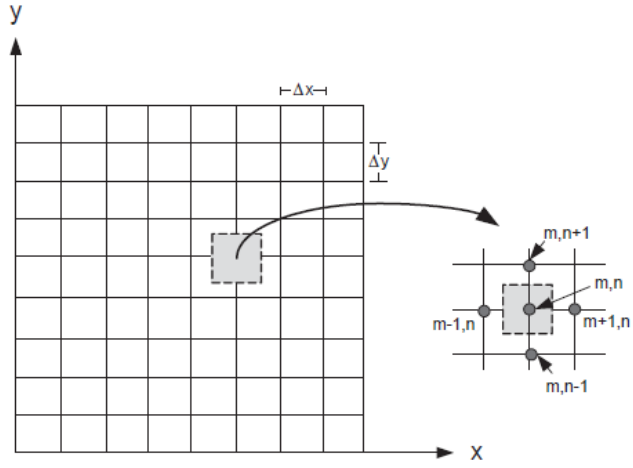


Fig. 3 Two-Dimensional Conduction Analysis Using a Nodal Network [4]

RESULTS

The procedure for obtaining a numerical solution to a one-dimensional, transient heat conduction problem is demonstrated using a representative layer within a fuel cell stack, as illustrated in Figure 4.

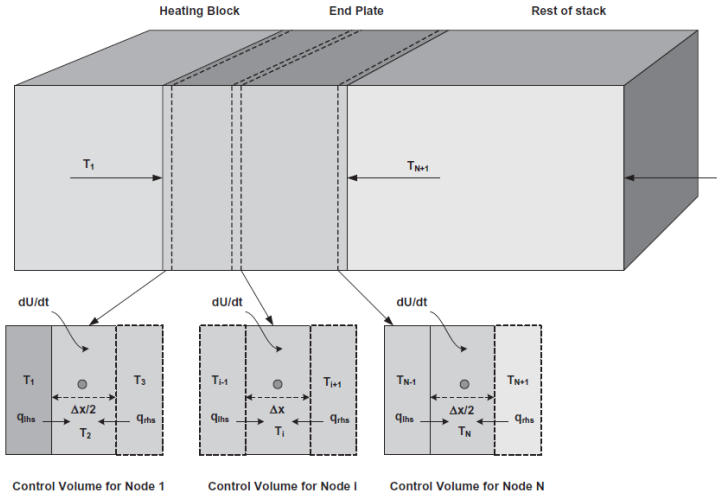


Fig. 4 Nodes uniformly distributed throughout the computational domain [4]

For the case of a uniformly distributed nodal arrangement, as illustrated in **Figure 4**, the spatial position of each node $x_i$can be defined as a function of the total layer thickness and the number of nodes. The nodal locations are given by:

$$x_i = \frac{(i-1)L}{(N-1)}, i = 1,2,\dots,N \qquad (12)$$

where $L$is the total thickness of the layer and Nrepresents the total number of nodes used in the numerical simulation. With this uniform discretization, the distance between two adjacent nodes, Δx, is constant and can be expressed as:

$$\Delta x = \frac{L}{N-1} \qquad (13)$$

To formulate the transient heat conduction problem, an energy balance is applied to each nodal control volume. The control volumes associated with the first node, the last node, and a typical internal node are shown schematically in **Figure 4**. Each control volume exchanges heat with its neighboring nodes through conduction and simultaneously stores thermal energy due to temperature variation with time.

The general energy balance for a nodal control volume can be written as:

$$\dot{q}_{\text{LHS}} + \dot{q}_{\text{RHS}} = \frac{dU}{dt} \quad (14)$$

where $\dot{q}_{\text{LHS}}$and $\dot{q}_{\text{RHS}}$denote the conductive heat transfer rates from the left-hand-side and right-hand-side neighboring nodes, respectively, and $\frac{dU}{dt}$represents the rate of change of internal energy within the control volume.

The conductive heat transfer rates from the adjacent nodes are approximated using Fourier's law as follows:

$$\dot{q}_{\text{LHS}} = \frac{kA}{\Delta x}(T_{i-1} - T_i) \quad (15)$$

$$\dot{q}_{\text{RHS}} = \frac{kA}{\Delta x}(T_{i+1} - T_i) \quad (17)$$

where $k$is the thermal conductivity of the material, $A$is the cross-sectional area normal to the heat flow direction, and $T_i$is the temperature at node i.

The rate of energy storage within the control volume is determined by the thermal mass and the temporal change in nodal temperature:

$$\frac{dU}{dt} = \rho c A \Delta x \frac{dT_i}{dt} \quad (16)$$

where $\rho$is the material density and $c$is the specific heat capacity.

Substituting Equations (6-28) through (6-30) into the energy balance equation yields:

$$\rho c A \Delta x \frac{dT_i}{dt} = \frac{kA}{\Delta x}(T_{i-1} - T_i) + \frac{kA}{\Delta x}(T_{i+1} - T_i) \quad (18)$$

Rearranging the above expression gives the governing equation for the temporal evolution of temperature at an internal node:

$$\frac{dT_i}{dt} = \frac{k}{\rho c(\Delta x)^2}(T_{i-1} + T_{i+1} - 2T_i), i = 2,3, \dots, N-1 \quad (19)$$

This formulation forms the basis for solving one-dimensional transient heat conduction problems using a nodal network approach and is particularly suitable for modeling thermal behavior within individual layers of a fuel cell stack.

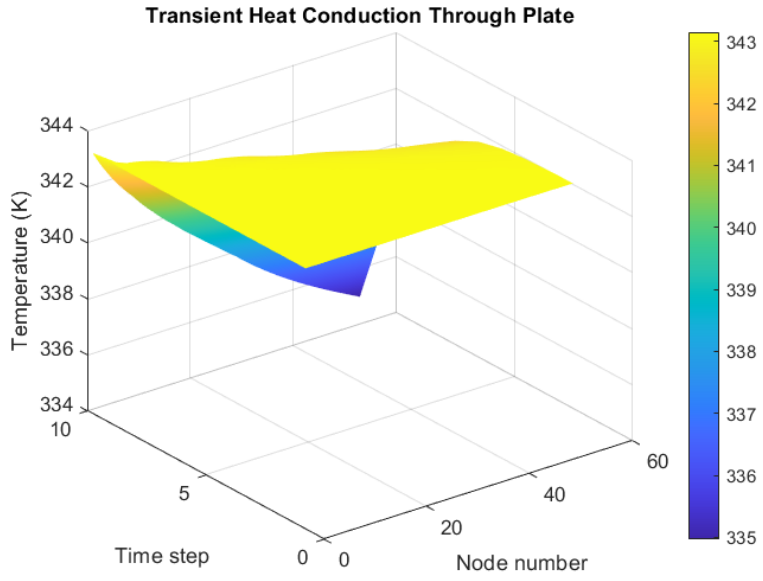


Fig. 5 Transient temperature distribution through the plate obtained using a one-dimensional nodal network approach for 10 secs.

Figure 5 illustrates the transient temperature distribution across the plate as a function of both spatial location and time. The results clearly demonstrate the temporal evolution of heat conduction from the initially uniform temperature toward a new thermal equilibrium. At the beginning of the simulation, all nodes are at the same initial temperature, whereas as time progresses, a temperature gradient develops due to heat loss at the convective boundary.

The temperature near the convective surface decreases more rapidly compared to the interior and adiabatic regions, indicating the dominant influence of convective heat transfer at the boundary. In contrast, the adiabatic side exhibits a slower temperature variation, as heat transfer is limited to conduction within the plate. This behavior confirms the physical consistency of the applied boundary conditions.

Overall, Fig. 5 highlights the capability of the nodal network approach to accurately capture transient heat conduction phenomena in fuel cell components. The results emphasize the importance of transient thermal analysis for understanding temperature gradients within fuel cell layers, which can significantly affect electrochemical performance, material durability, and overall system efficiency.

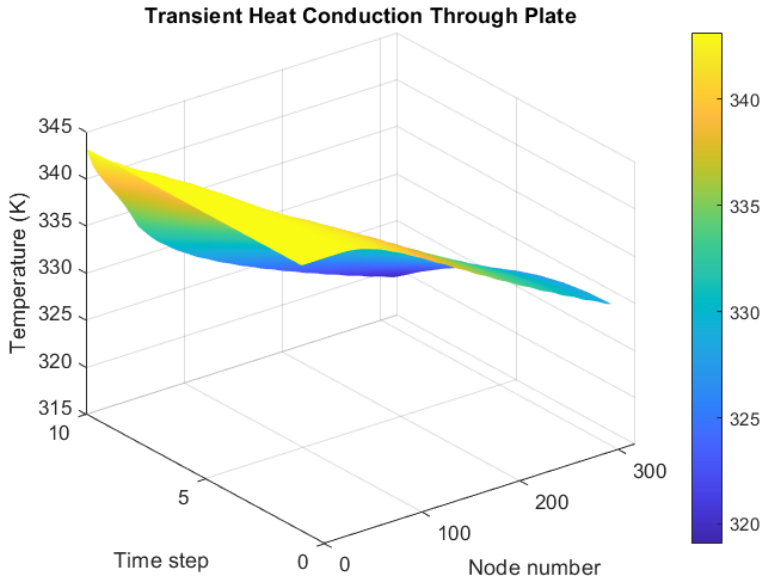


Fig. 6 Transient temperature distribution through the plate obtained using a one-dimensional nodal network approach for 150 secs.

Figure 6 illustrates the transient temperature distribution across the plate after extending the simulation time to 150 seconds. Compared to the earlier results obtained at shorter simulation times, a more pronounced thermal evolution is observed throughout the computational domain.

As the simulation time increases, thermal energy has sufficient time to propagate deeper into the plate via conduction. Consequently, the temperature gradients between adjacent nodes decrease, indicating a gradual approach toward a quasi–steady-state condition. The temperature near the convective boundary (outer surface) continues to decrease due to sustained heat loss to the surrounding fluid, while the interior nodes exhibit a smoother and more uniform temperature profile.

The extended simulation time highlights the dominant role of material thermal properties—such as thermal conductivity, density, and specific heat capacity—in governing the rate of heat diffusion. At 150 seconds, the temperature field shows reduced temporal variation compared to earlier time steps, suggesting that the system is approaching thermal equilibrium.

This behavior is particularly relevant for fuel cell applications, where prolonged operation leads to stabilized temperature distributions within stack components. Understanding this long-term thermal response is essential for predicting durability, managing thermal stresses, and ensuring uniform electrochemical performance across the fuel cell stack.

Overall, Fig. 6 demonstrates that increasing the simulation time allows the thermal system to evolve from an initial transient regime toward a more stable thermal state, emphasizing the importance of transient heat transfer analysis in fuel cell thermal management studies.

DISCUSSION

The transient heat conduction behavior of the fuel cell plate was investigated by extending the simulation time from 10 s to 100 s and further to 150 s. The resulting temperature distributions are presented in Fig. 5 and Fig. 6, respectively. These results provide insight into the temporal evolution of heat transfer within the plate and the progression toward thermal equilibrium.

At earlier times (10 s, not shown here), the temperature distribution is dominated by the initial condition, and only a limited portion of the plate is affected by convective cooling at the boundary. As the simulation time increases to 100 s (Fig. 5), thermal diffusion becomes more pronounced, and a smoother temperature gradient develops across the thickness of the plate. The influence of the convective boundary condition at the exposed surface becomes clearly observable, leading to a noticeable temperature reduction near the outer surface while the interior regions remain relatively warm.

When the simulation time is further extended to 150 s (Fig. 6), the temperature distribution approaches a quasi-steady state. The temperature gradient across the plate becomes more uniform, and the overall temperature level decreases compared to the 100 s case. This behavior indicates that the rate of heat removal by convection increasingly balances the conductive heat transfer within the solid. The thermal inertia of the plate, governed by its density and specific heat capacity, slows the cooling process, but over longer times the system progressively adjusts toward equilibrium with the surrounding fluid temperature.

Comparing Fig. 5 and Fig. 6 reveals that the most significant temperature changes occur during the early and intermediate stages of the transient process. Beyond 100 s, the rate of temperature change diminishes, suggesting that additional simulation time yields diminishing variations in the temperature field. This trend is characteristic of transient conduction problems in solids with convective boundary conditions and highlights the importance of time scale selection when modeling thermal behavior in fuel cell components.

From a fuel cell perspective, these results emphasize that thermal gradients within cell layers persist for extended periods during operation or startup. Such gradients can influence local reaction kinetics, water transport, and material durability. Therefore, accurate transient thermal modeling is essential for predicting real operating conditions and for designing effective thermal management strategies in fuel cell stacks.

CONCLUSION

In this study, transient one-dimensional heat conduction through a fuel cell plate was investigated using a nodal network approach. The governing energy balance equations were solved numerically to predict the time-dependent temperature distribution within the plate under convective boundary conditions.

The results demonstrate that simulation time has a significant impact on temperature profiles. While large temperature gradients are present at early times, prolonged operation leads to a more uniform temperature distribution as heat conduction and convection reach a balance. After 150 seconds, the system exhibits a smoother thermal field, indicating improved thermal stability.

The adopted numerical framework provides a reliable basis for analyzing heat transfer in fuel cell components and offers valuable insights for thermal management design. Future work may extend this model to multilayer fuel cell assemblies, include internal heat generation due to electrochemical reactions, and incorporate two- or three-dimensional effects to better represent real fuel cell operating conditions.

REFERENCES

[1] G. Jiao, G. Xia, T. Zhou, M. Peng, ve J. Wang, “Transient local effects analysis during typical accidents of a heat pipe cooled reactor”, *Nucl. Eng. Des.*, c. 436, s. 113988, May. 2025, doi: 10.1016/j.nucengdes.2025.113988.

[2] T. Tanaka, T. Ando, H. Ogura, ve Y. Inui, “Numerical modeling for commercial-scale tubular segmented-in-series solid oxide fuel cell providing

reliable calculation result in short calculation time", *Chem. Eng. J.*, c. 525, s. 170216, Ara. 2025, doi: 10.1016/j.cej.2025.170216.

[3] "On characteristics and research development of coupled fuel cell stack performance and stress - ScienceDirect". Erişim: 15 Aralık 2025. [Çevrimiçi]. Erişim adresi: https://www.sciencedirect.com/science/article/pii/S0306261925004490

[4] "PEM Fuel Cell Modeling and Simulation Using Matlab", ScienceDirect. Erişim: 15 Aralık 2025. [Çevrimiçi]. Erişim adresi: http://www.sciencedirect.com:5070/book/monograph/9780123742599/pem-fuel-cell-modeling-and-simulation-using-matlab

[5] M. V. Afraz, A. Sattari, Z. Alimohammadi, ve G. Karimi, "Dynamic thermal safety of Tesla-like lithium-ion battery packs: a transient analysis of cooling failures and overheating in electric vehicles", *Appl. Therm. Eng.*, c. 284, s. 129186, Oca. 2026, doi: 10.1016/j.applthermaleng.2025.129186.

[6] K. Yoshimura, N. Doda, M. Tanaka, T. Fujisaki, ve S. Murakami, "Development of 1D-CFD coupling method for natural circulation analyses through benchmark analyses of shutdown heat removal tests in EBR-II", *Ann. Nucl. Energy*, c. 226, s. 111896, Şub. 2026, doi: 10.1016/j.anucene.2025.111896.

Energy Balance Analysis of Layered Structures in Proton Exchange Membrane Fuel Cells

Hüseyin KAHRAMAN[*1]

[1]Mechanical Engineering Department/Faculty of Technology, Sakarya University of Applied Sciences, Türkiye
[*]huseyink@subu.edu.tr

ABSTRACT

Thermal management is a critical aspect of proton exchange membrane fuel cell (PEMFC) operation, directly influencing performance, durability, and operational stability. In multilayer fuel cell stacks, temperature distributions are inherently non-uniform due to differences in material properties, interfacial heat transfer, and convective heat losses at the stack boundaries. Accurate prediction of transient temperature evolution therefore requires a detailed layer-by-layer energy balance analysis.

In this study, a one-dimensional transient heat transfer model is developed to investigate the thermal behavior of layered PEM fuel cell structures. Energy balance equations are formulated for individual fuel cell layers, including end plates, gaskets, flow field plates, and the membrane electrode assembly (MEA). The model accounts for conductive heat transfer within solid layers, convective heat loss to the surrounding environment, and interfacial thermal resistances between adjacent layers. A nodal network approach is employed to discretize the stack in the through-plane direction, resulting in a system of time-dependent ordinary differential equations solved numerically.

Transient simulations are performed for different operating durations (60 s, 240 s, and 480 s) to capture both early-stage and long-term thermal responses. The results demonstrate that significant temperature gradients develop during the initial transient period, particularly near the end plates due to convective cooling. As operating time increases, thermal diffusion becomes dominant, leading to smoother temperature profiles and a gradual transition toward a quasi–steady-state condition. The findings highlight the importance of transient thermal analysis in fuel cell design and emphasize the need for accurate energy balance formulations when evaluating thermal behavior in layered fuel cell stacks.

Keywords – Fuel Cell, Heat Transfer, Transient Conduction, Nodal Network Method, Thermal Management.

INTRODUCTION

In Energy balances can be formulated around each individual layer of a fuel cell to enable a detailed investigation of heat diffusion as a function of both time and spatial position. Since a fuel cell stack consists of multiple layers with distinct thermal and physical properties, analyzing heat transfer at the layer level is essential for accurately predicting temperature distributions and transient thermal behavior [1].

Each layer within the fuel cells such as the end plate, gasket, contact interface, flow field, gas diffusion layer (GDL), catalyst layer, and membrane—can be treated as a control volume. By applying energy balance equations to these control volumes, the mechanisms governing heat conduction, convection, and thermal storage can be systematically described.

Figure 1 illustrates a representative fuel cell layer modeled as a control volume, highlighting the heat fluxes entering and leaving the layer as well as the energy stored within it [2].

Heat diffusion within fuel cell layers plays a critical role in overall cell performance and durability. Temperature gradients influence electrochemical reaction kinetics, water transport, membrane hydration, and mechanical stresses within the stack. Uneven temperature distributions may lead to localized hot spots, accelerated material degradation, or non-uniform current density distribution. Therefore, understanding how heat propagates through each layer is a prerequisite for effective thermal management and reliable fuel cell operation [3].

In this study, the modes of heat diffusion within individual fuel cell layers are examined by establishing energy balance equations tailored to the specific characteristics of each layer. Conductive heat transfer dominates within solid and porous layers, while convective heat transfer is particularly relevant at interfaces involving gas flow channels and cooling media. The transient nature of heat transfer is also considered, allowing temperature evolution to be evaluated over time under realistic operating conditions.

The analysis presented in this section systematically derives the energy balance formulations for the end plate, gasket, contact resistance region, flow field, gas diffusion layer, catalyst layer, and polymer electrolyte membrane. By resolving the thermal behavior of each layer independently and in relation to adjacent layers, a comprehensive understanding of heat transfer pathways within the fuel cell stack is achieved. This layer-by-layer energy balance approach provides a robust framework for predicting temperature distributions and supports the design and optimization of advanced thermal management strategies for fuel cell systems [3].

The end plates of a fuel cell stack are typically manufactured from metallic or high-strength polymeric materials and play a critical structural role within the system. Their primary function is to provide mechanical integrity by uniformly transmitting compressive forces across the entire fuel cell stack. Adequate and uniform compression is essential to ensure proper interfacial contact between the various layers, minimize contact resistance, and prevent gas leakage. Consequently, end plates must possess sufficient mechanical stiffness and strength to withstand operational loads while maintaining dimensional stability over extended periods of operation.

In addition to end plates, fuel cell stacks commonly include contact layers and gasket materials. Gasket layers are designed to seal the reactant gas pathways and prevent leakage of hydrogen, oxygen, or air from the flow channels. Proper gasket performance also contributes to maintaining consistent compression throughout the stack, which is necessary for reliable electrochemical and thermal behavior. Contact layers, often referred to as current collectors, serve as electrical interfaces that facilitate the transfer of

electrons from the gas diffusion layer (GDL) and bipolar plates to the external circuit.

From a thermal perspective, the end plates, contact layers, and gaskets are generally not directly involved in fluid transport processes. Unlike the flow field channels or porous layers, these components do not experience internal gas or liquid flow under normal operating conditions. As a result, heat transfer within these layers occurs predominantly through conduction [4].

One side of these layers is typically exposed either to an insulating material or directly to the ambient environment, while the opposite side is in contact with a conductive component such as a bipolar plate or current collector. This configuration creates a one-dimensional heat conduction path across each layer, governed by the material's thermal conductivity, thickness, and interfacial contact resistance.

An energy balance applied to these layers accounts for conductive heat transfer entering and leaving the control volume, as illustrated schematically in Figure 1. Since no internal heat generation or convective heat transfer is present within these components, the thermal analysis can be simplified to steady or transient conduction, depending on the operating conditions and time scale of interest [5].

Accurate modeling of heat transfer through the end plates, contact layers, and gaskets is important because these components influence overall stack temperature distribution and heat dissipation to the surroundings. Improper thermal design of these layers may lead to non-uniform temperature fields, increased thermal stresses, or inefficient heat removal, all of which can adversely affect fuel cell performance and durability [6].

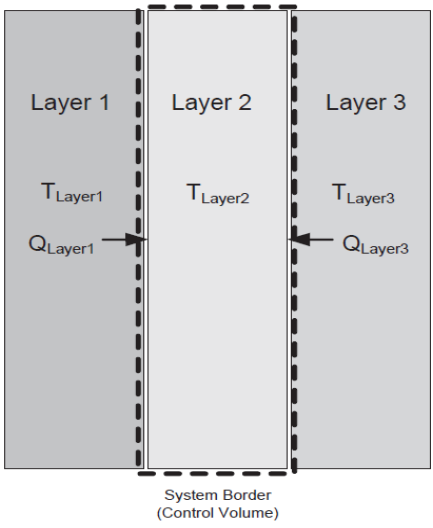


Fig. 1. Energy balance formulation for the flow field (bipolar) plate [7]

In a fuel cell stack, bipolar plates serve multiple functions, including separating the reactant gases of neighboring cells, providing electrical interconnection between cells, and offering structural support. These plates contain flow channels on both sides, forming the anode and cathode flow fields for adjacent unit cells. The geometry of the flow channels influences reactant distribution, mass transport, and overall cell performance. Bipolar plate materials are required to exhibit high electrical and thermal conductivity, gas impermeability, and strong resistance to corrosion and chemical degradation due to the harsh electrochemical environment. An illustration of the associated energy balance is presented in Figure 2.

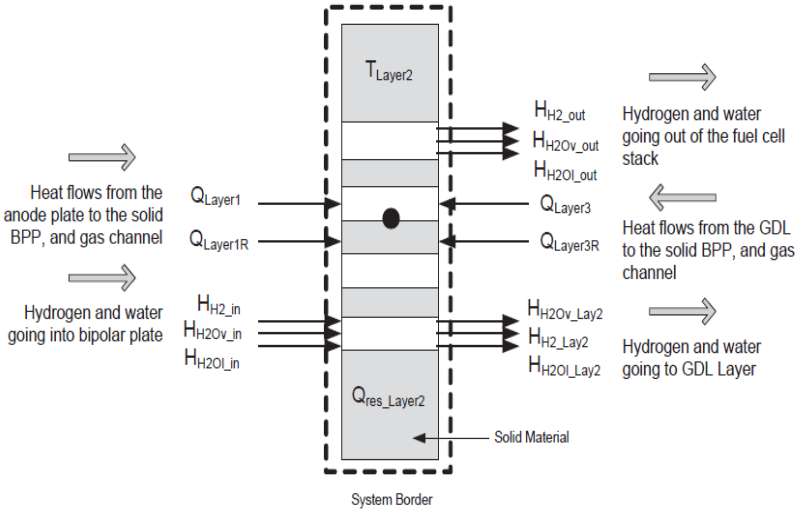


Fig. 2. Energy balance representation for an individual fuel cell layer [7]

MATERIALS AND METHOD

To The general energy balance for the end plate, contact, and gas diffusion layer (GDL) can be expressed as:

$$\rho_{\text{Layer2}} A_{\text{Layer2}} t_{\text{Layer2}} c_{p,\text{Layer2}} \frac{dT_{\text{Layer2}}}{dt} = Q_{\text{Layer1}} + Q_{\text{Layer3}} \qquad (1)$$

where $\rho_{\text{Layer2}}$is the density of Layer 2,
$A_{\text{Layer2}}$is the surface area of Layer 2,
$t_{\text{Layer2}}$is the thickness of Layer 2,
$c_{p,\text{Layer2}}$is the specific heat capacity of Layer 2,
$Q_{\text{Layer1}}$is the heat transfer rate from Layer 1 to Layer 2, and
$Q_{\text{Layer3}}$is the heat transfer rate from Layer 3 to Layer 2.

The left-hand side of Eq. (1) represents the rate of change of the control volume temperature of Layer 2.

The heat transfer from Layer 1 to Layer 2 is given by:

$$Q_{\text{Layer1}} = U_{\text{Layer1}} A\left(T_{\text{Layer1}} - T_{\text{Layer2}}\right) \qquad (2)$$

where $U_{\text{Layer1}}$is the overall heat transfer coefficient associated with Layer 1, $A$is the interfacial area, and Tdenotes temperature.

Similarly, the heat transfer from Layer 3 to Layer 2 can be written as:

$$Q_{\text{Layer3}} = U_{\text{Layer3}} A\left(T_{\text{Layer3}} - T_{\text{Layer2}}\right) \qquad (3)$$

If the heat exchange occurs between the layer and the surrounding environment, the overall heat transfer coefficient can be expressed as:

$$U_{\text{surr}} = \left(\frac{t_{\text{Layer2}}}{k_{\text{Layer2}}} + \frac{1}{h_{\text{surr}}}\right)^{-1} \qquad (4)$$

where $k_{\text{Layer2}}$is the thermal conductivity of Layer 2 and
$h_{\text{surr}}$is the convective heat transfer coefficient between the stack surface and the ambient air.

The overall heat transfer coefficient for heat transfer from Layer 1 to Layer 2 is given by:

$$U_{\text{Layer1}} = \left(\frac{t_{\text{Layer1}}}{k_{\text{Layer1}}} + \frac{t_{\text{Layer2}}}{k_{\text{Layer2}}}\right)^{-1} \qquad (5)$$

Likewise, the overall heat transfer coefficient for heat transfer from Layer 3 to Layer 2 can be written as:

$$U_{\text{Layer3}} = \left(\frac{t_{\text{Layer3}}}{k_{\text{Layer3}}} + \frac{t_{\text{Layer2}}}{k_{\text{Layer2}}}\right)^{-1} \qquad (6)$$

The bipolar plate experiences both conductive and convective heat transfer due to the presence of gas flow channels embedded within the plate structure. The fraction of the bipolar plate occupied by flow channels significantly influences the overall heat transfer characteristics. To account for this effect, an effective cross-sectional area for conduction, denoted as A_{1R}, is defined. This area represents the portion of the solid material that is in direct contact with the adjacent layers (Layer 1 and Layer 3).

The governing energy balance equation for heat transfer in an anode-side bipolar plate can be expressed as:

$$\left(\rho_{Layer2} c_{p,Layer2} A_{Layer2} t_{Layer2}\right) \frac{dT_{Layer2}}{dt} = Q_{Layer1} + Q_{Layer1R} + Q_{Layer3} + Q_{Layer3R} + Q_{\text{res,Layer2}} + \sum H_{\text{in}} - \sum H_{\text{out}} \quad (7)$$

where

ρ_{Layer2} is the density of Layer 2 (bipolar plate),

$c_{p,Layer2}$ is the specific heat capacity of Layer 2,

A_{Layer2} and t_{Layer2} are the area and thickness of Layer 2,

Q_{Layer1} and Q_{Layer3} are the heat transfer rates from adjacent layers to the gas channels,

$Q_{Layer1R}$ and $Q_{Layer3R}$ are the conductive heat transfer rates to the solid portion of the bipolar plate,

$Q_{\text{res,Layer2}}$ is the heat generated due to electrical resistance, and H represents the enthalpy associated with species entering and leaving the control volume.

The left-hand side of Eq. (7) represents the rate of change of thermal energy within the bipolar plate control volume, while the right-hand side accounts for all conductive, convective, resistive, and enthalpy-related heat transfer mechanisms.

The heat transfer from Layer 1 to the gas channels is given by:

$$Q_{Layer1} = U_{Layer1} A_{\text{void}} (T_{Layer1} - T_{Layer2}) \quad (8)$$

where A_{void} is the area associated with the flow channels.

The conductive heat transfer from Layer 1 to the solid portion of the bipolar plate is:

$$Q_{Layer1R} = U_{Layer1R} A_{1R} (T_{Layer1} - T_{Layer2}) \quad (9)$$

Similarly, the heat transfer from Layer 3 (GDL) to the gas channels is expressed as:

$$Q_{Layer3} = U_{Layer3} A_{\text{void}} (T_{Layer3} - T_{Layer2}) \quad (10)$$

and the conductive heat transfer from Layer 3 to the solid portion of the bipolar plate is:

$$Q_{Layer3R} = U_{Layer3R} A_{1R} \left(T_{Layer3} - T_{Layer2}\right) \quad (11)$$

Here, A_{void} denotes the area occupied by gas channels, and A_{1R} represents the effective solid contact area of the bipolar plate.

Due to the coexistence of solid conduction paths and gas-filled flow channels, the bipolar plate transfers different amounts of heat to solid material and the flowing gases. This dual-mode heat transfer behavior plays a critical role in determining local temperature gradients within the fuel cell stack. Accurate modeling of these mechanisms is essential for predicting thermal non-uniformities, managing heat removal, and ensuring reliable long-term operation of fuel cell systems.

Table 1. Material properties used for heat transfer calculations in fuel cell layer energy balance analysis [7]

Fuel Cell Layer	**Material**	**Thickness (m)**	**Area (m^2)**	**Density (kg/m^3)**	**Thermal Conductivity (W/m·K)**	**Specific Heat Capacity (J/kg·K)**
End plate	Polycarbonate	0.01	0.0064	1300	0.2	1200
Gasket	Conductive rubber	0.001	0.001704	1400	1.26	1000
Flow field plate	Stainless steel	0.0005	0.003385	8000	65	500
MEA	Carbon cloth / Pt/C / Nafion	0.001	0.0016	1300	26	864
Flow field plate	Stainless steel	0.0005	0.003385	8000	65	500
Gasket	Conductive rubber	0.001	0.001704	1400	1.26	1000
End plate	Polycarbonate	0.01	0.0064	1300	0.2	1200

Table 1 presents the material properties used in the heat transfer calculations for the fuel cell layers considered in Example 6-3. The table summarizes the key thermophysical parameters required for transient energy balance modeling, including layer thickness, effective heat transfer area, density, thermal conductivity, and specific heat capacity.

The selected materials represent a typical proton exchange membrane fuel cell (PEMFC) stack configuration, consisting of end plates, gaskets, flow field plates, and the membrane electrode assembly (MEA). Polycarbonate end plates are included due to their mechanical strength and relatively low thermal conductivity, which influences heat dissipation toward the surroundings. Conductive rubber gaskets are used to ensure sealing and electrical continuity while contributing to interfacial thermal resistance.

Stainless steel flow field plates are characterized by high density and thermal conductivity, enabling efficient heat spreading and structural support. The MEA, composed of carbon cloth, platinum catalyst, and Nafion membrane, exhibits intermediate thermal properties that strongly affect internal heat diffusion and electrochemical heat generation.

These material properties are critical inputs for accurately predicting transient temperature distributions within the fuel cell stack. Variations in thermal conductivity and heat capacity among layers directly influence the rate of heat diffusion, thermal gradients, and overall thermal response of the system. Therefore, the data provided in Table 1 form the basis for reliable numerical heat transfer analysis and realistic thermal modeling of fuel cell components.

RESULTS

Overall, Fig. 6 demonstrates that increasing the simulation time allows the thermal system to evolve from an initial transient regime toward a more stable thermal state, emphasizing the importance of transient heat transfer analysis in fuel cell thermal management studies.

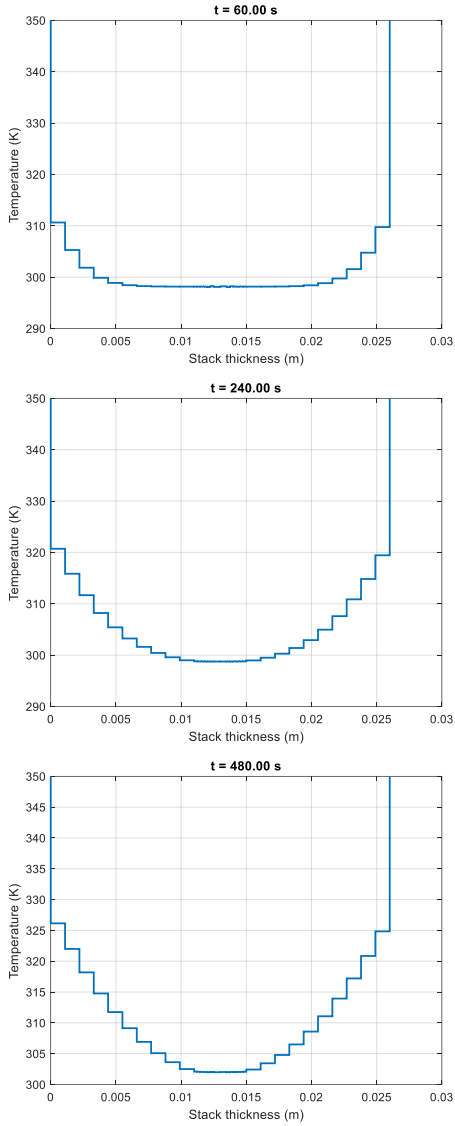


Fig. 3. Transient temperature distribution across the fuel cell stack at different times

Figure 3 presents the transient temperature distributions across the fuel cell stack at three different simulation times: 60 s, 240 s, and 480 s. These results illustrate the time-dependent thermal response of the multilayer fuel cell structure under conductive and convective heat transfer conditions.

At t = 60 s, the temperature distribution is still strongly influenced by the initial conditions. Pronounced temperature gradients are observed, particularly near the outer end plates, where convective heat losses to the surroundings dominate. The interior layers, including the flow field plates

and the MEA region, retain higher temperatures due to thermal inertia and limited heat penetration at this early stage. This indicates that heat diffusion across the stack is still in its early transient regime.

When the simulation time is extended to t = 240 s, thermal diffusion becomes significantly more developed. The temperature gradients across adjacent layers decrease, and the temperature profile becomes smoother throughout the stack. Heat transfer from the interior layers toward the end plates is more evident, reflecting the increasing role of conduction through the multilayer structure. The influence of material properties, such as thermal conductivity and heat capacity, becomes more pronounced, particularly in the metallic flow field plates, which facilitate faster heat spreading compared to polymer-based layers.

At t = 480 s, the system approaches a quasi–steady-state thermal condition. The temperature distribution shows minimal temporal variation compared to earlier times, indicating that conductive heat transfer within the stack is increasingly balanced by convective heat losses at the boundaries. The temperature profile across the layers becomes nearly stable, with reduced gradients between neighboring layers. This behavior suggests that the fuel cell stack has largely adapted to the imposed boundary conditions and ambient environment.

Overall, Fig. 3 demonstrates that the most significant thermal evolution occurs during the early and intermediate stages of operation. As time progresses, the rate of temperature change diminishes, and the stack transitions toward thermal equilibrium. From a fuel cell operation perspective, this highlights the importance of transient thermal analysis, particularly during startup and load changes, as temperature gradients persist for extended periods and can influence electrochemical performance, water management, and long-term durability of the fuel cell components.

DISCUSSION

The transient thermal behavior of the layered fuel cell stack was systematically analyzed by examining temperature distributions at three different time instants: 60 s, 240 s, and 480 s, as illustrated in Fig. 3. These time frames represent the early transient, intermediate transient, and near steady-state thermal regimes, respectively, and provide valuable insight into the dynamic heat diffusion mechanisms within the multilayer structure.

At the early stage (t = 60 s), the temperature field is strongly influenced by the prescribed boundary and initial conditions. Significant temperature gradients are observed near the end plates, where convective heat losses to the ambient environment are dominant. The interior layers, particularly the flow field plates and the membrane electrode assembly (MEA), retain higher temperatures due to their thermal inertia and limited heat penetration at short times. This behavior indicates that heat diffusion across the stack is still in its initial transient phase, and the thermal response is primarily governed by local material properties and boundary effects.

As the simulation progresses to t = 240 s, heat conduction becomes more effective across adjacent layers, resulting in a smoother temperature profile throughout the stack. The temperature differences between neighboring layers decrease, demonstrating enhanced thermal coupling within the multilayer system. Metallic flow field plates, characterized by relatively high thermal conductivity, play a key role in redistributing heat, while polymer-based components such as gaskets and end plates continue to act as thermal resistances. At this stage, both material properties and layer thicknesses significantly influence the overall temperature distribution.

At the longest simulation time considered (t = 480 s), the fuel cell stack approaches a quasi–steady-state thermal condition. The temperature gradients across the layers are substantially reduced, and temporal variations become minimal. This indicates that conductive heat transfer within the stack is increasingly balanced by convective heat losses at the outer boundaries. The system exhibits a stabilized thermal profile, suggesting that the dominant heat transfer mechanisms have reached equilibrium under the imposed operating conditions.

The comparison of these three time instants highlights that the most pronounced thermal changes occur during the early and intermediate stages of operation. Beyond a certain time scale, further increases in operating duration result in diminishing changes in the temperature field. This observation is particularly relevant for fuel cell systems, as startup and transient operating conditions are more likely to induce significant thermal gradients than steady-state operation.

From a practical perspective, persistent temperature gradients during transient operation can affect electrochemical reaction rates, membrane hydration, water transport, and mechanical stresses within the stack. Therefore, the results emphasize the necessity of transient thermal analysis in fuel cell design and operation. The nodal network-based energy balance approach employed in this study proves to be an effective and computationally efficient tool for capturing these time-dependent thermal effects and can support the development of improved thermal management strategies for layered fuel cell structures.

CONCLUSION

In this study, a comprehensive energy balance framework was developed to analyze transient heat transfer in layered proton exchange membrane fuel cell (PEMFC) structures. Each fuel cell layer was modeled as an individual control volume, allowing conductive and convective heat transfer mechanisms to be systematically represented within a one-dimensional nodal network formulation. The resulting governing equations were solved numerically to predict the time-dependent temperature distribution across the multilayer stack.

The numerical results demonstrate that thermal behavior in fuel cell layers is strongly time-dependent. Significant temperature gradients are observed during the early stages of operation, particularly near the end plates where convective heat losses to the surroundings are dominant. As the simulation time increases, heat diffusion through the solid layers becomes more effective, leading to smoother temperature profiles and a gradual transition toward a quasi–steady-state condition. At longer operating times, the balance between internal conduction and boundary convection governs the overall thermal response of the stack.

The analysis highlights the critical influence of material properties—such as thermal conductivity, density, and specific heat capacity—on heat propagation and thermal inertia within the fuel cell. Metallic flow field plates promote rapid heat spreading, while polymer-based components contribute to localized thermal resistance and delayed temperature equilibration. These effects underline the importance of layer-specific thermal modeling when evaluating fuel cell performance and durability.

Overall, the proposed energy balance and nodal network approach provides a computationally efficient and physically consistent tool for investigating transient heat transfer in fuel cell layers. The methodology can be readily extended to include internal heat generation from electrochemical reactions, multilayer coupling effects, and multidimensional geometries. Such extensions will further support the design and optimization of effective thermal management strategies for advanced fuel cell systems.

REFERENCES

[1] J.-R. Zhang *vd.*, "Analysis of oxygen transport and output performance of a proton exchange membrane fuel cell with order-structured catalyst layer using Sobol global sensitivity method", *J. Power Sources*, c. 659, s. 238395, Ara. 2025, doi: 10.1016/j.jpowsour.2025.238395.

[2] L. Han *vd.*, "Structure, design, and advanced characterization techniques of catalyst layers in proton exchange membrane fuel cells", *Chin. Chem. Lett.*, s. 111432, Haz. 2025, doi: 10.1016/j.cclet.2025.111432.

[3] S. Feng, F. Yan, H. Lu, Y. Su, ve H. Qin, "Study on the coupling effect of cathode slope converging flow field and gas diffusion layer porosity gradient on the performance of proton exchange membrane fuel cell", *Appl. Therm. Eng.*, c. 282, s. 128926, Oca. 2026, doi: 10.1016/j.applthermaleng.2025.128926.

[4] Y. Shen, J. Fu, C. Li, C. Chang, ve X. Sun, "Irreversible thermodynamics-driven optimization of proton exchange membrane fuel cell with gradient-structured cathode catalyst layers", *Renew. Energy*, c. 253, s. 123627, Kas. 2025, doi: 10.1016/j.renene.2025.123627.

[5] T. Huang *vd.*, "Novel cooling flow field for proton exchange membrane fuel cells based on sidewall lateral interconnection: Design and performance verification of the lateral interlinked channel", *Chem. Eng. J.*, c. 526, s. 170920, Ara. 2025, doi: 10.1016/j.cej.2025.170920.

[6] Y. Chen *vd.*, "Novel microporous layer prepared for proton exchange membrane fuel cells using different nano-cellulose composite carbon black", *J. Power Sources*, c. 665, s. 239024, Şub. 2026, doi: 10.1016/j.jpowsour.2025.239024.

[7] "PEM Fuel Cell Modeling and Simulation Using Matlab", ScienceDirect. Erişim: 15 Aralık 2025. [Çevrimiçi]. Erişim adresi: http://www.sciencedirect.com:5070/book/monograph/9780123742599/pem-fuel-cell-modeling-and-simulation-using-matlab

Homogeneous Charge Compression Ignition (HCCI): A Comprehensive Overview of Performance, Emissions, and Future Prospects

Adem TÜYLÜ[1]

1- Res. Asst.; Sakarya University of Applied Sciences, Faculty of Technology, Department of Mechanical Engineering. ademtuylu@subu.edu.tr, ORCID No: 0000-0001-9828-1573

ABSTRACT

Homogeneous Charge Compression Ignition (HCCI) technology, particularly within the framework of Low Temperature Combustion (LTC) methods, marks a major breakthrough for internal combustion engines. This approach merges the benefits of both Spark-Ignition (SI) and Compression-Ignition (CI) systems, resulting in improved thermal efficiency and extremely low emissions of NO_x and particulate matter (PM) by preventing flame spread and promoting consistent combustion across the cylinder. This characteristic leads to better fuel economy and reduced emissions, especially under partial load conditions. However, despite these benefits, HCCI faces challenges, including difficulty in controlling ignition timing, managing high pressure rise rates, and maintaining stable operation over a narrow range of engine loads. Additionally, elevated emissions of unburned hydrocarbons (HC) and carbon monoxide (CO) at high loads present major obstacles to the broader adoption of this technology. To overcome these issues, various approaches have been examined, including the implementation of exhaust gas recirculation (EGR), adjustments in valve timing, and the exploration of alternative fuels. These methods enhance combustion stability and expand the operational range of HCCI engines. Despite ongoing challenges, HCCI remains a promising solution for cleaner, more efficient combustion, especially with its capability to operate on renewable and alternative fuels, offering a sustainable alternative to conventional combustion systems for future transportation and power generation applications.

Keywords – HCCI, LTC, NO_x, EGR, Alternative Fuels.

INTRODUCTION

Internal combustion engines (ICEs) have remained the dominant power source for transportation and power generation for more than a century due to their robustness, reliability, and relatively high efficiency. However, the progressive depletion of petroleum-based fuel reserves, along with increasing concerns over environmental pollution and human health, has intensified the need for cleaner and more efficient combustion concepts. Conventional diesel engines, while offering high thermal efficiency, suffer from a fundamental trade-off between NO_x and PM emissions, which arises from mixing-controlled diffusion combustion. Although exhaust after-treatment technologies can mitigate these emissions, they inevitably increase system complexity, fuel consumption, and cost. As a result, improving in-cylinder combustion efficiency while simultaneously reducing emissions has become a central research objective in modern engine development (Tüylü, 2021; Tüylü et al., 2024).

In this context, low temperature combustion (LTC) strategies have emerged as promising alternatives to conventional diesel combustion. These strategies aim to suppress locally rich fuel–air regions and high flame temperatures by enhancing mixture homogeneity prior to ignition. Advanced combustion techniques like HCCI, partially premixed charge compression ignition (PCCI), and reactivity-controlled compression ignition (RCCI) have been thoroughly studied to minimize NO_x and soot emissions, achieving these goals without heavily relying on post-treatment technologies. These methods extend ignition delay and encourage better mixture homogenization prior to ignition. By extending the ignition delay and promoting premixing through early fuel injection, intake port injection, or dual-fuel operation, LTC concepts enable lean and low-temperature combustion conditions. Furthermore, the flexibility of these modes allows the utilization of renewable and alternative fuels, including alcohols such as ethanol and methanol, which can be produced from various feedstocks and exhibit favorable environmental characteristics (Calam et al., 2022; H. Duan et al., 2019; Krishnasamy et al., 2021).

Among advanced combustion concepts, HCCI represents a unique synthesis of SI and CI engine principles. In HCCI operation, a well-mixed and lean fuel–air mixture is formed prior to compression, and combustion is initiated by auto-ignition at multiple locations within the cylinder near top dead center. Unlike conventional diesel combustion, HCCI does not involve flame propagation or diffusion-controlled burning; instead, heat release occurs nearly simultaneously throughout the combustion chamber. This characteristic leads to high thermal efficiency and extremely low NO_x and PM emissions, particularly under partial load conditions. Nevertheless, challenges related to combustion phasing control, pressure rise rate, limited operating range, and increased HC and CO emissions remain significant barriers to practical implementation. Consequently, understanding the fundamentals, control strategies, advantages, and limitations of HCCI combustion is essential for advancing next-generation clean and efficient internal combustion engines, which will continue to play a vital role in future transportation systems. To better understand the evolution of combustion modes in internal combustion engines, it is important to recognize the distinct mechanisms of conventional diesel combustion (CDC), PCCI, RCCI and HCCI. Each of these modes offers different strategies for fuel injection and ignition control, with varying impacts on engine performance and emissions (Calam et al., 2022; H. Duan et al., 2019; Krishnasamy et al., 2021; Tüylü, 2021). Figure 1 illustrates the key differences in combustion and fuel injection strategies across these engine types. Within this framework, the present book chapter aims to comprehensively describe the HCCI combustion technology, to comparatively assess this combustion mode in relation to other low-temperature combustion (LTC) concepts, and

to elucidate the performance and emission characteristics associated with HCCI operation.

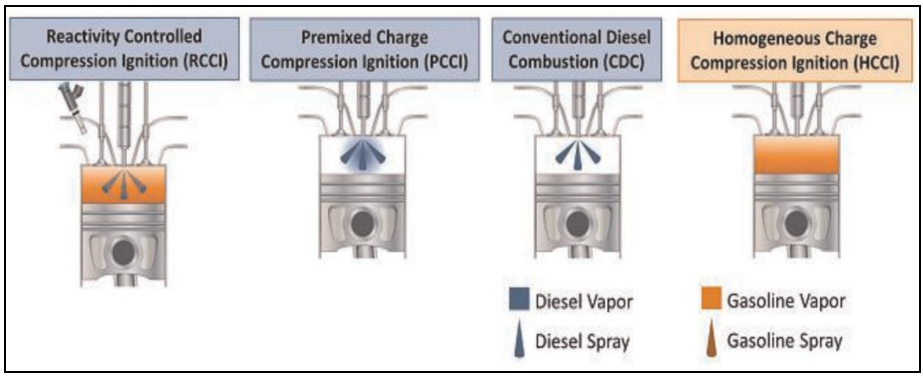


Figure 1: Schematic comparison of diesel LTC strategies and conventional diesel combustion (Krishnasamy et al., 2021)

HCCI COMBUSTION MODE

HCCI combustion is an advanced technology that combines the benefits of both SI and CI engines. In contrast to traditional CI engines, where combustion is initiated by fuel injected into compressed air, HCCI works by premixing the fuel and air before compression, with the mixture auto-igniting during the compression stroke. This results in the absence of flame propagation, reducing NO_x and PM emissions, and providing a highly efficient combustion process. The early efforts to apply HCCI combustion were initiated in the 1970s, with Onishi and colleagues pioneering its use in two-stroke engines. These experiments showed the potential to significantly reduce hydrocarbon emissions and improve fuel economy by utilizing internal residual gases and elevated charge temperatures. This concept was later extended to four-stroke engines, with studies in the 1980s exploring different compression ratios and fuel mixtures. However, despite its advantages in emission reductions and thermal efficiency, the HCCI technology has faced several challenges, particularly in combustion phasing control and its narrow operating range (Pandey et al., 2018).

The operation of HCCI combustion is governed by the thermodynamic properties of the fuel-air mixture, including the compression ratio, intake air temperature, and fuel properties. Key challenges of HCCI include the difficulty in controlling the ignition timing, as there is no spark to initiate combustion like in SI engines or controlled fuel injection like in CI engines. To address these issues, several control mechanisms have been explored, including EGR, variable valve timing (VVT), and the use of alternative fuels. EGR helps lower peak combustion temperatures by diluting the intake charge with residual gases, thus reducing NO_x emissions. VVT, when combined with variable compression ratios (VCR), allows for adjustments in cylinder pressure and temperature, facilitating better ignition timing control. Additionally, techniques like early or late direct injection are employed to improve fuel-air mixing, allowing for better control of combustion, especially under varying load conditions. These methods provide more flexibility in operating the HCCI engine over a broader range of conditions (Kumar & Rehman, 2016; Pandey et al., 2018; Parthasarathy et al., 2020; Polat et al., 2020).

As HCCI combustion continues to evolve, several generalized concepts and systems have been developed to further optimize the engine's performance and emissions. Figure 2 presents various generalized concepts of HCCI combustion, including Early Direct Injection HCCI, Late Direct Injection HCCI, and Premixed/Direct Injection HCCI. These systems are further divided into specific sub-concepts like Stratified Charge Compression Ignition (SCCI), Low Temperature Combustion (LTC), and External Mixture Preparation HCCI. Each strategy has been developed to tackle particular operational challenges, such as controlling the combustion process, optimizing ignition timing, and managing emission levels. By applying these generalized HCCI combustion concepts, researchers aim to achieve more efficient, cleaner, and flexible engine operations (Kumar & Rehman, 2016; Pandey et al., 2018; Parthasarathy et al., 2020; Polat et al., 2020).

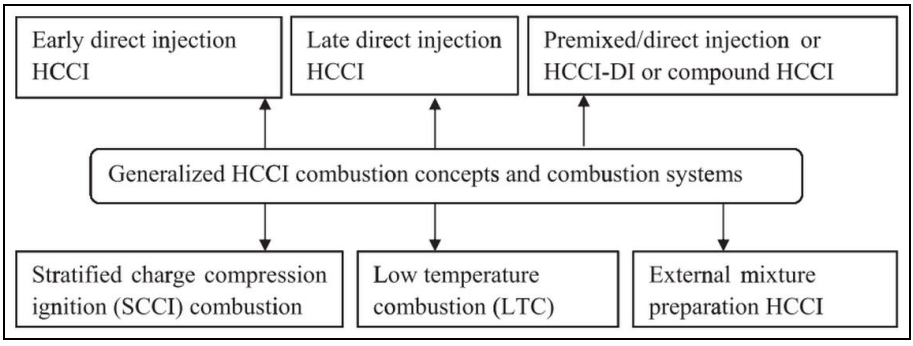


Figure 2: HCCI combustion concepts and combustion systems

HCCI vs. OTHER LTC COMBUSTION MODES

HCCI is an advanced combustion technology that provides considerable benefits in terms of emissions reduction and thermal efficiency. Unlike conventional engines, which rely on either spark ignition or compression ignition, HCCI operates by blending the fuel and air prior to entering the combustion chamber, where the mixture ignites spontaneously due to the elevated temperature and pressure during compression. This eliminates the need for spark plugs or direct fuel injection, which are traditionally used in SI and CI engines, respectively. The homogeneous mixture in HCCI results in more uniform combustion, leading to significantly lower peak combustion temperatures. This, in turn, reduces the formation of NO_x and PM, two of the most harmful pollutants emitted by conventional engines. Despite these advantages, HCCI combustion is not without its challenges. The primary issue is its narrow operating range. HCCI engines typically operate well at light to moderate loads but struggle under high load conditions, where controlling combustion phasing becomes complex. At high loads, achieving a stable combustion process can be difficult because there is no direct means to control ignition timing, as in SI engines, or fuel injection strategies, as in diesel engines. As a result, high pressure rise rates can occur, leading to knock or misfire, particularly at lower engine speeds and loads. Thus, while HCCI holds promise for significant emission reductions, its practical application is still limited by these operational constraints (Murugesa Pandian & Anand, 2018; Shim et al., 2020).

In contrast, PCCI shares similarities with HCCI in that it also uses a premixed fuel-air mixture for combustion, but with greater flexibility in controlling the ignition timing. PCCI achieves this by adjusting the timing of fuel injection, allowing it to manage the combustion process more precisely. This gives PCCI the ability to operate effectively across a wider range of engine loads compared to HCCI. By controlling the injection timing, PCCI engines can optimize the timing of ignition, leading to better control over combustion and allowing the engine to handle higher loads more efficiently. However, while PCCI offers improved flexibility, it is not without its own drawbacks. One of the challenges of PCCI is that the fuel-air mixture may not always be as homogeneous as in HCCI, particularly under high-load conditions. This may result in the formation of richer mixtures in localized areas within the combustion chamber, which may result in increased smoke emissions compared to HCCI. The tendency for localized combustion in these richer regions can lead to the production of soot and unburned hydrocarbons, especially if the fuel-air mixture is not optimally mixed. Nonetheless, PCCI represents a significant improvement over traditional diesel combustion, as it reduces the formation of NO_x and PM and the ability to adjust injection timing provides more control over emissions and

combustion performance across varying operating conditions (Murugesa Pandian & Anand, 2018; Shim et al., 2020).

RCCI offers further advancements by incorporating a dual-fuel system, typically combining a low-reactivity fuel such as gasoline with a high-reactivity fuel like diesel. This dual-fuel configuration allows for greater control over ignition timing and combustion phasing by adjusting the proportion of each fuel in the mixture. RCCI offers significant advantages in achieving excellent thermal efficiency while keeping NO_x and PM emissions at very low levels. By utilizing a high-reactivity fuel (like diesel) for ignition and a low-reactivity fuel (such as gasoline) to minimize NO_x and soot formation, this approach optimizes the combustion process. By optimizing the ratio of these fuels and controlling their injection timing, RCCI can operate effectively across a broader range of engine loads, achieving higher brake thermal efficiency and reducing the environmental impact. The dual-fuel setup also helps mitigate some of the challenges faced by both HCCI and PCCI, particularly in terms of combustion stability and load flexibility. Despite its benefits, RCCI comes with its own set of challenges, including the difficulty of managing two distinct fuels and the requirement for accurate control over fuel injection techniques. However, these complexities aside, RCCI stands out as one of the most promising combustion strategies for achieving both high efficiency and low emissions under various engine operating conditions, making it a flexible and sustainable choice for the future of engine technology. Table 1 below summarizes the key performance and emission characteristics of HCCI, PCCI, and RCCI, offering a direct comparison of their features, including fuel requirements, ignition control mechanisms, combustion temperature, thermal efficiency, and emission behavior. This table effectively highlights the distinctions and advantages of each combustion mode, aiding in the understanding of their operational ranges and suitability for various engine designs and operating conditions (Calam et al., 2022; Murugesa Pandian & Anand, 2018; Shim et al., 2020; Singh & Agarwal, 2018).

Table 1: Comparison of HCCI, PCCI, and RCCI Combustion Strategies (Calam et al., 2022; Singh & Agarwal, 2018)

Combustion Strategies	HCCI	PCCI	RCCI
Fuel Type	Blend of liquid or gaseous fuels	Blend of liquid or gaseous fuels	PFI of high octane fuel and DI of high cetane fuel
Injection Type	Port or early direct injection	Port or early direct injection	Port and direct injection
Injection Pressure	Low port injection or high direct injection pressure	Low port injection pressure or high direct injection pressure	Low port injection pressure and high direct injection pressure
Air/Fuel Mixture	Premixed	Partially premixed	Partially premixed
Aspiration	With or without boosted	With or without boosted	Boosted
Ignition Type	Auto ignition	Compression ignited	Compression ignited
Ignition Control	Temperature and pressure history	Injection timing	Injection timing
Power Output Control	Fuel flow control with lean air-fuel ratio	Fuel flow control with lean air-fuel ratio or charge dilution	Fuel reactivity stratification, air-fuel ratio stratification
Combustion Temperature	Low temperature	Low temperature	Low temperature
Thermal Efficiency	High	High	High
Mechanism Controlling the Fuel Burning Rate	Chemical kinetics	Chemical kinetics injection timing	Chemical kinetics and fuel reactivity
Emission Behaviour	Extremely low NO_x, soot, and CO2, high HC and CO	Extraordinarily low NO_x, soot and CO2, high HC and CO	Comparatively lower NO_x and PM, high HC and CO

Figure 3 illustrates the operating zones of various advanced combustion modes on the Phi-T diagram, comparing SI, CI, PCCI, RCCI, and HCCI. It reveals the differences in air-fuel ratio, ignition timing, ignition delay, flame propagation characteristics, and emissions across these modes. The diagram highlights the specific zones of each mode in terms of equivalence ratio and local temperature. HCCI functions within a range that reduces soot and NO_x emissions, all while achieving thermal efficiency that matches or even surpasses that of traditional diesel combustion. However, despite these advantages, HCCI still faces several technical challenges, including precise ignition timing control, homogeneous charge preparation, high pressure rise rates, knocking, and cold starts. These hurdles must be

overcome for HCCI to be more widely implemented and fully realize its potential benefits (X. Duan et al., 2021).

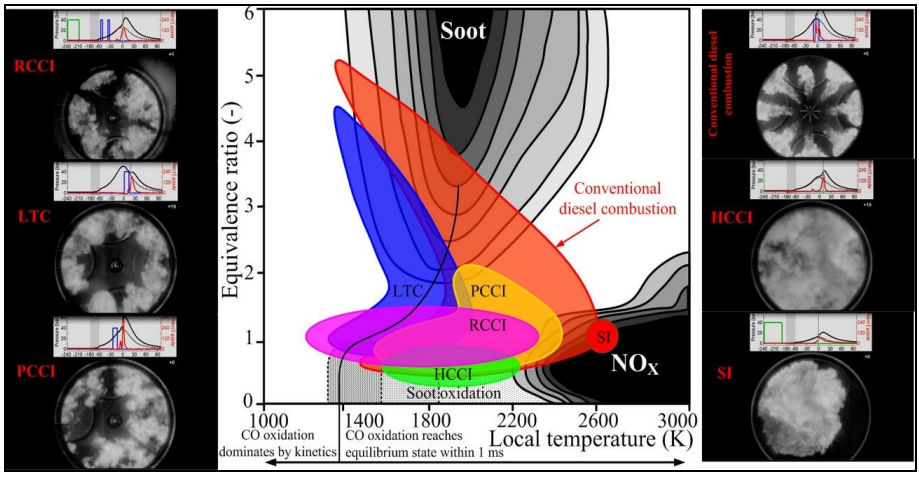


Figure 3: The operating regions of various combustion modes on the Phi-T diagram (X. Duan et al., 2021)

PERFORMANCE AND EMISSIONS OF HCCI COMBUSTION

HCCI combustion technology combines the advantages of conventional combustion systems, offering notable advancements in engine performance and emission reductions. In HCCI engines, combustion occurs without the formation of a flame front, as the entire air-fuel mixture burns simultaneously and slowly. This process is characterized by homogeneous air-fuel mixtures and spontaneous ignition across multiple points within the cylinder. As a result, HCCI engines exhibit lower NO_x and soot emissions in comparison with traditional diesel engines. When fueled by alternative sources like biodiesel (BDFs) or n-butanol, HCCI engines show even more significant improvements in emission characteristics. For instance, BDF-powered HCCI engines have demonstrated a 65–67% reduction in smoke and a 98–99% reduction in NO_x emissions at 80% load. However, these engines also face challenges, such as an increase in HC emissions, which are approximately 11 times higher at the same load. Similarly, n-butanol-fueled

HCCI engines maintain ultra-low NO_x and soot emissions without requiring EGR at low-to-mid engine loads. Despite these benefits, optimizing combustion phasing, particularly under full-load conditions, remains a significant challenge. In addition, maintaining a homogeneous air-fuel mixture and controlling CO and HC emissions, especially at higher engine loads, require careful calibration of engine parameters. Nevertheless, HCCI technology offers the potential for higher fuel efficiency and a broader operating range compared to conventional diesel engines. With the ability to operate on a wide range of fuels, including renewable and biodegradable options, HCCI represents a promising advancement in internal combustion engine technology, offering improved performance and reduced emissions (Aydoğan, 2020; Khandal et al., 2019; Tüylü, 2021; Zheng et al., 2015).

HCCI engines provide superior thermal efficiency while significantly reducing NO_x and PM emissions. However, they are associated with increased HC and CO emissions. The ability to operate with leaner air-fuel mixtures results in higher fuel efficiency, but incomplete combustion, due to flame extinction in the cylinder from insufficient oxygen and lower combustion temperatures, leads to the formation of unburned fuel molecules and thus higher HC emissions. Similarly, CO emissions arise from incomplete combustion because of insufficient oxygen during the combustion process. Compared to conventional diesel combustion, HCCI engines achieve more efficient energy conversion, significantly reducing soot and NO_x emissions. A homogeneous air-fuel mixture is formed, and as the cylinder temperature increases during compression, spontaneous ignition occurs, leading to lower NO_x emissions while maintaining high thermal efficiency. Despite these benefits, HCCI engines face challenges such as limited operating range, difficulties in controlling ignition timing, and elevated CO and HC emissions (Çelebi et al., 2021; Verma et al., 2022).

Studies on the effects of internal and external EGR have shown that internal EGR increases NO_x emissions from 30 ppm to 150 ppm, while external EGR reduces NO_x emissions from 150 ppm to as low as 10 ppm. These findings indicate that EGR is an effective method for controlling NO_x emissions in HCCI engines. Moreover, adjustments to the air-fuel ratio and EGR rate have been shown to control HC and CO emissions. In low-load conditions, HCCI engines exhibit significantly lower HC emissions compared to conventional engines, which increase HC emissions at higher loads. Thus, while HCCI technology offers a promising alternative with lower NO_x and particulate emissions, managing CO and HC emissions remains a key challenge for its widespread application (Verma et al., 2022).

ADVANTAGES AND DISADVANTAGES OF HCCI COMBUSTION

Despite these challenges, HCCI combustion offers numerous advantages over traditional internal combustion engine technologies. One of its primary benefits is the reduction of harmful emissions, specifically NO_x and soot, which are significantly lower than those in conventional diesel engines. Additionally, HCCI engines can operate at high compression ratios, which enhances their efficiency and power output. The flexibility in fuel choice is another advantage, as HCCI engines can run on a wide variety of fuels, including biodiesel and other renewable energy sources, making them more adaptable to future fuel innovations. The combustion process in HCCI engines is also marked by stable operation with minimal variation from cycle to cycle, leading to improved consistency and performance. Another key advantage is that HCCI engines do not require high-pressure fuel injection systems, which reduces the overall cost and complexity of the engine design. Finally, HCCI technology offers high thermal efficiency, making it an attractive option for both commercial and environmental purposes. In conclusion, while HCCI combustion technology offers considerable advantages, particularly in terms of emission reductions and fuel efficiency, it still faces significant hurdles in terms of control and operational range. Advances in fuel technology and engine design are crucial to overcoming these limitations and fully realizing the potential of HCCI combustion (Calam et al., 2022; X. Duan et al., 2021; Tüylü, 2021).

HCCI combustion technology, while promising, faces several challenges that impact its practical implementation. One major issue is the complexity of preparing a homogeneous air-fuel mixture, which is essential for optimal combustion. Additionally, controlling the controlled auto-ignition (CAI) process and managing the overall combustion dynamics can be difficult, especially under varying engine loads. The high pressure rise rates observed in HCCI engines, particularly during full-load operation, can lead to issues such as knocking and noise, limiting the engine's efficiency and lifespan. Furthermore, the limited operational range of HCCI engines, especially at high loads, is a significant drawback. Cold-starting is another challenge, as the engine may not operate optimally until it reaches the necessary temperature for spontaneous ignition. These issues can be addressed with expensive solutions, such as intake air heating systems, EGR, and fine-tuning compression ratios. However, a more economical approach involves the controlled use of alternative fuels with specific physical and chemical properties, which can alleviate these problems without incurring high costs (Calam et al., 2022; X. Duan et al., 2021; Tüylü, 2021).

CONCLUSION

Recent advancements in HCCI technology have brought about significant improvements in engine performance and emission reductions, combining the benefits of both spark ignition and compression ignition systems. HCCI engines can deliver high thermal efficiency and extremely low NO_x and particulate emissions, especially during partial load operation, thanks to the elimination of flame propagation and the uniform combustion process. However, despite these benefits, HCCI combustion remains challenging due to difficulties in controlling combustion phasing, managing pressure rise rates, and achieving stable operation across a broad range of engine loads. Moreover, the rise in unburned HC and CO emissions, especially under higher load conditions, poses a major challenge to the broader implementation of HCCI technology.

To address these challenges, several strategies have been investigated, such as EGR, VVT, and the use of alternative fuels with customized physical and chemical characteristics. While these methods show promise in addressing emission and performance issues, further optimization of engine parameters and combustion control mechanisms is necessary for the practical implementation of HCCI engines. The ability to operate on a wide range of fuels, including renewable and biodegradable options, further enhances the potential of HCCI technology as a sustainable and efficient alternative to conventional combustion systems. As research progresses, developing more efficient combustion control methods and improving fuel management strategies will be essential to unlock the full capabilities of HCCI engines in future transportation and power generation sectors.

REFERENCES

Aydoğan, B. (2020). An experimental examination of the effects of n-hexane and n-heptane fuel blends on combustion, performance and emissions characteristics in a HCCI engine. Energy, 192. https://doi.org/10.1016/j.energy.2019.116600

Calam, A., Halis, S., Aydoğan, B., & Haşimoğlu, C. (2022). Combustion characteristics of naphtha and n-heptane fuels in an auto-ignited HCCI engine at different lambda values and engine loads. Fuel, 327(June). https://doi.org/10.1016/j.fuel.2022.125183

Çelebi, S., Haşimoğlu, C., Uyumaz, A., Halis, S., Calam, A., Solmaz, H., & Yılmaz,

E. (2021). Operating range, combustion, performance and emissions of an HCCI engine fueled with naphtha. Fuel, 283(August 2020). https://doi.org/10.1016/j.fuel.2020.118828

Duan, H., Jia, M., Chang, Y., & Liu, H. (2019). Experimental study on the influence of low-temperature combustion (LTC) mode and fuel properties on cyclic variations in a compression-ignition engine. Fuel, 256(August), 115907. https://doi.org/10.1016/j.fuel.2019.115907

Duan, X., Lai, M. C., Jansons, M., Guo, G., & Liu, J. (2021). A review of controlling strategies of the ignition timing and combustion phase in homogeneous charge compression ignition (HCCI) engine. Fuel, 285(September 2020). https://doi.org/10.1016/j.fuel.2020.119142

Khandal, S. V., Banapurmath, N. R., & Gaitonde, V. N. (2019). Performance studies on homogeneous charge compression ignition (HCCI) engine powered with alternative fuels. Renewable Energy, 132(x), 683–693. https://doi.org/10.1016/j.renene.2018.08.035

Krishnasamy, A., Gupta, S. K., & Reitz, R. D. (2021). Prospective fuels for diesel low temperature combustion engine applications: A critical review. International Journal of Engine Research, 22(7), 2071–2106. https://doi.org/10.1177/1468087420960857

Kumar, P., & Rehman, A. (2016). Bio-diesel in homogeneous charge compression ignition (HCCI) combustion. Renewable and Sustainable Energy Reviews, 56, 536–550. https://doi.org/10.1016/j.rser. 2015.11.088

Murugesa Pandian, M., & Anand, K. (2018). Comparison of different low temperature combustion strategies in a light duty air cooled diesel engine. Applied Thermal Engineering, 142(July), 380–390. https://doi.org/10.1016/j.applthermaleng.2018.07.047

Pandey, S., Diwan, P., Sahoo, P. K., & Thipse, S. S. (2018). A review of combustion control strategies in diesel HCCI engines. Biofuels, 9(1), 61–74. https://doi.org/10.1080/17597269.2016.1257315

Parthasarathy, M., Ramkumar, S., Elumalai, P. V, & N, M. N. (2020). Control Strategies on HCCI Engine Performance and Emission characteristics by Combined Effect of Exhaust Gas Recirculation with Blend of Biodiesel and N-Heptane. Energy Sources, Part A: Recovery, Utilization, And Environmental Effects https://doi.org/10.1080/15567036.2020.1850924

Polat, S., Yücesu, H. S., Uyumaz, A., Kannan, K., & Shahbakhti, M. (2020). An experimental investigation on combustion and performance characteristics of supercharged HCCI operation in low compression ratio engine setting. Applied Thermal Engineering, 180(January). https://doi.org/10.1016/j.applthermaleng.2020.115858

Shim, E., Park, H., & Bae, C. (2020). Comparisons of advanced combustion technologies (HCCI, PCCI, and dual-fuel PCCI) on engine performance and emission characteristics in a heavy-duty diesel engine. Fuel, 262 (October2019), 116436. https://doi.org/10.1016/j.fuel. 2019.116436

Singh, A. P., & Agarwal, A. K. (2018). Low Temperature Combustion Engines and Mode Switching Strategies – A review. Journal of Energy and Environmental Sustainability 6, 42–47.

Tüylü, A. (2021). Kismi Homojen Dolgulu Direkt Enjeksiyonlu Bir Dizel Motorunun Deneysel Ve Numerik Olarak İncelenmesi, Master's Thesis.

Tüylü, A., Kolip, A., & Çelik, H. A. (2024). Etil Alkol Yakıtlı HCCI Motorun

Performans, NO ve İs Emisyon Özelliklerinin Deneysel Olarak İncelenmesi. Düzce Üniversitesi Bilim ve Teknoloji Dergisi, 12(2), 1036–1049. https://doi.org/10.29130/dubited.1184716

Verma, S. K., Gaur, S., Akram, T., Gautam, S., & Kumar, A. (2022). Emissions from homogeneous charge compression ignition (HCCI) engine using different fuels: a review. Environmental Science and Pollution Research, 29(34), 50960–50969. https://doi.org/10.1007/s11356-021-15602-x

Zheng, M., Han, X., Asad, U., & Wang, J. (2015). Investigation of butanol-fuelled HCCI combustion on a high efficiency diesel engine. Energy Conversion and Management, 98, 215–224. https://doi.org/10.1016/j.enconman.2015.03.098

Effects of Air Excess Ratio on Emissions in SI and CI Engines

Adem TÜYLÜ[1]

1- Res. Asst.; Sakarya University of Applied Sciences, Faculty of Technology, Department of Mechanical Engineering. ademtuylu@subu.edu.tr, ORCID No: 0000-0001-9828-1573

ABSTRACT

The excess air coefficient (λ) is a fundamental control parameter in internal combustion engines, as it governs mixture composition, combustion temperature, and oxidation kinetics, thereby exerting a decisive influence on engine performance and exhaust emission formation. This book chapter provides a comprehensive overview of the role of λ in regulating emissions in both spark-ignition (SI) and compression-ignition (CI) engines under conventional and alternative fuel operating conditions. First, the underlying physical and chemical mechanisms through which variations in λ affect the formation of major regulated pollutants, including nitrogen oxides (NO_x), carbon monoxide (CO), unburned hydrocarbons (HC), and particulate matter (PM), are discussed within a general combustion framework. The distinct sensitivities of SI and CI engines to mixture dilution are highlighted, emphasizing the trade-offs between combustion efficiency and emission reduction across different λ regimes. Subsequently, recent experimental and numerical studies are reviewed to illustrate how changes in λ influence combustion phasing, flame development, heat release behavior, and emission trends in a wide range of engine concepts, including lean-burn, dual-fuel, hydrogen-fueled, and ammonia-assisted systems. Particular attention is given to the identification of optimal λ ranges that balance thermal efficiency, combustion stability, and emission control. Overall, the chapter demonstrates that precise regulation of the excess air coefficient, often in combination with complementary in-cylinder strategies and after-treatment technologies, remains a key pathway for meeting increasingly stringent emission regulations while maintaining efficient engine operation.

Keywords – Excess Air Coefficient, SI, CI, Exhaust Emissions, NO_x.

INTRODUCTION

The excess air coefficient (λ) constitutes a key control parameter in internal combustion engines, as it directly determines combustion quality, thermal efficiency, and the formation of exhaust emissions. Defined as the ratio between the actual air supplied to the combustion chamber and the theoretical air demand required for complete stoichiometric combustion, λ provides a normalized measure of mixture composition that is independent of fuel type. Variations in λ modify oxygen availability, in-cylinder temperature evolution, and chemical reaction rates, thereby exerting a pronounced influence on both engine performance and pollutant generation. In the context of increasingly stringent emission legislation and growing environmental awareness, accurate regulation of λ has become a central objective of modern engine control strategies, particularly for minimizing regulated emissions such as CO, HC, and NO_x without compromising

efficiency (Deng et al., 2019; Shamil Ahmed Flamarz Al-Arkawazi, 2019; Sinkala et al., 2024; Zervas et al., 2003).

The role of the excess air coefficient differs fundamentally between SI and CI engines due to inherent differences in mixture preparation and ignition mechanisms. In SI engines, air and fuel are largely premixed prior to combustion, and ignition is initiated by a spark, which constrains stable operation to a narrow λ range near stoichiometric conditions. This operating regime is essential for maintaining combustion stability and achieving high conversion efficiency in three-way catalytic aftertreatment systems. In contrast, CI engines employ direct fuel injection into highly compressed air, resulting in spatially heterogeneous mixtures and predominantly lean global operation. Although this excess-air environment suppresses CO and HC emissions, it gives rise to a characteristic trade-off between nitrogen oxides and particulate matter as λ varies across the operating range. Consequently, the influence of λ on emissions and performance is strongly engine-type dependent, necessitating distinct optimization approaches for SI and CI combustion systems (Deng et al., 2019; Shamil Ahmed Flamarz Al-Arkawazi, 2019; Sinkala et al., 2024; Zervas et al., 2003).

On the other hand, strict emission regulations are imposed on both spark ignition and compression ignition engines, shaping vehicle emission control strategies worldwide. As illustrated in Figure 1, nitrogen oxide emissions are used as a representative indicator to exemplify the increasing stringency of Euro emission regulations for passenger vehicles with SI and CI engines. Beyond Europe, several countries have adopted vehicle emission standards that closely follow the regulatory structure and stringency of the Euro framework, leading to its widespread global influence. As shown in Figure 2, nearly 70% of global light-duty and heavy-duty vehicle sales are currently subject to Euro 6/VI standards or their previous iterations. Although the Euro 6/VI regulations were introduced nearly a decade ago and testing procedures have evolved since then, the emission limit values themselves have largely remained unchanged. In response, the forthcoming Euro 7/VII framework models the implementation of stricter NO_x limitations from 2027 onward, applying to all new light-duty and heavy-duty vehicles, while assuming zero-emission vehicle uptake consistent with the baseline scenario. This evolution further reinforces the central role of Euro standards in guiding future emission reduction pathways (Mulholland et al., 2022; Rodríguez et al., 2019).

Recent studies in the literature indicate that variations in the excess air coefficient exert a significant influence on combustion behavior and exhaust emission formation across different engine concepts and operating conditions. Changes in λ affect in-cylinder thermodynamic processes, combustion characteristics, and the formation trends of regulated pollutants, and these effects are often interrelated with other control parameters. Accordingly, the excess air coefficient is widely recognized as a central

variable in modern internal combustion engine research. In this context, both SI and CI engines show pronounced sensitivity of exhaust emissions to mixture composition, as the excess air ratio governs oxygen availability, combustion temperature, and reaction pathways. Consequently, the regulation of λ, often in combination with complementary in-cylinder measures and after-treatment systems, is widely regarded as a key approach for managing emission behavior in SI and CI engine technologies (Fayyazbakhsh et al., 2022; Gong et al., 2021; Shu et al., 2019).

In this book chapter, the effects of λ on exhaust emissions in spark ignition and compression ignition engines are first discussed and explained in a general framework, followed by an overview of recent studies investigating the impact of varying λ on engine performance and emission characteristics, thereby providing a structured basis for the detailed analyses presented in the subsequent sections.

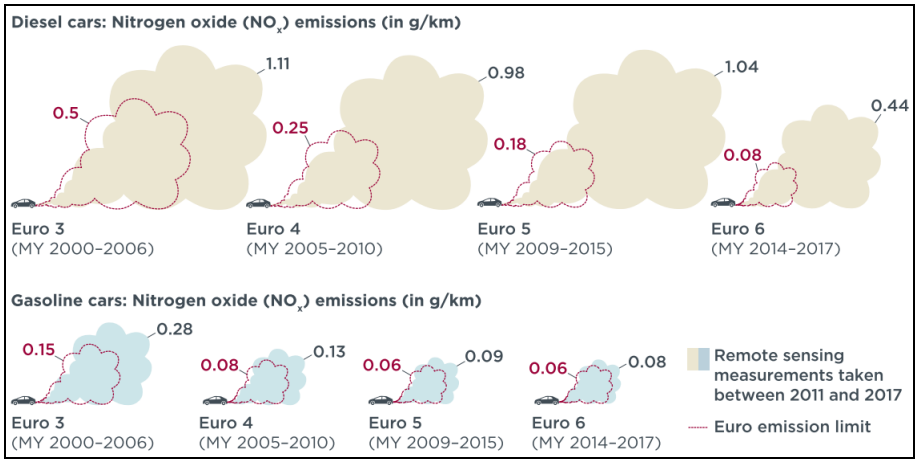


Figure 1: NO_x emissions from Euro 3 to Euro 6 for EU passenger vehicles (Rodríguez et al., 2019)

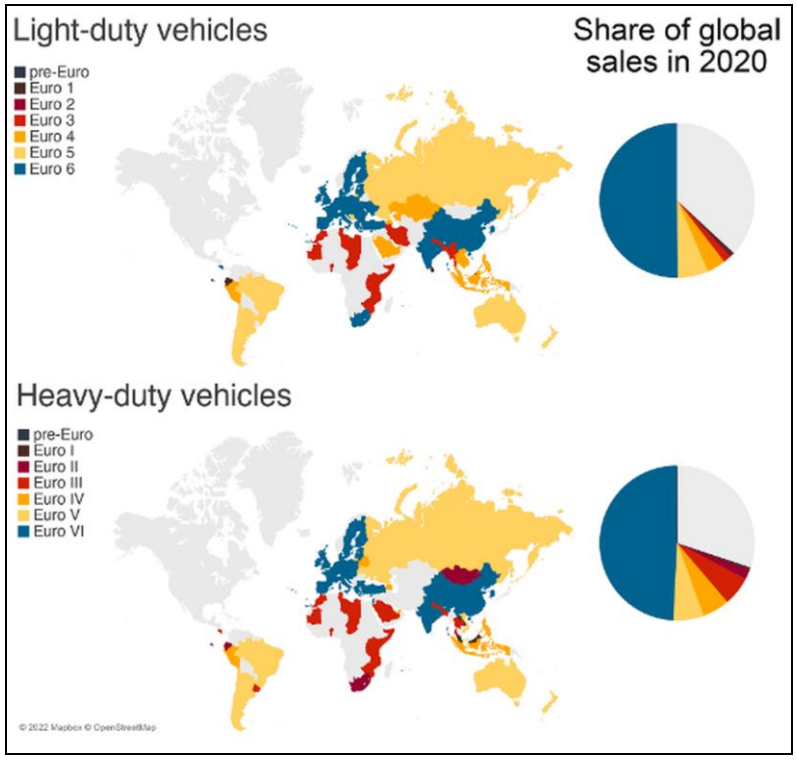


Figure 2: Countries applying Euro-based emission standards for light- and heavy-duty vehicles (Mulholland et al., 2022)

EFFECT OF THE EXCESS AIR COEFFICIENT ON EMISSIONS IN SI ENGINES

SI engines are a major source of regulated exhaust emissions, including NO_x, CO, and HC which originate primarily from incomplete or non-ideal combustion processes. The formation of these pollutants is strongly governed by mixture composition and in-cylinder combustion conditions, making the λ a key controlling parameter. As illustrated in Figure 3, under rich mixture conditions ($\lambda < 1.0$), limited oxygen availability restricts complete oxidation reactions, resulting in high CO emissions and

elevated HC levels due to incomplete fuel burnout and locally quenched flame zones. As λ increases toward stoichiometric conditions, improved oxidation efficiency leads to a sharp reduction in CO emissions, while HC concentrations also decrease as flame propagation becomes more stable and combustion completeness improves. In contrast, NOx emissions increase with rising λ and reach a maximum around λ ≈ 1.1, where sufficiently high combustion temperatures coincide with adequate oxygen availability, promoting thermal NO_x formation. Beyond this point, further increases in the excess air coefficient reduce combustion temperatures due to excess dilution, leading to a decline in NO_x emissions. Meanwhile, HC emissions tend to increase again under lean conditions as flame quenching, slower reaction rates, and partial misfire phenomena become more pronounced. These emission trends demonstrate the inherent trade-offs among regulated pollutants in SI engines and highlight the critical importance of precise λ control to balance emission reduction and combustion efficiency under increasingly stringent regulatory requirements (Majidi-Jirandehi et al., 2021).

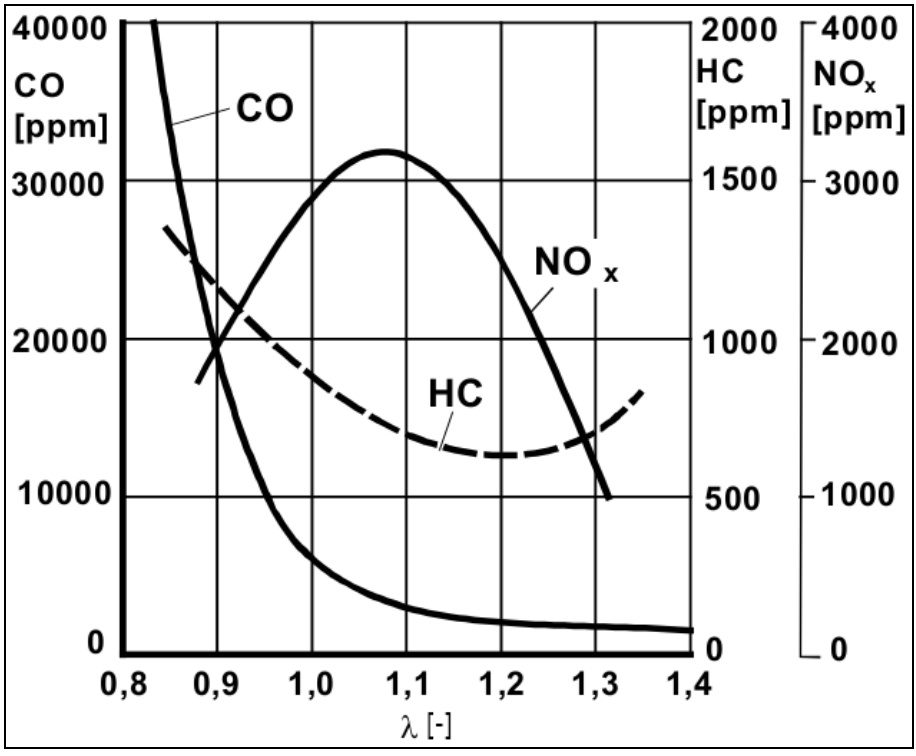


Figure 3: Variation of SI engine emissions depending on the excess air coefficient (IMKO, n.d.)

EFFECT OF THE EXCESS AIR COEFFICIENT ON EMISSIONS IN CI ENGINES

In CI engines, exhaust emission formation is strongly governed by the excess air coefficient (λ), which controls local mixture richness, oxygen availability, combustion temperature, and reaction residence time within the cylinder. As shown in Figure 4, CO emissions are relatively high at low excess air levels (λ ≈ 1–1.5) due to locally fuel-rich spray regions and limited oxygen availability that hinder the complete oxidation of carbon to CO2. As λ increases toward approximately 2, improved air entrainment and oxidation conditions promote more complete combustion, resulting in a minimum CO concentration. However, for λ values greater than 2, CO emissions rise again as highly diluted mixtures lead to lower combustion temperatures and slower oxidation kinetics, limiting the conversion of CO to CO2. HC emissions are lowest near λ ≈ 1, increase substantially in the range of λ ≈ 1–2, and continue to rise monotonically between λ ≈ 2 and 5. This behavior is primarily associated with prolonged ignition delay, reduced flame propagation, and incomplete fuel oxidation under increasingly lean conditions, where combustion may not fully complete within the available expansion time. NOx emissions reach their maximum levels at λ ≈ 1.1, where elevated in-cylinder temperatures coincide with sufficient oxygen concentration to favor thermal NOx formation. With further increases in λ, excess air dilution reduces peak combustion temperatures and shortens the effective high-temperature residence time, leading to a continuous decrease in NOx emissions. PM emissions are highest near λ ≈ 1–1.1 due to the dominance of locally rich combustion zones that promote soot nucleation and growth. A pronounced reduction in PM emissions occurs as λ increases from approximately 1.1 to 2, reflecting enhanced oxygen availability and intensified soot oxidation. For λ values between 2 and 5, PM emissions remain low and relatively stable, indicating the suppression of fuel-rich regions responsible for soot formation. Overall, Figure 3 illustrates that while increasing λ in CI engines is highly effective in reducing NOx and PM emissions, it simultaneously promotes HC formation and results in a non-monotonic CO response, revealing the fundamental emission trade-offs inherent to lean combustion operation (Reşitoğlu et al., 2015).

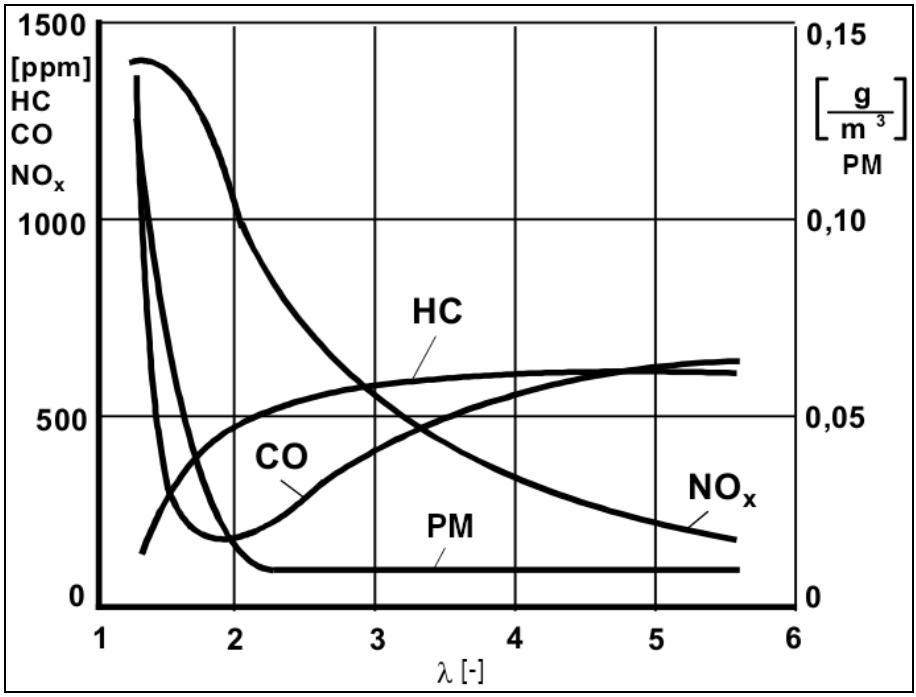


Figure 4: Variation of CI engine emissions depending on the excess air coefficient (IMKO, n.d.)

RECENT STUDIES ON EXCESS AIR COEFFICIENT IN INTERNAL COMBUSTION ENGINES

In this study (Shu et al., 2019), the influence of λ on the combustion and emission characteristics of a diesel pilot ignition natural gas engine was investigated using a CFD model coupled with a reduced chemical kinetic mechanism, which was validated against experimental data. The results showed that increasing λ significantly affects combustion phasing, heat release behavior, and in-cylinder pressure, with distinct responses under partial and full load conditions. From an emission perspective, NO_x emissions increased as λ rose from 1.0 to approximately 1.1 but decreased at higher λ values due to combustion temperature reduction, while unburned methane emissions exhibited a minimum followed by a sharp increase under

highly lean operation. Carbon monoxide emissions decreased with increasing λ up to moderate lean conditions and then showed limited variation. Overall, the findings confirm that λ plays a key role in governing combustion efficiency and emission trade-offs in lean-burn CI-based engine concepts.

Xin et al. investigated the role of λ in an ammonia–hydrogen fueled Miller-cycle spark-ignition engine, with a particular focus on combustion behavior, engine performance, and emission characteristics under both stoichiometric and lean-burn conditions. The results demonstrated that λ plays a decisive role in determining engine stability and efficiency, as performance gains achieved by increasing the ammonia energy share under near-stoichiometric operation diminished once λ exceeded approximately 1.4. At higher λ values, combustion dilution reduced exhaust temperatures and increased the tendency toward combustion instability, with misfire observed when the ammonia fraction exceeded a critical threshold. From an emissions perspective, lean operation combined with ammonia–hydrogen fueling led to elevated NOx emissions, indicating the necessity of exhaust aftertreatment solutions such as selective catalytic reduction. Overall, the study highlights the importance of carefully optimizing λ to balance efficiency, combustion stability, and emissions in ammonia-based SI engine applications (Xin et al., 2023).

Meng et al. experimentally examined the influence of λ on combustion behavior, performance, and emission characteristics of a hydrogen-fueled rotary engine under different load conditions. The results showed that reducing λ from lean to moderately lean conditions leads to a non-monotonic response in engine performance, with both work capacity and brake thermal efficiency first increasing and then decreasing. An optimal operating point was identified at $\lambda \approx 1.4$, where the highest brake thermal efficiency was achieved alongside favorable emission levels. Combustion analysis revealed that decreasing λ shortens the flame development period and advances the combustion phasing, while the flame propagation duration exhibits a minimum near the optimal λ value. From an emissions perspective, both HC and NO emissions initially increase and then decrease as λ is reduced, with NO emissions rising sharply near stoichiometric conditions due to elevated combustion temperatures. Overall, the study demonstrates that λ plays a critical role in balancing combustion stability, thermal efficiency, and emission formation in hydrogen-fueled rotary engines, highlighting lean combustion as an effective strategy for achieving efficient and low-emission operation (Meng et al., 2021).

Wang et al. investigated the effects of λ on combustion and emission characteristics of an ammonia–hydrogen fueled Wankel rotary engine using a three-dimensional numerical model validated by experimental data. The results showed that λ strongly governs flame propagation, in-cylinder pressure development, and NO_x formation. Reducing λ from 1.8 to 1.4

accelerated flame propagation, increased peak pressure and heat release rate, and significantly improved indicated mean effective pressure, while NOx emissions increased by approximately 47.4%. When λ exceeded 1.4, enhanced mixture reactivity promoted higher combustion intensity, whereas at $\lambda \leq 1.2$, excessively rapid flame propagation shortened combustion duration and reduced power output. From an emission perspective, NO_x formation exhibited a non-monotonic behavior with decreasing λ, reaching a minimum near $\lambda = 1.0$ due to the altered dominance of nitrogen- and hydrogen-related reaction pathways. Overall, the study identified $\lambda \approx 1.4$ as an optimal operating point that balances power performance and emission formation in ammonia–hydrogen rotary engine operation (Wang et al., 2024).

CONCLUSION

The excess air coefficient (λ) has been shown to be a governing parameter in controlling combustion behavior and exhaust emission formation in both spark-ignition and compression-ignition engines. As discussed throughout this chapter, variations in λ directly influence oxygen availability, in-cylinder temperature evolution, reaction kinetics, and combustion completeness, thereby shaping the formation trends of major regulated pollutants. In SI engines, operation near stoichiometric conditions remains essential for ensuring combustion stability and effective three-way catalyst performance, while deviations toward rich or lean mixtures introduce clear trade-offs between NO_x, CO, and HC emissions. In contrast, CI engines inherently operate under globally lean conditions, where increasing λ effectively suppresses NOx and particulate matter formation but promotes hydrocarbon emissions and non-monotonic CO behavior due to combustion dilution effects.

The reviewed experimental and numerical studies consistently demonstrate that neither excessively low nor excessively high λ values are favorable for achieving an optimal balance between efficiency, stability, and emission control. Instead, intermediate λ ranges often emerge as optimal operating points, particularly in advanced combustion concepts such as lean-burn, hydrogen-fueled, and ammonia-assisted engines. These findings highlight that precise λ regulation, supported by advanced control strategies and appropriate after-treatment technologies, is indispensable for meeting

increasingly stringent emission regulations. Overall, the excess air coefficient remains a central design and optimization variable in modern internal combustion engine research, playing a crucial role in the transition toward cleaner and more efficient engine technologies.

REFERENCES

Deng, B., Chen, Y., Liu, A., Xu, Z., Hu, S., Fu, J., Liu, X., Feng, R., & Zhou, L. (2019). The excess air coefficient effect on the performances for a motorcycle twin-spark gasoline engine: A wide condition range study. *Applied Thermal Engineering*, *150*(June 2018), 1028–1036. https://doi.org/10.1016/j.applthermaleng.2019.01.084

Fayyazbakhsh, A., Bell, M. L., Zhu, X., Mei, X., Koutný, M., Hajinajaf, N., & Zhang, Y. (2022). Engine emissions with air pollutants and greenhouse gases and their control technologies Permanent link This document is the Accepted Manuscipt version of the article that can be shared via institutional repository . *Journal of Cleaner Production*, *376*, 134260. https://doi.org/10.1016/j.jclepro.2022.134260

Gong, C., Si, X., & Liu, F. (2021). Combined effects of excess air ratio and EGR rate on combustion and emissions behaviors of a GDI engine with CO2 as simulated EGR (CO2) at low load. *Fuel*, *293*(November 2020), 120442. https://doi.org/10.1016/j.fuel.2021.120442

IMKO. (n.d.). *Institut für Maschinenmesstechnik und Kolbenmaschinen (IMKO). (n.d.). Verbrennungsmotoren. RWTH Aachen University, Germany.*

Majidi-Jirandehi, A. A., Soleymani, M. M., & Dehghani, H. (2021). Determine the Useful Life of Catalytic Converter and Standard Revision of Technical Inspection Centers. *Automot. Sci. Eng*, *11*(3), 3614–3619.

Meng, H., Ji, C., Yang, J., Wang, S., Chang, K., & Xin, G. (2021). Experimental study of the effects of excess air ratio on combustion and emission characteristics of the hydrogen-fueled rotary engine. *International Journal of Hydrogen Energy*, *46*(63), 32261–32272. https://doi.org/10.1016/j.ijhydene.2021.06.208

Mulholland, E., Miller, J., Bernard, Y., Lee, K., & Rodríguez, F. (2022). The role of NOx emission reductions in Euro 7/VII vehicle emission standards to reduce adverse health impacts in the EU27 through 2050. *Transportation Engineering*, *9*(x), 100133. https://doi.org/10.1016/j.treng.2022.100133

Reşitoğlu, I. A., Altinişik, K., & Keskin, A. (2015). The pollutant emissions from diesel-engine vehicles and exhaust aftertreatment systems. *Clean Technologies and Environmental Policy*, *17*(1), 15–27. https://doi.org/10.1007/s10098-014-0793-9

Rodríguez, F., Bernard, Y., Dornoff, J., & Mock, P. (2019). Recommendations for Post-Euro 6 Standards for Light-Duty Vehicles. *International Council on Clean Transportation*, *October*, 1–54.

Shamil Ahmed Flamarz Al-Arkawazi. (2019). *Analyzing and predicting the relation between air–fuel ratio (AFR), lambda (λ) and the exhaust emissions percentages and values of gasoline-fueled vehicles using versatile and portable emissions measurement system tool*.

Shu, J., Fu, J., Liu, J., Wang, S., Yin, Y., Deng, B., & Becker, S. M. (2019). Influences of excess air coefficient on combustion and emission performance of diesel pilot ignition natural gas engine by coupling computational fluid dynamics with reduced chemical kinetic model. *Energy Conversion and Management*, *187*(January), 283–296. https://doi.org/10.1016/j.enconman.2019.03.047

Sinkala, H., Özarslan, S., & Benarous, A. (2024). Effect of Excess Air Ratio on Emissions, Performance and Efficiency of Gasoline Fueled Engines: A Review. *Energy, Environment & Storage*, *4*(2), 32–45. https://doi.org/10.52924/pupx2065

Wang, S., Sun, Y., Yang, J., & Wang, H. (2024). Effect of excess air ratio and spark timing on the combustion and emission characteristics of turbulent jet ignition direct injection hydrogen engine. *International Journal of Hydrogen Energy*, *93*(May), 1166–1178. https://doi.org/10.1016/j.ijhydene.2024.11.052

Xin, G., Ji, C., Wang, S., Hong, C., Meng, H., & Yang, J. (2023). Experimental study on the effect of hydrogen substitution rate on combustion and emission characteristics of ammonia internal combustion engine under different excess air ratio. *Fuel*, *343*(February), 127992. https://doi.org/10.1016/j.fuel.2023.127992

Zervas, E., Montagne, X., & Lahaye, J. (2003). Emissions of regulated pollutants from a spark ignition engine. Influence of fuel and air/fuel equivalence ratio. *Environmental Science and Technology*, *37*(14), 3232–3238. https://doi.org/10.1021/es026321n

Comparison of the Performance of LiBr-H_2O and NH_3–H_2O Fluid Pairs in Absorption Cooling Systems

Gamze ŞAHİN[1]
Ahmet COŞKUN[2]

1- Gamze ŞAHİN; Suleyman Demirel University Faculty of Engineering Mechanical Engineering. gamzesahin099@gmail.com ORCID No: 0000-0002-4051-9881.
2- Prof. Dr. Ahmet COŞKUN; Suleyman Demirel University Faculty of Engineering Mechanical Engineering ahmetcoskun@sdu.edu.tr ORCID No: 0000-0002-3351-4898.

ABSTRACT

The continuous increase in global energy consumption, the need to reduce dependence on fossil fuels, and the necessity to limit greenhouse gas emissions have made the development of alternative heating and cooling systems with high energy efficiency imperative. In line with the growing energy demand and environmental concerns, absorption-based heating and cooling systems capable of utilizing low-temperature heat sources have emerged as environmentally sustainable alternatives to conventional mechanical vapor compression systems. This work focuses on evaluating the thermodynamic behavior of absorption cooling systems operating with LiBr–H_2O and NH_3–H_2O working pairs, with particular emphasis on how changes in generator temperature influence the COP under different absorber temperature levels. The findings of the study are consistent with similar studies reported in the literature and provide a comparative evaluation of the two systems. The results indicate that, for both systems, the COP increases with increasing generator temperature. However, at very high generator temperatures, the performance improvement is limited by the risk of crystallization in the LiBr–H_2O system, while in the NH_3–H_2O system it is constrained by the reduced absorption capacity of ammonia. Furthermore, an increase in absorber temperature hinders the absorption of refrigerant vapor into the solution, leading to a reduction in the heat transfer capacity of the system. The comparative analysis shows that the COP of the LiBr–H_2O system (0.723) is significantly higher than that of the NH_3–H_2O system (0.214), indicating that the LiBr–H_2O working pair is more suitable for systems operating at moderate temperature levels. Although both systems exhibit considerable potential for sustainable energy applications, it is concluded that, under the investigated operating conditions, the LiBr–H_2O working pair demonstrates superior thermodynamic efficiency.

Keywords – Absorption, LiBr–H_2O, NH_3–H_2O, Efficiency, Absorber, Generator.

INTRODUCTION

Due to the rapid economic and political developments occurring both globally and in our country, the demand for energy is increasing day by day. Particularly in developed and developing countries, rapid growth and transformation accelerate energy consumption. Considering the environmental impacts and limited reserves of fossil fuels, the importance of sustainable and environmentally friendly energy policies has become even more significant.

Heat pumps are energy-efficient devices that extract low-temperature environmental heat as a source and utilize it for heating or cooling

applications. In comparison with traditional fossil-fuel-based heating systems, these systems operate with substantially lower energy consumption and play an important role in mitigating greenhouse gas emissions. Absorption heat pumps operate based on a thermochemical cycle instead of a mechanical compressor and use thermal energy as their driving source. In these systems, heat transfer occurs through the chemical interaction between the refrigerant and the absorbent. Commonly used working pairs include water–lithium bromide (H_2O–LiBr) and ammonia–water (NH_3–H_2O). In the H_2O –LiBr pair, water serves as the refrigerant and lithium bromide as the absorbent, whereas in the NH_3–H_2O system, ammonia acts as the refrigerant and water as the absorbent.

Absorption systems have significant potential for utilizing low-temperature energy sources such as waste heat, solar energy, geothermal energy, or industrial process heat. Therefore, they are widely employed in applications such as building conditioning, industrial cooling, district heating systems, and cogeneration. Due to their low electricity demand and environmentally friendly characteristics, absorption heat pumps are considered an important option in sustainable energy conversion systems.

Advantages and Disadvantages of Absorption Systems

Advantages

- Since the system operates using heat instead of a mechanical compressor, electricity consumption is significantly lower.
- They enable energy recovery by utilizing waste heat and low-grade thermal energy.
- The absence of a mechanical compressor results in quieter operation.
- They are environmentally friendly due to the use of natural refrigerants.

Disadvantages

- The initial investment cost is high.
- Their coefficient of performance (COP) values are relatively low.
- The units are heavy and bulky, requiring a large installation area.
- There is a risk of crystallization in LiBr–H_2O working pairs.
- Since water is used as the refrigerant, very low operating temperatures cannot be achieved.

Literature studies have examined the performance characteristics of working pairs used in absorption systems using various methods. Horuz (1999) compared the LiBr–H_2O and NH_3–H_2O working pairs through diagrammatic and computer-aided analyses, concluding that the LiBr–H_2O pair is generally more advantageous in terms of overall performance. However, it was also noted that the risk of crystallization and the use of water as the refrigerant impose significant limitations, particularly in achieving low

operating temperatures. In the model developed by Talbi and Agnew (2000) for a LiBr–H_2O absorption refrigeration system, the exergy losses of the system components were calculated, and it was demonstrated that the highest exergy loss occurs in the absorber. Furthermore, it was stated that the thermal loads of the condenser and the evaporator are approximately 27.8% lower than those of the absorber and the generator, a situation attributed to the effect of the solution mixing temperature. In studies focusing on the use of renewable energy sources in absorption systems, Keçeciler et al. (2000) experimentally investigated a LiBr–H_2O–based absorption refrigeration system utilizing geothermal energy obtained from the Sivas Hot Springs. Their results indicated that, although this energy source is not suitable for electricity generation, it offers an economical alternative for cooling applications within the temperature range of 4–10 °C. A comparative evaluation of absorption heat pump and heat transformer configurations revealed that the use of LiBr–H_2O as a working solution leads to more favorable operating characteristics than NH_3–H_2O, particularly in terms of efficiency and operating pressure levels (Kurem and Horuz, 2001). In studies aimed at improving system performance, a numerical study conducted under Antalya climatic conditions demonstrated that the performance of solar-assisted single-effect LiBr–H_2O absorption systems is strongly influenced by the inlet temperature of the hot water supplied to the generator, which also affects the required collector area (Atmaca and Yiğit, 2003). Kaynaklı and Yamankaradeniz (2003) demonstrated that the thermal conditions of system components have opposing effects on absorption cycle performance. While elevated generator and evaporator temperatures contribute positively to efficiency, increases in absorber and condenser temperatures result in performance degradation for both NH_3–H_2O and LiBr–H_2O based systems. Overall, it was reported that systems using the H_2O–LiBr working pair yield better performance outcomes. Gomri (2010), who investigated multi-effect absorption systems, reported in analyses conducted for single-, double-, and triple-effect systems with a cooling capacity of 300 kW that the highest exergy efficiency and COP values 25.2% and 1.9, respectively were achieved in the triple-effect system. Kaushik and Arora (2010) reported that the COP values of single-effect and series-flow double-effect LiBr–H_2O absorption systems vary within the ranges of 0.6–0.75 and 1.0–1.28, respectively, and demonstrated that the highest irreversibility occurs in the absorber for both systems. Pastakkaya (2017) stated that solar-driven absorption heat pump systems are capable of meeting 84–99% of the cooling demand of buildings across different climatic regions, while providing significant economic and environmental benefits. Similarly, Gündüz and Cimşit (2024) reported that, in solar-assisted single-effect absorption systems, the LiBr–H_2O working pair offers higher solar utilization rates and efficiency values compared to the NH_3–H_2O working pair.

A review of the literature reveals that $LiBr–H_2O$ and $NH_3–H_2O$ working pairs are predominantly preferred in absorption systems, and the thermodynamic performance of these systems is mostly evaluated within the framework of energy and exergy analyses. In this study, based on a simple absorption cooling system using $LiBr–H_2O$ and $NH_3–H_2O$ working pairs, the COP variation depending on the generator temperature for different absorber temperatures was investigated.

WORKING PRINCIPLE AND THERMODYNAMIC ANALYSIS OF ABSORPTION COOLING SYSTEM

An absorption cooling system mainly consists of a generator, condenser, evaporator, absorber, an expansion valve, and a solution pump that ensures the circulation of the working solution. In the system, the refrigerant and the absorbent operate together to accomplish heat transfer. As illustrated in Figure 1, a typical absorption system cycle is shown, in which the refrigerant evaporates at low pressure in the evaporator by absorbing heat from the surroundings, thereby providing cooling or collecting energy from a heat source. The refrigerant vapor is then absorbed by the absorbent in the absorber, forming a solution. During this process, the released heat is rejected to the ambient environment or supplied to the heating system. Subsequently, the rich solution is pressurized by the pump and delivered to the generator. In the generator, the refrigerant component is vaporized using energy supplied from an external heat source (such as waste heat, solar energy, or natural gas). The vaporized refrigerant is condensed in the condenser, where it releases heat to the surroundings and changes into the liquid phase. The liquid refrigerant then passes through the expansion valve, where its pressure is reduced, and is directed back to the evaporator. In this way, the cycle is completed, and the system continues to transfer heat continuously.

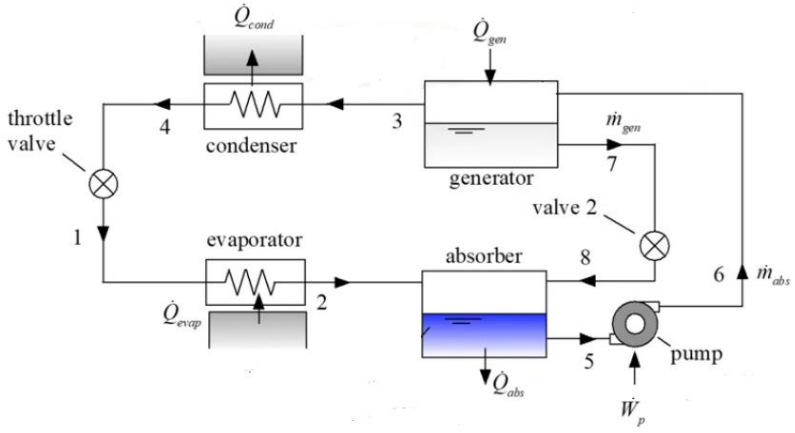


Figure 1: Absorption Cycle System (Klein and Nellis, 2011)

Assumptions for the LiBr–H_2O Absorption Cycle

- Ambient temperature $T_{amb} = 25$ ℃
- Generator temperature $T_{gen} = 90$ ℃
- Absorber temperature $T_{abs} = 32$ ℃
- Evaporator temperature $T_C = 5$ ℃
- Pump efficiency $\eta_p = 0{,}62$
- Approach temperature of condenser $\Delta T_{cond} = 4$ ℃
- Approach temperature of evaporator $\Delta T_{evap} = 3$ ℃
- Mass flow rate of water being cooled in evaporator $\dot{m}_w = 0{,}5\ kg/s$
- Inlet temperature of water being cooled in evaporator $T_{w,in}$= 12 ℃
- Exit temperature of water being cooled in evaporator $T_{w,out}$= 8 ℃

(Klein and Nellis, 2011)

Assumptions for the NH_3– H_2O Absorption Cycle

- Ambient temperature $T_{amb} = 25$ ℃
- Generator temperature $T_{gen} = 90$ ℃
- Generator pressure $P_{gen} = 1{,}6 \times 10^6\ Pa$
- Absorber temperature $T_{abs} = 32$ ℃
- Absorber pressure $P_{abs} = 400000\ Pa$
- Evaporator temperature $T_C = 5$ ℃
- Pump efficiency $\eta_p = 0{,}62$
- Approach temperature of condenser $\Delta T_{cond} = 4$ ℃
- Approach temperature of evaporator $\Delta T_{evap} = 3$ ℃
- Mass flow rate through the refrigeration cycle $\dot{m}_r = 0{,}05\ kg/s$

(Herold et al., 2016)

Thermodynamic Analysis

For any control volume, neglecting changes in kinetic and potential energies, the mass and energy balance equations are given as follows, respectively

$$\sum \dot{m}_{in} = \sum \dot{m}_{out} \tag{1}$$

$$\dot{Q} + \sum \dot{m}_{in}\, h_{in} = \dot{W} + \sum \dot{m}_{out}\, h_{out} \tag{2}$$

Here, "in" denotes the inlet and "out" denotes the outlet. $\dot{Q}$ and $\dot{W}$ represent the net heat and work inputs per unit time $\dot{m}$ denotes the mass flow rate, and h epresents the specific enthalpy.

$$\dot{Q}_{cond} = \dot{m}_r(h_3 - h_4) \tag{3}$$

$$\dot{Q}_{evap} = \dot{m}_r(h_2 - h_1) \tag{4}$$

$$\dot{Q}_{abs} + \dot{m}_{abs}h_5 = \dot{m}_r h_2 + \dot{m}_{gen}h_8 \tag{5}$$

$$\dot{Q}_{gen} + \dot{m}_{abs}h_6 = \dot{m}_r h_3 + \dot{m}_{gen}h_7 \tag{6}$$

$$\dot{W}_{pump} = \dot{m}_{abs}(h_6 - h_5) \tag{7}$$

Here, $\dot{Q}_{cond}$ denotes the condenser capacity, $\dot{Q}_{evap}$ the evaporator capacity, $\dot{Q}_{abs}$ the absorber capacity, $\dot{Q}_{gen}$ the generator capacity, and $\dot{W}_{pump}$ the pump power. In addition, $\dot{m}_{abs}$ represents the mass flow rate in the absorber, $\dot{m}_{gen}$ the mass flow rate in the generator, and $\dot{m}_r$ the refrigerant mass flow rate. The cooling performance is evaluated using the coefficient of performance (COP), which is defined as follows:

$$COP = \frac{\dot{Q}_{evap}}{\dot{Q}_{gen}} \tag{8}$$

RESULT AND DISCUSSION

A thermodynamic analysis of an absorption cooling system operating with LiBr–H_2O and NH_3–H_2O working pairs was performed, and the variation of the COP with generator temperature at different absorber temperatures was investigated.

The thermodynamic properties at all state points for the LiBr–H_2O and NH_3–H_2O working pairs are presented in Tables 1 and 2, respectively, while the calculated mass flow rates of the absorber, generator, and refrigerant for these working pairs are given in Table 3. In addition, the power capacities of the system components for the LiBr–H_2O and NH_3–H_2O working pairs are presented in Tables 4 and 5.

Table 1: Thermodynamic properties for the LiBr–H_2O Absorption Cycle

State No	Temperature T (°C)	Pressure P (Pa)	Enthalpy h (J/kg)	Mass flow rate $\dot{m}f_{LiBr}$ (kg/s)
1	5	872,5	121554	-
2	5	872,5	2,510E+06	-
3	90	4009	2,669E+06	-
4	29	4009	121554	-
5	32	872,5	73842	0,5369
6	32	4009	73845	0,5369
7	90	4009	261452	0,6799
8	-	872,5	261452	0,6799

Table 2: Thermodynamic properties for the NH_3–H_2O Absorption Cycle

State No	Temperature T (°C)	Pressure P (Pa)	Enthalpy h (J/kg)	Mass flow rate $\dot{m}_{NH_3H_2O}$ (kg/s)
1	-1,282	400000	121445	0,9786
2	2	400000	1,081E+06	0,9786
3	90	1,600E+06	1,462E+06	0,9786
4	29	1,600E+06	121445	0,9786
5	32	400000	-96438	0,5052
6	32,3	1,600E+06	-94076	0,5052
7	90	1,600E+06	168968	0,4625
8	51,8	400000	168968	0,4625

Table 3: The calculated mass flow rate values of the absorber, generator and refrigerant (kg/s)

	$LiBr–H_2O$	$NH_3–H_2O$
$\dot{m}_{abs}$	0,017	0,605
$\dot{m}_{gen}$	0,013	0,555
$\dot{m}_r$	0,004	0,05

Table 4: Power capacities of system components for the $LiBr–H_2O$ Absorption Cycle

System Components	Power Capacity (W)
Evaporator	8390
Condenser	8950
Generator	11591
Absorber	11032
Pump	0,053
COP	0,723

Table 5: Power capacities of system components and value of COP for the $NH_3–H_2O$ Absorption Cycle

System Components	Power Capacity (W)
Evaporator	47955
Condenser	67012
Generator	223664
Absorber	206034
Pump	1428
COP	0,214

The variation of COP with generator temperature at different absorber temperatures for the absorption cooling system operating with $LiBr–H_2O$ and $NH_3–H_2O$ working pairs is illustrated in Figures 2 and 3, respectively. As shown in the figures, the COP increases with increasing generator temperature. The generator temperature is a key parameter affecting the performance of absorption systems. As the generator temperature increases, the tendency of the refrigerant to vaporize from the rich solution also increases, resulting in a greater amount of refrigerant being separated from the solution. This enhances the purity and pressure of the refrigerant vapor circulating in the cycle, allowing a more favorable pressure difference to be established between the evaporator and the condenser. Consequently, while the cooling or heating effect achieved in the evaporator increases, the heat input supplied to the generator does not increase proportionally, leading to an improvement in the COP of the system. However, above a certain temperature

level, the risk of crystallization arises in the LiBr–H_2O system, whereas in the NH_3–H_2O system, excessive dilution of the solution leads to difficulties in ammonia absorption within the absorber. Therefore, further increases in generator temperature may cause the COP to level off or even decrease. Moreover, as the absorber temperature increases, the absorption of refrigerant vapor by the solution becomes more difficult. At higher temperatures, the absorbent absorbs a smaller amount of refrigerant. This results in a reduction in the refrigerant mass flow rate circulating in the cycle and a decrease in vapor generation. An increase in absorber temperature also raises the system pressures, thereby reducing the pressure difference between the evaporator and the condenser. As a result, the heat transfer capacity of the cycle decreases, leading to a reduction in the COP.

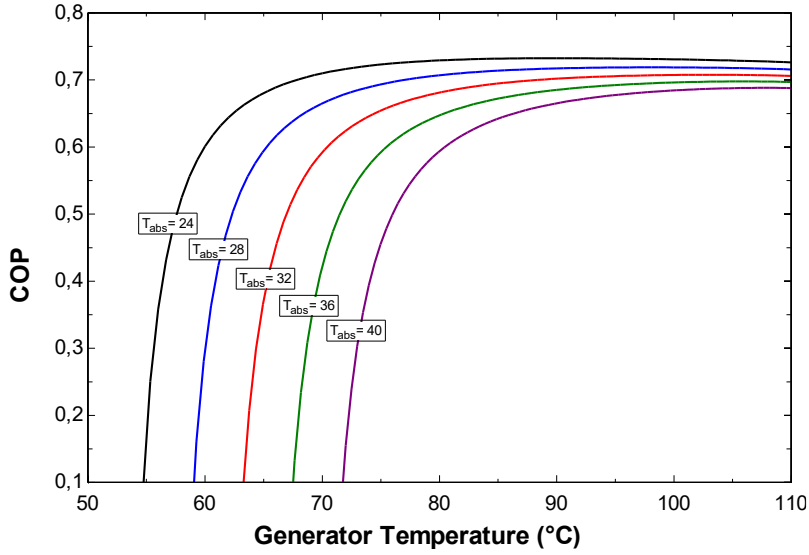


Figure 2: The effect of generator temperature variation on the COP value at different absorber temperatures for the LiBr–H_2O Absorption Cycle

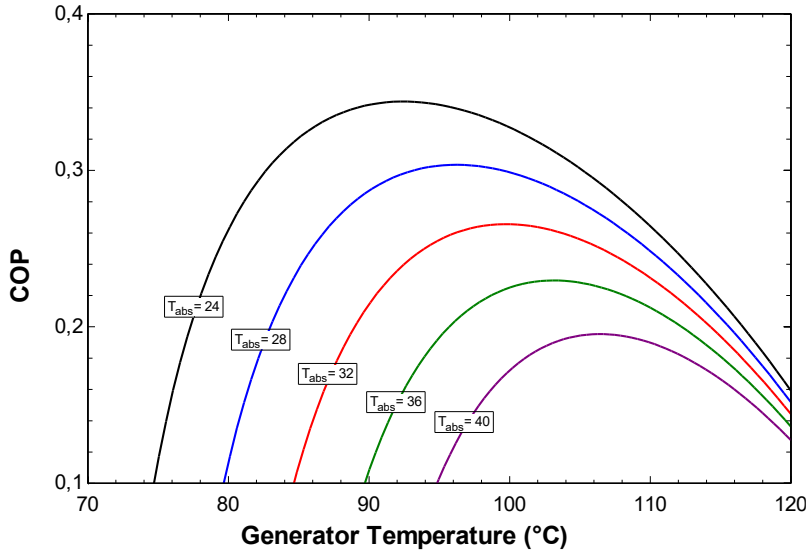


Figure 3: The effect of generator temperature variation on the COP value at different absorber temperatures for the NH_3–H_2O Absorption Cycle

CONCLUSION

In the thermodynamic analysis of absorption cooling systems operating with LiBr–H_2O and NH_3–H_2O working pairs, it was found that the generator and absorber temperatures are the primary parameters determining system performance. An increase in generator temperature facilitates the separation of the refrigerant, thereby increasing the COP; however, above a certain temperature level, limiting effects such as crystallization or weakened absorption become significant. In contrast, an increase in absorber temperature hinders the absorption of refrigerant vapor by the solution, resulting in a decrease in the COP. As a result of the analyses conducted, the COP of the LiBr–H_2O system was calculated as 0.723, while the COP of the NH_3–H_2O system was found to be 0.214. These results indicate that the LiBr–H_2O system is more efficient for heating and cooling applications operating at moderate temperature levels. Overall, it is concluded that absorption cooling systems are capable of effectively utilizing low-grade heat sources and represent a sustainable alternative to fossil-fuel-based systems.

REFERENCES

Atmaca I., Yigit A. (2003). Simulation of solar-powered absorption cooling system. Renewable Energy, 28(8), 1277-1293.

Gomri R. (2010). Investigation of the potential of application of single effect and multiple effect absorption cooling systems. Energy Conversion and Management, 51(8), 1629-1636.

Gündüz, A.H., Cimşit, C. (2024). Solar Fraction (SF) Analysis of Solar-Energy-Driven Absorption Cooling Systems. Journal of Engineering and Architecture Faculty, 39(4), 2103-2112.

Herold, K., Radermacher, R., Klein, S., (2016). Absorption Chillers and Heat Pumps. CRC Press, 346, Boca Raton.

Horuz, İ. (1999). A Comparison of the Ammonia–Water Pair and the Lithium Bromide–Water Pair in Absorption Cooling Systems. Uludağ University, Institute of Science, Master's Thesis, 147p, Bursa.

Kaushik S. C., Arora A. (2009). Energy and exergy analysis of single effect and series flow double effect water–lithium bromide absorption refrigeration systems. International Journal of Refrigeration, 32(6), 1247-1258.

Kaynaklı, Ö., Yamankaradeniz, R. (2003). Comparison of Single-Stage Absorption Cooling Systems Using H_2O–LiBr and NH_3–H_2O Solutions. DEU Faculty of Engineering Journal of Science and Engineering, 5(2), 73-87.

Keçeciler A., Acar H. İ., Doğan A. (2000). Thermodynamic analysis of the absorption refrigeration system with geothermal energy: an experimental study. Energy Conversion and Management, 41(1), 37-48.

Klein, S., Nellis, G. (2011). Thermodynamics. Cambridge University Press, 1072, New York.

Kurem, E., Horuz, İ. (2001). A comparison between ammonia-water and water-lithium bromide solutions in absorption heat transformers. International Communications in Heat and Mass Transfer, 28(3), 427-438.

Pastakkaya, B. (2017). Investigation of the Heating and Cooling Performance of a Solar-Energy-Sourced Absorption Heat Pump System for Different Degree-Day Regions. 13th National HVAC&R Engineering Congress, April 19-22, 2017, İzmir, 893-901.

Talbi M. M., Agnew B. (2000). Exergy analysis: an absorption refrigerator using lithium bromide and water as the working fluids. Applied Thermal Engineering, 20(7), 619-630.

Adsorption of Drugs with Zeolite Structures

Amani MARABEH[1]
İbrahim İNANÇ[2]

1- Yüksek Lisans Öğrencisi.; Ondokuz Mayıs Üniversitesi Nanobilim ve Nanoteknoloji Anabilim Dalı Başkanlığı - Nanobilim ve Nanoteknoloji (İngilizce) Yüksek Lisans Programı . amani.marabeh29@gmail.com, ORCID No: 0009-0003-0629-7888.
2- Dr. Öğr. Üyesi.; Ondokuz Mayıs Üniversitesi Mühendislik Fakültesi, Metalurji Ve Malzeme Mühendisliği Bölümü . ibrahim.inanc@omu.edu.tr ORCID No: 0000-0003-1988-1197.

ABSTRACT

Zeolites, a class of crystalline hydrated aluminosilicates, have emerged as highly promising materials for advanced drug delivery systems (DDS) due to their unique structural and chemical properties. Their well-defined porous architecture, high surface area, and tunable surface chemistry facilitate exceptional drug adsorption capacity and controlled release kinetics. This comprehensive review explores the fundamental principles of drug adsorption onto zeolites, detailing the various mechanisms, including physical adsorption, ion exchange, and specific chemical interactions. It further examines the diverse applications of zeolite- based DDS across various therapeutic areas, such as oncology, infectious diseases, and inflammation management, highlighting key case studies involving drugs like Doxorubicin,5-Fluorouracil, and Ibuprofen. Finally, the article discusses the recent advances, including the development of nanozeolites and functionalized composites, addresses the current challenges in clinical translation, and outlines the future perspectives for these versatile materials in pharmaceutical research.

Keywords – Zeolite, Drug Delivery, Adsorption, Nanozeolites, Pharmaceutical.

THE RISE OF ZEOLITES IN DRUG DELIVERY

The development of effective drug delivery systems is a cornerstone of modern pharmacology, aiming to enhance therapeutic efficacy while minimizing systemic toxicity. Traditional drug formulations often suffer from poor bioavailability, rapid degradation, and non-specific distribution. In response, porous materials have been extensively investigated as drug carriers, with zeolites standing out as a particularly versatile and robust class (Pandya et al., 2024).

Zeolites are microporous materials composed of a framework of silicon, aluminum, and oxygen, forming a three-dimensional structure with channels and cages of molecular dimensions (Kordala & Wyszkowski, 2024). This unique structure, combined with their inherent stability, biocompatibility, and low cost, positions them as ideal candidates for drug adsorption and controlled release (Servatan et al., 2020). The adsorption process, which is central to their function as DDS, involves loading the therapeutic agent into the zeolite pores and onto the external surface, followed by a controlled release triggered by environmental factors like pH, temperature, or ion concentration at the target site (Kadja, Culsum, & Putri, 2023).

This review provides an in-depth analysis of the principles governing drug adsorption by zeolites, the factors influencing their performance, and their current and future roles in pharmaceutical applications.

ZEOLITE FUNDAMENTALS AND STRUCTURAL PROPERTIES

The efficacy of zeolites as drug carriers is intrinsically linked to their physicochemical properties, which can be precisely engineered for specific applications.

Structure and Composition

Zeolites are characterized by their crystalline structure, which can be natural (e.g., Clinoptilolite) or synthetic (e.g., ZSM-5, Zeolite A, Zeolite Y) (Pérez-Botella, Valencia, & Rey, 2022). The framework is composed SiO and AlO tetrahedra linked by shared oxygen atoms. The substitution of Si^{4+} by Al^{3+} creates a net negative charge on the framework, which is balanced by exchangeable extra-framework cations (e.g., Na^{+}, K^{+}, Ca^{2+}) (M Servatan, 2020).

The *Silicon-to-Aluminum ratio (Si/Al)* is a critical parameter that dictates the zeolite's properties:

Low Si/Al: Indicates a higher concentration of Al^{3+} and thus a higher negative charge density, leading to increased hydrophilicity and a greater capacity for ion exchange (IMS Souza, 2023).

High Si/Al: Results in a more hydrophobic surface and greater thermal and chemical stability, which can be advantageous for certain drug types (G Dunkley, 2025).

Porosity and Surface Area

Zeolites possess a high internal surface area, often exceeding 500 m^2/g, primarily due to their microporous nature (pore size < 2 nm) (F Dernaika, 2024). The uniform and regular pore size distribution acts as a molecular sieve, allowing for selective loading and release of drug molecules based on their size and shape. The development of *nanozeolites* (particles <100 nm) has further enhanced their potential by increasing the external surface area and improving cellular uptake efficiency (FG Mayaki, 2024).

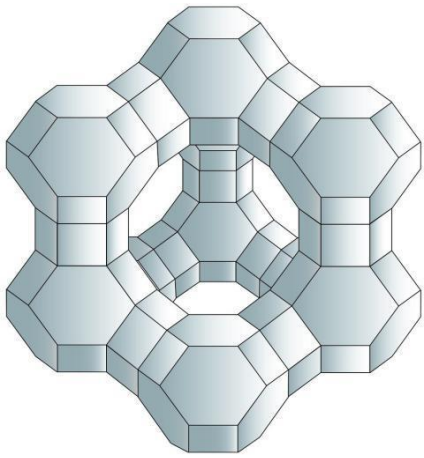


Figure 1**:** Idealized Crystalline Structure of a Zeolite, illustrating the interconnected pores and channels.

MECHANISMS AND FACTORS GOVERNING DRUG ADSORPTION

Drug adsorption onto zeolites is a complex process driven by multiple simultaneous interactions between the drug molecule, the zeolite surface, and the solvent medium.

Adsorption Mechanisms

The primary mechanisms responsible for drug loading include:

1. **Physical Adsorption (Physisorption):** Driven by weak van der Waals forces and hydrogen bonding between the drug and the zeolite surface. This is often the dominant mechanism for non-polar or weakly polar drugs (NR Mijailović, 2022).
2. **Ion Exchange:** The negatively charged zeolite framework attracts positively charged drug molecules (e.g., protonated amines). The drug cation replaces the extra- framework cations, leading to high loading capacity for ionic drugs (EA Serna-Galvis, 2024). This mechanism is particularly important for basic drugs.
3. **Specific Chemical Interactions:** In some cases, stronger interactions like pi-pi stacking (for aromatic drugs) or the formation of well-defined local chemical bonds (covalent or

ionic) can occur, leading to highly stable drug-zeolite complexes (M Servatan, 2013).

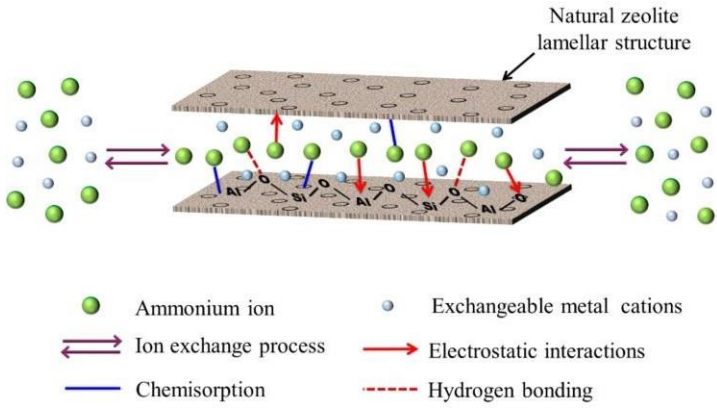


Figure 2: Schematic of Drug Adsorption Mechanisms, highlighting electrostatic, hydrogen bonding, and van der Waals interactions.

Key Factors Influencing Adsorption

The efficiency and extent of drug adsorption are governed by several factors, summarized in Table1.

Table 1: Key Factors Influencing Drug Adsorption onto Zeolites

Factor	Description	Effect on Adsorption
Drug Polarity/Ionization	The charge and hydrophilicity of the drug molecule, often dictated by the solution pH.	Polar, ionized drugs (especially cations) show higher adsorption via ion exchange (S Ambarish, 2024).
Zeolite Si/Al	Determines the framework charge density and hydrophilicity.	Low Si/Al (hydrophilic) favors polar drugs; high Si/Al (hydrophobic) favors non-polar drugs (G Dunkley, 2025).
Pore Size and Volume	The size and shape of the zeolite channels relative to the drug molecule.	Optimal adsorption occurs when the drug size is slightly smaller than the pore opening, allowing for maximum internal loading (M Servatan, 2020).
Temperature	Affects the equilibrium and kinetics of the adsorption process.	Adsorption is typically exothermic; thus, lower temperatures often favor higher loading capacity (NR Mijailović, 2022).
Loading Method	Techniques like solvent evaporation, impregnation, or ion exchange.	Ion exchange is highly effective for ionic drugs; impregnation is common for neutral drugs (MA Sayed, 2022).

APPLICATIONS IN ADVANCED DRUG DELIVERY SYSTEMS

Zeolite-based DDS have been successfully explored for the delivery of various therapeutic agents, demonstrating superior controlled release profiles and targeted action.

Anticancer Drug Delivery

Zeolites, particularly those with large pores like Faujasite (FAU) and Zeolite Y, have shown significant potential in oncology (J Hao, 2021).

- **Doxorubicin (DOX):** A widely used chemotherapy agent, DOX has been successfully loaded onto various zeolite structures. The porous nature of the zeolite protects DOX from premature degradation and allows for a sustained release profile, which is crucial for reducing the cardiotoxicity associated with the drug (V Khojaewa, 2019).
- **5-Fluorouracil (5-FU):** Studies using Zeolite ZSM-5 and H-Beta have demonstrated high loading capacity for 5-FU. The release of 5-FU from the zeolite carrier can be tuned by adjusting the Si/Al and the drug loading level, enabling targeted delivery to cancer cells (A Grela, 2021). Furthermore, zeolite-drug complexes have been investigated for oral administration, aiming for targeted release in the colon for colorectal cancer therapy (E Khodaverdi, 2016).

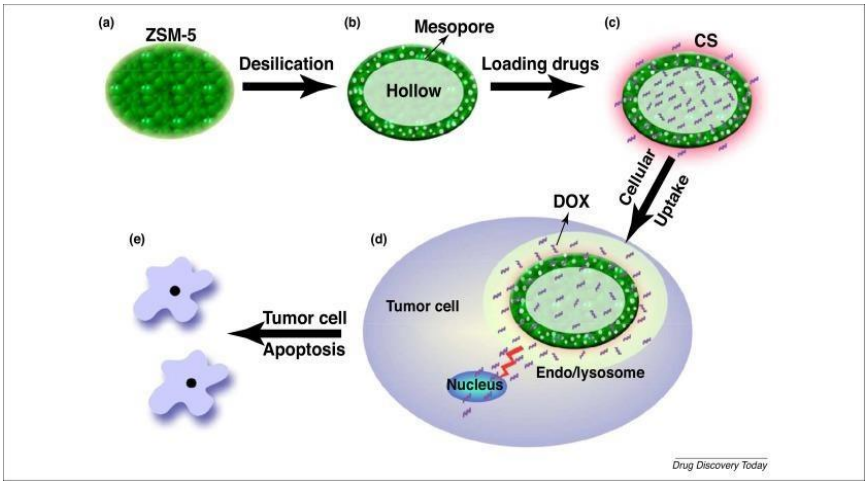


Figure 3: Schematic of Doxorubicin (DOX) Delivery via Zeolite Nanoparticles, showing loading, cellular uptake, and release within the tumor cell.

Anti-inflammatory and Antibiotic Delivery

Zeolites are also effective carriers for non-steroidal anti-inflammatory drugs (NSAIDs) and antibiotics, offering solutions to common formulation challenges.

- **NSAIDs (Ibuprofen, Indomethacin):** NSAIDs often cause gastric irritation. Loading these drugs onto synthetic zeolites allows for a pH-dependent release profile, where the drug is protected in the acidic environment of the stomach and released primarily in the neutral to basic environment of the intestine (E Khodaverdi, 2014). This controlled release minimizes local irritation and improves patient compliance.

- **Antibiotics (Levofloxacin):** Zeolite-A nanocomposites, often combined with biopolymers like chitosan, have been developed as advanced carriers for antibiotics such as Levofloxacin (J Brauer, 2025). The composite structure enhances the drug loading capacity and provides a sustained release, which is beneficial for treating chronic infections and reducing the risk of antibiotic resistance.

Clinoptilolite: A Natural Carrier for Drug Delivery

Clinoptilolite is the most abundant and extensively studied natural zeolite, prized for its high adsorption capacity, chemical inertness, and low toxicity, making it a promising material for pharmaceutical applications (I Siora, 2025). Its structure, characterized by a high Si/Al and a three-dimensional channel system, allows for the effective encapsulation of various drug molecules (V Hovhannisyan, 2017).

One of the primary applications of Clinoptilolite is in the delivery of non-steroidal anti- inflammatory drugs (NSAIDs). Studies have successfully employed Clinoptilolite as a carrier for drugs like *Diclofenac* and *Ibuprofen*, demonstrating a sustained release profile that helps mitigate the gastrointestinal side effects typically associated with these drugs (GTM Kadja, 2023).

Furthermore, Clinoptilolite has been explored as a carrier for antibiotics in topical therapies, such as *anti-acne treatments*, by utilizing its ion-exchange properties to load active metal ions or drug molecules (G Cerri, 2004). The natural origin and generally recognized as safe (GRAS) status of Clinoptilolite also make it a favorable candidate for oral drug delivery systems, where it can be used to encapsulate vitamins (e.g., Vitamin E) or other functional materials for slow release (M Servatan, 2020). Its use in clinical evaluations, such as the study of PMA-zeolite (a modified Clinoptilolite), further highlights its potential for safe and prolonged use in human applications (SK Pavelić, 2022).

SYNTHESIS, FUNCTIONALIZATION, AND COMPOSITES

To optimize drug adsorption and release characteristics, zeolites are often modified or integrated into composite materials.

Nanozeolites and Size Control

The synthesis of *nanozeolites* (Nanosized Zeolite Particles) is a major advancement. Their small size allows for better penetration into biological barriers and improved cellular uptake, which is critical for intravenous administration and targeted therapy (FG Mayaki, 2024). Synthesis methods, such as hydrothermal synthesis and template-assisted methods, are continuously being refined to control particle size, morphology, and crystallinity (T Pandya, 2024).

Surface Functionalization

The external surface of zeolites can be functionalized with various chemical groups or polymers to enhance drug loading, stability, and targeting capabilities. For instance, surface modification with silane coupling agents or organic polymers can alter the hydrophobicity and introduce specific binding sites for drug molecules or targeting ligands (P Zarrintaj, 2020). The integration of zeolites into polymeric materials, such as hydrogels or microparticles, creates hybrid systems that combine the high adsorption capacity of the zeolite with the structural integrity and biocompatibility of the polymer (A Grela, 2021).

CHALLENGES AND FUTURE PERSPECTIVES

Despite the significant progress, the clinical translation of zeolite-based DDS faces several challenges, which also define the future research directions in this field.

Current Challenges

1. **Low Drug Loading:** While zeolites have a high surface area, the size-exclusion effect of the micropores can limit the loading of large drug molecules, leading to a relatively low overall drug-to-carrier ratio in some cases (S Ambarish, 2024).
2. **Reproducibility and Scalability:** The synthesis of highly uniform nanozeolites with consistent pore size and surface

properties remains a challenge for large-scale pharmaceutical production (A Grela, 2021).

3. **Toxicity and Biocompatibility:** Although generally considered safe, the long-term *in vivo* fate, degradation products, and potential toxicity of certain synthetic zeolites, especially at the nanoscale, require further rigorous investigation and standardization (J Hao, 2021).

Future Directions

The future of zeolite-mediated drug adsorption is focused on intelligent, multi-functional systems:

Stimuli-Responsive Release: Developing zeolites that release their payload in response to specific physiological triggers, such as pH changes in tumors or inflammation sites, or elevated enzyme levels (GTM Kadja, 2023).

Machine Learning Optimization: Utilizing machine learning and computational modeling to predict optimal zeolite structures (Si/Al, pore size) and functionalization strategies for a given drug molecule, significantly accelerating the development process (AR Bertão, 2024).

Clinical Translation: Moving beyond in vitro and animal studies to human clinical trials. Preliminary clinical evaluations of defined zeolite materials, such as PMA- zeolite, have shown promising safety profiles, paving the way for broader clinical application (SK Pavelić, 2022). Furthermore, their use in tissue engineering and gene delivery systems is an emerging area of research (P Zarrintaj, 2020).

CONCLUSION

Zeolites represent a powerful and versatile platform for drug adsorption and controlled delivery. Their unique crystalline structure, high porosity, and tunable surface chemistry allow for the efficient loading of a wide range of therapeutic agents, from small-molecule drugs to proteins. The ability to control the release kinetics, minimize side effects, and potentially target specific tissues makes zeolite-based systems a compelling alternative to conventional formulations. While challenges related to scalability and clinical safety remain, ongoing research into nanozeolites, functionalized composites, and advanced computational design promises to unlock the full potential of these materials, ushering in a new era of precise and effective drug delivery.

REFERENCES

Ambarish, S. (2024). Zeolites: Microporous natural mineral carrier in controlled, targeted (nanocarriers) and gene delivery systems . German Journal of Pharmaceuticals and Biomedical Sciences. https://gjpb.de/index.php/gjpb/article/view/107

Ambarish, S. (2024). Zeolites: Microporous natural mineral carrier in controlled, targeted (nanocarriers) and gene delivery systems . German Journal of Pharmaceuticals and Biomedical Sciences. https://www.researchgate.net/publication/380690147

Bertao, A. R. (2024). Machine learning-assisted optimization of drug delivery systems based on zeolites. ACS Applied Materials & Interfaces, 16 (5), 6234–6247. https://doi.org/10.1021/acsami.3c182

Brauer, J. (2025). Emerging extra-large pore zeolites as adsorbents for pharmaceuticals. Microporous and Mesoporous Materials, 367 , 112822. https://www.sciencedirect.com/science/article/pii/S1387181125003476

Cerri, G. (2004). Zeolites in biomedical application: Zn-exchanged clinoptilolite-rich rock as active carrier for antibiotics in anti-acne topical therapy. Applied Clay Science, 27 (3–4), 141–150. https://www.sciencedirect.com/science/article/pii/S0169131704000547

Dernaika, F. (2024). Development and characterization of a zeolite-based drug delivery system for cannabidiol. Heliyon, 10 (3), e23641. https://pmc.ncbi.nlm.nih.gov/articles/PMC11409077/

Dunkley, G. (2025). The effect of Si/Al ratio and drug loading level on the molecular behaviour and controlled release of 5-fluorouracil from zeolite H-Beta. Journal of Materials Chemistry B . https://pubs.rsc.org/en/content/articlehtml/2025/tb/d5tb00256g

Grela, A. (2021). A review on the application of zeolites and mesoporous silica materials in the removal of non-steroidal anti-inflammatory drugs and antibiotics from water. Materials, 14 (17), 4994. https://pmc.ncbi.nlm.nih.gov/articles/PMC8433637/

Hao, J. (2021). Effects of zeolite as a drug delivery system on cancer therapy: A systematic review. Molecules, 26 (20), 6196. https://www.mdpi.com/1420-3049/26/20/6196

Hovhannisyan, V. (2017). Photodynamic dye adsorption and release performance of zeolite-based carriers. Scientific Reports, 7 , 45503. https://www.nature.com/articles/srep45503

Kadja, G. T. M. (2023). Recent advances in the utilization of zeolite-based materials for controlled drug delivery. Materials Today Chemistry, 28 , 101356. https://www.sciencedirect.com/science/article/pii/S2211715623001492

Khodaverdi, E. (2014). Evaluation of synthetic zeolites as oral delivery vehicle for non-steroidal anti-inflammatory drugs. Journal of Pharmaceutical Sciences, 103 (1), 327–335. https://pmc.ncbi.nlm.nih.gov/articles/PMC4069848/

Khodaverdi, E. (2016). Synthetic zeolites as controlled-release delivery systems for nonsteroidal anti-inflammatory drugs. Chemistry & Biodiversity, 13 (8), 1076–1085. https://onlinelibrary.wiley.com/doi/10.1111/cbdd.12716

Khojaewa, V. (2019). Zeolites as carriers of antitumor ribonuclease binase. Frontiers in Pharmacology, 10 , 442. https://www.frontiersin.org/articles/10.3389/fphar.2019.00442

Kordala, N. (2024). Zeolite properties, methods of synthesis, and selected applications. Molecules, 29 (5), 1069. https://www.mdpi.com/1420-3049/29/5/1069

Mayaki, F. G. (2024). Nanozeolites: Eco-friendly materials for healthcare applications. Physica Scripta, 99 (4), 045002. https://iopscience.iop.org/article/10.1088/1402-4896/ae26f4

Mijailović, N. R. (2022). Environmental and pharmacokinetic aspects of zeolite/pharmaceutical systems—Two facets of adsorption ability. Minerals, 12 (8), 837. https://www.mdpi.com/2073-4344/12/8/837

Pandya, T. (2024). Zeolite-based nanoparticles drug delivery systems in modern pharmaceutical research and environmental remediation. Heliyon, 10 (4), e24486. https://www.sciencedirect.com/science/article/pii/S2405844024124486

Pandya, T. (2024). Zeolite-based nanoparticles drug delivery systems in modern pharmaceutical research and environmental remediation. Heliyon, 10 (4), e24486. https://pmc.ncbi.nlm.nih.gov/articles/PMC11388657/

Pavelić, S. K. (2022). Clinical evaluation of a defined zeolite-clinoptilolite in a 12-week open-label study. Molecules, 27 (12), 3819. https://pmc.ncbi.nlm.nih.gov/articles/PMC9197155/

Pérez-Botella, E. (2022). Zeolites in adsorption processes: State of the art and future prospects. Chemical Reviews, 122 (16), 14175–14241. https://doi.org/10.1021/acs.chemrev.2c00140

Serna-Galvis, E. A. (2024). Removal and elimination of pharmaceuticals in water using zeolites. Environmental Research, 247 , 118087. https://pmc.ncbi.nlm.nih.gov/articles/PMC11602855/

Servatan, M. (2013). Factors affecting drug adsorption on beta zeolites. Journal of Separation Science, 36 (9), 1461–1470.

https://onlinelibrary.wiley.com/doi/10.1002/jssc.201201142
Servatan, M. (2020). Zeolites in drug delivery. PAR NSF . https://par.nsf.gov/servlets/purl/10211860
Servatan, M. (2020). Zeolites in drug delivery: Progress, challenges and opportunities. Materials Science and Engineering: C, 110 , 110760. https://www.sciencedirect.com/science/article/abs/pii/S1359644620300738
Siora, I. (2025). Comparative study of the structural characteristics of clinoptilolite-rich zeolites. Microporous and Mesoporous Materials, 355 , 112656. https://www.sciencedirect.com/science/article/abs/pii/S1387181125002938
Souza, I. M. S. (2023). Zeolites as ingredients of medicinal products. Pharmaceuticals, 15 (5), 1352. https://www.mdpi.com/1999-4923/15/5/1352
Sayed, M. A. (2022). Insight into the loading properties of Na^+ green zeolite. ACS Omega, 7 (3), 2478–2487. https://doi.org/10.1021/acsomega.1c06671
Zarrintaj, P. (2020). Zeolite in tissue engineering: Opportunities and challenges. Materials Chemistry and Physics, 248 , 122902. [https://onlinelibrary.wiley.com/doi/10.1002/mco2.5](https://onlinelibrary.wiley.com/doi/10.1002/mco2.
Kadja, G. T., Culsum, N. T., & Putri, R. M. (2023). Recent advances in the utilization of zeolite-based materials for controlled drug delivery. Results in Chemistry, 5, 100910.
Kordala, N., & Wyszkowski, M. (2024). Zeolite properties, methods of synthesis, and selected applications. Molecules, 29(5), 1069.

Pandya, T., Patel, S., Kulkarni, M., Singh, Y. R., Khodakiya, A., Bhattacharya, S., & Prajapati, B. G. (2024). Zeolite-based nanoparticles drug delivery systems in modern pharmaceutical research and environmental remediation. Heliyon, 10(16).
Pérez-Botella, E., Valencia, S., & Rey, F. (2022). Zeolites in adsorption processes: State of the art and future prospects. Chemical reviews, 122(24), 17647-17695.
Servatan, M., Zarrintaj, P., Mahmodi, G., Kim, S.-J., Ganjali, M. R., Saeb, M. R., & Mozafari, M. (2020). Zeolites in drug delivery: Progress, challenges and opportunities. Drug Discovery Today, 25(4), 642-656.

Integrating Human Factors into Operational Excellence: A FUCOM-Based Multidimensional Consistency-Optimized Approach in Socio-Technical Systems

Burcu YILMAZ KAYA[1]

1- Assoc. Prof. Dr.; Gazi University, Faculty of Engineering, Industrial Engineering Department. burcuyilmaz@gazi.edu.tr ORCID No: 0000-0002-5088-5842

ABSTRACT

Operational Excellence (OPEX) constitutes a strategic and systematic pursuit of organizational learning and capability building, aiming to embed excellence across all operational layers. While OPEX is broadly recognized for its emphasis on cost reduction, error minimization, and high-performance outcomes, its deep conceptual and methodological intersection with Human Factors Engineering (HFE) has remained insufficiently theorized and empirically underexplored. Addressing this critical gap, the present study integrates HFE principles into the evaluation of OPEX critical success factors (CSFs) through a rigorous mathematical solution approach. To achieve methodological precision, the Full Consistency Method (FUCOM) was applied, marking its first one-to-one use in the OPEX-HFE context. FUCOM's methodological superiority lies in its ability to derive maximally consistent weight structures with minimal comparison requirements, while eliminating logical inconsistencies that commonly affect other similar methods. Importantly, its mathematical formulation-requiring careful construction of consistency constraints and optimization of deviation-from-consistency (DFC- χ) functions-demands a high level of analytical expertise to implement correctly. In this study, FUCOM produced perfect coherence, with all χ values calculated as zero, ensuring fully reliable results. The consolidated ranking identified "C2: Error Prevention Mechanisms" as the most influential CSF, followed by standardized work and ergonomic workplace design. Overall, this research delivers the first mathematically validated decision model linking OPEX and HFE, significantly advancing the theoretical and methodological foundations of operational strategy design in socio-technical systems.

Keywords – OPEX, Human Factors Engineering, Efficiency Management, FUCOM.

INTRODUCTION

Operational excellence (OPEX) is a direction toward learning and developing an excellent culture in all aspects of an organization (Hajiagha, Mandiraji, Moradi, Garza-Reyes & Alaei, 2024). OPEX is very important to all kind of functions of various industries, hence has varying definitions particular to the application point, yet, the definition of Hammer (2004) as high-performance achievements through current tools and operating methods to ensure reduced costs and errors is one of the most valid ones for all type of problems.

OPEX aims to increase the efficiency of existing systems by determining and improving machine and human capacity utilization ratios, eliminating downtime, increasing usability, thereby reducing standard time, minimizing errors, and improving quality. In this context, as can be

understood from OPEX frame definition, OPEX philosophy and applications intersects with human factors engineering (HFE) applications, while it covers the same operation field and achieves the intersecting goals through HFE solutions.

Past studies have investigated the critical success factors (CSFs) of OPEX methodologies and their relationship to application domains and/or employed technology level. However, as also stated by Antony, Swarnakar, Gupta, Kaur, Jayaraman, Tortorella and Cudney (2023) those correlations varied across studies, leading to misunderstandings of the actual influence of CSFs on the OPEX implementation. Furthermore, association of the human factors engineering (HFE) related elements with OPEX CSFs has not been studied thoroughly, although the OPEX core relies on efficiency and optimal capacity utilization as in line with HFE. Where employing OPEX methodologies has the aim to solutions that will provide the best productivity and capacity utilization rates to realize the full potential of the organization, unfortunately direct implementation of OPEX can have negative impact and result in challenges associated with financial, social, and environmental factors, hence acknowledging these risks the CSFs have to be first identified and then dissected in terms of all phenomenal elements to the work structure.

In this book chapter, the identified OPEX CSFs via detailed extant literature reviews were dissected regarding HFE point-of-view with group decision-making (GDM), where, Full Consistency Method (FUCOM) was used to delineate the results as a novel and robust mathematical solution approach. This study addresses the following gaps in the existing literature: (*i*) OPEX CSFs are examined in relation to HFE, (*ii*) GDM is used to integrate the assessments of subject matter experts from different fields, and (*iii*) the FUCOM method, which has no precedent for use in OPEX research, is used for the first time with one-to-one correlation.

OPERATIONAL EXCELLENCE (OPEX)

OPEX is the persistent search for finding ways to improve efficiency and effectiveness across any business, despite of its size and nature (Antony, Sony, McDermott, Swarnakar, Galli, Doulatabadi & Kaul 2023).

Organizations can improve their products and processes by adopting OPEX initiatives, such as Lean and/or Six Sigma, to target long-term and sustainable growth (Antony, Swarnakar, Cudney & Pepper, 2021), although its successful implementation is not an easy task (Antony, Lizarelli &

Fernandes, 2020; Swarnakar, 2021). In the last three decades, OPEX methodologies have enabled businesses to increase productivity, financial savings, and customer satisfaction (Swarnakar, Singh, Antony, Tiwari, Cudney & Furterer, 2020).

Lean, Six Sigma, Lean Six Sigma, Agile, Kaizen, 5S, Total Quality Management could be listed as being heavily employed OPEX methodologies. With their ability to improve productivity, enhance product/service quality, improve customer experience and consequently aid organizational improvement, all types of OPEX methodologies, or initiatives as some literature instances called, have been selected and employed as proper to the industry in question. Out of all OPEX methodologies, Lean, Six Sigma, Lean Six Sigma (LSS) and Agile are three methodologies that are predominantly deployed in organizations (Antony et al., 2023; Antony, Sony, Jayaraman, Swarnakar, Tortorella, Garza-Reyes, Rathi, Gutierrez, McDermott & Lameijer, 2024; Naik, Sony, Antony, McDermott, Tortorella & Jayaraman, 2024;).

Commonly visited OPEX CSFs could be summarized as *organizational structure and culture, training and education, technical competence, effective communication, management/employee commitment, management/employee involvement, funds and resources, leadership, effective process monitoring, customer relation, supplier relation, clear objectives and goal identification, teamwork, personal capabilities, performance management systems, reward system, previous experience level*, and, *project prioritization and/or selection* (Soti et al., 2010; El Safty, 2012; Psomas, 2016; Marzagão and Carvalho, 2016; Antony et al., 2020; Antony et al., 2021; Barclay et al, 2021; Tanudiharjo et al., 2021; Yadav et al., 2021; Swarkarnar, 2021; Antony et al., 2023; Antony et al., 2024; Kharub, Sony, Gupta & Mcdermott, 2024; Ozbiltekin-Pala, Kazancoglu, Kumar, Garza-Reyes & Luthra, 2024; Edaki, 2025).

FUCOM

FUCOM was introduced to the MCDM literature by Pamucar, Stevic and Sremac (2018), and, is a criteria weighting methods employing decision maker assessments, hence, presents a handful of advantages while handling group decision making problems.

Unlike other decision-maker-centered multi-criteria decision-making methods (e.g. BWM, AHP, ANP, etc.), the FUCOM method calculates

criteria weights with complete consistency, rather than relying on calculations that accept discrepancies below a certain threshold, even if they don't guarantee complete consistency and accuracy, which can arise from other methods' pairwise comparison operations. In addition to these two major advantages, FUCOM also provides mathematically robust and reliable results through its unique computation capacity. The calculation mechanism of the FUCOM method relies on calculating criterion weights with mathematical accuracy based on a smaller number of pairwise comparisons and the solution of linear programming models representing these comparison data.

The calculation steps of the FUCOM method are explained below according to Pamucar, Stevic and Sremac (2018).

Step 1: FUCOM makes n-1 comparisons for n criteria. n criteria (C_j=C_1,C_2, …, C_n) are ranked from most important to least important by decision-maker(s) to provide the criteria rank regarding the importance of each one, as the first step. The criteria ranking is obtained according to the expected values of the weighting coefficients as shown in Equation (1). Here, k is the rank of the handled criterion, while if the decision maker considers certain criteria to be of equal importance, the "=" symbol can be used instead of ">".

$$C_{j(1)} > C_{j(2)} > \cdots C_{j(k)} \quad (1)$$

Step 2: In this step, the comparative priorities ($\varphi_{j/(k+1)}, k = 1,2,\ldots,n$) of the problem criteria whose importance rankings were determined by the decision maker are determined, and the comparative priority vector is obtained (Equation 2). Here, the comparative priority of the evaluation criteria ($\varphi_{j/(k+1)}$) represents the advantage of the kth ranked criterion to the (k+1)th ranked criterion.

$$\Phi = (\varphi_{1/2}, \varphi_{2/3}, \varphi_{3/4}, \ldots, \varphi_{k/(k+1)}) \quad (2)$$

The FUCOM method allows decision-makers to use integers, decimal values or the values from the predefined scale for the pairwise comparison of the criteria, meaning flexibility in criteria assessment as another benefit yielded by this method.

Step 3: The final importance weighting values (w_1, w_2, ⋯ w_n) of the criteria are calculated. The final values of criteria weighting coefficients should satisfy two conditions; (*i*) the ratio of the weighting coefficients is equal to the comparative priority among the observed criteria ($\varphi_{j/(k+1)}$), as it was defined in Step 2 (Equation 3); and, (*ii*) the final values of the weighting coefficients must ensure mathematical transitivity (Equation 4), where $\varphi_{k/(k+1)} x \varphi_{(k+1)/(k+2)}$ should be equal to $\varphi_{k/(k+2)}$, and, since $\frac{w_k}{w_{k+1}} = \varphi_{k/(k+1)}$ and $\varphi_{(k+1)/(k+2)} = \frac{w_{k+1}}{w_{k+2}}$, that $\frac{w_k}{w_{k+1}} \otimes \frac{w_{k+1}}{w_{k+2}} = \frac{w_k}{w_{k+2}}$, which leads to Equation (4) presented hereinafter.

$$\frac{w_k}{w_{k+1}} = \varphi_{k/(k+1)} \quad (3)$$

$$\frac{w_k}{w_{k+2}} = \varphi_{k/(k+1)} \otimes \varphi_{(k+1)/(k+2)} \quad (4)$$

Complete consistency is achieved if the conditions shown in Equation (3) and Equation (4) are satisfied. In other words, if these conditions are met, the deviation from full consistency (DFC) will be minimal. In this case, maximum consistency is achieved, and the DFC value calculated according to the computed criteria weighting coefficients will be $\chi = 0$. The final importance weights of the problem criteria will be obtained by solving the linear programming model represented in Equation (5) which depends on the constraints developed regarding the relationships represented in Equations (1) - (4).

$$\begin{aligned} &\min \chi \\ &\text{s.t.} \\ &\left| \frac{w_{j(k)}}{w_{j(k+1)}} - \varphi_{k/(k+1)} \right| \leq \chi, \quad \forall j \\ &\left| \frac{w_{j(k)}}{w_{j(k+1)}} - \varphi_{k/(k+1} \otimes \varphi_{(k+1)/(k+2)} \right| \leq \chi, \quad \forall j \\ &\sum_{j=1}^{n} w_j = 1, \ \forall j \\ &w_j \geq 0, \ \forall j \end{aligned} \quad (5)$$

Solving the linear programming model in Equation (5) calculates the final importance weights of the criteria (w_1, w_2, $\cdots$ w_n) and the degree of DFC (χ).

SCIENTIFIC LITERATURE REVIEW

An extant scientific literature analysis considering OPEX operations and problem dimension was researched separately emphasising the human factors engineering applications, whereas, the related literature of FUCOM applications in OPEX studies was also examined, however this method has a wide and rich literature of itself, unfortunately applications of FUCOM on the handled problem was found to be very restricted. Hence, the OPEX and multi-criteria decision making related literature is also dissected apart from the rest, to be able to analyse the employment of the proposed solution point-of-view to the handled kind of problems.

Existing OPEX literature can generally be grouped under headings such as investigated *sector types* (Swartz & Singh, 2022; Al Zaabi, Antony,

Arturo & Tortorella, 2023; Pirrone, Bernasconi & Friedli, 2023; Al Zaabi, Antony, Garza-Reyes, Tortorella, Sony & Jayaraman, 2025; Barros, Sousa & Dinis-Carvalho, 2025; Kandil, 2025; Kazancoglu, Garza-Reyes, Elgharbaoui & Bennasi., 2025; Kumar, Mittal, Verma & Anthony, 2025; Thomas, Sony, McDermott & Bhat, 2025; Sony et al., 2025; Bilancia, Costa, Amrani & Staudacher, 2026), *application organization dimensions* (Kourriche & Aboutafail, 2024; Larsson, Linderson, Birkie & Bellgran, 2025), *critical success factor analysis* (Köhler, Friedli and Basu, 2020; Antony et al., 2023; Kharub et al., 2024; Ozbiltekin-Pala et al., 2024; Edaki, 2025), *application barriers introduction* (Antony et al., 2023; Antony et al., 2024; Hajiagha et al., 2024; Joshi, Sharma, Bartwal, Joshi & Prasad, 2024; Shokri, Antony, Garza-Reyes & Cherrafi, 2025), *roadmap development based on overall application results* (Arndt, 2011; Sony & Naik, 2019; Carvalho, Sampaio, Rebentisch, McManus, Carvalho & Saraiva, 2023; Naik et al., 2024; Castillo & Merz, 2025; Madzik, Falat, Jayaraman, Sony & Antony, 2025; Swarnarkar & Khalfan, 2025), etc. However, it has been observed that topics e.g. the analysis of the impact of digitalization on OPEX applications, the link between OPEX applications and Industry 4.0, the analysis of the relationship between OPEX applications and ergonomic improvements are found to not be well-explained by these numerous valuable studies available in the current literature. However, it is clear that ergonomics, or human factors engineering, which aims to increase human productivity and improve system efficiency, interacts with OPEX applications in an intertwined manner. This study aims to fill this significant gap in the literature, create a new perspective on OPEX research, and finally, find robust and valid solutions using GDM techniques that minimize data loss with appropriate mathematical methods. Some recent studies performed under different OPEX application pillars addressed above and specifically created specific to the solution approach of this chapter were summarized hereinafter.

As some recent examples of the related literature grouped under the sectoral OPEX applications, Al Zaabi et al. (2023) analyzed energy industry and OPEX application in their review paper through 32 articles and identified applied OPEX methodologies, publication themes, trends, gaps, benefits, motivating factors and limitations of OPEX methodologies' application in the energy sector, and, mitigations for future research. Carvalho et al. (2023) investigated the relationship between OPEX and organizational performance through their previously proposed theoretical framework and appending empirical validation into it, by presenting the summary of ten industrial case studies. As key findings a better understanding of the balance needed to maintain high operational performance levels while dealing with change was addressed. Pirrone, Bernasconi and Friedli (2023) analysed OPEX applications in pharmaceutical companies and its synergy with digital transformation through 25 interviews with OPEX functions and digitalization experts over a

timespan of three years, to examine whether companies have adapted and applied the methodologies. In another study, Al Zaabi et al. (2025) examined energy industry in terms of implemented OPEX methodologies, motivation drivers, deployment approaches, sustainability factors, benefits and challenges in Oman with the proposed qualitative analysis methodology, involving interviews with 18 energy industry experts from variosly sized companies. Swarnakar and Khalfan (2025) investigated the relationship between hard and soft OPEX applications within ten manufacturing organizations across diverse industries to analyze their impact on corporate performance levels via qualitative research performed through literature reviews, semi-structured interviews, organizational websites, internal documents, newspaper articles, social reports, relevant conference presentation booklets, business press and annual reports. A cross-case comparison approach was employed using a 5-point Likert scale to analyze the data obtained from 45 semi-structured interviews with senior managers, literature findings and secondary resources. Barros, Sousa and Dinis-Carvalho (2025) compared the most referenced OPEX models for education industry in terms of adoption in University Research Centres in their review paper. They performed a Systematic Literature Review (SLR) to consolidate knowledge about various OPEX models in University Research Centres applications. Kandil (2025) evaluated Lean Six Sigma 4.0 integration strategies as an OPEX application in tourism industry in Egypt, Tunisia and Morocco. The intuitive fuzzy (IF) hybrid average and geometric operator was employed as a novel way to tackle the instability caused by the IF hybrid average operator to calculate criteria weights and evaluate three alternative countries, where, a systematic literature review of the relevant works were also performed to form the problem space.

The heading covering the recent instances from the existing OPEX literature concentrated on the enlightment of the critical success factors or the barriers and challenges, or both, are presented hereinafter. As the studies investigated the critical success criteria, Köhler Fiedli and Basu (2020) analysed OPEX enablers for quality control labs in pharmaceutical quality systems, based on a database of the University of St.Gallen comprising 53 pharmaceutical quality control labs via independent t-test and multiple linear regressions. The study results show that 0,35 of the variances in lab performance can be explained by enabler implementation which means highly effective quality control labs exhibit significantly higher OPEX enabler implementation. In another study, Antony et al. (2023) analyzed the critical success factors for OPEX applications in manufacturing organizations in their review paper. They analyzed thirty studies and presented that OPEX implementation in manufacturing organizations is positively and strongly influenced by employee involvement, training, education, management, and required resources, where; it is moderately influenced by leadership, project prioritization, organizational culture,

effective communication, customer relation, supplier relation, performance management, rewarding system, and teamwork. As examples focused on barriers and challenges faced, Thomas et al. (2025) examined OPEX methodologies applications in ophthalmology and the barriers and enablers influencing this implementation. They interpreted patterns and gaps according to the reported 25 peer-reviewed case studies that implemented OPEX strategies in ophthalmology. As the results indicated, OPEX methodologies in ophthalmology Lean is the dominant methodology, where, in overall applications primarily focused on reducing surgical complications (Six Sigma) and optimising patient flow (Lean). Bilancia et al. (2026) analyzed OPEX applications in luxury fashion companies through their proposed qualitative approach based on a series of interviews with three French and Italian luxury fashion companies to understand the distinctive characteristics of the luxury fashion supply chain and provide insights to the foundation of a structured performance framework.

Under the recent instances of the related literature grouped as country specific studies, Kourriche and Aboutafail (2024) investigated Morocco market through 40 firms operating in different industries to assess the level of maturity of OPEX dimensions and benchmarked Small-Medium (SMEs) and Large (LEs) Enterprises. Doyer and Odendaal (2025) aimed on to identification of research and implementation gaps of OPEX applications in African continent. They analyzed ten integrated OPEX approaches with a systematic eight-step process and screened 353 publications through PRISMA-SCR checklist. Their review study indicated that total quality management (20.8%), business process reengineering (12.9%) and Lean Six Sigma (12.9%) were the most prevalent approaches where Six Sigma and total productive maintenance were the least prevalent ones. They also analyzed the industry type and country distribution as a novel synthesis of OPEX applications in Africa. In another study, Kazancoglu et al. (2025) investigated OPEX implementation within the Moroccan automotive industry with a qualitative, exploratory design using semi-structured interviews with eight key experts from Moroccan automotive industry. The study results showed that Moroccan automotive firms adopt OPEX practices in a phased, structured manner, beginning with foundational methodologies such as Lean manufacturing, 5S and Kaizen, before progressing to advanced tools like Six Sigma.

As another pillar of the wide literature of the handled subject, the heading of the sustainability and OPEX relation analysis has also an emerging importance. As an instance grouped under this heading, Naik et al. (2024) developed a framework to assess the sustainability of OPEX initiatives in an organisation based on operational, economic, environmental, and social pillars of sustainability concept and identified the enablers for the sustainability of OPEX initiatives. In another study, Castillo and Merz (2025) proposed a multidimensional model to evaluate OPEX

methodologies' sustainability in manufacturing environments including production and intralogistics processes, which appended a fourth sustainability dimension to the known-three-pillars as governance. Shokri et al. (2025) investigated the sustainability subject and the interaction and prioritisation of barriers to green deployment of OPEX applications in manufacturing sector, through fuzzied versions of different multi-criteria decision-making techniques to model the interaction and prioritisation of analyzed barriers.

The studies employed multi-criteria decision-making methods to characterise and investigate the handled problem space in scientific OPEX analysis are presented, hereinafter. As a recent instance, Ozbiltekin-Pala et al. (2024) analyzed critical factors to strategic alignment of OPEX and Industry 4.0 technologies for manufacturing industries with the Analytical Hierarchy Process (AHP) and the Grey Relational Analysis (GRA) methods. Pansare, Yadav and Nagare (2024) investigated the relationship among Industry 4.0 technologies, OPEX strategies and reconfigurable manufacturing system practices to identify and prioritize the performance metrics, where, the Stepwise Weight Assessment Ratio Analysis (SWARA) was used to compute the selected practices' weights, while the Weighted Aggregated Sum Product Assessment (WASPAS) method was used to prioritize the performance metrics. Kumar et al. (2025) explored OPEX achievements via digital technologies in food industry. They performed a systematic literature review to create the problem space and the AHP method to determine the relative importance of the primary factors impacting digital technology employment for OPEX achievement. However, as a robust solution method FUCOM has its own wide application area in various application fields, the application of this novel method to OPEX studies is very narrow. To the best of our knowledge, there is only one instance from the related literature. Kanik, Eriskan, Soysal and Ömürgönülsen (Early access) handled OPEX in last-mile delivery processes and employed IFS, FUCOM and COPRAS methods to prioritise feasible strategies built on 27 criteria classified under SWOT dimensions. While this study offers valuable solutions for last-mile delivery, its approach and correlation to the OPEX literature is limited, yet, it is valuable as the first study in the OPEX-literature to provide a solution using the FUCOM method. The in-depth literature review shows that since there is no other study that relates OPEX criteria to human factors and engineering pillars, and, addresses them using the FUCOM method, this chapter will make valuable contributions to fill a significant gap of the related literature.

REAL WORLD APPLICATION

In this section, OPEX CSFs are contemplated in HFE perspective, in this means, their suitability to be achieved through HFE applications via e.g. method study, standard time determination, cognitive ergonomics enhancement, physical ergonomic arrangements, etc. without the need for any additional implementation, i.e. new cost centers, is evaluated by GDM according to the opinions of decision-making team (DMT) consisting of five field experts. With that regard, CSFs that link OPEX principles with HFE applications will be assayed in line with the core of OPEX where lies the philosophy of cost reduction without any extra investments and best allocation of sources to maximize the utilization ratios.

Five DMs to represent differing aspects of HFE and OPEX applications were selected (Table 1) to assess the CSFs and validate the identified problem criteria regarding the extant literature review presented in the previous section. Eight criteria were determined to dissect the handled problem space, which are "C_1: *Ergonomic Work Design* (OPEX correspondence in waste reduction and process stability): Ergonomic workstations designed according to anthropometric data, movement economy principles, and optimal workload distribution reduce unnecessary physical effort, lower cycle times, and improve productivity, which also strengthens operational flow.", "C_2: *Error Prevention Mechanisms* (OPEX correspondence in defect reduction and quality improvement): Visual cues, colour coding, standardized layouts, warning systems, and error-proofing (poka-yoke) principles reduce human errors and process variability. These HFE strategies directly enhance quality performance and contribute to zero-defect operations.", "C_3: *Cognitive Workload Management & Reduction of Task Complexity* (OPEX correspondence in process reliability and operational efficiency): High cognitive workload decreases decision accuracy and increases error probability. HFE methods e.g. task simplification, cognitive task analysis, interface optimization, etc. improve reliability and reduce processing time.", "C_4: *Standardized Work & User-Centered Procedure Design* (OPEX correspondence in standardization and variation reduction): When work procedures are designed based on user behaviour, cognitive capabilities, and ergonomic principles, adherence increases and variability decreases, which reinforces OPEX objectives by ensuring consistency and predictable process outcomes.", "C_5: *Competency/Safety Trainings* (OPEX correspondence in continuous improvement and zero-defect mindset): Human reliability improves when employees receive ergonomics-based training, scenario-based learning, and behavioral safety programs. Enhanced competency supports sustained OPEX performance and fosters a culture of improvement.", "C_6: *Human Machine Interaction & Risk Analysis* (OPEX correspondence in reliability and downtime reduction): HFE contributes to safer machine interfaces, improved

alarm management, and intuitive controls, hence, reducing safety-related incidents minimizes downtime, enhances equipment availability, and increases overall operational reliability.", "*C_7: Employee Involvement* (OPEX correspondence in Kaizen and continuous improvement): Participatory ergonomics and structured feedback mechanisms empower employees to contribute to process design and improvement. This engagement enhances organizational learning and strengthens the root-cause analysis culture central to OPEX.", "*C_8: Fatigue Management* (OPEX correspondence in performance consistency and quality stability): Fatigue negatively affects reaction time, decision-making, and musculoskeletal performance, hence, ergonomically designed shift schedules, rest cycles, and balanced task allocation support consistent operational performance and reduce variation."

Table 1: DM Demographics

Expertise	Title	Background	Experience(years)
Manufacturing	Chief	Mechanical Engineering (Ph.D.)	20
HFE	Academician	Industrial Engineering (Ph.D.)	19
OHS	Expert	Chemical Engineering	5
Manufacturing	Expert	Mechanical Engineering	8
Logistics	Chief	Industrial Engineering	15

DMT performed the CSFs comparison based on the scale [1, 9] (Pamucar et al., 2018) using Equation (1) – Equation (4), where, Table 2 presents the criteria rank and comparative priority values established according to each DM, hereinafter.

Table 2: DMs Assessments

	DMs	C1	C2	C3	C4	C5	C6	C7	C8
Criteria Rank	DM1	4	1	5	8	7	3	6	2
	DM2	1	5	6	2	7	4	8	3
	DM3	4	3	7	6	5	1	8	2
	DM4	2	1	6	3	7	4	8	5
	DM5	5	6	7	1	4	3	8	2
Comparative Priority	DM1	5	1	6	7	7	5	6	4
	DM2	1	6	7	4	7	5	8	4
	DM3	5	4	7	6	6	1	8	4
	DM4	3	1	6	3	6	4	8	5
	DM5	1	5	5	6	6	7	8	9

DMT assessments presented in Table 2 then are employed to constitute particular mathematical models (Equation 5) to calculate the final importance weights of problem criteria according to each DM which will be

depended on the constraints developed regarding the relationships represented in Equations (1) - (4). To explain the computation mechanism of FUCOM more preciously, calculations made regarding Equation (1) – (4) considering the comparison values of DM1 presented in Table 2 to constitute the linear programming model regarding Equation (5) are respectively presented in Equation (6), Equation (7) and Equation (8), hereinafter.

$$\varphi_{\frac{C_2}{C_8}} = 4; \quad \varphi_{\frac{C_8}{C_6}} = \frac{5}{4}; \quad \varphi_{\frac{C_6}{C_1}} = \frac{5}{5};$$
$$\varphi_{\frac{C_1}{C_3}} = \frac{6}{5}; \quad \varphi_{C_3/C_7} = \frac{6}{6}; \quad \varphi_{C_7/C_5} = \frac{7}{6}; \quad \varphi_{C_5/C_4} = \frac{7}{7} \tag{6}$$

$$\frac{w_2}{w_8} = 1; \quad \frac{w_8}{w_6} = 1{,}25; \quad \frac{w_6}{w_1} = 1; \quad \frac{w_1}{w_3} = 1{,}2;$$
$$\frac{w_3}{w_7} = 1; \quad \frac{w_7}{w_5} = 1{,}167; \quad \frac{w_5}{w_4} = 1 \tag{7}$$

$$\frac{w_2}{w_6} = 1 * 1{,}25; \quad \frac{w_8}{w_1} = 1{,}25 * 1; \quad \frac{w_6}{w_3} = 1 * 1{,}2;$$
$$\frac{w_1}{w_7} = 1{,}2 * 1; \quad \frac{w_3}{w_5} = 1 * 1{,}167; \quad \frac{w_7}{w_4} = 1{,}167 * 1 \tag{8}$$

The final importance weights of the problem criteria will be obtained by solving the linear programming models for each DM assessment values presented in Table 2. The linear programming model constituted to compute the final importance weighting values regarding DM1 assessments is represented in Equation (9) hereinafter, to provide an instance for the readers.

$$\min \chi$$

s.t.

$$\left|\frac{w_2}{w_8} - 1\right| \leq \chi \quad \left|\frac{w_8}{w_6} - 1{,}25\right| \leq \chi \quad \left|\frac{w_6}{w_1} - 1\right| \leq \chi$$
$$\left|\frac{w_1}{w_3} - 1{,}2\right| \leq \chi \quad \left|\frac{w_3}{w_7} - 1\right| \leq \chi \quad \left|\frac{w_7}{w_5} - 1{,}167\right| \leq \chi$$
$$\left|\frac{w_6}{w_3} - 1{,}2\right| \leq \chi \quad \left|\frac{w_2}{w_6} - 1{,}25\right| \leq \chi \quad \left|\frac{w_8}{w_1} - 1{,}25\right| \leq \chi$$
$$\left|\frac{w_1}{w_7} - 1{,}2\right| \leq \chi \quad \left|\frac{w_5}{w_4} - 1\right| \leq \chi \quad \left|\frac{w_3}{w_5} - 1{,}167\right| \leq \chi$$
$$\left|\frac{w_7}{w_4} - 1{,}167\right| \leq \chi$$
$$\sum_{j=1}^{8} w_j = 1, \quad w_j \geq 0, \ \forall j \tag{9}$$

The importance levels regarding eight CSFs in the criteria set were determined in terms of the optimal weighting scores (Table 3) calculated through the solutions of five different linear programming models , which were constituted separately according to particular DM assessments presented in Table 2. The weighting values, the DFC values (χ), and, the overall weighting scores were calculated using MS Excel macro software features. The importance levels, particular weighting values, and overall weighting scores related to problem criteria were computed and presented in Table 3, hereinafter. The overall weighting scores were computed by taking the arithmetic mean of the weight values calculated particular to each DM assessment.

Table 3: FUCOM results: Weighting Values (W_i), Overall Weighting Scores (OWC), DFC (χ) Values

W_i	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	Sum	χ
W_1	0,088	0,441	0,073	0,063	0,063	0,088	0,073	0,110	1,000	0,000
W_2	0,439	0,073	0,063	0,110	0,063	0,088	0,055	0,110	1,000	0,000
W_3	0,087	0,109	0,062	0,072	0,072	0,435	0,054	0,109	1,000	0,000
W_4	0,129	0,388	0,065	0,129	0,065	0,097	0,049	0,078	1,000	0,000
W_5	0,079	0,068	0,059	0,473	0,079	0,095	0,053	0,095	1,000	0,000
OWC	0,165	0,216	0,064	0,170	0,068	0,160	0,057	0,100	1,000	

RESULTS AND DISCUSSION

Five DM profiles from different areas of expertise were included in the study to reflect a different perspective in the solution space for evaluating OPEX CSFs. In this context, it is observed that the priority order of HFE arrangements that will provide OPEX CSFs, determined as a result of the evaluations of the DMs, has changed (Table 2, Table 3, Figure 1).

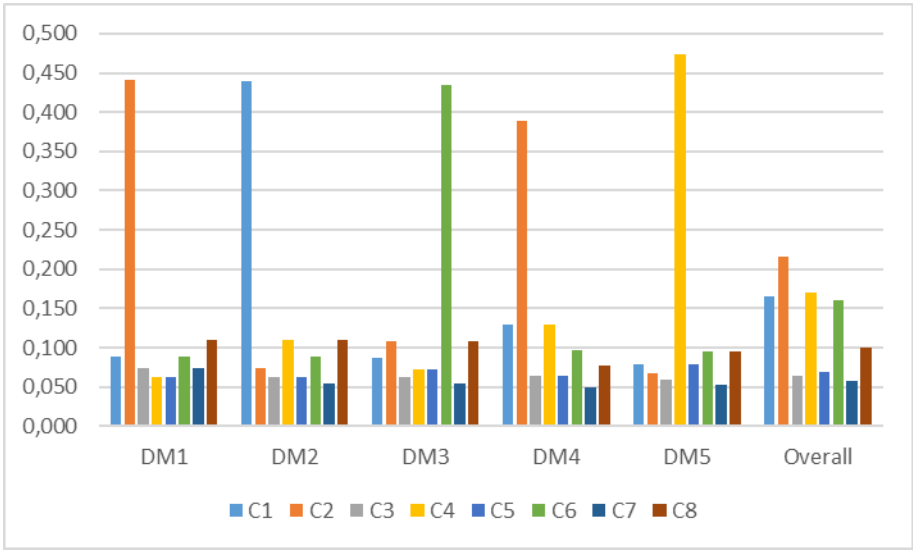


Figure 1: HFE-based OPEX CSF importance value trade-offs

"*C_2: Error Prevention Mechanisms*" was identified as to be considered first among all HFE-related OPEX CSFs according to DM1 and DM4, while, the CSFs determined to be the first to be planned in operational strategies to be provided emergently due to their high priority levels were identified as "*C_1: Ergonomic Work Design"* according to DM2, "*C_6: Human Machine Interaction & Risk Analysis"* according to DM3, and, "*C_4: Standardized Work & User-Centered Procedure Design"* according to DM5, respectively, for the other DMs (Table 3, Figure 1, Figure 2).

The research results showed that the deviation values which are denoted as DFC (χ) values were calculated as 0 for all assessments performed using the FUCOM method (Table 3). This shows that the inconsistency rate in the calculations is zero and completely reliable results have been obtained, as one of the biggest benefits provided by FUCOM method. According to the research results, the final priority ranking of HFE-based OPEX CSFs was found to be "*C_2: Error Prevention Mechanisms*, "*C_4: Standardized Work & User-Centered Procedure Design", "C_1: Ergonomic Work Design"*, "*C_6: Human Machine Interaction & Risk Analysis"*, "*C_8: Fatigue Management"*, "*C_5: Competency/Safety Trainings"*, "*C_3: Cognitive Workload Management & Reduction of Task Complexity"*, and "*C_7: Employee Involvement"* (Table 3, Figure 1, Figure 2). The fluctuation in the priority level of CSFs is presented in Figure 2, compiled from numerical result values and analysis assessment data synthesized according to DM evaluations, hereinafter.

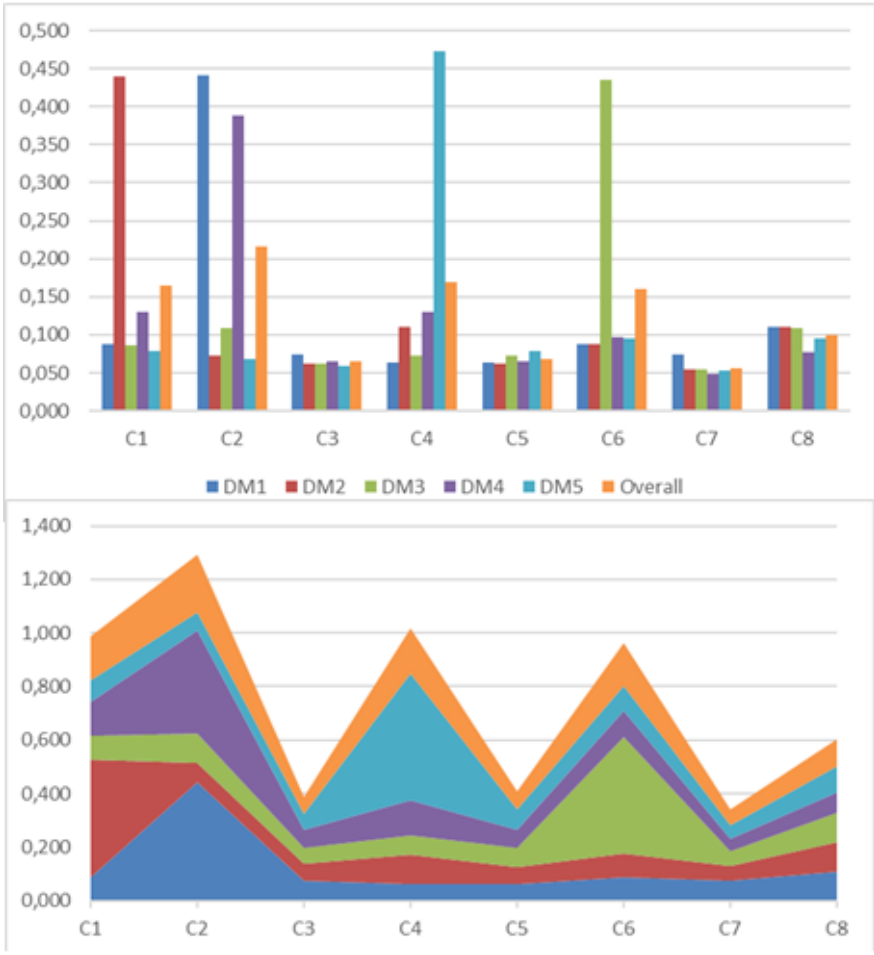


Figure 2: Overall Criteria Importance Benchmarking for Particular DM Perspectives

As limitation of this study all DM profiles were handled equally which means that the weights of the DM assessments were assumed to have an equal influence on the outcome. Sensitivity analysis might be performed by changing the DM impacts through assigning differing weight scores to particular assessments. Furthermore, this study examined OPEX CSFs based on HFE in the perspective of general purpose of use. As a suggestion for future studies, different results which might be obtained by examining sector-specific assessment data and problem input sets might be analyzed.

REFERENCES

Al Zaabi, Y., Antony, J., Garza-Reyes, J. A., & Tortorella, G. (2023). Operational excellence methodologies in the energy sector: A systematic literature review. Total Quality Management & Business Excellence, 34(9-10), 1173-1195.

Al Zaabi, Y., Antony, J., Garza-Reyes, J. A., Tortorella, G. D. L., Sony, M., & Jayaraman, R. (2025). Unlocking the potential: a study on the role of operational excellence in Oman's energy sector. International Journal of Quality & Reliability Management, 42(1), 157-182.

Antony, J., Lizarelli, F. L., & Fernandes, M. M. (2020). A global study into the reasons for Lean Six Sigma project failures: Key findings and directions for further research. IEEE Transactions on

Engineering Management, https://doi.org/10.1109/TEM.2020.3009935.

Antony, J., Swarnakar, V., Cudney, E., & Pepper, M. (2021). A meta-analytic investigation of lean practices and their impact on organizationalnal performance. Total Quality Management & Business Excellence, 1–27.

Antony, J., Sony, M., McDermott, O., Swarnakar, V., Galli, B., Doulatabadi, M., & Kaul, R. (2023). An empirical study into the reasons for failure of sustaining operational excellence initiatives in organizations. The TQM Journal, 35(7), 1569-1587.

Antony, J., Swarnakar, V., Gupta, N., Kaur, J., Jayaraman, R., Tortorella, G. L., & Cudney, E. (2023). Critical success factors for operational excellence initiatives in manufacturing: a meta-analysis. Total Quality Management & Business Excellence, 34(9-10), 1152-1172.

Antony, J., Sony, M., Jayaraman, R., Swarnakar, V., Tortorella, G. D. L., Garza-Reyes, J. A., rathi, R., Gutierrez, L., McDermott, O. & Lameijer, B. A. (2024). Global perspectives on operational excellence: unveiling critical failure factors and sustainable pathways. International Journal of Quality & Reliability Management, 41(10), 2604-2626.

Arndt, M. (2011). Proficiency testing as an essential precondition for operational excellence. The 3rd International Proficiency Testing Conference - PTconf 2011, Iasi, Romania, Sep 27-30.

Barros, A., Sousa, R.M., Dinis-Carvalho, J. (2025). A Study on the Adoption of Operational Excellence Principles in the Organization and Management of University Research Centres. In: McDermott, O., et al. Advances in Operational Excellence in the Higher Education Sector. ILSSforHEI 2024. Lecture Notes in Management and Industrial Engineering. Springer, Cham. https://doi.org/10.1007/978-3-031-84816-2_7

Bilancia, A., Costa, F., Zouggar Amrani, A., Portioli Staudacher, A. (2026). Key Characteristics of Luxury Fashion Supply Chain and Impact on Operational Performance: A Qualitative Study. In: Mizuyama, H., Morinaga, E., Nonaka, T., Kaihara, T., von Cieminski, G., Romero, D. (eds) Advances in Production Management Systems. Cyber-Physical-Human Production Systems: Human-AI Collaboration and Beyond. APMS 2025. IFIP Advances in Information and Communication Technology, vol 767. Springer, Cham. https://doi.org/10.1007/978-3-032-03542-4_35

Carvalho, A. M., Sampaio, P., Rebentisch, E., McManus, H., Carvalho, J. Á., & Saraiva, P. (2023). Operational excellence, organizational culture, and agility:

bridging the gap between quality and adaptability. Total Quality Management & Business Excellence, 34(11-12), 1598-1628.
Castillo, G. C. S., & Merz, C. K. Multidimensional model for assessment sustainability of OPEX tools. 2025 Conference on Production Systems and Logistics-CPSL-Annual, Lima, Peru, Mar 18-21.
Doyer, I., & Odendaal, P. (2025). The operational excellence research landscape in Africa: a scoping review. International Journal of Lean Six Sigma. Early Access, https://doi.org/10.1108/IJLSS-07-2024-0167.
Edaki, L. (2025). Critical Success Factors of Lean Six Sigma in Higher Education Institutions. In: McDermott, O., et al. Advances in Operational Excellence in the Higher Education Sector. ILSSforHEI 2024. Lecture Notes in Management and Industrial Engineering. Springer, Cham. https://doi.org/10.1007/978-3-031-84816-2_21
Hajiagha, S. H. R., Mahdiraji, H. A., Moradi, S., Garza-Reyes, J. A., & Alaei, S. (2024). Unveiling the relation between the challenges and benefits of operational excellence and industry 4.0: A hybrid fuzzy decision-making approach. The TQM Journal, 36(1), 51-70.
Hammer, M. (2004), Deep Change: How Operational Innovation Can Transform Your Company. Harvard Business Review, 82(4), 84-93.
Joshi, S., Sharma, M., Bartwal, S., Joshi, T., & Prasad, M. (2024). Critical challenges of integrating OPEX strategies with I4. 0 technologies in manufacturing SMEs: A few pieces of evidence from developing economies. The TQM Journal, 36(1), 108-138.
Kandil, T. (2025). Integrated Lean Six-Sigma 4.0 operational excellence: a fuzzy social networks hybrid MCDM model to assess tourism organisations' resilience. International Journal of Lean Six Sigma. Early Access, https://doi.org/10.1108/IJLSS-07-2024-0168.
Kanık, Z. B., Erişkan, S., Soysal, M., & Ömürgönülşen, M. (2025). A hybrid approach based on qualitative and quantitative techniques for analyzing last-mile parcel delivery. Opsearch, Early Access, https://doi.org/10.1007/s12597-025-01017-6.
Kazancoglu, Y., Garza-Reyes, J. A., Elgharbaoui, M., & Abbana Bennani, C. (2025). An exploratory study on the practice of operational excellence in the automotive industry in Morocco. International Journal of Lean Six Sigma. Early acces, https://doi.org/10.1108/IJLSS-06-2024-0139.
Kharub, M., Sony, M., Gupta, H., & McDermott, O. (2024). Operational excellence in African markets: unveiling its drivers, processes and impacts through a mixed-methods approach. International Journal of Lean Six Sigma. Early Access. https://doi.org/10.1108/IJLSS-02-2024-0020.
Kourriche, I., & Aboutafail, M. O. (2024). Assessing the level of maturity of operational excellence in Morocco: a comparative study between SMEs and LEs. Engineering, Technology & Applied Science Research, 14(1), 12812-12816.
Köhler, S., Friedli, T., & Basu, P. (2020). Operational excellence in pharmaceutical quality control labs: driver of an effective quality system. Journal of Pharmaceutical Innovation, 15(1), 138-154.
Kumar, V., Mittal, A., Verma, P., & Anthony, J. (2025). Enhancing OPEX in the food industry: integrating digital technologies for sustainable transformation.

Industrial Management & Data Systems, 1-34. Early Access. https://doi.org/10.1108/IMDS-01-2025-0029.
Larsson, J., Linderson, S., Birkie, S. E., & Bellgran, M. (2025). Developing capabilities for operational excellence in multisite production companies. Production & Manufacturing Research, 13(1), 2566920.
Madzík, P., Falát, L., Jayaraman, R., Sony, M., & Antony, J. (2025). Exploring research trends in Lean, Six Sigma and Lean Six Sigma methodologies through a hybrid artificial intelligence approach. Production Planning & Control, 1-25.
Naik, S., Sony, M., Antony, J., McDermott, O., Tortorella, G. L., & Jayaraman, R. (2024). Operational excellence framework for sustainability in the organisation: a design science approach. Production Planning & Control, 35(11), 1215-1231.
Ozbiltekin-Pala, M., Kazancoglu, Y., Kumar, A., Garza-Reyes, J. A., & Luthra, S. (2024). Analyzing critical factors of strategic alignment between operational excellence and Industry 4.0 technologies in smart manufacturing. The TQM Journal, 36(1), 161-177.
Pirrone, L., Bernasconi, M., & Friedli, T. (2023). Investigating The Role Of Digitalization In OPEX Programs–A Case Study From The Pharmaceutical Industry. 4th Conference on Production Systems and Logistics (CPSL), Tech Monterrey, Queretaro, MEXICO, FEB 28-MAR 02.
Shokri, A., Antony, J., Garza-Reyes, J. A., & Cherrafi, A. (2025). Interaction and prioritisation of barriers to green deployment of operational excellence projects in the manufacturing sector. International Journal of Quality & Reliability Management. Early Access. https://doi.org/10.1108/IJQRM-08-2024-0250.
Sony, M., & Naik, S. (2019). Six Sigma with CK theory for innovations in operational excellence: a case study. Benchmarking: An International Journal, 26(7), 2105-2121.
Sony, M., Ramadan, M. A., Antony, J., Al Dhaheri, M. K., McDermott, O., & Cudney, E. A. (2025). Assessing ISO 18404 standard applicability in the service sector: a qualitative study. The TQM Journal, 37(3), 726-746.
Swarnakar, V., Singh, A. R., Antony, J., Tiwari, A. K., Cudney, E., & Furterer, S. (2020). A multiple integrated approach for modelling critical success factors in sustainable LSS implementation. Computers & Industrial Engineering, 150, Article 106865.
Swarnakar, V. (2021). Assessment of critical failure factors for implementing Lean Six Sigma in manufacturing industry: A case study. International Journal of Industrial Engineering, 3(1), 1–17.
Swarnakar, V., & Khalfan, M. (2025). Unveiling the impact of operational excellence practices on corporate performance: an exploratory research. The TQM Journal. Early Access. https://doi.org/10.1108/TQM-06-2024-0224.
Swartz, B., & Singh, S. (2022). A tale of three regions: quality culture and its impact on pharmaceutical production. South African Journal of Industrial Engineering, 33(3), 137-149.
Thomas, A., Sony, M., McDermott, O., & Bhat, S. (2025). Exploring operational excellence in ophthalmic services: a scoping review and future research directions. International Journal of Quality & Reliability Management. Early Access. https://doi.org/10.1108/IJQRM-05-2025-0156.

Industrial Energy Management: ISO 50001, Energy Auditing, and a PDCA-Based Implementation Approach[1]

Eren UYAN[2]

1- This chapter is derived from the author's PhD thesis conducted at Newcastle University, UK. The content has been revised, updated, and adapted to the book chapter format.

2- Assist. Prof.; Yalova University, Altınova Vocational School of Higher Education, . eren.uyan@yalova.edu.tr ORCID No: 0000-0002-0861-5240

ABSTRACT

Energy management is increasingly recognized as a cornerstone of low-carbon manufacturing, enabling industrial organizations to reduce energy costs, improve competitiveness, and mitigate greenhouse gas (GHG) emissions. This chapter presents a structured roadmap that links energy management practices with the implementation of an Energy Management System (EnMS), emphasizing ISO 50001 and its Plan-Do-Check-Act (PDCA) cycle. The chapter clarifies the roles of leadership commitment, energy policy, energy teams, energy planning, and performance monitoring in building an organizational culture of continuous improvement. A central focus is placed on the technical and managerial function of energy auditing as a prerequisite for establishing credible energy baselines and energy performance indicators, and for identifying, quantifying, and prioritizing energy saving measures. The discussion distinguishes between a plant energy audit aimed at uncovering efficiency opportunities and an internal EnMS audit aimed at verifying conformance and effectiveness of the management system. In addition, the chapter outlines typical audit processes and classifies audits into walk-through and detailed levels, highlighting how audit depth affects data needs, accuracy of savings estimates, and decision readiness. Overall, the chapter provides practitioners and researchers with an integrated perspective on how energy audits can be embedded into ISO 50001-aligned management routines to sustain long-term energy, cost, and emission reductions.

Keywords –Energy Management; Energy Audit; ISO 5000; PDCA; Energy Efficiency, Continuous Improvement.

INTRODUCTION

As a result of the growing global concerns about climate change, stronger environmental policies, and rising energy costs, energy management has become a strategic priority for energy-using sectors. Industrial plants are among the largest energy consumers and emitters of greenhouse gases (GHG) worldwide, and thus they present opportunities to improve energy efficiency and reduce GHGs (Sola & Mota, 2020). The activities in industrial sectors involve extensive application of technology and complex processes, requiring knowledge and structure reliant on organisational management practices. As such, achieving energy efficiency in industry is a challenging undertaking, requiring a systematic approach and management of energy use (Schulze et al., 2016; Sola & Mota, 2020). Implementing an energy management within an organisation is one of the most effective approaches for reducing energy consumption, associated

emissions and associated costs (Schulze et al., 2016). The aim of this chapter is to provide a practical and structured understanding of industrial energy management and its implementation by defining its main building blocks, including energy management systems (EnMS), ISO 50001, and energy auditing. Furthermore, the chapter explains the interrelationships between these elements within a PDCA-based continuous improvement framework.

ENERGY MANAGEMENT

Energy management can be defined as a combination of energy practices, efficiency techniques, and management of related processes aimed at reducing energy use and GHGs emissions of a company (Amundsen, 1999; Kannan & Boie, 2003). Alternatively, it can be referred to as:

"*the proactive, organised and systematic coordination of a company`s use of energy to meet the requirements, taking into account environmental and economic objectives* (IMO, 2011)*"*

Energy management assists a plant to reduce energy costs by means of improved energy performance and optimised use of energy-related assets and energy sources (DOE, 2016a). Therefore, implementing some form of energy management is a primary step towards saving energy and energy costs, reducing GHG emissions, and staying competitive in the market (DOE, 2016a). Implementation of any technical and management practices which will save energy and its cost and reduce GHGs emissions in an organisation can be regarded as energy management. According to a research carried out by (Caffal, 1995), industries can save up to 40% of total energy use by adopting energy management practises.

There can be different plans or methods to develop an energy management program in an organisation. Also, there can be various means to improve the energy performance of a facility or site. This can be by means of using more energy efficient equipment, switching to low carbon energy sources, avoiding the excess energy consumption which is determined to be unnecessary in a process or system, or changing power consumption patterns within the industrial site which is called as demand response techniques. The appropriate means and associated energy saving potentials, which will increase the energy performance of the organization, can be identified through conducting an energy audit within the plant. Once they are identified, quantified, and ensured that they are technically feasible (e.g. do

not cause any disruption on the production), they can be prioritised for implementation with regards to various criteria such as economic and environmental merits. Thereafter, an action plan can be prepared to implement the selected energy saving measures.

At the heart of Energy Management is continuous improvement which means that energy management benefits should be continuous, rather than one-time results (DOE, 2016a). After implementing the actions which will improve the energy performance of the organisation, it is of importance to ensure that the implemented actions are sustainable. A plant should proactively and strategically manage energy use across the plant to ensure obtained benefits are continuous and long term (DOE, 2016a). Otherwise, the first initiative for energy management and effort spent in the energy audit can fail. Therefore, energy management should be a part of the culture of an organisation such as quality and safety. In this regard, an EnMS is essential to maintain the success of the implemented energy saving actions as well as continuous improvement is ensured.

ENERGY MANAGEMENT SYSTEMS

An EnMS can be defined as (DOE, 2016b):

an interacting series of processes that enables an organization to systematically achieve and sustain energy management actions and energy performance improvements. It provides the processes and systems needed to incorporate energy considerations and energy management into daily operations as part of an organizational strategy for continually improving energy performance.

Therefore, an EnMS establishes the structure and discipline to implement technical and management strategies which lead to substantial energy, cost, and GHGs emission reductions and to sustain those savings over time (ISO, 2018). By employing an EnMS, energy management becomes a part of the organisational culture of a company and ongoing management of the major energy uses and compliance with energy related legal and other requirements are sustained (DOE, 2016a).

It establishes clear responsibilities, documented procedures, ongoing training, internal audits for conformance, corrective and preventive action, management reviews, and continual improvement (Kannan & Boie, 2003).

The major benefits of implementing an EnMS are summarised as follows (DOE, 2016a; SEAI, 2015):

- It integrates energy management culture in day-to-day operations,
- It ensures that senior managers commit to energy efficiency and that all staff play a role in the process; thus, energy management becomes an integrated business approach rather than single initiatives,
- It enables the plant to become proactive rather than reactive in dealing with energy issues,
- It provides a systematic framework for energy improvement, and it standardises processes so that improvements are sustained over time,
- It ensures that a process of continual improvement is sustained,
- It helps a plant comply with energy efficiency and emission reduction obligations,
- It reduces energy costs and reduces the risk to energy price fluctuations.

In addition to the above, a company can optimise its industrial systems and improve overall monitoring of system inefficiencies by implementing an EnMS (Nulty, 2015). Thanks to the after-effects of this such as enhanced production and capacity utilization, reduced pollution and resource use, and lower operation and maintenance costs, implementing an EnMS increases the competitiveness of the organisation (Nulty, 2015). Taking these and the above-stated benefits into account, it is essential to implement an EnMS for a successful energy management program within an organisation.

An EnMS can be implemented according to either available ISO 50001 energy management standard or a custom EnMS (Nulty, 2015). ISO 50001 has been adopted by many countries as their national EnMS standards; therefore, organisations who wishes to implement an energy management program within their facilities can follow the structured framework of ISO 50001. This will also help them fulfilling the requirements to be certified by ISO 50001. A recent study by (McKane et al., 2017) estimates that in 2030 ISO 50001 will result in about 16 EJ of annual primary energy savings, and 1000 Mt of avoided annual CO_2 energy savings given that a 50% uptake level of ISO 50001 in the industrial and commercial sectors by 2030 is realized.

ISO 50001 AND THE PDCA FRAMEWORK

ISO 50001is a voluntary international standard developed by the International Organisation for Standardization (ISO) that specifies requirements to establish and implement an EnMS. The ISO 50001 follows the Plan-Do-Check-Act (PDCA) management cycle, which is the common element found in all of ISO`s management systems and standards, and structures energy management processes around this PDCA framework, which actions are planned, implemented, checked, and the results used to continually improve both energy performance and the EnMS *(DOE, 2016b; ISO, 2018)*. The standard addresses the following *(IMO, 2016)*:

- Energy use and consumption evaluations through performing energy reviews and development of energy policies.
- Measurement, documentation and reporting of energy use.
- Design and procurement practices for energy-using equipment, systems, and processes.
- Development of an energy management plan and other factors effecting energy performance that can be monitored and influenced by the organisation.
- Energy use and consumption evaluations through performing energy reviews and development of energy policies.
- Measurement, documentation and reporting of energy use.
- Design and procurement practices for energy-using equipment, systems, and processes.
- Development of an energy management plan and other factors effecting energy performance that can be monitored and influenced by the organisation.

To do these, the ISO 50001 provides a framework of requirements enabling organisations to (DOE, 2016b; ISO, 2018):

- Develop a policy for more efficient use of energy.
- Fix targets and objectives to meet the policy.
- Use data to better understand and make decisions concerning energy consumption.
- Measure the results.
- Review the effectiveness of the policy.
- Continually improve energy management

Although there can be minor differences between the ISO 50001 and custom EnMS models, the main elements are like the EnMS model of the ISO 50001 as can be seen in Figure 1. The successful implementation of an EnMS, whether the ISO 50001 or a customised approach, can be realised through following a sequence of key steps (Nulty, 2015):

1. Management decision
2. Planning
3. Implementation
4. Checking
5. Review

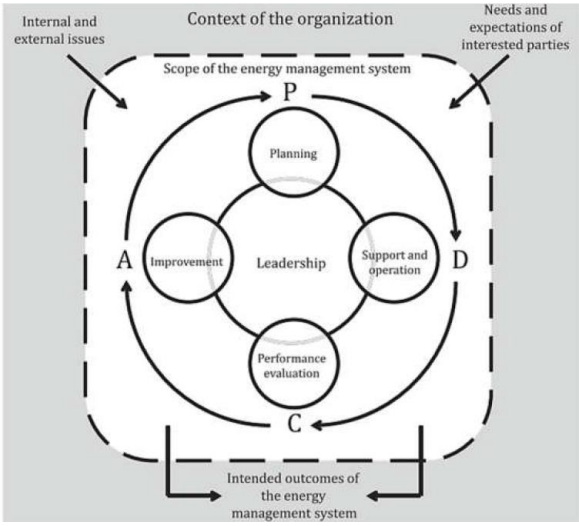


Figure 1 Energy Management System Model (ISO, 2018)

Plan

Context of the organisation

Understanding the organisation and its context

A company should analyse its organisational context to determine external and internal issues that are associated with the purpose of the EnMS to be implemented and that can affect the energy performance of the plant as well as the EnMS (ISO, 2018).

Internal issues can include (ISO, 2018):

- core business objectives and strategy;
- asset management plans;
- financial resources affecting the organisation;
- sustainability consideration;
- contingency plans for interruptions in energy supply;
- maturity of existing technology;
- operational risks and liability considerations;

External issues can include:

- issues related to interested parties such as existing national or sector objectives,
- requirements or standards;
- restrictions or limitations on energy supply, security and reliability;
- energy costs or the availability of types of energy;
- effects of weather;
- effects of climate change;
- effects of GHGs emissions.

Understanding the needs and expectations of interested parties

The organisation should determine the interested parties and their requirements in terms of its energy performance and the EnMS so that these can be addressed through the EnMS. The organisation should ensure its access to the applicable legal and other requirements associated with its energy performance and understand how those requirements apply to them. Also, these legal and other requirements should be reviewed at defined intervals.

Determining the scope of the EnMS

The organisation should determine the boundries and applicability of the EnMS to establish its scope. While doing this, the needs and expectations of interested parties mentioned above should be considered. Also, it should be ensured that the scope and boundries of the EnMS include the authority of the organisation to control its energy efficiency, energy use and energy consumption. The EnMS scope and boundries are maintained as documented information (ISO, 2018).

Leadership

The top management of the organization should show leadership and commitment for continual improvement of its energy performance and the effectiveness of the EnMS (ISO, 2018).

Leadership and commitment

A greater involvement of the top management in the EnMS is of importance for the success and effectiveness of the EnMS. Therefore, the top management should demonstrate leadership and commitment in terms of continual improvement of the energy performance of the company and the effectiveness of the EnMS. The EnMS should be a part of the overall business strategy of the company (Howell, 2014a; ISO, 2018)

Also, the top management must involve in the communication and continued review and approval of the management system and provide

leadership and support when and where needed (Howell, 2014a). This is essential for the success and sustainability of the energy management because without the enthusiasm, commitment, and support of the top management neither the EnMS can be implemented successfully nor the any improvement in energy performance can be realized (Howell, 2014a).

Energy Policy

The commitment of the top management and the overall intention of the organization for managing energy must be formally expressed, documented, communicated, and understood by everyone in the organization, including on-site contractors (Eccleston et al., 2011a). For this reason, it is the responsibility of the top management to establish, implement, and maintain an energy policy which will state the organisation`s vision for energy management (Howell, 2014a). While the energy policy may vary from company to company, it should involve basic common commitments (Howell, 2014a). ISO 5000:20018 requires the top management to establish an energy policy that (ISO, 2018):

- "is appropriate to the purpose of the organization;
- provides a framework for setting and reviewing objectives and energy targets
- includes a commitment to ensure the availability of information and necessary resources
- to achieve objectives and energy targets;
- includes a commitment to satisfy applicable legal requirements and other requirements
- related to energy efficiency, energy use and energy consumption;
- includes a commitment to continual improvement of energy performance and the
- EnMS;
- supports the procurement of energy efficient products and services that impact energy performance.
- supports design activities that consider energy performance improvement (ISO, 2018)"

According to ISO 50001:2018, the energy policy should (ISO, 2018):

- "be available as documented information;
- be communicated within the organisation;
- be available to interested parties, as appropriate;
- be periodically reviewed and updated as necessary (ISO, 2018)"

Energy Manager and Energy Team

Once the top management commitment is obtained, an energy manager (also called as energy director or energy champion) is needed to be attained (Smith & Parmenter, 2016a). The energy manager can be a member of the engineering staff in a large team, or an electrician, a maintenance supervisor (Smith & Parmenter, 2016a). It is the responsibility of the energy manager to ensure that the EnMS is planned, developed, implemented, and sustained with continuous improvement (Howell, 2014a).

One would expect that the energy manager alone cannot do all his energy management responsibilities. For this reason, an energy team, which is leaded by the energy manager, can be created. This will also increase the involvement of the people from various departments in the organisation.

Developing Energy Plan

The energy manager and energy team need to develop a plan to achieve the goals expressed in the energy policy. This energy plan should include the specific processes essential to improving energy performance (Eccleston et al., 2011b). These include:

- energy review process
- energy baselining
- energy performance indicators
- objectives
- targets
- actions

Energy Review and Baselining

This step can be considered as the status quo of the energy performance of the organisation. ISO 50001-2011 defines Energy Review as follows (ISO, 2018):

"Energy review is a process that assesses the organisation`s current energy performance and identifies opportunities for improvement as an input to developing the energy baseline of energy performance indicators."

The outcomes of energy review and baselining processes will provide the basis and a starting point to perform actions for energy performance improvements. In the energy review process, the following activities are done (Parrish & Ledewitz, 2012) :

- setting a scope and boundary for EnMS
- understanding past and present energy consumption, energy types
- identifying significant energy uses

- developing a set of energy efficiency measures to reduce the energy consumption of significant energy users
- developing an energy baseline
- developing energy performance indicators to assess the effectiveness of the energy efficiency measure
- developing objectives and targets for the organisation that support energy policy.
- developing action plans to achieve the objectives and targets

As stated above, ISO 50001-2018 requires the identification of significant energy users and appropriate energy efficiency measures to apply on them. Focusing on significant energy uses within the facility can be regarded as a wise approach because they can offer considerable energy saving potentials. However, the collective effect of less energy users should not be underestimated. The sum of the energy saving potentials identified in less energy consuming systems might be considerable. Therefore, especially at the planning of establishing an energy management system, it will be rational to conduct a detailed energy audit which covers all the energy consuming systems and all energy aspects in an organisation so that a thorough picture of the baseline energy performance can be taken.

From a technical point of view, the review part of the planning phase can be considered as an energy audit. The energy review and baselining steps can be achieved by conducting an energy auditing in the organisation. In fact, energy auditing can be regarded as a precondition for establishing an EnMS. Figure 2 illustrates the role of the auditing procedure in the overall ISO 50001 EnMS structure. Therefore, it is clear that energy auditing fits into the planning stage (Kluczek & Olszewski, 2017).

Conducting an energy audit is crucial for understanding energy consumption characteristics of an organisation and identifying energy and energy cost saving potentials. A qualified energy auditor can conduct a detailed energy audit and assess all relevant aspects related to the energy performance of the organisation and establish energy performance indicators. The key output of the energy audit is a detailed energy performance analysis of the organisation together with a comprehensive list of the identified energy performance improving measures prioritized with respect to their technical, economic, and environmental merits. These measures can include technical measures such as operations and maintenance, retrofit and modifications in processes and systems, shifting low carbon energy uses, etc. Besides the technical measures, there can be several factors such as organisational barriers which discourages efficient use of energy such as lack of awareness of the employees. Identification and assessment these aspects should also be included in the energy audit.

Thereafter, based on the energy auditing results, realistic and meaningful objectives and targets that supports the energy policy can be created (Smith & Parmenter, 2016b). Besides, an energy management action plan to achieve these objectives and targets is developed.

Finally, the outcome of the energy planning phase is "*an action plan for energy performance improvement*", which have been authorised, funded, and staffed by top management, shall be used in the implementation phase (Eccleston et al., 2011b).

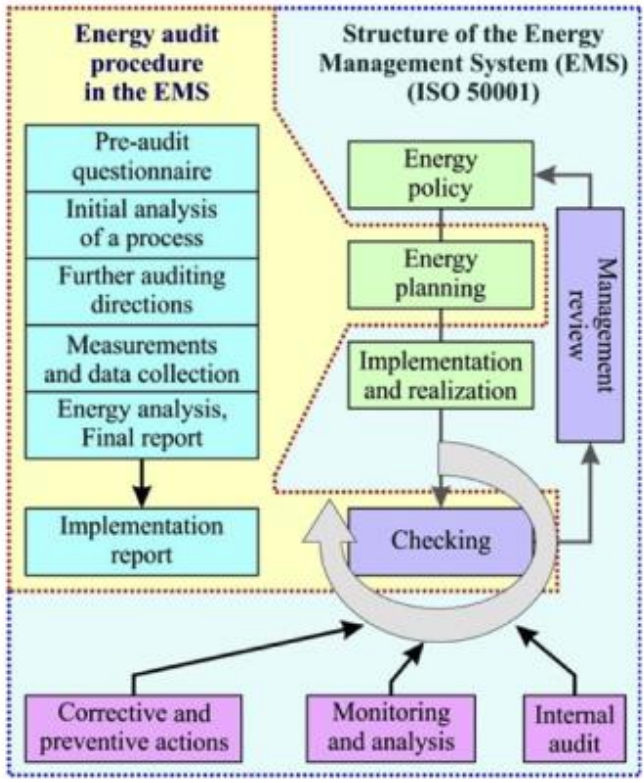


Figure 2: Structure of energy audit inscribed in the ISO 50001 energy management system (EnMS) **(Klu**czek & Olszewski, 2017)

Implementation

In the planning phase, a major energy action plan to improve their energy performance is developed. The execution of these actions is the responsibility of the energy manager and energy team. It should be ensured that the roles and responsibilities are clearly defined and agreed upon for each person that will be involved in implementing the energy action plan. (DOE, 2014). Besides, it should be ensured that the people such as employees associated with the energy management action plan are aware of

the changes to their operations. They should be motivated for successful implementation of the actions, which can be accomplished by means of training and awareness activities (DOE, 2014).

When all the above have been done and the action plans have been implemented, the next phase of energy management system, i.e. checking, takes place to check the results of these actions to see the improvement (DOE, 2014).

Checking

In this phase the EnMS, the progress of the implemented actions in the previous phase are checked to ensure that they are progressing as planned (DOE, 2014). This is done in accordance with the predefined milestones developed in the energy management action plan (DOE, 2014). To be able to check the progress of the implemented energy management actions, a sufficient and regular data gathering must be in place for an effective evaluation of the progress (DOE, 2014). The scope of data and frequency of collection are prescribed in the action plan (DOE, 2014). Ongoing checking can be accomplished through monitoring, analyses, etc.

When the projects of the implemented actions are completed, the results must be evaluated comparing it against the estimated savings or expectations from the implemented. If the actions are not producing the expected results, then, appropriate corrections are done (DOE, 2014).

In achieving the above, an internal energy audit can be conducted. The objectives and scope of the audit can be established in a way to check the progress and results of the undertaken actions based on previous energy audit. This can be considered as a follow-up audit. Before starting it, it is essential to do a review of the energy management action plan prepared and implemented in the previous phases so that the performance of the already implemented actions and EnMS can be examined (Eccleston et al., 2011b). The focus of the internal audit is primarily on to evaluate how well the planned and established EnMS is being implemented in the organisation and to assess its overall effectiveness with respective to the organisation`s energy policy and energy objectives and targets established in the planning phase.

The outcomes of the internal audit is verification of that the energy action plans have been undertaken within the agreed upon timeframe as well as validating their effectiveness (Eccleston et al., 2011b). Based on the results, corrective and preventive actions can be taken if there are any nonconformities or ineffectiveness.

Review

ISO 50001 requires that the top management reviews the energy management system at planned intervals for its suitability, adequacy, and

effectiveness (Howell, 2014b). To do this, a management review meeting is organised and the top management is briefed on the progress and results of the energy management actions, energy management and a review of the EnMS program itself (DOE, 2014). The management review meeting agenda can include the following items (Howell, 2014b):

1. Energy management action plan reviews, energy audit results
2. Evaluation of legal and other compliance and any changes to legal or other requirements
3. The energy performance of the organisation
4. The status of corrective and preventative actions
5. The performance of the EnMS
6. The extent to which goals and objectives have been met
7. Recommendations for improvement

Management reviews are essential parts of the check and act phases of continually improvement of an organisation of EnMS. The output of the review phase indicates the necessary changes and actions for improvement. (Howell, 2014b).

ENERGY AUDIT

It is clear from the above that the ISO 50001 provide only general requirements for the operational level for companies on how to realize energy efficiency in their plants (Doerr et al., 2013; ISO, 2018). It defines neither specific performance criteria regarding energy efficiency (Introna et al., 2014) nor energy efficiency means. The standard just proposes a management model that contributes to develop and implement the energy policy and to establish targets, goals and action plan outcomes of the analysis and control of energy consumption data (Introna et al., 2014). Despite this, carrying out an energy review is a requirement and measuring and analysing energy efficiency and its influencing factors and identifying and implementing appropriate measures to achieve energy efficiency is the responsibility of the organisation (Doerr et al., 2013; ISO 50001, 2011). This is foundation to build an efficient EnMS with low energy consumption, low-pollution and low-emissions in an industrial organisation (Alhourani & Saxena, 2009)

Bearing the above in mind, it is important to note that energy audit is crucial for understanding energy consumption characteristics of a plant and identifying energy and energy cost saving potentials within its processes. In fact, as also stated earlier, an energy audit is the initial step and a precondition to correctly establish and implement an EnMS. However, it is relevant to note that an energy audit defined in the above form is different from an internal audit of an EnMS which can be seen in Figure 2. As also explained earlier, the purpose of an internal audit of an EnMS is to evaluate the processes, procedures and implementation of the EnMS to check whether they are appropriate to the organisation, implementation status and conforming to requirements of the EnMS standard (DOE, 2016b).

Energy audit is highly implemented key tool in the area of energy management (Saidur, 2010). In an energy audit, energy consumption characteristics of a plant or an energy using system are examined so as to ensure that energy is being used in an efficient manner. If it is not used efficiently, areas with inefficiencies where energy is wasted are identified for improvement. Together with this, related cost and emissions saving potentials are realised. In a sense, it is similar to financial accounting (EMSD, 2007).

The objective of an energy audit can vary from one organisation to another and it can also have various degrees of complexity (CIPEC, 2011; Hasanbeigi & Lynn, 2010). Nonetheless, the underlying reasons to conduct an energy audit are usually to understand how energy is consumed within a plant and to identify potentials for energy savings (Hasanbeigi & Lynn, 2010). In this respect, a typical energy audit process involves (CIPEC, 2011):

- Data collection and review
- Plant surveys and system measurements
- Observation and review of operating practices
- Data Analysis

The information obtained through the audit by means of data collection, surveys and measurements is further analysed in order to identify energy saving potentials (ESPs), energy cost saving potentials (ECSO), and GHG emissions reduction potentials. A plant can conduct energy audit by the energy auditor and energy team established in the planning phase or can outsource an energy auditor. Because recommendations suggested by the auditors depends on auditor's experience and knowledge (Kluczek & Olszewski, 2017), it is very important for an energy auditor to have a full understanding of energy using systems.

In general, energy audits are classified in two major groups ((AG, 2011; APO, 2008; CIPEC, 2011; EMSD, 2007; Hasanbeigi & Lynn, 2010; Thumann & Younger, 2008):

1. Preliminary Audit (or Walk-through Audit)
2. Detailed Audit

In a preliminary audit, the readily available or easily obtained data is used to simple analysis of energy use in a plant. There is no need for a lot of measurements in this audit type. It can be regarded as relatively quick exercise (APO, 2008):

1. To determine the energy use in the plant
2. To estimate the scope for saving
3. To identify the most like for attention
4. To identify immediate savings
5. To set a reference point
6. To identify areas where more detailed analysis is needed

As for detailed energy audit, more detailed data and information are required because all major energy consuming systems (i.e. significant energy users) are evaluated. A detailed audit provides the most accurate estimation of energy and costs savings and economic analyses are conducted for identified potentials (APO, 2008). For this reason, it is better to conduct a detailed energy audit when establishing an energy management program or system.

In the literature, there are various generic methodologies to perform a detailed energy audit (AG, 2011; APO, 2008; CIPEC, 2011; EMSD, 2007; Hasanbeigi & Lynn, 2010; Thumann & Younger, 2008). Various handbooks and guidelines are readily available which energy managers, plant owners, researchers can use. It is not rationale to expect an energy auditor to follow the exact same steps and procedures for energy auditing different plants. The methodology for energy audit can show difference from one plant to another. Besides, intellectual profundity of methods for energy analysis and measurements is dependent on the experience and knowledge of the energy auditor (Kluczek & Olszewski, 2017).

RECOMMENDATIONS, IMPLICATIONS, AND CONCLUSIONS

This chapter provided a practical understanding of the core elements of energy management including EnMS, ISO 50001, and energy audit, and and their interrelationships within a PDCA-based framework. Given the complexity of industrial energy consumption, sustainable achievements in energy efficiency in industrial organisations can only be realised through the integration of energy management into organisational processes and the reinforcement of continuous improvement principles.

ISO 50001–aligned Energy Management Systems (EnMS) offer a comprehensive PDCA-based framework that facilitates organisations in systematically planning, implementing, monitoring, and revieweing energy performance improvement actions. Within this framework, energy auditing functions as a critical technical foundation by establishing energy baselines, identifying significant energy users, and facilitating the prioritisation of energy saving potentials and measures.

It is also important to clearly distinguish between between internal EnMS audits and plant energy audits. The focus of plant energy audits is to understand the level of energy use in the plant and to identify energy saving potentials. On the other hand, internal EnMS audits assess the the progress and effectiveness of the EnMS implementation thereby supporting corrective actions and continual improvement.

Overall, by clarifying the interrelationships between energy management, ISO 50001, and energy auditing, this chapter offers practicable framework for the understanding, implementation and maintenance of industrial energy management practices.

REFERENCES

AG. (2011). *Energy Efficiency Opportunities Assessment Handbook*. Australian Government - Deparment of Resources, Energy and Tourism. www.energyeffciencyopportunities.gov.au

Alhourani, F., & Saxena, U. (2009). Factors affecting the implementation rates of energy and productivity recommendations in small and medium sized companies. *Journal of Manufacturing Systems*, *28*(1), 41–45. https://doi.org/10.1016/j.jmsy.2009.04.001

Amundsen, A. (1999). *Joint management of energy and environment. Cleaner production, environmental management systems and energy management facilitating sustainable industrial development in the food processing industry* [PhD Thesis]. Centre for Environmental Studies, Erasmus

University, Rotterdam, The Netherlands. Fredrikstad (Norway): Oestfold Research Foundation.
APO. (2008). *Working Manual on Energy Auditing in Industries*. Asian Productivity Organisation.
Caffal, C. (1995). *Energy management in industry, Analysis Series*. Centre for the Analysis and Dissemination of Demonstrated Energy Technologies (CADDET).
CIPEC. (2011). *Energy savings toolbox: An energy audit manual and tool*. Natural Resources Canada. Office of Energy Efficiency - Canadian Industry Program for Energy Conservation (CIPEC). http://publications.gc.ca/site/eng/384235/publication.html
DOE. (2014). *DOE eGuide for ISO 50001*. Deparment of Energy. https://www.energy.gov/eere/amo/articles/doe-eguide-lite
DOE. (2016a). *eGuide, Why Energy Management*. Https://Ecenter.Ee.Doe.Gov/_layouts/Ecenter/Ppc.Eguide/Home.Aspx.
DOE. (2016b). *eGuide—ISO 50001 Energy Management*. AMO Energy Resources Center - DOE. https://ecenter.ee.doe.gov/_layouts/ecenter/ppc.eguide/home.aspx#
Doerr, M., Wahren, S., & Bauernhansl, T. (2013). *Methodology for energy efficiency on process level*. Forty sixth CIRP conference on manufacturing systems 2013.
Eccleston, C., March, F., & Cohen, T. (2011a). *Inside Energy: Developing and Managing an ISO 50001 Energy Management System*. CRC Press. https://doi.org/10.1201/b11694
Eccleston, C., March, F., & Cohen, T. (2011b). *Inside Energy: Developing and Managing an ISO 50001 Energy Management System*. CRC Press. https://doi.org/10.1201/b11694
EMSD. (2007). *Guidelines on Energy Audit*. Electrical & Mechanical Services Department - The Government of the Hong Kong Special Administrative Region. http://www.emsd.gov.hk/filemanager/en/content_718/Guidelines_on_Energy_Audit.pdf
Hasanbeigi, A., & Lynn, P. (2010). *Industrial Energy Audit Guidebook: Guidelines for Conducting an Energy Audit in Industrial Facilities*. Ernest Orlando Lawrence Berkeley National Laboratory.
Howell, M. T. (2014a). *Effective Implementation of an ISO 50001 Energy Management System (EnMS)*. ASQ Quality Press.
Howell, M. T. (2014b). *Effective implementation of an ISO 50001 energy management system (EnMS)*. ASQ Quality Press.
IMO. (2011). *Main events in IMO's work on limitation and reduction of greenhouse gas emissions from international shipping*. IMO. http://www.imo.org/en/OurWork/Environment/PollutionPrevention/AirPollution/Documents/Main%20events%20IMO%20GHG%20work%20-%20October%202011%20final.pdf
IMO. (2016). *Energy Management Plans and Systems—IMO Train the Trainer (TTT) Course on Energy Efficient Ship Operation*. International Maritime Organisation (IMO). http://www.imo.org/en/OurWork/Environment/PollutionPrevention/AirPoll

ution/Documents/Air%20pollution/M6%20energy%20management%20plan%20and%20system%20final.pdf
Introna, V., Cesarotti, V., Benedetti, M., Biagiotti, S., & Rotunno, R. (2014). Energy Management Maturity Model: An organizational tool to foster the continuous reduction of energy consumption in companies. *Journal of Cleaner Production*, *83*(0), 108–117. https://doi.org/10.1016/j.jclepro.2014.07.001
ISO. (2018). *ISO 50001: 2018, Energy Management System*. ISO.
ISO 50001,. (2011). *ISO 50001: 2011, Energy Management System*. ISO.
Kannan, R., & Boie, W. (2003). Energy management practices in SME—case study of a bakery in Germany. *Energy Conversion and Management*, *44*(6), 945–959. https://doi.org/10.1016/S0196-8904(02)00079-1
Kluczek, A., & Olszewski, P. (2017). Energy audits in industrial processes. *Journal of Cleaner Production*, *142*, 3437–3453. https://doi.org/10.1016/j.jclepro.2016.10.123
McKane, A., Therkelsen, P., Scodel, A., Rao, P., Aghajanzadeh, A., Hirzel, S., Zhang, R., Prem, R., Fossa, A., Lazarevska, A. M., Matteini, M., Schreck, B., Allard, F., Villegal Alcántar, N., Steyn, K., Hürdoğan, E., Björkman, T., & O'Sullivan, J. (2017). Predicting the quantifiable impacts of ISO 50001 on climate change mitigation. *Energy Policy*, *107*, 278–288. https://doi.org/10.1016/j.enpol.2017.04.049
Nulty, H. M. (2015). *AN INTRODUCTION TO ENERGY MANAGEMENT SYSTEMS: ENERGY SAVINGS AND INCREASED INDUSTRIAL PRODUCTIVITY FOR THE IRON AND STEEL SECTOR* (p. 34). OECD - DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INNOVATION STEEL COMMITTEE. https://www.oecd.org/sti/ind/DSTI-SU-SC(2014)14-FINAL-ENG.pdf
Parrish, K., & Ledewitz, J. (2012). *Early Lessons Learned from Building an ISO 50001: Conformant Energy Management System for MIT*. 13.
Saidur, R. (2010). A review on electrical motors energy use and energy savings. *Renewable and Sustainable Energy Reviews*, *14*(3), 877–898. https://doi.org/10.1016/j.rser.2009.10.018
Schulze, M., Nehler, H., Ottosson, M., & Thollander, P. (2016). Energy management in industry – a systematic review of previous findings and an integrative conceptual framework. *Journal of Cleaner Production*, *112*, 3692–3708. https://doi.org/10.1016/j.jclepro.2015.06.060
SEAI. (2015). *Energy Management Systems* [Governmental]. Sustainable Energ Authority of Ireland. http://www.seai.ie/Your_Business/Large_Energy_Users/Energy_Management_Standard/Energy_Management_System/
Smith, C. B., & Parmenter, K. E. (2016a). *Energy management principles: Applications, benefits, savings* (2nd edn).
Smith, C. B., & Parmenter, K. E. (2016b). *Energy management principles: Applications, benefits, savings* (2nd edn). Elsevier. http://linkinghub.elsevier.com/retrieve/pii/0198759382900820
Sola, A. V. H., & Mota, C. M. M. (2020). Influencing factors on energy management in industries. *Journal of Cleaner Production*, *248*, 119263. https://doi.org/10.1016/j.jclepro.2019.119263

Thumann, A., & Younger, W. J. (2008). *Handbook of energy audits* (7th ed). Fairmont Press ; Distribution by CRC Press/Taylor & Francis.

Digital Work Study: A Multi-Dimensional Framework and Research Agenda

Erkan Sami KOKTEN[1]

1- Assist. Prof. Dr.; Karabuk University Faculty of Engineering and Natural Sciences Industrial Engineering. erkansamikokten@karabuk.edu.tr ORCID No: 0000-0003-3428-4534

ABSTRACT

Although classical work study is one of the fundamental tools of the industrial engineering discipline, its dependence on observers, labor-intensive nature, and low scalability structure may make it insufficient in meeting the requirements of modern production systems, such as speed and flexibility. This study aims to develop a multi-dimensional framework for the concept of digital work study, based on existing literature, and categorizes the methods in the literature into four dimensions: data source, analysis method, output type, and stage of development. The findings indicate that image processing and computer vision–based approaches enable non-intrusive data acquisition, allowing cycle times to be extracted without altering operator behavior, and achieve notably high levels of accuracy in action recognition. In contrast, IMU-based wearable systems have been found to present certain limitations in practical use due to issues such as operator comfort and calibration, despite offering advantages such as higher sensitivity. Hybrid solutions stand out as systems that provide high accuracy even under occlusion conditions by combining image and sensor data. The majority of the studies reviewed are positioned at development stages 2 and 3. However, it has been determined that there are shortcomings in determining elements such as standard time, performance rating, or tolerance rate. The study emphasizes that integrating Predetermined Motion Time Systems (PMTS)-based (MTM/MODAPTS) work measurement with digital segmentation outputs may be of critical importance in terms of decision support, line balancing, and workforce planning. Consequently, digital work study should be adopted not only as a technology that replaces the stopwatch method, but as a standardization approach for Industry 4.0 that combines algorithmic observation, ergonomic risk management, and real-time production analytics.

Keywords – Digital Work Study, Time Study, Action Recognition, PMTS, Industry 4.0, Computer Vision, IMU, Hybrid Sensing.

INTRODUCTION

Work study is a set of techniques aimed at systematically analyzing human, machine, and material resources to ensure that a specific work is performed in the most efficient manner (Waseem et al., 2021). As one of the fundamental tools of industrial engineering, this discipline plays a crucial role in enhancing productivity in production systems, establishing standard times, and refining work processes. In the classical sense, work study consists of two main components: method study and work measurement, and is based on the Scientific Management approach.

Frederick W. Taylor, one of the pioneers of the Scientific Management concept, laid the foundations of work study with his work in the field of work measurement and time study. Taylor proposed that a job be divided into smaller components for analysis, the time for each component be measured, and the most efficient working method be determined. According to Taylor, the main reason for inefficiency in production systems was not material waste but rather the ineffective use of human labor. Complementing this approach, Frank and Lilian Gilbreth developed the concept of motion study, examining the hand and body movements performed by workers; they proposed systematic analysis methods to eliminate unnecessary movements and enable work to be performed with minimal fatigue. The basic motion classification developed by the Gilbreths, called Therblig, allows for the micro-level examination of operations, enabling the identification of inefficient movements (Fang, 2015).

The primary goal of method study is to analyze the movements required to perform a task, eliminate unnecessary or repetitive movements, and reorganize the remaining movements in the most efficient order and method. This reduces the job content, lowers operator fatigue, and increases overall system efficiency (Waseem et al., 2021). Time study, on the other hand, aims to determine the time required for a skilled worker to perform a specific job under defined conditions and at a specific performance level, i.e., the standard time. Within this scope, methods such as time study, work sampling, and Predetermined Motion Time Systems (PMTS) have been developed (Ji et al., 2022; Nesterak et al., 2025). In PMTS systems, tasks are divided into basic movements, and standard time is calculated by assigning predefined durations (TMU-Time Measurement Unit) to each movement. MTM (Methods-Time Measurement), MODAPTS (Modular Arrangement of Predetermined Time Standards), and similar systems are typical examples of this approach (Mares-Castro et al., 2025; Wu et al., 2016).

The classic role of work study in industry is to eliminate inefficiencies. In this regard, work study serves fundamental objectives such as increasing productivity, standardizing work processes, controlling costs, planning the workforce, and ensuring a fair distribution of the workload (Gurav & Upadhyay, 2025; Mor et al., 2019). In this respect, classical work study is a fundamental improvement tool that scientifically analyzes how production systems work (method) and how long tasks should take (time). However, traditional work study methods have significant limitations. In the classical approach, the process is conducted by a qualified observer and relies on manual time measurements (Faccio et al., 2019; Waseem et al., 2021). This makes the method labor-intensive and time-consuming, limiting its applicability in large-scale or highly variable production environments. Furthermore, observers' subjective assessments and workers' tendency to change their behavior when they realize they are being observed (the

Hawthorne effect) can negatively affect the accuracy and repeatability of measurement results (Bortolini et al., 2020; Nesterak et al., 2025).

With the spread of Industry 4.0 and smart manufacturing paradigms, the need for digitalization in the field of work studies has become increasingly apparent. Today's production systems exhibit a more complex and dynamic structure due to increasing product diversity, shorter product life cycles, and the need for mass customization (Bortolini et al., 2020; Li et al., 2025). In this context, manual time study methods may be insufficient in providing the necessary speed, flexibility, and data integrity. At the same time, the need to reduce human error and subjectivity in the measurement process has accelerated the shift towards more objective and automated data collection systems (Gurav & Upadhyay, 2025).

In line with these developments, new digital concepts have emerged in the literature that support or partially replace classical work study approaches. The concepts of “digital time study” or “automated time study” refer to the recording of work element times without human intervention through the use of sensors and algorithms (Ji et al., 2022; Li et al., 2025). “Smart work measurement,” implemented with Internet of Things (IoT)-based sensors and energy monitoring systems, enables the collection and analysis of production data in near real time (Jung et al., 2020; Sitthidetchtamroung & Khamkanya, 2024). In addition, automatic motion analysis approaches based on computer vision and deep learning techniques can analyze cycle times and inefficient movements without requiring intervention by operators, by tracking their body postures and hand movements (Anadani et al., 2021; Hassani et al., 2023; Panahi et al., 2025). These technologies provide the basis for creating digital twins of work processes and using big data analytics in continuous improvement activities (Mares-Castro et al., 2025; Nesterak et al., 2025).

CLASSICAL WORK STUDY AND ITS LIMITATIONS

Work study is a managerial and analytical tool that aims to systematically examine the activities carried out in an organization to ensure the most efficient use of available resources (Waseem et al., 2021). The classical work study approach, developed by pioneers such as Taylor and Gilbreth, is based on scientific principles of method study, which determines how a job should be done, and time study, which defines how long it takes to

complete the job. This approach is used as an effective tool for increasing productivity, establishing standards, and controlling costs in production systems. However, as industrial environments have become increasingly complex, variable, and digitized, some methodological limitations inherent in classical work study methods have become more apparent. The classical work study approach is based on the principle of dividing an operation into analyzable and measurable parts. In this context, a work process consists of work elements defined based on the basic movements performed by the operator or specific changes occurring in the workpiece (Ji et al., 2022). The process of breaking down a job into its elements enables the systematic examination of complex operations and the separate evaluation of each step. The concept of standard time, which is the final output of time study, refers to the time required for a skilled operator with average ability to complete a specific task under defined working conditions and at a normal working pace (Ji et al., 2022; Le & Nguyen, 2021). Since observed times can vary naturally from cycle to cycle, standard time is based on average values representing normal working conditions, not on individual measurements. Therefore, the raw times measured during the time study process are converted to standard time, taking into account the operator's performance level. Another fundamental element considered in calculating standard time is rest allowances. Normal time is calculated by adding tolerance allowances necessary for the operator to recover from physiological and mental fatigue, meet personal needs, and cope with unavoidable delays. In the literature, standard time is generally expressed as the product of normal time and tolerance allowances (Le & Nguyen, 2021). This approach aims to create more realistic standards by incorporating the human factor into the time measurement process.

Traditional time and motion study applications mainly rely on a well-trained analyst observing an operator for a certain number of cycles and manually recording the times with a stopwatch (Ji et al., 2022). However, this manual process is subject to various factors that limit the reliability of the data and the applicability of the method. First, manual time study is highly dependent on the observer. Analyst distraction, cognitive fatigue from the monotonous timing process, and prolonged observation can lead to measurement errors (Ji et al., 2022). Additionally, operators who consciously or unconsciously alter their work behavior when aware of being observed are a significant source of bias, defined in the literature as the Hawthorne effect. This effect limits the accuracy of standard times by preventing observed cycle times from reflecting actual operating conditions.

Another significant limitation of classical methods occurs in the analysis of short-cycle and high-repetition tasks. Manually measuring micro-operations lasting less than a second is both time-consuming and prone to error. Although predetermined time standards such as MTM-1 offer high theoretical accuracy, the implementation of these systems through manual

video analysis requires expertise, can be costly, and is prone to human error (Andreopoulos et al., 2024). This situation makes it particularly challenging to identify ergonomic risks and inefficient micro-movements, which negatively impact overall production efficiency (Mares-Castro et al., 2025). Furthermore, classical work study methods cannot adequately represent time-dependent performance changes such as learning curves and fatigue. The labor-intensive nature of manual measurements often necessitates data collection within a limited time frame, making it challenging to comprehensively analyze changes in operator performance throughout the day or over the long term (Ji et al., 2022). The need for long-term observation causes fatigue not only in operators but also in analysts, negatively affecting measurement consistency. Manual data entry and analysis processes are not only prone to errors but also unsustainable in large-scale, highly variable modern production environments (Le & Nguyen, 2021). These limitations demonstrate that classical work study approaches are inadequate in meeting the speed, flexibility, and data integrity requirements of today's production systems. Structural problems, including labor intensity, high costs, and subjectivity, have made the development of computer vision, sensor-based monitoring, and AI-supported digital work study approaches inevitable (Faccio et al., 2019; Mares-Castro et al., 2025).

CONCEPTUAL FRAMEWORK OF DIGITAL WORK STUDY

The digitization of traditional methods used in industrial process optimization represents a new approach developed to increase data accuracy and minimize human error. In this context, the concept of digital work study is related to the development of classical methods through the use of technological tools. This section will define digital work study and evaluate the components that constitute this concept.

1. Digital Work Study

Digital work study is the measurement and analysis of the time and method required to perform a work task using sensors, video recording devices, software, and artificial intelligence algorithms, rather than manual tools (such as stopwatches and paper forms), based on human observation (Ji et al., 2022). This approach aims to eliminate the subjectivity arising from the human factor in determining the start and end points of work elements.

While traditional time studies are prone to analyst fatigue and errors due to monotony, digital methods automate the process using computer vision and deep learning techniques (e.g., CNN), thereby reducing human intervention and error (Ji et al., 2022; Mares-Castro et al., 2025). Furthermore, these systems have the potential to minimize the impact of psychological factors on the data, such as employees changing their behavior when they realize they are being monitored.

2. Components of the Digital Work Study

The digital work study process consists of an integrated flow where data is converted from its raw form to a standard time format.

Data Collection: The first step in the process is converting physical movements in the workplace into digital signals. At this stage, cameras that record the operator's movements in 3D (e.g., Microsoft Kinect, RGB-D cameras) or IoT-based smart plugs and sensors that monitor power consumption on the machine are used (Bortolini et al., 2020; Jung et al., 2020). These devices continuously and non-stop record operator data or machine status as a data stream.

Data Processing: The collected raw data (video frames or sensor signals) is processed to remove noise and make it analyzable. For example, skeleton tracking algorithms detect the operator's joint points (such as the hand, elbow, and shoulder) to extract the biomechanical characteristics of movements (Bortolini et al., 2020; Mares-Castro et al., 2025).

Time/Task Element Extraction: The processed data is divided into basic movements (Therbligs or MTM codes) such as reaching, grasping, and joining. Deep learning models (MS-G3D or TadTR) classify the actions in the video stream and automatically determine the start and end times of each work element (Li et al., 2025). This stage enables the automatic calculation of the cycle time.

Standard Time Production: In the final stage, the observation time measured by the algorithms is combined with the work pace and rest periods (tolerances) to calculate the standard time. For example, using synthetic evaluation methods, the operator's speed and fatigue are taken into account and compared with predetermined time standards (PTS) to produce the final standard time (Li et al., 2025).

3. Digital Work Study and Industry 4.0 Relationship

The concepts of Digital Work Study and Industry 4.0 are often confused in the literature. Although there is a hierarchical relationship between these two concepts, they are distinct from each other. Industry 4.0 is a broad paradigm encompassing the digitization of production systems and the integration of technologies such as cyber-physical systems, IoT, and big data analytics (Bortolini et al., 2020). This framework provides the necessary infrastructure and vision for production to become autonomous and data-

driven. Digital work study is the application of Industry 4.0 technologies (sensors, AI, big data) to the solution of a specific problem, namely work measurement and method analysis. It is positioned as an enabling technology used to achieve smart manufacturing goals (Hassani et al., 2023; Mares-Castro et al., 2025). Therefore, digital work study is a vertical application area within the Industry 4.0 ecosystem that supports human-centered production and efficiency.

CLASSIFICATION OF DIGITAL WORK STUDY APPROACHES

Digital work study applications form a complex structure in terms of data collection infrastructures, analysis algorithms used, and the scope of the output produced. This makes it challenging to compare the findings of different studies and to assess the level of development in the field comprehensively. Therefore, in this section, digital work study approaches reported in the literature are classified along four main dimensions: (i) data source, (ii) analysis technique, (iii) output type, and (iv) stage of development. The proposed classification aims to make the strengths and weaknesses of the methods comparable by bringing together the technical components of digital work study systems under a common terminology. The literature review underlying this study was conducted in the Web of Science and Google Scholar databases. Keyword combinations such as “digital work study,” “automatic time study,” “smart work measurement,” and “computer vision time study” were used in the search process, with studies published between 2015 and 2025 considered. Studies focused on motion analysis in the fields of sports, medicine, and rehabilitation were excluded from the scope. The search was limited to time measurement and digital work study applications in production and assembly processes.

1. Classification by Data Source

The primary data source used in digital work studies is one of the fundamental elements that determine the system's cost, accuracy, suitability for workplace conditions, and ease of implementation. In the literature, data sources are generally grouped under video-based, wearable sensor-based, and hybrid approaches.

1.1. Video-based approaches

Video-based systems rely on the collection of image data through fixed cameras (CCTV, RGB, or RGB-D) placed in the production environment and the extraction of operator movements from this data. These studies often use markerless motion capture and pose estimation techniques to track the operator's skeleton model; motion/action classification is performed using deep learning-based architectures (CNN-based structures) (Ji et al., 2022; Mares-Castro et al., 2025).

The advantage of video-based approaches is that they are "non-intrusive" because they do not require any device to be attached to the operator and can be set up with relatively low hardware costs. This feature offers an advantage in that it minimizes the impact on the natural working posture. However, system performance is highly sensitive to environmental conditions in practice. Factors such as occlusion, insufficient lighting, and complex backgrounds can lead to data loss and detection errors on assembly lines (Mares-Castro et al., 2025; Panahi et al., 2025).

1.2. Wearable sensor-based approaches

In wearable systems, the kinematics of movement (acceleration, angle, speed) and muscle activity can be measured directly using inertial measurement units (IMUs), accelerometers, and/or electromyography (EMG) sensors placed on the operator's body (Mazloumi et al., 2021). This approach offers high sensitivity, particularly in assessing ergonomics and physical loads. For example, it has been reported that IMU data can be used to identify musculoskeletal risks based on lumbar and upper limb movement patterns, as well as to detect kinematic differences in workers with specific health conditions, such as back pain (Mazloumi et al., 2021).

However, there are some difficulties in integrating wearable approaches with time studies. Sensor-induced positional drift can be observed during prolonged use, and wearing the devices can affect worker comfort, reducing acceptance levels. Furthermore, the calibration and maintenance requirements of sensors can limit the system's sustainability in field applications (Bortolini et al., 2020).

1.3. Hybrid approaches

Hybrid approaches can integrate multiple data types to reduce the limitations of a single data source and increase detection accuracy. A common example in the literature is the combination of video data with sensor data; in some applications, it has been reported that the inclusion of environmental signals such as sound or ground vibration in the system improves event detection performance (Dong et al., 2025; Pan et al., 2022). The distinguishing feature of hybrid systems is that they allow for the simultaneous assessment of not only cycle time but also ergonomic risk and fatigue indicators. However, integration, synchronization, and data fusion

problems create an additional methodological burden in hybrid systems. Table 1 summarizes a comparison of digital work study approaches based on the data source.

Table 1: Comparison of digital work study approaches by data source

Data Source	Advantages	Disadvantages	Typical Applications
Video-Based Systems	Low hardware cost, non-intrusive, natural working behavior	Field of view occlusion, light and background sensitivity, and high computational requirements	Assembly lines, manual workstations, ergonomics, and posture analysis
Wearable Sensor-Based Systems	High measurement accuracy, physical load, and fatigue detection	Positional drift, impact on operator comfort, and calibration requirement	Ergonomic risk analysis, musculoskeletal system assessments
Hybrid Systems	High accuracy, time + ergonomics + fatigue integration	System complexity, data synchronization	High-risk tasks, R&D environments, digital twin applications

The comparison of video, IMU, and hybrid work study systems is presented in Table 2. The table indicates that video-based solutions have the lowest hardware cost, primarily due to the low cost of camera infrastructure (Kim et al., 2016). In IMU-based systems, the cost increases in direct proportion to the number of sensor units and their placement; therefore, they fall into the medium-high cost category. Hybrid systems, on the other hand, are determined by the cost of software and integration investment rather than individual hardware. Therefore, they are classified as variable costs.

Video systems offer the advantage of non-intrusive operation. They do not require wearable devices, do not alter worker behavior, and reduce the Hawthorne effect. Therefore, they are the systems with the lowest level of intervention. IMU, on the other hand, is intrusive. Attaching sensors to the arm, waist, and legs takes time, can affect usage time, and can interrupt the workflow (Faccio et al., 2019; Kim et al., 2016; Panahi et al., 2025). Hybrid systems require a relatively moderate level of intervention. Although sensor placement is reduced, calibration or setup requires effort.

Video systems perform quite well in terms of action recognition or measurement accuracy (Ji et al., 2022; Wang et al., 2018). IMU systems, on the other hand, achieve medium to high accuracy (Faccio et al., 2019). Hybrid systems offer accuracy values comparable to those of video systems. Panahi et al. (2025) noted that using video alone may fail when objects occlude each other. They reported that with a hybrid system, sub-assemblies could be detected with 93% accuracy and an F1 score of 0.93, even in complex and obstructed images. This demonstrates that hybrid approaches are more reliable in complex factory environments where video or sensors alone are insufficient.

Table 2: Comparison of video, IMU, and hybrid work study systems

System	Hardware Cost	Installation and Intervention Level	Action Recognition/Accuracy
Video	Low	Low	High
IMU (Wearable Sensor)	Medium/High	High	Medium/High
Hybrid Systems	Variable (Cost depends on software integration)	Medium	High

2. Classification by Analysis Technique

The way the collected data is processed is a critical factor that determines the autonomy level and generalizability of the digital work study system. In the literature, analysis approaches can generally be classified as rule-based, machine learning/deep learning-based, and model-based approaches.

2.1. Rule-based systems

In rule-based systems, movement classification is performed based on predefined thresholds or logical rules. An example of this classification method is labeling a movement as a "reach" if the joint angle is above a certain value. The advantage of these systems is that their implementation and computational costs are relatively low. However, due to the variety of movements in production environments, differences between operators, and environmental variability, rule-based approaches are often limited in capturing complex and variable patterns (Sitthidetchtamroung & Khamkanya, 2024).

2.2. Machine learning and deep learning-based systems

In machine learning and especially deep learning-based systems, models are trained on large datasets to extract features from raw data and automate motion/action classification. CNN, LSTM, and ConvLSTM can be used to detect work elements and assign durations from video data. In this context, some studies have focused on cycle time and/or workload estimation by automatically extracting operational segments from video frames (Li et al., 2025). Modeling human–object interaction is reported as an important trend that enables a more accurate definition of actions such as screwing, picking, and placing (Hassani et al., 2023). Although deep learning-based approaches have high accuracy potential, there are discussions regarding issues such as the need for labeled data, dataset biases, generalizability to different line and product configurations, and interpretability.

2.3. Model-based approaches

In model-based approaches, physical principles, biomechanical models, or simulation-based structures are incorporated into the analysis. Simulation of human movement in a digital twin environment or estimation of joint angles and movement patterns from indirect measurements using physics-based graphical models can be evaluated within this scope (Bortolini et al., 2020; Dong et al., 2025). The primary objective of this approach is to partially compensate for deficiencies in sensor or image data and to generate predictions with high physical consistency. However, model calibration and computational load are critical factors determining its applicability on an industrial scale.

3. Classification by Output Type

Digital work study systems can be evaluated in four basic output types based on the scope of the information they provide. In the first group, the system only detects the start and end times of the task to determine the total cycle time. In the second group, the work process is broken down into subcomponents, and the task element times are reported. Some approaches aim to relate this breakdown to PMTS/MTM coding (Li et al., 2025). In the third group, ergonomic risk indicators (risky postures, excessive joint strain) are generated along with time information (Mares-Castro et al., 2025). In the fourth and most comprehensive output set, fatigue levels and the associated rest share/recovery need are estimated by monitoring impairments in movement kinematics (Long & Yang, 2025). This classification demonstrates that digital work study is not limited to time measurement alone; it has expanded to encompass ergonomics and occupational health and safety dimensions.

4. Classification by Level of Development

The industrial applicability and degree of automation of digital work study systems may depend not only on the sensors or algorithms used, but also on the extent to which the system automatically produces end-to-end output. Therefore, the approaches in the literature have been evaluated at five levels according to their degree of development (Table 3).

Stage 1: Data is collected digitally using cameras or sensors. However, start-end detection and the analysis process are performed mainly manually (Petersdorff, 2011).

Stage 2: The system can automatically mark the presence or absence of work or cycle boundaries. However, the distinction between work elements is limited and requires human verification (Jung et al., 2020; Panahi et al., 2025).

Stage 3: Actions are automatically classified from video or sensor data, and time assignments related to work elements are made (Hassani et al., 2023; Ji et al., 2022).

Stage 4: The system aims to calculate standard times by going beyond the observed periods and considering factors such as performance level (rating) and task difficulty (Li et al., 2025).

Stage 5: The system operates in an integrated manner with MES/digital twin. Ergonomic risk and fatigue indicators are monitored in real-time, standards are dynamically updated, and predictive alerts are generated (Mares-Castro et al., 2025; Nesterak et al., 2025).

Table 3: Stages of development of digital work study systems and expected minimum functions

Stage	Description	Expected Minimum Functions
Stage 1	Digital data collection	Data recording with sensors/cameras, manual analysis
Stage 2	Semi-automatic time measurement	Detection of presence/absence of work, limited automation
Stage 3	Automatic task element recognition	Motion/action classification, duration assignment
Stage 4	Standard time generation	Standard time calculation considering rating and task difficulty
Stage 5	Real-time adaptive standard	MES/digital twin integration, dynamic standards, and predictive alerts

5. Discussion of Classifications and Comparative Evaluation

The multi-dimensional classification proposed in this study reveals the dimensions in which digital work study approaches are more decisive for different problem types. While the data source is a critical determinant in terms of the system's applicability in the field, cost, and employee acceptance, the analysis technique has a direct impact on the system's autonomy level and generalizability. Output type is one of the fundamental axes defining the scope of digital work study, revealing whether the system only measures cycle time or also encompasses human factors such as ergonomics and fatigue. Finally, the development stage dimension enables the comparison of different studies not only in terms of technical accuracy but also in terms of industrial integration and sustainability. In this context, it is observed that the vast majority of studies in the literature are concentrated in the Stage 2-3 range, while solutions for Stage 4 and 5 are still limited in number and experimental in nature.

The pipeline depicted in Figure 1 is adapted from recent computer vision work on measurement frameworks, where data acquisition precedes algorithmic processing and industrial integration (Ji et al., 2022; Panahi et al., 2025). RGB-based pose estimation and IMU-based motion signals are commonly processed through action segmentation pipelines, producing temporal annotations such as motion elements and cycle time (Li et al., 2025). These outputs are increasingly connected to digital twin environments and manufacturing execution systems (Bortolini et al., 2020; Mares-Castro et

al., 2025), supporting line balancing, workforce allocation, and ergonomic monitoring.

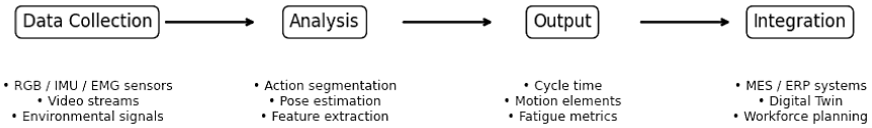


Figure 1: Flowchart for the digital work study system

THE RELATIONSHIP BETWEEN DIGITAL WORK STUDY AND PMTS

Work measurement in engineering is crucial for enhancing productivity and establishing standards. In this context, PMTS, particularly MTM and MODAPTS, are among the most commonly used and reliable methods for analyzing manual work elements (Wu et al., 2016). However, with the age of digitalization and the rise of computer vision and sensor-based technologies, the way classic PMTS methods are applied is changing. The purpose of this section is to examine the relationship between digital work study and classical PMTS methods.

1. Strengths of PMTS

The most significant advantage of PMTS methods, particularly MTM and MODAPTS, is that they divide work into universally defined standard movements (such as reaching, grasping, and placing) and assign a predetermined time value to each movement. These systems can measure a task without the need for a stopwatch. The structured data and common language provided by these methods enable the standardization not only of the task duration but also of the task execution method.

2. Strengths of Digital Methods

Although classic PMTS methods provide high accuracy, they are time-consuming to implement, costly, and dependent on the analyst's expertise (Mares-Castro et al., 2025; Wu et al., 2016). Furthermore, manual data entry and analysis are prone to human error. In contrast, computer vision and AI-supported digital methods offer the capacity to process large data sets, speed, and flexibility. Digital systems can automatically analyze

video data to identify operational frames and process thousands of data points in a short time without human intervention (Li et al., 2025).

3. Integration of PMTS and Digital Methods

Recent studies have shown that digital measurement and analysis systems can generate PMTS codes automatically or semi-automatically. In this context, it has been reported that deep learning-based motion recognition models can identify basic motion pairs (MODAPTS) with high accuracy, particularly in assembly operations, and that the system can automatically calculate the predefined durations corresponding to these motions. For example, Li et al. (2025) demonstrated that MODAPTS codes can be digitally extracted using the MS-G3D and TadTR architectures and that the associated PTS durations can be generated without human intervention. Similarly, it is reported that processing camera data with classifiers such as K-Nearest Neighbor (KNN) yields classification accuracies of up to 100% for MTM-2 movement categories (e.g., reaching, grasping) (Mares-Castro et al., 2025).

These findings underscore the need for a future framework for hybrid work studies. In this framework, digital tools (such as sensors, cameras, and computer vision applications) should take on labor-intensive tasks, including data collection, segmentation, and duration determination. Meanwhile, PMTS logic should serve as a functional syntax for interpreting, classifying, and standardizing the obtained motion sequences. In such a division of labor, the manual segmentation process is delegated to software, shifting the observer's role to a strategic position focused on evaluating the rationality of the operation, identifying sources of variation, and optimizing the task design (Wu et al., 2016).

Furthermore, the potential of the hybrid approach is not limited to time measurement. The integration of deep learning and PMTS enables real-time monitoring of ergonomic risks and instant assessment of indicators such as operator fatigue, risky joint angles, or wrist hyperextension. Thus, digital work study outputs can integrate with production execution systems and digital twin platforms to create the infrastructure for dynamically updating standard times and implementing predictive improvement mechanisms (Bortolini et al., 2020; Mares-Castro et al., 2025). In this context, it is considered that PMTS and digital methods are not mutually exclusive but rather complementary structures. Integrated models, where data collection speed is provided by digital systems and standardization logic is provided by PMTS, will give a new direction to the discipline of work study.

RESULTS AND DISCUSSION

This study examines research on digital work studies under a multi-dimensional classification framework, systematically evaluating the conceptual development of the field. The study structures the concept of digital work study along the dimensions of data source, analysis technique, output type, and stage of development, showing where the literature is concentrated and revealing the complementary relationship between PMTS and digital methods. The proposed classification offers a common language that enhances the comparability of various approaches.

The current literature reveals that manual time studies are increasingly inadequate in meeting the speed, flexibility, and multi-productivity requirements of modern production environments due to their labor-intensive structure, observer dependency, and low scalability. In contrast, video and sensor-based data collection methods enable the generation of unobtrusive data with minimal impact on the operator's work behavior. Furthermore, integrating PMTS-based standardization approaches with digital measurement systems can create a new analytical framework that provides decision support functions in process design, workforce planning, and performance evaluation.

Digital work study presents a new field of research that combines Industry 4.0 technologies with human-centered production, converging the disciplines of ergonomics, work study, and data science. In this context, increasing academia-industry collaborations, data sharing, and establishing joint testing platforms are critical. In conclusion, the concept of digital work study should not be considered merely as an update to time measurement technology. Digital work study also represents a new standardization architecture for production systems. In this regard, the study offers a human-centered methodological perspective in the context of digitalization.

REFERENCES

Anadani, O., Bergström, H., Fåhraeus, G., Helgesson, O., Svensson, S., & Tärnåsen, H. (2021). Utilizing Computer Vision for the Analysis of Manufacturing Processes, Evaluating Automation Possibilities for the AviX Software Suite. CHALMERS UNIVERSITY OF TECHNOLOGY, UNIVERSITY OF GOTHENBURG Gothenburg.

Andreopoulos, E., Gorobets, V., & Kunz, A. (2024). Automated transcription of mtm motions in a virtual environment. International Congress on Information and Communication Technology, 243–259.

Bortolini, M., Faccio, M., Gamberi, M., & Pilati, F. (2020). Motion Analysis System (MAS) for production and ergonomics assessment in the manufacturing processes. Computers and Industrial Engineering, 139(October 2018), 105485. https://doi.org/10.1016/j.cie.2018.10.046

Dong, Y., Kim, S. E., Rose, J., & Noh, H. Y. (2025). Physics-Informed Graphical Modeling of Human–Structure Interactions for Gait and Motion Analysis. Journal of Sound and Vibration, 618(PB), 119351. https://doi.org/10.1016/j.jsv.2025.119351

Faccio, M., Ferrari, E., Gamberi, M., & Pilati, F. (2019). Human Factor Analyser for work measurement of manual manufacturing and assembly processes. International Journal of Advanced Manufacturing Technology, 103(1–4), 861–877. https://doi.org/10.1007/s00170-019-03570-z

Fang, Y. (2015). A method of meta-mechanism combination and replacement based on motion study. Production and Manufacturing Research, 3(1), 310–323. https://doi.org/10.1080/21693277.2015.1093437

Gurav, V., & Upadhyay, A. (2025). A COMPUTER VISION-BASED MOTION TIME STUDY FOR MANUFACTURING ASSEMBLY LINE. Industrial Engineering Journal, 54(3), 42–65.

Hassani, I. El, Masrour, T., Hajji, T., Belmarouf, C., & Talhaoui, Y. (2023). Artificial Intelligence for Work Measurement: A Promising Approach to Improving Productivity and Reducing Waste. Lecture Notes in Networks and Systems, 771 LNNS, 89–109. https://doi.org/10.1007/978-3-031-43524-9_7

Ji, J., Pannakkong, W., & Buddhakulsomsiri, J. (2022). A Computer Vision-Based Model for Automatic Motion Time Study. Computers, Materials and Continua, 73(2), 3557–3574. https://doi.org/10.32604/cmc.2022.030418

Jung, W. K., Kim, H., Park, Y. C., Lee, J. W., & Ahn, S. H. (2020). Smart sewing work measurement system using IoT-based power monitoring device and approximation algorithm. International Journal of Production Research, 58(20), 6202–6216. https://doi.org/10.1080/00207543.2019.1671629

Kim, C. Y., Hong, J. S., & Chun, K. J. (2016). Validation of feasibility of two depth sensor-based Microsoft Kinect cameras for human abduction-adduction motion analysis. International Journal of Precision Engineering and Manufacturing, 17(9), 1209–1214. https://doi.org/10.1007/s12541-016-0145-2

Le, M. T., & Nguyen, D. T. (2021). Building a support system for time study to calculate the standard time at production line. Proceedings of 2021 International Conference on System Science and Engineering, ICSSE 2021, 437–441. https://doi.org/10.1109/ICSSE52999.2021.9538472

Li, N., Zhang, D., Li, Y., & Zhang, Q. (2025). Repetitive assembly basic action detection and standard work measurement based on deep learning. International Journal of Computer Integrated Manufacturing, 38(5), 696–714. https://doi.org/10.1080/0951192X.2024.2372281

Long, D., & Yang, S. (2025). LightPose: A lightweight fatigue-aware pose estimation framework. Journal of Industrial Information Integration, 48(October), 100988. https://doi.org/10.1016/j.jii.2025.100988

Mares-Castro, A., Calzada-Ledesma, V., Becerra-Rodríguez, M. B., Santiago-Montero, R., & Estrada-Monje, A. (2025). Intelligent Motion Classification via Computer Vision for Smart Manufacturing and Ergonomic Risk Prevention in SMEs. Applied Sciences (Switzerland), 15(20), 1–19. https://doi.org/10.3390/app152010914

Mazloumi, A., Kazemi, Z., Abedzadeh, S., & Foroushani, A. R. (2021). Design and fabrication of portable motion recorder: comparing lumbar kinematics in workers with and without low back pain in assembly line of a car manufacturing industry. Journal of Health and Safety at Work, 11(1), 55–71.

Mor, R. S., Bhardwaj, A., Singh, S., & Sachdeva, A. (2019). Productivity gains through standardization-of-work in a manufacturing company. Journal of Manufacturing Technology Management, 30(6), 899–919. https://doi.org/10.1108/JMTM-07-2017-0151

Nesterak, J., Szelągowski, M., & Radziszewski, P. (2025). Workplace performance measurement: digitalization of work observation and analysis. Journal of Intelligent Manufacturing, 36(5), 3569–3585. https://doi.org/10.1007/s10845-024-02419-x

Pan, P., McDonald, T., Smidt, M., & Dias, R. (2022). Use of vision and sound to classify feller-buncher operational state. International Journal of Forest Engineering, 33(2), 129–138. https://doi.org/10.1080/14942119.2022.2037927

Panahi, R., Louis, J., Podder, A., Swanson, C., Pless, S., & Mirhasani, S. (2025). Monitoring installation of partially occluded subassemblies in modular construction factories using BIM, ray tracing, and computer vision. Construction Robotics, 9(1), 1–18. https://doi.org/10.1007/s41693-024-00148-4

Petersdorff, H. A. von. (2011). Automated real-time performance feedback and time study using computer vision technology (Issue October). Stellenbosch University.

Sitthidetchtamroung, N., & Khamkanya, A. (2024). Mathematical Model Development for AI Gesture Classification in Industrial Motion and Time Study. 8th International Conference on Information Technology 2024, InCIT 2024, 793–797. https://doi.org/10.1109/InCIT63192.2024.10810546

Wang, P., Liu, H., Wang, L., & Gao, R. X. (2018). Deep learning-based human motion recognition for predictive context-aware human-robot collaboration. CIRP Annals, 67(1), 17–20. https://doi.org/10.1016/j.cirp.2018.04.066

Waseem, M., Ghani, U., Habib, T., Noor, S., & Khan, T. (2021). Productivity Enhancement at Molding Compound Manufacturing Plant by Applying Time and Motion Analysis. Mehran University Research Journal of Engineering and Technology, 40(4), 761–774. https://doi.org/10.22581/muet1982.2104.07

Wu, S., Wang, Y., BolaBola, J. Z., Qin, H., Ding, W., Wen, W., & Niu, J. (2016). Incorporating motion analysis technology into modular arrangement of predetermined time standard (MODAPTS). International Journal of Industrial Ergonomics, 53, 291–298. https://doi.org/10.1016/j.ergon.2016.03.001

Machine Learning–Based Housing Price Prediction in King County: An Interpretable XGBoost Framework

Ersin NAMLI[1]
Mustafa GÜLER[2]

1- Assoc. Prof. Dr .Ersin Namlı; Istanbul University-Cerrahpasa, Industrial Engineering Department. enamli@iuc.edu.tr ORCID No: 0000-00001-5980-9152
2- Assist. Prof. Dr Mustafa Güler,. Istanbul University-Cerrahpasa, Department of Engineering science. m.guler@iuc.edu.tr ORCID No:000-0002-9289-5088

ABSTRACT

The ability to reliably predict housing prices is fundamental to informed financial decision-making and the stability of the broader macroeconomic environment. This study develops a high-performance predictive framework for residential property valuation in King County, Washington, using an optimized Extreme Gradient Boosting (XGBoost) regressor. Utilizing a dataset of 21,613 observations, the research employs a systematic pipeline that includes unsupervised anomaly detection via Isolation Forest, model-based feature selection, and hyperparameter optimization through RandomizedSearchCV.

The experimental results demonstrate that the final tuned model achieves strong predictive power, yielding an R^2 of approximately 0.90 and a Mean Absolute Percentage Error (MAPE) of 12%. Feature importance analysis indicates that construction quality (grade), interior living area (sqft_living), and precise geospatial coordinates (latitude and longitude) are the primary drivers of price formation. However, residual diagnostics reveal significant heteroscedasticity in the upper price tail, suggesting that luxury property valuations are influenced by intangible attributes and architectural uniqueness not fully captured by standard structural descriptors. The findings confirm the utility of machine learning for automated valuation while highlighting the necessity of segmented modeling or variance-stabilizing transformations for the high-end market.

Keywords – Housing Price Prediction; XGBoost; Machine Learning; Feature Selection; Anomaly Detection.

INTRODUCTION

The real estate sector remains one of the most critical components of both individual wealth and the global economy. Accurately predicting house prices is not only essential for optimizing the financial decisions of buyers and sellers but also plays a vital role in maintaining macroeconomic stability. Historical data suggests that sudden collapses in real estate prices can jeopardize the stability of the banking system and trigger systemic financial crises (Deng et al., 2019). Therefore, understanding and forecasting price movements in the housing market serves as a strategic tool for managing broader economic risks.

Price formation in the housing market is a multi-layered process shaped by more than just physical characteristics; it involves complex evaluation procedures (Plante, 2022). Determining a property's value requires a combined assessment of basic elements—such as property type and location—alongside intrinsic features like sunlight exposure and views, as well as extrinsic factors like proximity to services and infrastructure

(Cardone et al., 2024; Loro et al., 2024). However, the sheer volume of market information, combined with the "bounded rationality" of human actors, makes manual price analysis difficult and time-consuming (Isler et al., 2021). Because of these limitations, real estate agencies and investors increasingly need technological solutions that minimize error margins, strengthen negotiation power, and prevent the risk of overpayment (Hernes et al., 2024).

Today, machine learning algorithms offer the capacity to model the intricate relationships within large datasets in ways that traditional methods cannot (Maia & David, 2024). These advanced techniques provide a wide range of solutions, from analyzing how environmental regulations like energy performance certificates affect prices (Dell'Anna, 2025) to processing thousands of online listings to create objective market maps (Hromada, 2015). This study aims to address the challenges of house price prediction by exploring the capabilities of the XGBoost algorithm within a methodological framework to develop high-accuracy forecasting models.

Literature Review

Table 1: General Machine Learning Trends in Housing Price Prediction

Model/Technique	Key Features & Practices	Typical Use Cases	Performance Trends	Notable Limitations
Linear Regression	Simple, interpretable, low cost	Baseline, small datasets	Underperforms on nonlinearity	Limited accuracy
Random Forest	Ensemble, robust to overfitting, feature importance	Medium/large datasets, rental/sales	High accuracy, good for moderate complexity	Can be slow, less effective on high-dimensional data
Gradient Boosting	Sequential ensemble, handles nonlinearity	Complex feature sets	High accuracy, stable	Sensitive to tuning
XGBoost	Regularized boosting, parallelizable, handles missing data	High-dimensional, structured data	Superior accuracy, fast, robust	Interpretability, computational cost
LightGBM	Efficient boosting, handles large data	Large-scale, tabular	High accuracy, fast	May overfit, less interpretable
ANN/LSTM	Deep learning, captures complex patterns	Temporal/spatial data	High accuracy for time series	Requires large data, less interpretable
AutoML (Google, H2O)	Automated model selection/tuning, explainability	Diverse datasets	Competitive, user-friendly	Black-box risk, limited customization
Hybrid/Stacked Models	Combines strengths of multiple models	Complex, multi-source data	Best accuracy, robust	High complexity, interpretability

Recent housing price prediction studies increasingly favor ensemble learning and, where appropriate, deep learning approaches because they capture non-linear relationships and complex feature interactions more effectively than classical linear models. Table 1 synthesizes common choices: linear regression remains a lightweight baseline, while tree ensembles (Random Forest, Gradient Boosting, XGBoost, and LightGBM) dominate structured/tabular settings due to strong accuracy and robustness, especially when paired with careful preprocessing and tuning. Deep learning (ANN/LSTM) is most competitive when temporal or spatial dependencies are explicitly modeled. AutoML tools offer pragmatic model selection and tuning for diverse datasets but raise transparency and customization concerns. Hybrid/stacked frameworks often achieve the best accuracy at the cost of greater complexity and reduced interpretability. (Chen & Li, 2022; Kalaivani et al., 2023; Sood et al., 2023; Yan & Zong, 2020; Abdul-Rahman et al., 2021; Leo Joseph et al., 2024).

Table 1 suggests a clear trade-off frontier. For typical tabular housing datasets, boosted tree methods (particularly XGBoost/LightGBM) tend to provide state-of-the-art performance, but their sensitivity to hyperparameters and limited transparency necessitate rigorous tuning and post-hoc explainability. In contrast, Random Forest remains a robust alternative under moderate complexity and noisy features, while LSTM-based architectures are better suited to price series or dynamic market signals. Consequently, model selection should be driven by data modality (tabular vs. temporal/spatial), computational constraints, and the required level of interpretability for deployment.

Table 2: Comprehensive Table of Real Estate Volatility & Forecasting Studies

Study (Author, Year)	Region / Market	Forecasting / Modeling Approach	Key Findings (High-level)	Evaluation Metrics (as stated)
Behera et al., 2024	Global (financial volatility literature)	Review of ML-based volatility forecasting (regression, ensembles, deep learning)	Highlights that combining ML with richer information sources can improve volatility forecasting;	Varies by reviewed studies (not fixed in the reference metadata).
Gooi et al., 2018	Malaysia (real estate stock volatility)	Smooth Transition Exponential Smoothing (STES) vs ad-hoc and GARCH-type benchmarks	Reports strong forecasting performance of STES for Malaysian real estate stock volatility in comparative settings.	MAE (additional metrics not specified in the provided reference text).
Liow & Ibrahim, 2010	International (securitized real estate markets)	Volatility decomposition and correlation analysis (Component-GARCH family)	Meaningful volatility co-movements and correlations across international securitized real estate markets.	Not specified in the provided reference text.
Cotter & Stevenson, 2007	U.S. (REIT returns)	GARCH-family modeling of daily REIT volatility dynamics	Shows that market sentiment is a key driver of daily REIT volatility dynamics.	Not specified in the provided reference text.
Wheaton, 2015	U.S. (real estate markets)	Conceptual/empirical decomposition of real estate market volatility	Decomposes volatility into components related to demand- and supply-side factors.	Not specified in the provided reference text.
Alburshaid & Al-Alawi, 2024	Real estate forecasting literature	Review of forecasting methods (statistical + ML)	Surveys forecasting approaches and motivates hybridization and richer data integration	MAE, MSE, RMSE, MAPE
Zheng et al., 2025	China (policy shock; spillovers to upstream firms)	Quasi-natural experiment on volatility spillover effects ("Three Red Lines" policy)	Finds that the policy shock increases stock price volatility of upstream firms.	Not specified in the provided reference text.
Li & Kao, 2022	U.S. (COVID-19 period housing values)	Spatial analysis / spatial modeling of housing value changes	Finds spatially uneven housing value changes during COVID-19	Not specified in the provided reference

				text.
Ecevit et al., 2022	Turkey (Istanbul)	DT / RF / LR with spatial features for price prediction	Reports that Random Forest outperforms alternatives and improves predictive performance.	Accuracy, Precision, Recall, F1-Score
Simlai, 2014	U.S. (Boston housing)	Spatial ARCH (SARCH) model for conditional heteroskedasticity	Introduces SARCH to capture conditional spatial variability of housing prices	Not specified in the provided reference text.
Cheong et al., 2011	UK / broader market linkages	Empirical analysis of real estate risk and market volatility linkages	Shows that real estate sector risk can influence broader market volatility, particularly in stress periods.	Not specified in the provided reference text.
Wong et al., 2007	Hong Kong (spot & forward real estate markets)	Bivariate GARCH for volatility transmission	Finds volatility transmission between the real estate forward and spot markets.	Not specified in the provided reference text.
Gu et al., 2017	Global real estate markets	Multivariate GARCH with spatial effects (MGARCH + spatial structure)	Demonstrates time-varying correlations and contagion-like dependence patterns.	Not specified in the provided reference text.
Yao & Fotheringham, 2016	Scotland (house prices)	Local spatiotemporal mixed-model approach	Shows that explicitly modeling spatiotemporal relationships improves local prediction.	Not specified in the provided reference text.
Sachdeva et al., 2022	U.S. (housing neighborhoods)	Multiscale Geographically Weighted Regression (MGWR)	Quantifies intrinsic neighborhood value and identifies multiscale housing submarkets.	Not specified in the provided reference text.
Behera et al., 2023	Macroeconomic volatility (big data context)	Big-data analytics for macro indicators and volatility insight	Highlights the usefulness of large-scale/unstructured signals in understanding volatility regimes and macroeconomic dynamics.	Not specified in the provided reference text.

Across the reviewed studies in Table 2, two dominant methodological directions emerge. First, volatility-focused work in securitized real estate and REITs commonly relies on GARCH-family models (including component and multivariate extensions) to characterize time-varying risk and cross-market transmission. Second, housing-market valuation and forecasting increasingly incorporate spatial structure via econometric (e.g., SARCH) and geographically weighted approaches (e.g., MGWR), reflecting strong submarket heterogeneity and local dependence. Recent reviews further suggest that machine learning and hybrid pipelines can enhance performance when rich auxiliary signals (macroeconomic indicators, sentiment, or unstructured data) are available, although gains must be balanced against interpretability requirements and the availability of transparent evaluation metrics.

METHODOLOGY

Extreme Gradient Boosting (XGBoost) Regressor

In this study, ensemble machine learning techniques are utilized to maximize predictive performance and analyze complex data structures by combining multiple models into a single, superior predictive framework. The foundation of this approach is the boosting principle, which sequentially builds a series of "weak learners" models only slightly better than random guessing to develop a single, highly accurate "strong learner". According to Schapire & Freund (2014), this algorithm operates by training models in succession, where each new model is designed to correct the errors of its predecessors by assigning higher weights to misclassified examples. The final prediction is then determined by a weighted vote of all weak learners in the ensemble. Gradient boosting further generalizes this process by training subsequent models directly on the residuals (errors) of the previous models, typically employing decision trees as the base learners.

For the methodological framework of this research, XGBoost (Extreme Gradient Boosting) was selected as a highly optimized implementation of the gradient boosting algorithm. As highlighted by Wade (2020), XGBoost distinguishes itself through its computational speed and the inclusion of regularization terms that penalize model complexity to prevent overfitting. The core objective of the algorithm is to minimize a learning objective that combines a loss function, which measures prediction error,

and a regularization term, which measures model complexity. By optimizing this combined objective, XGBoost effectively manages the bias-variance tradeoff to produce a model that is both accurate and robust for generalization.

The predictive model was developed using Extreme Gradient Boosting (XGBoost**)** in a regression setting. XGBoost is a scalable implementation of gradient-boosted decision trees, where the final prediction is formed as an additive ensemble of weak learners (trees). At iteration *t*, the model prediction is updated as

$$\widehat{y_i^{(t)}} = \widehat{y_i^{(t-1)}} + f_t(x_i)$$

where f_t is the newly added regression tree. The model is optimized by minimizing a regularized objective function:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} \ell\left(y_i, \widehat{y_i^{(t)}}\right) + \sum_{k=1}^{t} \Omega(f_k)$$

where $\ell(\cdot)$ denotes the training loss (squared error in this work), and $\Omega(f)$ is a regularization term controlling model complexity . At each boosting step, XGBoost fits the new tree to the negative gradient of the loss function, enabling the model to capture non-linear relationships and higher-order interactions among housing attributes such as living area, quality grade, and location. Model performance was evaluated on a held-out test set using multiple regression metrics including MAE, MAPE, RMSE, Pearson's R and R^2

To mitigate the influence of atypical transactions and non-representative observations, an Isolation Forest (IF) algorithm was employed as an unsupervised anomaly detection step prior to model training. Isolation Forest identifies outliers by exploiting the principle that anomalous instances are easier to isolate than regular observations. In addition to this to improve parsimony and reduce noise, a model-based feature selection strategy was applied using the baseline XGBoost model's split-based feature importance scores. Predictors were ranked by importance and the top 12 variables were retained for the subsequent modeling stages. Finally, hyperparameter optimization was performed using RandomizedSearchCV with 3-fold cross-validation, searching over n estimators, "max_depthmax", and learning rate. The best configuration was selected according to the cross-validated scoring criterion default R^2 in scikit-learn and the resulting tuned model was used to generate final test-set predictions and reports.

.

RESULTS AND DISCUSSION

Dataset Description

This study uses the Housing Price Dataset hosted on Kaggle, which contains residential property sale records with structural characteristics (e.g., bedrooms, bathrooms, living area), quality/amenity indicators (e.g., grade, condition, view, waterfront), and geospatial identifiers (zipcode, latitude, longitude) designed for supervised price prediction tasks.

In the version used in our experiments (Housing.csv, provided in the dataset package), there are 21,613 observations and 21 variables, including the target price and 20 predictors: bedrooms, bathrooms, sqft_living, sqft_lot, floors, waterfront, view, condition, grade, sqft_above, sqft_basement, yr_built, yr_renovated, zipcode, lat, long, sqft_living15, and sqft_lot15, along with an identifier (id) and transaction date (date). These variables jointly capture property size, build/renovation information, neighborhood proxies (the "15" variables), and location effects that are central to real estate valuation modeling.

The dataset corresponds to King County, Washington (Seattle metro area) house sales and is widely used as a benchmark for regression models in housing valuation research.

1. Exploratory Analysis of Price Drivers

1.1 Correlation Structure of Cleaned Features

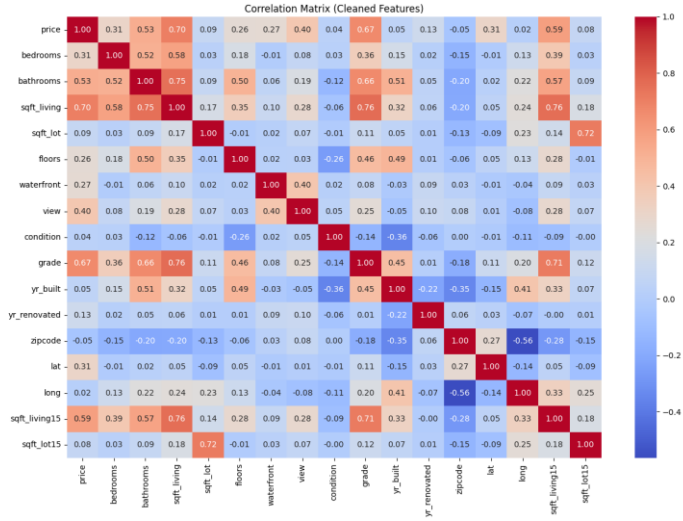


Figure 1. Correlation Matrix (Cleaned Features).

Figure 1 shows that price is most strongly associated with interior living area (sqft_living; r ≈ 0.70) and construction/finish quality (grade; r ≈ 0.67). Neighborhood proxy variables such as sqft_living15 also show a substantial association (r ≈ 0.59). Several predictors display high intercorrelations (e.g., sqft_living with bathrooms and grade), indicating redundancy that may distribute importance across correlated variables in tree-based models.

1.2 Nonlinear and Interaction Patterns: Size and Quality

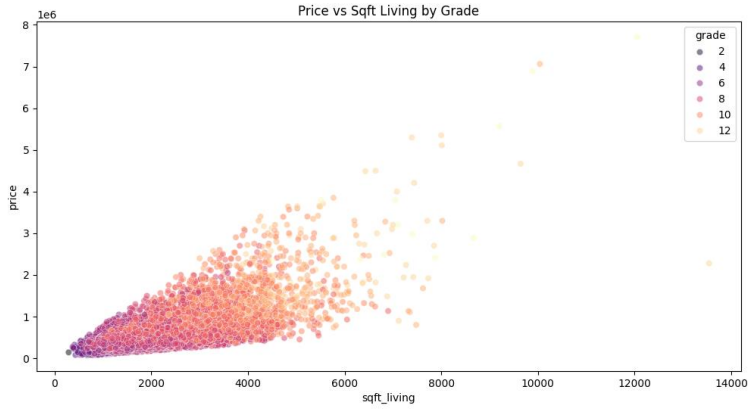


Figure 2. Price vs. Sqft Living by Grade.

As illustrated in Figure 2, price increases with living area, but the relationship is stratified by grade. For a given sqft_living, higher-grade properties systematically command higher prices, suggesting strong interaction effects where the marginal value of size depends on build quality. Dispersion increases at larger sizes, reflecting heterogeneity in the luxury segment.

2. Amenity and Location Effects

2.1 Waterfront and View Premiums

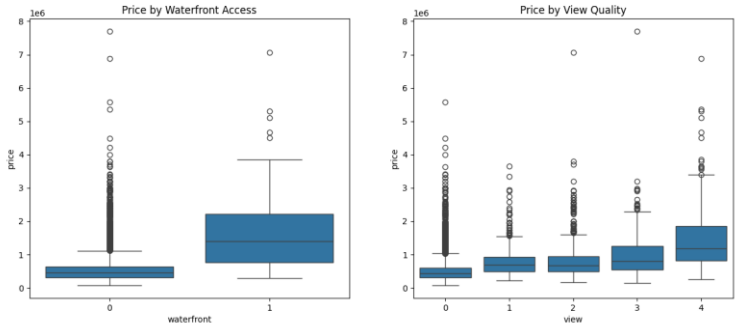


Figure 3. Price by Waterfront Access and Price by View Quality.

Figure 3 indicates pronounced amenity premiums. Waterfront access is associated with a markedly higher median price and broader distribution, while view quality exhibits an increasing median price pattern across categories. The widening spread for higher view levels suggests that amenities interact with size, quality, and micro-location to generate substantial valuation variance.

2.2 Spatial Price Concentration

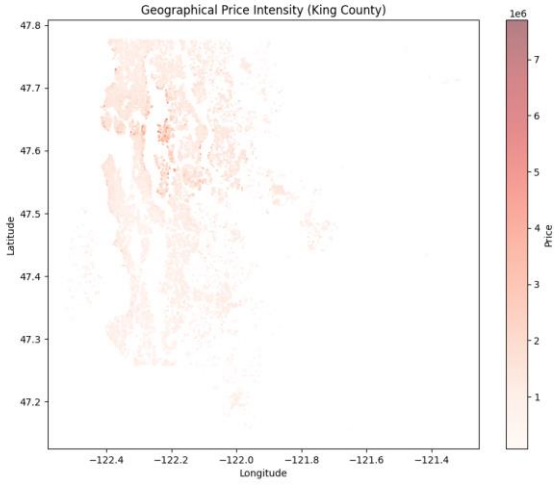


Figure 4. Geographical Price Intensity (King County).

The geographical intensity plot (Figure 4) reveals clear clustering of high-price observations, consistent with neighborhood-level segmentation. This supports incorporating location signals (e.g., latitude/longitude and appropriately encoded zipcode/neighborhood proxies) to capture spatial structure in housing valuations.

3. Model Performance and Interpretation

3.1 Comparative Model Performance Across Stages

Table 3. Model Performance Across Modeling Stages (Test Set).

Stage	MAE	MAPE	R	R2	RMSE
Before Feature Selection	61,509.76	12.26	0.944	0.892	99,480.50
Selected Features	61,829.16	12.29	0.944	0.890	99,995.87
After Parameter Optimisation	60,050.09	12.09	0.949	0.900	95,512.59

Table 3 summarizes predictive performance on the held-out test set (n = 4,193). Hyperparameter optimization produces the strongest generalization, improving RMSE and R^2 relative to the baseline. Feature selection alone does not yield improvements, suggesting that tree-based learners can leverage weak predictors via interactions or that removed variables contained complementary signal.

3.2 Feature Contribution in the Final Model

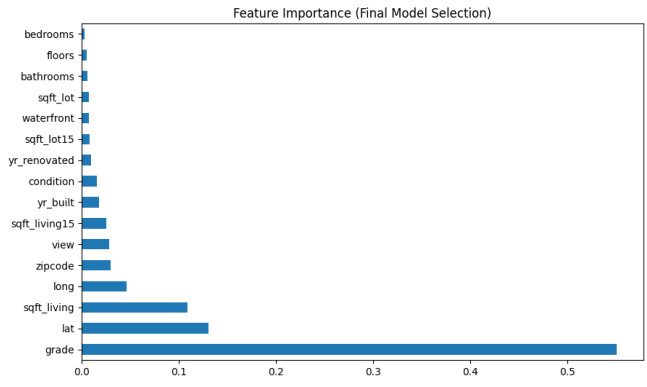


Figure 5. Feature Importance (Final Model Selection).

Figure 5 shows a concentrated importance profile. Grade is the dominant predictor, followed by latitude and sqft_living, indicating that quality, location, and size jointly explain most variance in price. Secondary contributions arise from other location-related signals (longitude, zipcode) and amenity/structure variables (view, yr_built, condition), while bedrooms exhibits comparatively limited incremental importance due to redundancy with size and bathroom counts.

4. Error Diagnostics

4.1 Residual Distribution

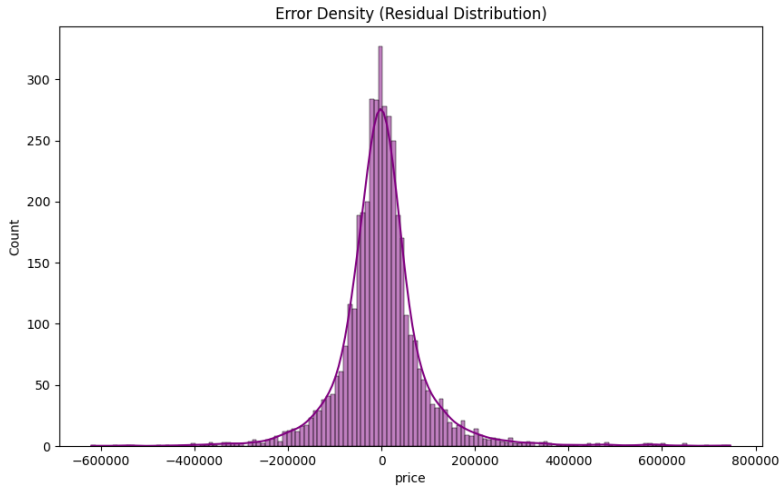


Figure 6. Error Density (Residual Distribution).

Residuals are centered near zero, indicating limited global bias, but exhibit heavy tails. This pattern is expected in housing markets with a long-tailed price distribution, where rare luxury properties and unique amenity combinations can produce large errors.

4.2 Actual–Predicted Agreement and Heteroscedasticity

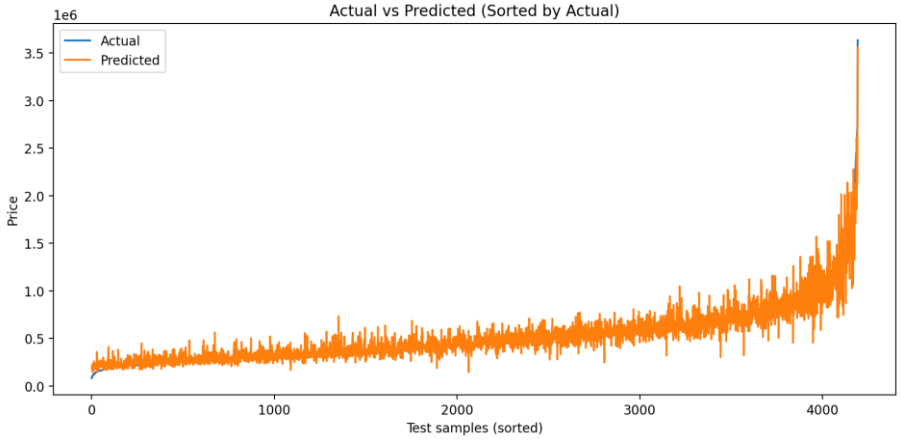


Figure 7. Actual vs. Predicted (Sorted by Actual).

Figure 7 demonstrates that predictions track the overall pattern of actual prices across the test range, with larger deviations emerging in the upper tail. This is consistent with increased volatility and scarcity of comparable examples for high-end properties.

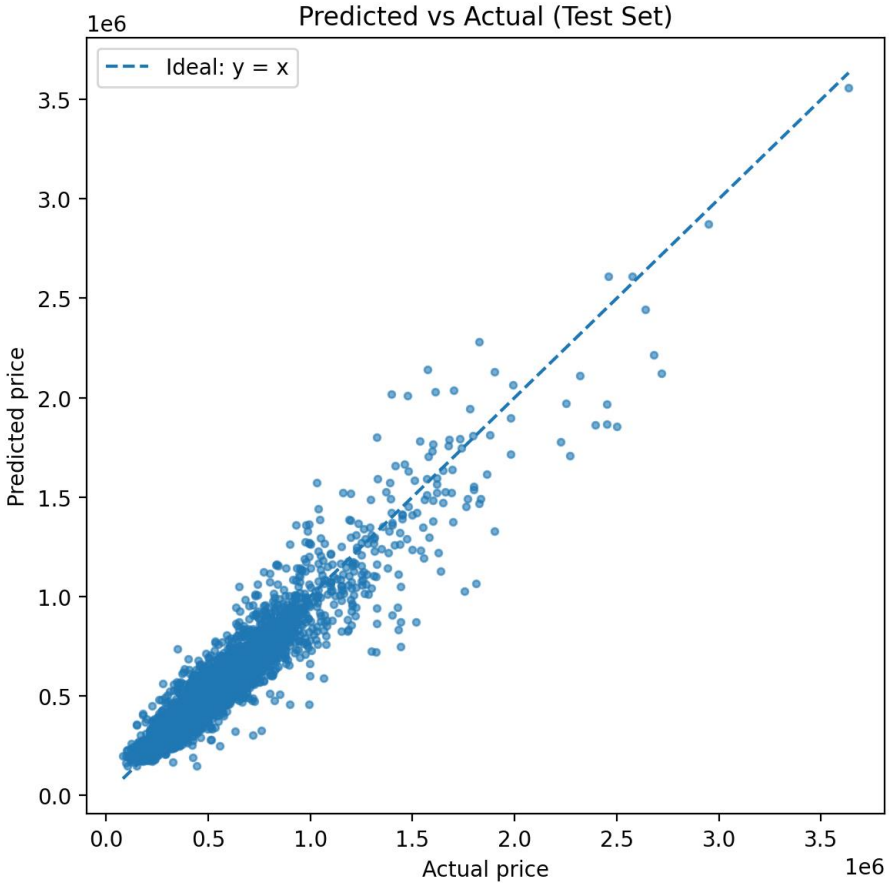


Figure 8. Predicted vs. Actual (Test Set).

The scatter plot (Figure 8) shows most observations clustered near the ideal y = x line, supporting strong overall fit. Dispersion increases at higher price levels, indicating heteroscedasticity and higher uncertainty in expensive segments.

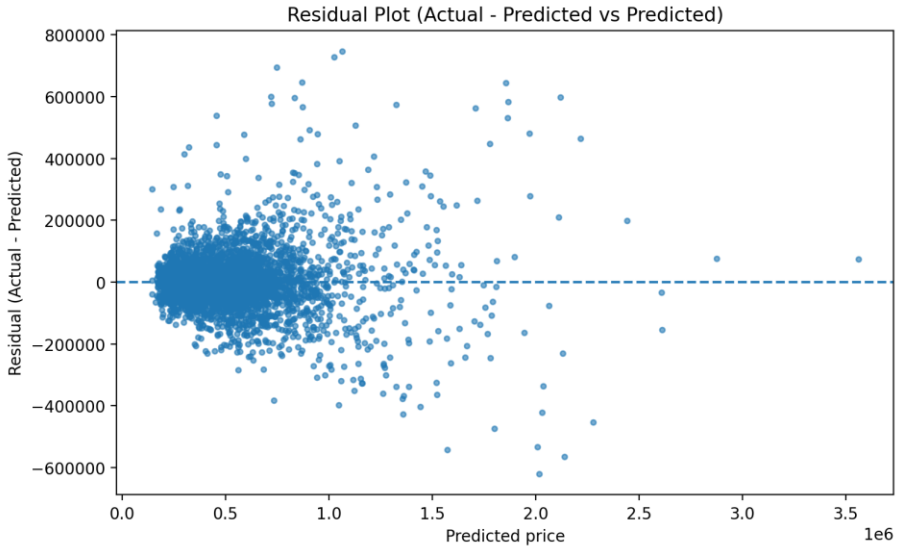


Figure 9. Residual Plot (Residuals vs. Predicted).

Figure 9 further confirms heteroscedasticity: residual variance increases with predicted price. While residuals remain broadly centered around zero (no strong monotonic bias), the widening spread suggests that variance-stabilizing transformations (e.g., log(price)) or segment-specific modeling may further improve calibration in the upper tail.

Analysis and Strategic Insights

The empirical results indicate that King County housing prices are driven by a hierarchy of structural, quality, and geospatial factors. Feature redundancy was notable among size-related variables; retaining "sqft_living" while deprioritizing correlated sub-components improved model parsimony and interpretability without degrading predictive performance. Incorporating Isolation Forest for anomaly detection and removing approximately 3% extreme observations reduced residual dispersion and improved out-of-sample stability.

From a market perspective, geospatial position (latitude/longitude) and construction quality (grade) consistently outweighed traditional count-based metrics (e.g., bedrooms), aligning with a strong spatial segmentation in local demand. While the optimized model achieved strong overall accuracy (MAPE ≈ 12%, $R^2 \approx 0.90$), residual variance increased markedly in the upper price tail, suggesting that luxury valuations depend on unobserved, intangible attributes not captured by the available numerical features.

Finally, the performance gain from baseline to the tuned model confirms the value of hyperparameter optimization in capturing nonlinear interactions between location and quality. Future work should consider log-target modeling and/or segmented approaches for the high-end market to address the observed heteroscedasticity

CONCLUSION

The objective of this study was to develop a high-performing predictive framework for residential property valuation in King County using an optimized XGBoost regressor. Through a systematic pipeline—including redundancy-aware feature handling, anomaly filtering via Isolation Forest, and hyperparameter optimization—the model achieved an out-of-sample R^2 of 0.90 on the held-out test set, indicating strong explanatory power for housing prices.

Results show that living area and construction quality (grade), together with precise geospatial positioning, dominate price formation, consistent with micro-market segmentation. Residual diagnostics reveal heteroscedasticity, with substantially larger errors in the upper price tail, suggesting that extreme valuations depend on unobserved attributes (e.g., architectural uniqueness, premium finishes, or other intangible factors) not fully represented in standard structural descriptors.

Future work should consider log-scale modeling of the target to stabilize variance and improve relative accuracy across price strata, as well as segmented modeling strategies that explicitly address the high-end market. Overall, the findings confirm the utility of machine learning for automated valuation while highlighting persistent challenges in modeling the extreme luxury tail with purely quantitative features.

REFERENCES

Abdul-Rahman, S., Mutalib, S., Zulkifley, N. H., & Ibrahim, I. (2021). Advanced machine learning algorithms for house price prediction: Case study in Kuala Lumpur. International Journal of Advanced Computer Science and Applications, 12(2), 550–557.

Alburshaid, E. A., & Al-Alawi, A. I. (2024). Forecasting methods for the real estate market: A review. In Proceedings of the 2024 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS 2024) (pp. xx–xx).

Brar, S. S. (Year). Housing Price Dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/sukhmandeepsinghbrar/housing-price-dataset

Behera, I., Nanda, P., Mitra, S., & Kumari, S. (2024). Machine learning approaches for forecasting financial market volatility. Intelligent Systems Reference Library. Springer.

Behera, I., Nanda, P., Mitra, S., Senapati, S., et al. (2023). Analyzing market volatility: Big data's insight into macroeconomic indicators. In Proceedings of the 2023 International Conference on Data Science and Informatics (ICDSI 2023) (pp. xx–xx).

Cardone, B., Di Martino, F., & Senatore, S. (2024). Real estate price estimation through a fuzzy partition-driven genetic algorithm. Information Sciences, 667, 120442. https://doi.org/10.1016/j.ins.2024.120442

Cheong, C. S., Olshansky, A., & Zurbruegg, R. (2011). The influence of real estate risk on market volatility. Journal of Property Investment and Finance, 29(2), 145–158.

Chen, D., & Li, R. Y. M. (2022). Predicting housing price in Beijing via Google and Microsoft AutoML. In Current State of Art in Artificial Intelligence and Ubiquitous Cities (pp. xx–xx). Springer.

Cotter, J., & Stevenson, S. (2007). Uncovering volatility dynamics in daily REIT returns. Journal of Real Estate Portfolio Management, 13(4), 327–344.

Dell'Anna, F. (2025). Machine learning framework for evaluating energy performance certificate (EPC) effectiveness in real estate: A case study of Turin's private residential market. Energy Policy, 198, 114407. https://doi.org/10.1016/j.enpol.2024.114407

Deng, Y., Zeng, Y., & Li, Z. (2019). Real estate prices and systemic banking crises. Economic Modelling, 80, 111–120. https://doi.org/10.1016/j.econmod.2018.11.029

Ecevit, M. I., Erdem, Z., & Dag, H. (2022). Reviewing the effects of spatial features on price prediction for real estate market: Istanbul case. In Proceedings of the 7th International Conference on Computer Science and Engineering (UBMK 2022) (pp. xx–xx).

Gooi, L.-M., Choo, W.-C., Nassir, A. M. D., & Ng, S.-I. (2018). Volatility forecasting of real estate stock in Malaysia with smooth transition exponential smoothing. International Journal of Economics and Management, 12(1), 61–77.

Gu, H., Liu, Z., & Weng, Y. (2017). Time-varying correlations in global real estate markets: A multivariate GARCH with spatial effects approach. Physica A: Statistical Mechanics and Its Applications, 473, 460–472. https://doi.org/10.1016/j.physa.2017.02.045

Hernes, M., Tutak, P., & Siewiera, M. (2024). Prediction of residential real estate price on primary market using machine learning. Procedia Computer Science, 246, 3142–3147. https://doi.org/10.1016/j.procs.2024.09.358
Hromada, E. (2015). Mapping of real estate prices using data mining techniques. Procedia Engineering, 123, 233–240. https://doi.org/10.1016/j.proeng.2015.10.090
Isler, O, Flew, T., Erol, I., & Dulleck, U. (2021). Market news and credibility cues improve house price predictions: An experiment on bounded rationality in real estate. Journal of Behavioral and Experimental Finance, 31, 100550. https://doi.org/10.1016/j.jbef.2021.100550
Kalaivani, K. S., Kanimozhiselvi, C. S., Bilal, Z. H. M., et al. (2023). A comparative study of regression algorithms on house sales price prediction. In Proceedings of the 2023 2nd International Conference on Augmented Intelligence and Sustainable Systems (ICAISS 2023) (pp. xx–xx).
Leo Joseph, L. M. I., Reddy, E. H., Srinidhi, M., et al. (2024). Predicting real-time house prices: A machine learning approach using XGBoost algorithm. In Asia Pacific Conference on Innovation in Technology (APCIT 2024) (pp. xx–xx).
Li, X., & Kao, C. (2022). Spatial analysis and modeling of the housing value changes in the U.S. during the COVID-19 pandemic. Journal of Risk and Financial Management, 15(9), 404. https://doi.org/10.3390/jrfm15090404
Liow, K. H., & Ibrahim, M. F. (2010). Volatility decomposition and correlation in international securitized real estate markets. Journal of Real Estate Finance and Economics, 40(2), 182–205.
Loro, S., Lo Verso, V. R. M., Fregonara, E., & Barreca, A. (2024). Influence of daylight on real estate housing prices: A multiple regression model application in Turin. Journal of Building Engineering, 96, 110413. https://doi.org/10.1016/j.jobe.2024.110413
Maia, W. F., & David, S. A. (2024). Machine learning applied to estate pricing for residential rentals in dynamic urban markets—The case of São Paulo city. Engineering Analysis with Boundary Elements, 169, 105988. https://doi.org/10.1016/j.enganabound.2024.105988
Plante, M. (2022). Price formation through real estate exchange: Informing and mediating evaluation with attachment. Management Accounting Research, 56, 100793. https://doi.org/10.1016/j.mar.2022.100793
Sachdeva, M., Fotheringham, S., & Li, Z. (2022). Do places have value? Quantifying the intrinsic value of housing neighborhoods using MGWR. Journal of Housing Research, 31(2), 159–183.
Schapire, R. E., & Freund, Y. (2014). Boosting: Foundations and algorithms. MIT Press.
Simlai, P. (2014). Estimation of variance of housing prices using spatial conditional heteroskedasticity (SARCH) model with an application to Boston housing price data. The Quarterly Review of Economics and Finance, 54(3), 317–325.
Sood, K., Shang, S., & Nijad, N. (2023). Predictive analytics for real estate during COVID-19 pandemic. In 2023 IEEE World AI IoT Congress (AIIoT 2023) (pp. xx–xx).
Wade, C. (2020). Hands-on gradient boosting with XGBoost and scikit-learn. Packt Publishing.
Wheaton, W. C. (2015). The volatility of real estate markets: A decomposition. Journal of Portfolio Management, 41(6), 89–101.

Wong, S. K., Chau, K. W., & Yiu, C. Y. (2007). Volatility transmission in the real estate spot and forward markets. Journal of Real Estate Finance and Economics, 34(2), 163–178.
Yan, Z., & Zong, L. (2020). Spatial prediction of housing prices in Beijing using machine learning algorithms. In ACM International Conference Proceeding Series (pp. xx–xx).
Yao, J., & Fotheringham, A. S. (2016). Local spatiotemporal modeling of house prices: A mixed model approach. The Professional Geographer, 68(3), 412–423.
Zheng, D., Gao, J., Zhang, W., & Kwon, J. (2025). When real estate trembles: Spillover effects on stock price volatility under China's "Three Red Lines" policy. Finance Research Letters, 56, 104157.

Interpretable and Explainable Machine Learning in Healthcare Decision Support Systems

Ferda BOZKURT[1]

1- Assistant Professor; Sakarya University of Applied Sciences, Sakarya Vocational School, Department of Computer Programming.
fbozkurt@subu.edu.tr ORCID No 0000-0002-8398-3780

ABSTRACT

The increasing adoption of machine learning techniques in healthcare has significantly enhanced the predictive capabilities of clinical decision support systems. However, the lack of transparency in many high-performing models has raised serious concerns regarding trust, accountability, and ethical responsibility. In response, interpretable and explainable machine learning approaches have emerged as essential components of modern healthcare analytics. This chapter presents a comprehensive review of interpretable and explainable machine learning methods applied to healthcare decision support systems. It discusses conceptual foundations, methodological frameworks, clinical applications, evaluation strategies, ethical challenges, and future research directions.

Keywords – Interpretable Machine Learning; Explainable Artificial Intelligence (XAI); Healthcare Decision Support Systems; Clinical Data Analytics; Model Transparency; Decision Trees; Feature Importance; SHAP; LIME; Ethical and Trustworthy AI in Healthcare.

INTRODUCTION

The healthcare sector is experiencing a paradigm shift driven by the integration of digital technologies, large-scale data collection, and artificial intelligence. Electronic health records, medical imaging systems, wearable sensors, and remote monitoring devices generate vast volumes of heterogeneous data. Leveraging these data effectively requires advanced computational methods capable of extracting clinically meaningful insights.

Machine learning (ML) has become a central technology in healthcare analytics, supporting tasks such as disease diagnosis, prognosis prediction, risk stratification, and treatment optimization (Esteva et al., 2019). Despite their success, many machine learning models—particularly deep neural networks and ensemble methods—are inherently opaque. This opacity conflicts with the fundamental requirements of healthcare practice, where clinicians must understand and justify decisions.

Recent studies emphasize that interpretability and explainability are not optional features but core requirements for deploying artificial intelligence in clinical environments (Doshi-Velez and Kim, 2017). As a result, interpretable and explainable machine learning has emerged as a critical research area aimed at bridging the gap between predictive performance and clinical usability.

This chapter aims to provide a structured and comprehensive overview of interpretable and explainable machine learning approaches in healthcare decision support systems.

CLINICAL DECISION SUPPORT SYSTEMS: BACKGROUND AND EVOLUTION

Clinical decision support systems (CDSS) are computer-based tools designed to assist healthcare professionals in clinical decision-making by integrating patient-specific information with medical knowledge (Shortliffe and Cimino, 2014). A 2008 study addressed the role of early CDSS in tackling clinical problems in areas such as diabetes management.

Early Rule-Based Systems

Early CDSS relied on symbolic artificial intelligence and expert systems that encoded medical knowledge as explicit rules. These systems were highly interpretable but limited in scalability and adaptability (Musen et al., 2014). Empirical studies from the late 2000s demonstrated that, despite their conceptual appeal, rule-based CDSS had only modest effects on clinical outcomes, particularly in chronic disease management. In the context of diabetes care, early implementations of computerized reminders and decision support tools were often poorly integrated into clinical workflows, developed reactively to meet care standards, and appended to existing health information systems, which constrained their real-world effectiveness (Lester et al., 2008).

Similarly, CDSS embedded within Electronic Medical Records showed limited adoption and impact in ambulatory settings due to usability issues, insufficient contextual relevance, and misalignment with clinicians' documentation practices (Schnipper et al., 2008). Beyond disease-specific applications, early CDSS were also deployed for broader patient monitoring in settings such as nursing homes, where alert-based systems supported condition surveillance but remained largely descriptive rather than predictive (Alexander, 2008).

At the same time, qualitative studies highlighted that physician expectations, attitudes, and resistance to workflow disruption were major barriers to CDSS acceptance, underscoring the socio-technical nature of CDSS implementation (Varonen et al., 2008; Toth-Pal et al., 2008). To address these limitations, research began to emphasize modular, interoperable CDSS architectures and improved knowledge management strategies, aiming to decouple inference mechanisms from knowledge representation and enable greater system flexibility (Kim et al., 2008; Wu et al., 2008).

Overall, the limitations of early symbolic CDSS—particularly their rigidity, maintenance burden, and dependence on manually encoded rules—motivated a gradual shift toward data-driven and machine learning–based approaches, which promised improved scalability while raising new challenges related to transparency and interpretability.

Data-Driven CDSS

The emergence of data-driven approaches marked a fundamental shift from traditional knowledge engineering toward machine learning–based clinical decision support systems (Shortliffe and Sepúlveda, 2018; Topol, 2019). Unlike rule-based CDSS that rely on manually encoded expert knowledge, data-driven systems learn predictive patterns directly from large-scale clinical datasets such as electronic health records, laboratory results, medical imaging, and wearable sensor data (Esteva et al., 2019). This capability enables the modeling of complex, nonlinear relationships among clinical variables and supports more personalized, adaptive, and context-aware decision-making (Obermeyer and Emanuel, 2016; Topol, 2019).

Empirical studies have shown that machine learning–based CDSS can outperform traditional approaches in tasks such as disease risk stratification, outcome prediction, and treatment recommendation across multiple clinical domains (Rajkomar et al., 2018; Miotto et al., 2016). However, the increasing use of complex models—including ensemble methods and deep neural networks—has intensified concerns regarding interpretability, transparency, and clinical trust (Doshi-Velez and Kim, 2017). Many high-performing models operate as "black boxes," offering limited insight into how input features influence predictions, which poses challenges for clinical accountability, patient safety, and regulatory compliance (Holzinger et al., 2019).

The lack of interpretability has been repeatedly identified as a major barrier to clinician adoption of data-driven CDSS, as clinicians require explanations to validate recommendations, detect potential errors, and assess bias (Tonekaboni et al., 2019; Ribeiro et al., 2016). In response, research attention has increasingly shifted toward integrating explainability mechanisms into machine learning–based CDSS, leading to the development of Explainable Artificial Intelligence (XAI). Techniques such as feature attribution methods, local surrogate models, and visual explanation tools aim to balance predictive accuracy with clinical interpretability (Guidotti et al., 2018; Samek et al., 2019). Consequently, the evolution of data-driven CDSS reflects an ongoing trade-off between performance and transparency, highlighting the necessity of interpretable and trustworthy AI for effective clinical decision support.

MACHINE LEARNING PARADIGMS IN HEALTHCARE

Machine learning techniques applied in healthcare can be broadly categorized into several paradigms based on the availability of labeled data, learning objectives, and model complexity. Each paradigm offers distinct advantages and limitations in clinical decision support applications.

Supervised Learning

Supervised learning algorithms are the most widely adopted machine learning techniques in healthcare, as they are designed to learn mappings between labeled input data and known clinical outcomes. Common supervised learning methods include logistic regression, decision trees, support vector machines, random forests, and neural networks. These algorithms are extensively used for classification and regression tasks such as cancer detection, cardiovascular risk prediction, disease progression modeling, and mortality forecasting (Shickel et al., 2018; Rajkomar et al., 2018).

In clinical settings, supervised learning benefits from its relatively straightforward evaluation using labeled datasets and clinically meaningful outcome variables. Models such as decision trees and logistic regression offer a degree of interpretability, which is critical for clinical acceptance. However, supervised learning approaches are highly dependent on the availability and quality of annotated data, which can be costly, time-consuming, and prone to labeling bias. Moreover, models trained on retrospective datasets may exhibit limited generalizability across institutions or patient populations (Obermeyer and Emanuel, 2016).

Unsupervised Learning

Unsupervised learning methods operate without labeled outcome variables and aim to identify latent structures or patterns within clinical data. Techniques such as clustering, dimensionality reduction, and topic modeling are commonly employed for patient stratification, phenotype discovery, and exploratory data analysis (Miotto et al., 2016). These approaches are particularly valuable in complex, heterogeneous diseases where predefined labels may not fully capture disease subtypes or progression trajectories.

In healthcare, unsupervised learning has been applied to discover novel disease phenotypes, identify subgroups of patients with similar clinical characteristics, and reduce the dimensionality of high-dimensional data such as genomic profiles and electronic health records. Despite their exploratory power, unsupervised methods pose interpretability challenges, as the clinical meaning of derived clusters or latent representations is not always clear. Validation often requires extensive post hoc clinical analysis and expert

interpretation, which can limit their immediate applicability in routine clinical decision-making (Beam and Kohane, 2018).

Ensemble and Deep Learning Models

Ensemble learning methods, including random forests and gradient boosting machines, combine multiple base learners to improve predictive accuracy and robustness. Deep learning architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), are capable of modeling highly complex and nonlinear relationships in large-scale, multimodal healthcare data. These models have achieved state-of-the-art performance in applications such as medical image analysis, natural language processing of clinical notes, and disease risk prediction (Esteva et al., 2019).

However, the high predictive performance of ensemble and deep learning models often comes at the cost of reduced transparency and interpretability. Their internal representations and decision-making processes are typically opaque to clinicians, leading to the characterization of such systems as "black boxes" (Samek et al., 2019). For instance, while deep learning models have demonstrated high accuracy in lung cancer screening and radiological image interpretation, the lack of clear explanations for model predictions raises concerns regarding clinical trust, accountability, and patient safety. Consequently, the deployment of these models in clinical practice increasingly requires the integration of explainability techniques—such as feature attribution methods and visualization-based explanations—to provide insights into model behavior and support informed clinical decision-making (Ribeiro et al., 2016; Holzinger et al., 2019).

INTERPRETABILITY AND EXPLAINABILITY: CONCEPTUAL FOUNDATIONS

In the context of artificial intelligence and clinical decision support systems, the terms interpretability and explainability are often used interchangeably; however, they represent conceptually distinct yet complementary notions. Understanding this distinction is particularly critical in healthcare, where trust, accountability, and patient safety are central requirements.

Interpretability

Interpretability refers to the extent to which a model's internal mechanisms and decision-making processes can be directly understood by humans without the need for additional explanatory tools (Lipton, 2018). An interpretable model is inherently transparent: its structure, parameters, and logic are accessible and intelligible to domain experts. Classical examples include linear regression, decision trees, and rule-based expert systems, where the contribution of each input feature to the final prediction can be explicitly traced.

In healthcare applications, interpretability enables clinicians to understand how a model operates and why a specific output is produced, facilitating clinical validation and integration into medical reasoning. However, there is often a trade-off between interpretability and predictive performance, as highly interpretable models may struggle to capture complex, nonlinear relationships present in large-scale clinical data.

Explainability

Explainability, by contrast, focuses on providing human-understandable explanations for model outputs, even when the underlying model itself is complex or opaque (Guidotti et al., 2018). Explainable models do not necessarily expose their internal structure; instead, they offer post hoc explanations that describe why a particular decision was made, which features were most influential, and how reliable the prediction may be.

Explainability is therefore particularly relevant for black-box models such as deep neural networks and ensemble methods, which achieve high predictive accuracy but lack inherent transparency. Explainable Artificial Intelligence (XAI) frameworks aim to bridge this gap by introducing techniques such as feature attribution, local surrogate models, and visualization-based explanations. According to widely accepted XAI principles, explainable systems should support transparency, informativeness, meaningful user interaction, and auditability, especially in safety-critical domains like healthcare (Guidotti et al., 2018; Holzinger et al., 2019).

Importance of Interpretability and Explainability in Healthcare

The importance of interpretability and explainability in healthcare extends beyond technical considerations and directly impacts clinical, ethical, and regulatory dimensions.

First, explainability enhances *clinical trust and adoption*. Clinicians are more likely to rely on AI-driven recommendations when they can understand the rationale behind model outputs and assess whether they align with clinical knowledge and patient-specific context (Tonekaboni et al., 2019).

Second, explainability supports *error detection and model validation*. By revealing influential features and decision pathways, explainable models allow clinicians and developers to identify spurious correlations, data leakage, or hidden biases that could compromise patient safety (Doshi-Velez and Kim, 2017).

Third, interpretability and explainability are essential for *ethical and legal accountability*. In the event of an incorrect diagnosis or adverse clinical outcome, understanding how an AI system arrived at a decision is critical for determining responsibility and ensuring ethical use. Healthcare AI systems must therefore provide explanations that are not only technically sound but also comprehensible to medical professionals and patients.

Finally, explainability is increasingly required for *regulatory compliance*. Regulatory bodies and policy frameworks emphasize transparency and accountability in medical AI systems. For example, the European Commission has issued guidelines highlighting the necessity of explainable and trustworthy AI in healthcare applications, particularly in relation to patient rights, safety, and risk management (European Commission, 2020).

The conceptual difference between black-box models and interpretable models in healthcare can be illustrated as follows.

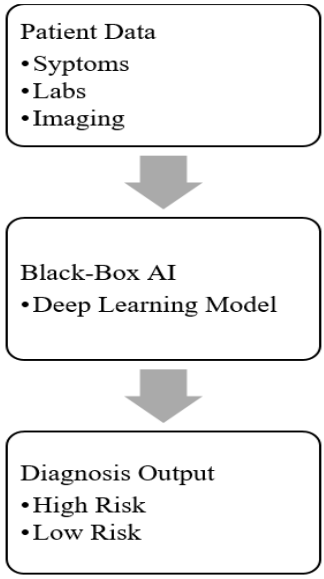


Figure 1: Black-Box Clinical Decision Support Model in Medical Reminder Support

Figure 1 illustrates a black-box clinical decision support model used in medical diagnosis, in which patient data such as symptoms, laboratory

results, and imaging findings are processed by a complex machine learning model. In this setting, the reasoning behind the model's predictions remains unclear, the influence of individual clinical features on the final decision is unknown, and the lack of transparency results in limited clinical trust. Consequently, clinicians may find it difficult to validate model outputs or confidently integrate them into clinical decision-making processes.

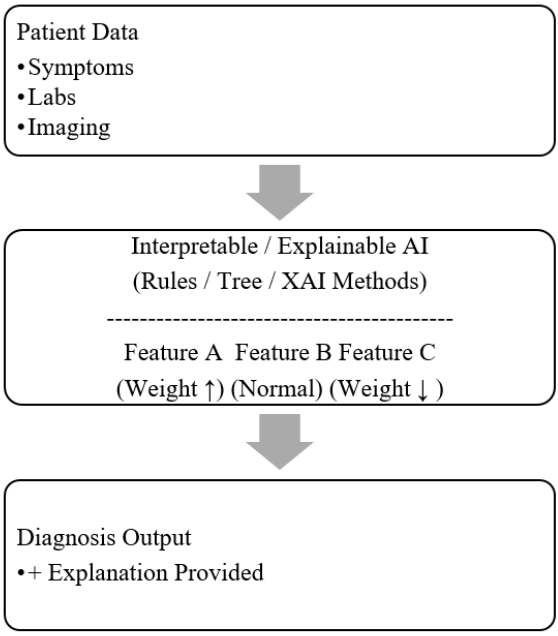


Figure 2: Interpretable and Explainable Clinical Decision Support Model in Medical Diagnosis

Figure 2 presents an interpretable and explainable clinical decision support model designed for medical diagnosis. In contrast to black-box systems, this model provides clear reasoning for its predictions, offers feature-level insight into how individual clinical variables contribute to the decision, and thereby supports improved clinical trust. By making the decision-making process transparent and understandable, such systems facilitate clinical validation, error detection, and more effective adoption in routine healthcare practice.

INHERENTLY INTERPRETABLE MACHINE LEARNING MODELS

Inherently interpretable machine learning models are characterized by their transparent internal structures, which allow humans to directly understand how inputs are transformed into outputs without the need for post hoc explanation techniques. In healthcare, such models are particularly valuable because they support clinical reasoning, facilitate validation, and align with ethical and regulatory requirements for transparency.

Decision Trees

Decision trees represent decision-making processes as hierarchical, tree-like structures composed of internal decision nodes and terminal leaf nodes. Each internal node corresponds to a test on a clinical feature, while each branch represents an outcome of that test, ultimately leading to a prediction at the leaf level. This structure closely resembles clinical decision-making workflows, making decision trees highly intuitive and interpretable for clinicians (Breiman et al., 1984).

In healthcare applications, decision trees have been widely used for cancer diagnosis, cardiovascular disease risk prediction, and the classification of respiratory disorders, among other tasks (Kourou et al., 2015). One of their primary advantages lies in their visual nature, which allows clinicians to trace the exact path from input features to the final decision. This transparency facilitates clinical validation, supports shared decision-making, and enhances trust in model recommendations.

Despite their interpretability, decision trees may suffer from overfitting, particularly when trees become overly complex. Moreover, single decision trees may exhibit limited predictive performance compared to ensemble-based methods, motivating their frequent use as base learners in more complex models.

Rule-Based Models

Rule-based models generate explicit IF–THEN rules that describe decision logic in a form that is readily understandable by clinicians. These rules can be derived from expert knowledge, data-driven learning processes, or a combination of both. Because rule-based systems explicitly encode decision logic, they align naturally with clinical guidelines and established medical protocols (Huysmans et al., 2011).

In healthcare, rule-based approaches have been applied to diagnostic support, treatment recommendation, and clinical alert systems. Variants such as fuzzy logic systems extend traditional rule-based models by handling uncertainty and imprecision in clinical data, making them particularly suitable for complex medical domains where binary logic may be insufficient. As a result, rule-based and fuzzy systems are often regarded as

reliable and transparent solutions within the broader framework of explainable artificial intelligence in medicine.

However, the main limitation of rule-based models lies in their scalability and maintenance burden. As the number of rules increases, systems may become difficult to manage and update, especially in rapidly evolving clinical environments.

Linear and Logistic Regression

Linear and logistic regression models remain among the most widely used techniques in clinical research and healthcare analytics due to their strong statistical foundations and inherent interpretability. In these models, coefficients directly quantify the relationship between input features and the outcome variable, providing clear insights into feature importance and directionality (Steyerberg et al., 2010).

In medical applications, logistic regression is particularly popular for binary outcome prediction tasks such as disease presence or absence, risk stratification, and prognostic modeling. The simplicity and transparency of regression models allow clinicians to easily interpret how individual variables influence predictions, which is especially critical in sensitive healthcare contexts involving diagnosis, prognosis, and treatment decisions.

While regression models may be limited in their ability to capture complex nonlinear interactions, their interpretability, robustness, and ease of validation continue to make them a cornerstone of clinical decision support systems and risk prediction models.

THE ACCURACY - INTERPRETABILITY TRADE-OFF

One of the fundamental challenges in interpretable machine learning is achieving an effective balance between predictive accuracy and model transparency. In many healthcare applications, highly accurate predictions are essential for early diagnosis and risk stratification; however, these predictions must also be understandable to clinicians in order to support trust, accountability, and safe clinical decision-making.

Simple and inherently interpretable models—such as linear regression, logistic regression, decision trees, and rule-based systems—offer clear and transparent decision logic. Their internal mechanisms can be directly examined, allowing clinicians to understand how individual clinical variables contribute to model outputs. However, these models often struggle

to capture complex, nonlinear interactions and high-dimensional relationships that are common in real-world healthcare data, potentially limiting their predictive performance.

In contrast, complex models such as ensemble methods and deep neural networks are capable of modeling intricate nonlinear patterns and interactions, often achieving state-of-the-art performance in tasks such as medical image analysis, disease prediction, and outcome forecasting. Despite their high accuracy, these models are typically opaque, making it difficult to explain how predictions are generated. As argued by Rudin (2019), reliance on black-box models in high-stakes domains like healthcare raises serious concerns, as lack of interpretability can undermine clinical trust, hinder error detection, and complicate ethical and legal accountability.

To mitigate this trade-off, hybrid approaches have been proposed that seek to combine the strengths of both paradigms. These approaches include using interpretable models where possible, constraining complex models to improve transparency, or augmenting high-performance models with explainability techniques. Examples include generalized additive models with interaction terms, attention-based neural networks with interpretable weights, and post hoc explanation methods applied to black-box predictors. Such strategies aim to preserve predictive performance while providing sufficient insight into model behavior to support clinical use.

Overall, the accuracy–interpretability trade-off represents a central design consideration in healthcare AI systems. Rather than pursuing accuracy alone, modern CDSS increasingly emphasize trustworthy performance, where predictive power is balanced with interpretability and explainability to meet clinical, ethical, and regulatory requirements.

Table 1: Comparison of Machine Learning Models in Terms of Accuracy and Interpretability in Healthcare

Model Type	Predictive Accuracy	Interpretability	Clinical Trust
Linear Regression	Low	Very High	High
Logistic Regression	Low–Moderate	Very High	High
Decision Trees	Moderate	High	Moderate–High
Ensemble Models	High	Low	Moderate
Deep Learning Models	Very High	Very Low	Low

As summarized in Table 1, there is a clear trade-off between predictive accuracy and interpretability in healthcare machine learning models. While deep learning models achieve very high accuracy, their low interpretability can limit clinical trust, whereas simpler models such as linear and logistic regression offer high transparency but lower predictive performance.

EXPLAINABLE ARTIFICIAL INTELLIGENCE FOR BLACK-BOX MODELS

Explainable Artificial Intelligence (XAI) for black-box models encompasses a set of techniques designed to make the predictions of complex machine learning and deep learning systems understandable to human users, without requiring direct access to the internal parameters or architecture of the model. Black-box models—such as deep neural networks, ensemble methods, and complex nonlinear classifiers—often achieve high predictive performance but lack transparency, which poses significant challenges in safety-critical domains such as healthcare, cardiology, and clinical decision support systems.

In medical applications, the inability to explain why a model produces a particular prediction may reduce clinical trust, hinder adoption by healthcare professionals, and raise ethical and legal concerns. Consequently, XAI methods aim to bridge the gap between predictive accuracy and interpretability by providing post-hoc explanations that clarify how input features influence model outputs.

Model-Agnostic Explanation Methods

Model-agnostic explanation methods generate explanations without relying on access to the internal structure, parameters, or training process of the underlying model. These approaches treat the predictive model as a black box and derive explanations solely based on the relationship between input features and output predictions. As a result, model-agnostic methods can be applied consistently across a wide range of machine learning algorithms, including traditional classifiers, ensemble models, and deep learning architectures.

Two of the most widely adopted model-agnostic explanation techniques in the literature are LIME and SHAP.

LIME (Local Interpretable Model-Agnostic Explanations)

LIME is a post-hoc explanation technique that approximates the behavior of a complex black-box model in the local neighborhood of a specific instance using an interpretable surrogate model (Ribeiro et al., 2016). The core assumption of LIME is that, although a model may be globally complex and nonlinear, it can be reasonably approximated by a simple and interpretable model—such as a linear regression or decision rule—within a small region around the instance being explained.

To generate explanations, LIME perturbs the input data around the instance of interest and observes how the black-box model's predictions change. Based on these perturbed samples, LIME trains a local surrogate model that mimics the original model's behavior in that vicinity. The

coefficients of the surrogate model are then used to identify the most influential features contributing to the prediction.

LIME primarily provides local explanations, making it particularly suitable for instance-level interpretation, such as explaining why a specific patient was classified as high-risk or low-risk. By highlighting key features responsible for individual predictions, LIME enhances transparency and supports clinical reasoning, although it does not guarantee global consistency across the entire dataset.

SHAP (SHapley Additive exPlanations)

SHAP is a theoretically grounded explanation framework based on cooperative game theory, specifically Shapley values (Lundberg and Lee, 2017). In SHAP, each feature is treated as a "player" in a cooperative game, and the prediction of the model is considered the "payout." The Shapley value quantifies the average marginal contribution of each feature to the prediction across all possible feature combinations.

One of the key strengths of SHAP is that it satisfies important theoretical properties such as local accuracy, consistency, and missingness, ensuring reliable and stable explanations. SHAP provides a unified framework capable of generating both local explanations (for individual predictions) and global explanations (by aggregating feature contributions across the dataset).

In healthcare applications, SHAP is particularly valuable because it offers feature-level insight into model decisions, enabling clinicians to understand which physiological, demographic, or clinical variables drive predictions. This transparency improves interpretability, enhances clinical trust, and facilitates regulatory compliance in medical AI systems.

Model-Specific Explanation Methods

In contrast to model-agnostic techniques, model-specific explanation methods are designed for particular model families, most notably deep neural networks. These approaches leverage internal information such as gradients, activations, or convolutional feature maps to explain how predictions are formed.

Prominent examples include Gradient-weighted Class Activation Mapping (Grad-CAM) and Saliency Maps, which are widely used in medical image analysis tasks such as echocardiography, cardiac MRI, and CT imaging.

Grad-CAM generates heatmaps that highlight the regions of an input image most relevant to a model's prediction by analyzing gradient information from convolutional layers. Saliency Maps, on the other hand, identify important input pixels by computing gradients of the output with respect to the input. These methods primarily provide local, visual

explanations and are particularly effective for interpreting deep learning models in imaging-based cardiology applications.

APPLICATIONS IN HEALTHCARE

Explainable machine learning (XML) techniques have been increasingly adopted across a wide range of healthcare applications, driven by the need for transparency, accountability, and clinician trust in artificial intelligence–based decision support systems. In contrast to purely black-box approaches, explainable models enable clinicians and stakeholders to understand the rationale behind predictions, facilitating safer and more effective clinical adoption.

The following domains represent the most prominent application areas in which explainable machine learning techniques have been actively investigated and applied:

- Cancer risk prediction (Kourou et al., 2015)
- Cardiovascular disease diagnosis (Rajkomar et al., 2018)
- Sleep disorder and respiratory disease analysis
- Hospital readmission prediction

These four areas are the key application areas in the XAI literature where high clinical risk, intensive data usage, and interpretability budget are most prominent.

Beyond core diagnostic and prognostic tasks, explainable AI has been increasingly applied to more specialized and emerging healthcare domains:

- Antimicrobial Prescription Management
- Autism Spectrum Disorder (ASD) Diagnosis
- COVID-19 Diagnosis and Clinical Decision Support

These areas are generally discussed in reviews under the heading *emerging applications of XAI in healthcare*.

Core Clinical Prediction and Diagnosis Applications

Cancer Risk Prediction and Diagnosis

One of the earliest and most prominent applications of explainable machine learning is cancer risk prediction and diagnosis. Machine learning techniques have been widely employed to classify cancer subtypes, predict disease progression, and identify high-risk patients using heterogeneous data

sources, including genomic profiles, medical imaging, and structured clinical records (Kourou et al., 2015).

In this context, explainability plays a pivotal role by allowing clinicians to identify the most influential biomarkers and clinical variables driving model predictions. Feature attribution methods, such as SHAP or LIME, enable the interpretation of complex predictive models by highlighting which genetic markers, imaging features, or laboratory measurements contribute most strongly to risk stratification. This transparency facilitates validation of model outputs against established oncological knowledge and supports clinical confidence in AI-assisted diagnostic systems.

Moreover, explainable models help bridge the gap between data-driven predictions and clinical reasoning, enabling oncologists to integrate machine learning insights into personalized treatment planning and early intervention strategies.

Cardiovascular Disease Diagnosis and Prognosis

Explainable machine learning approaches have been extensively applied to the diagnosis, risk stratification, and prognosis of cardiovascular diseases, which remain one of the leading causes of morbidity and mortality worldwide. Large-scale studies have demonstrated that machine learning models can outperform traditional risk scoring systems in predicting cardiovascular outcomes, detecting arrhythmias, and supporting early diagnosis using heterogeneous clinical data (Rajkomar et al., 2018).

In cardiology, explainability is particularly critical due to the high-risk nature of clinical decisions and the widespread reliance on complex data sources, including electrocardiograms (ECG), echocardiography, cardiac MRI, and longitudinal electronic health records. Explainable AI techniques such as SHAP enable clinicians to interpret the influence of structured clinical variables—such as age, blood pressure, cholesterol levels, and comorbidities—on individual risk predictions. For imaging-based models, Grad-CAM and related visual explanation methods highlight diagnostically relevant regions within cardiac images, facilitating alignment between model outputs and clinician visual reasoning.

By enhancing transparency and trust, explainable models support clinical validation, improve acceptance of AI-assisted diagnostic tools, and contribute to safer integration of machine learning systems into cardiology workflows (Rajkomar et al., 2018).

Sleep Disorder and Respiratory Disease Analysis

Explainable machine learning has increasingly been applied to the analysis of sleep disorders and respiratory diseases, including obstructive sleep apnea (OSA), chronic obstructive pulmonary disease (COPD), and other sleep-related breathing conditions. Machine learning models utilize

multimodal physiological signals—such as oxygen saturation, airflow, respiratory effort, electroencephalography (EEG), and electrocardiography (ECG)—to identify abnormal patterns and support automated screening and diagnosis.

Given the temporal and high-dimensional nature of sleep and respiratory data, explainability is essential for ensuring that predictive models rely on clinically meaningful physiological features rather than spurious correlations. As highlighted by Doshi-Velez and Kim (2017), interpretability is crucial for validating model behavior and assessing whether explanations faithfully represent the underlying predictive mechanisms. Explainable AI techniques enable clinicians to identify which physiological signals, time segments, or frequency components contribute most significantly to classification outcomes, thereby enhancing transparency and interpretability.

Furthermore, explainable approaches facilitate clinical validation by allowing model predictions to be examined in light of established sleep medicine knowledge and diagnostic criteria. This alignment between data-driven insights and clinical expertise is essential for building trust and supporting the adoption of AI-based decision support systems in sleep laboratories and respiratory care settings (Doshi-Velez and Kim, 2017).

Hospital Readmission Prediction

Hospital readmission prediction represents a critical application area for explainable machine learning, with significant implications for healthcare quality, patient outcomes, and cost reduction. Machine learning models have been widely used to identify patients at high risk of unplanned readmission by analyzing demographic characteristics, comorbid conditions, laboratory results, treatment history, and discharge-related factors. However, the opaque nature of many high-performing predictive models limits their direct adoption in clinical practice.

Explainable artificial intelligence addresses this challenge by providing transparent, feature-level insights into model predictions, enabling clinicians to understand why a patient is classified as high risk for readmission. As emphasized by Doshi-Velez and Kim (2017), rigorous interpretability is essential for validating model behavior and ensuring that explanations faithfully reflect the underlying decision-making process. In addition, large-scale studies in clinical machine learning highlight the importance of interpretability for integrating predictive models into real-world healthcare systems and supporting clinician trust (Rajkomar et al., 2018).

By revealing the contribution of key variables—such as length of hospital stay, prior admission history, medication complexity, and discharge planning factors—explainable models support targeted clinical interventions and personalized care strategies. Consequently, explainability enhances both

the usability and clinical relevance of machine learning–based readmission prediction systems.

Emerging and Specialized Application Areas

Beyond core diagnostic and prognostic tasks, explainable artificial intelligence (XAI) has increasingly been applied to more specialized and emerging healthcare domains, particularly in areas where clinical accountability, ethical responsibility, and regulatory compliance are critical.

Antimicrobial Prescription Management

Explainable machine learning models have been developed to predict antimicrobial resistance patterns and to support recommendations for appropriate therapeutic agents. In this context, explainability is essential to ensure clinician trust, facilitate clinical validation, and support responsible antibiotic stewardship. Transparent explanations enable healthcare professionals to understand how patient characteristics, microbiological data, and treatment histories influence model recommendations, thereby reducing the risk of inappropriate antibiotic use and helping combat the global challenge of antimicrobial resistance (Samek et al., 2021; Tjoa and Guan, 2020).

Autism Spectrum Disorder (ASD) Diagnosis

The integration of machine learning and explainable artificial intelligence has emerged as a critical advancement for supporting early screening and clinical decision support in autism spectrum disorder (ASD). Rather than serving as standalone diagnostic tools, explainable models assist clinicians and caregivers by making the contribution of behavioral, neurological, and developmental features transparent, thereby facilitating informed clinical assessment and earlier intervention. This capability supports personalized care strategies while preserving clinician oversight in the diagnostic process.

Transparency is particularly important in the context of ASD due to the complexity and heterogeneity of symptoms, as well as the ethical and developmental implications associated with early screening and diagnostic decision-making. Explainable AI helps ensure that model outputs can be critically evaluated and aligned with established clinical understanding, reinforcing responsible and trustworthy use of machine learning in ASD-related applications (Doshi-Velez and Kim, 2017; Samek et al., 2021).

COVID-19 Diagnosis and Clinical Decision Support

During the COVID-19 pandemic, explainable artificial intelligence techniques were widely employed to improve diagnostic accuracy and transparency in machine learning–based screening tools. These systems leveraged imaging data (e.g., chest X-rays and CT scans), clinical variables,

and laboratory results to support the rapid detection and management of SARS-CoV-2 infection. However, many COVID-19–related AI models were developed under urgent conditions using limited and heterogeneous datasets, resulting in early-stage and rapidly developed models that were sometimes prone to overfitting and limited generalizability.

In this context, explainability played a critical role by enabling clinicians to validate model predictions against medical expertise and established clinical knowledge. By revealing which features or image regions most strongly influenced predictions, XAI techniques increased transparency and supported informed clinical judgment, thereby enhancing confidence in AI-assisted decision-making during a public health emergency (Tjoa and Guan, 2020; Rajkomar, Dean, and Kohane, 2018).

EVALUATION OF EXPLAINABILITY

Evaluating the effectiveness of explainable AI methods remains a major challenge in healthcare applications. Several key dimensions are commonly used to assess explainability:

Fidelity measures how accurately an explanation reflects the true behavior of the underlying predictive model. High-fidelity explanations closely approximate the original model's decision-making process and are essential for reliable interpretation.

Stability assesses the robustness of explanations to small perturbations in input data. Stable explanations produce consistent interpretations for similar inputs, which is particularly important in clinical settings where slight measurement variations are common.

Comprehensibility refers to how easily an explanation can be understood by human users, especially clinicians who may not have technical expertise in machine learning. Explanations should be concise, intuitive, and aligned with clinical reasoning.

Clinical usefulness evaluates the extent to which explanations contribute to improved clinical decision-making, trust, and workflow integration. An explanation may be technically sound but clinically irrelevant if it does not support actionable insights.

A study examining the evaluation of XAI applications in cardiology reported that 37% of studies relied primarily on comparisons with existing literature, 11% involved clinicians as domain experts, and 1% performed no explicit evaluation of explainability. These findings highlight substantial

gaps in standardized evaluation practices and underscore the need for more rigorous, clinician-centered assessment frameworks (Doshi-Velez and Kim, 2017).

ETHICAL, LEGAL, AND SOCIAL IMPLICATIONS

The deployment of AI systems in healthcare raises significant ethical, legal, and social concerns, particularly regarding bias, fairness, accountability, and patient privacy (Char et al., 2018). Bias in training data can lead to unequal performance across demographic groups, potentially exacerbating health disparities. Explainable AI plays a crucial role in identifying and mitigating such biases by revealing how sensitive attributes influence model predictions.

Fairness, non-discrimination, and justice are central ethical principles that must be considered in the design and deployment of medical AI systems. Explainability supports accountability by enabling clinicians, patients, and regulators to scrutinize AI-driven decisions.

From a legal perspective, regulatory frameworks increasingly emphasize transparency and explainability in medical AI systems. Compliance with data protection regulations and emerging AI governance policies requires clear documentation of model behavior and decision logic. The integration of ethical and legal considerations is therefore essential for promoting the safe, reliable, and socially acceptable use of artificial intelligence in healthcare.

FUTURE RESEARCH DIRECTIONS

Future research in explainable AI for healthcare is expected to focus on several key directions:

Hybrid interpretable–deep learning models aim to combine the high predictive performance of deep learning with inherent interpretability, reducing reliance on post-hoc explanations.

Clinician-centered explanation interfaces will prioritize usability and workflow integration, enabling healthcare professionals to interact with and interrogate model predictions effectively.

Standardized explainability metrics are needed to objectively compare different XAI methods and establish benchmarks for clinical validation.

Integration of domain knowledge into machine learning models represents a promising avenue for improving both predictive accuracy and interpretability by embedding expert medical reasoning directly into model structures.

The combination of real-time data from wearable sensors with existing health information systems will enable continuous patient monitoring and personalized interventions, creating new opportunities and challenges for explainable AI in digital health.

RESULTS AND DISCUSSION

Interpretable and explainable machine learning approaches are essential for the safe, ethical, and effective deployment of healthcare decision support systems. By enhancing transparency, accountability, and clinical trust, XAI methods bridge the gap between advanced computational analytics and real-world medical practice. Moving forward, multidisciplinary collaboration, standardized evaluation frameworks, and continuous methodological innovation will be critical to overcoming barriers to adoption and fully realizing the transformative potential of artificial intelligence in healthcare.

REFERENCES

Alexander, G. L. (2008). A descriptive analysis of a nursing home clinical decision support system. Journal of Gerontological Nursing, 34(5), 17-23.

Beam, A. L., and Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317-1318.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). Classification and regression trees. Belmont, CA: Wadsworth.

Char, D. S., Shah, N. H., and Magnus, D. (2018). Implementing machine learning in health care—addressing ethical challenges. New England Journal of Medicine, 378(11), 981-983.

Doshi-Velez, F., and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.

Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., and Dean, J. (2019). A guide to deep learning in healthcare. Nature Medicine, 25(1), 24-29.

European Commission. (2020). Ethics guidelines for trustworthy AI. Brussels: European Union.

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., and Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), 1-42.

Holzinger, A., Langs, G., Denk, H., Zatloukal, K., and Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1312.

Huysmans, J., Dejaeger, K., Mues, C., Vanthienen, J., and Baesens, B. (2011). An empirical evaluation of the comprehensibility of decision table, tree and rule based predictive models. Decision Support Systems, 51(1), 141-154.

Kim, Y., Patel, V. L., and Zhang, J. (2008). A cognitive analysis of decision support in medical informatics. Journal of Biomedical Informatics, 41(4), 519-529.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., and Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. Computational and Structural Biotechnology Journal, 13, 8-17.

Lipton, Z. C. (2018). The mythos of model interpretability. Communications of the ACM, 61(10), 36-43.

Lundberg, S. M., and Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765-4774.

Miotto, R., Li, L., Kidd, B. A., and Dudley, J. T. (2016). Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. Scientific Reports, 6, 26094.

Musen, M. A., Middleton, B., and Greenes, R. A. (2014). Clinical decision-support systems. In Shortliffe, E. H., and Cimino, J. J. (Editors), Biomedical Informatics (pp. 643-674). London: Springer.

Obermeyer, Z., and Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216-1219.

Rajkomar, A., Dean, J., and Kohane, I. (2018). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347-1358.

Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135-1144.

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215.

Samek, W., Wiegand, T., and Müller, K. R. (2019). Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. arXiv preprint arXiv:1708.08296.

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., and Müller, K. R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 109(3), 247-278.

Schnipper, J. L., Gandhi, T. K., Wald, J. S., Grant, R. W., Poon, E. G., Volk, L. A., and Middleton, B. (2008). Effects of an electronic medication reconciliation application and process redesign on potential adverse drug events. Archives of Internal Medicine, 168(8), 771-780.

Shortliffe, E. H., and Cimino, J. J. (2014). Biomedical informatics: Computer applications in health care and biomedicine. London: Springer.

Shortliffe, E. H., and Sepúlveda, M. J. (2018). Clinical decision support in the era of artificial intelligence. JAMA, 320(21), 2199-2200.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., and Kattan, M. W. (2010). Assessing the performance of prediction models. Epidemiology, 21(1), 128-138.

Tjoa, E., and Guan, C. (2020). A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE Transactions on Neural Networks and Learning Systems, 32(11), 4793-4813.

Tonekaboni, S., Joshi, S., McCradden, M. D., and Goldenberg, A. (2019). What clinicians want: Contextualizing explainable machine learning for clinical end use. Proceedings of the Machine Learning for Healthcare Conference, 359-380.

Topol, E. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56.

Varonen, H., Kortteisto, T., and Kaila, M. (2008). What may help or hinder the implementation of computerized decision support systems: A qualitative study of physicians' experiences. Family Practice, 25(3), 162-167.

Wu, J., Roy, J., and Stewart, W. F. (2008). Prediction modeling using EHR data: Challenges, strategies, and a comparison of machine learning approaches. Medical Care, 46(10), S106-S113.

Functional Bioplastics: Next-Generation Materials with Active, Smart, and Barrier Properties

H. Duygu BİLGEN[1]

1- Prof.; Mersin University, Engineering Faculty, Department of Environmental Engineering, hduygubilgen@mersin.edu.tr, ORCID No: 0000-0002-9510-8131,

ABSTRACT

Bioplastics are now playing a more functional and strategic role in food packaging, going beyond being merely an environmentally friendly alternative to petroleum-based plastics. Initially developed as passive packaging materials, bioplastics have evolved into active packaging systems in recent years, gaining the potential to extend food shelf life, slow microbial spoilage, and reduce food quality losses.

The use of natural additives that provide antimicrobial and antioxidant properties is increasing in active bioplastic packaging. Third-generation bio-based raw materials and additives derived from agricultural and industrial waste enable both the improvement of functional properties and the support of a circular economy approach. Furthermore, the utilization of industrial and agricultural waste in the production of active bioplastic packaging holds significant potential for reducing plastic waste, decreasing carbon and water footprints, and developing sustainable food packaging systems.

Keywords – Active Packaging, Antimicrobial Packaging, Bioplastics, Food Preservation, Shelf-Life Extension.

INTRODUCTION

For many years, food packaging has used four main families of petroleum based plastics: PE, PP, PET, and PVC. However, due to their end of life release into the environment and the resulting microplastic pollution, there has been a growing aim to gradually replace food contact packaging with more sustainable options. In this respect, bioplastics offer a competitive solution. Furthermore, today, bioplastics are going beyond being just an alternative to petroleum-based plastics and are taking on a more strategic role in the food industry, offering innovative solutions to sustainability and food safety issues.

Initially developed solely for environmentally friendly packaging, bioplastics have evolved from passive materials into active functional systems within the food sector (Fig.1). The demand for biodegradable bioplastic food packaging has significantly increased with the Single-Use Plastics Directive, which is part of the European Union's Plastics Strategy and came into effect in 2019. However, ensuring environmental sustainability alone is no longer considered sufficient in food packaging, issues such as the growing world population, food security, and reducing food waste are becoming increasingly critical.

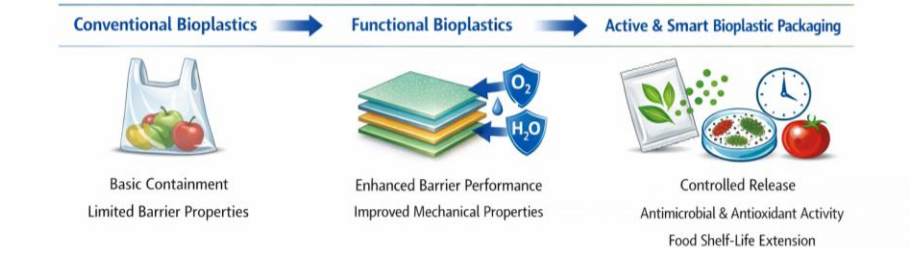


Figure 1. Evolution of bioplastics from passive packaging materials to active and functional food packaging systems.

In this context, the concept of active packaging, which has emerged as a significant innovation in food preservation, encompasses the use of bioplastics not only as passive packaging materials but also as functional systems thanks to improved quality, safety, and integrity properties. Active bioplastic packaging extends shelf life, slows microbial spoilage, and directly contributes to reducing food waste.

Antimicrobial bioplastics

Antimicrobial bioplastics are frequently encountered in active food packaging. These packaging systems have been developed to increase food safety and extend the shelf life of foods. Unlike traditional passive packaging, antimicrobial bioplastics, in addition to their biodegradability, interact with the packaged food or packaging atmosphere, preventing the growth of microorganisms that cause spoilage and foodborne illnesses.

Microbial contamination is commonly encountered in raw meat, fish, dairy products, and ready-to-eat foods. Moisture content and oxygen permeability are important factors in the packaging of these products. Antimicrobial bioplastics, produced by incorporating antimicrobial agents into bioplastic matrices, provide a continuous and localized antimicrobial effect throughout the storage period, offering a significant alternative to the direct addition of traditional preservatives to food. This approach prevents taste changes caused by preservatives and simultaneously ensures long-term preservation of food through the controlled release of antimicrobial agents. This also reduces food waste. Antimicrobial bioplastic active packaging protects vitamins (A, E), essential fatty acids, and bioactive compounds found in food against oxidation (Gómez-Estaca et al., 2014). Controlled antioxidant release also reduces lipid oxidation, inhibits microbial growth and preserves nutrients and their flavor (Fig.2).

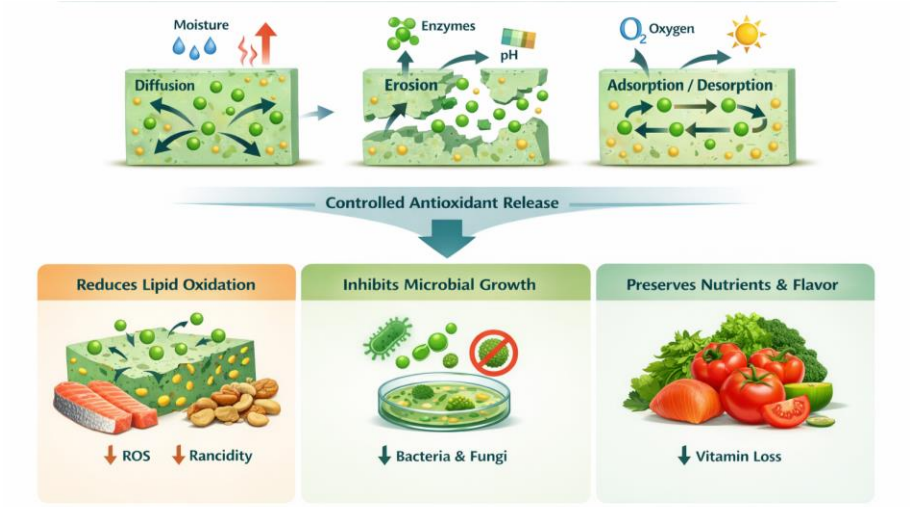


Figure 2. Mechanisms of antioxidant release from bioplastic matrices and their effect on food preservation

Antimicrobial bioplastics, produced by integrating natural antioxidants such as phenolic compounds, flavonoids, tocopherols, and plant extracts into bioplastic matrices, provide a continuous and localized antimicrobial effect throughout the storage period, offering a significant alternative to the direct addition of traditional preservatives to food (Gómez-Estaca et al., 2014).

Many academic studies focus on enhancing starch-based bioplastics, the most preferred type of bioplastic today, with biodegradable functional additives. Some of these studies, which demonstrate strong antimicrobial performance, are as follows: Adding nisin and natamycin to a tapioca starch matrix supplemented with rice bran nanoparticles provided significant antimicrobial activity against Saccharomyces cerevisiae and Listeria innocua (Berti et al., 2021). The use of ferulic acid and cinnamic acid in cassava starch-based films has been shown to effectively inhibit the growth of Escherichia coli and L. innocua (Ordóñez et al., 2021). Amylose-rich hydroxypropyl starch films enriched with pomegranate peel extract exhibited strong antimicrobial activity against Staphylococcus aureus and Salmonella species (Ali et al., 2019). Tapioca starch films containing nisin and potassium sorbate showed high efficacy against L. innocua and the yeast species Zygosaccharomyces bailii, with the simultaneous use of both preservatives yielding the best results (Basch et al., 2013). Tapioca starch films reinforced with chitosan nanoparticles provided stronger inhibition, particularly against Gram-positive bacteria, depending on the concentration (Shapi'i et al., 2020). The use of chitosan, Surinamese cherry leaf extract, and especially natamycin in cassava starch-based films revealed a positive

antifungal effect against Aspergillus flavus and Aspergillus parasiticus (Chakravartula et al., 2020). The addition of Thymus vulgaris essential oil and propolis ethanolic extract to a corn starch matrix produced an antimicrobial effect against E. coli and Listeria monocytogenes (Ardjoum et al., 2023). Films made from red fruit seed starches with tamarind and licorice root showed high bacterial inhibition against S. aureus and E. coli, especially more pronounced against E. coli (Chowdhury et al., 2023). In esterification-modified corn starch films, mandarin peel essential oil and purple corn cob anthocyanins showed high bacteriostatic capacity against E. coli and S. aureus (Chen et al., 2023).

The inhibition of the growth of different microorganisms with a wide variety of additives and the extension of food shelf life with active packaging have been proven both in laboratory studies and in practice. Therefore, antioxidant bioplastic packaging is the preferred choice for many foods found on supermarket shelves today.

Improving the oxygen and moisture barrier properties of bioplastics

The primary factors limiting the widespread use of bioplastics in food packaging are their poor oxygen and water vapor barrier properties compared to conventional plastics.

In recent years, the development of nanofillers, bio-based additives, and multilayer structures has significantly reduced this disadvantage. Reducing oxygen permeability and improving the moisture barrier enhances food preservation by limiting microbial growth and oxidative reactions. These engineering solutions strengthen the applicability of bioplastics on an industrial scale (Fig.3).

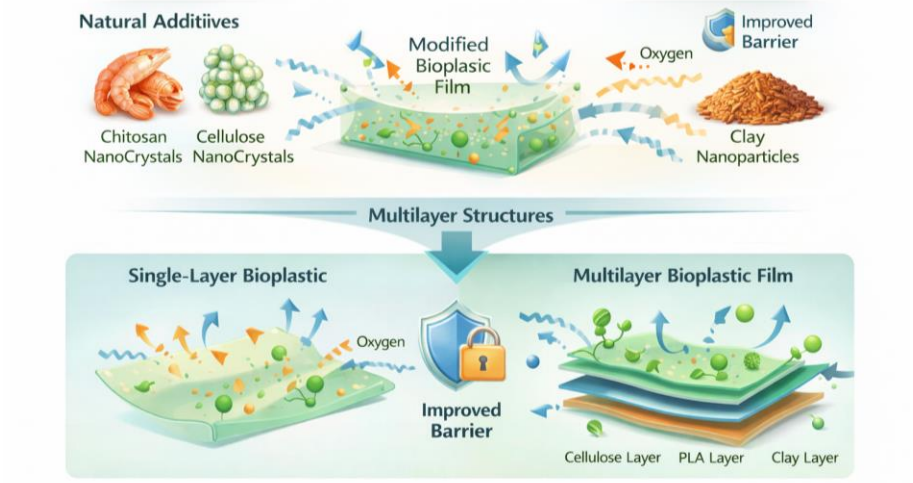


Figure 3. Improvement of oxygen and water wapor berrier properties through biodegredable additives and multilayer structures

Many crosslinking and coupling agents are used to improve the water resistance, as well as the degradation and tensile strength, of bioplastic films intended for packaging.

Shusprajhaa et al. (2025) reported that banana starch, which has a higher amylose content (25-30%) compared to other starches, improves the water vapor barrier properties and thermal stability of bioplastic films, while also positively affecting their film-forming properties, tensile strength, and biodegradability. Banana peels, bananas rejected from export lines, and farm waste, which are agricultural wastes, can be used as resources at this stage.

In their study, Chang et al. (2021) developed a multiphase and multilayer film assembly consisting of thermoplastic starch (TPS) and its maleated counterpart (MTPS) with poly(butylene adipate-co-terephthalate) (PBAT) to develop a barrier film with very high mechanical properties. This multilayer bioplastic film was produced by the researchers using reactive extrusion, compression molding, and dip coating processes. They reported that adding a co-mixture of TPS and PBAT to the innermost layer strengthened the interfacial bond of the multilayer film. As a result, an improvement of 86.8% in the moisture barrier and 74.3% in the oxygen barrier was achieved compared to single-layer TPS and PBAT bioplastic films.

Enhancing functionality with natural additives

The integration of naturally biodegradable materials into bioplastic matrices provides a dual benefit, both in terms of improving the functional properties of bioplastic packaging materials and in terms of environmental impact. It is essential that the additives used to reinforce 100% biodegradable bioplastic films are also environmentally friendly.

Cellulose and cellulose nanocrystals, when incorporated into bioplastic matrices, improve the barrier performance of packaging by reducing oxygen and water vapor permeability, thereby increasing mechanical strength (Trache et al., 2017; Siracusa et al. 2008).

Lignin, due to its natural phenolic structure, forms an effective barrier against UV radiation while also providing antioxidant functionality to packaging materials, making it an important bio-based additive (Marcoaldi et.al., 2025; Yang et al., 2016).

Chitosan, thanks to its positively charged amino groups, interacts with the cell membranes of microorganisms, exhibiting broad-spectrum antimicrobial activity and is widely preferred in active food packaging (Shahbazi & Shavisi, 2022). The combined use of chitosan and nanocomposite additives has been supported by various experimental studies to improve the oxygen and water vapor barrier in bio-based films, thereby

preserving the quality of packaged food (Dıblan, 2025; Al-Hilifi et al., 2023).

Phenolic compounds contribute to extending the shelf life of food in bioplastic packaging systems by delaying lipid oxidation and suppressing microbial growth due to their strong antioxidant and antimicrobial activities. Therefore, they are preferred additives in terms of oxidative stability and microbial control (Figueroa-Lopez et al., 2020).

Additives derived from biological waste, such as fish scales and bones, improve the mechanical strength and gas barrier properties of bioplastic matrices thanks to their calcium, phosphate, and collagen content, while also offering sustainable solutions in line with the circular economy approach (Zhang et al., 2019).

In active packaging applications where essential oils are integrated into bioplastic film systems, these oils provide controlled release. Through these mechanisms, they offer both antimicrobial and antioxidant effects, playing an important role in the production of active bioplastic packaging (Atarés & Chiralt, 2016).

Circular economy based active bioplastic packaging

The circular economy approach aims not only at replacing fossil-based plastics in bioplastic technologies but also at transforming waste into high value-added functional materials. Waste materials defined as third-generation raw materials with no or very limited uses are gaining importance in this context. The necessity of sustainable bioplastic materials makes biological waste from agricultural, industrial, and food processing sources the most suitable resources for the production of active bioplastic packaging.

Fish processing waste, shellfish shells, agricultural residues, lignocellulosic industrial by-products, and bio-based wastes generated from food production processes are emerging as third-generation bioplastic raw materials. These types of waste raw materials include wastes that cannot be used for other purposes or have very limited uses. These wastes can be used both as the main raw material in the production of biopolymer matrices and as active additives, enabling the development of closed-loop systems in accordance with circular economy principles (Zhang et al., 2019; Ahmed et.al, 2025).

Additives derived from industrial and agricultural waste provide bioplastic packaging not only with environmental sustainability but also with active functionality. For example, calcium phosphate and collagen-based components obtained from fish scale and bone waste improve the mechanical strength and gas barrier properties of bioplastic matrices, while phenolic compounds and lignin derivatives isolated from agricultural waste provide antioxidant, UV barrier, and antimicrobial properties to bioplastic films (Yang et al., 2016; Figueroa-Lopez et al., 2020).

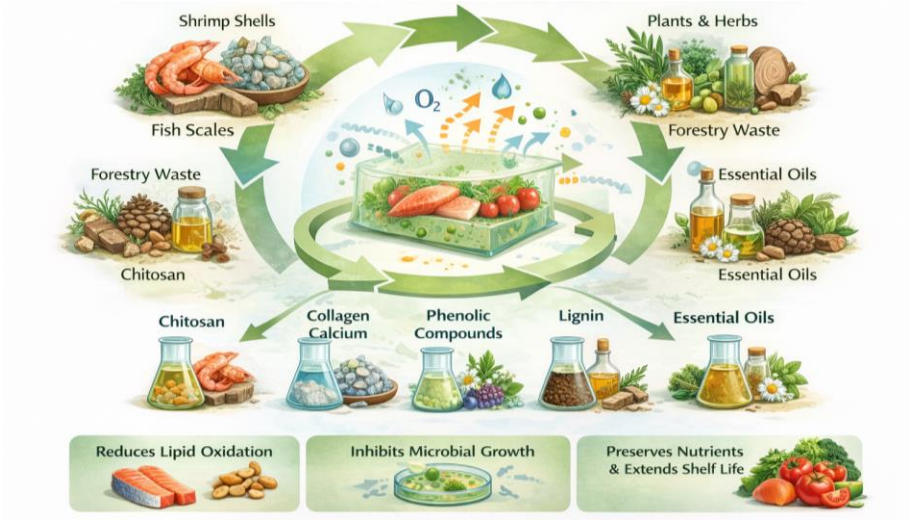


Figure 4. Circular economy based active bioplastic packaging using bio-waste derived additives.

The use of waste based biodegradable additives in active bioplastic packaging offers significant advantages in terms of food safety and shelf life extension. Bio-based active components derived from industrial waste act at the packaging-food interface through controlled release mechanisms, slowing lipid oxidation, suppressing microbial spoilage, and thus contributing to the reduction of food waste (Atarés & Chiralt, 2016).

This approach represents a circular production model where waste is transformed into innovative and functional materials instead of being disposed of, unlike the raw material-product-waste process in traditional linear production models. Therefore, active bioplastic packaging systems based on third-generation raw materials are located at the intersection of circular economy, food safety, and environmental sustainability goals (Geueke et al., 2018).

Today, researchers are also conducting studies on the production of bioplastic packaging materials from a wide variety of agricultural wastes (Ahmed et.al, 2025;Marotta et al.,2025). These studies aim to contribute to the development of biodegradable packaging solutions that reduce carbon footprint and environmental pollution by adopting a circular economy approach and utilizing various plant-based wastes (stems, leaves, fruits and vegetables, seeds and husks, roots and tubers) for the production of bioplastic films.

CONCLUSION

Bioplastics are not only passive alternatives to petroleum-based plastics, but also active, multifunctional materials suitable for sustainable food packaging systems, leading to a decisive advancement in sustainable food packaging technologies. Active bioplastic packaging methods increases food safety and shelf life with antimicrobial and antioxidant properties, and also offers a solution to microplastic contamination, which is associated with traditional petroleum-based plastics and can cause food pollution. The oxygen and moisture barrier properties of food packaging can be improved through bio-based additives, nanocomposites, and multilayer structures, thus increasing the industrial applicability of bioplastic materials.

The use of third-generation raw materials derived from agricultural, industrial, and food processing waste ensures that circular economy principles are properly implemented in the food packaging sector from a sustainability perspective. This perspective simultaneously reduces plastic waste, lowers carbon and water footprints, and contributes to closed-loop material systems. Biodegradable additives derived from waste, such as lignin, phenolic compounds, chitosan, and mineral-rich fishing by-products, emerge as sustainable resources while significantly increasing packaging functionality.

Research and development efforts towards the future of biodegradable food packaging should focus on optimizing the controlled release mechanisms of antimicrobial and antioxidant active substances that will be effective in food preservation, ensuring their long-term stability and compatibility, and scaling up production processes. Similarly, comprehensive life cycle assessments should reveal the environmental and toxicological sustainability of these materials throughout their entire life cycle.

In conclusion, increasing the production and use of active bioplastic packaging based on the circular economy will lead to a reduction in environmental pollution caused by traditional plastics, the prevention of microplastic pollution, and positive developments in food safety and food waste studies. At this stage, interdisciplinary collaborations between materials scientists, food engineers, and environmental engineers, the application of laboratory-scale innovations to industry, and their transformation into commercial solutions will be critical to supporting a sustainable food packaging future.

REFERENCES

Ahmed M.M., Bin Nabi H., Mia S., Ahmad I., Zzaman W. (2025) Valorization of plant-based agro-waste into sustainable food packaging materials: Current approaches and functional applications, Applied Food Research,5(2), 101368, https://doi.org/10.1016/j.afres.2025.101368.

Al-Hilifi, S.A., Al-Ibresam, O.T., Al-Hatim, R.R., Al-Ali, R.M., Maslekar, N., Yao, Y., & Agarwal, V. (2023). Development of chitosan/whey protein hydrolysate composite films for food packaging application. Journal of Composites Science, 7(3), 94. https://doi.org/10.3390/jcs7030094

Ali, A., Chen, Y., Liu, H., Yu, L., Baloch, Z., Khalid, S., Zhu, J., Chen, L. (2019). Starchbased antimicrobial films functionalized by pomegranate peel. International Journal of Biological Macromolecules, 129, 1120–1126. https://doi.org/10.1016/j. ijbiomac.2018.09.068

Ardjoum, N., Chibani, N., Shankar, S., Salmieri, S., Djidjelli, H., Lacroix, M. (2023). Incorporation of Thymus vulgaris essential oil and ethanolic extract of propolis improved the antibacterial, barrier and mechanical properties of corn starch-based films. International Journal of Biological Macromolecules, 224, 578–583. https://doi. org/10.1016/j.ijbiomac.2022.10.146

Atarés, L., Chiralt, A. (2016). Essential oils as additives in biodegradable films and coatings for active food packaging. Trends in Food Science & Technology, 48, 51–62. https://doi.org/10.1016/j.tifs.2015.12.001

Basch, C. Y., Jagus, R. J., Flores, S. K. (2013). Physical and antimicrobial properties of tapioca Starch-HPMC edible films incorporated with Nisin and/or Potassium sorbate. Food and Bioprocess Technology, 6(9), 2419–2428. https://doi.org/10.1007/s11947- 012-0860-3

Berti, S., Jagus, R. J., & Flores, S. K. (2021). Effect of rice bran addition on physical properties of antimicrobial biocomposite films based on Starch. Food and Bioprocess Technology, 14(9), 1700–1711. https://doi.org/10.1007/s11947-021-02669-0

Chang C.C., Trinh B. M., Mekonnen T. H. (2021) Robust multiphase and multilayer starch/polymer (TPS/PBAT) film with simultaneous oxygen/moisture barrier properties, Journal of Colloid and Interface Science,593, 290-303, https://doi.org/10.1016/j.jcis.2021.03.010.

Chakravartula S.N., S., Lourenço, R. V., Balestra, F., Quinta Barbosa Bittante, A. M., Sobral, P. J. do A., Dalla Rosa, M. (2020). Influence of pitanga (Eugenia uniflora L.) leaf extract and/or natamycin on properties of cassava starch/ chitosan active films. Food Packaging and Shelf Life, 24, Article 100498. https://doi. org/10.1016/j.fpsl.2020.100498

Chen, Q., Zhang, P., You, N., Xu, Y., Zhang, Y., Luan, P., Lin, B., Wang, Z., & Zhang, L. (2023). Preparation and characterization of corn starch-based antimicrobial indicator films containing purple corncob anthocyanin and tangerine peel essential oil for monitoring pork freshness. International Journal of Biological Macromolecules, 251, Article 126320. https://doi.org/10.1016/j.ijbiomac.2023.126320

Chowdhury, M. A., Hossain, N., Badrudduza, M. D., Rana, M. M. (2023). Development and characterization of natural sourced bioplastic for food packaging applications. Heliyon, 9(2), Article e13538. https://doi.org/10.1016/j.heliyon.2023.e13538

Dıblan, S. (2025). The effect of humidity and temperature conditions on mechanical and barrier properties in nanocrystalline cellulose-reinforced bio-based films. Food and Health, 11(4), 301-317. https://doi.org/10.3153/FH25026

Figueroa-Lopez, K. J., Vicente, A. A., & Reis, M. A. M. (2020). Incorporation of phenolic compounds into biodegradable films for active food packaging. Food Packaging and Shelf Life, 25, 100515. https://doi.org/10.1016/j.fpsl.2020.100515

Geueke, B., Groh, K., & Muncke, J. (2018). Food packaging in the circular economy: Overview of chemical safety aspects. Journal of Cleaner Production, 193, 491–505. https://doi.org/10.1016/j.jclepro.2018.05.005

Gómez-Estaca J. , López-de-Dicastillo C. , Hernández-Muñoz P. , Catalá R. , Gavara R. (2014), Advances in antioxidant active food packaging, Trends in Food Science & Technology, Volume 35, Issue 1, 42-51. https://doi.org/10.1016/j.tifs.2013.10.008

Marcoaldi C., Acampora V., Venezia V., Prieto C., Grappa R., Silvestri B., Luciani G., Lagaron J. M. (2025), Challenges in lignin integration within biopolymer matrices: Toward stable and effective lignin nanoparticles as additives for sustainable food packaging, Industrial Crops and Products, 224, 120336. https://doi.org/10.1016/j.indcrop.2024.120336.

Marotta, A., Borriello, A., Khan, M. R., Cavella, S., Ambrogi, V., & Torrieri, E. (2025). Boosting Food Packaging Sustainability Through the Valorization of Agri-Food Waste and By-Products. Polymers, 17(6), 735. https://doi.org/10.3390/polym17060735

Ordonez, R., Atar´es, L., Chiralt, A. (2021). Physicochemical and antimicrobial properties of cassava starch films with ferulic or cinnamic acid. Lebensmittel Wissenschaft & Technologie, 144, Article 111242. https://doi.org/10.1016/j. lwt.2021.11124

Shahbazi, Y., & Shavisi, N. (2022). Chitosan-based active food packaging films: Antimicrobial and antioxidant properties. Food Chemistry, 373, 131438.https://doi.org/10.1016/j.foodchem.2021.131438

Shapi'i, R. A., Othman, S. H., Nordin, N., Kadir Basha, R., & Nazli Naim, M. (2020). Antimicrobial properties of starch films incorporated with chitosan nanoparticles: In vitro and in vivo evaluation. Carbohydrate Polymers, 230, Article 115602. https://doi. org/10.1016/j.carbpol.2019.115602

Shuprajhaa T., Kumar Paramasivam S., Pushpavalli S., Anandakumar S., Naik R., (2025) Influence of additives on the development, mechanical, functional characteristics and biodegradability of banana starch-based bio plastic films, International Journal of Biological Macromolecules,295, 139544. https://doi.org/10.1016/j.ijbiomac.2025.139544.

Siracusa, V., Rocculi, P., Romani, S., & Rosa, M. D. (2008). Biodegradable polymers for food packaging: A review. Trends in Food Science & Technology, 19(12), 634-643. https://doi.org/10.1016/j.tifs.2008.07.003

Trache, D., Hussin, M. H., Haafiz, M. K. M., & Thakur, V. K. (2017). Recent progress in cellulose nanocrystals: Sources and production. Nanoscale, 9(5), 1763–1786. https//doi.org/10.1039/ C6NR09494E

Yang, W., Owczarek, J. S., Fortunati, E., Kozanecki, M., Mazzaglia, A., Balestra, G. M., & Kenny, J. M. (2016). Antioxidant and antibacterial lignin nanoparticles

in biodegradable films. ACS Sustainable Chemistry & Engineering, 4(8), 3397–3408. https://doi.org/10.1021/acssuschemeng.6b00251
Zhang, Y., Liu, X., Wang, Y., & Zhao, Y. (2019). Fish scale-based biocomposites for sustainable packaging applications. Journal of Cleaner Production, 228, 1161–1172. https://doi.org/10.1016/ j.jclepro.2019.04.311

Fundamental Principles of Construction Project Management

Abdullah Emre KELEŞ[1]
Hümeyra TOPÇUDAN[2]

1- Assoc. Prof. Dr.; Adana Alparslan Türkeş Science and Technology University. Faculty of Engineering, Department of Civil Engineering. aekeles@atu.edu.tr ORCID No: 0000-0002-5472-2450
2- Adana Alparslan Türkeş Science and Technology University. Faculty of Engineering, Department of Civil Engineering. topcudanhumeyra@gmail.com ORCID No: 0000-0002-1356-226X

ABSTRACT

People need various resources to sustain their lives. The effective and rational use of these resources is of great importance in construction projects. Construction projects have multidisciplinary systematic nature that require the simultaneous contribution of different areas of expertise. Within this system, disciplines such as engineering, architecture, economics, business, and law; to work together and in interaction for the purpose that achieve the project's goals. The construction sector is one of the most intensive application areas the discipline of project management due to its structure, which involves high-cost investments, long project durations, complex production processes, and the simultaneous interaction of numerous stakeholders. In construction projects, the effective, efficient, and economical use of limited resources such as time, cost, quality, labor, materials, and equipment is to emerge as a fundamental requirement for achieving project objectives. Therefore, it to be of great important that decisions made during the project management process are evaluated not only technically but also within the framework of economic and managerial principles.

In this study, the fundamental elements affecting the construction project management process were examined using a literature review method; topics such as integration management, scope management, time management, cost management, quality management, human resources management, communication management, risk management, and procurement management were examined within the scope of project management knowledge areas. The role of these knowledge areas in construction projects was presented. Furthermore, economic and managerial tools that are among the fundamental concepts in project management, such as production, productivity, economically, cost, capital, efficiency, profitability, capacity, and break-even point analysis, and their effects on construction projects have been addressed.

Keywords – Project Management, Project Management, Time Management, Cost Management, Scope Management.

INTRODUCTION

The needs that many people require throughout their lives vary according to their level of education, economic power, culture, the value judgments of the society in which they live, and the level of development. Social scientists have classified these needs. Maslow's Hierarchy of Needs is a psychological theory proposed by Abraham Maslow in 1943, in his paper "A Theory of Human Motivation." It suggests that humans have a hierarchy of needs that must be fulfilled in a specific order, starting from the most basic physiological needs and progressing to higher levels of psychological

and self-fulfilment needs. The hierarchy is often depicted as a pyramid with five levels, arranged from the bottom to the top (Zoysa, et al.,2025). Below, are presented in a classified manner needs within Maslow's Hierarchy of Needs.

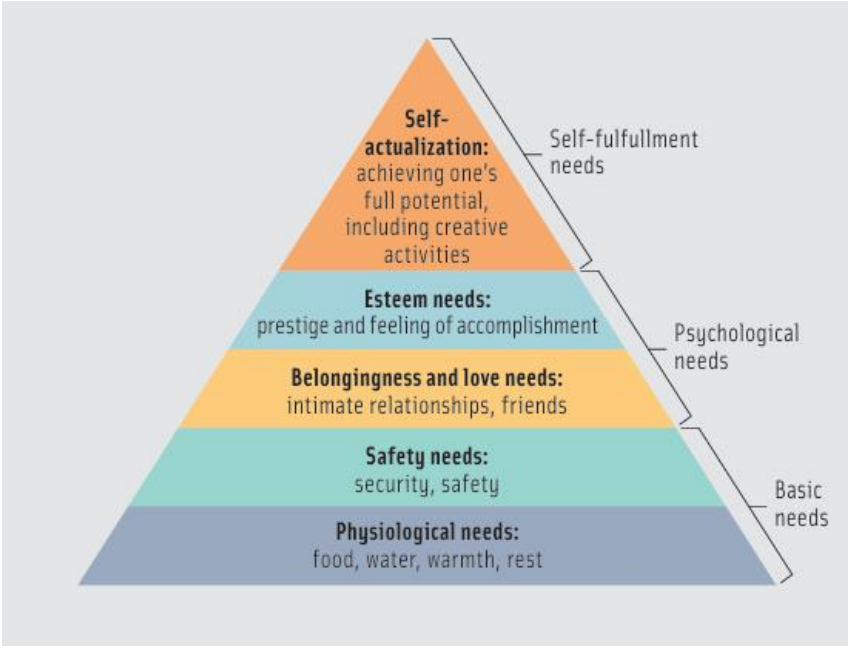


Figure 1. Maslow's Hierarchy Pyramid (Poston, 2009)

In general, higher-level needs do not emerge unless lower-level needs are satisfied. For example, in an individual or a society whose physiological needs have not yet been met, the need for safety or social needs do not arise. In addition, some of the mentioned needs are compulsory, that is, vital in nature, while others are aimed at increasing individuals' levels of entertainment, rest, or satisfaction. Considering the differences in people's levels of needs and their sense of satisfaction, it is possible to say that human needs are virtually unlimited. Every individual uses many resources, such as time, energy, and money, which are personal or social in nature, even while carrying out ordinary activities to meet their daily needs. The totality of things used to meet needs is referred to as resources. In businesses where a wide variety of resource utilization is an issue, it is a necessity to have at a sufficient level understanding and application of the basic principles about of economic behavior at every stage of the production activities carried out. Here is referred to economic behavior as acting appropriately to be moved principles that will enable the use efficient of the owed to resources. Therefore "economic behavior", by accepting a level of consciousness that

must be achieved at the individual level, efforts must be made to implement it. Therefore “economic behavior”, by accepting a level of consciousness that must be achieved at the individual level, it is necessary to show effort in terms of implementation this. In this matter inadequacies, make unavoidable a significant extent wasting resources and deviations from production targets. The need for shelter is also among physiological needs, and the construction industry has been carrying out projects to meet people's shelter needs from the past to the present. Project management in the construction industry is a concept that emerged to prevent resource waste and deviations from production targets during the company’s project process. Considering the size of project dimensions, the being found of geographical differences in the projects to be implemented, and the fact that teams from different countries will come together to work on the projects, the concept of project management become important. Project management is a person or organization within the project organization that plans, executes, and controls the activities of other participants and is in a contractual relationship with the owner or is part of the owner's own organization (Kolaylıoğlu and Avunduk, 2006).

The past of project management dates back to very ancient history. The construction of projects such as the Great Wall of China, the Egyptian pyramids can be given as examples of project management. These is thought to be applied that project management according to the techniques of the period in which they were built. Looking at dates close to the present day, the Gantt chart developed by Henry Gantt in 1917, PERT (Project Evaluation and Review Technique) and CPM (Critical Path Method) methods in 1958, programs such as MS Project and Primavera in subsequent years have been used in project management (Öztürk and Bulak, 2024). The use of methods such as Gantt charts, CPM, PERT enables the project management process to be better managed compared to the previous method, while each new method was developed in to emerging of different needs. Unlike the Gantt chart, these tools enabled the analysis of more complex relationships between the tasks and the uncertainty of each task’s duration. With such precision, the critical path could be better calculated and understood. Still, the use of Gantt chart at the time didn’t decline, quite the opposite, it increased (Wilson, 2003); the tools were used in combination, and the outcome was visualised through the Gantt chart. Still, the Gantt chart remained an important tool both in planning and controlling project schedules; so much so that popular project management software, such as MS Project and Primavera, uses the Gantt chart as the main visual representation of projects (Geraldi, & Lechler, 2012). The development and change in programs such as MS Project and Primavera will be provided in line with the needs that must be met in the project management process.

That important to in the construction industry is managed time, cost and resources correctly and is realized of the project management process by

ensuring quality. To project be completed a successful form using the project management methods of the factors that influence to project management and success are be evaluated and directed. In this study aims to present by literature review of the fundamental factors that influence project management and its success.

LITERATURE REVIEW

To meet social needs and finding an solution to a problem or aimed at taking advantage of a specific opportunity; a team, within a specified timeframe and limited budget using certain sources, while prioritizing customer satisfaction and quality, provided that potential risks are managed, the process of initiating, executing, controlling, and completing a unique the plan in accordance with defined objectives and goals within a specified scope is called a "project" (Kazan, H., 2010). In short, it is the entirety of planned activities implemented temporarily to meet a specific need, solve a problem, or to establish a new product, service, or structure. Project management is a systematic process that encompasses planning, execution, monitoring, control, and completion within the constraints of time, cost, and quality to ensure a project achieves its defined objectives. In other words, it is the application of knowledge, skills, tools, and techniques to achieve project objectives (Project Management Institute [PMI], 2021). Project management involves not only planning, but also monitored, directed, and evaluated activities.

Basic Concepts in Project Management

Production: It can be defined as making a change to on a physical asset that will increases its value. It is the conversion of taken as input resources into output through various by undergoing processes.

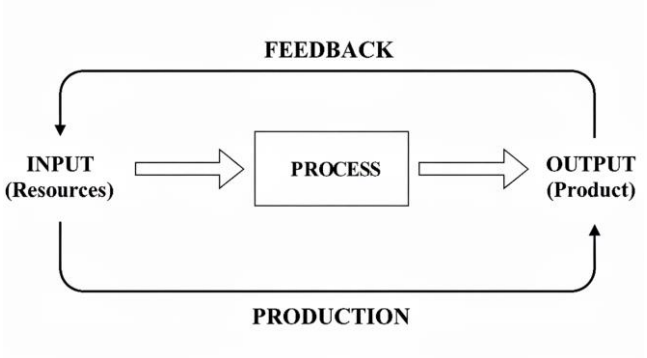


Figure 2. Production Process Diagram

Production factors, in other words resources, are examined in five groups: Men, Machines, Money, Materials, Management-5M.

Productivity: It refers to the level of production achieved by utilizing production factors in the most appropriate manner.

Efficiency: In its most basic sense, it is the relationship between the resources used for production and the goods and services produced. It is the ratio of the product (output), obtained as a result of production activity to the physical quantity of the production factors used (input).

Economically: It is the ratio of the monetary value of the used inputs to the monetary value of the outputs.

Profitability: Profitability is the ratio of operating profit earned in a given period to the capital employed in the business during that period.

Capital: It has many meanings used in economics, accounting, and business. To define it in the field of business, capital refers to all the assets owned by the business.

Finance: It is the task of providing funds to the business as if they were its own resources and foreign source.

Cost: Cost is defined as the monetary amount paid for the inputs used to produce a specific product. Especially in economies where free market conditions prevail, the price of a product is not determined by the manufacturer's preference but according to their prices of competing producers. Since price increases are not allowed, the healthiest way to increase the profit obtained from the sale of a product is to reduce the

product's cost. In order to reduce product costs, firstly it is required to know the expenses (costs) that constitute the cost. The process of determining expense items is called cost analysis. For every business, being able to compete by pricing its product realistically and accurately determining the profit or loss generated from its commercial activities within a given period is directly related to the presence of a realistic cost accounting system within the enterprise.

Break-Even (Profit) Point Analysis: Break-even point (BEP) analyses are one of the analyses used to explain the relationships between production quantity, cost, and profit. A break-even point defines when an investment will generate a positive return and can be determined graphically or with simple mathematics (Kampf, 2016). The BEP is which the business makes neither a profit nor a loss in other words is the level of production where revenues equal expenses. What amount of profit is expected at which level of production, at identifying the discount that can be offered when setting the product sales price, the BBN is determined important.

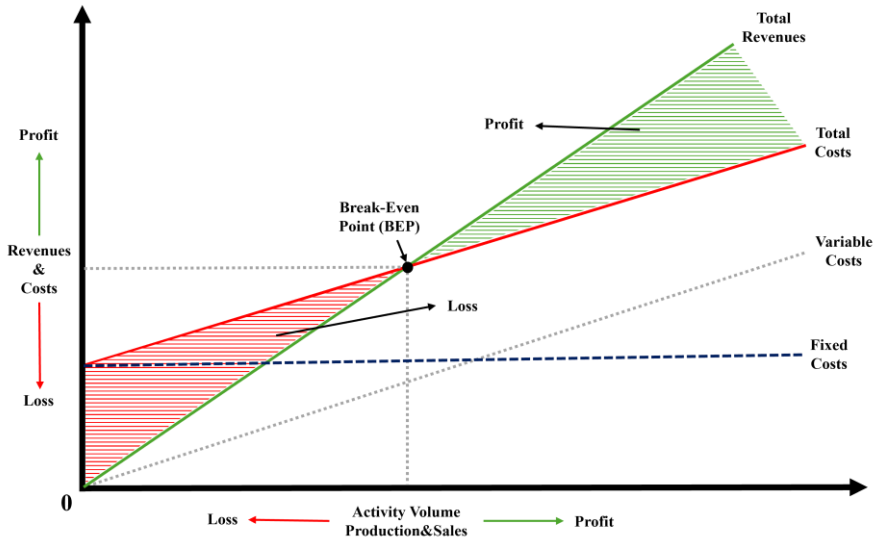


Figure 3. Break-Even Point Analysis Chart

Capacity: Capacity represents production power. In other words, capacity is the highest output amount that a business can achieve within a period of time.

Law of Diminishing Yield: İncreasing productivity, it is known that is possible by combining production factors in the most appropriate way. If one production factor is continuously increased while the others are kept constant, the level of production first increases in parallel with the increased

production factor and then begins to decline after reaching a certain level. This phenomenon is called the “law of diminishing yield.”

Project management consists of five basic stages: planning, organization, orientation (decree-command), coordination and control.

Planning: The activities that make up the project (the activities, or in other words, the work packages) and the relationships between them are identified and duration, resource, and cost analyses related to these project activities are done carried out at this stage.

Organization: The gathering of personnel, equipment, location, and other requirements, and the defining and structuring of relationships, that will allow for the implementation and completion of works to achieve the objectives outlined in the plan is referred organization.

Orientation (Decree-Command): The operation of the organizational phase is ensured by the proper establishment of the command-control system. The sound establishment and operation of the command-control chain important factors affecting is production efficiency. Consequently, the communication system should be required set up to avert misunderstandings and facilitate rapid communication.

Coordination: Coordination is the arrangement of relationships between existing units and ongoing activities within an organization, in a manner that is geared towards a common goal, avoids conflict, and ensures mutual complementarity.

Control (Check): Control is the completion of activities to determine whether the work performed is in accordance with the plan and schedule, to identify any irregularities, if any, to determine the solutions and methods, and to communicate them to the relevant and authorized units or persons for the necessary corrections to be made. Successful project management is achieved by correctly implementing these stages.

As in many sectors, there is a competitive environment in the construction industry. In this competitive environment, protecting businesses for assets and ensuring continuity are one of the important goals. Increasingly competitive conditions, limited resources, and the complexity of work processes create challenges in implementation. Project management is important for the effective management of the project implementation process. Project management is the process of planning, organizing, and controlling of the resources such as time, cost, materials, and labor to achieve a specific goal. In the construction industry in particular, project management is playing a critical role in completing the project within the planned time frame, within the specified budget, and at the expected quality

standards. The project management triangle is one of the most important concepts affecting project success. Every project has time, budget, and scope constraints. The cost constraint is a project's financial constraints. In other words, it is the amount monetary you want to spend on the project. Scope refers to the tasks required to achieve the project's objectives. Time refers to the time required to complete the project. The project management triangle ensures that everyone involved in the project remains committed to the project objectives and that everyone acts in line with these objectives (Özer, et all, 2020).

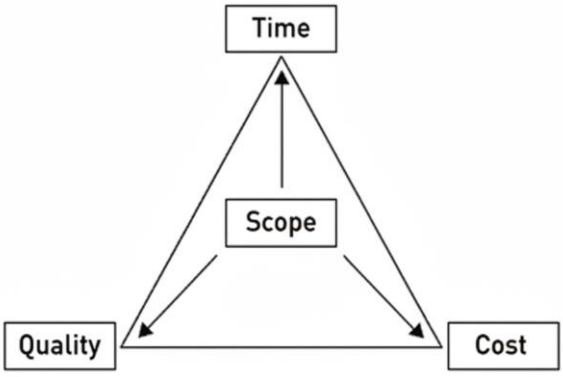


Figure 4. Project Management Triangle

The most important reason for the emergence of the management to method known as project management in construction is that projects have a specific duration and be completed by the end of that period. The beginning of this process is determining customer expectations and requirements. The middle of the process involves designing and meeting requirements. The result is the physical construction of the building. The project has a beginning and an end. As in other sectors, there is no repetition or continuation (Hellard, 1993). The fundamentals of project management, situations that may be encountered in project management in the construction sector, and the planning techniques and tools to be used in project management are among at the top of the topics that must be examined for an effective project management process (Yüksel and Aytekin, 2020). The Project Management Body of Knowledge (PMBOK), a resource published by the Project Management Institute (PMI), a leader in project management, classified project management knowledge areas as follows (Yurdal, 2022).

Project Management Knowledge Areas

Scope Management: It is defined project tasks and determining their boundaries. Project Scope Management encompasses the processes that ensure all tasks necessary for the successful completion of the project are included within the project scope. Project Scope Management is

fundamentally related of definition and auditing what will be included in the project and what is excluded (Kazan, 2010).

Time Management: Project Time Management is a method of consciously controlling the time spent on specific activities with the aim of increasing production and productivity. Time management can be supported by various skills, tools, and techniques used to complete specific tasks and projects. These skills, tools, and techniques; includes planning, allocation, goal setting, delegation, time analysis, observation, arrangement, scheduling, prioritization, and similar activities (Kazan, 2010). It is the process of planning, sequencing, and determining the duration of project activities. In short, it defines the processes that ensure the project is completed on time.

Cost Management: It involves establishing, monitoring, and controlling the project budget. To the process of completing the project within the specified time frame at the desired quality level and at the minimum cost referred to project cost management. By the effective project cost management for project success requires of reduced project costs, improved quality, and completes the project within the predetermined time frame and targeted budget limits (PMI, 2008). To ensure these matters estimating project costs is an important issue in cost management is met.

Quality Management: Quality is the sum of characteristics that enable a product or service to meet specified or potential requirements (ISO 8402). A quality management process compatible with the ISO 9000 series, to the project ensures the establishment of criteria that meet quality requirements, both in terms of managerial and the characteristics of the product obtained. Quality management to secure that the product or service complies with specified standards. Quality management processes in projects; by the company implementing the project, include all activities carried out to determine quality policies, objectives, and responsibilities in order to achieve the necessary quality standards for the project (Kazan, H., 2010).

Human Resources Management: Construction projects are organizations in where many employees work as they are labor-intensive projects. Due to its constantly changing structure in the process of construction projects, the teams and individuals involved are converting in this line (Yurdal, 2022). For the efficient use of human power, equipment, and materials the importance of human resources management is clearly seen.

Communication Management: It ensures effective information flow among project stakeholders. In the execution of projects many teams are on duty. Due to the constantly changing nature of construction projects, planning communication between teams should be one of the primary

objectives. By managing the information flow between stakeholders correctly within the framework of the plans made, it becomes one of the major factors in increasing project efficiency (Yurdal, 2022).

Risk Management: Potential risks must be identified, analyzed, and controlled. It is accomplishing planning of solutions by drawing a road map on how to resolve risks that may resultant. Finally, it plays a major role in the successful completion of the project by auditing the roadmap drawn under the heading of risk management (Yurdal, 2022). Risks identified and searching for solutions is often possible head off before they arise (Memioğlu, 2020).

Supply Management: It is a strategic management process that plans, organizes, and controls the timely, cost-effective, and high-quality level procurement of goods, services, or resources required by a project or business from external sources. The identification of needs in a project or production involves planning which products will be purchased externally and which will be produced internally, followed by research for supplier selection. Supplier selection is crucial for projects. Constantly changing suppliers as it poses a risk for projects also to damage the sense of belonging that is difficult to establish between the supplier and the project (Yurdal, 2022).

Stakeholder Management: When looking at the parties involved in a construction project, there are many individuals, companies, and organizations such as the client, main contractor, subcontractors, suppliers, designers, etc. The parties that are involved or may become involved are referred to as stakeholders (Yurdal, 2022). Stakeholder management is the process of systematically establishing relationships with all parties involved in a construction project or directly or indirectly affected by the project. This systematic process aims to achieve project objectives by balancing the needs and expectations of stakeholders and reducing project risks. Effectively managing communication among stakeholders facilitates the implementation of environmental, social, and economic sustainability principles throughout the project, helping to create long-term value. (Keleş et all, 2025).

Integration Management: It comprised all topics related to managing of tasks that could impact the entire project. There are also different definitions in integration management. In the context of project management, integration deciding where to focus resources and efforts within a specified timeframe, anticipating potential issues, addressing problems before they become critical, and ensure to coordination among activities to ensure project success (Kazan, 2010). Integration management consists of activities that include processes that define, combine, and coordinate various

applications included within the project in order to successfully manage project expectations, and actions that integrate these processes (Kurşunoğlu, 2017).

Project management knowledge areas (integration, scope, time, cost, quality, human resources, communication, risk, procurement, stakeholders) even if it seems be processes that need to be worked on separately, but in practice, the concepts of scope, cost, communication, risk, human resources, time, procurement, stakeholders, and quality are intertwined and each affected the other. Therefore, integration management plays an important role in project management.

CONCLUSIONS

As in all sectors around the world, businesses in the construction industry, in an era of rapid changes, by renewing constantly their products, processes and systems are protected their companies and adapted to this age. Otherwise, in increasingly competitive conditions, businesses face consequences such as project failures and financial losses. There are many factors that affect success and productivity in a project, such as cost, time, and quality. To be affected by these from factors as little as possible, since the emergence of the concept of project management, the aim has been to ensure that projects are managed correctly and efficiently. Therefore, the need for project management has increased over time, and project management knowledge areas (integration, scope, time, cost, quality, human resources, communication, risk, procurement, stakeholders) have become more important today. Conscious management of these knowledge areas will also positively affect the management of projects in the construction sector.

In construction project management, it is important to track activities that have a direct impact on project productivity, especially those on the critical path, on a daily and weekly basis. Since the financial provisions of these progressions are significant amount, small changes in productivity it means to be reflected as large amounts, making the factors affecting productivity of great importance in project management. It is clear that project success does not come solely from keeping project costs to a minimum, nor are projects completed in the shortest time considered more successful. Project success is not solely dependent on the concept of cost and process. Depending on the project to be implemented, it will be possible to ensure success for businesses by correctly planned and managed quality, time, and cost constraints. Therefore, an integrated system of project

management that includes concepts such as quality, cost, time, risk, supply, and communication will facilitate the management process for businesses. By this study, concepts affecting project management, project management knowledge areas (integration, scope, time, cost, quality, human resources, communication, risk, procurement, stakeholders), and the effects on project success of project management in the construction sector, presented based on a literature review of previous studies and articles.

The results of the study that treating project management knowledge areas independently limits project performance and in contrast, it has been observed that managing these information areas with an integrated system approach significantly increases project success. In this context, it was concluded that integrated management practices conducted within the time-cost-quality triangle are decisive in achieving increased efficiency, cost control, and sustainable quality targets in construction projects. The study is considered to being evaluated besides to provide a theoretical framework for project management practices in the construction industry, serving guidance nature for practitioners and academic studies. It is intended to contribute to future studies related to project management.

REFERENCES

de Zoysa, M.S.N., Coorey, S.B.A., Thayaparan, M., de Silva, L., (2025). A human centred approach to post disaster resettlement housing: Case sudy on walk-up apartments in Sri Lanka. In: Waidyasekara, K.G.A.S., Jayasena, H.S., Wimalaratne, P.L.I. and Tennakoon, G.A. (eds). Proceedings of the 13th World Construction Symposium, 15-16 August. https://doi.org/10.31705/WCS.2025.3

Geraldi, J., Lechler, T., (2012). Gantt Charts revisited: A critical analysis of its roots and implications to the management of projects today. International Journal of Managing Projects in Business, 5(4), 578-594. https://doi.org/10.1108/17538371211268889

Hellard, R.B., (1993). Total Quality in Construction Projects Achieving Profitability with Customer Satisfaction. Thomas Telford House, London.

Kampf, R., Majerčák, P., Švagr, P., (2016) Application of Break-Even Point Analysis, Primjena Break-Even Point Analize, scientific journal for the sea and maritime affairs, Vol. 63 No. 3 Special Issue. https://doi.org/10.17818/NM/2016/SI9

Kazan, H., (2010). Project Management Course Notes, Istanbul University Faculty of Open and Distance Education, Faculty of Business Administration.

Keleş, A.E., Gülek, G.G. Górecki, J., (2025). Use of Project Management Knowledge Areas in Civil Infrastructure Projects: Implications for

Sustainability Assessment and Risk Analysis. Sustainability, 17, 9129. https://doi.org/10.3390/su17209129.
Kolaylıoğlu Ö., Avunduk H., (2006). Project Management and the Project Manager in the Construction Industry. Dokuz Eylul University, Institute of Social Sciences, Department of Civil Engineering, Department of Business Administration, Master's Thesis.
Kömürlü, R., Toltar, L., (2018). Project Management in Construction: Its Impact on Project Success. Journal of Architecture and Life (Mimarlık ve Yaşam Dergisi),3(2),2018,249–258. https://dergipark.org.tr/en/download/article-file/607595
Kurşunoğlu, Z., (2017). A Study on Success Criteria in Project Management. Gebze Technical University, Institute of Social Sciences, Department of Business Administration, Master's Thesis, Gebze.
Memioğlu, C., (2020). Risk Management in Contracts within the Scope of the Tendering Process in the Construction Industry. Master's Thesis, Kocaeli University, Institute of Natural Sciences, Kocaeli, 651058.
Özer, G., Gök, M.Ş., Merter, A.K., (2020). Project Cost Estimation: Alternative Approaches. Journal of Yaşar University, 2020, 15 (Special Issue), 272–279. https://dergipark.org.tr/tr/download/article-file/1049959
Öztürk, K., Bulak, M.E., (2024). Project Management and BIM (Building Information Modeling). Uskudar University, Institute of Natural Sciences, Department of Industrial Engineering, Engineering Management Master's Program, Master's Thesis, Istanbul.
PMI. 2008. "A guide to the project management body of knowledge (PMBOK guide)." Project Management Institute Publication., USA.
PMI. 2021. "A Guide to the Project Management Body of Knowledge (PMBOK® Guide)." Philadelphia, PA, USA: Project Management Institute.
Poston, R.A., (2009) An Exercise in Personal Exploration: Maslow's Hierarchy of Needs, Association of Surgical Technologists, August. https://www.ast.org/pdf/308.pdf
Wilson, J., (2003). Gantt charts: A centenary appreciation. European Journal of OperationalResearch, 149(2), 430-437.
Yurdal, T., (2022). Supply Chain Management and Supplier Selection Approaches in the Construction Industry. Master's Thesis, Kocaeli University, Institute of Natural Sciences, Department of Architecture.
Yüksel, H., Aytekin, O., (2020). Productivity Analyses in Construction Project Management. Eskişehir Osmangazi University, Institute of Natural Sciences, Department of Civil Engineering, Division of Building Science, Master's Thesis.

Numerical Analysis of Different Bullet Impact Tests on Aluminum Plates of Varying Thicknesses

Serkan ERDEM [1]
Mehmet Erbil ÖZCAN [2]

1- Assoc. Prof.; Department of Mechanical Engineering, Faculty of Engineering, Firat University serdem@firat.edu.tr ORCID No: 0000-0002-2504-0135
2- Assoc. Prof.; Firat University, Faculty of Engineering, Department of Mechanical Engineering meozcan@firat.edu.tr ORCID No: 0000-0003-1641-8279

ABSTRACT

This study numerically investigates the ballistic effects of different types of projectiles on aluminum alloy plates of varying thicknesses. Target plates made from Aluminum 7075 and Aluminum 2024 alloys, commonly used in the aerospace and defense industries, with thicknesses of 1 mm, 3 mm, and 5 mm were selected. The target plates were modeled with dimensions of 300 mm × 500 mm, and the impact of G1 and G7 projectiles on these plates at constant speed was analyzed under two separate simulation scenarios. Numerical analyses were performed using ANSYS software based on the finite element method. As a result of the simulations, the maximum and minimum stress values generated in the plates for both projectile types were determined, and damage formation and stress distributions were evaluated in detail. The results revealed the effects of projectile type, plate thickness, and alloy type on ballistic performance. Significant changes in maximum stress values and damage levels were observed, particularly with increasing plate thickness. Furthermore, it was determined that differences in the mechanical properties of 7075 and 2024 alloys were reflected in the ballistic behavior. The findings of this study were discussed in comparison with similar studies in the literature, and it was assessed that the results could contribute to the selection of appropriate materials and thicknesses in defense and aerospace applications.

Keywords – Ballistic Analysis, Aluminum Alloys, Finite Element Method, Bullet Impact, Plate Thickness.

INTRODUCTION

In the military, clothing and armor materials are crucial not only for providing protection suitable for seasonal conditions, but also for protecting the user from external factors such as bullets and shrapnel during combat. Therefore, testing and analyzing specialized armor materials for both civilian use in special missions and military law enforcement has become increasingly important in recent times.

In this context, it is necessary to evaluate criteria such as ergonomics, weight, flexibility, and user comfort during long-term use, as well as determining the ballistic resistance of armor materials. Furthermore, examining the behavior of these materials under different temperature, humidity, and mechanical loading conditions is crucial for obtaining reliable results that represent real-world mission conditions. Experimental tests and numerical analyses conducted in accordance with national and international standards contribute to objectively demonstrating the effectiveness of armor

systems and provide a scientific foundation for the design processes of new generation protective equipment.

Aluminum and aluminum alloys are frequently used in lightweight armor and armor support structures due to their advantages such as high specific strength (strength/weight ratio), good formability, and corrosion resistance. However, mechanical limitations of pure aluminum, such as its relatively low hardness and yield strength, limit its use as the sole primary ballistic barrier material; therefore, alloyed aluminum series have begun to be used as material groups exhibiting desired properties for different purposes.

Types and characteristics of bullets and plates used

G1 Type Bullet Specifications

G1-type bullet is generally designed for hunting, military, and short-to-medium range shooting. This type of ammunition has higher air resistance, faster velocity loss, and is more susceptible to wind deflection. Therefore, the ballistic coefficient (BC) value is highly velocity-dependent**;** a single G1 BC value can yield inaccurate results across a wide range of velocities.

G7 Type Bullet Specifications

The G7 model is based on an aerodynamically advanced reference bullet defined in line with modern ballistic requirements. The G7 bullet type has a long, slender, and pointed nose, a constricted tail**,** and a geometric structure with high aerodynamic efficiency. This geometry is very similar to bullet currently used, especially for sniper ammunition, long-range shots, and military and tactical applications. This bullet type has a lower drag coefficient. Velocity is maintained for longer periods, and wind deflection is reduced. It also exhibits more stable flight during the supersonic-transonic transition.

Specifications of Aluminum 7075 and 2024 Series Plates

2024 aluminum alloy is considered one of the most common and well-known aluminum alloys in the aerospace industry. Particularly due to its high strength-to-density ratio, it stands out as an ideal engineering material in applications where lightness and strength are critical. This alloy can be safely used in aerospace structures thanks to its high static strength and superior fatigue resistance.

The 2024 aluminum alloy, whose properties are given in Table 1 and Table 2, belongs to the 2000 series of aluminum alloys, and its main alloying element is copper. The high copper content gives the alloy superior mechanical properties, but also results in relatively low corrosion resistance. Therefore, in

practical applications, the 2024 alloy is usually protected with a coating layer of high-purity aluminum or subjected to surface treatments such as anodic oxidation.

This alloy offers high strength and fatigue resistance, as well as good machinability and cutting properties. These characteristics facilitate the production of aerospace components with complex geometries. However, the corrosion resistance and weldability of 2024 aluminum alloy are limited. It is particularly susceptible to thermal cracking and stress corrosion cracking during welding processes. Therefore, specialized methods such as resistance spot welding, seam welding, or flash welding are generally preferred for joining the alloy by welding.

2024 aluminum alloy is defined under international standards such as AMS 4035, QQ-A-250/4 and GB/T 3190-2017, and is widely used in high-performance structural aerospace applications, primarily in aircraft fuselage components.

7075 aluminum alloy belongs to the 7000 series of aluminum alloys and is one of the most important engineering materials, distinguished by its high strength. The main alloying elements of this alloy are zinc (Zn), magnesium (Mg), and copper (Cu). The high zinc content, in particular, gives 7075 alloy mechanical strength close to that of steel, making it widely used in high-performance applications such as the aerospace and defense industries.

7075 aluminum alloy is ideally suited for applications where weight is a critical design parameter, thanks to its high specific strength. Being a heat-treatable alloy, it allows for optimization of its mechanical properties at different temper states (e.g., T6 and T73). The highest tensile and yield strengths are achieved at the T6 temper, while the T73 temper provides greater resistance to stress corrosion cracking.

In addition to its mechanical properties, 7075 alloy exhibits good fatigue resistance. However, its formability and ductility are relatively limited as a result of its high strength. A risk of cracking can occur, particularly during cold forming processes. On the other hand, its machinability is quite good, making it suitable for the production of high-precision parts using machining methods.

One of the most significant disadvantages of 7075 aluminum alloy is its limited corrosion resistance. Due to its susceptibility to stress corrosion cracking, surface protection methods are of great importance. In this context, anode oxidation, coating, and appropriate heat treatment strategies are

frequently applied. Furthermore, the alloy has low weldability, and conventional fusion welding methods are generally not recommended.

Aluminum alloy 7075, whose chemical composition is given in Table 1 and mechanical properties in Table 2, is defined under international standards such as ASTM B209, AMS 4045 and EN 573, and has a wide range of applications, primarily in aircraft structural elements, landing gear components, military equipment and machine parts requiring high strength.

Table 1Chemical composition of Al7075 and Al2024 materials.

Material	Al	Cu	Mg	Si	Zn	Fe	Mn	Cr	Ti
Al2024	90.9	4.9	1.8	0.5	0.25	0.5	0.9	0.1	0.15
Al7075	90.02	1.2	2.1	0.4	5.1	0.5	0.3	0.18	0.2

Table 2Chemical composition of Al7075 and Al2024 materials.

Material	*Yield Strength [MPa]*	*Tensile Strength [MPa]*	*Elasticity Modulus [GPa]*	*Poisson Ratio*	*Density [kg/m³]*
Al 2024-T3	324	469	73	0.33	2780
Al 7075-T6	503	572	71.7	0.33	2810
Structural Steel	250	460	200	0.30	7850

Among the alloys frequently cited in the literature for armor applications are high-strength or armor-developed series such as AA5083, AA7075, AA2139, and AA7039 . Studies have shown that alloy selection—particularly cutting/melting process history, heat treatment status, and microhardness distribution—is decisive in ballistic performance. For example, AA5083 exhibits good integration and impact performance compared to other aluminum alloys of similar thickness, while AA6061 (of the same thickness) has been reported to have relatively weaker performance (Akbari et al., 2023).

Aluminum sheet bases with ceramic front surface Hybrid solutions, where ceramics are combined with panels or UHMWPE/composite supports, have been widely studied in the literature. These approaches aim for the ceramic to absorb most of the kinetic energy on the front surface through a bullet-breaking mechanism, while the aluminum/back layer dissipates the particles and impact, providing structural support. This study has shown that ceramic + aluminum/composite back layer combinations offer better performance than single-metal combinations; however, parameters such as layer order, bonding quality, and the rigidity of the back support material play a critical role (TEOMAN et al., 2023) .

In their study, Alwan et al. analyzed the ballistic impact on a three-layer circular sandwich plate using different types of projectiles and 6061-T651

aluminum alloy, ZK61m magnesium alloy, and 1100-H12 aluminum alloys. The simulation focused on residual velocity and energy absorption values for each material variation and validated the configurations with previous experiments using the Johnson-Cook material model in ABAQUS/Explicit. In addition to affecting energy absorption, stress and strain in axial, shear, circumferential, and radial deformations also showed how strong each plate arrangement was against ballistic loads. The aim of the research was to determine the strongest plate design (Alwan et al., 2022) . In a study with target plate thicknesses ranging from 0.5 to 3 mm, Gupta et al. [8] examined aluminum target plates struck with blunt, pointed, and hemispherical steel projectiles at different velocities (using a pneumatic gun). At thinner plate thicknesses, pointed bullet were more effective than blunt bullets, but hemispherical projectiles caused the most deformation. Hemispherical bullets have the highest ballistic limiting velocity. Experimental results and ABAQUS finite element analysis results showed good correlation (Gupta et al., 2007) .

In recent years, aluminum foams and honeycomb-shaped sandwich panels have attracted attention due to their superior energy absorption and impact damping properties compared to other profiles. Variables such as the density, cell geometry, and core thickness of the aluminum foam directly affect its energy absorption capacity and damage propagation. In ceramic and aluminum foam-supported systems, the density parameter can increase the toughness and ballistic resistance of the backing layer. Experimental and numerical studies in this field show that foam-based options perform particularly well against multi-label and low- to medium-caliber threats (Brekken et al., 2022) .

Such as SiC/Al and B4C/Al reinforced and functionally gradient aluminum compounds, are produced using powder metallurgy and particle reinforcement methods. The resulting materials offer a hard exterior and ductile interior, providing good resistance to penetration. Many studies have reported that the distribution of SiC or B4C particles in the matrix and the layer thicknesses significantly affect ballistic performance, and that heat treatment and compression steps also shape the ultimate strength (Übeyli et al., 2014) .

Dynamic impact analyses performed using the finite element (FE) method are a cost-effective tool for parameter selection and optimization compared to real experiments. In experiments conducted using this method, the accuracy and reliability of the models increase if physical phenomena such as the high strain rate behavior of the material, fracture criteria, particle-matrix interactions, and multiple damage accumulation are correctly modeled. Many studies in the literature have validated FE models with experimental tests and shown that modeling approaches can reduce the number of experiments in the design process (Sasank Sekhar Murty et al., 2018) .

This study investigates the ballistic behavior of aluminum alloy plates of varying thicknesses under the impact of different types of bullets. Impact analyses were performed on plates made of Aluminum 7075 and Aluminum 2024 alloys with thicknesses of 1 mm, 3 mm, and 5 mm, using G1 and G7 bullets. Two simulations were conducted where G1 and G7 projectiles impacted aluminum plates of 1,3 and 5 mm thickness and 300 mm x 500 mm cross-sectional area, made of 7075 and 2024 alloys – the most commonly used alloys in the aerospace and defense industries – at a constant speed. In both simulations, the targets were hit, maximum and minimum stresses were determined, and the results were analyzed. Numerical analyses were carried out using ANSYS software based on the finite element method; the effects of bullet type, material type, and plate thickness on ballistic performance were evaluated comparatively. Based on the findings, the maximum stresses and damage analyses of different parameters were examined in detail, and the results were discussed in relation to studies in the literature.

METHOD

G1 and G7 type bullets were used with 300x500 mm aluminum 7075 and aluminum 2024 series plates, modeled in Solidworks, and saved as a parasolid. Plate thicknesses of t=1 mm, t=3 mm, and t=5 mm were selected. Subsequently, this parasolid record was imported into Ansys Workbench and a simulation was performed, and the results obtained are shown below.

Border conditions

The projectile is launched from a plate with fixed supports on all four sides as shown in Figure 1, positioned so that the bullet will strike the center of the plate at a speed of 900 m/s. A 2 mm gap is left between the plate and the bullet. Mesh properties are given in Table 3.

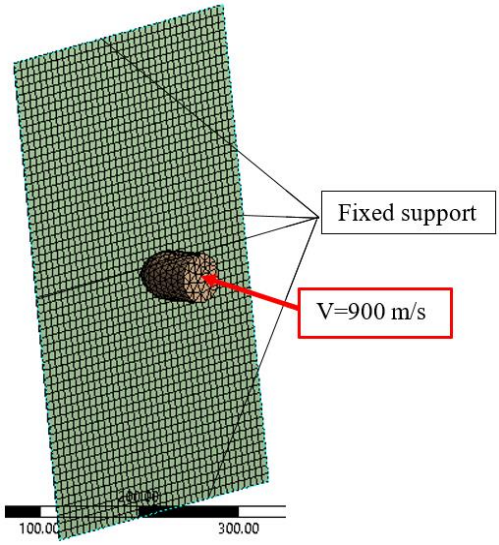


Figure 1Finite element model and boundary conditions used in the analysis.

Table 3Mesh properties.

	G1 (t=1 mm)	G1 (t=3 mm)	G1 (t=5 mm)	G7 (t=1 mm)	G7 (t=3 mm)	G7 (t=5 mm)
Node	5131	5098	5107	4584	4584	4584
Element	5939	5745	5790	2250	2250	2250

a) Analysis performed for G1 bullet type and t=1mm.

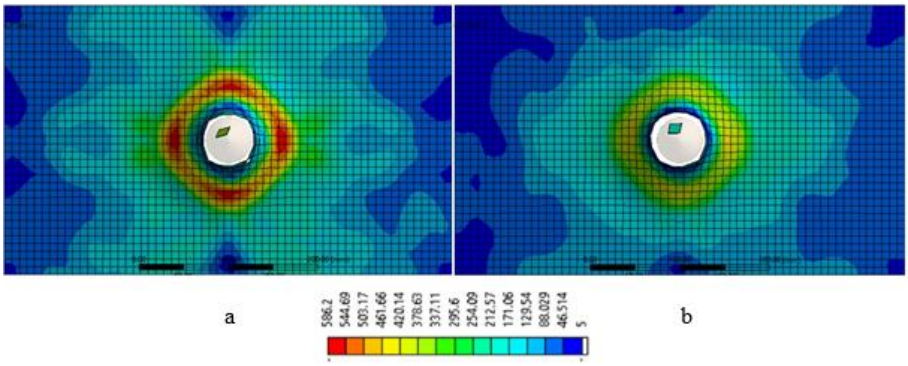


Figure 2a) Stress analysis (MPa) of projectiles fired at Al7075 and b) Al2024 armor materials.

Figure 2 shows the distribution of plastic deformations in the plate for Al7075 and Al2024 plate materials at a G1 bullet velocity of 900 m/s, obtained according to von Mises stresses and presented comparatively. The figure is

given for a single scale value for these two materials. Accordingly, it is seen that the plastic deformation occurs in the form of a circle around the hole. Although the shape of the zones is the same for different materials, the size of the zone in the 2024 material is larger, which can be attributed to its lower yield stress. This situation is observed much more clearly in Figure 3, obtained for different scale values. The scale is obtained by starting with the yield value of each material. In Al7075 material, plastic deformation occurs in a very small area around the hole. This can be explained by the high mechanical properties of this material.

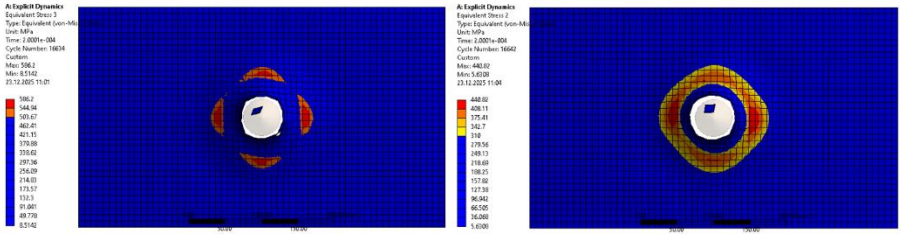


Figure 3Single-scale representation of plastic deformation occurring in the plate for Al7075 and Al2024 plate materials in the G1 bullet type.

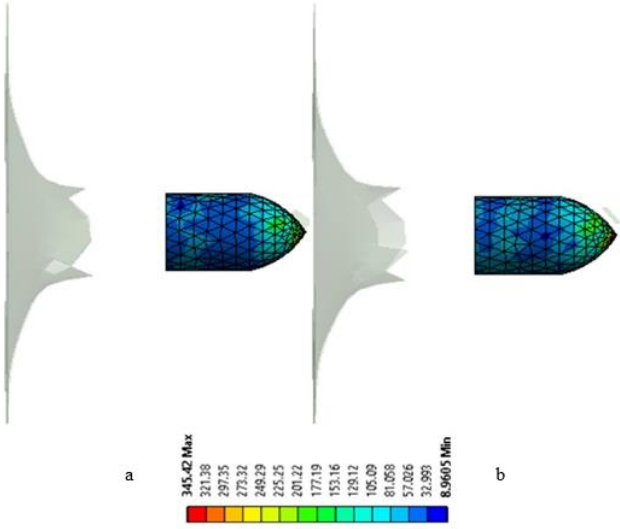


Figure 4a) Stress analysis of projectiles fired at Al7075 and b) Al2024 armor materials at the moment of exit from the armor.

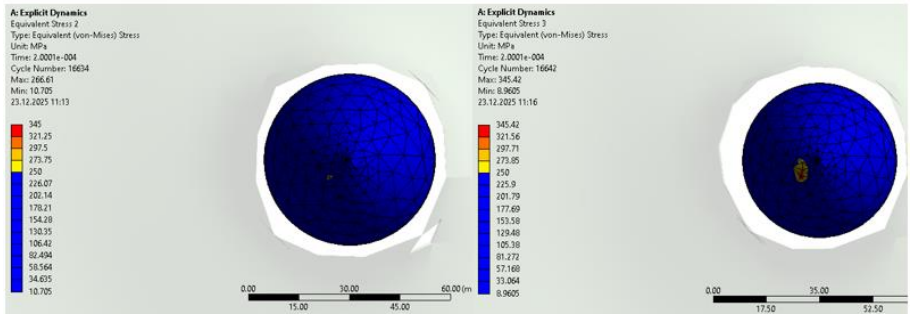


Figure 5Single-scale representation of plastic deformation occurring in the bullet for Al7075 and Al2024 plate materials.

The plastic deformation of the structural steel projectile when the G1 bullet penetrates 1 mm thick Al7075 and Al 2024 materials is shown in Figure 4. Similarly, starting from the yield stress value of the structural steel material, Figure 5 is obtained. Examination of Figures 4 and 5 reveals no plastic deformation in the bullet, but stress concentration occurs in a small area. Otherwise, no deformation is observed in the bullet.

b) Analysis performed for G1 bullet type and t=3 mm.

The plastic deformations in Al7075 and Al2024 plates resulting from the firing of a G1 bullet at a speed of 900 m/s are shown in Figure 6 for a single scale. It was observed that increasing the plate thickness from t=1 to t=3 mm did not significantly affect the resulting plastic deformation.

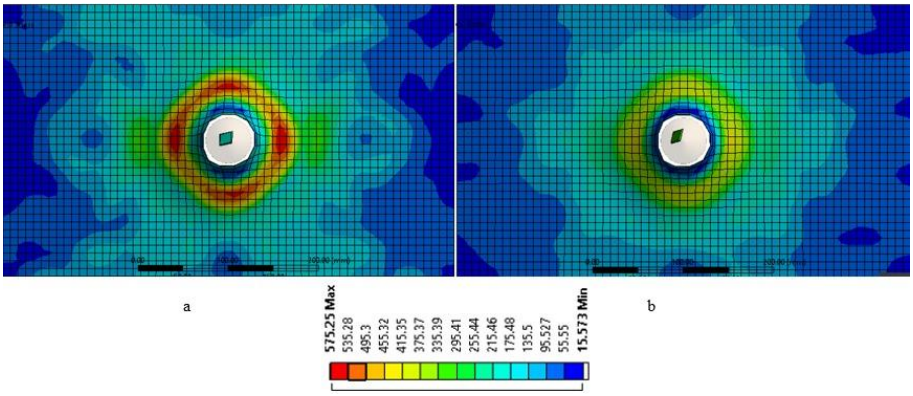


Figure 6a) Stress analyses of bullet fired at Al7075 and b) Al2024 armor materials.

c) Analysis performed for G1 bullet type and t=5 mm.

The plastic deformation caused by the bullet hitting the plates, resulting in a thickness of t=5 mm, is given in Figure 7 for a single scale. The plastic deformation area was obtained with a circular profile geometry, but in terms of magnitude, it was obtained in Al2024 plate material, which is much clearer in separate scales as mentioned above.

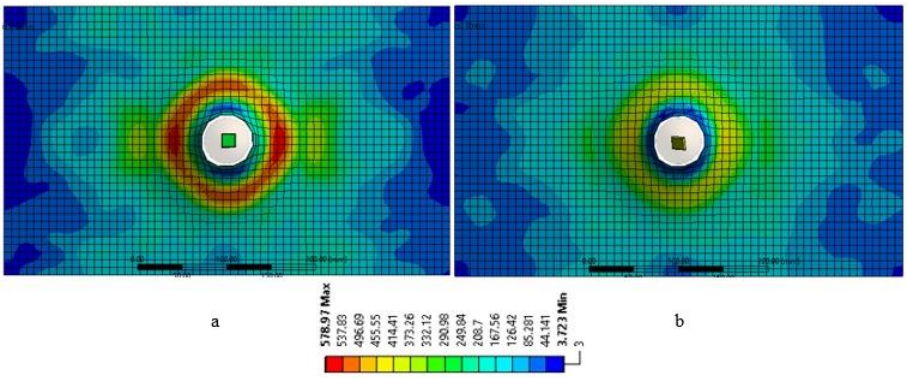


Figure 7a) Stress analyses of bullet fired at Al7075 and b) Al2024 armor materials.

d) Analysis performed for G7 bullet type and t = 1 mm.

The plastic deformation in Al7075 and Al2024 plates resulting from the firing of a G7 bullet at a speed of 900 m/s is shown in Figure 8. The plastic deformation obtained in Figure 8 is for a single scale, and the resulting plastic deformation is circular, as in the case of the G1 bullet. In Figure 9, plastic deformation is obtained for separate scales for the yield strengths of each material. As seen in the figure, the plastic deformation area in Al7075 material is considerably smaller than that in Al2024 material. This is because the mechanical properties of Al7075 material are superior to those of Al2024 material.

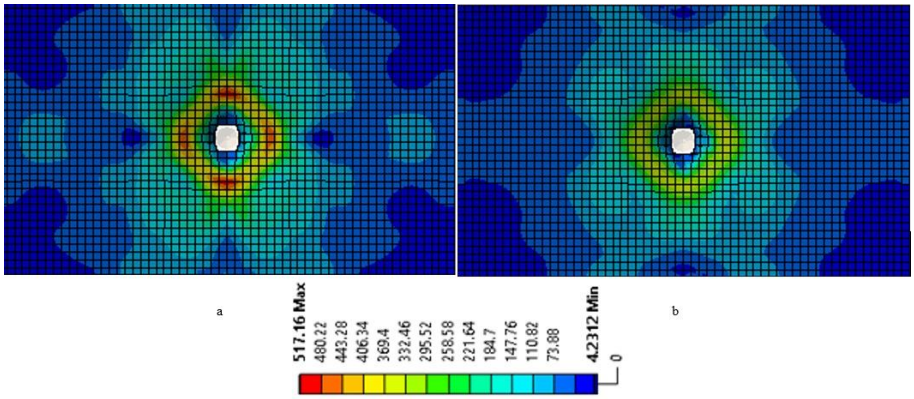


Figure 8a) Stress analyses of bullet fired at Al7075 and b) Al2024 armor materials.

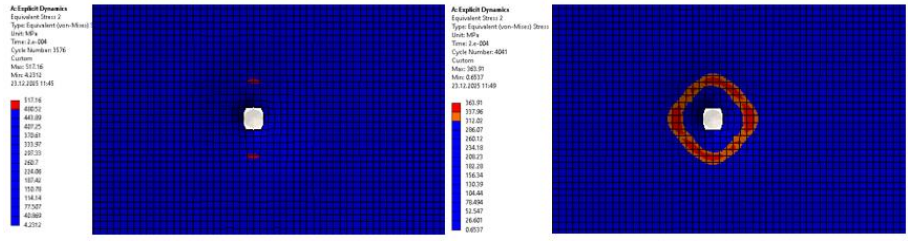


Figure 9Single-scale representation of plastic deformation occurring in the plate for Al7075 and Al2024 plate materials in the G7 bullet type.

The plastic deformation of the G7 bullet when fired at Al7075 and Al2024 plates is shown in Figure 10 for a single scale. No deformation was observed in the bullets. The graph in Figure 11 was obtained by considering the yield strength of the bullet material, structural steel. When the geometries are zoomed in sufficiently, it is seen that there is no deformation in the bullets, but stress concentration is observed in a small area.

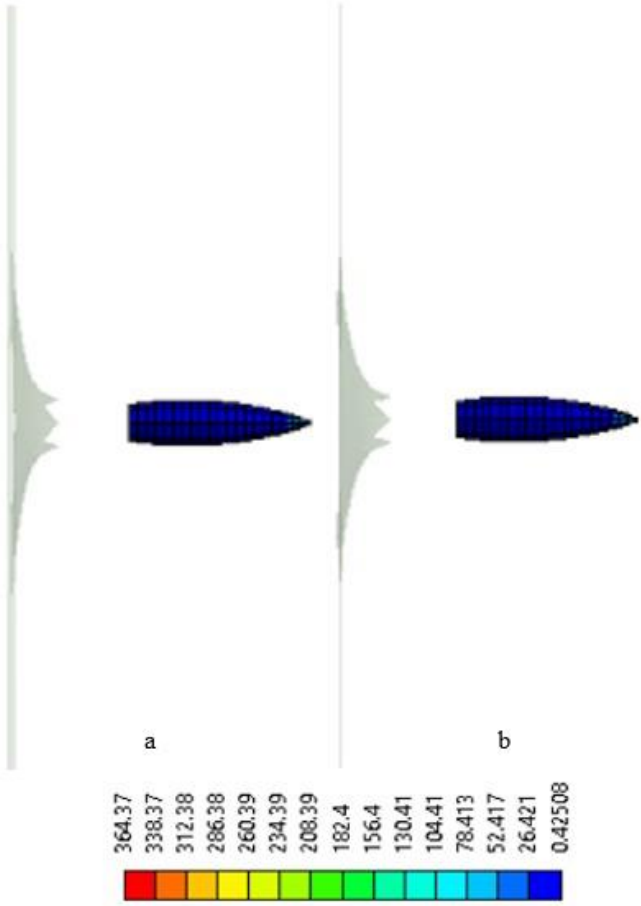


Figure 10a) Stress analysis of projectiles fired at Al7075 and b) Al2024 armor materials at the moment of exit from the armor.

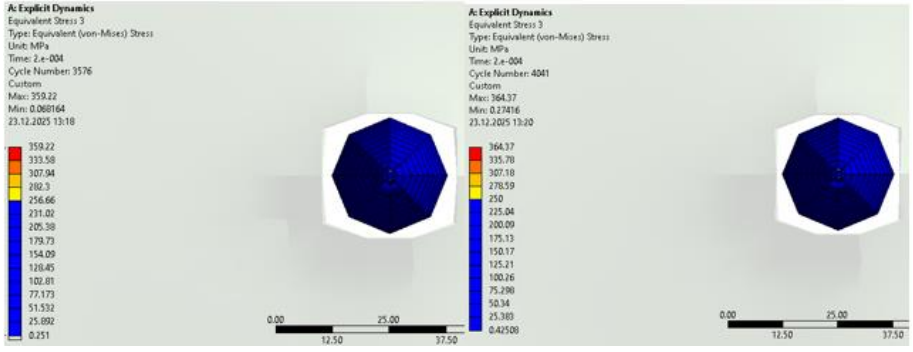


Figure 11. Single-scale representation of plastic deformation occurring in the bullet for Al7075 and Al2024 plate materials.

e) Analysis performed for G7 bullet type and t = 3 mm.

The plastic deformation caused by a G7 bullet fired at 900 m/s on a plate made of Al7075 and Al2024 materials is shown in Figure 12 for a single scale. Increasing the plate thickness from t=1 mm to t=3 mm does not significantly affect the plastic deformation.

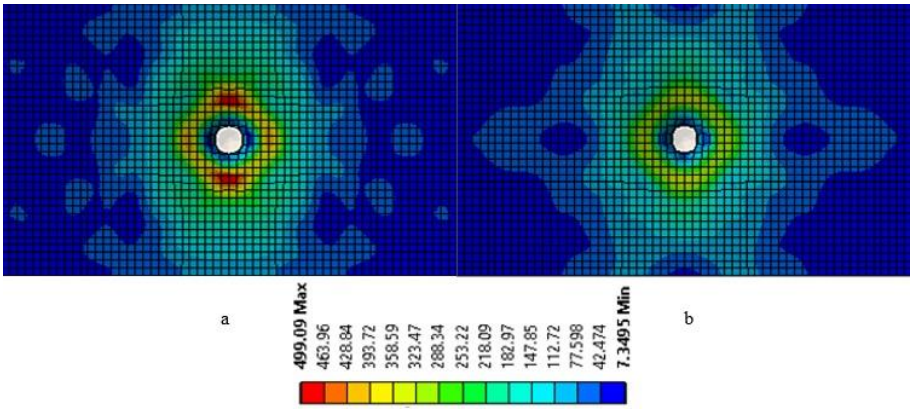


Figure 12a) Stress analyses of bullets fired at Al7075 and b) Al2024 armor materials.

f) Analysis performed for G7 bullet type and t=5 mm.

The plastic deformation caused by the firing of a G7 bullet onto plates with a thickness of t=5 mm is shown in Figure 13 for a single scale. Due to the superior mechanical properties of the Al7075 material used, the resulting plastic deformation was much smaller compared to Al2024. However, when the thicknesses were examined, no noticeable difference was observed in the plastic deformations of either the Al7075 or Al2024 plates. This situation can be explained by the fact that the thickness of the plates used was insufficient to provide resistance against the bullet.

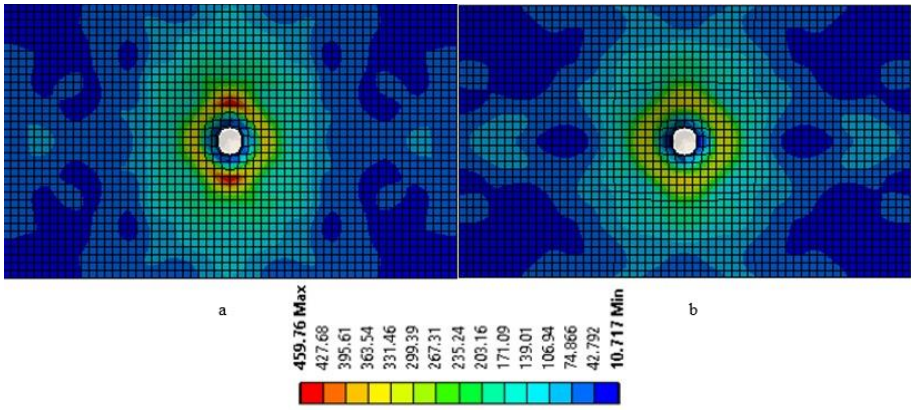


Figure 13a) Stress analyses of bullets fired at Al7075 and b) Al2024 armor materials.

RESULTS

In this study, a simulation was performed of two bullets, G1 and G7, being fired at a speed of 900 m/s at Al7075 and Al2024 plates with thicknesses of t=1 mm, t=3 mm, and t=5 mm. The results obtained can be summarized as follows.

- Plastic deformation was achieved for all materials and thicknesses using both types of bullets.
- The area of plastic deformation resulting from the use of G7 bullet was found to be smaller compared to the area resulting from the use of G1 bullets.
- When comparing plate materials, the plastic deformation observed in Al7075 plate material was lower than that of Al2024 material when using both G1 and G7 ammunition. This can be explained by the superior mechanical properties of Al7075.
- The effect of increasing plate thickness could not be observed for either bullet. This uncertainty stemmed from the plate thickness not being sufficiently large.

In future studies, the effect of varying and increasing the plate thicknesses (t=1, t=3, and t=5 mm used in this study) on plate and projectile behavior will be investigated.

REFERENCES

Akbari, T., Ansari, A., & Rahimi Pishbijari, M. (2023). Influence of aluminum alloys on protection performance of metal matrix composite armor reinforced with ceramic particles under ballistic impact. *Ceramics International* , *49* (19), 30937–30950. doi: 10.1016/j.ceramint.2023.07.046

Alwan, FHA, Prabowo, A.R., Muttaqie, T., Muhayat, N., Ridwan, R., & Laksono, F.B. (2022). Assessment of ballistic impact damage on aluminum and magnesium alloys against high velocity bullets by dynamic FE simulations. *Journal of the Mechanical Behavior of Materials* , *31* (1), 595–616. doi: 10.1515/jmbm-2022-0064

Brekken, KA, Vestrum, O., Dey, S., Reyes, A., & Børvik, T. (2022). Experimental Tests and Numerical Simulations on the Ballistic Impact Response of a Highly Inhomogeneous Aluminum Foam. *Materials* , *15* (13). doi: 10.3390/ma15134651

Gupta, N.K., Iqbal, MA, & Sekhon, G.S. (2007). Effect of projectile nose shape, impact velocity and target thickness on deformation behavior of aluminum plates. *International Journal of Solids and Structures* , *44* (10), 3411–3439. doi: 10.1016/j.ijsolstr.2006.09.034

Sasank Sekhar Murty, A., Lahari Ramya, P., Sai Gowtham, D., & Sudhakar, I. (2018). Simulation and validation of ballistic performance of armor grade AA7075 aluminum alloy using friction stir processing. *Materials Today: Proceedings* , *5* (9), 20186–20192. doi: 10.1016/j.matpr.2018.06.388

TEOMAN, A., GÖDE, E., TONBUL, K., ŞEKER, H. İ., & KUSHAN, MC (2023). Civilian (Covert Armored) Vehicle Armoring and Ceramic Composite Armor Applications. *Defence Sciences Journal* , *2* (43), 473–492. doi: 10.17134/khosbd.1248358

Übeyli, M., Balci, E., Sarikan, B., Öztaş, MK, Camuşcu, N., Yildirim, RO, & Keleş, Ö. (2014). The ballistic performance of SiC-AA7075 functionally graded composite produced by powder metallurgy. *Materials and Design* , *56* , 31–36. doi: 10.1016/j.matdes.2013.10.092

Nano-TiO_2 as Tribological Modifiers in Metal Matrix Composites: A State-of-the-Art Review

Ali KAFALI[1]
Mehmet Tayyip ÖZDEMİR[2]

1- Research Assistant, PhD; Karabuk University, Faculty of Engineering and Natural Sciences, Mechatronics Engineering Department, Karabuk, Türkiye. alikafali@karabuk.edu.tr ORCID No: 0009-0000-9998-2371.
2- Asst. Prof. Dr.; Karabuk University, Faculty of Engineering and Natural Sciences, Mechanical Engineering Department, Karabuk, Türkiye. tayyipozdemir@karabuk.edu.tr ORCID No: 0000-0002-4290-0045

ABSTRACT

Due to superior mechanical strength and enhanced tribological characteristics, MMCs strategically play important roles in modern engineering applications. Nanotechnological approaches and nano-scale reinforcement integration are highly welcome in recent years to extend the limits of the performance of these materials, especially for optimizing friction and wear behaviors. The present chapter contains a comprehensive "state-of-the-art review" on the contemporary research in using TiO_2 nanoparticles as a tribological modifier in metal-matrix composite materials. The influence of Nano-TiO_2 reinforcement on friction coefficient values, wear rates, microhardness values, and tribofilm formation rates has been studied in relation to the type of metal matrix used, volume fraction levels, and fabrication conditions being diverse in nature. Synthesized data provide evidence that upon attaining a homogeneous dispersion of nanoparticles within the matrix, wear resistance significantly improves, while friction coefficient is minimized. Emphasis is provided toward the role of the oxide-based protective films formed at the contact surfaces by the tribochemical reactions as the key mechanism to improved frictional behavior. Upon this, existing gaps in the literature are identified and recommendations are suggested regarding potential future research routes concerning high-temperature resistance, hybrid reinforcement strategies, and advanced characterization techniques.

Keywords – Metal Matrix Composites, Tribology, Nano-TiO_2, Wear Resistance, Friction Coefficient.

INTRODUCTION

In today's material research and engineering, besides the high strength of the components, another important characteristic of these components needs to be high tribological performance. The growing importance of friction/wear problems in such components makes the development of innovative material solutions even more necessary. Even here, besides high specific strength and high tribological performance, the high strength and specific strength of MMCs are attractive. By incorporating reinforcement elements (ceramics) into the metal matrix components, MMCs can develop 30% higher wear strength and 20% higher lightness compared with traditional alloys. The past few years have paid growing attention to the use of nanotechnologies and nano-sized reinforcement elements not only for developing more advanced MMCs with super Tribological performance but also for improving the Tribological performance of existing MMCs (Ren et al., 2024). In such applications, besides titanium dioxide nanoparticles with average dimensions around 100 nm (Nano-TiO_2), titanium dioxide

nanoparticles with dimensions smaller than 100 nm show high hardness (1000-1200HV), high thermal stability, and high chemical inertness. Compared to micron-sized particles (d <10μm), these nanoparticles show higher surface area and higher surface area ratios (100 times higher surface area), making them highly effective reinforcement elements in matrices (Stawarz et al., 2019). These nanoparticles dissolve in high uniformity in matrices and show increased wear strength and low friction coefficient because of the high load-carrying capability and inability of dislocation movement. In such applications, these nanoparticles act as an effective "tribological modifier" (improving agent) (Qi et al., 2013).

There is a lack of systematic knowledge about it in the current literature, with some gaps awaiting closure despite very interesting findings. Firstly, the literature analysis showed that it had primarily centered on particular metal matrices, such as aluminium, accounting for about 65% of the literature, or particular production methods. Comparison experiments among different methods, such as stir casting, powder metallurgy, or friction stir welding, are limited to date. Secondly, notwithstanding their crucial role in optimizing the particular property, the impact of particular levels of doping with nano-TiO_2 (in the 0.5 to 5% range) and phase compositions, such as anatase, rutile, and brookite, of the nanomaterial has remained unclear to date. Thirdly, performance comparison experiments for nano-TiO_2 against other usual nanofiller materials, such as Al_2O_3, SiC, and/or CNT, and its efficiency in other couple materials, such as Cu, Mg, and/or Ti, had remained unexplored to date. The aim of this book chapter will be to present the particular application of nano-TiO_2 as a tool to optimize the particular property of MMCs through a review of current state-of-the-art knowledge. In this study, the role of nano-TiO_2 additions will be investigated to understand their effects on wear mechanisms, CoF, and surface Hardness properties, focusing on parameters such as the additive levels, base materials, and processing methods.

In addition, the aforementioned manufacturing approaches hindering nanoparticle agglomeration, and thereby handling equal distribution, and their influence on the resultant performances would also be discussed. The tribological property of the created material in varied environmental states (dry friction, presence of a liquid, and high temperature) would also be examined. The systematic review of the current body of knowledge would identify the factors behind the performances of the respective materials and the gaps in the current knowledge on nano-TiO_2-reinforced MMC systems, thereby fulfilling the objectives of the proposed study by mapping a guideline document for the current and future researchers and practitioners on the aforementioned topic, namely the related new research areas of using a combination of nano- and micro-reinforcement and surface technologies.

BASIC CONCEPTS

Fundamentals and Importance of Tribology

Tribology, which has been derived from the Greek work "tribos," to rub, refers to "the science and technology of interacting surfaces moving relative to, or slipping against, each other with associated friction, wear, and lubrication effects, considered within the context of materials science, mechanics, and physics principles"(Adetunla et al., 2023). Surface interactions cannot be avoided within the operational processes of any kind of mechanical system. Consequently, the forces of friction and wear resistance due to surface interactions lead to material depletion of components, damage to dimensional tolerances, and, ultimately, system failure. Realizing that approximately 23% of overall energy consumption worldwide has been ascribed to surface interactions, any effort to reduce these surface interactions or lose these energies would be very important from the point of improving efficiency, saving energy, and prolonging the economy of components (Holmberg & Erdemir, 2017). Two fundamental parameters stand out in evaluating the tribological performance of materials:

• Coefficient of Friction (μ): This is the ratio of the resistance (friction force) to relative motion between two contact surfaces to the applied normal load. Reducing this coefficient, a dimensionless quantity, directly contributes to reducing energy losses and heat generation in the system (Černe et al., 2022).

• Wear Rate: Defined as the mass or volume loss from the material under a given load and sliding distance. High wear resistance indicates the material's capacity to maintain surface integrity and withstand harsh operating conditions (Villanueva et al., 2024).

The relationship between these two parameters is complex; a low coefficient of friction does not always guarantee a low wear rate. Therefore, simultaneously optimizing these two properties according to application requirements in the design of metal matrix composites (MMCs) is a fundamental goal of materials engineering.

Tribological Performance Indicators: Wear Resistance and Friction Coefficient

Tribological properties of materials depend on the interaction between the various properties of the structure. Moreover, the process can be affected by external factors such as the load applied, sliding velocity, temperature, and nature of the other surface. In scientific literature, the determination of the properties depends on the test rigs ASTM G99 (pin-on-disk), ball-on-disk, and reciprocating tribometer. In the theory, frictional and wear process principles are expressed using the Archard-Holm theory (Kundu et al., 2016). According to the theory, the actual contact area between two surfaces only makes up a small portion (1-10%) of the nominal contact area, which

depends on the load applied and the material's hardness. Thus, the adjustment of the hardness values and parameters of the surface roughness plays a major role in the adjustment of the resultant frictional force and the loss of material. Description of the factors responsible for influencing the properties of the process and the underlying principles are expressed below:

Temperature Effect: Temperature is among the most important parameters influencing the durability of any system involving friction. Upon considering various braking systems and high-speed trains, the operating temperatures, which range from 200°C to 700°C, indicate that as the temperature of the system increases, there would be diminished resistance to wear due to softening, thermal expansion, oxidation, and changes in microstructural phases of the materials. For instance, research has revealed that for aluminium matrix composites, which exhibit coefficients of friction between 0.4 and 0.5, the value increases to between 0.5 and 0.6 for temperatures of 300°C. According to the study conducted by Verma et al. on the role of braking materials, achieving thermal stability could minimize wear by 2-3 times (Verma et al., 2016). Che et al. and Chen et al. explained that by choosing right materials and optimizing different phases of materials, like ceramics, the adverse effects of temperatures could be overcome (Che et al., 2018; Chen et al., 2022).

Load, Sliding Speed, and Surface Roughness: As normal force enhances pressure on the contact surfaces, it causes plastic deformation, which might result in an increase in wear rates by 50-150%. The sliding speed influence is linked with flash temperature and tribofilm formation dynamics at the friction surface. In their research work on 5052 aluminium alloys, Dou and Xia observed that as temperature rose, the friction coefficient generally increased by 20-30%. The influence of load and speed, however, remained less significant (Dou and Xia, 2019). Additionally, surface roughness (Ra) is one factor that directly impacts the actual contact area. In support of the impact of surface roughness on friction coefficient, increase in Ra from 0.2μm to 2.0μm was shown by Liang et al. to increase the friction coefficient by 25%. The high stresses created due to aspirate or surface roughness contact points on surfaces increase the tendency towards adhesive wear and micro-cracks due to surface asperities (Liang et al., 2018).

Self-lubricating materials and surface treatments are one of the key methods in this regard. Torres et al. showed that the effect of solid lubricant phases such as graphite, MoS_2, and h-BN in the composition resulted in a reduction in friction of 40-60% at high temperatures (Torres et al., 2018). In this regard, self-lubrication materials function as contact-free agents in a vacuum and high temperatures, as liquid lubricants fail to perform in such conditions. In such cases, their mechanism is also applicable to nano-reinforcements, wherein particles like Nano-TiO_2 create a protective and load-carrying film over the wear surface and reduce frictional force and

improve wear resistance. In modern engineering practices, composites and their advanced tribological capabilities are being synthesized simultaneously in combination with nano-reinforcements, solid lubricants, and surface coatings.

Metal Matrix Composites: Definition and General Characteris

Metallic matrix composites (MMC) can be considered a separate type of composite materials, having the reinforcing elements embedded in the metal matrices, distinguished by their improved mechanical properties. With regards to the favorable properties like lightness, high tensile strength, resistance to corrosion, and enhanced resistance to wear and tear, MMC has important applications. Tensile strength, resistance to corrosion, and enhanced resistance to wear and tear are the key factors that contribute largely to its applications. Aluminium Matrix Composites (AMC) has important applications in the automotive sector because they are light in weight and possess enhanced strength (ERGÜL & KURT, 2021). Similarly, the use of Titanium Matrix Composites (TMC) in the medical industry can be attributed to its resistance to corrosion due to chemical agents and its bio-compatibility (Escaich et al., 2020). The Matrix Composites consist of the matrix that has several elements. However, regardless of the matrices that are composed of aluminium, titanium, or magnesium, the elements acting as reinforcement agents are silicon carbide (SiC), boron nitride (BN), or alumina (Al_2O_3) (Bodunrin et al., 2015). Even the Al6061 matrix with the addition of silicon carbide (SiC) and Al_2O_3 shows improved hardness, with the potential for a wide range of applications (KOCAMAN & ATEŞ, 2023). Along with this, the introduction of boron nitride reinforcement in titanium based MMCs contributed greatly to the enhancement of its mechanical performance (TOPCU & KARAMAN, 2019).. On the other hand, techniques being employed in these composite developments are also quite varied. Powder metallurgy, casting techniques, as well as many other engineering techniques, are mostly favored in MMC production (Bodunrin et al., 2015). The choice of reinforcement materials in these production techniques plays a major role in defining the end performance of these composite materials. For instance, in an experimental investigation of the different experiments of various reinforcement rate values, it was demonstrated that the 20% rate of reinforcement offered the best results in terms of mechanical properties (PUL & BAYDAROĞLU, 2020). There are many developments that are currently being emphasized in order to ensure optimal functionality of metal-based composite materials. In this regard, a lot of studies are currently being done in respect to different experiments of different types of material reinforcements in respect to their effects on different properties of these MMCs. As a result, in addition to a broad approach in the manufacturing of MMCs, because of different types of composite materials in use in different

applications, there is a heightened possibility of these advanced materials (Srinivas & Mohan, 2020).

Nanoparticle Reinforced Composite Systems

Nanoparticles reinforced composite materials are an immense study area in the fields of both engineering and material science. The properties of the composite material are based on some parameters that include the type of nanoparticles and their size. The strength and coordination of nanoparticles are of utmost importance, especially in medical science and their electric properties (Phan & Haes, 2019; Zhang et al., 2017). The stability of the nanoparticle composite material is highly remarkable. The advantages of such materials create an impressive effect on their respective operational area. The stability is associated with physical and chemical properties and locking and dispersions of nanoparticles (Yu et al., 2025). The uses of nanoparticle reinforced composite materials in medical science operations like the delivery of drugs and tissue engineering are newly invented and exponentially growing (Bernal-Chávez et al., 2021). The sterilization of nanoparticles is most valuable and important for effective and proper medical operations. The different sterilization techniques improve the use of nanoparticles in biotic conditions with minimal potential of contamination (Bernal-Chávez et al., 2021; Tapia-Guerrero et al., 2020).

Also, the mechanical properties in composite nanoparticles have a direct relationship with the integrity of the added structure where adding metal-composed nanoparticles increases the mechanical strength of the composite matrix (Saatchian et al., 2025). Moreover, the optimization of physical and chemical features of nanomaterials in composites is important. Different forms of nanoparticles can provide enhancements in the features of a structure, such as strength and durability (Shivalingam et al., 2024). For example, the addition of ZnO/Si3N4 nanoparticles can be helpful, along with research on alloys which optimize different features (He et al., 2018). Thus, composite systems with nanoparticles have great potential in materials science. Future research will be directed towards optimization in stability, efficiency, and development of additional functions/properties of such composites (Thoniyot et al., 2015).

TITANIUM DIOXIDE (TiO_2) NANOPARTICLES

Structural and Physical Properties

Titanium Dioxide (TiO_2) nanoparticles have been in the limelight lately because of the peculiar structure and physical properties that the nanoparticles pose. These have resulted in extensive use in the fields of materials science research, photocatalysis research, and biomedicine research. The structure that the TiO_2 nanoparticles pose can be extremely diverse based on the procedure that they have been synthesized under, the anatase and rutile form that the particulate matter comprises, and the final sizes that have been obtained. The extent to which control may be obtained on the properties that the particulate matter presents to the system is an important factor that affects the optimization toward the already existing different uses. The procedure that the TiO_2 nanoparticles have been synthesized under may influence the final sizes and the final shapes that the particulate matter may pose. This will greatly have an influence on the final properties that the particulate matter may pose to the system. For example, the procedure that the TiO_2 nanoparticles have been synthesized under may utilize the hydro-thermal procedure and the sol-gel procedure to synthesize the particulate matter depending on the sizes that the system aims to achieve. This may be an important factor in the application related to the UV shielding effect (Tavakolian et al., 2024). Techniques such as Transmission Electron Microscopy (TEM) analysis, UV-VISIBLE spectroscopy analysis, are usually used to analyze the characteristics exhibited by the particles. These would allow for the analysis of characteristics such as the size distribution of the particles, as well as the surface characteristics exhibited by the particles (Ram et al., 2021). It has been found to be true that the size of the particles can be designed to vary from a few nanometres, resulting in an improvement in the surface area to volume ratio and interaction with other materials present in the surrounding areas (Irudayaraj et al., 2023). The final characteristics exhibited by the TiO_2 nanoparticles, along with the high temperature stability, characteristics, and responses are very significant on an atomic level interaction basis. For instance, the presence of the TiO_2 particles within the polymer matrix in case of PLA biopolymer polymers would undergo significant improvement in the high temperature stability characteristics along with the corresponding strength properties (H. E. Ali et al., 2022). Additionally, the high temperature stability characteristics were found to vary in a considerable manner when compared to the other properties in the biopolymer composition in an unaltered form. These aspects can be attributed to the reinforcement offered by the nanoparticles in preventing the flow characteristics exhibited by the biopolymers at high temperatures (Gallah et al., 2020).

Moreover, the photocatalytic capabilities of TiO_2, having a wide band gap with the ability to create electron-hole pairs after being irradiated with

UV light, have also gained immense attention for use in the degradation and disinfection methods for pollutants (Ghamarpoor et al., 2023; J. Wu et al., 2021). Some literature has also shown the significant use of TiO_2 for reducing organic pollutants through the photocatalytic process, having vast applications for water and air purifications techniques (Nguyen et al., 2021). The photocatalytic ability is also largely dependent on surface area and the crystalline structure of TiO_2 nanoparticles, where anatase TiO_2 has significant photocatalytic ability over the rutile structure of TiO_2 (Karunakaran et al., 2020). Moreover, the antimicrobial ability of TiO_2 nanoparticles has gained significant attention for their applications. Due to high reactivity, these particles or materials have the ability to attack or destroy the bacterial cell membrane itself (de Freitas Rosa Remiro et al., 2021). Such capabilities have immense applications for preparing materials for the healing process of wounds for use in the healthcare field for preparing bio-coated materials that remain uninfected (Rashid et al., 2021). Therefore, from now onwards, titanium dioxide nanoparticles will also possess immense intrinsic properties for their effective applications in various domains because of their high reactivity on their surfaces and their varied properties. Moreover, their varied and adjustable properties through synthesis and modification techniques will also help TiO2 nanoparticles to efficiently perform their roles in various domains, ranging from material science, technology, to their applications in pollution and medical treatment, performing better in their respective domains compared to other nanoparticles.

Compatibility and Interfacial Interaction of Nano-TiO_2 with Metal Matrices

In the case of nano-TiO2, due to its versatility and photocatalytic properties, there has been significant study related to the interaction and compatibility of nano-TiO2 with different metals, and it has an important role in the creation of advanced metal matrix composites. Adding nano-TiO2 to the metal composites not only helps to improve the mechanical properties of the material, it also helps to improve the functional properties of composites, thereby increasing the applications of composites in aerospace, automobile, and construction materials (Malaki et al., 2019; Venugopal & Karikalan, 2021). Among the significant advantages related to the addition of nano-TiO2 in the case of the metal composites, the role of the mechanical properties of the resulting composite material is of prime interest and benefit to the researcher. It has been observed that the addition of nano-reinforcement to the material causes a significant improvement in the strength and brittleness of the metal composites (A. Ali, 2021). The excellent mechanical properties that the composites have are believed to be the result of a strong binding force between the TiO_2 nanoparticles and the metal matrix, and this binding force played a crucial role in the easier flow of the

mechanical stress in the materials (Malaki et al., 2019). The mechanical properties that the materials have are only among the benefits that the integration of nano-TiO_2 to the composite materials has. There are still other benefits that can be listed as follows: Adding nano-TiO_2 to the concrete mixture may provide self-cleaning ability to the construction materials (Bautista-Gutierrez et al., 2019). This may have a significant impact on the construction materials since there is a need to ensure that the construction materials are clean and free from any form of pollution.

Besides this, its photocatalytic function contributes to environmental sustainability, through the decomposition of pollutants, by Bautista-Gutierrez et al (Bautista-Gutierrez et al., 2019). Moreover, compatibility with different metal matrices, such as aluminium and magnesium, has been shown, thus indicating a broad scope of application in diverse alloy compositions. Anand & S., Venugopal & Karikalan discussed that during the integration of nano-TiO_2 into the metal matrix, potential problems include porosity and agglomeration during processing (Anand & S, 2020; Venugopal & Karikalan, 2021). Some stir casting processes have revealed that the control in the distribution of TiO_2 in the metal matrix is an important factor for maximizing improvements in mechanical properties and homogeneity of properties. Thus, from this discussion, it can be concluded that the addition of nano-TiO_2 to metal matrix composites provides major advances in mechanical performance and also extends functional improvements, especially ecological benefits. Such shaping of the microstructure by proper machining methods could give further possibilities for the development of new material variants to meet current engineering requirements.

FACTORS AFFECTING TRIBOLOGICAL PERFORMANCE

Effect of Nano-TiO_2 Content and Distribution

Nano-TiO_2, an promising additive in lubricants, displays a significant potential in the improvement of tribological properties. The incorporation of the nanomaterial may depend upon the concentration, distribution, and interaction with the lubricating fluid. Primarily, the material has the property of a reinforcing agent, adjusting the mechanical properties of polymeric nanocomposites and adding new functions that are hard to obtain using conventional materials. These benefits may appear in the form of improved wear resistance and coefficient of frictions for the material being lubricated (Cazan et al., 2021). Notably, in the automotive engines, the combination of

the material with an alumina nanomaterial has displayed better results in the improvement of tribological properties, enhancing the service life and the efficiency of the components (M. K. A. Ali et al., 2016). Various research studies have demonstrated that the optimal concentrations of the material in the lubricant have a significant influence upon its performance properties. As demonstrated in various research works, the addition of 3.0% of the material weight in the water-based nano-lubricants can significantly improve the tribological properties during micro-rolling titanium foils (Ma et al., 2023; H. Wu et al., 2018). This is supported by findings that particular concentrations allow for optimum interaction in the lubrication zone and hence give far better performance in load-carrying and lubrication properties than conventional lubricants (H. Wu et al., 2017). The other critical factor that relates to the tribological performance of TiO_2 is the aspect of nanoparticle size and its distribution within a lubricant. Spatial disposition of the TiO_2 may either work to increase or reduce lubrication efficiency. For example, smaller sizes of particles provide better dissolution in the lubricant matrix, therefore making friction and wear reduction more effective (Gulzar et al., 2017). This is of importance because this disposition directly dictates how efficient the lubricant will be in protecting the surfaces from wear; also, it develops a need for better methodology to ensure that the dispersion of TiO_2 is stable and homogeneous within the base lubricant (Gulzar et al., 2017). Another aspect has been emphasized in research: the idea that the interaction mechanism of nano-TiO_2 is not solely a size or concentration dependence but also has important chemical interactions. Studies have indicated that surface-functionalized TiO_2 nanoparticle Valley can have an enhanced affinity for the lubricant base and further improve their performance as nanoparticle additives (Fang-cheng et al., 2022). Improved lubrication effects have been attributed to synergistic interactions between TiO_2 and other additives, supporting the idea that composite systems can outperform single-additive formulations in tribological applications (Xia et al., 2016).

Surface Film Formation and Tribochemical Reactions

The formation of surface films and tribochemical changes are very significant in materials science and engineering. Surfaces directly alter the chemical and physical properties of materials; hence, much effort has been made to study and control these properties for advanced applications. The relationship between surface change, film formation, and tribochemical interactions is a fundamental factor in the application of many products. Surface film formation results due to interactions between the material and the environment. For instance, some of the important parameters in film development are surface roughness. Toker et al., in the study to establish the influence of surface roughness of 316L stainless steel on the precipitation of

oxides in liquid, affirmed that surface roughness did not have a positive effect on such reactions (TOKER et al., 2020).

On the one hand, it is mentioned that smooth surfaces provide functions such as maintaining aesthetic value, inhibiting discoloration, and inhibiting the retention of plaque (KAM HEPDENİZ et al., 2019). Tribo-chemical reactions, in which chemical changes occur on the surface, relate closely to friction, wear, and the formation of a surface film. Surface coating and modification techniques are frequently utilized in order to improve the tribological properties of the surface. This is especially pertinent when considering that a study carried out by Çelik on the structural and tribological characteristics of CrN films on pure titanium shows that such films increase strength and endurance of the compound (ÇELİK, 2016). Furthermore, from research carried out by Kılıçay, it is seen that surface modification processes increase wear resistance; hence, improvement of tribological characteristics takes place (KILIÇAY, 2020). Regarding antimicrobial characteristics of surfaces, an important role is played by the creation of antimicrobial compounds on surface layers and related chemical processes. This is especially important regarding research carried out by Zor et al., where an investigation on the characteristics of different types of wooden surface layers and antimicrobial levels shows that information gathered from an examination of surface layers of a coating is of great importance for use of the whole surface itself (ZOR & ŞEN, 2023). Chemistry of both and relations between both compounds on a surface coating decides film formation on a surface layer and thus affects tribo-chemistry processes itself.

In conclusion, surface film formation and tribochemical reactions are crucial processes in materials science and engineering applications. Modification of surface properties, coating thickness, and chemical structure are key components influencing material performance. Therefore, a deep understanding of surface film formation and tribochemical processes is critical for materials engineering,

GAPS IN THE LITERATURE AND FUTURE PERSPECTIVES

Nano-TiO_2/Matrix Metal Composites, or nano TiO_2/Matrix Metal Composites, possess immense potential to produce modern engineering components due to their advanced wear-resistance and low friction coefficient. Although the available literature indicates optimism, particularly for the soft matrix compounds aluminium and copper, the following categories of potential are of much value for future research: High-temperature purposes: The extent of stability of nano TiO_2 at higher temperatures remains unexplored. Multiples of reinforcements: As a hybrid reinforcement agent along with other nanoparticles, such as graphene, SiC, and MoS_2, nano TiO_2 could lead to an improvement in the anti-friction properties of the hybrid components. Surface modification methods: Rather than using the nano-TiO2 dispersion in the matrix, the oriented dispersion of nano-TiO2 on the surface or the integration of the nano-TiO2 into the surface-by-surface engineering can provide optimized friction characteristics, particularly in the first contact even. Recyclable and environmentally friendly matrices: Following the sustainable criteria, the utilization of recycled metals with the incorporation of TiO2 can be possible from the viewpoints of both economy and the environment. Advanced characterization techniques: The research works involving the aid of advanced micro-structural characterization techniques, like in-situ test systems, TEM, AFM, and XPS, would provide a clearer insight into the tribological aspects. Extensive research on the a forementioned aspects would allow the nano-TiO2 doped MMCs to enter the broader range of applications in the industrial scale and thus provide marked advancements in the area of tribology.

RESULTS

This chapter provides a critical analysis of various research studies that focus on the application of titanium dioxide (TiO2) nanoparticles (also referred to as ‘modifiers’) to enhance or optimize the tribological properties of MMCs. The chapter includes an in-depth study conducted by the authors of this chapter on the significant aspects affecting the performance of TiO2-modified MMCs, such as the level of volume fraction, uniformity of distribution, metal matrix type, and manufacturing method. These are the key elements for exploring how TiO2 affects the coefficients of friction,

wear rates, and surface roughness of MMCs. The primary contributions to the body of knowledge from previous investigations into TiO2-modified MMCs are:

• The distribution of nano-sized titanium dioxide (TiO_2) reinforcement in a composite material matrix affects how the composite behaves under friction and wear. The lack of agglomerates also has the same effect. The optimum ratios of components in the composite allows for the immobilization of dislocation and increases the stiffness, significantly reducing the friction coefficient and increasing wear resistance.

• The major reason that the tribofilm mechanism is responsible for reducing the wear of a material is due to the development of an oxide layer on the contact surface created by friction. When subjected to heat and pressure, the nano-sized titanium dioxide particles form a tribofilm due to frictional forces which reduce the bond between surfaces and therefore, reduce the amount of material lost from contact through metal-metal interface.

• Manufacturing Method(s) Affecting Effectiveness of Nano TiO_2: The performance of nano TiO_2 relies heavily on the manufacturing process' ability to maintain microstructural integrity. The [previously] published articles agree that both powder metallurgy and ultrasonically assisted stir casting processes provide for improved wettability of the nanoparticles, allowing for better bonding strength and success.

• Thermal Stability and Missing Data: The number of studies conducted under high temperatures and extreme environmental conditions (such as corrosive environments) remains limited. Although current data indicate that the thermal stability of nano TiO_2 positively contributes to high-temperature wear resistance, more comprehensive experimental data sets are needed in this area.

Future Research Routes: In the future, the development of "hybrid composite" systems using nano TiO_2 in combination with other reinforcements such as graphite, h-BN, or CNT, and their support through surface engineering (coating/heat treatment), stands out as the most promising research area for pushing the tribological limits of MMCs.

REFERENCES

Adetunla, A., Afolalu, S., Jen, T.-C., and Ogundana, A. (2023). The Advances of Tribology in Materials and Energy Conservation and Engineering Innovation. *E3S Web of Conferences*, *391*, 01014. https://doi.org/10.1051/e3sconf/202339101014

Ali, A. (2021). Wear Behavior of Al6061/TiO2 Composites Synthesized by Stir Casting Process. *Journal of Advanced Engineering Trends*, *41*(2), 113–125. https://doi.org/10.21608/jaet.2021.55413.1081

Ali, H. E., Elbarbary, A. M., Abdel-Ghaffar, A. M., and Maziad, N. A. (2022). Preparation and characterization of polyvinyl alcohol/polylactic acid/titanium dioxide nanocomposite films enhanced by γ-irradiation and its antibacterial activity. *Journal of Applied Polymer Science*, *139*(24). https://doi.org/10.1002/app.52344

Ali, M. K. A., Xianjun, H., Mai, L., Qingping, C., Turkson, R. F., and Bicheng, C. (2016). Improving the tribological characteristics of piston ring assembly in automotive engines using Al2O3 and TiO2 nanomaterials as nano-lubricant additives. *Tribology International*, *103*, 540–554. https://doi.org/10.1016/j.triboint.2016.08.011

Anand, N., and S, S. K. (2020). Development and influence of tribomechanical properties on magnesium based hybrid metal matrix composites-a review. *Materials Research Express*, *7*(3), 036520. https://doi.org/10.1088/2053-1591/ab7d08

Bautista-Gutierrez, K. P., Herrera-May, A. L., Santamaría-López, J. M., Honorato-Moreno, A., and Zamora-Castro, S. A. (2019). Recent Progress in Nanomaterials for Modern Concrete Infrastructure: Advantages and Challenges. *Materials*, *12*(21), 3548. https://doi.org/10.3390/ma12213548

Bernal-Chávez, S. A., Del Prado-Audelo, M. L., Caballero-Florán, I. H., Giraldo-Gomez, D. M., Figueroa-Gonzalez, G., Reyes-Hernandez, O. D., González-Del Carmen, M., González-Torres, M., Cortés, H., and Leyva-Gómez, G. (2021). Insights into Terminal Sterilization Processes of Nanoparticles for Biomedical Applications. *Molecules*, *26*(7), 2068. https://doi.org/10.3390/molecules26072068

Bodunrin, M. O., Alaneme, K. K., and Chown, L. H. (2015). Aluminium matrix hybrid composites: a review of reinforcement philosophies; mechanical, corrosion and tribological characteristics. *Journal of Materials Research and Technology*, *4*(4), 434–445. https://doi.org/10.1016/j.jmrt.2015.05.003

Cazan, C., Enesca, A., and Andronic, L. (2021). Synergic Effect of TiO2 Filler on the Mechanical Properties of Polymer Nanocomposites. *Polymers*, *13*(12), 2017. https://doi.org/10.3390/polym13122017

ÇELİK, İ. (2016). CrN KAPLANAN SAF TİTANYUMUN YAPISAL ve TRİBOLOJİK ÖZELLİKLERİNİN İNCELENMESİ. *Anadolu University Journal of Science and Technology-A Applied Sciences and Engineering*, *17*(3). https://doi.org/10.18038/btda.94015

Černe, B., Bergant, Z., Šturm, R., Tavčar, J., and Zorko, D. (2022). Experimental and numerical analysis of laminated carbon fibre-reinforced polymer gears with implicit model for coefficient-of-friction evaluation. *Journal of Computational Design and Engineering*, *9*(1), 246–262. https://doi.org/10.1093/jcde/qwab083

Che, Q., Zhang, G., Zhang, L., Qi, H., Li, G., Zhang, C., and Guo, F. (2018). Switching Brake Materials To Extremely Wear-Resistant Self-Lubrication Materials via Tuning Interface Nanostructures. *ACS Applied Materials* and *Interfaces*, *10*(22), 19173–19181. https://doi.org/10.1021/acsami.8b02166

Chen, Y., Gao, F., Fu, R., Su, L., Han, X., and Yang, J. (2022). Braking-experimental study on the relationship between friction material type and brake disc temperature. *Industrial Lubrication and Tribology*, *74*(8), 975–984. https://doi.org/10.1108/ILT-03-2022-0107

de Freitas Rosa Remiro, P., de Sousa, C. P., Alves, H. C., Bernardo, A., and Aguiar, M. L. (2021). In Situ Evaluation of Filter Media Modified by Biocidal Nanomaterials to Control Bioaerosols in Internal Environments. *Water, Air,* and *Soil Pollution*, *232*(5), 176. https://doi.org/10.1007/s11270-021-05105-3

Dou, S., and Xia, J. (2019). Analysis of Sheet Metal Forming (Stamping Process): A Study of the Variable Friction Coefficient on 5052 Aluminum Alloy. *Metals*, *9*(8), 853. https://doi.org/10.3390/met9080853

ERGÜL, E., and KURT, H. (2021). Comparison of Properties by Applying ANFIS, ANN and Taguchi Approaches to Al Matrix Composites. *Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi*, *13*(2), 406–416. https://doi.org/10.29137/umagd.833300

Escaich, C., Shi, Z., Baron, L., and Balazinski, M. (2020). Machining of Titanium Metal Matrix Composites: Progress Overview. *Materials*, *13*(21), 5011. https://doi.org/10.3390/ma13215011

Fang-cheng, L., Mei-ying, Z., Jing-xuan, S., Xiu-quan, L., Hao-ou, R., Qing, X., Sheng-hui, W., Sheng-dong, Y., and Xiao-bin, C. (2022). Effect of <scp> TiO $_2$ </scp> modified with dopamine and Silane coupling agent on the electrical properties of <scp>PMIA</scp> insulation paper. *Polymer Engineering* and *Science*, *62*(9), 2998–3009. https://doi.org/10.1002/pen.26080

Gallah, H., Mighri, F., Ajji, A., and Bandyopadhyay, J. (2020). Flexible electrospun PET/TiO 2 nanofibrous structures: Morphology, thermal and mechanical properties. *Polymers for Advanced Technologies*, *31*(7), 1612–1623. https://doi.org/10.1002/pat.4890

Ghamarpoor, R., Fallah, A., and Jamshidi, M. (2023). Investigating the use of titanium dioxide (TiO2) nanoparticles on the amount of protection against UV irradiation. *Scientific Reports*, *13*(1), 9793. https://doi.org/10.1038/s41598-023-37057-5

Gulzar, M., Masjuki, H. H., Kalam, M. A., Varman, M., Zulkifli, N. W. M., Mufti, R. A., Zahid, R., and Yunus, R. (2017). Dispersion Stability and Tribological Characteristics of TiO 2 /SiO 2 Nanocomposite-Enriched Biobased Lubricant. *Tribology Transactions*, *60*(4), 670–680. https://doi.org/10.1080/10402004.2016.1202366

He, Q., Li, A., Wang, Z., Zhang, Y., Kong, L., and Yang, K. (2018). Tribological behavior of ZnO-Si 3 N 4 nanoparticles-based lubricating grease. *Journal of Experimental Nanoscience*, *13*(1), 231–244. https://doi.org/10.1080/17458080.2018.1511923

Holmberg, K., and Erdemir, A. (2017). Influence of Tribology on Global Energy Consumption, Costs and Emissions. *Friction*. https://doi.org/10.1007/s40544-017-0183-5

Irudayaraj, N., Rajamani, S., Mahalingam, P., Karunakaran, J., M, A. H., and Chandran, D. E. (2023). Comparative Evaluation of Mechanical Properties

and Microleakage of Cention N and Titanium Dioxide Nanoparticles-Enriched Cention N: An In Vitro Study. *Cureus*. https://doi.org/10.7759/cureus.51209

KAM HEPDENİZ, Ö., SEÇKİN KELTEN, Ö., ve GÜRDAL, O. (2019). Cam İyonomer İçerikli Dört Farklı Restoratif Materyalin Yüzey Pürüzlülüklerinin Değerlendirilmesi. *Süleyman Demirel Üniversitesi Sağlık Bilimleri Dergisi*, *10*(1), 13–17. https://doi.org/10.22312/sdusbed.468329

Karunakaran, C., Navamani, P., and Gomathisankar, P. (2020). Tuning the optical, electrical and photocatalytic properties of nanoparticulate TiO 2 through anatase-coating on rutile. *Advances in Natural Sciences: Nanoscience and Nanotechnology*, *11*(2), 025013. https://doi.org/10.1088/2043-6254/ab92fa

KILIÇAY, K. (2020). Plazma Transferli Ark (PTA) Yöntemi ile Üretilen WC Kompozit Kaplamalarda Enerji Girdisinin Mikroyapısal ve Tribolojik Özelliklere Etkisi. *Bilecik Şeyh Edebali Üniversitesi Fen Bilimleri Dergisi*, *7*(1), 251–264. https://doi.org/10.35193/bseufbd.702519

KOCAMAN, R., ve ATEŞ, S. (2023). Al6061 Matrisli SiC Al2O3 ve Kömür Cürufu Tozu Takviyeli Hibrit Kompozitlerin Sertlik ve Aşınma Davranışlarının İncelenmesi. *Uluslararası Muhendislik Arastirma ve Gelistirme Dergisi*. https://doi.org/10.29137/umagd.1287314

Kundu, S., Das, S. K., and Sahoo, P. (2016). Influence of load and temperature on tribological behaviour of electroless Ni-P deposits. *IOP Conference Series: Materials Science and Engineering*, *149*, 012046. https://doi.org/10.1088/1757-899X/149/1/012046

Liang, G., Schmauder, S., Lyu, M., Schneider, Y., Zhang, C., and Han, Y. (2018). An Investigation of the Influence of Initial Roughness on the Friction and Wear Behavior of Ground Surfaces. *Materials*, *11*(2), 237. https://doi.org/10.3390/ma11020237

Ma, L., Ma, L., Ma, X., Zhou, C., Jiang, Z., and Zhao, J. (2023). Tribological Properties and Lubrication Mechanisms of Water-Based Nanolubricants Containing TiO2 Nanoparticles during Micro Rolling of Titanium Foils. *Materials*, *17*(1), 170. https://doi.org/10.3390/ma17010170

Malaki, M., Xu, W., Kasar, A. K., Menezes, P. L., Dieringa, H., Varma, R. S., and Gupta, M. (2019). Advanced Metal Matrix Nanocomposites. *Metals*, *9*(3), 330. https://doi.org/10.3390/met9030330

Nguyen, T. T., Dao, K. A., Luong, X. D., and Ung, T. D. T. (2021). Enhanced Electrochemical Water Splitting Activity Using Annealed TiO2 Nanoparticles As Photoanodes. *Journal of Electronic Materials*, *50*(11), 6459–6466. https://doi.org/10.1007/s11664-021-09175-1

Phan, H. T., and Haes, A. J. (2019). What Does Nanoparticle Stability Mean? *The Journal of Physical Chemistry C*, *123*(27), 16495–16507. https://doi.org/10.1021/acs.jpcc.9b00913

PUL, M., ve BAYDAROĞLU, V. (2020). B4C/SiC Katkılı Alüminyum Esaslı Kompozitlerin Mekanik Özelliklerin İncelenmesi ve Balistik Performanslarının Modellenmesi. *Politeknik Dergisi*, *23*(2), 383–392. https://doi.org/10.2339/politeknik.525011

Qi, X., Jia, Z., and Yang, Y. (2013). Influence of the dispersion of nano titanium dioxide on the tribological performance of fabric self-lubricating liner. *Journal of Applied Polymer Science*, *130*(3), 2100–2105. https://doi.org/10.1002/app.39343

Ram, V. V., Arivarasu, L., Rajeshkumar, S., and Thangavelu, L. (2021). Green

Synthesis and Characterisation of Mucuna Pruriens Mediated Titanium Dioxide Nanoparticles Using Transmission Electron Microscope. *Journal of Pharmaceutical Research International*, 362–370. https://doi.org/10.9734/jpri/2021/v33i62B35624

Rashid, M. M., Forte Tavčer, P., and Tomšič, B. (2021). Influence of Titanium Dioxide Nanoparticles on Human Health and the Environment. *Nanomaterials*, *11*(9), 2354. https://doi.org/10.3390/nano11092354

Ren, J.-Y., Ji, G.-C., Guo, H.-R., Zhou, Y.-M., Tan, X., Zheng, W.-F., Xing, Q., Zhang, J.-Y., Sun, J.-R., Yang, H.-Y., Qiu, F., and Jiang, Q.-C. (2024). Nano-Enhanced Phase Reinforced Magnesium Matrix Composites: A Review of the Matrix, Reinforcement, Interface Design, Properties and Potential Applications. *Materials*, *17*(10), 2454. https://doi.org/10.3390/ma17102454

Saatchian, E., Naseri, S., Hosseini, S., and Gholamhosseinian, H. (2025). Promising application of nano-WO3/epoxy composite in intensity-modulated brachytherapy: a simulation study. *Radiation Oncology Journal*, *43*(1), 22–29. https://doi.org/10.3857/roj.2024.00339

Shivalingam, C., Gurumoorthy, K., Murugan, R., and Ali, S. (2024). Herbal-Based Green Synthesis of TB-ZnO-TiO(II) Nanoparticles Composite From Terminalia bellirica: Characterization, Toxicity Assay, Antioxidant Assay, and Antimicrobial Activity. *Cureus*. https://doi.org/10.7759/cureus.55686

Srinivas, K. S., and Mohan, M. M. (2020). Experimental Investigation of Mechanical Properties of Ceramic Reinforced Al-7075 Metal Matrix Hybrid Composites. *Materials Science Forum*, *979*, 34–39. https://doi.org/10.4028/www.scientific.net/MSF.979.34

Stawarz, S., Witek, N., Kucharczyk, W., Bakar, M., and Stawarz, M. (2019). Thermo-protective properties of polymer composites with nano-titanium dioxide. *International Journal of Mechanics and Materials in Design*, *15*(3), 585–599. https://doi.org/10.1007/s10999-018-9432-7

Tapia-Guerrero, Y. S., Del Prado-Audelo, M. L., Borbolla-Jiménez, F. V., Giraldo Gomez, D. M., García-Aguirre, I., Colín-Castro, C. A., Morales-González, J. A., Leyva-Gómez, G., and Magaña, J. J. (2020). Effect of UV and Gamma Irradiation Sterilization Processes in the Properties of Different Polymeric Nanoparticles for Biomedical Applications. *Materials*, *13*(5), 1090. https://doi.org/10.3390/ma13051090

Tavakolian, A. H., Abdullah, A., and Hoshmand Moini, S. (2024). Synthesis mechanism of TiO 2 nanoparticles via high frequency electrical arc discharge. *Physica Scripta*, *99*(12), 125932. https://doi.org/10.1088/1402-4896/ad8bfa

Thoniyot, P., Tan, M. J., Karim, A. A., Young, D. J., and Loh, X. J. (2015). Nanoparticle–Hydrogel Composites: Concept, Design, and Applications of These Promising, Multi-Functional Materials. *Advanced Science*, *2*(1–2). https://doi.org/10.1002/advs.201400010

TOKER, S. M., BATTAL, E., DEMİR, Z., ve ÇEVİK, K. E. (2020). Mikrodeformasyon ile Yüzey Özellikleri Değiştirilen 316L Paslanmaz Çeliğin Sentetik Vücut Sıvısı ile Etkileşimi. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, *8*(4), 2455–2467. https://doi.org/10.29130/dubited.707023

TOPCU, İ., ve KARAMAN, E. (2019). Çok Duvarlı Karbon Nanotüp Takviyeli Düzenli/ Düzensiz Şekilli Ti-6Al-4V Kompozitlerin Aşınma Davranışlarının İncelenmesi. *Düzce Üniversitesi Bilim ve Teknoloji Dergisi*, *7*(3), 1249–1260. https://doi.org/10.29130/dubited.545434

Torres, H., Rodríguez Ripoll, M., and Prakash, B. (2018). Tribological behaviour of self-lubricating materials at high temperatures. *International Materials Reviews*, *63*(5), 309–340. https://doi.org/10.1080/09506608.2017.1410944

Venugopal, S., and Karikalan, L. (2021). Microstructure and physical properties of hybrid metal matrix composites AA6061-TiO2-SiC via stir casting techniques. *Materials Today: Proceedings*, *37*, 1289–1294. https://doi.org/10.1016/j.matpr.2020.06.462

Verma, P. C., Ciudin, R., Bonfanti, A., Aswath, P., Straffelini, G., and Gialanella, S. (2016). Role of the friction layer in the high-temperature pin-on-disc study of a brake material. *Wear*, *346–347*, 56–65. https://doi.org/10.1016/j.wear.2015.11.004

Villanueva, E., Albizuri, J., Caballero, P., Guraya, T., and Vicario, I. (2024). Prediction of Wear Rate by a New Direct Method Using the Friction Coefficient Curve. *Journal of Manufacturing and Materials Processing*, *9*(1), 6. https://doi.org/10.3390/jmmp9010006

Wu, H., Zhao, J., Luo, L., Huang, S., Wang, L., Zhang, S., Jiao, S., Huang, H., and Jiang, Z. (2018). Performance Evaluation and Lubrication Mechanism of Water-Based Nanolubricants Containing Nano-TiO2 in Hot Steel Rolling. *Lubricants*, *6*(3), 57. https://doi.org/10.3390/lubricants6030057

Wu, H., Zhao, J., Xia, W., Cheng, X., He, A., Yun, J. H., Wang, L., Huang, H., Jiao, S., Huang, L., Zhang, S., and Jiang, Z. (2017). A study of the tribological behaviour of TiO2 nano-additive water-based lubricants. *Tribology International*, *109*, 398–408. https://doi.org/10.1016/j.triboint.2017.01.013

Wu, J., Yi, S., Wang, Y., Yao, J., and Gao, W. (2021). Polymer-based TiO 2 nanocomposite membrane: synthesis and organic pollutant removal. *International Journal of Smart and Nano Materials*, *12*(2), 129–145. https://doi.org/10.1080/19475411.2021.1901792

Xia, W. Z., Zhao, J. W., Wu, H., Jiao, S. H., Zhao, X. M., Zhang, X. M., and Jiang, Z. Y. (2016). A Novel Nano-TiO2 Additive Oil-in-Water Lubricant for Hot Steel Rolling. *Materials Science Forum*, *861*, 201–206. https://doi.org/10.4028/www.scientific.net/MSF.861.201

Yu, Y., Zhang, C., Yang, X., Sun, L., and Bian, F. (2025). Microfluidic Synthesis of Magnetic Nanoparticles for Biomedical Applications. *Small Methods*, *9*(4). https://doi.org/10.1002/smtd.202401220

Zhang, D., Gökce, B., and Barcikowski, S. (2017). Laser Synthesis and Processing of Colloids: Fundamentals and Applications. *Chemical Reviews*, *117*(5), 3990–4103. https://doi.org/10.1021/acs.chemrev.6b00468

ZOR, M., ve ŞEN, F. (2023). Kuarternize Karboksimetil Selüloz Esaslı Farklı Ahşap Kaplamaların Antimikrobiyal ve Yüzey Özelliklerinin İncelenmesi. *Bartın Orman Fakültesi Dergisi*, *25*(2), 247–254. https://doi.org/10.24011/barofd.1291350

Frequency-Adjustable Antenna Design for RF Energy Harvesting Applications

Fettah KINAL[1]
Mahmut Ahmet GOZEL[1]
Mesud KAHRİMAN[1*]
Ozlem COSKUN[1]

[1]Department of Electrical and Electronics Engineering/Suleyman Demirel University, Turkey * (mesudkahriman@sdu.edu.tr)

ABSTRACT

In contemporary wireless communication applications, reconfigurable antennas have attracted considerable attention due to their ability to meet changing requirements in frequency, polarization, and radiation pattern. Designs of reconfigurable antennas are realized using varactor diodes, PIN diodes, radio-frequency microelectromechanical systems (RF-MEMS), FET diodes, and liquids with different dielectric constants. In general, the use of active magnetic control (AMC) changes the antenna capacitance and thus tunes the antenna's resonance frequency. This study is implemented using a varactor diode. Varactors change their capacitance depending on the applied voltage, which facilitates control. The focus here is also on different types of reconfigurable antennas. Numerous parameters such as antenna dimensions and the dielectric material used are taken into account, and the resonance frequencies of the antennas are reported. In the designed antenna, the desired results are obtained for capacitance values between 2.03 pF and 37.35 pF. Using a single varactor stage at the center, rather than three varactors, yields results much closer to the desired outcome. Consequently, it is observed that, in reconfigurable antenna designs, employing a single central varactor stage is more suitable and closer to the desired result compared to a three-varactor structure.

Keywords: RF Energy Harvesting, Frequency-Tunable Antenna, Varactor Diode.

INTRODUCTION

In recent years, there has been a high demand for devices that are low-profile, efficient, compact, high-speed, planar, inexpensive, and easily integrable. These devices have various applications in different roles within the modern telecommunications sector. The antenna is one of the most crucial parts of evolving communication systems and plays an increasingly important role in modern electronic devices. They are mostly used for applications that require fixed functional characteristics such as polarization, beamwidth, and far-field properties. However, antenna systems with such fixed functional features may increase circuit complexity and reduce efficiency. In addition, integrating multiple radio communications and antenna systems into a single device may also degrade system performance [1,2].

Frequency-reconfigurable antennas focus on frequency tuning within the desired frequency bands and on efficient spectrum utilization. Pattern-reconfigurable antennas aim to steer signals in the desired direction, save energy, improve channel capacity, increase gain, and reduce in-band interference [3–6]. Reconfiguration of an antenna is achieved by altering its frequency, polarization, or radiation properties. This alteration can be

accomplished through various techniques that redistribute antenna currents, thereby modifying the electromagnetic fields across the antenna's effective aperture. Reconfigurable antennas can meet complex system requirements by modifying their geometries and electrical behaviour, thereby adapting to changes in environmental conditions or system requirements [7].

There are four main techniques to achieve tunability in antennas: electrical, optical, mechanical, and material-based modifications [8]. In the literature, different techniques have been employed—using PIN diodes, varactor diodes, capacitive loading, RF-MEMS, various smart materials, and optical switches—to design reconfigurable antennas, and reconfigurable/tunable antenna designs have been proposed [9]. When a PIN diode is used in antenna design, the operating frequencies of these antennas can be easily changed by controlling the on/off states of the P-N junction diodes. Alternatively, using a varactor diode enables operation at multiple resonance frequencies with a single varactor [10].

LITERATURE REVIEW

The frequency-reconfigurable antenna designed by Majid et al. operates in the 1.98–3.59 GHz range using five PIN diodes. This antenna was designed with Taconic RF35 substrate and a surface-mount RF capacitor. The five switches (PIN diodes) used in the antenna design allow nine different states via different combinations. Among these states, there is one in which all diodes are on and the patch is smoothly shorted to ground, resulting in operation at 3.5 GHz. In another state, when all diodes are off, the patch slot acts as a feed network and resonates at 1.98 GHz. In these nine different states, the return loss is less than -10 dB. In the design, eight different resonance frequency bands can be obtained by changing the effective length of the slots. In simulations and measurements, directional radiation is highest at frequency F1, bidirectional radiation is lowest at frequencies F2–F4, and omnidirectional radiation is obtained at frequencies F5–F9. The gain of the designed antenna varies between 4.8 and 0.2 dBi. In the proposed values, a directional gain of 4.8 dBi is achieved at F1, while an omnidirectional gain of 0.2 dBi is achieved at F9. This study represents a significant investigation into frequency-reconfigurable antenna design and the gain values obtained with different operating modes [11].

The antenna design developed by Khan et al. enables frequency reconfiguration through the use of angles and PIN diodes. The antenna is divided into two by the feed transmission line and equipped with PIN diodes. Using a total of eight PIN diodes, the antenna operates over a wide

frequency range of 2.43–3.3 GHz. The antenna material has a dielectric constant of 4.5 and a loss tangent of 0.002. These properties are important parameters affecting antenna performance. The return loss being better than 10 dB in both cases indicates high efficiency and low losses. The measured gain is 4.5 dBi when the PIN diodes are on and 5.5 dBi when they are off. This study highlights the advantages achieved with frequency-reconfigurable antenna design using PIN diodes. The wide frequency range, low losses, and high gains show that this antenna can be successfully used in various applications [12].

In the study by Alhamad et al., frequency-reconfigurable antennas were examined for designs suitable for 4G, 4G/5G, and 5G frequencies. These antenna designs have tuneable features using multiple PIN and varactor diodes to support the 5G spectrum. In one antenna design implemented using varactors, both N-type and P-type varactors were used. A varactor functions as a component with voltage-dependent capacitance. The varactor model SMV2019-040LF, with a capacitance varying between 0.23–1.43 pF, allows antenna operation at seven different frequencies. For example, when C = 1.43 pF, the S11 value is -15.17393 dB at 2.888 GHz, whereas when C = 0.23 pF, the S11 is -17.87984 dB at 4.584 GHz.

In another antenna design implemented using PIN diodes; an FR-4 substrate was used. The design included PIN diodes of model HSMP-3860. For modelling the PIN diodes, a reverse bias voltage of -1 V was used, and under this condition, the PIN diode was considered off. Simulation results supported selecting the PIN diode to be modelled in the off state [13].

In the study conducted by Ahmad et al., a frequency-tuneable antenna design is presented. The design starts with a truncated-corner square patch, and then 6 mm-wide slots are added on all four sides of the antenna. These slots create an inductive effect and reduce the resonance frequency from 3.05 GHz to 2.41 GHz. A 1 mm-wide slot is opened symmetrically at the center of the antenna, and after the current path is interrupted, the resonance frequency is reduced to 2.04 GHz. Subsequently, the resonance frequency is further lowered from 2.04 GHz to 1.21 GHz.

A feature of this design is the addition of a parallel inductor near the PIN and the probe to increase the return loss and reduce reflection by bringing the input impedance closer to 50 ohms. Then, a varactor is placed in the 1 mm central slot region. When a DC bias voltage is applied, the varactor changes its capacitance to manipulate the connection impedance and phase relationships between the upper and lower slots. FR-4 is used as the dielectric material. The varactor model is SMV123. When DC forward bias is applied, the capacitance is 9.63 pF and the frequency is 1.010 GHz, whereas at 7.5 V, the capacitance is 1.72 pF and the frequency is 1.160 GHz. Due to impedance mismatch at zero bias, the return loss is better than -14 dB. In addition, the losses are better than -6 dB across all frequency bands.

This design demonstrates effective performance over a wide frequency range, minimizing return loss and enhancing antenna reliability [14].
In the study by Zammit et al., a frequency-tuneable antenna was realized using shorting posts and varactors. The antenna dimensions were set to 80 mm × 40 mm, and a 20 mm × 22 mm antenna was equipped with two shorting posts, each with a radius of 0.45 mm. This arrangement ensures that the antenna resonates at 1.9 GHz. In particular, the current density around the shorting posts was arranged to be controlled via a capacitor. In this way, the resonance frequency of the antenna is reduced and can be tuned according to the desired antenna frequency range. A Philips BB184 varactor diode was placed in front of each shorting post. These varactors are tuneable between 2 pF (10 V) and 14 pF (1 V). The parasitic inductance and resistance of the varactor were determined as 0.6 nH and 0.65 Ω, respectively. The designed antenna was simulated with CST between 0 V and 10 V. In the system, the shorting post was shorted to the ground plane of the antenna. Slots of 1 mm were opened to feed the varactor diodes, and 100 pF capacitors were placed. Three capacitors were sufficient for this arrangement. The capacitors were aligned with the probe feed and the shorting posts. Rogers 4003 was chosen as the substrate material, with εr = 3.38 and thickness of 1.524 mm. The designed antenna operates in the 0.55–1.5 GHz range. An HP8510 network analyser was used to measure the return loss [15].

FREQUENCY-ADJUSTABLE ANTENNA DESIGN AND SIMULATION

The configuration and design parameters of the proposed antenna are presented in Figure 1. In this study, FR-4 with a relative dielectric constant (εr = 4.3) and a loss tangent of 0.019 was selected as a suitable substrate material. It is fed by a 50 ohm compatible microstrip line.

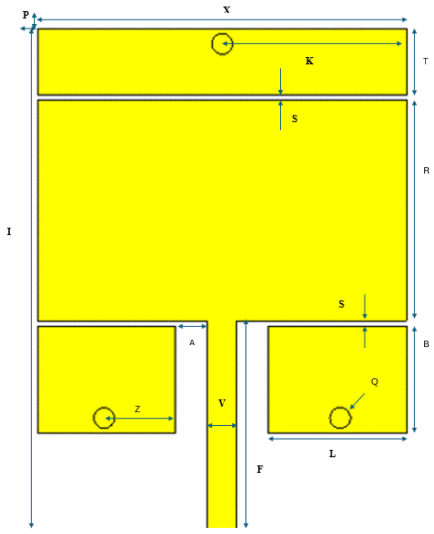


Figure 1. Geometry of the designed antenna

The dimensions of the antenna are shown in Table 1. Electrical dimension and gain values were examined during the design, and as the most suitable result, the FR4 dimension on the back surface of the antenna was kept shorter than the grounding surface.

Table 1. Antenna dimentions

Parameters		**Parameters**	
X	36 mm	P	1 mm
K	18 mm	T	6.5 mm
S	0.5 mm	R	21.5 mm
I	49 mm	B	10.5 mm
Z	6.73 mm	A	3.1 mm
F	20.5 mm	L	13.46 mm
Q	1 mm	V	2.87 mm

Subsequently, the antenna structure was exported from CST in DWG format, and the copper regions were filled in black in AutoCAD. In the next step, the antenna design was printed at a 1:1 scale on greaseproof paper using a laser printer and cut to exact size as shown in Figure 2. Using a laminator, the design was transferred onto the FR-4 board with the aid of heat and pressure. After laminating, the ground plane was completely covered with plastic tape and the board was immersed in an etchant. The areas without print and tape were etched away by the acid.

Figure 2. The designed antenna after printing

Next, an SMA connector was attached as shown in Figure 3, and the measurement stage commenced.

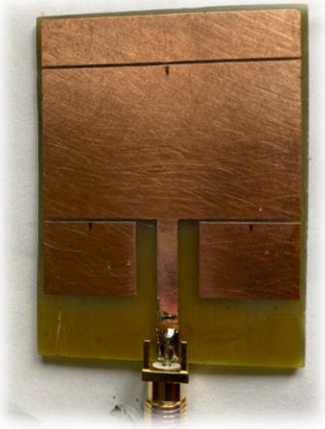


Figure 3. SMA connector

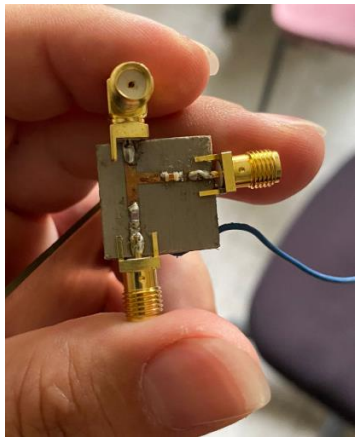


Figure 4. Bias-T

A Bias-T allows the application of both DC voltage and an RF signal to the antenna. The operating principle of the Bias-T is based on placing an inductor in the DC path and a capacitor in the RF path. In this arrangement, the inductor in the DC path prevents the RF signal from reaching the power supply, while the capacitor prevents the DC voltage from reaching the spectrum analyser (Figure 4). The measurement setup is given in Figure 5. The measurement results were plotted with the help of MATLAB.

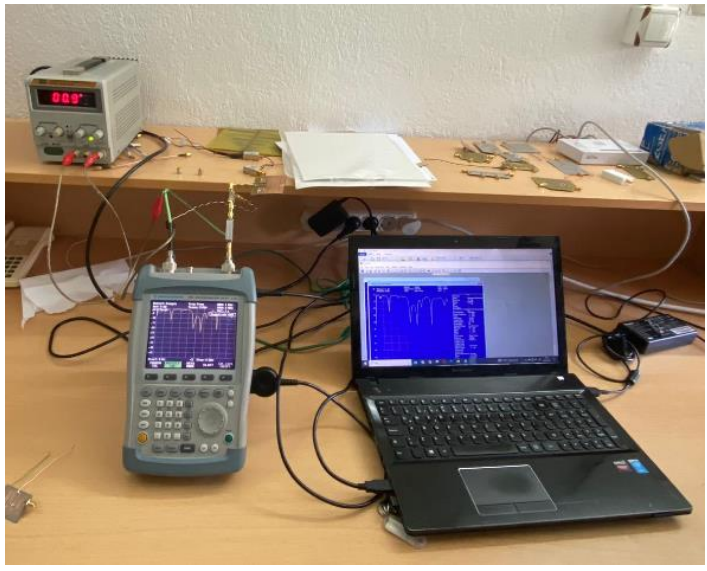


Figure 5. Prepared measurement setup

RESULTS

First, measurements were taken without grounding the parts separated from the patch by opening a slot, and then with grounding (Figure 6).

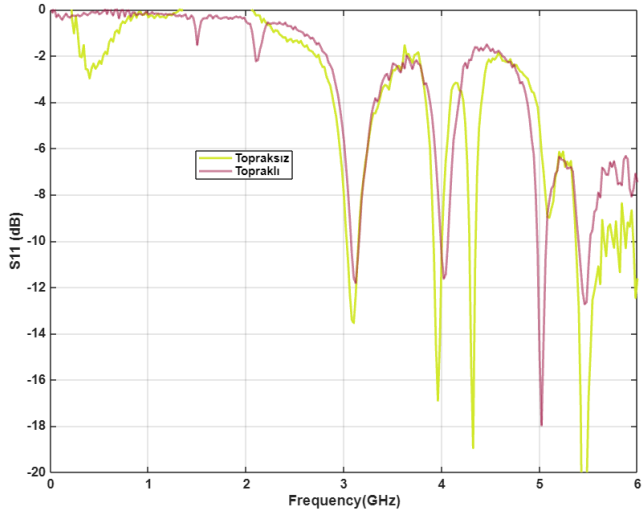


Figure 6. S_{11} measurement result without and with ground connection

The S_{11} graph obtained using a single varactor, and the variation of a single central varactor between 2.03 pF and 37.35 pF are shown in Figures 7 and 8.

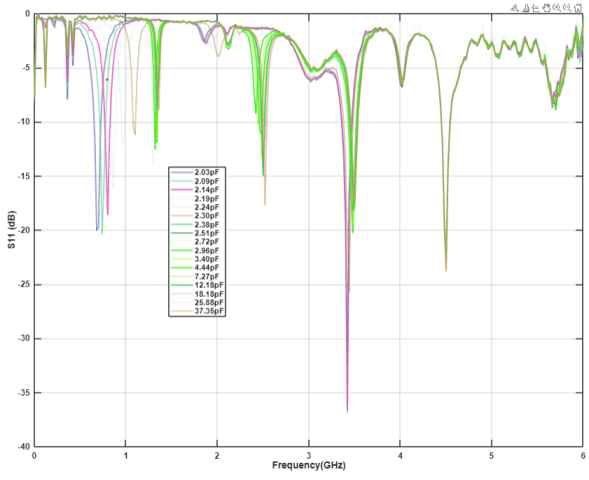


Figure 7. S_{11} graph for a single central varactor

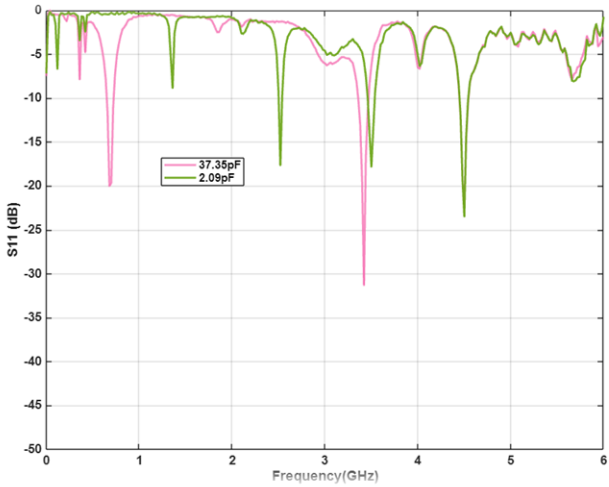


Figure 8. Variation of a single central varactor between 2.03 pF and 37.35 pF

The S_{11} graph obtained using three varactors, and the variation of three varactors between 2.03 pF and 37.35 pF are shown in Figures 9 and 10.

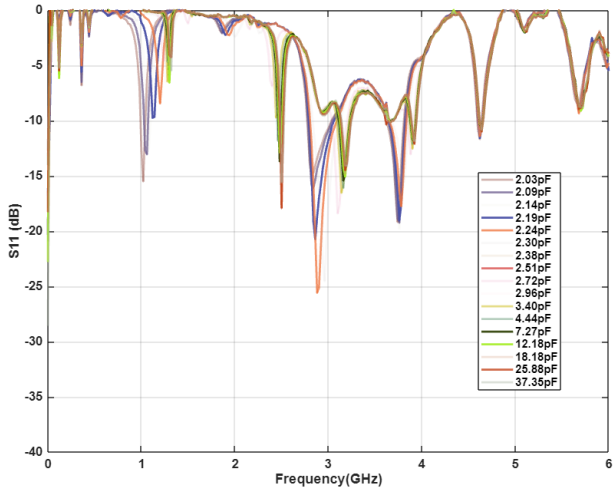


Figure 9. S_{11} graph with three varactors

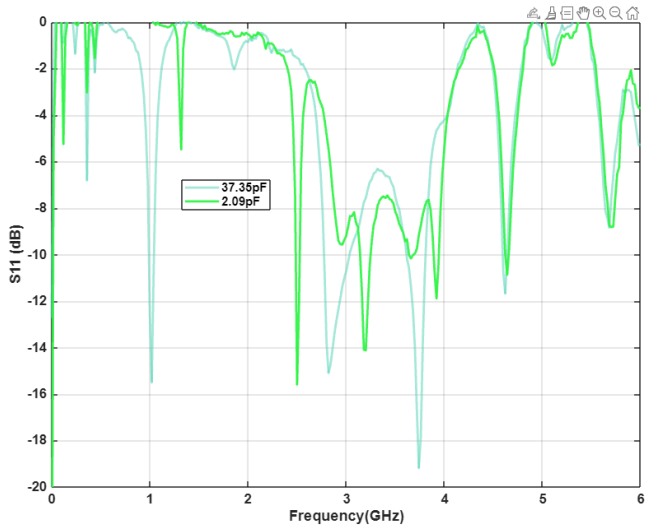


Figure 10. Variation of three varactors between 2.03 pF and 37.35 pF

CONCLUSIONS

In recent years, due to the realization of multi-functional requirements in wireless and satellite communication systems, there has been a shift toward frequency-reconfigurable antennas. This enables cost and space savings by using only one antenna capable of operating in multiple bands, instead of many antennas. This study is implemented using a varactor diode. In an antenna employing a varactor diode, supplying the varactor with a rectenna can make control easier than manual adjustment and enables remote control. Varactors stand out as key components offering a wide range of applications in high-frequency circuit design. In this study, a new design was realized using three varactors. The aim was to shift a center frequency by approximately 1.5 GHz with a capacitance varying between 2.03 pF and 37.35 pF. As a result, it was observed that employing a single central varactor stage in reconfigurable antenna designs is more suitable and closer to the desired outcome than a three-varactor structure. Considering its dimensions and radiation characteristics, the proposed antenna is thought to be usable in many mm-wave applications, particularly RF energy harvesting.

REFERENCES

[1] A. Iqbal, J. J. Tiang, S. K. Wong, M. Alibakhshikenari, F. Falcone, and E. Limiti, Miniaturization trends in substrate integrated waveguide (SIW) filters: A review, *IEEE Access*, vol. 8, pp. 223287–223305, 2020.

[2] A. Ghaffar, A. Altaf, A. Aneja, X. Jun Li, S. Khan, M. Alibakhshikenari, F. Flalcone, and E. Limiti, A frequency reconfigurable compact planar inverted-F antenna for portable devices. *Hindawi International Journal of Antennas and Propagation*, pp.1-9, 2022.

[3] I. F. Akyildiz, W. Y. Lee, M. C. Vuran and S. Mohanty, NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey. *Computer Networks,* 50 (3), pp.2127-2159, 2006.

[4] J. Dong, Y. Li and B. Zhang, A survey on radiation pattern reconfigurable antennas. *7th International Conference on Wireless Communications, Networking and Mobile Computing*, pp.1-4, 2011.

[5] M. Borhani, P. Rezaei, A. Valizade, Design of a reconfigurable miniaturized microstrip antenna for switchable multiband systems. *IEEE Antennas and Wireless Propagation Letters*, 15, pp. 822-825, 2016.

[6] S. W. Lee and Y. Sung, Compact frequency reconfigurable antenna for LTE/WWAN mobile handset applications. *IEEE Transactions on Antennas and Propagation*, 63(10), pp. 4572–4577, 2015.

[7] C. G. Christodoulou, Y. Tawk, S. *A. Lane and S. R. Erwin, Reconfigurable antennas for wireless and space applications. Proceedings of the IEEE*, 100 (7), pp. 2250-2261, 2012.

[8] Y. Costantine, G. Tawk, and C. G. Christodoulou, Design of reconfigurable antennas using graph models. *Synthesis Lectures on Antennas Springer,* pp.148, 2013.

[9] J. H. Lim, C. W. Song, Z. J. Jin and T. Y. Yun, Frequency reconfigurable planar inverted-F antenna using switchable radiator and capacitive load. *IET Microwaves Antennas Propagation*,7(6), pp.430-435, 2013.

[10] S. Yang, Y. Chen, C. Yu, Y. Gong, and F. Tong, Design of a low-profile, frequency-reconfigurable, and high gain antenna using a varactor-loaded AMC ground, *IEEE Access,* vol. 8, pp. 158635-158646, 2020.

[11] H. A. Majid, M. K. A. Rahim, M. R. Hamid, N. A. Murad, and M. F. Ismail, Frequency-reconfigurable microstrip patch-slot antenna, *IEEE Antennas and Wireless Propagation Letters,* vol. 12, pp. 218-220, 2013.

[12] M. S. Khan, A. Iftikhar, A. D. Capobianco, R. M. Shubair, and B. Ijaz, Pattern and frequency reconfiguration of patch antenna using PIN diodes, *Microwave and Optical Technology Letters,* vol. 59, no. 9, pp. 2180-2185, 2017.

[13] R. Alhamad, E. Almajali, and S. Mahmoud, Electrical reconfigurability in modern 4G, 4G/5G and 5G antennas: A critical review of polarization and frequency reconfigurable designs, *IEEE Access,* 2023.

[14] M. Al Ahmad, S. Kabeer, A. A. Sanad, and L. J. Olule, Compact single-varactor diode frequency-reconfigurable microstrip patch antenna, *IET Microwaves, Antennas & Propagation,* vol. 15, no. 9, pp. 1100-1107, 2021.

[15] J. A. Zammit and A. Muscat, A small tunable antenna using multiple shorting posts and varactor diodes, in *2008 3rd International Symposium on Communications, Control and Signal Processing*, IEEE, pp. 83-86, 2008.

Accelerated Pavement Testing on Concrete Pavement

Muhammet Vefa AKPINAR[1]

1- Prof.Dr.Muhammet Vefa Akpınar; Karadeniz Technical University, Civil Engineering Department. mvakpinar@ktu.edu.tr ORCID No: 0000-0001-7912-8274

ABSTRACT

This study compares Accelerated Pavement Testing (APT) systems, a device used to predict the road performance and service life of highway pavements. The impact of real traffic and thermal loads on concrete roads was studied using the Accelerated Pavement Test (APT) facility within the Transportation Department of the Civil Engineering Faculty at Karadeniz Technical University (KTU). This testing process allowed the effects typically observed over 20 years on unreinforced concrete roads with joints to be simulated 8 months. The results helped to analyze and interpret vertical deformations and cracks that develop over time on concrete pavements due to thermal changes.

Keywords – Concrete, Pavement, Accelerated Pavement, Testing, Transportation.

INTRODUCTION

The performance and eventual failure of a pavement system from the cumulative damage inflicted by repeated heavy traffic loads across all pavement layers. Evaluating the lifespan of a pavement cannot be achieved by analyzing the performance of a single layer or material in isolation. Accelerated Pavement Testing (APT) remains the most reliable method for realistically assessing the long-term performance and overall lifespan of the entire pavement structure.(1-10)

Accelerated Pavement Testing, APT, testing are performed on a test track using real vehicles or via specialized equipment equipped with a load frame that applies moving wheel loads in either a linear or circular pattern. These load frames can be placed over an actual pavement section or, more commonly, installed over pavement sections built within dedicated test facilities. The key benefit of these specialist facilities is their ability to maintain controlled conditions, effectively reducing the influence of external variables.

The first engineering applications of Accelerated Pavement Testing (APT) were carried out in the late 1950s by AASHTO, established in the USA. These tests in the USA aimed to determine the vertical deformations of data stored under the heavy traffic loads of highways and airport roads, and to design roads accordingly. This was generally done on trial roads built with big government subsidies. Later, with the contribution of the USA, these studies began in Australia, France, and England. Beyond trial roads, this research began in laboratory settings in the 1980s. The USA, taking the first step in this field, made significant investments in APT. These APT systems can be tested in the field (Table 1 and Figure 1) or in laboratories specifically designed for these tests.

Table 1. Accelerated Pavement Testing facilities.

Research center	Name	Institution
Advanced Transportation Research and Engineering Laboratory (ATREL); Rantoul, IL	ATREL	University of Illinois at Urbana-Champaign
Caltrans Accelerated Pavement Testing (CAL-APT) Heavy-Vehicle-Simulator (HVS) Program; Richmond, CA	CAL-APT	California Department of Transportation
U.S. Army Cold Regions Research and Engineering Laboratory (CRREL), Frost Effects Research Facility; Hanover, NH	CRREL-HVS	Cold Regions Research and Engineering Laboratory
FHWA Pavement Test Facility (PTF); McLean, VA	FHWA-PTF	Federal Highway Administration
Florida-Accelerated Pavement Testing and Research Facilitya (APTRF); Gainesville, FL	FL-APTRF	Florida Department of Transportation
Indiana DOT/Purdue APT Facility; West Lafayette, IN	INDOT/Purdue	Indiana Department of Transportation
Kansas-Accelerated Pavement Testing (APT); Manhattan, KS	KS-APT	Kansas State University
Louisiana Transportation Research Center (LTRC) Pavement Research Facility (PRF); Port Allen, LA	LTRC-PRF	Louisiana Transportation Research Center
Minnesota Road Research Project (Mn/ROAD); Minneapolis, MN	Mn/ROAD	Minnesota Department of Transportation
National Center for Asphalt Technology (NCAT) Pavement Test Track (PTT); Auburn University, AL	NCAT-PTT	National Center for Asphalt Technology
Ohio-Accelerated Pavement Loading Facility (APLF); Lancaster, OH	OH-APLF	Ohio University/Ohio State University
Penn State (PS) Pavement Durability Facilityb (PDF); College Park, PA	PS-PDF	Pennsylvania State University
Texas Mobile Load Simulator (TxMLS); Austin, TX	TxMLS	Texas Department of Transportation

U.S. Army Corps of Engineers Waterways Experiment Station (WES) HVS; Vicksburg, MS	WES-HVS	Engineers Research and Development Center
WesTrackc; Reno, NV	WesTrack	Nevada Automotive Test Center

METHOD

Developed in 2009, the Full-Scale Mobile Loading Simulator KTU-APT offers numerous benefits, including ground water level table changes. Its trafficking system is uni-directional, replicating real heavy 40 ton truck loading conditions to provide reliable data for theoretical modeling. The maximum load capacity per set of dual wheels on the MLS full-scale machines is 7.5 tons (75 kN), operating at a top speed of 6 m/s (22 km/h). This allows for the application of 6,000 axle loads per hour (Figure 1 and 2). The device also supports lateral movement to simulate wander during operation. Powered by contactless linear induction motors (LIMs), the system eliminates the need for a gearbox, resulting in a drive system free from wear and fatigue.

C25 class concrete, as specified in the tender contracts, was used. This project included a reference section of a jointed unreinforced concrete road, incorporating various proportions of polypropylene fibers. The road will feature a yellow concrete coating measuring 5 m x 15 m x 0.20 m, placed over 95-98% compacted base layer soil. Joint cuts were made at 4-meter intervals with a depth of 5 cm (h/4).

Figure 1. APT torque applied on the concrete surface.

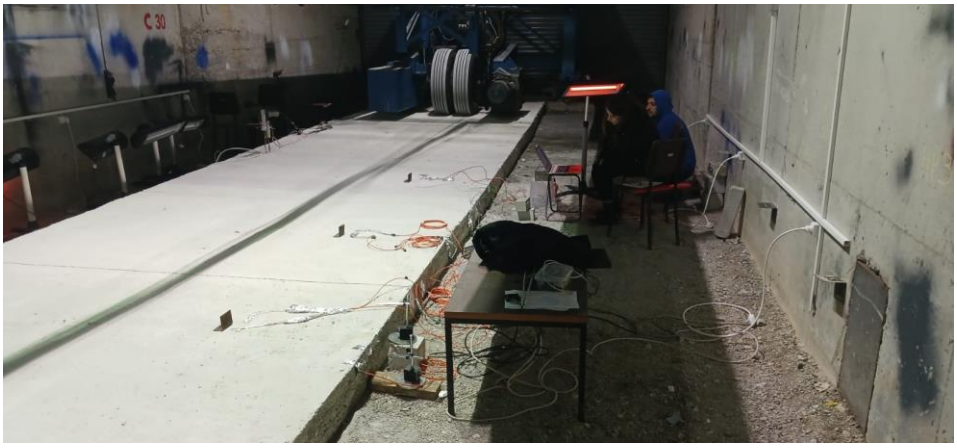


Figure 2. APT Facility. Asphalt over concrete overlay.

RESULTS AND DISCUSSION

One of the most important features that distinguishes the APT system from other systems is its movement mechanism. In many systems worldwide, since cable systems are used, the drive torque is not obtained in the experiments (Figure 3). The loads carried in the system are tested like the idler axle of a truck. However, in order to obtain data close to reality, real tire-road adhesion must be achieved. In the MVA APT system, the 30 kW electric motor is directly connected to the axle where the test will be performed with a special flanged mechanism. In this way, the movement of the system is provided by this axle. In this way, values close to the adhesion and deformation that can occur between a tire connected to the axle of the differential of a real truck and the road platform are obtained in the system.

Thermal cracking takes place when there are significant temperature variations (top and bottom) within a concrete pavement structure. Such temperature differences cause the cooler sections to contract more than the warmer ones, leading to a restriction in contraction. When this restriction generates tensile stresses that surpass the concrete's tensile strength, thermal cracks form. It's important to note that temperature-induced cracking can also occur in concrete components that are not classified as mass concrete.

When cementitious materials hydrate, heat is generated for several days following the placement of concrete. The surface experiences tensile stresses, potentially leading to cracking caused by the uncontrolled temperature difference across the concrete pavement section. Thermal cracking generally occurs at early stages (1 year) but, in rarer cases, can happen when concrete surfaces are suddenly exposed to extreme temperature fluctuations (less than 4 hours).

Over time, concrete members expand when subjected to heat and contract when exposed to cold temperatures. If such bulk volume changes caused by temperature variations are restricted, cracking may even occur at joints (Figure 4) . This type of cracking, often referred to as temperature cracking, typically develops later and presents a long-term concern. Mid-Plate carcking can also develope in the middle of the plate due low thermal changes.

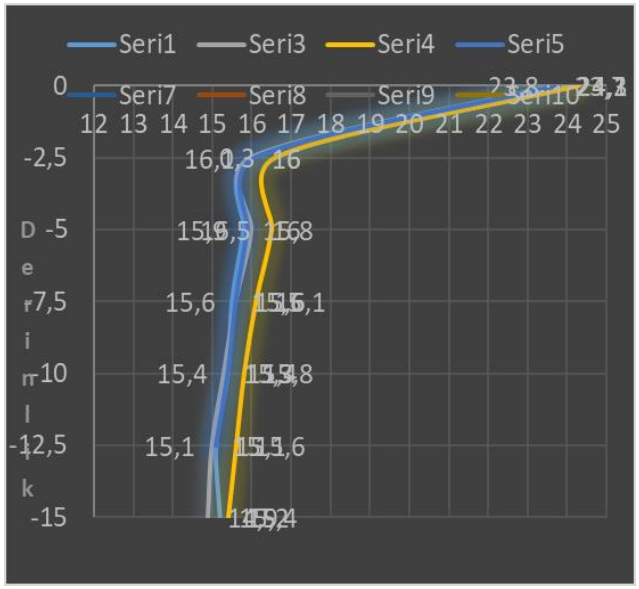


Figure 3. Thermal changes throught the 15 cm thickness of concrete pavement

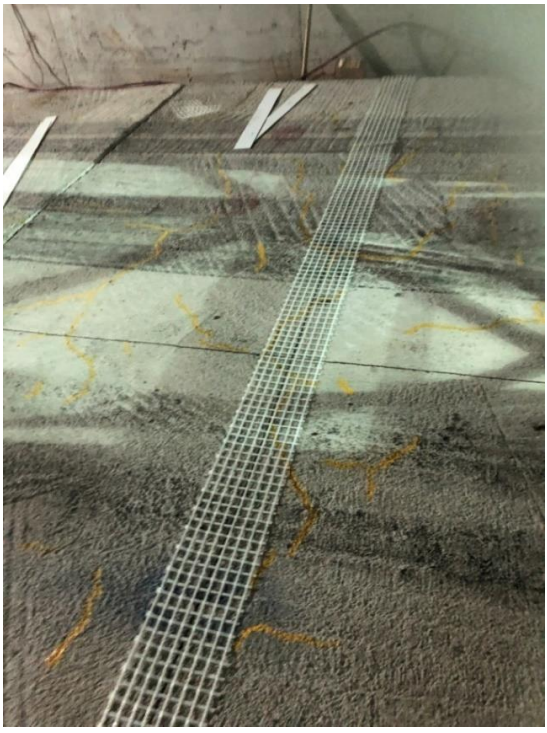


Figure 4. Concrete pavement jont reinforcement.

CONCLUSION

APT, facility will allow for a lower budget allocation for repairing deformations in the road platform, facilitating the construction of new roads. The common goal of all APT systems with different technical specifications is to determine the estimated lifespan of roads to be used, to detect potential damage in advance, to provide civil engineers with the opportunity to develop new solutions, and most importantly, to produce answers to pavement road problems that would otherwise take a very long time to solve, in relatively short periods, based on tests and experiments. Furthermore, this study is important for other researchers as it forms the cornerstone of the infrastructure for Turkey's AASHTO mechanistic empirical studies.

REFERENCES

American Association of State Highway and Transportation Officials (AASHTO). 1993. Guide for Design of Pavement Structures. American Association of State Highway and Transportation Officials. Washington, DC.

American Concrete Institute (ACI). 2002. Guide for Design of Jointed Concrete Pavements for Streets and Local Roads. ACI 325.12R-02. Report by ACI Committee 325. American Concrete Institute. Farmington Hills, MI.

American Concrete Pavement Association (ACPA). 1992. Design and Construction of Joints forConcrete Streets. IS061P. American Concrete Pavement Association. Skokie, IL.

Federal Highway Administration (FHWA). 2007a. Best Practices for Dowel Placement Tolerances. FHWA-HIF-07-021. Federal Highway Administration. Washington, DC.

FHWA (2014). Glossary - Distress Identification Manual for The Long-Term Pavement Performance Program (Fifth Revised Edition) , May 2014 - FHWA-HRT-13-092.

Mallick, R. B. and El-Korchi, T. (2018). Pavement Engineering Principles and Practice. New York: CRC Press.

Bustos, M., de Solminhac, H.E., Darter, M.I., Caroca, A., and Covarrubias, J.P., "Calibration of Performance Models for Jointed Concrete Pavements Using Long-Term Pavement Performance Database," Transportation Research Record 1629, TRB, pp. 108-116, Washington, DC, 1998.

Moody, E.D., "Transverse Cracking Distress in Long-Term Pavement Performance Jointed Concrete Pavement Sections," Transportation Research Record 1629, TRB, pp. 6-12, Washington, DC, 1998.

Snyder, M. B. 2017. Design and Construction of Highway Pavement Joint Systems. Dowel and Tie Bar Design Considerations. Part 1. American Concrete Pavement Association, Rosemont, IL.

Tayabji, Shiraz D. 1986. “Dowel Placement Tolerances for Concrete Pavements.” Transportation Research Record 1062. Transportation Research Board, Washington, DC.

http://www.aapaq.org/documents/ST/5_AAPA_2010_Study_Tour_APT_v5x6.pdf
http://www.utexas.edu/research/ctr/pdf_reports/1814_S.pdf
http://www.mrr.dot.state.mn.us/research/apt/misc/numertoc.pdf

Assessment of Indoor Particulate Matter Concentrations in Supermarkets and Fitness Centers

Özlem ÖZDEN ÜZMEZ[1]
Sıray SAĞDIÇ[2]

1- Assoc. Prof. Dr.; Eskişehir Technical University Engineering Faculty Environmental Engineering Department. oozden@eskisehir.edu.tr ORCID No: 0000-0003-4310-788X
2- Environmental Engineer; Eskişehir Technical University Engineering Faculty Environmental Engineering Department. siraysagdic222@gmail.com ORCID No: 0009-0005-8244-9040

ABSTRACT

This study aims to evaluate indoor air quality (IAQ) in enclosed environments characterized by differing occupancy levels and activity patterns, specifically fitness centers and supermarkets. The research focuses on determining the influence of human activity intensity and foot traffic on particulate matter (PM) levels, with particular emphasis on PM_{10} concentrations. These indoor settings contain notable emission sources that contribute to particulate pollution, underscoring the importance of assessing IAQ within such environments.

Simultaneous indoor and outdoor measurements were conducted during two distinct occupancy periods—low and peak—in both facility types. Data collection was performed using two DustTrak™ II Aerosol Monitors. The findings reveal that indoor PM_{10} concentrations in the fitness center were markedly higher than those measured in the supermarket, indicating that physical activities constitute a more substantial source of particulate emissions compared to typical supermarket operations. Furthermore, PM_{10} concentrations increased during peak occupancy periods in both environments, demonstrating a direct relationship between activity level and particulate matter accumulation.

Indoor/outdoor (I/O) ratios exceeding 1 suggest that particulate concentrations predominantly originate from indoor sources. In the fitness center, I/O ratios were greater than 1 during both occupancy periods, whereas in the supermarket, I/O ratios remained below 1 regardless of occupancy level. These results collectively emphasize that activity intensity and environment type play decisive roles in shaping indoor particulate pollution levels.

Keywords – Air Quality, Indoor Air Pollution, Supermarket, Fitness Center, PM_{10}.

INTRODUCTION

In the last decade, as a reflection of lifestyle changes and increasing levels of urbanization, interest in air quality control has shifted from outdoor to indoor environments. It is known that people spend approximately 90% of their daily lives indoors. Scientific findings demonstrate that inadequate indoor air quality (IAQ) adversely affects human health, and therefore, monitoring and controlling IAQ is of critical importance for public health.

As a result of human activities, various particulate matter (PM) is generated indoors, including dust accumulated on floors and other surfaces, tracked in dirt, textile fibers, skin flakes and hair debris, as well as particles originating from household furniture (Morawska et al., 2013). These dust particles can be resuspended into the air due to mechanical interactions such

as cleaning, turbulence caused by air conditioning and heating systems, and human movement. Particles entering indoor spaces from outdoors may also resuspend into the air after settling on interior surfaces.

Indoor air quality (IAQ) is shaped both by indoor sources and by urban air pollution, highlighting the importance of examining indoor-to-outdoor pollutant concentration ratios (I/O) within living spaces and environmental conditions. Furthermore, IAQ is related to outdoor air quality (OAQ). However, the actual strength of this relationship varies considerably due to existing factors such as building characteristics, ventilation rates, pollutant type, presence of indoor sources, heating, ventilation and air conditioning (HVAC) filtration, door and window usage, seasonal periods, and other local urban microenvironments within buildings. Therefore, poor IAQ can be held responsible for respiratory diseases, allergies, and even cancer (Isinkaralar et al., 2025; WHO, 2021).

In this study, simultaneous indoor and outdoor PM_{10} measurements were conducted in selected indoor microenvironments, including supermarkets and fitness centers, where occupants typically spend extended periods of time. Measurement campaigns were carried out during two distinct occupancy conditions—high-traffic and low-traffic periods—to evaluate the influence of human activity intensity on indoor air quality. This approach enabled the characterization of PM_{10} exposure levels associated with typical indoor occupancy patterns. Furthermore, indoor-to-outdoor (I/O) ratios were derived from the paired PM_{10} datasets to assess the extent to which outdoor particulate pollution contributes to indoor concentrations. Through this analysis, the relative roles of indoor activity and outdoor air quality in shaping the indoor particulate pollution profile were systematically determined.

INDOOR AIR POLLUTION

Enclosed spaces are defined as environments where people spend most of their time, such as homes, schools, commercial and administrative office buildings, and government buildings, excluding industrial facilities. Indoor air pollution began to attract attention in the 1970s following the oil crisis, during which energy restrictions were implemented. As gasoline prices and the cost of energy increased, energy conservation measures were prioritized; building ventilation was reduced, and constructions were designed to be airtight. In the same period, synthetic products and plastics—capable of dispersing and accumulating in indoor air—began to replace natural materials.

The widespread use of computers, printers, and other electronic devices also increased the heat and electromagnetic radiation load within buildings (Zencirci & Işıklı, 2017).

The impacts of indoor air quality (IAQ) on human health have been receiving increasing attention, as pollutants emitted from numerous sources into indoor air can lead to both acute and chronic health issues. These diseases often go unnoticed by individuals, resulting in various problems. In this context, in order for aesthetically appealing and high-rise buildings to meet indoor air quality standards, certain requirements must be fulfilled. Otherwise, they will not only indicate poor and unhealthy living conditions but also be signs of potential illnesses.

This condition has been referred to as "Sick Building Syndrome (SBS)." Research has shown that occupants of such buildings frequently suffer from sleep problems, lack of concentration, fatigue, headaches, depression, drowsiness, poor attention span, impairment of taste and smell, as well as irritations in the respiratory system and eyes. The causes of these complaints include inadequate ventilation, pollutants infiltrating from the external environment, indoor-originated pollutants, construction materials used in buildings, insufficient lighting, inappropriate humidity and temperature, odors, noise, and electromagnetic pollution (Tran et al., 2020; Mannan & Al-Ghamdi, 2021). Moreover, the reliance on artificial lighting instead of daylight, improper use of colors that otherwise provide positive energy to individuals, and inadequate or improper indoor lighting are also highlighted as factors requiring further examination (Demirarslan & Başak, 2018).

Particulate Matter

Particulate matter (PM) is an air pollutant composed of liquid and solid particles suspended in the air. This pollutant is a mixture that can exhibit different physical and chemical properties depending on its location. Common chemical components of particulate matter include sulfates, nitrates, ammonium, sodium, potassium, calcium, magnesium, and chloride ions, as well as other inorganic ions, organic and elemental carbon, crustal material, particle-bound water, metals (including cadmium, copper, nickel, vanadium, and zinc), and polycyclic aromatic hydrocarbons (PAHs). In addition, microbial components and allergens may also be present (WHO, 2013).

Particulate matter is classified according to its aerodynamic diameter. PM_{10} refers to particles with a diameter of less than or equal to 10 μm, whereas $PM_{2.5}$ refers to particles with a diameter of less than or equal to 2.5 μm. As particle size decreases, their potential to penetrate deeper regions of the lungs increases. While PM_{10} primarily affects the upper respiratory tract, $PM_{2.5}$ can reach the alveoli.

According to the World Health Organization (WHO), the daily guideline values for indoor air concentrations of PM_{10} and $PM_{2.5}$ are 50 µg/m³ and 25 µg/m³, respectively (WHO, 2013). The effects of indoor air quality on individuals manifest both as reduced work performance and various health problems (Arıkan & Tekin, 2020). Figure 1 shows an example of a comparison with the thickness of a human hair to get an idea of the different particle matter sizes.

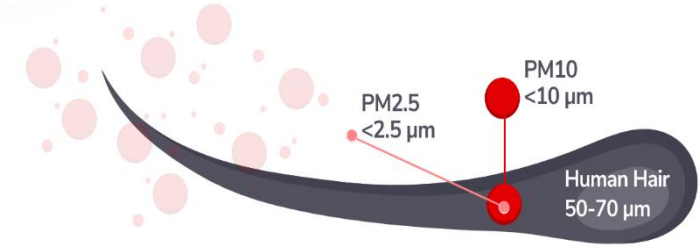


Figure 1: Particulate Matter (PM) Size Comparison

Characteristics of PM_{10}

PM_{10} refers to particulate matter that passes through an inlet with 50% efficiency for particles with an aerodynamic diameter of 10 µm, as defined by EN 12341 (Regulation on Air Quality Assessment and Management, 2008).

PM_{10} consists of solid or liquid particles suspended in the air that are smaller than 10 micrometers (µm) in diameter. These particles can remain airborne for long periods and easily reach the lungs through the respiratory system. PM_{10} originates from various sources, including dust, pollen, mold spores, smoke, industrial waste, and combustion by-products (Seinfeld & Pandis, 2016; Manisalidis et al., 2020; WHO, 2021).

The sources of PM_{10} are generally divided into two categories: natural and anthropogenic (human-related) sources. Natural sources include soil transport, sea salt, forest fires, and volcanic activities, while anthropogenic sources include motor vehicles, industrial facilities, coal-based heating systems, and, particularly in indoor environments, human activity, cleaning operations, building materials, and cooking processes (EEA, 2019; Kelly & Fussell, 2019). Indoor PM_{10} levels are largely influenced by outdoor air quality, but indoor sources can also significantly affect these concentrations. PM_{10} particles can be generated indoors through activities such as smoking, burning candles, cooking with gas stoves, and resuspension of dust from carpets and other surfaces. Moreover, some studies have shown that factors such as window type (especially older windows), heating method (particularly coal usage), and seasonal variations have noticeable impacts on indoor PM_{10} concentrations (Abdel-Salam, 2021). Indoor PM_{10} levels are typically lowest

during summer, while wintertime emissions from heating and inadequate ventilation lead to increased concentrations.

In supermarkets and fitness centers, PM_{10} pollution is mostly caused by resuspended dust from indoor activities, insufficient ventilation, and infiltration of outdoor air. In particular, increased air movement during exercise in fitness centers can resuspend dust from the floor, leading to sudden spikes in PM_{10} levels.

Unlike ambient air quality, universally binding legal standards specifically for indoor air quality are less common. Therefore, the most recent guidelines from international health organizations and specific national indoor air quality regulations are used as primary references. Table 1 presents the PM_{10} limit values explicitly defined for indoor environments by South Korea's Indoor Air Quality Control Act, Hong Kong's Indoor Air Quality Objectives and China's national standard.

Table 1: Examples of National Indoor PM_{10} Limit Values in Some Countries

Country / Standard	PM_{10} Limit Value	Time Period	Description
South Korea (Indoor Air Quality Control Act)	100 $\mu g/m^3$	Maintenance Standard	A mandatory legal standard for public-use facilities and multi-family housing.
Hong Kong EPD (IAQ Objectives, 2019)	100 $\mu g/m^3$	8-hour average	Defined for "Good Class" indoor air quality certification in offices and public places.
China (GB/T 18883-2022)	50 $\mu g/m^3$	24-hour average	The newly revised national standard specifically for indoor air quality, effective from 2023.

EXAMPLES OF LITERATURE STUDIES CONDUCTED IN INDOOR ENVIRONMENTS COVERED IN THE STUDY

Observations and on-site measurements conducted in supermarkets revealed that high PM_{10} concentrations in the bakery section were linked to flour usage, particularly during the kneading process when flour is manually spread on the counter and dough is pressed. Furst et al. (2024) found that

prolonged exposure to flour is associated with various respiratory diseases, especially baker's asthma, which is a well-documented occupational illness. PM_{10} concentrations in the bakery area exceeded the Portuguese indoor air quality regulation of 50 $\mu g/m^3$ per day, with peak values reaching 515 $\mu g/m^3$. Concentrations increased rapidly starting from 06:00, when production began, and decreased only during the workers 10:30 break. The highest average concentration (123 $\mu g/m^3$) was recorded between 06:00 and 08:00, corresponding to the period of most intense production. A decrease was observed after production ended at 15:00 until 19:00, when cleaning activities took place. PM_{10} levels were also elevated in the storage area between 13:00–16:00 (truck unloading times) and peaked again around 17:30 and 19:30. Outdoor monitors recorded average PM_{10} concentrations of 25.4 $\mu g/m^3$ (optical) and 19.1 $\mu g/m^3$ (gravimetric). PM_{10} levels increased during morning commuting (06:00), lunchtime return (14:00), and store closing (19:00). Variability in indoor PM_{10} concentrations was attributed to resuspension of dust, human traffic, stocking, packaging, and ventilation efficiency (Furst et al., 2024).

Using a DustTrak Aerosol Monitor (Model 8530), Rumchev et al. (2021) found that PM_{10} was the dominant particulate fraction in the dough preparation area, with concentrations reaching up to 181 $\mu g/m^3$. The highest PM_{10} levels occurred in zones characterized by intensive mechanical activity and during peak operational periods, underscoring the significant influence of flour handling and particle resuspension on airborne particulate emissions (Rumchev et al., 2021).

Slezakova et al. (2018) reported that particulate matter concentrations in fitness centres were strongly influenced by ventilation type, occupancy density, and the intensity of physical activity. In naturally ventilated gyms, PM_{10} levels ranged between 5–1080 $\mu g/m^3$ and $PM_{2.5}$ between 5–777 $\mu g/m^3$, whereas in mechanically ventilated facilities, PM_{10} ranged between 11–106 $\mu g/m^3$ and $PM_{2.5}$ between 5–104 $\mu g/m^3$. During periods of user presence, concentrations increased up to 16-fold for PM_{10} and 12-fold for $PM_{2.5}$, with levels rising as much as 35-fold during cleaning activities. These findings underscore the critical role of ventilation performance and indoor activity patterns in shaping particulate exposure levels (Slezakova et al., 2018).

MATERIALS AND METHODS

Study Area and Measurement Points

This study was conducted in two different indoor environments located in the Aşağı Söğütönü District of Eskişehir province in Türkiye: a supermarket and a fitness center. Both are located in the same neighborhood and on the same street. Measurements were carried out simultaneously under both indoor and outdoor conditions during periods of high and low occupancy. Each measurement period was set to one hour, during which PM_{10} particulate matter concentrations were continuously monitored.

During the low-occupancy period, approximately 40 customers were recorded entering and exiting the supermarket, whereas this number rose to about 250 during the high-occupancy period. In the fitness center, the low-occupancy period included one trainer and five athletes, while the high-occupancy period involved two trainers and twenty athletes. These fluctuations in user density were considered when interpreting the measurement data. Figure 2 shows the locations of the supermarket and fitness center which were in the same neighborhood.

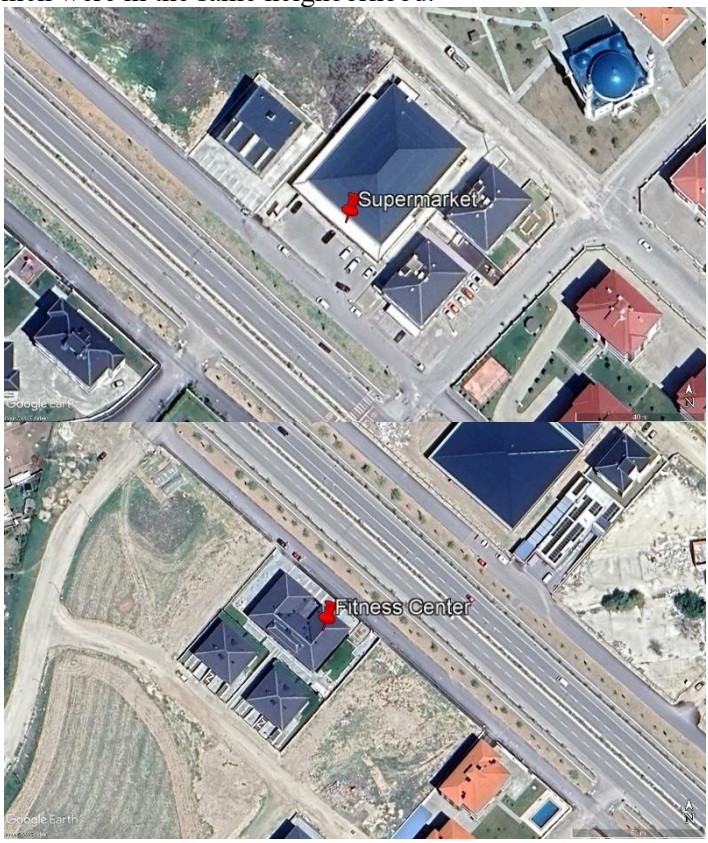


Figure 2: The Locations of The Supermarket and Fitness Center Considered as The Study Area

Measurement Program

As part of the indoor air quality (IAQ) assessment, PM_{10} concentrations were measured simultaneously in both indoor and outdoor environments. The measurements were performed using two DustTrak™ II Aerosol Monitors, and minute-by-minute data recordings were obtained. Measurements were conducted in both the fitness center and supermarket during periods of low (non-crowded) and high (crowded) user occupancy. Each measurement period was planned to last one hour, during which simultaneous outdoor measurements were also taken. Indoor/Outdoor (I/O) ratios were calculated using PM concentrations obtained from simultaneously performed measurement results. For the supermarket, the low-occupancy measurement period was conducted between 09:00 a.m. and 10:00 a.m., when the number of customers was relatively low. The high-occupancy period corresponded to 6:00–7:00 p.m., aligning with after-work hours when customer density increased. Visual representation of the measurement study is presented in Figure 3.

Figure 3: Supermarket Indoor and Outdoor Measurement Points

PM_{10} measurements in the fitness center were also conducted during two different time periods based on occupancy levels. The low-occupancy measurement was carried out between 10:30 and 11:30 a.m., when the number of users was relatively low. The high-occupancy measurement took place between 8:00 and 9:00 p.m., corresponding to the period when the fitness center was the most crowded. In both measurement sessions, simultaneous outdoor measurements were also performed. This allowed for the calculation of Indoor/Outdoor (I/O) PM_{10} concentration ratios. Visual representation of the measurement study is presented in Figure 4.

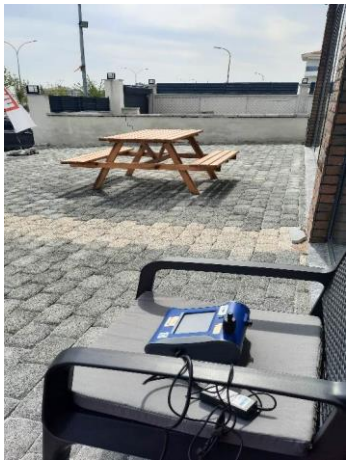


Figure 4: Fitness Center Indoor and Outdoor Measurement Points

Measurement Method

As previously mentioned, the measurements were conducted using two DustTrak™ II Aerosol Monitors. The operating principle of the device is based on the continuous drawing of aerosols into the instrument using a diaphragm pump. A portion of the aerosol flow is separated before entering the detection chamber and passed through a HEPA filter. This filtered air is then reinjected around the inlet as a sheath (protective) flow to prevent contamination of the optical components. The remaining portion, known as the sample flow, enters the detection chamber directly.

Within this chamber, the aerosol is illuminated by a laser light sheet generated by a laser diode. The light emitted from the diode first passes through a collimation lens, then through a cylindrical lens, forming a thin layer of light. A gold-coated spherical mirror collects a significant portion of the light scattered by the particles and focuses it onto a photodetector. The voltage generated at the photodetector is proportional to the mass concentration of the aerosol over a wide concentration range.

RESULTS

PM_{10} Concentration Results

Low-Occupancy Period Results for the Supermarket

Indoor PM_{10} measurements for the supermarket were carried out in the early hours of the day, during the period right after the market was opened. This time frame generally represents the period with the fewest customers, during which supermarket employees perform opening preparations such as shelf arrangement and cleaning activities. During the measurement period, approximately 40 customers were observed entering and leaving the market. The PM_{10} concentration data obtained during this period are presented in Figure 5.

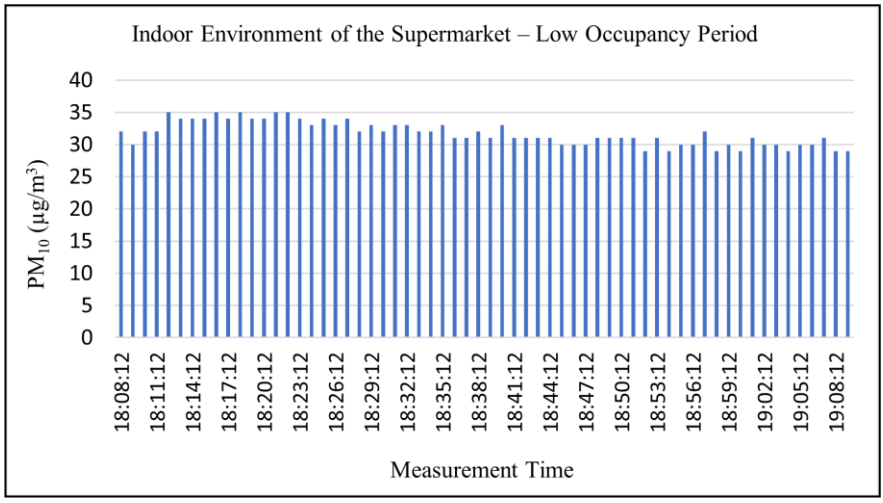


Figure 5: Indoor PM_{10} Concentrations of the Supermarket – Low-Occupancy Period

The supermarket structure includes two automatic doors used for both entry and exit, directly opening to an outdoor parking area. The parking area experiences varying levels of vehicle traffic throughout the day. In the morning hours, when employees arrive and park their vehicles, particulate matter can be transported into the air due to vehicle emissions. PM_{10} particles emitted from vehicle exhausts can be carried into the indoor environment through the repeated opening and closing of these doors. Furthermore, it is known that there is a correlation between outdoor particulate matter concentrations and indoor levels. Therefore, outdoor pollutants can directly affect indoor air quality. The measured outdoor PM_{10} concentrations are shown in Figure 6.

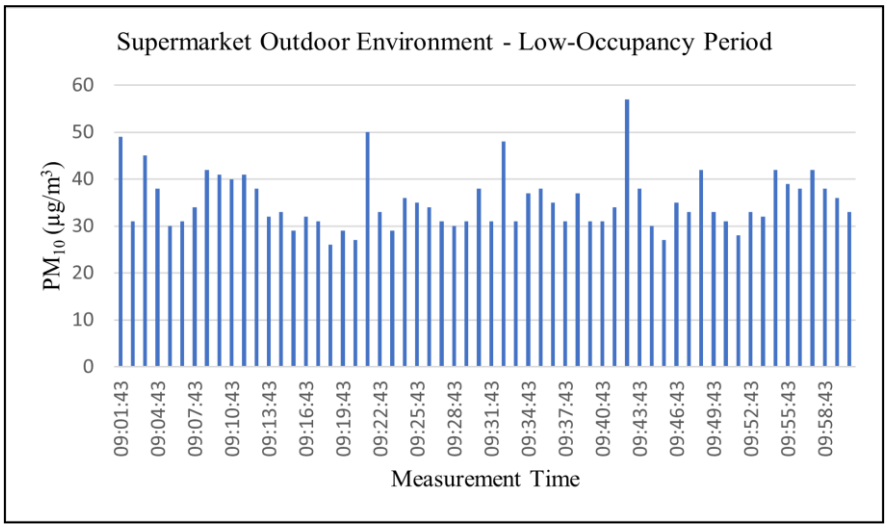


Figure 6: Outdoor PM_{10} Concentrations of the Supermarket – Low-Occupancy Period

High-Occupancy Period Results for the Supermarket

The high-occupancy period for the supermarket corresponds to the evening hours, which overlap with post-work hours when shopping traffic intensifies. During this period, approximately 250 customer entries and exits occurred inside the market nearly six times higher than during the low occupancy period. The graph created from the measured data is presented in Figure 7.

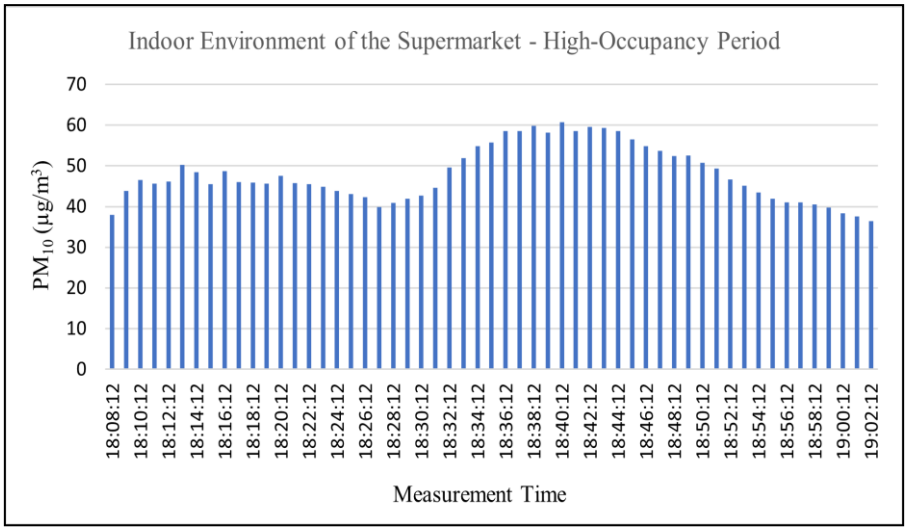


Figure 7: Indoor PM_{10} Concentrations of the Supermarket – High-Occupancy Period

The main doors of the supermarket open directly onto the parking area. During peak hours, a noticeable increase in vehicle traffic was observed around the supermarket. This condition likely contributed to higher indoor particulate matter levels. Emissions from vehicles—especially exhaust gases and brake-wear dust—have the potential to be transported indoors.

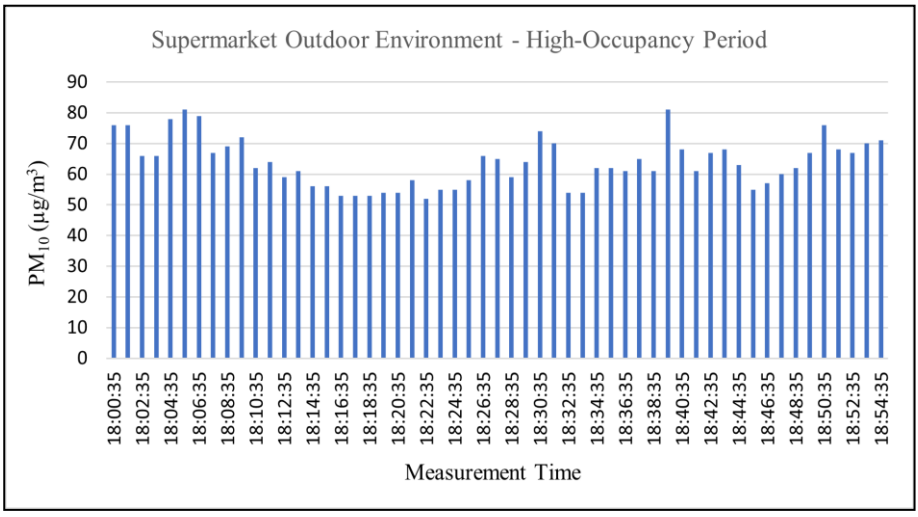


Figure 8: Outdoor PM_{10} Concentrations of the Supermarket – High-Occupancy Period

Low-Occupancy Period Results for the Fitness Center

Measurements in the fitness center were conducted during off-peak hours, when user density was low. During this time, only one trainer and five athletes were present. The small number of individuals performing exercises limited the level of biological and physical pollutants generated indoors. During the measurements, the effect of the rubber flooring material on indoor air quality was also considered. While rubber floors are chosen for their durability and slip resistance, they also have the potential to generate small particles through friction. The results of the measurements are shown in Figure 9.

The fitness center is located approximately 20 meters from the heavily used Söğüt–Uludere Road in Eskişehir city. One fade of the building directly faces this road, which is frequently used throughout the day by municipal buses, worker shuttles, and private vehicles. This intense traffic flow may lead to high concentrations of exhaust-related air pollutants in the outdoor environment. Measured outdoor PM_{10} concentrations are shown in Fig. 10.

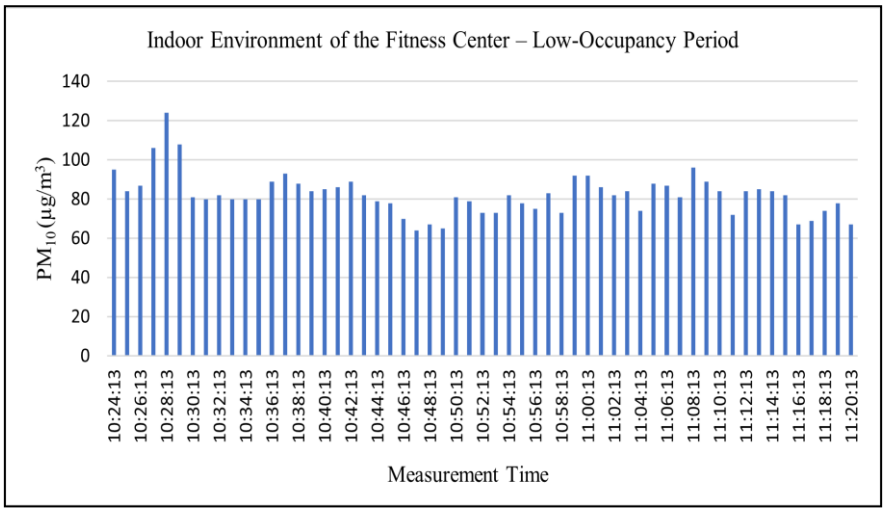


Figure 9: Indoor PM_{10} Concentrations of the Fitness Center – Low-Occupancy Period

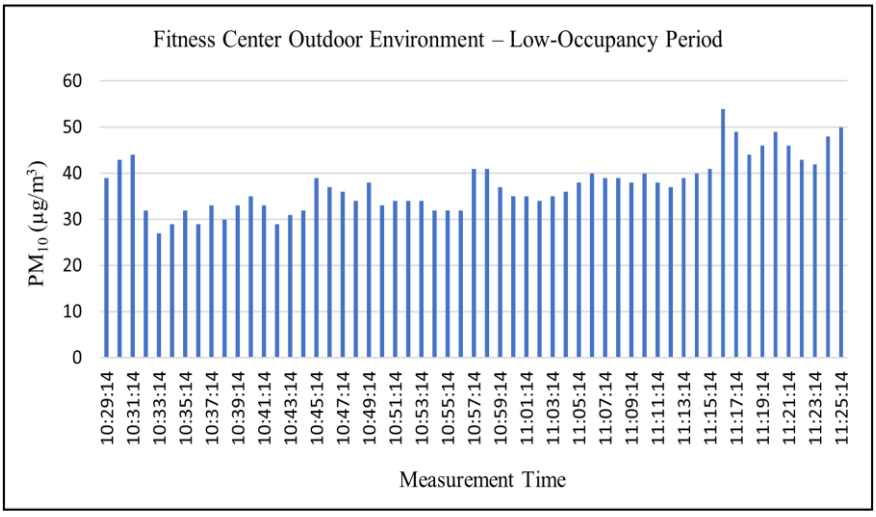


Figure 10: Outdoor PM_{10} Concentrations of the Fitness Center – Low-Occupancy Period

High-Occupancy Period Results for the Fitness Center

During PM_{10} measurements conducted in the high-occupancy period of the fitness center, twenty-two people were actively exercising. The type and intensity of the exercise performed play a determining role in indoor air quality. Activities such as aerobics, weightlifting, jumping, and treadmill running—characterized by high movement intensity—can cause accumulated dust and particles in the air to resuspend. The measurement results are shown in Figure 11.

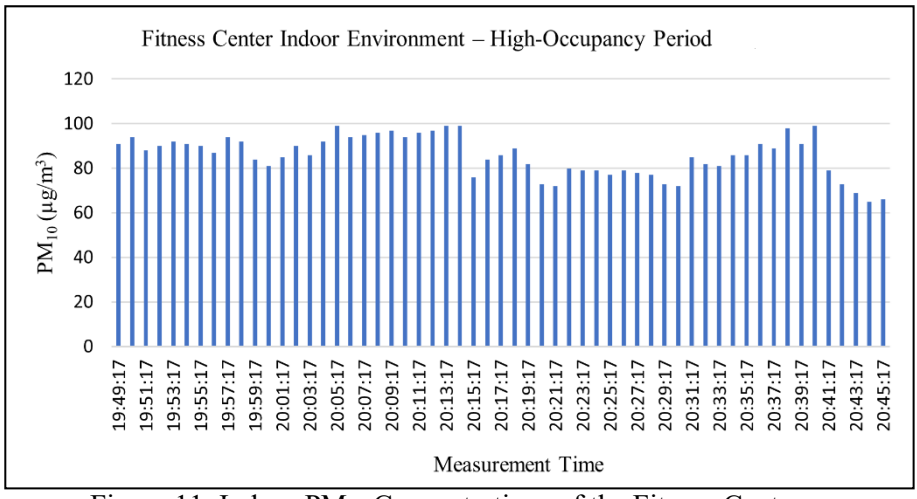


Figure 11: Indoor PM_{10} Concentrations of the Fitness Center – High-Occupancy Period

In addition, emissions from vehicles waiting in the parking area in front of the gym also contributed to the indoor PM_{10} levels. When both physical activities and environmental factors are considered together, the variation in particulate matter concentration is directly related not only to the number of users but also to the nature of the activities and outdoor conditions.

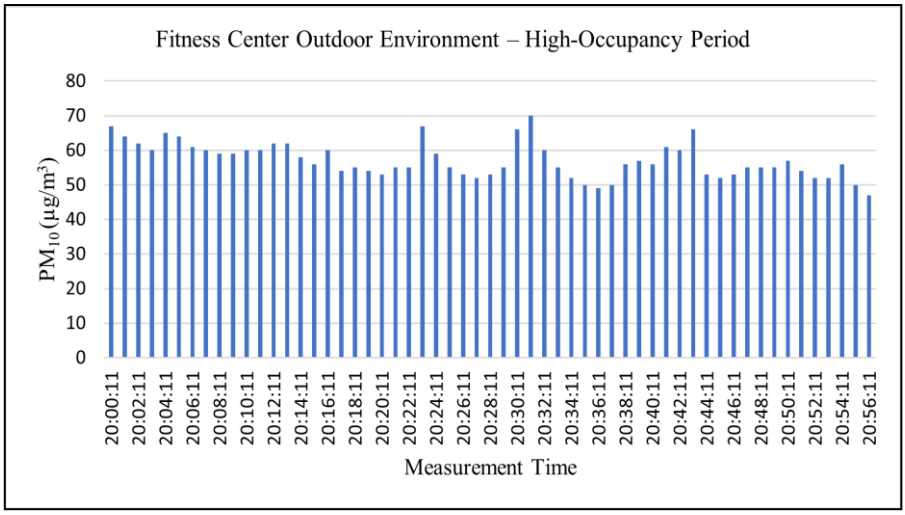


Figure 12: Outdoor PM_{10} Concentrations of the Fitness Center – High-Occupancy Period

Comparison of Indoor PM_{10} Concentrations

Figure 13 illustrates the mean PM_{10} concentrations and their corresponding standard deviations measured in the indoor environments of the supermarket and fitness center. During both high- and low-occupancy periods, PM_{10} levels in the fitness center exceeded those observed in the supermarket, reaching approximately 87 µg/m³ and 83 µg/m³, respectively. These results highlight the substantial contribution of elevated physical activity to increased particulate matter generation in indoor settings.

Indoor PM_{10} concentrations in the supermarket were recorded at 48 µg/m³ during peak hours and 32 µg/m³ during low-occupancy periods. These levels were notably lower than those measured in the fitness center, indicating that routine supermarket activities—such as customer circulation and shelf stocking—generate comparatively less dust resuspension. The reduced concentrations during low-traffic periods further reflect an improvement in indoor air quality when the space is less occupied.

These results demonstrate that both occupancy level and the nature of activities significantly influence indoor air quality, underscoring the need for effective ventilation strategies in environments with high physical exertion. Nonetheless, indoor air quality is shaped by more than human-related factors. Elements such as flooring materials, sports equipment, and cleaning products may also contribute to increased particulate generation. Additionally, poor or insufficient ventilation can hinder the removal of contaminants, further diminishing air quality. Therefore, achieving better indoor air quality requires addressing both user-related sources and structural or technical considerations.

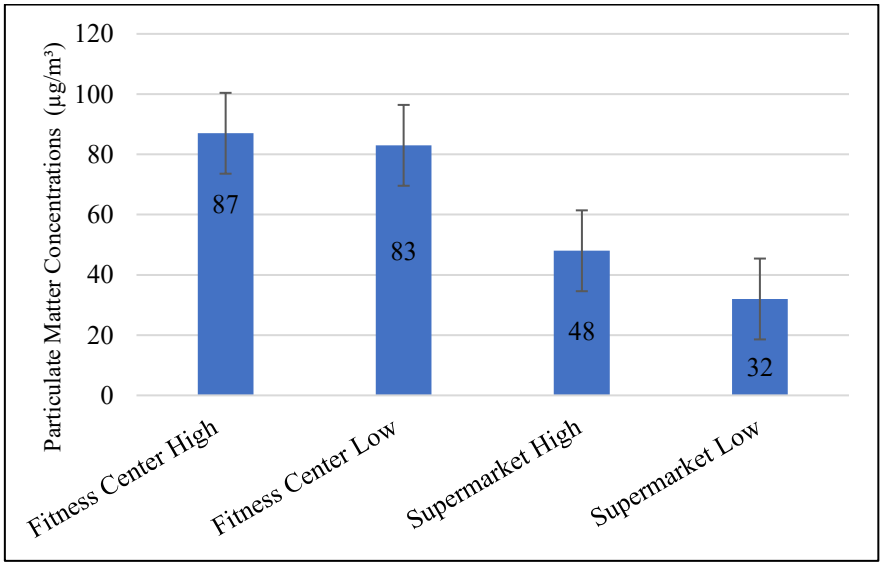


Figure 13: Average Indoor PM_{10} Concentrations

Indoor/Outdoor Concentration Ratios (I/O)

The ratios derived by dividing indoor concentrations by outdoor concentrations obtained from simultaneous measurements are presented in Figure 14. Values lower than one suggest that the outdoor environment is the primary source of pollutants entering the indoor space.

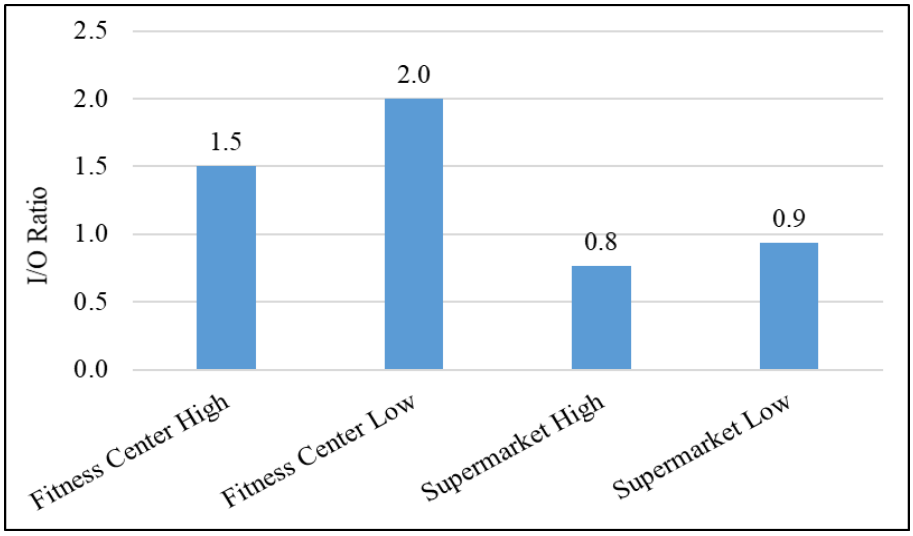


Figure 14: Indoor/Outdoor Ratios

During the high-occupancy period in the fitness center, the I/O ratio was calculated as 2, while during the low-occupancy period, it was 1.5. Since these ratios are greater than one, the outdoor environment did not act as a pollution source for the indoor space. This shows that indoor activities were the primary contributors to PM_{10} levels. Factors leading to higher ratios include the intensity of human activity, the type of sports equipment used, floor material, insufficient ventilation, or low performance of the ventilation system.

In the supermarket, the I/O ratio was measured as 0.9 during high-occupancy periods and 0.8 during low-occupancy periods. Although these values are below "1", they are relatively close, indicating that the outdoor environment contributes to indoor PM_{10} levels. This also suggests that both indoor and outdoor sources play a comparable role in determining the observed concentrations. Given that the supermarket is located in a developing area, it is possible that the I/O ratio may exceed the value "1" in the future. Therefore, enhancing the ventilation system could help maintain acceptable indoor air quality.

Integrated Evaluation of PM_{10} Concentrations in Indoor and Outdoor Environments

Figure 15 shows the combined indoor and outdoor PM_{10} concentrations for the supermarket and the fitness center. In the fitness center, average indoor PM_{10} concentration was 87±8 µg/m^3 during peak hour and 83±10 µg/m^3 during off-peak hour, while average outdoor concentration was 59±13 µg/m^3 and 42±28 µg/m^3, respectively. In both cases, indoor concentrations were higher than outdoor ones. This indicates that user movement, exercise intensity, materials used, and floor types act as significant indoor pollution sources.

In the supermarket, indoor average PM_{10} concentration was 48±7 µg/m^3 during high-occupancy period and 32±2 µg/m^3 during low-occupancy period, while average outdoor concentrations were 64±8 µg/m^3 and 35±6 µg/m^3, respectively. Although the supermarket showed a more balanced distribution of indoor and outdoor PM_{10} levels, outdoor concentrations were higher during peak hours. This indicates that supermarkets generate less indoor particulate matter compared to fitness centers, but are more influenced by external factors such as vehicle traffic.

Overall, indoor PM_{10} concentrations in the fitness center exceeded those in the supermarket during both low- and high-occupancy periods and remained higher than the corresponding outdoor levels.

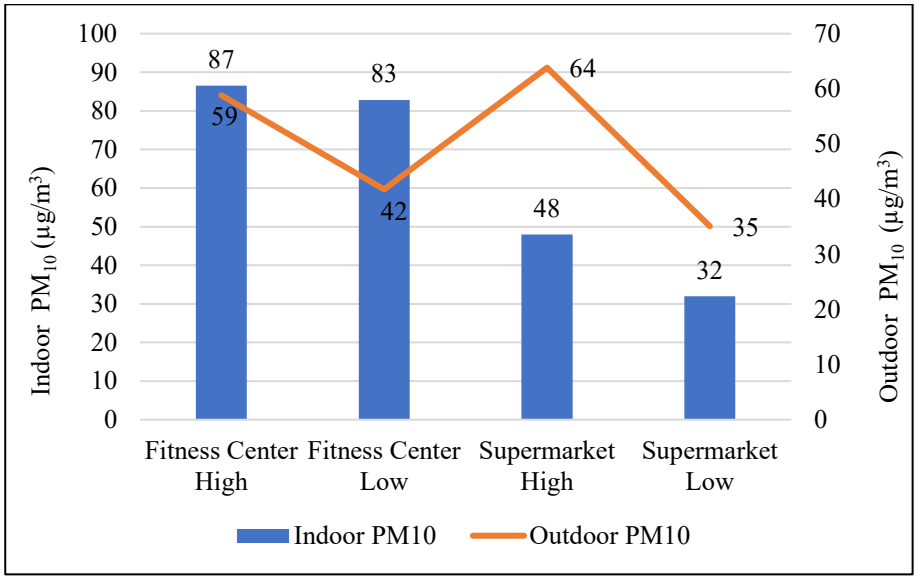


Figure 15: Combined Evaluation of Indoor and Outdoor PM_{10} Concentrations

CONCLUSION AND EVALUATION

In this study, indoor air quality was evaluated in two enclosed environments: a fitness center and a supermarket. The results revealed distinct characteristics for each setting. In the fitness center, indoor PM_{10} concentrations were consistently higher than outdoor levels during both peak and off-peak hours. This indicates that internal sources, specifically the resuspension of dust caused by intense physical activity, human movement, and the nature of rubber flooring materials, are the primary drivers of particulate matter generation. Conversely, in the supermarket, indoor concentrations were generally lower than or closer to outdoor levels, suggesting that the indoor environment is relatively better controlled, but is significantly influenced by the infiltration of outdoor pollutants, particularly from vehicle traffic during peak hours. Finally, this study focused on short-term (1-hour) peak exposure periods and the high short-term concentrations observed, particularly in the fitness center, highlight a potential acute exposure risk for occupants, emphasizing the need for continuous monitoring and improved ventilation strategies.

To improve indoor air quality and mitigate health risks in such high-occupancy environments, specific measures should be implemented. For fitness centers, where internal generation is dominant, increasing the air exchange rate and using high-efficiency particulate air (HEPA) filtration systems are essential to remove suspended particles. Additionally, replacing dry cleaning methods with frequent wet mopping can significantly reduce dust resuspension from flooring. For supermarkets, which are more affected by external sources, installing air curtains at entrances and maintaining positive indoor pressure can prevent the infiltration of outdoor pollutants. Furthermore, upgrading HVAC filters to capture fine particles from vehicle emissions is recommended to protect customers and staff.

REFERENCES

Abdel-Salam, M.M.M. (2021). *Seasonal variation in indoor concentrations of air pollutants in residential buildings*. Journal of the Air & Waste Management Association, 71(6), 761–777. https://doi.org/10.1080/10962247.2021.1895367

Arıkan, İ., ve Tekin, Ö. F. (2020). Partiküler madde ve karbondioksit için iç ortam hava kalitesi indeksi (İHKİ) hesaplaması: okul örneği. Eskişehir Türk Dünyası

Uygulama ve Araştırma Merkezi Halk Sağlığı Dergisi, 5 (2), 188-195. https://doi.org/10.35232/estudamhsd.699656

Demirarslan, K. O. ve Başak, S. (2018). Hasta bina sendromu kavramı literatür araştırması ve çeşitli mekânların iç hava kalitelerinin karşılaştırılması. Mühendislik Bilimleri ve Tasarım Dergisi, 6 (2), 190-201. https://doi.org/10.21923/jesd.340029

Environmental Protection Department (HKEPD). (2019). A Guide on Indoor Air Quality Certification Scheme for Offices and Public Places. The Government of the Hong Kong Special Administrative Region. Retrieved November 28, 2025, from https://www.iaq.gov.hk/wp-content/uploads/2021/04/new-iaq-guide_eng.pdf

Furst, L., Cipoli, Y., Galindo, N., Yubero, E., Viegas, C., Pena, P., Nunes, T., Feliciano, M., Alves, C. (2024). Comprehensive analysis of particulate matter, gaseous pollutants, and microbiological contamination in an international chain supermarket. Environmental Pollution, 363, 125236. https://doi.org/10.1016/j.envpol.2024.125236

Isinkaralar, O., Rajfur, M., Isinkaralar, K., Świstołowski, P. (2025). Contamination degree and health implications of indoor air pollution: Operating field measurements in market environments. Science of the Total Environment, 969, 178946. https://doi.org/10.1016/j.scitotenv.2025.178946

Kelly, F. J., & Fussell, J. C. (2019). Improving indoor air quality, health and performance within the built environment. *Building and Environment, 186*, 107314. https://doi.org/10.1016/j.atmosenv.2018.11.058

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and health impacts of air pollution: A review. *Frontiers in Public Health, 8*, 14. https://doi.org/10.3389/fpubh.2020.00014

Mannan, M., & Al-Ghamdi, S. G. (2021). Indoor air quality in buildings: A comprehensive review on the factors influencing air pollution in residential and commercial buildings. *International Journal of Environmental Research and Public Health, 18*(6), 3276. https://doi.org/10.3390/ijerph18063276

Ministry of Environment (South Korea). (2023). Indoor Air Quality Control Act. Statutes of the Republic of Korea. Retrieved November 28, 2025, from https://elaw.klri.re.kr/eng_service/main.do

Morawska, L., Afshari, A., Bae, G. N., Buonanno, G., Chao, C. Y. H., Hänninen, O., Hofmann, W., Isaxon, C., Jayaratne, E. R., Pasanen, P., Salthammer, T., Waring, M., & Wierzbicka, A. (2013). Indoor aerosols: From personal exposure to risk assessment. Indoor Air, 23(6), 462–487. https://doi.org/10.1111/ina.12044

Rumchev, K., Zhao, Y., Lee, A. (2021). Case report: occupational dust exposure among bakery workers in Perth, western Australia. Front. Public Health, 9,723154. https://doi.org/10.3389/fpubh.2021.723154

Seinfeld, J.H. and Pandis, S.N. (2016) Atmospheric Chemistry and Physics: From Air Pollution to Climate Change. John Wiley & Sons, Hoboken.

Slezakova, K., Peixoto, C., Oliveira, M., Delerue-Matos, C., Pereira, M. do C., Morais, S. (2018). Indoor particulate pollution in fitness centres with emphasis on ultrafine particles. *Environmental Pollution, 233*, 180–193. https://doi.org/10.1016/j.envpol.2017.10.050

State Administration for Market Regulation (SAMR). (2022). GB/T 18883-2022: Standards for Indoor Air Quality. Standardization Administration of China. Retrieved November 29, 2025, from https://www.codeofchina.com/standard/GBT18883-2022.html

Tran, V. V., Park, D., & Lee, Y. C. (2020). Indoor air pollution, related human diseases, and recent trends in the control and improvement of indoor air quality. *International Journal of Environmental Research and Public Health, 17*(8), 2927. https://doi.org/10.3390/ijerph17082927

World Health Organization. (2021). WHO global air quality guidelines: Particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. https://www.who.int/publications/i/item/9789240034228

World Health Organization (WHO). Regional Office for Europe (2013). Health effects of particulate matter: policy implications for countries in eastern Europe, Caucasus and central Asia. World Health Organization. Regional Office for Europe. https://iris.who.int/handle/10665/344854

Zencirci, S. A. ve Işıklı, B. (2017). Hava Kirliliği. Türk Dünyası Uygulama ve Araştırma Merkezi Halk Sağlığı Dergisi, 2(2), 24–36.

Understanding Landslide Hazards: Causes and Assessment Methods

Samet BERBER[1]

1- Assistant Professor.; Balıkesir Üniversitesi Mühendislik Fakültesi Jeoloji Mühendisliği Bölümü. sametberber@balikesir.edu.tr ORCID No: https://orcid.org/0000-0002-8747-9346

ABSTRACT

Landslide hazard analysis represents a critical component of natural hazard assessment, integrating multidisciplinary knowledge to predict where, when, and how large future landslides may occur. This chapter provides a comprehensive examination of landslide hazard assessment concepts, methodologies, and applications. We begin by establishing fundamental definitions, clearly distinguishing between landslide susceptibility, hazard, and risk—three related but distinct concepts that serve different purposes in risk management frameworks. The complex interplay between controlling factors (topography, geology, soil properties, hydrology, vegetation) and triggering factors (rainfall, earthquakes, human activities) is explored, forming the foundation for understanding slope failure mechanisms. Temporal probability and magnitude assessment methods are examined, including statistical, physically-based, and hybrid approaches that enable quantitative hazard estimation. Geographic Information Systems (GIS) have revolutionized spatial landslide hazard mapping, facilitating integration of diverse datasets and application of sophisticated analytical techniques ranging from qualitative expert assessment to machine learning algorithms. Despite significant advances, persistent challenges remain, including data limitations, temporal prediction uncertainties, climate change impacts, and effective integration into decision-making processes. Emerging technologies such as satellite-based monitoring, artificial intelligence, and real-time early warning systems offer promising opportunities for enhancing assessment capabilities and expanding global coverage, ultimately contributing to more resilient communities in landslide-prone regions.

Keywords – Landslide Hazard, Susceptibility Mapping, Temporal Probability, GIS-Based Analysis, Risk Assessment.

INTRODUCTION

Landslides represent one of the most widespread and damaging geological hazards affecting communities worldwide, causing thousands of casualties and billions of dollars in economic losses annually (Petley, 2012). These mass movement phenomena occur across diverse geological, topographical, and climatic settings, from mountainous terrains to coastal cliffs, affecting both densely populated urban areas and remote rural communities. The increasing frequency and impact of landslide events in recent decades can be attributed to multiple factors, including intensified rainfall patterns associated with climate change, accelerated urbanization in hazard-prone areas, deforestation, and inadequate land-use planning (Froude and Petley, 2018).

The systematic assessment of landslide hazards has evolved significantly over the past several decades, transitioning from qualitative expert-based evaluations to sophisticated quantitative analyses employing advanced computational techniques and extensive spatial datasets. This evolution reflects both the growing recognition of landslides as a significant societal threat and the technological advances that have revolutionized our ability to analyze and predict these phenomena (Guzzetti et al., 1999). Modern landslide hazard analysis integrates multiple disciplines, including geology, geomorphology, hydrology, soil mechanics, remote sensing, and geographic information systems (GIS), creating a comprehensive framework for understanding and predicting slope failures.

The importance of rigorous landslide hazard assessment cannot be overstated, as it forms the foundation for effective risk management strategies, land-use planning decisions, and the development of early warning systems. Accurate hazard assessment enables policymakers, urban planners, and disaster management authorities to make informed decisions regarding infrastructure development, residential zoning, and the allocation of resources for mitigation measures (Fell et al., 2008). Furthermore, as climate change continues to alter precipitation patterns and increase the frequency of extreme weather events, the need for robust landslide hazard assessment methodologies becomes increasingly urgent.

This chapter provides a comprehensive examination of landslide hazard analysis, encompassing fundamental concepts, analytical methodologies, and practical applications. We begin by establishing clear definitions and distinguishing between related but distinct concepts such as hazard, susceptibility, and risk. Subsequently, we explore the complex interplay of controlling and triggering factors that govern landslide occurrence, followed by detailed discussions of temporal probability assessment and magnitude estimation. The chapter then examines GIS-based mapping techniques that have revolutionized spatial landslide hazard analysis, before concluding with an examination of current challenges and future research directions in this dynamic field.

DEFINITION OF LANDSLIDE HAZARD

Landslide hazard is formally defined as the probability of occurrence of a potentially damaging landslide phenomenon within a specified area and time period (Varnes, 1984). This definition encapsulates three critical components that distinguish hazard assessment from other forms of landslide analysis: spatial location, temporal probability, and the potential for damage. The spatial component identifies where landslides may occur, the temporal component addresses when they might happen, and the damage potential relates to the magnitude or intensity of the anticipated event.

The concept of landslide hazard fundamentally differs from simple landslide occurrence or inventory mapping. While landslide inventories document past slope failures and provide valuable historical records, hazard assessment is inherently forward-looking, seeking to predict future landslide activity (Guzzetti et al., 2012). This predictive nature introduces significant complexity, as it requires understanding not only where landslides have occurred previously but also the conditions under which future failures might develop. The hazard assessment process must therefore integrate knowledge of past landslide patterns with an understanding of the dynamic factors that may trigger future events.

Quantifying landslide hazard requires specification of both magnitude and frequency. Magnitude refers to the size, volume, velocity, or destructive potential of a landslide, while frequency relates to how often landslides of various magnitudes occur within a given timeframe (Corominas et al., 2014). This magnitude-frequency relationship is analogous to that used in other natural hazard assessments, such as earthquakes or floods, and follows similar probabilistic principles. Large, catastrophic landslides typically occur less frequently than smaller events, but their consequences can be disproportionately severe, necessitating careful consideration in hazard assessment frameworks.

The temporal component of landslide hazard can be expressed in various forms, ranging from qualitative estimates (such as low, medium, or high probability) to quantitative probabilities derived from statistical analysis of historical data. For instance, hazard might be expressed as an annual probability (the likelihood of occurrence in any given year) or as a return period (the average time interval between events of a particular magnitude). The choice of temporal framework depends on available data, the intended application, and the decision-making context. Short-term hazard assessments might focus on imminent threats associated with specific triggering events like rainstorms, while long-term assessments consider multi-decadal or centennial timeframes relevant for land-use planning.

Several classification schemes have been developed to standardize landslide hazard assessment and communication. These typically involve categorizing areas into hazard zones based on the combination of spatial

probability and expected magnitude. Common classification systems range from simple three-level schemes (low, medium, high) to more complex multi-level frameworks that incorporate both probability and consequence considerations (Fell et al., 2008). The selection of an appropriate classification system should balance scientific rigor with practical applicability, ensuring that hazard information can be effectively communicated to and utilized by decision-makers and affected communities.

HAZARD, SUSCEPTIBILITY AND RISK

Understanding the distinctions between landslide susceptibility, hazard, and risk is essential for effective communication and appropriate application of assessment methodologies. These three concepts, while related and often confused in practice, represent fundamentally different aspects of landslide analysis and serve distinct purposes in the broader framework of landslide risk management (Guzzetti et al., 1999; van Westen et al., 2006).

Landslide susceptibility refers to the spatial likelihood or predisposition of an area to produce landslides based on local terrain conditions and inherent geological characteristics. Susceptibility assessment answers the question "where are landslides likely to occur?" without specifying when they might occur or their expected magnitude (Brabb, 1984). This spatial probability is controlled by predisposing factors such as slope angle, lithology, soil properties, land cover, and geomorphological characteristics. Susceptibility maps typically categorize terrain into zones of varying likelihood for landslide occurrence, from very low to very high susceptibility, based on the presence and combination of favorable conditions for slope failure. Critically, susceptibility assessment does not incorporate temporal considerations or specific triggering mechanisms, making it a static representation of spatial probability that remains constant unless the underlying terrain characteristics change.

Landslide hazard, as previously defined, extends beyond susceptibility by incorporating both temporal probability and magnitude considerations. Hazard assessment answers not only "where" but also "when" and "how large" landslides might occur (Varnes, 1984; Guzzetti et al., 2005). This requires analysis of triggering factors such as rainfall, earthquakes, or human activities, and their expected frequency and intensity. Hazard assessment is inherently dynamic, as it must account for time-varying triggering conditions and their relationship to slope instability. For example, a slope might have

high susceptibility due to steep gradients and weak geological materials, but the actual hazard depends on the frequency and intensity of rainfall events or seismic activity capable of triggering failure. Quantitative hazard assessment typically produces probability estimates for landslides of various magnitudes within specified time periods, expressed as annual probabilities or return periods.

Landslide risk represents the expected degree of loss resulting from landslide occurrence, incorporating not only the hazard itself but also the vulnerability of exposed elements and the value of elements at risk (Varnes, 1984). Risk assessment answers the question "what are the potential consequences?" and is typically quantified as the product of hazard probability, vulnerability (the degree of loss to elements at risk), and the value or number of elements exposed (Fell et al., 2008). Risk assessment requires comprehensive analysis of buildings, infrastructure, population, and economic assets in landslide-prone areas, along with their susceptibility to damage from different landslide types and magnitudes. This multi-dimensional assessment produces quantitative estimates of expected losses, which might be expressed in monetary terms (economic risk), casualties (individual or societal risk), or other consequence metrics relevant to decision-making.

The progression from susceptibility to hazard to risk represents an increasingly complex and comprehensive analysis, each building upon the previous level. Susceptibility mapping requires understanding of terrain characteristics and their relationship to slope stability, hazard assessment adds temporal and magnitude considerations, and risk assessment incorporates the socioeconomic dimension of exposed communities and infrastructure. This hierarchical relationship has important practical implications: susceptibility assessment can be conducted with relatively limited data and is appropriate for preliminary screening and broad-scale planning, hazard assessment requires additional temporal data and is necessary for more detailed land-use decisions, and full risk assessment is typically reserved for detailed site-specific studies where comprehensive consequence analysis is justified by the magnitude of potential impacts or the importance of decisions to be made (van Westen et al., 2006).

CONTROLLING AND TRIGGERING FACTORS OF LANDSLIDES

Landslide occurrence is governed by the complex interaction between controlling factors, which predispose slopes to failure by determining their inherent stability, and triggering factors, which initiate movement by exceeding critical stability thresholds (Popescu, 1994). Understanding this distinction and the multifaceted relationships between various factors is fundamental to effective hazard assessment and forms the basis for both deterministic and statistical modeling approaches.

Controlling Factors

Controlling factors, also termed preparatory or predisposing factors, determine the intrinsic stability conditions of slopes and their susceptibility to failure. These factors are generally static or change slowly over geological or human timescales, establishing the baseline stability state upon which triggering factors act (Guzzetti et al., 1999).

Topographic factors constitute the most fundamental controls on slope stability, with slope angle being the primary determinant of gravitational driving forces. Steeper slopes experience greater shear stress and are consequently more susceptible to failure, with critical thresholds varying according to material properties and environmental conditions. However, slope stability analysis must consider additional topographic parameters beyond simple slope angle, including slope aspect (which influences solar radiation, moisture conditions, and weathering patterns), slope curvature (which affects water accumulation and stress distribution), and elevation (which correlates with various climatic and vegetation characteristics). Recent studies have demonstrated that topographic wetness index and stream power index, which incorporate upslope contributing area and slope gradient, provide valuable insights into hydrological conditions influencing stability (Moore et al., 1991).

Geological factors encompass lithology, structural features, and stratigraphy, all of which profoundly influence slope stability. Different rock types exhibit varying degrees of resistance to weathering, different shear strength characteristics, and different responses to water infiltration. Sedimentary rocks, particularly those containing clay-rich layers, are generally more susceptible to landsliding than massive crystalline rocks, though weathering and fracturing can reduce the strength of any lithology. Structural discontinuities, including bedding planes, joints, faults, and foliation, create planes of weakness that may act as potential failure surfaces when unfavorably oriented relative to slope geometry (Selby, 1993). The spatial distribution and orientation of these discontinuities, along with their infilling materials and aperture characteristics, must be carefully considered in stability assessment.

Soil properties, including grain size distribution, plasticity, cohesion, friction angle, and permeability, directly determine the mechanical behavior

of slope materials. Fine-grained soils with high clay content are particularly problematic due to their low permeability, high water retention capacity, and susceptibility to strength loss upon wetting. Residual soils developed through in-situ weathering of bedrock inherit characteristics from both the parent material and the weathering processes, creating complex spatial variability in engineering properties. Colluvial deposits, formed by previous mass movements or hillslope transport processes, are often particularly susceptible to failure due to their loose structure and heterogeneous composition (Terzaghi, 1950).

Hydrological conditions play a dual role, acting as both controlling and triggering factors. Groundwater levels, drainage patterns, aquifer characteristics, and soil moisture regimes establish baseline hydrological conditions that fundamentally influence slope stability. Elevated groundwater increases pore pressures, reducing effective stress and therefore shear strength, while also adding weight to potentially unstable masses. Springs, seeps, and areas of groundwater convergence represent particularly vulnerable zones where elevated moisture conditions persist. Surface drainage patterns determine water accumulation and infiltration patterns, with concave slopes and valley heads typically experiencing greater water convergence than convex or planar slopes (Montgomery and Dietrich, 1994).

Land cover and vegetation influence slope stability through multiple mechanisms with competing effects. Root systems provide mechanical reinforcement, binding soil particles and creating a composite material with enhanced shear strength. Forest canopies intercept rainfall, reducing infiltration and soil moisture, while evapotranspiration actively removes soil water. However, vegetation also adds weight to slopes, and tree movement during windstorms can transmit forces to root systems and underlying soils. Deforestation typically destabilizes slopes, as the beneficial effects of roots and evapotranspiration are lost while root decay creates preferential flow paths, though the temporal evolution of these effects is complex and varies among forest types and environmental settings (Sidle and Ochiai, 2006).

Triggering Factors

Triggering factors are dynamic agents that induce the transition from marginally stable to actively failing conditions by exceeding critical stability thresholds. These factors act on timescales ranging from seconds to months and represent the immediate causes of landslide initiation (Wieczorek, 1996).

Rainfall is the most common landslide trigger globally, with intensity, duration, antecedent conditions, and cumulative amounts all influencing triggering effectiveness. Intense short-duration rainfall can trigger shallow rapid landslides through rapid infiltration and transient pore pressure increases, while prolonged moderate rainfall typically triggers deeper-seated movements through gradual groundwater rise and progressive strength

reduction. Empirical rainfall thresholds, relating rainfall characteristics to landslide occurrence, have been established for numerous regions worldwide, typically expressed as relationships between rainfall intensity and duration or as cumulative rainfall amounts (Guzzetti et al., 2007). These thresholds exhibit significant regional variability reflecting differences in soil properties, antecedent moisture conditions, and slope characteristics, necessitating local calibration for effective application.

Earthquakes trigger landslides through the application of transient inertial forces that temporarily increase driving stresses on slopes. Seismically-induced landslides range from disrupted slides and rock falls to large catastrophic rock avalanches, with the spatial distribution and severity depending on ground motion characteristics, topographic amplification effects, and material properties. Strong earthquakes can trigger thousands or tens of thousands of landslides over areas exceeding 10,000 square kilometers, causing more casualties and economic losses than all other earthquake effects combined in some events (Keefer, 1984). Earthquake-induced landslides preferentially occur on steep slopes composed of weak, fractured, or poorly consolidated materials, particularly when groundwater conditions are elevated.

Human activities increasingly function as landslide triggers, particularly in rapidly developing regions where urbanization extends into mountainous terrain. Excavation and loading modify slope geometry and stress distributions, potentially reducing factors of safety below critical thresholds. Deforestation for agriculture or development removes stabilizing vegetation. Poor drainage design concentrates surface water and increases infiltration. Leaking water supply and sewage systems elevate groundwater levels. Vibrations from traffic, construction, or blasting can provide the final impetus for failure in marginally stable slopes. The distinction between human activities as triggers versus controlling factors can be ambiguous, as activities like deforestation modify long-term stability conditions while simultaneously acting as immediate failure triggers (Sidle and Ochiai, 2006).

TEMPORAL PROBABILITY AND MAGNİTUDE ASSESSMENT

The temporal dimension of landslide hazard assessment distinguishes it from susceptibility mapping and represents one of the most challenging aspects of quantitative hazard analysis. Accurate estimation of when landslides might occur and their expected magnitude is essential for effective risk management, early warning systems, and long-term planning decisions (Guzzetti et al., 2005).

Temporal Probability Assessment

Temporal probability estimation requires analysis of landslide frequency data in relation to triggering factors and their recurrence characteristics. Several approaches have been developed for temporal probability assessment, each with distinct data requirements, underlying assumptions, and applicability constraints.

Statistical analysis of historical landslide records provides the most direct approach to temporal probability estimation when sufficient data are available. This method involves analyzing landslide occurrence frequency over time, potentially stratified by magnitude classes, location, or landslide type. Poisson probability models are commonly applied, assuming that landslides occur randomly and independently through time with a constant average rate. Under these assumptions, the probability of n landslides occurring within time period t can be calculated using the Poisson distribution (Coe et al., 2004). However, the assumption of temporal independence is often violated for rainfall-triggered landslides, which may exhibit clustering during wet periods or following major triggering events. More sophisticated time-series approaches account for temporal clustering and non-stationarity in occurrence rates.

Physically-based approaches estimate temporal probability through explicit modeling of triggering mechanisms and their recurrence characteristics. For rainfall-triggered landslides, this involves combining rainfall threshold models with rainfall frequency analysis. The probability of threshold exceedance within a given time period provides an estimate of landslide probability, assuming that threshold exceedance reliably indicates landslide occurrence (Guzzetti et al., 2007). This approach requires: identification of empirical or physically-based rainfall thresholds for landslide initiation, statistical characterization of rainfall event frequency and magnitude, and assessment of antecedent moisture conditions and their influence on threshold values. Similar approaches can be applied to earthquake-triggered landslides by combining ground motion thresholds with seismic hazard assessments derived from probabilistic seismic hazard analysis frameworks (Jibson et al., 2000).

The exceedance probability method represents a hybrid approach combining susceptibility assessment with trigger frequency analysis. This method estimates temporal probability as the product of spatial susceptibility

probability and the annual probability of critical triggering conditions occurring. For example, if a slope has 0.3 probability of failure given critical rainfall conditions, and critical rainfall occurs on average once every 10 years (annual probability 0.1), the annual landslide probability is approximately 0.03 or 3% (Guzzetti et al., 1999). This approach provides an intuitive framework but requires careful consideration of conditional probabilities and independence assumptions.

Event-based versus continuous probability formulations represent fundamentally different temporal frameworks. Event-based probability addresses the likelihood of landslide occurrence during specific triggering events, such as individual rainstorms or earthquakes, while continuous probability estimates the likelihood of occurrence over extended periods regardless of triggering mechanism. Event-based probabilities are relevant for early warning systems responding to imminent threats, while continuous probabilities are more appropriate for long-term planning and risk assessment. The relationship between these formulations depends on trigger frequency and the proportion of time that critical triggering conditions exist.

Magnitude Assessment

Landslide magnitude can be characterized through various metrics, including volume, area, velocity, intensity, and destructive potential, with the appropriate metric depending on landslide type and assessment objectives. Magnitude estimation is inherently more uncertain than spatial or temporal probability assessment due to the extreme variability in landslide sizes and the complex controls on failure magnitude (Malamud et al., 2004).

Frequency-magnitude relationships provide a statistical framework for magnitude assessment, describing the relationship between landslide size and occurrence frequency. Studies have demonstrated that landslide inventories often exhibit power-law or exponential distributions, with small landslides occurring much more frequently than large ones. These relationships can be characterized by parameters that describe the scaling behavior and allow probabilistic magnitude estimation. However, frequency-magnitude distributions vary among different landslide types, geological settings, and triggering conditions, requiring careful consideration of appropriate analogs and transferability (Stark and Hovius, 2001).

Empirical magnitude estimation relates landslide size to observable characteristics through statistical analysis of inventory data. Relationships have been established between landslide volume and area, runout distance and volume, and maximum depth and planimetric dimensions. These empirical relationships enable magnitude estimation from readily observable characteristics and support rapid post-event assessment. However, the significant scatter in these relationships reflects the complex controls on landslide geometry and behavior, limiting precision for individual predictions while providing useful probabilistic bounds (Hungr et al., 1999).

Physically-based magnitude estimation employs slope stability analysis and runout modeling to estimate potential failure volumes and extent. Limit equilibrium or finite element analysis can define the volume of material likely to be involved in failure, while dynamic runout models simulate post-failure motion and deposition. These approaches provide more mechanistic insights than purely empirical methods but require detailed site characterization and involve significant parameter uncertainty. The integration of physically-based magnitude estimation with probabilistic frameworks allows quantification of magnitude uncertainty and supports risk-informed decision making (Hungr et al., 2005).

GIS-BASED LANDSLIDE HAZARD MAPPING

Geographic Information Systems have revolutionized landslide hazard assessment by providing powerful platforms for spatial data management, analysis, and visualization. GIS-based approaches enable integration of diverse datasets, application of sophisticated analytical techniques, and production of spatially distributed hazard assessments at scales ranging from site-specific to regional or national (van Westen et al., 2008).

Data Requirements and Sources

Effective GIS-based hazard mapping requires compilation of comprehensive spatial datasets representing both controlling and triggering factors. Topographic data, derived from digital elevation models (DEMs), provide the foundational dataset from which numerous derivative parameters can be calculated, including slope angle, aspect, curvature, topographic wetness, and stream power indices. The resolution and quality of DEMs critically influence subsequent analysis, with high-resolution LiDAR-derived DEMs enabling detailed characterization of subtle terrain features, while coarser SRTM or ASTER DEMs may be more appropriate for regional assessments (Tarolli, 2014).

Geological and soil maps provide information on lithology, structural features, and soil properties that fundamentally control slope stability. These data are often available from geological surveys at various scales, though the generalization inherent in small-scale regional maps may obscure important local variations. Remote sensing data, including multispectral and hyperspectral imagery, synthetic aperture radar (SAR), and interferometric

SAR (InSAR), provide complementary information on land cover, surface characteristics, and ground deformation (Tofani et al., 2013). InSAR techniques, in particular, have emerged as powerful tools for detecting subtle slope movements that may precede catastrophic failure.

Landslide inventory data form a critical component of most GIS-based hazard assessments, documenting the location, type, size, and where possible, age of past landslide events. Inventories can be compiled through field mapping, aerial photograph interpretation, satellite image analysis, or historical records, each approach offering different coverage, detail, and temporal perspectives. Multi-temporal inventories, documenting landslides from multiple time periods, provide particularly valuable insights into spatial and temporal patterns of activity (Guzzetti et al., 2012).

Mapping Methodologies

GIS-based landslide hazard mapping encompasses a spectrum of approaches varying in complexity, data requirements, and theoretical foundations. These methodologies can be broadly categorized as qualitative, quantitative statistical, and physically-based approaches.

Qualitative methods, including geomorphological mapping and heuristic assessment, rely on expert knowledge to identify and delineate hazard zones based on interpretation of terrain characteristics and landslide evidence. Expert geomorphologists integrate multiple lines of evidence to assess slope stability and classify terrain into hazard zones, often using matrix approaches that combine controlling factors according to subjective weighting schemes. While these approaches benefit from expert experience and can incorporate features difficult to quantify numerically, they suffer from subjectivity, limited reproducibility, and inability to provide rigorous probability estimates (van Westen et al., 2008).

Statistical methods employ mathematical relationships between past landslide occurrence and controlling factors to predict future landslide probability. These data-driven approaches range from simple bivariate statistical methods to sophisticated multivariate techniques. Bivariate statistical analysis examines the relationship between landslide occurrence and individual controlling factors independently, calculating weights for each factor class based on observed landslide density. These weights are then combined to produce susceptibility indices. While conceptually straightforward, bivariate methods assume independence among factors, which is often violated in reality (Süzen and Doyuran, 2004).

Multivariate statistical methods overcome independence assumptions by simultaneously analyzing multiple controlling factors and their interactions. Logistic regression has been widely applied, modeling the probability of landslide occurrence as a function of multiple predictor variables. This technique provides probability estimates, allows statistical testing of variable significance, and can incorporate both continuous and

categorical variables (Atkinson and Massari, 1998). Other multivariate approaches include discriminant analysis, which identifies the combination of variables that best separates stable from unstable areas, and weights-of-evidence modeling, which applies Bayesian probability theory to combine multiple evidence layers.

Machine learning techniques represent the most recent evolution in statistical landslide hazard mapping, employing algorithms capable of learning complex non-linear relationships from training data. Artificial neural networks, support vector machines, random forests, and more recently deep learning approaches have demonstrated impressive predictive performance in numerous applications. These methods can handle high-dimensional data, automatically identify relevant features, and capture complex interactions among controlling factors (Reichenbach et al., 2018). However, they generally function as "black boxes" with limited interpretability, require large training datasets, and may suffer from overfitting without careful validation.

Physically-based methods employ slope stability models grounded in mechanics and hydrology to assess factor of safety or failure probability. The infinite slope model, assuming planar failure surfaces parallel to the ground surface, is commonly implemented in GIS frameworks due to its relative simplicity and minimal data requirements. More sophisticated approaches employ limit equilibrium or finite element analysis to consider complex slope geometries and failure mechanisms. These methods require detailed characterization of material properties, hydrological conditions, and loading scenarios, typically limiting application to detailed site-specific or small regional assessments (Pack et al., 1998).

Distributed hydrological modeling coupled with slope stability analysis represents the most comprehensive physically-based approach, simulating spatial and temporal variations in soil moisture and groundwater conditions and their influence on stability. These models can simulate response to specific rainfall scenarios or sequences, enabling event-based hazard assessment and early warning applications. However, the intensive data requirements, computational demands, and parameter uncertainty associated with these approaches have limited widespread application (Baum et al., 2010).

VALIDATION AND UNCERTAINTY ASSESSMENT

Rigorous validation is essential for establishing the reliability and appropriate application of landslide hazard maps. Validation involves comparing model predictions against independent landslide data not used in model development, typically through techniques such as random partitioning of inventory data, temporal validation using landslides from different time periods, or spatial validation using landslides from different areas (Chung and Fabbri, 2003).

Receiver Operating Characteristic (ROC) analysis and the area under the ROC curve (AUC) provide standard metrics for assessing predictive performance, evaluating how well models distinguish between landslide and non-landslide locations across the range of predicted probabilities. Success rate curves, comparing predictions against training data, assess model fitting, while prediction rate curves, using independent test data, evaluate true predictive capability. High success rates with low prediction rates indicate overfitting, where models memorize training patterns without capturing generalizable relationships (Frattini et al., 2010).

Uncertainty assessment acknowledges that landslide hazard predictions are inherently uncertain due to incomplete data, model simplifications, and natural variability. Sources of uncertainty include input data quality and resolution, landslide inventory completeness and accuracy, model structure and parameterization, and temporal extrapolation assumptions. Probabilistic approaches, including Monte Carlo simulation and Bayesian methods, provide frameworks for propagating and quantifying uncertainty, producing probability distributions rather than point estimates of hazard (Cascini et al., 2013).

CHALLENGES AND FUTURE DIRECTIONS

Data limitations represent a fundamental constraint on hazard assessment, particularly in developing regions where landslide impacts are often most severe. Comprehensive landslide inventories require intensive effort to compile and may be incomplete or biased toward larger, more accessible, or more damaging events. High-quality spatial datasets on geology, soils, and material properties remain unavailable for many regions,

limiting application of sophisticated assessment methods. Even where data exist, issues of scale compatibility, temporal consistency, and spatial resolution create analytical challenges (Guzzetti et al., 2012).

Temporal prediction remains the most elusive aspect of landslide hazard assessment. While spatial prediction has achieved impressive success through modern statistical and GIS methods, reliable forecasting of when landslides will occur continues to challenge researchers. The complex, non-linear relationships between triggering factors and slope response, combined with the influence of antecedent conditions and site-specific characteristics, make accurate temporal prediction extremely difficult. Even sophisticated physically-based models struggle to predict exact timing of failure, particularly for slow-moving or reactivating landslides (Corominas et al., 2014).

Climate change introduces profound uncertainties into hazard assessment by altering the frequency and intensity of triggering factors. Changing precipitation patterns may increase or decrease landslide hazard depending on regional climate trends, while permafrost degradation in high-latitude and high-altitude regions destabilizes previously frozen slopes. Sea level rise increases coastal landslide activity through toe erosion and elevated groundwater. Existing hazard assessments based on historical data may not adequately represent future conditions under altered climate regimes, necessitating development of non-stationary assessment approaches (Gariano and Guzzetti, 2016).

Scale dependencies affect both the selection of appropriate methodologies and the transferability of results. Assessment approaches effective at local scales may be impractical or inappropriate for regional applications, while regional-scale assessments may lack the detail necessary for site-specific decisions. The optimal scale for analysis depends on the intended application, available data, and the spatial scale of relevant processes. Developing multi-scale frameworks that enable consistent assessment across scales while recognizing scale-specific processes and uncertainties represents an ongoing challenge (van Westen et al., 2008).

Integration of hazard assessment into decision-making processes remains incomplete despite decades of research. Communication gaps between scientists and end-users, uncertainties in hazard predictions, institutional barriers, and competing priorities often limit the practical application of hazard assessments. Developing effective translation mechanisms that convey scientific findings in forms useful for planning and emergency management represents a critical challenge requiring engagement beyond traditional academic research (Calò et al., 2014).

RESULTS AND DISCUSSION

Landslide hazard analysis has evolved into a mature scientific discipline integrating knowledge from multiple fields to address a critical societal challenge. This chapter has examined the conceptual foundations, methodological approaches, and practical applications that collectively constitute modern landslide hazard assessment, while acknowledging persistent challenges and emerging opportunities.

The clear distinction between susceptibility, hazard, and risk provides essential conceptual clarity, recognizing that these related concepts serve different purposes within the broader risk management framework. Susceptibility mapping identifies where landslides are likely based on terrain characteristics, hazard assessment adds critical temporal and magnitude dimensions, and risk analysis incorporates consequences to exposed communities and infrastructure. This hierarchical framework enables appropriate selection of assessment approaches matching available data, resources, and decision-making needs.

Understanding the complex interplay between controlling and triggering factors forms the foundation for both statistical and physically-based assessment approaches. Controlling factors establish baseline stability conditions determined by topography, geology, soil properties, hydrology, and land cover, while triggering factors such as rainfall, earthquakes, and human activities induce failure by exceeding critical thresholds. This conceptual framework guides data collection, model development, and interpretation of assessment results.

GIS technology has revolutionized landslide hazard mapping by enabling integration of diverse spatial datasets, application of sophisticated analytical techniques, and production of spatially distributed assessments across multiple scales. The spectrum of GIS-based methodologies, from qualitative expert assessment through statistical data-driven approaches to physically-based slope stability modeling, provides flexibility to match assessment approaches to available data, required detail, and intended applications. No single approach is universally optimal; rather, method selection should consider the specific context, available resources, and decision-making requirements.

Despite significant progress, substantial challenges remain. Data limitations, particularly in developing regions, constrain assessment reliability and coverage. Temporal prediction remains elusive, limiting our ability to provide precise forecasts of landslide timing. Climate change introduces non-stationarity that challenges traditional assessment paradigms based on historical data. Bridging the gap between scientific assessment and practical application in decision-making processes requires continued attention to communication, uncertainty quantification, and stakeholder engagement.

Looking forward, emerging technologies including advanced satellite monitoring, artificial intelligence, and crowdsourcing offer exciting opportunities to enhance assessment capabilities and expand spatial coverage. Integration of multi-hazard perspectives, development of climate-adaptive assessment frameworks, and advancement of real-time early warning systems represent priority directions for future research and application. As landslide impacts continue to grow with expanding development in hazardous terrain and changing environmental conditions, the importance of rigorous, scientifically-grounded hazard assessment will only increase.

REFERENCES

Atkinson, P.M., & Massari, R. (1998). Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy. Computers & Geosciences, 24(4), 373-385.

Baum, R.L., Godt, J.W., & Savage, W.Z. (2010). Estimating the timing and location of shallow rainfall-induced landslides using a model for transient, unsaturated infiltration. Journal of Geophysical Research: Earth Surface, 115(F3), F03013.

Brabb, E.E. (1984). Innovative approaches to landslide hazard and risk mapping. Proceedings of the 4th International Symposium on Landslides, Toronto, 1, 307-324.

Calò, F., Ardizzone, F., Castaldo, R., Lollino, P., Tizzani, P., Guzzetti, F., Lanari, R., Angeli, M.G., Pontoni, F., & Manunta, M. (2014). Enhanced landslide investigations through advanced DInSAR techniques: The Ivancich case study, Assisi, Italy. Remote Sensing of Environment, 142, 69-82.

Cascini, L., Calvello, M., & Grimaldi, G.M. (2013). Groundwater modeling for the analysis of active slow-moving landslides. Journal of Geotechnical and Geoenvironmental Engineering, 136(9), 1220-1230.

Chung, C.J.F., & Fabbri, A.G. (2003). Validation of spatial prediction models for landslide hazard mapping. Natural Hazards, 30(3), 451-472.

Coe, J.A., Michael, J.A., Crovelli, R.A., Savage, W.Z., Laprade, W.T., & Nashem, W.D. (2004). Probabilistic assessment of precipitation-triggered landslides using historical records of landslide occurrence, Seattle, Washington. Environmental & Engineering Geoscience, 10(2), 103-122.

Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J.P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M.G., Pastor, M., Ferlisi, S., Tofani, V., Hervás, J., & Smith, J.T. (2014). Recommendations for the quantitative analysis of landslide risk. Bulletin of Engineering Geology and the Environment, 73(2), 209-263.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., & Savage, W.Z. (2008). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. Engineering Geology, 102(3-4), 85-98.

Frattini, P., Crosta, G., & Carrara, A. (2010). Techniques for evaluating the performance of landslide susceptibility models. Engineering Geology, 111(1-4), 62-72.

Frigerio, S., Schenato, L., Bossi, G., Cavalli, M., Mantovani, M., Marcato, G., & Pasuto, A. (2014). A web-based platform for automatic and continuous landslide monitoring: The Rotolon (Eastern Italian Alps) case study. Computers & Geosciences, 63, 96-105.

Froude, M.J., & Petley, D.N. (2018). Global fatal landslide occurrence from 2004 to 2016. Natural Hazards and Earth System Sciences, 18(8), 2161-2181.

Gariano, S.L., & Guzzetti, F. (2016). Landslides in a changing climate. Earth-Science Reviews, 162, 227-252.

Gill, J.C., & Malamud, B.D. (2014). Reviewing and visualizing the interactions of natural hazards. Reviews of Geophysics, 52(4), 680-722.

Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1999). Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology, 31(1-4), 181-216.

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., & Ardizzone, F. (2005). Probabilistic landslide hazard assessment at the basin scale. Geomorphology, 72(1-4), 272-299.

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C.P. (2007). Rainfall thresholds for the initiation of landslides in central and southern Europe. Meteorology and Atmospheric Physics, 98(3-4), 239-267.

Guzzetti, F., Mondini, A.C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang, K.T. (2012). Landslide inventory maps: New tools for an old problem. Earth-Science Reviews, 112(1-2), 42-66.

Hungr, O., Evans, S.G., & Hazzard, J. (1999). Magnitude and frequency of rock falls and rock slides along the main transportation corridors of southwestern British Columbia. Canadian Geotechnical Journal, 36(2), 224-238.

Hungr, O., Corominas, J., & Eberhardt, E. (2005). Estimating landslide motion mechanism, travel distance and velocity. Landslide Risk Management, 99-128.

Intrieri, E., Gigli, G., Casagli, N., & Nadim, F. (2013). Brief communication: Landslide Early Warning System: toolbox and general concepts. Natural Hazards and Earth System Sciences, 13(1), 85-90.

Jibson, R.W., Harp, E.L., & Michael, J.A. (2000). A method for producing digital probabilistic seismic landslide hazard maps. Engineering Geology, 58(3-4), 271-289.

Keefer, D.K. (1984). Landslides caused by earthquakes. Geological Society of America Bulletin, 95(4), 406-421.

Malamud, B.D., Turcotte, D.L., Guzzetti, F., & Reichenbach, P. (2004). Landslide inventories and their statistical properties. Earth Surface Processes and Landforms, 29(6), 687-711.

Montgomery, D.R., & Dietrich, W.E. (1994). A physically based model for the topographic control on shallow landsliding. Water Resources Research, 30(4), 1153-1171.

Moore, I.D., Grayson, R.B., & Ladson, A.R. (1991). Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. Hydrological Processes, 5(1), 3-30.

Pack, R.T., Tarboton, D.G., & Goodwin, C.N. (1998). The SINMAP approach to terrain stability mapping. Proceedings of the 8th Congress of the International Association of Engineering Geology, Vancouver, British Columbia, Canada, 1157-1165.

Petley, D. (2012). Global patterns of loss of life from landslides. Geology, 40(10), 927-930.

Popescu, M.E. (1994). A suggested method for reporting landslide causes. Bulletin of the International Association of Engineering Geology, 50(1), 71-74.

Reichenbach, P., Rossi, M., Malamud, B.D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. Earth-Science Reviews, 180, 60-91.

Selby, M.J. (1993). Hillslope Materials and Processes (2nd ed.). Oxford University Press, Oxford.

Sidle, R.C., & Ochiai, H. (2006). Landslides: Processes, Prediction, and Land Use. American Geophysical Union, Washington, DC.

Stark, C.P., & Hovius, N. (2001). The characterization of landslide size distributions. Geophysical Research Letters, 28(6), 1091-1094.

Süzen, M.L., & Doyuran, V. (2004). Data driven bivariate landslide susceptibility assessment using geographical information systems: A method and application to Asarsuyu catchment, Turkey. Engineering Geology, 71(3-4), 303-321.

Tarolli, P. (2014). High-resolution topography for understanding Earth surface processes: Opportunities and challenges. Geomorphology, 216, 295-312.

Terzaghi, K. (1950). Mechanism of landslides. In S. Paige (Ed.), Application of Geology to Engineering Practice (pp. 83-123). Geological Society of America, Berkeley Volume.

Tofani, V., Raspini, F., Catani, F., & Casagli, N. (2013). Persistent Scatterer Interferometry (PSI) technique for landslide characterization and monitoring. Remote Sensing, 5(3), 1045-1065.

van Westen, C.J., van Asch, T.W.J., & Soeters, R. (2006). Landslide hazard and risk zonation—why is it still so difficult? Bulletin of Engineering Geology and the Environment, 65(2), 167-184.

van Westen, C.J., Castellanos, E., & Kuriakose, S.L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. Engineering Geology, 102(3-4), 112-131.

Varnes, D.J. (1984). Landslide hazard zonation: A review of principles and practice. UNESCO, Paris, 63 pp.

Wieczorek, G.F. (1996). Landslide triggering mechanisms. In A.K. Turner & R.L. Schuster (Eds.), Landslides: Investigation and Mitigation (pp. 76-90). Transportation Research Board Special Report 247, National Academy Press, Washington, DC.

Adsorption Study of the System of Waste Foundry Sand/Methylene Blue*

Gültuğ Gamze KUTLU[1]
Sedef DİKMEN[2]
Semra MALKOÇ[3]

1-MSc Degree; Eskişehir Technical University Graduate Education Institute, Department of Environmental Engineering, Environmental Sciences. gultugkutlu@gmail.com ORCID No: 0000-0002-6478-5092.
2-Prof. Dr.; Eskişehir Technical University, Faculty of Science, Department of Physics. sdikmen@eskisehir.edu.tr ORCID No: 0000-0002-6164-4710
3-Prof. Dr.; Eskişehir Technical University, Faculty of Engineering, Department of Environmental Engineering. satik@eskisehir.edu.tr ORCID No: 0000-0002-8092-411X
*This study was derived from Gültuğ Gamze Kutlu's master's thesis.

ABSTRACT

This study investigates the adsorption of cationic dyestuffs, specifically Methylene Blue (MB), onto waste foundry sand (WFS) in a batch system. The research examines adsorption equilibrium, thermodynamics, and mechanisms. Adsorption tests were conducted at three different temperatures (25, 35, and 45 °C) with varying initial concentrations of dyestuff solutions.

MB adsorption onto WFS was best described by the pseudo-second-order kinetic model (r^2=0.993) rather than the Lagergren first-order model (R^2 = 0.647), with a calculated q_e (16.36 mg/g) close to the experimental value, indicating kinetics governed by surface active-site interactions.

Temperature has fewer effects at low MB concentrations. However, at higher concentrations, adsorption capacities at 25-35 °C were generally higher than at 45 °C, and removal efficiency decreased with increasing initial concentration. Equilibrium data were evaluated using Langmuir, Freundlich, and Dubinin–Radushkevich (D-R) models; overall, no model showed a consistently strong fit, but the D-R model provided the best fit at 25 °C and indicated predominantly physical adsorption, showing low average adsorption energies (E=1.707-3.096 kJ/mol). Thermodynamic analysis showed that adsorption was spontaneous ($\Delta G°<0$) and endothermic ($\Delta H°>0$), with increased interfacial disorder ($\Delta S° > 0$) and more favorable adsorption at higher temperatures.

FTIR results supported electrostatic interactions between cationic MB and negatively charged surface sites, while SEM images confirmed surface coating and partial pore filling after adsorption.

Keywords – Adsorption, Waste Foundry Sand, Dyestuff, Isotherm Of Adsorption, Kinetics Of Adsorption.

INTRODUCTION

Environmental pollution is one of the most important problems affecting the health and vital activities of all living beings. The decreasing availability of clean water resources and the increasing variety of industrial processes resulting from population growth, along with the resulting waste, necessitate a search for new and clean water sources or the reuse of existing ones. Undoubtedly, water is an indispensable life source for the survival of living beings, and it is known that the amount of clean water resources is decreasing every day. The textile industry, one of the most important sectors in wastewater generation, has a significant impact on the depletion of clean water resources. Wastewater generated in the textile industry and discharged into water sources severely damages the ecosystems in those water sources. The suspended solids, undissolved particles, heavy metals, and especially the

dyes contained in this wastewater are difficult to treat. Humans have continuously researched the ease and low cost of this treatment process. Waste from foundries, where casting materials used in many areas that facilitate human life are produced, poses a problem for people in terms of disposal. Casting sands are among the most significant of these wastes. In foundries that use sand casting, tons of waste casting sand are generated depending on the production intensity. This study investigated the removal of process waste, which has become a problem for the sector, using the adsorption method. Methylene Blue (MB), which is widely used in the textile industry, was used as the dye, and Foundry System Sand was used as the adsorbent. Adsorption studies were developed using different parameters. Evaluations were made using isotherm, kinetic, and thermodynamic models, and the type of adsorption that occurred was investigated.

Coloured substances used to give objects an aesthetic appearance or to protect them from external factors are called dyes, and the process of applying these substances is called dyeing (Külcü, 2012). Dyes are substances that can give permanent or semi-permanent colour to materials such as fabric, cotton, nylon, fibres, and silk in textiles, and their structure is coloured. The basic properties of dyes in textile dyeing processes are their ability to bind to the fibres and their coloured structure (Ölmez, 1999). Although the concepts of dye and dyestuff are often confused, they differ in terms of effect and structure: dye substances are generally of natural origin and organic, while dyes are mostly insoluble mixtures and inorganic in structure.

Dyes, even at low concentrations, present in wastewater not only detract from the aesthetic appearance of the water but also cause discoloration. They disrupt ecological balance by altering the flora and fauna of the natural environment (Banat et al., 1996; Aksu, 2005; Vijayaraghavan & Yun, 2008; Kaushik & Malik, 2009). Basic (cationic) azo dyes, widely used in the textile industry, are also used in synthetic formulations across many industries. These dyes are reduced by hepatic enzymes in the liver and bacteria in the intestines of mammals. The aromatic molecules formed by this reduction in living organisms are toxic (Namasivayam et al., 1996).

In the textile industry, large volumes of water are consumed in the preparation of dyes, the dyeing of fabrics for production, and the finishing processes of the resulting products. A portion of the water consumed is discharged as wastewater and must undergo treatment before discharge. Increased consumer demand in the textile sector has led to numerous industrial developments and higher production speeds. This has led to the use of synthetic dyes rather than natural dyes in production. These synthetic dyes, which were widely used following rapid production, have led to the generation of highly toxic wastewater (Caner, 1998).

The total volume of wastewater generated from textile dyeing and finishing processes in Turkey is approximately 150 million m³/year (İnce and Tezcanlı, 1999).

Minimizing treatment costs for drinking and irrigation water sourced from rivers and groundwater requires removing dyes and toxic pollutants. Consequently, textile wastewater must undergo treatment before discharge to the environment. A range of treatment methods, including chemical, physical, and biological processes, as well as their combinations, have been developed to eliminate pollutants from textile wastewater. Fig. 1 shows methods for effectively removing dye from textile industry wastewater.

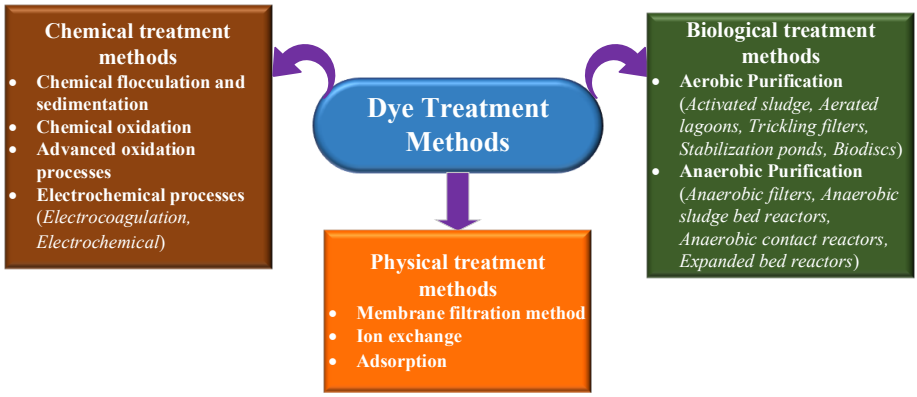


Figure 1: Dye treatment methods used in textile wastewater

Casting is the process of melting metals and metal alloys in furnaces of various designs, adding alloying additions to achieve the desired chemical composition, pouring the resulting mixture into moulds made of sand, metal, ceramic, or polystyrene, and allowing it to solidify to form a product (Bhardwaj &Kumar 2017).

In foundries, silica sand, zircon sand, chromite sand, and olivine sand are some of the types of sand used for sand casting. Although zircon sand, chromite sand, and olivine sand have better technical properties than silica sand, silica sand's higher refractoriness, abundance in nature, and low cost make it the most preferred type of sand. Silica sand is the main component of casting mould sand. Silica sand constitutes approximately 90% of the prepared mould sand. The resulting casting mold is expected to have properties such as creating a smooth surface on the cast parts, being heat-resistant, providing the desired strength, and allowing the gases formed during the casting process to pass through. To ensure all these properties, some additives are added to the prepared moulds along with silica sand (Yasak, 2001; Bhardwaj and Kumar 2017).

LITERATURE REVIEW

To recover the large quantities of waste generated during the casting process, rather than disposing of it, a study was conducted on the adsorption of methylene blue, malachite green, and crystal violet cationic dyes from aqueous solutions of waste casting sand. The adsorption equilibrium, kinetics, and thermodynamics of methylene blue, malachite green, and crystal violet with waste casting sand were investigated. The results show that the adsorption tests conform to the Langmuir Isotherm model. This indicates that the adsorption process of cationic dyes with waste casting sand occurs in a single layer. At 20 °C, the q_{max} values for MB, MG, and CV were 12.38 and 23.44, 8.31 mgg^{-1}, respectively (Gürkan and Çoruh 2018a).

In a study by Jada and Akbour (2014), the column method was chosen for methylene blue (MB) removal using quartz sand due to its easy availability and low cost. The data obtained in the study show the effects of various parameters, such as ionic strength, flow rate, medium pH, and temperature, on the removal of MB dye. It was observed that the amount of MB retained in the porous quartz sand decreased significantly as the flow rate increased. However, a slight decrease in the amount of retained MB occurred when the ionic strength of the aqueous phase increased. It was observed that the amount of MB retained on the quartz increased as the medium temperature or aqueous phase pH increased. However, this decrease was both metal-dependent and metal-independent. The adsorption of MB onto quartz sand particles occurs endothermically and is irreversible.

Firdaus et al. (2020) extracted silica from abundant natural beach sand to remove hazardous dyes from textile-industry wastewater. The method used in the study proved to be fast, simple, and low-cost for adsorbing synthetic dyes. Based on these adsorption experiments, it was concluded that the adsorption of dyes onto silica minerals occurs via electrostatic attraction and van der Waals forces between the dyes and silica. Under the determined optimum conditions, more than 90% of the dyes were adsorbed onto silica.

In another study, a silica-sand/anionized-starch composite (SS-CMS) was simply prepared. CMS-SS has been utilized as an effective adsorbent for removing cationic dyes (methyl blue and crystal violet) and the metal ion [Cu(II)] from water in relevant single and binary systems. Compared with anionized starch without silica sand, CMS-SS achieved significantly higher adsorption capacities of approximately 653.31, 1246.40, and 383.08 mg/g for MB, CV, and Cu (II), respectively (Li et al., 2020).

The chemical composition of fly ash waste from thermal power plants is like that of waste foundry sands. However, the dye adsorption capacities of fly ash remained lower than those reported in other studies using foundry sand. Similarly, industrial waste coal gangue was also unsuccessful in removing various dyes (Yener et al., 2011). Table 1 presents the dye adsorption capacities of various adsorbents found in the literature. This table

also compares the adsorption capacities of clays, frequently used as adsorbents in dye removal, with those of other silicon-based waste materials.

Table 1: Adsorption Capacities of Various Adsorbents (Kutlu 2021)

Adsorbent	Dyestuff	Adsorption Capacity (mg/g)	References
Waste Foundry Sand	Methylene Blue	12.4	(Gürkan and Çoruh, 2018a)
	Malachite Green	28.4	
	Crystal Violet	8.31	
Waste Foundry Sand	Malachite Green	23.3	(Gürkan and Çoruh, 2018b)
Bentonite	Malachite Green	179	(Bulut et al., 2008)
Ca-Bentonite	Congo Red	107	(Lian et al., 2009)
Bentonite	Malachite Green	50.2	(Yıldırım, 2017)
	Methylene Blue	78.8	
Thermal Power Plant Fly Ash	Acid Violet 7	7.00	(Yener et al., 2011)
	Basic Blue 41	22.0	
	Acid Violet 17	18.0	
	Basic Yellow 45	10.0	
Thermal Power Plant Fly Ash	Methylene Blue	19.6	(Aigbe et al., 2021)
Industrial waste coal gangue	Cationic Blue X-GRRL	2.17	(Zhoua et al., 2019)
	Cationic Red X-5GN	1.04	
	Crystal Violet	47.5	
	Malachite Green	40.2	
	Basic Fuchsine	70.3	

ADSORPTION ISOTHERMS

Graphs showing the change in the amount of adsorbate with initial concentration at a constant temperature are called adsorption isotherms. Adsorption isotherms are used to quantify the relationship between the amount of adsorbate per unit mass of adsorbent (mg/g) and the initial concentration (mg/L). In short, mathematical equilibrium is explained using adsorption isotherms (Sarıkaya, 2005). Many isotherm models are used in adsorption studies, along with the Langmuir, Freundlich, Dubinin-Radushkevich (D-R) and Temkin isotherm models.

Langmuir isotherms model

According to the Langmuir isotherm model, adsorption occurs as a single-layer, homogeneous, physical process (Ho and Wang, 2004). The Langmuir isotherm is one of the most important relationships used in interpreting chemical adsorption and equilibrium adsorption behaviour for many cases, and in determining the total surface area on solid surfaces (Karaoğlu, 2007).

The Langmuir isotherm model is expressed as follows:

$$q_e = \frac{q_m\, a_L\, C_e}{1 + a_L\, C_e} \quad (1)$$

Q_e: Amount of substance adsorbed onto a unit adsorbent (mg/g)

C_e: Equilibrium concentration of the substance remaining in the solution after adsorption (mg/L)

a_L: Langmuir isotherm constant (L/mg), which depends on the adsorption energy

q_m: Single-layer adsorption capacity (mg/g)

K_L: Langmuir constant related to adsorption capacity (L/mg).

substance

Here, q_m is the maximum adsorption capacity of the adsorbent. This isotherm does not clarify the equilibrium state in heterogeneous adsorption systems. To assess adsorption suitability, the RL separation factor is calculated. A constant between 0 and 1 indicates the existence of a suitable isotherm model (Şahan, 2007).

$$R_L = \frac{1}{1 + K_L\, C_0} \quad (2)$$

Freundlich isotherm model

The Freundlich isotherm has derived the following mathematical equation to explain the adsorption of various adsorbent molecules from the solution phase:

$$q_e = K_F . C_e^{\frac{1}{n}} \quad (3)$$

K_F: Freundlich constant (L/mg)

C_e: Equilibrium concentration of the substance remaining in the solution after adsorption (mg/L)

Q_e: Amount of substance adsorbed onto a unit adsorbent (mg/g)

n: Defined as the isotherm parameter expressing the adsorption density.

If the logarithm of both sides of this equation is taken to make it linear:

$$\ln q_e = \ln K_F + \frac{1}{n} \ln C_e \quad (4)$$

the equation is obtained. By plotting the $logq_d$ and $logC_d$ graph, a straight line is obtained. The slope of this line is 1/n. $LogK_F$ is the ordinate intersection point. The heterogeneity factor 1/n is in the range of 0-1 and takes values close to zero depending on the heterogeneity of the surface. From this, the isotherm constants n and K_F can be calculated (McKay et al., 1985; Attard and Barnes, 1998).

Dubinin-Radushkevich (D-R) iso therm model

This isotherm model is used for adsorptions occurring on homogeneous and heterogeneous surfaces in the low concentration range. Multilayer adsorption occurs on ions bound to the most ideal regions in terms of energy (Ho, 1999).

The nonlinear and linear equations of the D-R isotherm model are shown in Eq. 6 (Mathematically expressed as shown below (Ho, 1999; Apak, 1997):

$$\ln q_e = \ln q_m - \beta\varepsilon^2 \quad (5)$$

β: Constant related to the average free energy of adsorption per 1 mole of adsorbed substance ($mol^2\ J^{-2}$)

q_e: Amount of substance adsorbed onto a unit adsorbent (mol $g^{-}1$)

q_m: Single-layer adsorbent capacity (mol g^{-1})

Ɛ: Polanyi potential (mol KJ^{-1})

The value of Ɛ, which is the Polanyi potential, can be calculated with the following equation:

$$\varepsilon = RT\ln(1 + \frac{1}{C_e}) \quad (6)$$

R: Ideal gas constant (R=8.314 $Jmol^{-1}K^{-1}$)

T: Absolute temperature (K).

ADSORPTION KINETICS

The rate of the adsorption process is determined by which step of the adsorption process is used, and the appropriate adsorbed substance-adsorbent contact time (retention time) is obtained by determining the adsorption kinetics. Analysis of kinetic data is necessary to understand the factors that affect the rate of the adsorption mechanism (Temkin and Pyzhev, 1940).

Lagergren first-order kinetic model

Lagergren developed the first-order kinetic model, one of the most widely used models in the adsorption kinetics of a component dissolved in an aqueous solution onto an adsorbent. This developed model is expressed as follows:

$$\frac{1}{q_t} = \left(\frac{k_1}{q_{max}}\right)\left(\frac{1}{t}\right) + \frac{1}{q_{max}} \quad (7)$$

If the graph of 1/t to $1/q_t$ is plotted in the above equation, the line with the slope value k1/qmax is obtained (Müjdeci, 2019).

Pseudo-second-order kinetic model

The pseudo-second-order equations, based on the adsorption capacity of the solid phase, were derived by Ho and McKay. It is a model that encompasses all stages of the adsorption process. The model is expressed as follows:

$$\frac{1}{q_t} = \frac{1}{k_2\, q^2{}_{max}}\, t \quad (8)$$

When the graph of t against t/q_t in this equation is plotted, a line with a slope of $1/q_{max}$ and an ordinate intercept of $1/k_2q^2{}_{max}$ is obtained (Müjdeci, 2019; Sadegh et al., 2015).

Intraparticle diffusion model

In the intraparticle diffusion model (Weber-Morris Model), the rate is

$$q_t = k_p\, t^{1/2} + C \quad (9)$$

where;

C: intercept

q_t: amount of substance adsorbed on a unit adsorbent at time t ($mg.g^{-1}$)

$t_{1/2}$: half-life ($min^{1/2}$)

k_p: intraparticle diffusion rate constant ($mg.g^{-1}\ min^{-2}$) is defined.

ADSORPTION THERMODYNAMICS

In adsorption studies, temperature-dependent parameters are the most important indicators in determining the type and level of interactions between the adsorbed and the adsorbent, and in explaining the adsorption mechanism (i.e., physical or chemical) (Tran et al., 2016). Chemical and physical adsorption can be determined using parameters such as enthalpy change (ΔH^o), standard Gibbs free energy change (ΔG^o), and entropy change (ΔS^o). A negative ΔH^o value indicates that the adsorption process is exothermic, while a positive ΔH^o value indicates that the process is endothermic (Zhou et al., 2012). Generally, when the ΔH^o is less than 40 kJ.mol^{-1}, absolute physical adsorption is observed, while when it is in the range of 80-200 kJ.mol^{-1}, chemical adsorption is observed (Zhou et al., 2012; Erdem, 2010). A negative ΔG^o ($\Delta G^o<0$) indicates suitability and spontaneous occurrence of the system, while a positive ΔG^o ($\Delta G^o>0$) indicates that the system is not spontaneous and is unsuitable. Therefore, the ΔG^o is an important factor in determining the degree of spontaneousness of the process. In general, the ΔG^o ranges from −20 to 0 kJ.mol^{-1} for physical adsorption and from −80 to −400 kJ.mol^{-1} for chemical adsorption (Ammendola et al., 2017; Erdem, 2010). The sign and magnitude of the ΔS^o provide information about whether the distribution of the adsorbate at the solid/liquid interface during the adsorption process is less random ($\Delta S^o<0$) or more random ($\Delta S^o>0$) (Ammendola et al., 2017).

The relationship between thermodynamic parameters for the adsorption process is as follows:

$$\Delta G^o = \Delta H^o - \Delta S^o \quad (10)$$

ΔG^o: Standard Gibbs free energy change (kJ/mol)
ΔH^o: Standard enthalpy change (kJ/mol)
T: Absolute temperature (K)
ΔS^o: Standard entropy change (kJ/molK)

$$K_C = \frac{C_a}{C_e} \quad (11)$$

The equilibrium constant Kc is calculated using Eq. 11. In here,
K_c: Equilibrium constant,
C_a: Concentration of adsorbed substance at equilibrium (mg/L),
C_e: Concentration of substance remaining in solution at equilibrium (mg/L).

By plotting K_c, which is found with Eq. 11, against C_e, K_c is found and the Gibbs free energy of adsorption is obtained by putting it into equation (12).

$$\Delta G^o = -RT \ln K_C^o \quad (12)$$

Here, R is the gas constant (8.314 J/mol K).

MATERIALS AND METHODS

Materials

The waste foundry sand used as an adsorbent in the experiments was obtained from a foundry casting area operating in Eskişehir. The samples were ground into powder in a laboratory mortar. The samples were sieved through a 125 μm Fritsch sieve to prepare them for adsorption experiments.

The methylene blue used in the experiments was 99% pure and sourced from Merck. It was used to prepare the dye solutions without any purification. Its chemical formula is $C_{16}H_{18}ClN_3S$.

1.0 g of Methylene Blue (Sigma) was weighed, dissolved in deionized water, and diluted to 1000 mL to prepare a 1000 mg/L stock solution. The 25-50-100-250 and 500 mgL-1 solutions used in determining the adsorption isotherms were prepared by diluting the stock solution.

Characterization

In this study, XRF, XRD, BET, FTIR, and SEM were used to characterize waste foundry sand that had lost its functional properties after casting.

A Bruker S8 Tiger model XRF spectroscope was used to determine the metal oxide composition of the waste material used in the experiments. Samples were prepared for analysis by the melting method. For this purpose, approximately 1.0 g of sample was mixed with 9.0 g of lithium tetraborate and melted in a platinum crucible at 1200 °C to form a glass disc. The oxide percentage components of the sample were determined.

XRD analyses were performed using a Bruker D8 Advance model X-ray diffractometer with CuK_α radiation.

The specific surface area of waste foundry sand was determined using a Quantachrome Nova 2200 Surface & Pore Size Analyzer operating according to the BET method with N_2 gas adsorption at 77 K. Before analysis, the sample was left overnight in an oven at 100 °C to remove moisture and then subjected to degassing for 2 hours in the measuring device. The specific surface areas of the sample were determined using the N_2 gas adsorption technique in a liquid nitrogen environment (–196 °C).

To determine the relevant functional groups before and after the adsorption of MB onto the surface of the waste sample, Fourier Transfom-Infrared (FT-IR) spectra of the samples were obtained in the 4000-400 cm^{-1} wavelength range using the ATR (Attenuated Total Reflectance) mode (Perkin Elmer Spectrum 100) of an FT-IR spectrophotometer. Additionally, to improve the clarity of the surface morphology of the sample, they were coated with Au-Pd under a voltage of 5 kV for 60 seconds using a Zeiss Ultra Plus Scanning Electron Microscope (SEM), and then SEM images were taken before and after adsorption at different magnifications.

Adsorption Studies

In this study, waste foundry sand was used as an adsorbent. The adsorption of Methylene Blue (MB) onto this selected waste material was investigated using the batch method. In batch adsorption studies, after determining the optimal adsorbent amount (adsorbent dose) and contact time, the effects of initial concentration and temperature on the adsorption capacity (amount of dye adsorbed) were investigated. In the second stage, the optimum contact time was determined based on the optimum adsorbent amount obtained. In the third stage, the effect of different initial concentrations on the adsorption capacity was investigated. In the final stage, adsorption experiments were conducted at three temperatures to assess the impact of temperature on adsorption capacity.

Adsorption experiments were performed in a temperature-controlled shaking incubator (Zhicheng ZHWY-2102C). After the adsorption process was completed, samples were taken from the liquid phase using a pipette, diluted as needed, and the concentration was measured using a UV-Vis spectrophotometer (Shimadzu UV-Vis 2600). Based on these values, adsorbent capacities (q_e) were calculated using the Eq. 13.

$$q_e = \frac{C_0 - C_e}{m} \times V \tag{13}$$

Here, C_0 and C_e represent the initial, and equilibrium amounts of methylene blue in the solution (mgL^{-1}), V represents the solution volume (L), and m represents the mass of the adsorbent (g).

The percentage removal efficiencies of methylene blue from aqueous solutions onto WSS were calculated using Eq. 14.

$$Removal\ efficiency\ (\%) = \frac{C_0 - C_e}{C_0} \times 100 \tag{14}$$

- *Optimum adsorbent amount*

To determine the adsorbate-adsorbent ratio, adsorption efficiency, and adsorption capacity, the adsorbent amount was varied at a constant initial concentration (C_0 = 50 mg/L). For this purpose, adsorption experiments were carried out by weighing 0.025, 0.050, 0.075, 0.100 and 0.125 g samples into 25 mL dye solutions in Erlenmeyer flasks, with a solid/liquid ratio in the range of 1-5 g/L. For these experiments, the contact time was kept constant at t = 60 min, the medium temperature at T = 25 °C and the stirring speed at 125 rpm.

- *Optimum contact time*

0.050 g of WSS (to ensure a solid/liquid ratio of 2 g/L) was weighed and added to 25 mL of MB solution with a constant initial concentration C_0=50 mg/L in flasks. Then, the mixtures were stirred at 125 rpm at a constant temperature of 25 °C for the specified contact times. After the adsorption time was complete, the flasks were removed from the shaker,

samples were drawn using a pipette, and the concentration was measured using a UV spectrophotometer.

- *Initial concentration*

In the first two stages of the experimental studies, two critical parameters, namely the optimum adsorbent amount and contact time, were investigated, and their effect on dye adsorption, i.e., the optimum conditions for MB adsorption, was determined. Based on the determined optimum conditions, the adsorption capacities and efficiencies of MB were investigated using dye solutions of 25, 50, 100, 250, and 500 mg/L, prepared by diluting a 1000 mg/L stock MB solution. 0.050 g of waste foundry sand was weighed and placed in Erlenmeyer flasks, to which MB solutions at the specified concentrations were added. The mixture was stirred for the specified optimum time. Then, the Erlenmeyer flasks were removed from the stirrer, and the absorbance values of samples taken from the liquid phase by pipetting were measured using a UV-Vis spectrophotometer. Adsorption capacities were calculated according to these concentrations and isotherm plots were obtained.

- *Adsorption isotherms at different temperatures*

Finally, adsorption isotherm studies were carried out by adsorbing MB solutions with initial concentrations ranging from 25-500 mg/L at three different temperatures (25 °C, 35 °C, and 45 °C) and at an adsorbent dose of 2 g/L. 0.05 g of adsorbent (waste foundry sand) was added to erlenmeyer flasks containing MB solutions (25 mL), and the flasks were placed in a temperature-controlled shaker. After setting the desired temperature, the mixtures were stirred at 125 rpm for the specified contact times. Then, a small amount of sample was drawn from the liquid phase through the Erlenmeyer flasks using a pipette, and the concentrations of MB remaining in the solution were determined using a UV-Vis spectrophotometer.

The adsorption amounts of MB were calculated using the Freundlich, Langmuir, and D-R isotherm models. Isotherm constants were calculated from the equations and slopes of the graphs, and graphs were plotted showing the amount of MB adsorbed by 0.05 g of adsorbent versus the MB concentration remaining in the solution phase at equilibrium.

RESULTS AND DISCUSSION

Characterization

The metal oxide composition obtained from the XRF analysis of the WFS is given in Table 2. The waste casting sand removed from the mould after casting contains burnt bentonite, burnt coal dust, resin from core sand, silica sand, iron, and nutrient residues. The specific surface area of WFS is 6.29 g/m^2.

Table 2: Metal Oxide Compound of WFS

Metal oxide	Mass (%)	Metal oxide	Mass (%)
SiO_2	89.48	MgO	0.254
Al_2O_3	3.304	P_2O_5	0.037
Fe_2O_3	1.359	SO_3	0.687
K_2O	0.295	TiO_2	0.246
CaO	0.469	L O I	3.866

The XRD pattern of the WFS is shown in Fig. 2. In the XRD graphs of the foundry sand, which is a mixture containing many minerals, the most intense peaks (approximately 2θ=21°, 26.7°, and 50°) belong to the main component, SiO_2. According to the XRD pattern, additional peaks indicate that, in addition to SiO_2, oxide compounds such as Al_2O_3 and Fe_2O_3 are present in the waste sand used as adsorbents. This is due to other residues such as bentonite, burnt coal dust, resin from core sand, and iron.

Fig. 3 shows the FTIR spectrum of cast core sand casting moulds (casting sands separated from burnt core sands), referred to as WFS. The 3417 cm^{-1} and 1546 cm^{-1} ranges represent the stretching and bending vibrations of the surface hydroxyl groups, respectively.

The absorption bands at 490 cm^{-1} and 461 cm^{-1} in the spectrum correspond to the Si–O–Al and Si–O–Si bending vibrations, respectively. Sharp absorption bands appear at 777 cm^{-1} and 1082 cm^{-1}. These correspond to the symmetrical and asymmetrical stretching vibrations of the Si–O bonds, respectively. The band at 520 cm^{-1} is due to the presence of Al–O stretching vibration. At 3417 cm^{-1}, a peak characteristic of the deformation of the O–H bond is observed. Typical deformations for C–H bonds are known to occur in the 3000 cm^{-1} region for hydrocarbons (Mastella, 2014). Deformation of C–O bonds, however, was observed in the 1709 cm^{-1} region (C=O). These results suggest the presence of aliphatic and formaldehyde-like structures in WFS. The presence of these organic compounds can be explained by the mixing of machine oils from the casting line with the casting sand during the moulding process. Furthermore, the spectrum

confirms the presence of silica (SiO_2) originating from burnt bentonite and silica sand within the WFS.

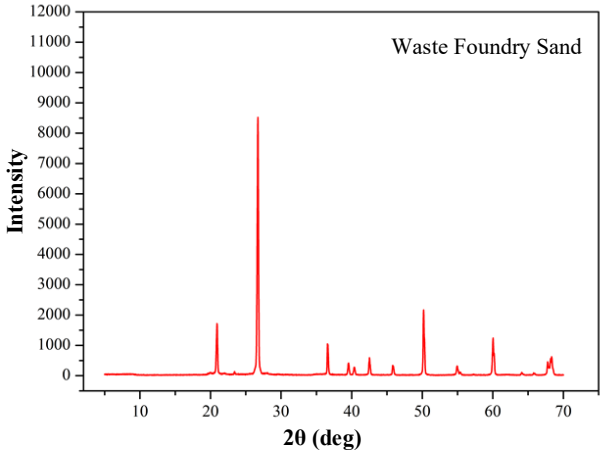


Figure 2: XRD Analysis of WFS

The image of the waste foundry sand obtained from the examination with Scanning Electron Microscopy (SEM) is given in Fig. 4. When the SEM image of the waste foundry sand, taken at 10 K× magnification, is examined, layered structures are observed due to the presence of bentonite. These results are consistent with other characterization techniques such as XRD (phase analysis) and XRF (chemical analysis).

The optimum adsorbent amounts

The amount of adsorbent is an important parameter when evaluating the efficiency and economic aspects of the adsorption process. To determine the optimum adsorbent dose and optimize adsorption conditions for MB WFS adsorption, studies were conducted at an initial concentration of 50 mg/L and adsorbent doses between 1-5 g/L, and the results obtained are given in Fig. 5.

Optimum contact times

In Fig. 6, the MB adsorption capacity on WFS increases immediately from the start of the experiment. The capacity reaches its maximum value of 20.7 mg/g after 15 minutes. Subsequently, at 60 minutes, the adsorption capacity values do not change. Therefore, the optimal contact time for MB is 60 minutes.

Initial concentration

In experiments conducted to investigate the effect of initial concentration (C_0) on adsorption capacity (q), the graph plotted using five different initial concentrations of MB is given in Fig. 7. The adsorption

capacity value increased rapidly from approximately 12 mg/g to approximately 23 mg/g at a concentration of 50 mg/L, decreased to approximately 22 mg/g at 100 mg/L and 250 mg/L, and increased again at 500 mg/L, reaching its highest value (approximately 26 mg/g). The percentage removal values reached their highest values of approximately 98% at 25 mg/L and 50 mg/L and then decreased with increasing concentration.

Adsorption capacity at different temperatures depending on initial concentration

In MB adsorption on WFS at different initial concentrations, the optimal adsorbent dose of 0.05 g was used, and the contact times determined in the corresponding experiments were used. Under these defined conditions, the change in adsorption capacity across medium temperatures (25, 35, and 45 °C) was investigated. The resulting graphs are given in Fig. 8.

Fig. 8 shows the MB adsorption capacity graphs at three temperatures as a function of the initial concentration. No effect of temperature is observed at low MB concentrations, whereas differences in adsorption capacity are observed as the initial concentration increases. It was observed that the adsorption capacities at 25 and 35 °C were higher than those at 45 °C (the highest temperature) across the different concentrations. The removal efficiency reached its highest value at approximately 95% at 25°C, approximately 97% at 35°C, and approximately 98% at 45°C, and then decreased with increasing concentration.

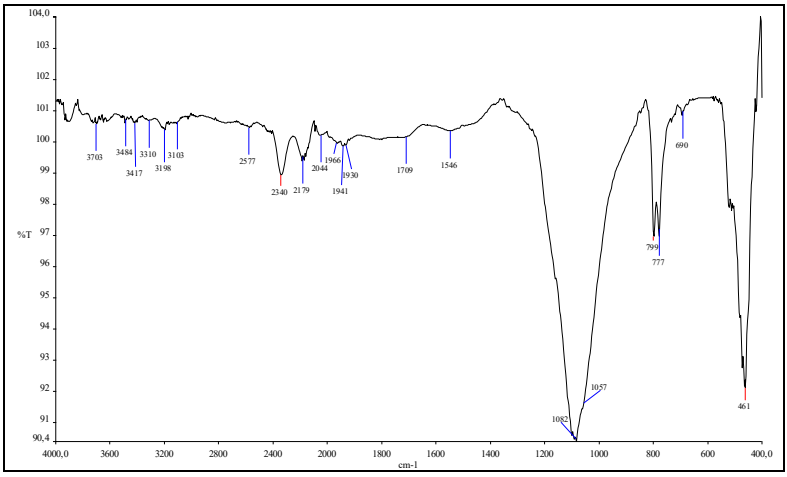


Figure 3: FTIR Spectra of WFS.

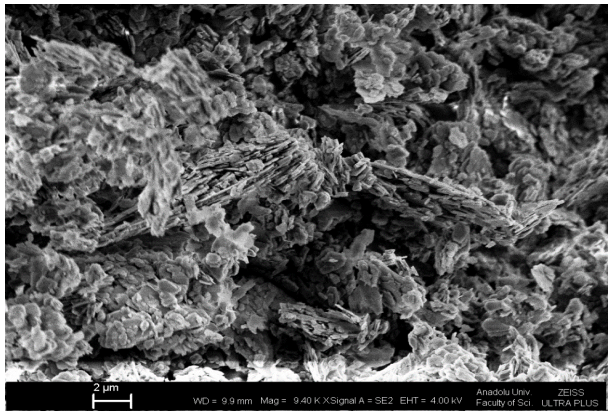


Figure 4: SEM Images of WFS

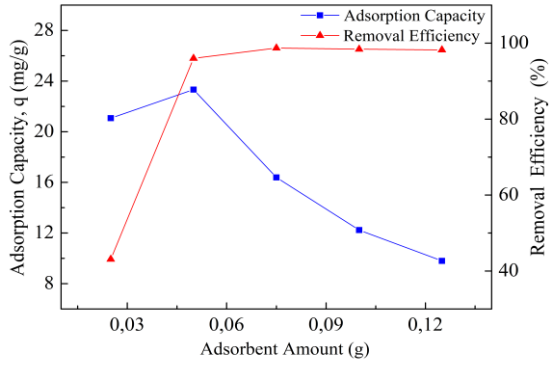


Figure 5: Adsorption Capacity and Removal Efficiency of WFS for MB Removal as a Function of Adsorbent Amount (C_0 = 50 mgL^{-1}, V=25 mL, t=60 min, T=25 °C).

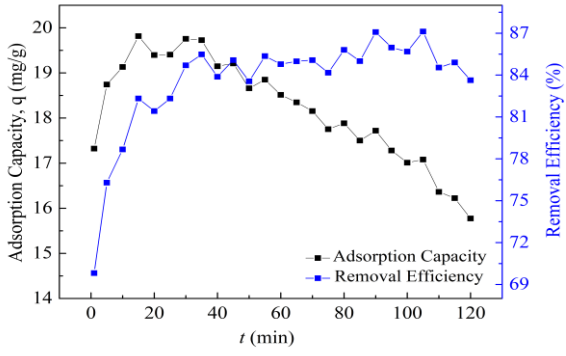


Figure 6: Time-depend Changes in the Adsorption Capacity (mg/g) and Removal Efficiency (%) of MB onto WFS (C_0=50 mgL^{-1}, V=25 mL and T=25 °C)

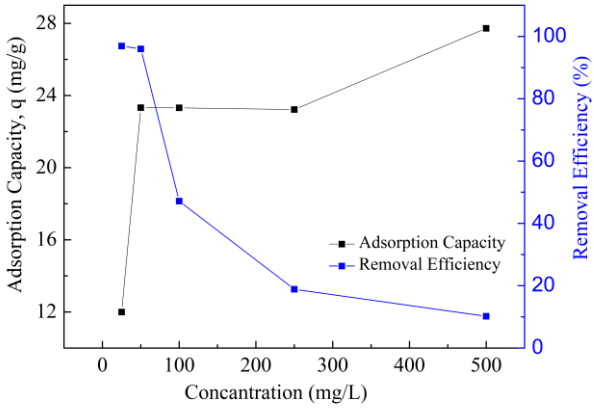


Figure 7: Adsorption Capacity and Removal Efficiency of MB by WFS as a Function of Initial Concentration (t=60 min; T=25°C)

Kinetic and Isotherm Models of Adsorption

- *Lagergren first-order kinetic model*

The first-order Lagergren rate equation was applied to the experimental data obtained from the adsorption of MB onto WFS. Since the r^2 (correlation coefficient) values obtained from the lnq1-qt graph plotted against t are considerably less than 1, this graph is not included here. However, the kinetic parameters calculated from the slopes and intersection points of the lines obtained from this graph are given in Table 3.

- *Pseudo-second-order kinetic model*

Fig. 9 shows the t-versus-t/q_t plots for the rate expression (Eq. 8), using q_t (mg g^{-1}) data from MB adsorption on WSF. As shown in Table 3, a high correlation coefficient (0.993) was obtained for MB adsorption on WSF, indicating suitability for the pseudo-second-order kinetic model. The agreement between the adsorption amounts (q_2) calculated using this model and the experimental data ($q_{max,e}$) also supports this situation. The agreement with the pseudo-second-order kinetic model suggests that chemical adsorption may be the rate-determining step in the process.

Table 3: Adsorption kinetic parameter of MB dye on WFS

Lagergren -first-order			**Pseudo-second-order**		
k_1 (min^{-1})	**q_e (mg g^{-1})**	**r^2**	**q_e g mg^{-1} min^{-1}**	**k_2 mg g min^{-1}**	**r^2**
0.00144	19.79	0.647	16.36	0.0155	0.993

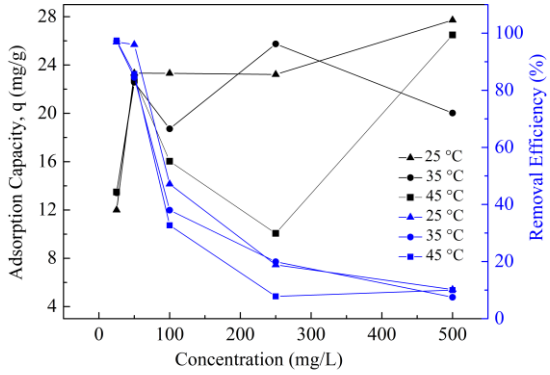


Figure 8: Effect of Temperature on the Adsorption of MB at Different Initial Concentrations onto WFS (t=60 min)

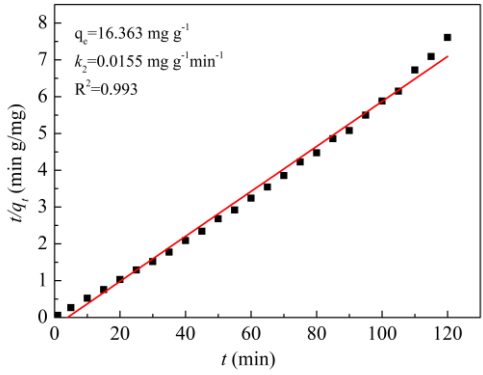


Figure 9: Pseudo-second-order Kinetics of MB adsorption onto WSF.

Adsorption isotherms of MB at different temperatures

In this part of the study, the variation of MB adsorption capacities on WFS with equilibrium concentration at different temperatures and determined optimum conditions was investigated. Isotherm graphs are given in Fig. 10. Examining the isotherm curves clearly shows that a very unstable adsorption process occurs. WFS fine fraction may contain metal oxides, bentonite residue, lime components and binder resin particles. This can cause the liquid phase remaining at the top to be turbid if not adequately filtered. The release of ions/organic matter into the solution can alter the pH of the medium. Because the adsorption of MB, a cationic dye, is highly sensitive to pH and surface charge, it can induce point fluctuations. MB primarily adsorbs on the surface of pure silica with silanol groups, but its low adsorption capacity and heterogeneous surface make reproducibility difficult. Even the most minor defects disrupt the isotherm.

- *Langmuir Isotherm Model*

The Langmuir isotherm is a model used to describe monolayer adsorption on a surface. The curve of 1/Ce versus $1/q_e$ has a slope of $1/q_m$, and its extrapolation yields a straight line with slope $1/q_mK_L$. The Langmuir isotherm is shown in Fig. 11. The isotherm parameters K_L (L/mg) and q_m (mg/g) are calculated from the slopes and shifts of this graph.

Isotherm data and corresponding correlation coefficients (r^2) are presented in Table 6. Examining the values in Table 6, it is evident that the adsorption process does not conform to the Langmuir isotherm model, as the correlation coefficients between MB and WFS are not close to 1 at any temperature. As is known, this isotherm model indicates that adsorption occurs in a single layer and that the adsorbent surface is homogeneous. Based on these results, it is reasonable to conclude that MB adsorption does not occur in a single layer.

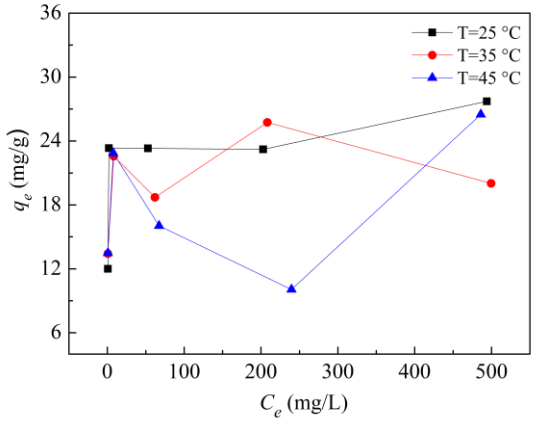


Figure 10: Adsorption isotherm of MB onto WFS at different temperatures.

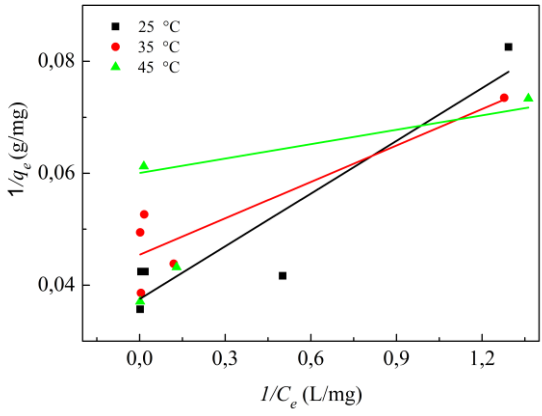


Figure 11: Langmuir isotherm model for MB adsorption onto WFS at different temperatures.

- *Freundlich Isotherm Model*

Another isotherm applicable to adsorption data is the Freundlich isotherm, which suggests that multilayer adsorption may occur rather than single-layer adsorption. If the $\ln C_e$ and $\ln q_e$ values are plotted using Eq. (4), the slope of the resulting line is equal to 1/n. The ordinate intersection point gives $\ln K_F$. Accordingly, because linearity was not observed when applying the Freundlich adsorption isotherm model to the experimental MB adsorption data, the corresponding graphs are not included in this section. However, the obtained Freundlich isotherm parameters and the r2 values for the isotherm lines are presented in Table 6. The Freundlich isotherm model does not fit the adsorption of MB on WFS. In general, the Freundlich isotherm is used to explain adsorption on heterogeneous surfaces. According to the Freundlich isotherm model, adsorption is assumed to occur on surfaces with different adsorption energies and characteristics. If the amount adsorbed (q_e) increases rapidly with concentration (C_e), whereas the adsorbent surface fills more slowly, the adsorption process is said to conform to the Freundlich isotherm. High adsorption capacity (K_F) values indicate that the adsorbents used have low capacity, and high adsorption density (n) values may indicate that the adsorption process did not occur well throughout the entire concentration range studied (Acisli et al., 2020). Hong et al. (2009) studied the removal of methylene blue from aqueous solutions by adsorption onto bentonite. Adsorption studies were conducted at 15, 25, 35, and 45 °C using initial concentrations of 50-1000 mg/L methylene blue solutions, an adsorbent dose of 0.1 g/L, and a contact time of 4.5 hours.

- *Dubinin-Radushkevich (D-R) isotherm model*

The D-R equation has been used in many studies to determine whether chemical or physical adsorption occurs from water on solid surfaces. Physical and chemical adsorption can occur simultaneously at solid-liquid interfaces. The D-R isotherm is an empirical adsorption model used to describe the adsorption mechanism, typically assuming a heterogeneous surface and a variable adsorption potential. In Figure 12, the Polonian potential ε ($molKJ^{-1}$) is first calculated using Eq. 6. Then, ε^2 is plotted against $\ln q_e$ using Eq. 5, and the values of K_{DR} ($mol^2\ J^{-2}$) and q_m (mgg^{-1}) are calculated from the slope of the line and the shift of the y-axis, respectively. The D-R isotherm allows differentiation between physical and chemical adsorption using the free sorption energy value (E, J/mol). E values were calculated using the following Eq 15. All obtained values are given in Table 4.

When Table 4 is examined, it is evident that MB adsorption onto WFS at all temperatures does not conform to the D-R isotherm models, as the r^2 values are less than 1.

Table 4: Isotherm parameters calculated for the adsorption of MB onto WFS.

Model	Constants	Temperature		
		25 °C	35 °C	45 °C
q_e, exp (mg/g)				
Langmuir				
	q_m (mg/g)	26.656	22.010	16.651
	K_L (L/g)	1.194	2.094	7.006
	r^2	0.816	0.760	–0.276
Freundlich				
	q_m (mg/g)	26.656	22.010	16.652
	K_F (L/mg)	16.026	15.641	15.608
	n	11.542	15.420	50.075
	r^2	0.512	0.671	0.711
D-R				
	q_m (mg/g)	17.576	12.171	27.119
	E (kj/mol)	1.707	2.135	3.096
	r^2	0.894	0.676	–0.210

$$E = \frac{1}{\sqrt{2K_{DR}}} \quad (15)$$

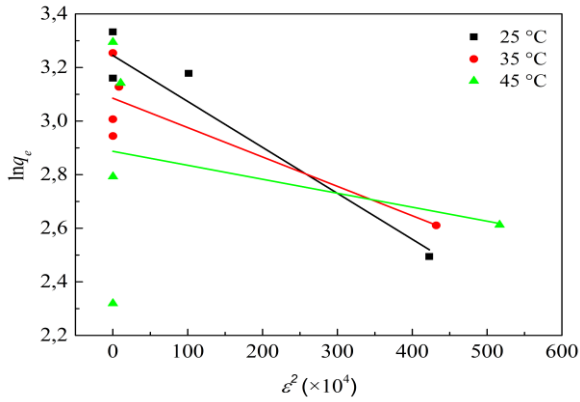


Figure 12: D-R isotherm model for MB adsorption onto WFS at different temperatures.

Adsorption Thermodynamics

Table 5 shows the adsorption thermodynamics values for WFS for the methylene blue dye. For different adsorption temperatures (25, 35, and 45 °C), the ΔG° values are negative (ΔG°<0). This means that adsorption occurs spontaneously at all three temperatures. Furthermore, as the temperature increases, ΔG° becomes even more negative. In other words, the process becomes more favourable at higher temperatures. Since ΔH°=69.40 kJ/mol (ΔH°>0), the process is endothermic (proceeds by absorbing heat). The entropy value (ΔS°=233.41) J/molK is calculated as positive. This indicates

that the disorder of the system increases during adsorption. In dye adsorption studies, this is often explained by reasons such as changes in the structure of the dye at the solid surface-solution interface.

Table5: Thermodynamic parameters for MB adsorption on WFS.

Temperature (°C)	ΔG^o (kJ/mol)	ΔH^o (kJ/mol)	ΔS^o(J/mol-K)
25	–0.4397	69.40	233.41
35	–1.8925		
45	–5.1468		

CONCLUSIONS

The results obtained are as follows:

The factors to consider when determining the optimum adsorbent amount are that it should not cause solid accumulation in the solution and that each particle should be able to easily contact the dye molecule. Based on these evaluations, the optimal adsorbent dose for all dyes was 2 g/L (solid to liquid ratio of 0.05 g/25 mL).

The time-dependent behaviour of MB adsorption was determined by monitoring the contact time between WFS and dye solutions for 1-240 minutes. MB adsorption initially occurs rapidly, and the adsorption rate decreases as equilibrium is approached. The contact time for MB was determined as 60 minutes.

Initial MB concentration was fixed at 50 mg/L, and the effect of adsorbent amount on MB removal was examined by varying the WFS dose from 1 to 5 g/L (0.025-0.125 g). The removal efficiency increased markedly from 43.1% to 96.0% when the dose was raised from 1 to 2 g/L, and then approached a plateau at higher dosages (98.2-98.7% for 3-5 g/L), indicating that additional adsorbent provided limited improvement under the studied conditions. The maximum removal efficiency (98.7%) was obtained at 3 g/L. In contrast, the adsorption capacity (q, mg/g) showed a decreasing trend at higher dosages, consistent with the reduced driving force per unit mass and partial underutilization of available adsorption sites when excess adsorbent is present. Based on these results, 2 g/L can be considered an effective dose to achieve high removal with lower adsorbent consumption.

The adsorption of MB on WFS was evaluated at 25, 35, and 45 °C using the optimum adsorbent dose (0.05 g) and selected contact time (60

min). Temperature had no significant effect at low initial MB concentrations, but expansions in the range became apparent. At higher concentrations, the adsorption capacities obtained at 25 and 35 °C were generally higher than those at 45 °C. For removal change, the highest values were observed at low concentrations (approximately 95% at 25 °C, 97% at 35 °C and 98% at 45 °C), followed by a decrease in initial concentration.

In evaluating the kinetic data, the Lagergren first-order model showed low fit (r^2=0.647) and therefore could not adequately explain MB adsorption. In contrast, the pseudo-second-order model provided high linear fit (r^2=0.993), and the calculated equilibrium adsorption capacity (q_e=16.36 mg/g) was found to be close to the experimental value. These results indicate that the adsorption kinetics of MB on WFS are more consistent with the pseudo-second-order model, and that the rate is controlled by surface interactions and steps related to the active sites.

Equilibrium data regarding the adsorption of MB onto WFS were evaluated using Langmuir, Freundlich, and D-R isotherm models. Overall, all improvements of the models did not provide a strong fit. The Langmuir model showed moderate to weak distributions (r^2=0.816 and 0.760 at 25 and 35 °C, respectively) and an unacceptable fit at 45 °C (r^2=−0.276), suggesting that monolayer adsorption on a homogeneous surface is not a sufficient explanation. Similarly, the improvement of the Freundlich model showed low aspect ratios (r^2=0.512-0.711), indicating that it did not satisfactorily capture the concentration-dependent equilibrium behaviour. The D-R model allowed for optimal fit at 25 °C (r^2=0.894), with the individual's higher performance compromise constraint weakening at 35 °C (r^2=0.676) and weakening at 45 °C (r^2=−0.210). The mean adsorption energy values for D-R were low (E=1.707-3.096 kJ/mol), which is accelerated by adsorption (physical adsorption) where weak accelerations predominate.

When thermodynamic parameters are evaluated, the adsorption of MB onto WFS occurs spontaneously at all temperatures ($\Delta G^\circ<0$); furthermore, $\Delta H^\circ>0$ indicates that the process is endothermic, and $\Delta S^\circ>0$ shows that the disorder at the solid-solution interface increases during adsorption, and adsorption becomes more favourable as the temperature increases.

REFERENCES

Acisli, O., Acar, İ. Ve Khataee, A. (2020). Preparation of a fly ash-based geopolymer for removal of a cationic dye: Isothermal, kinetic and themodynamic studies. Journal of Industrial and Engineering Chemistry, 83, 53-63.

Aigbe, U.O., Ukhurebor K.E., Onyancha R.B., Osibote O.A., Darmokoesoemo H. ve Kusuma H.S. (2021). Fly ash-based adsorbent for adsorption of heavy metals and dyes from aqueous solution: a review. Journal of Materials Research and Technology, 14, 2751-2774.

Aksu, Z. (2005). Application of biosorption for the removal of organic pollutants: A review. Process Biochemistry, 40, 997-1026.

Ammendola P., Raganati F. ve Chirone R. (2017). CO_2 Adsorption on a Fine Activated Carbon in a Sound Assisted Fluidized Bed: Thermodynamics and Kinetics. Chemical Engineering Journal, 322, 302-313.

Apak, R. (1997). Temel Analitik Kimya, ISBN 975404. 372-8, 461-503.

Attard, G. ve Barnes, C. (1998). Surfaces. 1-36, 236. England: Oxford Science Publications Oxford.

Banat, I. M., Nigam, P., Singh, D. ve Marchant, R. (1996). Microbial Decolorization of Textile-Dye-Containing Effluents A Review. Bioresource Technology, 58, 217-227.

Bhardwaj, B. ve Kumar, P. (2017). Waste foundrys and inconcrete: A review. Construction and Building Materials, 156, 661–674.

Bulut, E., Ozacar, M. ve Sengil, I.A. (2008). Adsorption of malachite green onto bentonite: Equilibrium and kinetic studies and process design. Microporous and Mesoporous Materials, 115, 234-246.

Caner, O. (1998). Beyaz Çürükçül Mantar Phanerochaete chrysosporium ile Tekstil Endüstrisi Atıksularının Arıtılması. İstanbul: Doktora Tezi, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü.

Erdem, B. (2010) Heterohalkalı aromatik bileşiklerin doğal killere adsorpsiyonunun incelenmesi. Eskişehir: Anadolu Üniversitesi Fen Bilimleri Enstitüsü, Doktora Tezi.

Firdaus M. L., Madina, F.E., Yulia S., Elvia, R., Ishmah, S. N., Eddy, D. R. and Cid Andres A.P. (2020). Silica extraction from beach sand for dyes removal isotherms: Isotherms, kinetics and thermodynamics. Rasayan Journal of Chemistry, 13 (1), 249-254.

Gürkan, E. H. and Çoruh, S. (2018a). Atık döküm kumu ile katyonik boyarmaddelerin adsorpsiyon çalışmaları. CBÜ Fen Bil. Dergisi, 13 (1), 515-521.

Gürkan, E. H. and Çoruh, S. (2018b). Using waste foundry sand for the removal of malachite green dye from aqueous solutions-kinetıc and equilibrium studies. Environmental Engineering and Management Journal, 123-133.

Ho, Y. S. (1999). Pseudo Second – Order Model for Sorption Processes Process Biochemistry, 34, 451.

Ho, Y.S. and Wang, C.C. (2004). Pseudo-isotherms for the sorption of cadmium ion onto tree fern, Process Biochemistry, 39, 759-763.

İnce, N.H. and Tezcanlı, G. (1999). Treatability of textile dye-bath effluents by advanced oxidation: preparation for reuse. Water Science and Techology, 40 (1), 183-190.

Jada, A. and Ait Akbour, R. (2014). Adsorption and removal of organic dye at quartz sand-water interface. Oil & Gas Science and Technology, 69(3), 405-413.

Karaoğlu, M. H. (2007). Sulu çözeltilerden bazı boyar maddelerin fizikokimyasal yöntemlerle giderimi. Doktora Tezi, Balıkesir Üniversitesi, Fen Bilimleri Enstitüsü, Kimya Anabilim Dalı.

Kaushik, P. and Malik, A. (2009), Fungal dye decolourization: Recent advances and future potential. Environment International, 35, 127-141.

Kutlu, G. G. (2021). Sulu çözeltilerden bazı boyar maddelerin gideriminde atık döküm kumlarının kullanımı. Eskişehir: Yüksek Lisans Tezi, Eskişehir Teknik Üniversitesi.

Külcü, A. (2012). Yüksek Lisans Tezi. Metil viyole içeren sulu çözeltilerden renk giderimi için thamnidium elegans'ın biyosorpsiyon karanteristiklerinin incelenmesi. Eskişehir: Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü.

Li, P., Gao, B. and Yang, H. (2020). Evaluation of the selective adsorption of silica-sand/anionized-starch composite for removal of dyes and cupper(II) from their aqueous mixtures, International Journal of Biological Macromolecules, 149, 1285-1293.

Lian, L., Guo, L. and Guo, C. (2009). Adsorption of congo red from aqueous solutions onto Ca-bentonite. Journal of Hazardous Materials, 161(1), 126-131.

Mastella, M.A., Gislon, S., Pelisser, F., Ricken, C., Silva, L.D., Angioletto, E. and Montedo, O.R.K. (2014) Mechanical and toxicological evaluation of concrete artifacts containing waste foundry sand. Waste Management, 34(8), 1495-1500.

McKay, G., Otterburn, M.S. and Aga, A.J. (1985). Fuller's earth and fired clay as adsorbents for dyestuffs. Water, Air and Soil Pollution. 24, 307-322.

Müjdeci, B. (2019). Sepiyolit destekli aspergillus terreus ile kongo kırmızısı giderimi. Eskişehir: Yüksek Lisans Tezi, Eskişehir Teknik Üniversitesi.

Namasıvayam, C., Muniasamy, N., Gayatri, K., Rani, M. and Ranganathan, K. and (1996). Removal of Dyes from Aqueous Solution by Cellulosic Waste Orange Peel, Bioresource Technology, 57(1), 37-43.

Ölmez, T. (1999). Yüksek Lisans Tezi. Tekstil Endüstrisinde Reaktif Boya Banyolarında Ozon ile Renk Giderimi. İstanbul: İ.T.Ü. Fen Bilimleri Enstitüsü.

Sadegh, H., Shahryari-ghoshekandi, R., Agarwal, S., Tyagi, I., Asif, M., and Gupta, V. K. (2015). Microwave-assisted removal of malachite green by carboxylate functinalized multi-walled carbon nanotubes: kinetics and equilibrium study. Jornal of Molecular Liquids, (206), 151-158.

Sarıkaya, Y. (2005). Fizikokimya. Ankara: Gazi Kitapevi.

Şahan, A. (2007). Farmosetik Maddelerin Aktif Çamur Arıtma Prosesinde Abiyotik Gideriminin İncelenmesi. Adana: Yüksek Lisans Tezi Çukurova Üniversitesi Fen Bilimleri Enstitüsü, Çevre Mühendisliği Anabilim Dalı.

Temkin, M. and Pyzhev, V. (1940). Kinetics of ammonia synthesis on promoted iron catalyst. 327-356. Acta Physicochimica URSS, 12, 327-356.

Tran, H. N., You, S. J. and Chao, H. P. (2016). Thermodynamic parameters of cadmium adsorption onto orange peel calculated from various methods. A

Comparison Study. Journal of Environmental Chemical Engineering, 4(3), 2671-2682.
Vijayaraghavan, K. and Yun, Y. S. (2008), Bacterial biosorbents and biosorption, Biotechnology Advances, 26, 266-291.
Yasak, E. (2001). Kalıp kumlarının özelliklerinin iyileştirilmesi. İstanbul Teknik Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.
Yener, J., Arca, N. and Kopaç T. (2011). Çeşitli boyar maddelerin termik santral uçucu külleri üzerine adsorpsiyonunun incelenmesi. Zonguldak Karaelmas Üniversitesi Fen Edebiyat Fakültesi Kimya Bölümü, Zonguldak.
Yıldırım, A. (2017). Çitosan/kil kompozitlerinin sentezi, karakterizasyonu ve adsorpsiyon uygulamaları. Dicle Üniversitesi, Fen Bilimleri Enstitüsü. Doktora Tezi.
Zhou, X., Yi, H., Tang, X., Deng, H. and Liu, H. (2012). Thermodynamics for the adsorption of SO_2, NO and CO_2 from flue gas on activated carbon fiber. Chemical Engineering Journal, 200- 202, 399-404.
Zhoua, L., Zhoua, H., Hua, Y., Yanb, S., Yang, J. (2019). Adsorption removal of cationic dyes from aqueous solutions using ceramicadsorbents prepared from industrial waste coal gangue. Journal of Environmental Management, 234, 245-252.

Performance Degradation of Photovoltaic Panels Caused by Dust Accumulation: A Case Study in Karabük Region

Melike ADAL ÇELİK[1]
Yunus Emre KARASU[2]

[1]Department of Electrical and Electronics Engineering, Graduate School of Postgraduate Education, Karabük University, Karabük, Türkiye
E-mail: melikeadal@icloud.com
[2]Karabük University, Electrical & Electronics Engineering, Karabuk, Türkiye
E-mail: yekarasu@karabuk.edu.tr
ORCID: 0000-0001-7611-9102

ABSTRACT

The performance of photovoltaic (PV) systems is significantly affected by environmental conditions, among which dust accumulation remains one of the most critical yet region-dependent factors. While numerous studies have investigated the impact of dust on PV module efficiency, most of them focus on arid or desert climates, leaving industrial regions relatively underexplored. Karabük, an industrial city located in the Black Sea region of Türkiye, presents a distinct atmospheric environment characterized by industrial emissions and particulate matter. This study examines the influence of dust accumulation on the electrical performance of photovoltaic modules, with a particular focus on current–voltage (I–V) and power–voltage (P–V) characteristics. The effects of dust on key electrical parameters such as short-circuit current, open-circuit voltage, maximum power output, and fill factor are analyzed based on findings reported in the literature and discussed within the context of Karabük's environmental conditions. The main contribution of this work lies in contextualizing existing dust-related performance degradation studies for an industrial region, highlighting region-specific implications and identifying limitations of current approaches. The results emphasize the necessity of region-adapted maintenance strategies and provide a foundation for future experimental investigations in industrial environments.

Keywords—Photovoltaic Systems, Dust Accumulation, I–V Characteristics, P–V Characteristics, Industrial Regions, Karabük.

INTRODUCTION

The increasing global demand for clean and sustainable energy has accelerated the deployment of photovoltaic (PV) systems across diverse geographical regions. Solar energy is widely recognized as a promising renewable energy source due to its abundance, scalability, and environmentally friendly nature. However, the actual performance of PV systems under real-world operating conditions often deviates from their rated values as a result of various environmental and operational factors.

Among these factors, dust accumulation on the surface of PV modules has been identified as one of the most critical causes of performance degradation. Dust particles deposited on the glass surface reduce optical transmittance, limiting the amount of solar irradiance reaching the photovoltaic cells and leading to deterioration in electrical output. Numerous studies reported in the literature indicate that dust-induced power losses may range from a few percent to over 30%, depending on dust density, exposure duration, and environmental conditions.

Existing research on dust-related photovoltaic performance degradation has largely concentrated on desert and semi-arid regions, where sandstorms and

coarse dust deposition dominate. While these studies provide valuable insights, their findings may not be directly transferable to industrial environments. In industrial regions, airborne particulate matter typically consists of finer particles originating from industrial emissions, vehicular traffic, and combustion processes. Such particles tend to adhere more strongly to module surfaces and may result in different degradation mechanisms compared to desert dust.

Karabük, located in the northern part of Türkiye, represents a typical industrial city characterized by intensive iron and steel production and elevated levels of atmospheric particulate matter. Despite the increasing interest in deploying photovoltaic systems in industrial regions, limited attention has been given to evaluating how industrial dust affects photovoltaic performance. This gap highlights the need for region-specific analyses that account for local environmental conditions rather than relying solely on desert-based findings.

In this context, the present study investigates the influence of dust accumulation on the electrical performance of photovoltaic modules, with particular emphasis on variations in current–voltage (I–V) and power–voltage (P–V) characteristics. By synthesizing findings from existing literature and interpreting them within the environmental framework of the Karabük region, this study aims to provide a region-oriented perspective on dust-induced performance degradation.

The main contributions of this study are as follows: (i) a comprehensive discussion of dust-induced photovoltaic performance degradation with emphasis on industrial environments, (ii) a comparative interpretation of existing literature results in the context of Karabük's environmental characteristics, (iii) an analytical evaluation of changes in I–V and P–V characteristics caused by dust accumulation, and (iv) identification of limitations in current studies and recommendations for future region-specific experimental investigations. Through these contributions, the study seeks to support the development of more effective maintenance and mitigation strategies for photovoltaic systems operating in industrial regions.

FIGURES AND ANALYSIS OF I–V AND P–V CHARACTERISTICS

A. Effect of Dust on I–V Characteristics

The current–voltage (I–V) characteristics provide critical insight into the electrical behavior of photovoltaic modules under varying operating conditions. Figure 1 illustrates a comparison between the ideal I–V curve of a clean photovoltaic module and the curve obtained under dust-affected conditions.

As observed in the figure, dust accumulation leads to a noticeable reduction in the short-circuit current (I_{sc}), primarily due to decreased solar irradiance reaching the active cell surface. The slope of the I–V curve near the short-circuit region is significantly affected, indicating increased optical losses caused by dust deposition.

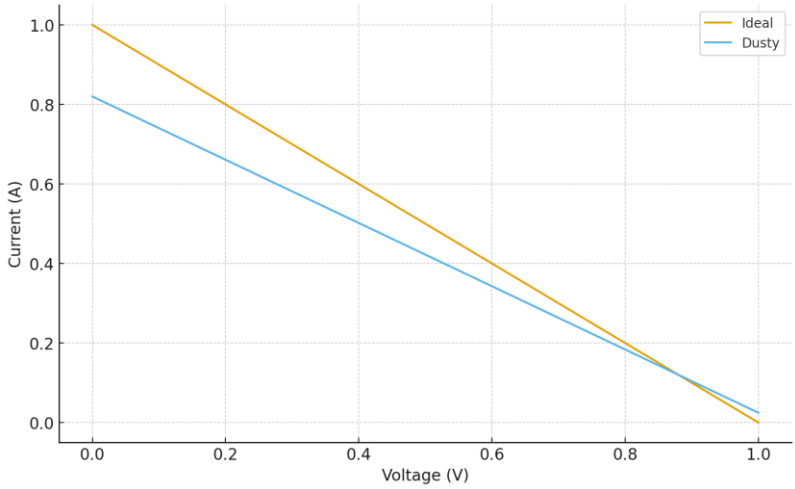


Fig. 1. Ideal versus dust-affected I–V curves

In contrast, the open-circuit voltage (V_{oc}) exhibits relatively minor variation, which is consistent with findings reported in the literature. This behavior suggests that dust predominantly influences current-related parameters rather than voltage-dominated characteristics. Such degradation patterns are particularly critical for systems operating in industrial regions where fine particulate matter adheres strongly to module surfaces.

B. Effect of Dust on P–V Characteristics

The power–voltage (P–V) characteristics reflect the combined impact of dust accumulation on both current and voltage parameters. Figure 2 presents a comparison of P–V curves under clean and dust-affected conditions.

The maximum power point (MPP) is observed to shift towards lower voltage and power levels under dust accumulation, resulting in a significant reduction in maximum power output (P_{max}).

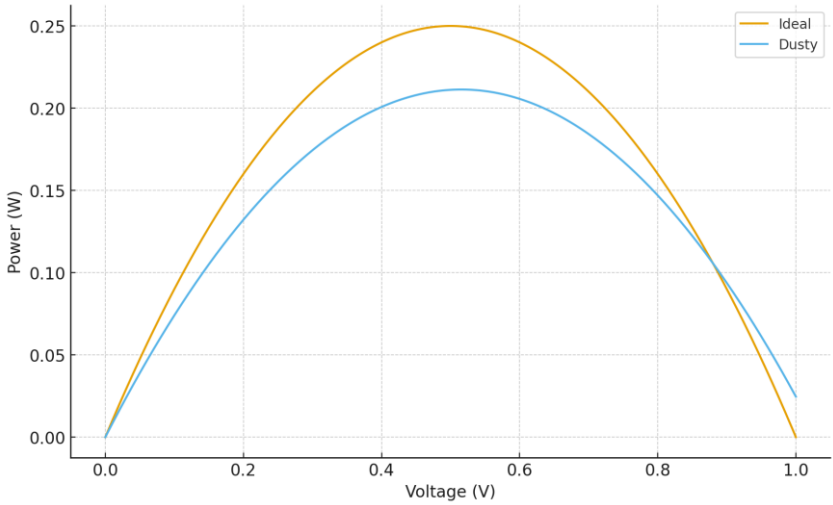


Fig. 2. Ideal versus dust-affected P–V curves

This reduction is attributed to the combined effect of decreased short-circuit current and fill factor degradation. The distortion of the P–V curve highlights the potential for considerable energy yield losses if dust mitigation strategies are not implemented, especially in regions with persistent industrial emissions such as Karabük.

C. Fill Factor Degradation

Dust increases scattering and nonuniform shading on the module surface, leading to fill factor reductions typically in the 3–15% range.

D. Electrical Parameter Degradation

Dust accumulation affects not only optical transmission but also the electrical parameters of photovoltaic modules. The short-circuit current (I_{sc}) is the most sensitive parameter, as it is directly proportional to irradiance. Experimental studies report reductions in I_{sc} ranging from 5% to over 30% depending on dust density.

The open-circuit voltage (V_{oc}) is less affected; however, prolonged dust accumulation combined with increased module temperature may cause slight voltage reductions. Maximum power (P_{max}) experiences compounded losses due to simultaneous reductions in current and fill factor.

Non-uniform dust distribution may result in localized shading, increasing mismatch losses and accelerating degradation mechanisms such as hot-spot formation.

Table I. Electrical Parameter Variations

Parameter	**Clean Module**	**Dust-Affected Module**	**Typical Reduction**
I_{sc}	Nominal	Decreased	5–35%
V_{oc}	Nominal	Slightly decreased	1–5%
P_{max}	Nominal	Significantly decreased	8–30%
Fill Factor	High	Reduced	3–15%

DISCUSSION / COMPARATIVE ANALYSIS

The results discussed in this study demonstrate that dust accumulation significantly alters the electrical performance of photovoltaic modules, with the magnitude and nature of degradation depending strongly on environmental conditions. While literature studies conducted in desert and semi-arid regions report severe power losses due to coarse dust particles, industrial regions such as Karabük present a different degradation profile. Unlike experimental studies conducted under controlled desert conditions, this work provides a literature-based analytical interpretation tailored to an industrial region, offering practical insights for photovoltaic installations in environments similar to Karabük.

In industrial environments, airborne particulate matter typically consists of finer particles originating from combustion processes, vehicular emissions, and industrial activities. These particles exhibit stronger adhesion to module surfaces, leading to persistent optical losses even under moderate wind or rainfall conditions. Consequently, dust-induced performance degradation in industrial regions may occur more gradually but remain continuous over time.

Table II. Environmental Factors Comparable To Karabük

Parameter	Karabük Profile	Effect on PV
Industrial emissions	Medium–High	↑ Dust deposition
Summer rainfall	Low	↑ Deposition
Winter rainfall	Medium–High	↓ Deposition
Vehicular traffic	Moderate	↑ Fine particles

When compared with findings reported by Mani and Pillai [1] and Sayyah et al. [2], the degradation trends discussed in this study align with observed reductions in short-circuit current and maximum power output. However, the industrial context of Karabük suggests that standard cleaning intervals recommended for desert regions may not be directly applicable.

These findings highlight the necessity of region-specific maintenance strategies for photovoltaic installations. In the case of Karabük, more frequent cleaning schedules or the use of anti-soiling coatings may be required to mitigate long-term performance losses. Furthermore, the lack of experimental data specific to industrial regions remains a limitation of the present study and underscores the need for future experimental investigations.

For reference, typical power loss ranges reported in the literature under different dust densities are summarized in Table III. These values are mainly derived from studies conducted in desert and semi-arid regions and are therefore presented here to illustrate general trends rather than exact loss levels for industrial environments such as Karabük.

Table III. Typical Power Loss Levels Due To Dust (From Literature)

Dust Density (mg/m²)	Power Loss (%)
20–40	5–8
50–80	8–15
100–150	15–25
>150	25–35

MITIGATION TECHNIQUES AND MAINTENANCE STRATEGIES

Various mitigation strategies have been proposed to reduce dust-related performance degradation. Periodic manual or automated cleaning remains the most effective method. Literature suggests that cleaning intervals between two and four weeks provide optimal performance recovery in dusty environments.

Hydrophobic and anti-soiling coatings have gained attention due to their ability to reduce particle adhesion. These coatings decrease surface energy, allowing dust to be removed more easily by wind or light rainfall.

Module tilt angle optimization also plays a critical role. Increased tilt angles reduce dust retention but may compromise annual energy yield depending on geographical latitude. Therefore, tilt optimization should balance soiling mitigation and irradiance capture.

Emerging technologies such as robotic cleaning systems and electrostatic dust removal techniques offer promising solutions, particularly for large-scale PV installations.

CONCLUSION AND FUTURE WORK

This study has examined the impact of dust accumulation on the electrical performance of photovoltaic modules, with a particular focus on variations in I–V and P–V characteristics. By synthesizing findings from existing literature and interpreting them within the environmental context of Karabük, an industrial region in Türkiye, the study highlights the region-dependent nature of dust-induced performance degradation.

The analysis indicates that dust accumulation primarily affects current-related parameters, leading to significant reductions in short-circuit current and maximum power output, while open-circuit voltage remains relatively less sensitive. The distortion of I–V and P–V curves under dust-affected conditions demonstrates the potential for substantial energy yield losses if appropriate mitigation strategies are not implemented.

Unlike studies conducted in desert and rural environments, the industrial characteristics of Karabük suggest that finer particulate matter with higher adhesion properties may cause persistent performance degradation over time. This observation emphasizes the importance of adopting region-specific maintenance strategies rather than directly applying findings from arid-region studies.

The present study is subject to certain limitations. The analysis is based on a literature-driven approach and does not include experimental measurements conducted directly in the Karabük region. Additionally, variations in dust

composition, seasonal effects, and long-term exposure conditions were not experimentally evaluated.

Future work should focus on experimental investigations involving real-time performance monitoring of photovoltaic modules installed in industrial regions. Such studies should include detailed characterization of dust composition, deposition rates, and cleaning intervals to develop optimized mitigation strategies. The findings of this study provide a foundation for future region-specific research aimed at improving the long-term performance and reliability of photovoltaic systems operating in industrial environments. Overall, the findings of this study contribute to the understanding of dust-induced photovoltaic performance degradation in industrial regions and support the need for region-adapted maintenance planning. The outcomes are expected to be useful for researchers and practitioners working on photovoltaic system deployment in similar industrial environments.

REFERENCES

[1] M. Mani and R. Pillai, “Impact of dust on solar photovoltaic performance,” Renewable & Sustainable Energy Reviews, 2010.

[2] A. Sayyah, M. Horenstein, and M. Mazumder, “Energy yield loss caused by dust deposition,” Solar Energy, 2014.

[3] S. Kalogirou, Solar Energy Engineering, Academic Press, 2014.

[4] H. Elminir et al., “Effect of dust on transparent covers,” Energy Conversion and Management, 2006.

[5] S. Mekhilef et al., “Effect of dust, humidity, and temperature on PV performance,” Energy Reports, 2020.

[6] J. Kaldellis and M. Kapsali, “Simulating the dust effect on the energy performance of photovoltaic generators,” Solar Energy, 2011.

[7] A. Kimber, L. Mitchell, S. Nogradi, and H. Wenger, “The effect of soiling on large grid-connected photovoltaic systems,” IEEE Journal of Photovoltaics, 2006.

[8] M. Sarver, A. Al-Qaraghuli, and L. Kazmerski, “A comprehensive review of the impact of dust on the use of solar energy,” Renewable & Sustainable Energy Reviews, 2013.

[9] T. Ilse et al., “Techno-economic assessment of soiling losses and mitigation strategies,” IEEE Journal of Photovoltaics, 2019.

[10] H. Sayigh, Solar Energy Engineering, Elsevier, 2018.

Use Of Interlayer In Asphalt Concrete

Murat KARACASU[1]

Ümmügülsüm ÜNAL[2]

1. Prof.Dr. Department of Civil Engineering, Faculty of Engineering and Architecture, Eskisehir Osmangazi University, Campus of Batı Meselik, Odunpazari, Eskisehir, Turkey. muratk@ogu.edu.tr ORCID No: 0000-0001-9721-0984
2. Student. Department of Civil Engineering, Faculty of Engineering and Architecture, Eskisehir OsmangaziUniversity, Campus of Batı Meselik, Odunpazari, Eskisehir, Turkey
gulsumunal128@gmail.com ORCID No: 0009-0006-7023-9393

ABSTRACT

Asphalt concrete pavements are expected to perform well under traffic loads and adverse environmental conditions throughout their expected service life. Asphalt concrete pavements must be strong enough to resist cracking under repeated traffic loads. If the pavementis unable to do so, the pavements tend to deformation, separate and crack. This type of deterioration impacts the pavement's service life, performance, stability and costs of maintenance. The performance of asphalt concrete is determined by the properties of its components, as well as the behaviour and properties of the bitumen and aggregate mixture. This study investigated the effects of using an interlayer in the mixture to improve the performance of asphalt concrete and address these problems The Marshall Design Test method used for this study. The values of bulk specific gravity, void ratio, bitumen filled void ratio and Marshall stability were obtained using single and double layers The results obtained the optimum bitumen ratio.

Keywords – Asphalt Concrete, Flexible Pavement, Marshall Design Method, Layer

INTRODUCTION

Asphalt concrete is constituted of a mixture of aggregate and bituminous binder. The aggregate used in the mixture includes coarse aggregates, fine aggregates,mineral fillers and air voids. Aggregates are prepared in accordance with the gradation limits specified in the standards, and then mixed with bitumen. Aggregate and bitumen have certain physical properties. The aggregates influence the load-bearing capacity of the pavement. They must be resistant to disruption under compression and heavy traffic loads. The bitumen provides adhesion between the aggregate particles, preventing the flexible pavement from disintegrating under traffic loads, while also filling the mixture voids to provide an impermeable layer (Tunç,2001).

In flexible pavement structures, damage caused by traffic loads and adverse environmental conditions can adversely affect performance. Behnood and Gharehveran (2019) have stated in their study that there are three main types of damage that affect the performance of asphalt pavements: rutting, fatigue cracking and thermal cracking. Rutting develops slowly as the load repetition increases. This is the result of the pavement suffering repeated and temporary loading, which causes a reduction in volume and lateral movement in one or more of its layers, leading to surface deformation in the longitudinal direction. Fatigue cracks are a form of damage caused by the repeated occurrence of traffic-related stresses, and in the later stages they lead to alligator skin cracking. Thermal cracks, on the other hand, are formed as a result of abrupt temperature changes and take the

form of closely spaced cracks perpendicular to the road (Albayati, 2023;Li, 1999; Behnood ve Gharehveran,2019).

A reflection crack is a type of crack that appears in new asphalt and is caused by an pre-existing crack in the underneath layer. The stresses in the existing layer are intensified by traffic loads and the effects of daily or seasonal temperature changes, creating reflection cracks. Without another layer to restrict its progression, the crack quickly becomes visible in the upper layer, causing early damage. Therefore, a special layer is required to delay or prevent cracks from reaching the surface (Aschuri ve Yamin, 2011;Hu,2010). Various methods are employed to prevent reflection cracks in conjunction with the use of layers. These include using a stress-relieving membrane interlayer (SAMI), reinforcing the bitumen membrane, increasing the thickness of the bitumen membrane, modifying the bitumen materials, using geosynthetics (e.g. geotextiles, geogrids, geocells, geocomposites and geonets), using cracked plate approaches (e.g. crack and seat, break and seat and rubblization) and using steel reinforcement mesh and surface treatments (Nunn ve Potter, 2014; Dhakal vd., 2016; Deilami ve White 2020). Many studies in the literature evaluate the performance of intermediate layers. Among these methods, many studies have been published in the literature aimed at evaluating the performance of intermediate layers.

Greene et al. (2012) have evaluated the effect of asphalt rubber membrane interlayer (ARMI) and its performance in preventing reflective cracking and in reducing rutting. The behaviour of the coating layer under load and its response to damage could be observed more quickly using accelerated coating tests (APTs). Coatings containing ARMI reached a depth of approximately 12.5 mm 20 to 50 times faster than ARMI-free coatings. Pavements containing ARMI achieved a depth of around 12.5 mm in 20 to 50 times less time than pavements without ARMI. Samples without ARMI proved to be more resistant to reflective cracking than samples with ARMI. Laboratory tests (CSIC tests) revealed that cracking initially occurred within the ARMI and subsequently extended to the surface. It was concluded that using ARMI increases the potential for tyre track formation and does not significantly impact the reduction of reflection cracks. It was concluded that the use of ARMI increases the potential for rutting and also found to have a significant impact on reducing reflection cracks. The use of ARMI was found to have a significant impact on reducing reflection cracks and it was also concluded that it increases the potential for rutting.

Dempsey (2002) evaluated the effectiveness of the ISAC (strain-absorbing composite layer) material in preventing reflection cracks. The creation of a laboratory road section involved the use of a Portland cement concrete (PCC) slab with an asphalt concrete overlay. It evaluated the number of cycles until cracking occurred. The same test conditions, the ISAC system applied to the AC overlay improved performance compared to the control pavement. The use of ISAC resulted in a significant decrease in

pavement stress and a dramatic increase in the number of cycles. Laboratory evaluations have demonstrated the effectiveness of the ISAC system for base isolation. Field performance results also indicated that the ISAC system is effective in reducing reflective cracking in both airport and road pavement systems.

Ogundipe et al. (2013) investigated the performance of the stress-absorbing membrane interlayer (SAMI) under traffic loads. As part of this study, pavements containing and not containing SAMI were constructed and subjected to traffic loading. As part of this study, a pavement including and excluding SAMI was constructed and subjected to traffic loading. According to the findings obtained at the end of the study, SAMI delays the formation of reflection cracks. In addition, the shear rigidity of the interface between the pavement and the SAMI prevents deformations in the cracked lower layer from reaching to the pavement under traffic loads.

Xiao et al. (2024) investigated the effect of placing a welded galvanised steel mesh between the base and surface layers of asphalt pavements on crack formation. Laboratory studies were conducted on a mixture of unreinforced asphalt, glass fibre geogrid, polyester glass fibre geotextile, and welded galvanised steel mesh. This mixture was prepared according to the Marshall design method. The performance of these mixtures was then compared in terms of high- and low-temperature deformation resistance, pressure, flexibility and tensile strength. The reinforced mixtures reduced the amount of rutting and delayed crack development and increased fatigue life compared to the unreinforced mixtures. The high stability and strength of the welded galvanised steel mesh distribute the applied stresses uniformly, consequently increasing the structure's overall deformation and shear resistance. Steel wire mesh displayed the highest dynamic stability, resisting deformation at high temperatures and demonstrating high tensile strength to prevent crack progression. It also resisted cracking at low temperatures and limited the deformation of the aggregate under load, delaying the cracking process and increasing fatigue resistance.

Spadoni et al. (2021) investigated the physical and mechanical properties of four geocomposites, which were made by combining a geosynthetic material with a bituminous membrane. The performance of these geocomposites in asphalt pavements examined in terms of crack propagation and interlayer adhesion. Geocomposites A and D consist of a nonwoven polyester fabric combined with a glass grid; Geocomposite B contains multifunctional glass fibres; and Geocomposite C contains an anisotropic glass grid. The geocomposites were evaluated for crack propagation resistance and interlayer shear strength under static and dynamic loads. According to the laboratory test results, the geocomposites increased crack propagation resistance and delayed reflection cracking from the lower to the upper layer. However, the interlayer shear strength test showed that adhesion between the asphalt layers was negatively affected. Field

applications have also indicated that geocomposites reduce the stability of asphalt pavement layers.

Silva et al. (2023) investigated the impact of various geosynthetic materials on the shear bond strength of asphalt pavement layers. The study examined the effect of four geogrids and one polyester fabric on this strength. The geosynthetics were evaluated based on their covering properties, aperture size and presence, nervure width, and tensile properties. According to the test results, all samples conformed to the minimum requirement of 0.56 MPa ISS. Control samples recorded higher interface shear strength (ISS) values than samples with a geosynthetic interlayer. The presence of geogrid apertures creates interlocking and increases ISS.

Bitumen-coated materials improved ISS performance. However, small geotextile fibres passing through the aperture of the geogrid had an adverse effect on ISS performance.

The adhesion of asphalt emulsion and geosynthetic materials to different pavement layers is the focus of an investigation by Delbono et al. (2014). The study involved the using two emulsions with different melting points and a aperture size 40 mm polyester-based bituminous layer with a polypropylene nonwoven geotextile. The softening and flash points of the materials were determined and their FTIR spectra and SEM images examined. The test results for bituminous and non-bituminous polypropylene samples, as well as samples coated with asphalt emulsion, showed that the properties of the geotextile polypropylene change with temperature when in contact with geosynthetic asphalt emulsion, and that the melting point range increases depending on the emulsion type. According to the test results of bituminous and non-bituminous polypropylene samples and samples coated with asphalt emulsion, when in contact with geosynthetic asphalt emulsion, the polypropylene properties of the geotextile change with temperature and the melting point range increases depending on the emulsion type. When placed between geosynthetic layers, adhesion of the geosynthetic base material increases when it melts; however, discontinuity is observed at the interface when it does not melt.

The effect of geotextile intermediate layers on the bonding properties of asphalt pavements was the investigation of Zhang et al. (2022) in their study. Within the scope of the study, researchers evaluated the effect of various factors on nonwoven polyester geotextiles. These factors include binder type, tack coat types and appication rate, surface paving temperature, exposure period to water and temperature of the enviroment According to the results of the research, the adhesion performance of the geotextile intermediate layer is more affected by water exposure than temperature. At higher temperatures, the bonding layer is more easily processed into the geotextile and a whole is able to be bonded with the asphalt upper layer. However, the performance of the geotextile decreases with increasing exposure to water. As the rate of adhesive layer increases, the performance

of the intermediate layer initially improves but then decreases. When the quantity of adhesive is excessive, the layers slide against each other, which negatively affects the performance of the intermediate layer.

The Marshall Mix Design method was developed by Bruce Marshall of the Mississippi Department of Highways in the 1930s. Marshall indicated that the selected bitumen ratio provided the highest density, and that the stability and flow values also provided the minimum acceptable levels (Asphalt Institute ,2014). The Marshall method is widely used to produce durable, stable and flexible mixtures. It is based on two fundamental principles: density and void analysis, and stability and flow testing (Gartner, 1989). Flow value indicates the mixture's capacity to deformation under external loads, while stability indicates its load-bearing capacity (Xu et al., 2016).

This study was implemented in accordance with the Marshall Mix Design method (TS EN 3720, TS 12697, AASHTO T45, ASTM D1559, D6927 and D5581). After the Marshall test procedures were completed, the optimum bitumen ratio for the mixture was determined by average the values obtained from the equations for practical specific gravity, void ratio, bitumen-filled void ratio, and Marshall stability.

Material and Method

Aggregate

In this study, limestone was used as the aggregate. A selection has been made within the Type-1 gradation limits for the abrasion layers specified in the Republic of Turkey General Directorate of Highways Technical Specification (HTS)Highway Technical Specifications. Figure 1. displays the aggregate particle size distribution curve, while Table 1. describes the aggregate's physical properties.

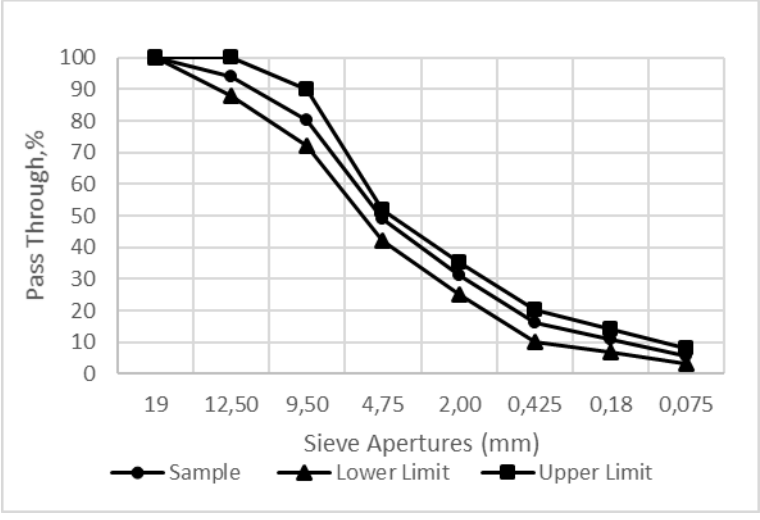


Figure 1. Aggregate Particle Size Distribution Curve

Table 1. Physical properties of aggregate

Proporties		Value	Standard
Specific Gravity			
Course Aggregate	Apparent specific gravity	2.715	
	Specific gravity (g/cm^3)	2.690	ASTM C127
	Saturated surface dry gravity	2.699	
Fine Aggregate	Apparent specific gravity	2.728	
	Specific gravity (g/cm^3)	2.593	ASTM C128
	Saturated surface dry gravity	2.643	
Mineral filler	Specific gravity (g/cm^3)	2.787	ASTM C128
Compacted Bulk Density (g/cm^3)		1.470	ASTM C 29
Loose Bulk Density (g/cm^3)		1.351	ASTM C 29
Los Angeles Abrasion (%)		25.87	ASTM C 131
Flakiness Index (%)		11,67	ASTM D 4791

Bitumen

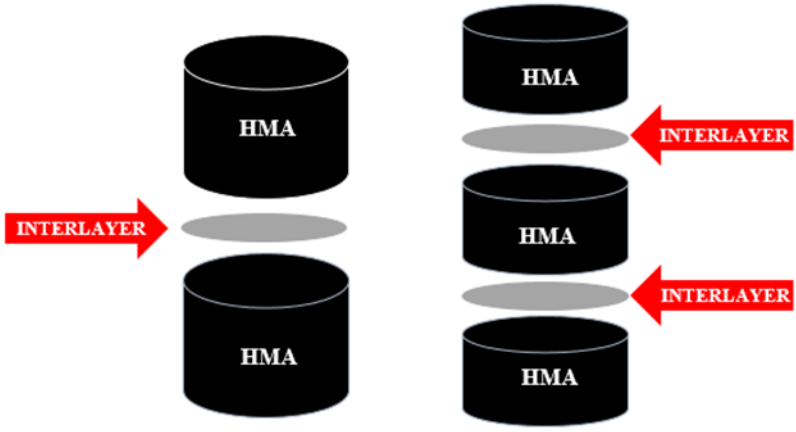
The bitumen with a penetration grade of 50/70 used in the mixture was sourced from the Tupras (Izmit) refinery facilities. The properties of the 50-70 penetration bituminous binder are listed in Table 2.

Table 2. 50/70 Properties of bitumen

Proporties	Value	Standard
Spesific gravity(g/cm^3)	1.026	ASTM D70 09e1
Penetration (25°C)	50/70	ASTM D5
Softening point (°C)	48	ASTM D36/D36M 09
Ductility (25°C)	>100 cm	ASTM D113 07
Loss of Heating (%)	0.05	ASTM D6-95
Flash Point (°C)	301	ASTM D92 05a
Viscosity (at 135°C, cP)	438	ASTM D4402-06
Viscosity(at 165 °C, cP)	135	ASTM D4402-06

Interlayer

Within the scope of this study, intermediate layers were used to divide the asphalt concrete samples into two or three equal parts, in two different positions. These samples were placed in cylindrical moulds with a diameter of 100 mm. The intermediate layers were proportionate to these moulds.



a) **b)**

Figure 2. a) One Layer b) Double Layer

The optimal bitumen ratio for the asphalt concrete samples was determined using the Marshall design method. In the study, 1,150 g of aggregate was prepared according to the gradation criteria set out in the Highway Technical Specifications. Aggregate and bitumen were left in an oven at 160°C for 2-3 hours (Karacasu, 2021). In a heated mixer, a mixture was prepared by adding bitumen to the aggregate at percentages ranging from 3.5% to 6.5% of the aggregate's weight. The mixture is placed in sample moulds with a shovel. For a single-layer specimen, half of the mixture forms the base layer and the other half forms the upper layer. The layer is then placed on the base layer. For the double-layer specimen, the mixture is divided into three equal portions and placed into the moulds to form the base, middle, and top layers. Double-layered asphalt concrete samples are prepared by placing a layer on the base and middle layers. Control samples are prepared using the same aggregate and bitumen ratio. The samples in the moulds are given 75 blows to their upper and lower surfaces with a Marshall hammer and the samples are then left to cool for at least three hours. Sixty-three samples are prepared for each bitumen ratio, with three samples for each ratio. The samples are removed from the mould using a lifting jack. Their height is measured and their weight in air and water is recorded. After standing in a 60 °C water bath for 30–45 minutes, the samples were tested for Marshall strength. The specimens were loaded at a speed of 51 mm/min to obtain the resistance and flow values

The specimens were loaded at a speed of 51 mm/min to obtain the resistance and flow values.

Figure 3. Preparation of Samples

The optimum bitumen ratio was determined by calculating the Marshall stability, void ratio, practical specific gravity, bitumen-filled void ratio, aggregate-interior void ratio and flow values of the control, single-layer and double-layer samples prepared with different bitumen ratios using Excel tables.

4. Test Results and Evaluation

The bitumen ratios that maximise Marshall strength and practical specific gravity, achieve a void ratio of 4% and a void-filled ratio (VFA) of 70%, were determined using graphs. The optimum bitumen ratio is found by taking the average of these four values. The optimum ratio is placed on the bitumen-flow ratio graph to check its suitability within the specification limits (2–4 mm).

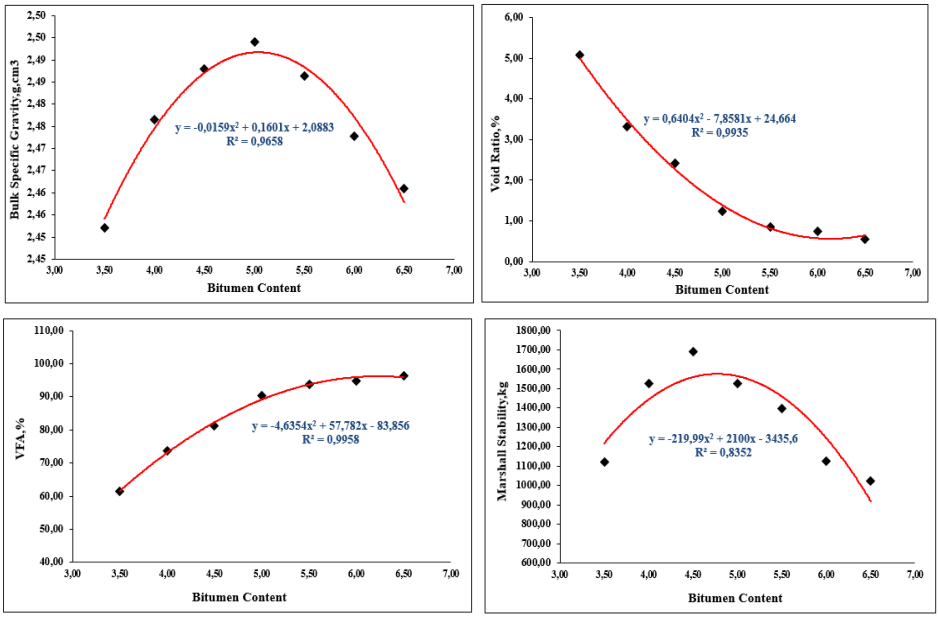


Figure 1. Bitumen content graphs of control samples

The practical specific gravity of a compacted bituminous mixture is defined as the weight of the sample divided by its volume (Texas Department of Transportation, 2007). As the bitumen content increases, the PÖA value initially increases and then decreases. The PÖA-bitumen ratio graph shows the bitumen ratio corresponding to the peak of the curve. In the void ratio–bitumen ratio graph, void percentages decrease as bitumen content increases.

Air voids provide sufficient space for thermal expansion and enable adequate compaction under traffic loads. However, an excessive void ratio reduces the asphalt mixture's resistance to water and fatigue life, while an insufficient void ratio reduces its resistance to rutting (Jiang et al., 2017). When the mixture's void content is in the range of 3–5%, its air and water permeability is generally low. This increases the mixture's durability and reduces the possibility of moisture-related deterioration. A void ratio of %3 or lower leads to poor mix stability and surface bleeding problems, while a void ratio of %5 or higher reduces tensile strength, durability and fatigue life. To maintain stability and avoid exceeding the lower and upper limits, the design asphalt content is typically set at 4% air voids (Quintus and Hughes, 2019).
VFA (void-filled aggregate) is the volume of void space between the aggregate particles in the mixture that is filled with asphalt binder, as a percentage of the mixture's total volume. VMA is the part of the mixture

containing asphalt binder and void space (VFA), representing the effective asphalt content. The VFA value rises as the amount of asphalt binder rises (Materials Manual, 2025; Pratico, 2014).
VFA (void-filled asphalt) is the volume of void space between the aggregate particles in the mixture that is filled with asphalt binder, as a percentage of the mixture's total volume. VMA is the part of the mixture containing asphalt binder and void space (VFA), representing the effective asphalt content. The VFA value rises as the amount of asphalt binder rises (Materials Manual, 2025; Pratico, 2014). In the HTS, the VFA value for the type 2 wearing course is specified as being in the range of 65–75% (HGM, 2013).

Table 3. OBC Value

Variable	**Wearing Course Type -2**		**Bitumen Content %**	**Equation Value of Bitumen Ratio**
	min	**max**		
Bulk Specific Gravity,(g/cm^3)	-	-	5,03	2,49
Void Ratio,%	3	5	3,82	4
VFA,%	65	75	3,85	70
Marshall Stability, kg	900	-	4,77	1576
Flow ,mm	2	4	-	2,86
VMA ,%	14	16	-	12,64
CONTROL AVERAGE			4,37	

The OBC value has been calculated for the bitumen ratio that gives a VFA value of 70%. If the VFA value is below the specified value, the aggregate particles are less likely to bond with each other, resulting in separation of the coating. Conversely, if the VFA value exceeds the specified criteria, the resistance to repeated traffic loads is reduced (Yildirim et al., 2018).
Stability is defined as the resistance to continuous dynamic and static loads, as well as the tensile and deformation forces caused by the acceleration or deceleration of the wheels. Marshall stability is defined as the peak resistance load obtained during loading at a constant deformation rate. The stability and flow results depend on the class and quantity of the binder and the structure of the aggregate (MS-2 Asphalt Mix Design Method, 2014). In HTS, the minimum Marshall stability value for the wearing course types1 and 2 is 900 kg. The bitumen ratio that maximises Marshall stability has also been determined by a formula.

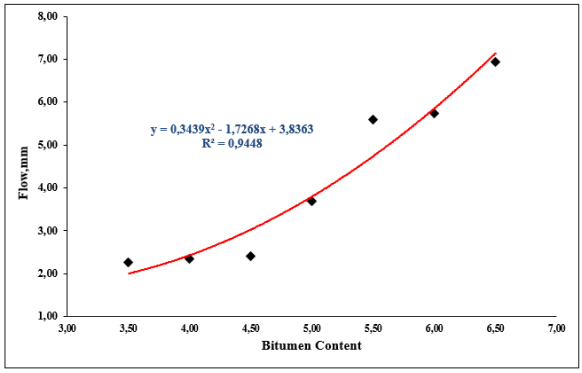


Figure 5. Flow Rate-Bitumen Content curve graph of control samples

The Marshall Quotient (MQ) is calculated as the ratio of stability to flow, providing an approximate indication of the ratio of load to deformation under test conditions. MQ can be used as an indicator of a material's resistance to permanent deformation in service conditions (Hınısoğlu & Agar, 2004).

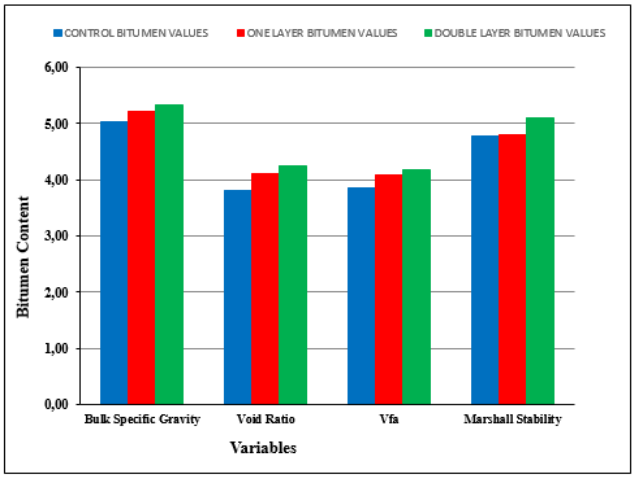


Figure 2. Marshall parameters bitumen ratio values

The bitumen ratios for the practical specific gravity, void ratio, bitumen-filled void ratio and Marshall stability values of the single and double layer samples are shown in Figure 8. The optimum bitumen ratios for these values are shown in Figure 9. The bitumen ratio values for all parameters indicate that the control samples have a lower ratio than the others. The presence of a layer increases the optimum bitumen ratio.

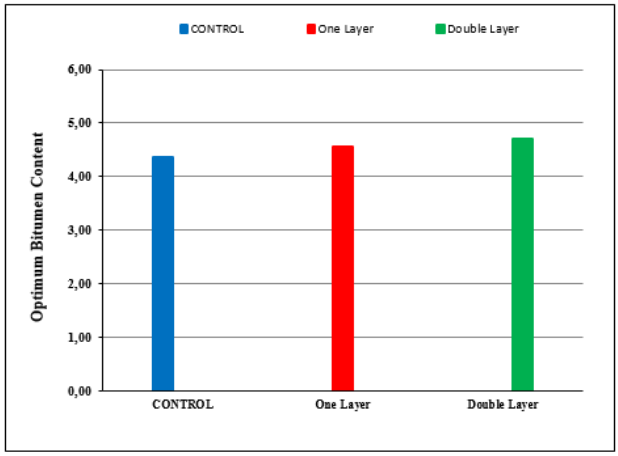


Figure 3.Control, one layer and double layer OBC

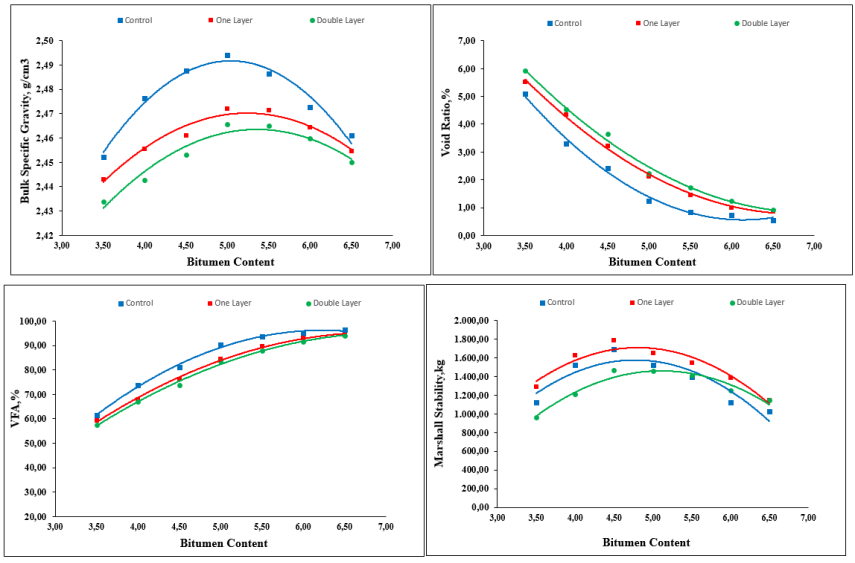


Figure 4. Marshall parameters of the samples

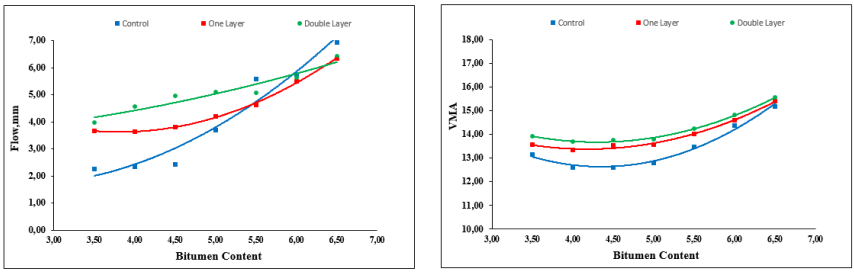


Figure 5. a) Flow –Bitumen Content of the samples b) VMA graph

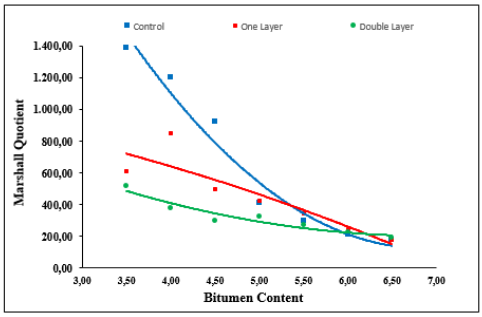


Figure 6. Marshall Quotient parameters of the samples

RESULTS AND DISCUSSION

When control samples and interlayer samples are compared, the following results can be obtained:
* Control samples have lower bitumen content values than the others for all parameters. The presence of layers has increased the optimum bitumen content values.
* In terms of BSG values, control samples have higher values. This is due to the intermediate layers increasing the sample volume. Void ratios have also increased due to the intermediate layer.
* Bitumen-filled void ratios are slightly higher in control samples.
* MS values are high in single-layer samples. Double-layer sample values are low.
* In terms of flow values, single and double-layered samples yielded better results. This is because the shear strength of the samples increased due to the intermediate layers used. This is an important feature.

REFERENCE

Albayati, A. H. (2023). A review of rutting in asphalt concrete pavement. Open Engineering, 13(1), 20220463.
Aschuri, I., & Yamin, A. (2011). The Use of Stress Absorbed Membrane Interlayer (SAMI) to Reduce Reflection Crack on Road Pavement. In Proceedings of the Eastern Asia Society for Transportation Studies Vol. 8 (The 9th International Conference of Eastern Asia Society for Transportation Studies, 2011) (pp. 282-282). Eastern Asia Society for Transportation Studies.
Asphalt Mix Design Methods MS-2(2014). Seventh Edition,USA
Awan, H. H., Hussain, A., Javed, M. F., Qiu, Y., Alrowais, R., Mohamed, A. M., ... & Alzahrani, A. M. (2022). Predicting marshall flow and marshall stability of asphalt pavements using multi expression programming. Buildings, 12(3), 314.

Behnood, A., & Gharehveran, M. M. (2019). Morphology, rheology, and physical properties of polymer-modified asphalt binders. European Polymer Journal, 112, 766-791.
Deilami, S., & White, G. (2020). Review of reflective cracking in composite pavements. International journal of pavement research and technology, 13(5), 524-535.
Dempsey, B. J. (2002). Development and performance of interlayer stress-absorbing composite in asphalt concrete overlays. Transportation Research Record, 1809(1), 175-183.
Delbono, H. L., & Giudice, C. A. (2014). Adherence in a pavement rehabilitated with a polymeric grid used as interlayer. Construction and Building Materials, 54, 454-459.
Dhakal, N., Elseifi, M. A., & Zhang, Z. (2016). Mitigation strategies for reflection cracking in rehabilitated pavements–A synthesis. International Journal of Pavement Research and Technology, 9(3), 228-239.
Gartner, W. (1989). Asphalt concrete mix design: Development of more rational approaches (Vol. 1041). ASTM International.
Greene, J., Choubane, B., Chun, S., & Kim, S. (2012). Effect of asphalt rubber membrane interlayer (ARMI) on instability rutting and reflection cracking of asphalt mixture. Florida Department of Transportation.
Karayolları Genel Müdürlüğü, 2013. Karayolları Teknik Şartnamesi, Ankara.
Hınıslıoğlu, S., & Ağar, E. (2004). Use of waste high density polyethylene as bitumen modifier in asphalt concrete mix. Materials letters, 58(3-4), 267-271.
Hu, S., Zhou, F., & Scullion, T. (2010). Reflection cracking–based asphalt overlay thickness design and analysis tool. Transportation Research Record, 2155(1), 12-23.
Jiang, Y., Xue, J., & Chen, Z. (2017). Influence of volumetric property on mechanical properties of vertical vibration compacted asphalt mixture. Construction and Building Materials, 135, 612-621.
Karacasu, M. (2021). Yol Üstyapı Deney Tasarımları ve Superpave. Nobel Akademik Yayıncılık.
Kuloglu, N. (1999). Effect of astragalus on characteristics of asphalt concrete. Journal of Materials in Civil Engineering, 11(4), 283-286.
Li, Y. (1999). Asphalt pavement fatigue cracking modeling. Louisiana State University and Agricultural & Mechanical College.
Materials Manual (2025). Washington State Department Of Transportation Materials Manual M 46-01.47 .March.
Nunn, M. E., & Potter, J. F. (2014). 33 ASSESSMENT OF METHODS TO PREVENT REFLECTION CRACKING. Reflective Cracking in Pavements: State of the Art and Design Recommendations, 360.
OgundipeO.M., Thom,N.H. & Collop,A.C. (2013) Evaluation of performance of stress-absorbing membrane interlayer (SAMI) using accelerated pavement testing, International Journal of Pavement Engineering.
Pratico, F. G., Vaiana, R., & Moro, A. (2014). Dependence of volumetric parameters of hot-mix asphalts on testing methods. Journal of materials in civil engineering, 26(1), 45-53.
Silva, M. P. S., Santos, K. M., & Correia, N. S. (2023). Interface shear bond analysis of different geosynthetic paving interlayers. In Geosynthetics: Leading the Way to a Resilient Planet (pp. 2153-2158). CRC Press.

Spadoni, S., Ingrassia, L. P., Paoloni, G., Virgili, A., & Canestrari, F. (2021). Influence of geocomposite properties on the crack propagation and interlayer bonding of asphalt pavements. Materials, 14(18), 5310.
Texas Department of Transportation (2007).Determining Density Of Compacted Bitüminous Mixtures. Tex–207–F. Texas, USA.
Tunç, A., (2001). Yol Malzemeleri ve Uygulamaları, Atlas Yayınevi, İstanbul, 840s
Von Quintus, H. L., & Hughes, C. S. (2019). Design & Construction of Heavy-Duty Pavements. Quality Improvement Series 123.
Xu, B., Chen, J., Zhou, C., & Wang, W. (2016). Study on Marshall Design parameters of porous asphalt mixture using limestone as coarse aggregate. Construction and Building Materials, 124, 846-854.
Xiao, Q., Zhao, X., Li, Z., & Zhu, M. (2024). Performance evaluation of welded galvanized steel wire mesh reinforced asphalt pavements. Construction and Building Materials, 455, 139178.
Yıldırım, Z. B., Karacasu, M., & Okur, D. V. (2018). Atık lastik ve cam lif ile modifiye edilmiş bitümün asfalt betonu performansına etkileri. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 18(3), 1019-1027
Zhang, H., Sun, B., Li, Y., & Li, Y. (2022). Adhesive property and road performance evaluation of asphalt overlay pavement with geotextile interlayer. Advances in Materials Science and Engineering, 2022(1), 3084668.

Vibrational Characteristics of Multi-Layered Sandwich Plates Featuring Hybrid Auxetic Core and FG Face Layers

Kerim Gökhan AKTAŞ[1]

Fatih PEHLİVAN[2]

1- Assoc. Prof.; Karabük Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi Makine Mühendisliği Bölümü. kerimgokhanaktas@karabuk.edu.tr ORCID No: 0000-0002-8076-6799

2- Assoc. Prof.; Karabük Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi Makine Mühendisliği Bölümü. fatihpehlivan@karabuk.edu.tr ORCID No: 0000-0003-2675-6124

ABSTRACT

This document contains the free vibration behavior of a multilayer hybrid sandwich plate structure with various auxetic core geometries and functionally graded (FG) surface layers. The proposed configuration consists of five different layers: two FG surface layers, two different inner auxetic core layers, and a central layer separating the cores. Three unique hybrid design combinations, star-4/re-entrant (SR), honeycomb/star-4 (HS), and re-entrant/honeycomb (RH), were evaluated. While the inner core layers and central plate are composed of pure nickel, the outer coatings are designed as Functionally Graded Materials (FGM) transitioning from nickel (Ni) at the inner interface to silicon nitride (Si_3N_4) at the free surfaces following a power-law distribution. Free vibration analysis was performed using the Finite Element Method (FEM) to obtain the first six natural frequencies of the system. This research systematically examines the effect of various critical parameters, including power-law index, temperature effects, boundary conditions, geometric configurations, and aspect ratios, on the dynamic response. This study expands upon sandwich panel analyses, which are generally limited to homogeneous core structures in the literature, by revealing the synergistic effect of hybrid auxetic arrangements and functionally graded surfaces. The findings will serve as a reference, particularly in the aerospace industry for fuselage structures under high thermal loads, in the defense industry for energy-dissipating armor systems, and in precision mechanical engineering for optimizing components requiring vibration isolation.

Keywords –Auxetic structures, functionally graded material (FGM), free vibration, finite element analysis (FEA), hybrid sandwich panels, natural frequency.

INTRODUCTION

In the design of modern engineering structures, sandwich plate systems have become very common due to their properties such as lightweight, high specific rigidity, and multifunctional performance (Kamble, 2024; Tarlochan, 2021). These structures are widely used, particularly in aerospace engineering, defense industry, marine, automotive, and precision mechanical systems, because they exhibit superior performance under vibration, shock, and thermal loads (Jang & Roh, 2025; Palomba et al., 2022; Prusty et al., 2025). The dynamic behavior of sandwich structures is determined by the material properties and geometries of the surface and core layers (Jin et al., 2016; Nilsson & Nilsson, 2002). Accurate modeling of this behavior is critical for reliable and optimized designs.

Free vibration analysis is a fundamental tool in determining the dynamic characteristics of sandwich structures. Natural frequencies and mode shapes form the basic inputs for many engineering applications such as resonance risk assessment, noise and vibration control, and fatigue life estimation (Kihm et al., 2018; Pelayo et al., 2015; Ren et al., 2021). Therefore, material distribution, core architecture, boundary conditions (BC), and environmental effects that influence the free vibration behavior of sandwich plate systems are extensively researched in the literature (Burlayenko & Sadowski, 2020; Cho, 2022; Guo et al., 2022).

In recent years, the use of auxetic cellular structures in place of foam cores has attracted attention. Auxetic materials are structures with a negative Poisson ratio, characterized by their expansion rather than contraction in the transverse direction under tension (Carneiro et al., 2013; Prawoto, 2012). This unusual mechanical behavior offers auxetic cores high energy absorption capacity and enhanced vibration damping potential (Parhi & Roy, 2025; Sebaq & Liu, 2024). Different auxetic geometries have been investigated both theoretically and experimentally by researchers, and significant advantages of these structures have been demonstrated (Alomarah et al., 2020; Farshbaf et al., 2025; Najafi et al., 2021, 2022).

However, many studies have focused on sandwich structures with a single type of auxetic core geometry (Aktas et al., 2024; Eroğlu et al., 2025; Han et al., 2022). Yet, different auxetic geometries exhibit significantly different behaviors in terms of stiffness, mass distribution, and wave propagation characteristics (Dong et al., 2025; Zhang et al., 2020). This suggests that dynamic performance can be tuned more effectively if multiple auxetic core types are used within the same sandwich structure to create a hybrid architecture. However, studies on the free vibration behavior of hybrid configurations where auxetic cores are combined in successive layers are extremely limited in the literature.

On the other hand, the use of Functionally Graded Materials (FGMs) in the surface layers of sandwich structures is widely investigated to reduce stress concentrations caused by abrupt material transitions and improve thermomechanical performance. FGMs are generally metal-ceramic based composites (Esen et al., 2025; Patel, 2025). This concept offers significant advantages, especially in structures exposed to high temperature gradients

(Demirbas, 2017). The free vibration and buckling behavior of FGM surface layers, modeled using power law, exponential and sigmoidal functions, has been addressed by many researchers using classical plate theory, first and higher-order shear deformation theories, and the Finite Element Method (Daikh & Zenkour, 2019; Ebrahimi & Rastgo, 2008; Garg et al., 2021; Mashat et al., 2014; Radaković et al., 2020; Sah & Ghosh, 2021).

However, studies on sandwich structures using FGM layers and auxetic cores together have mostly been limited to the assumption of a single layer auxetic core (Mahesh, 2022; Pham et al., 2023; Thang et al., 2024). In particular, the effect of the interaction of FGM surface layers with hybrid core systems consisting of different auxetic geometries on the free vibration characteristics has not been comprehensively explored in the literature. Furthermore, the effects of thermal environments, power law index, boundary conditions, and geometric parameters on the dynamic response of such multilayer hybrid structures are an open area of research.

This section aims to address the literature gaps by comprehensively investigating the free vibration behavior of a multilayer hybrid sandwich plate structure consisting of two functionally graded surface layers, two different auxetic core layers, and a central layer separating these cores. Three unique hybrid auxetic core combinations, indented/star-4, star-4/honeycomb, and honeycomb/indented, are considered. The core layers and central plate are made of pure nickel, while the surface layers are modeled in an FGM structure with a power law distribution varying from nickel on the inner surfaces to silicon nitride on the free surfaces. Free vibration analysis was performed using the Finite Element Method, and the first six natural frequencies of the system were obtained. The study also systematically analyzed the effects of power law index, temperature rise, boundary conditions, aspect ratios, and hybrid core arrangement on dynamic behavior. This work offers a fresh perspective on the sandwich plate literature and serves as an important reference, particularly in the design of lightweight structures operating at high temperatures and in advanced engineering applications requiring vibration isolation.

MATHEMATICAL FORMULATIONS

Figure 1 presents an illustration of the multi-layered hybrid sandwich plate design. The design consists of three different core layer combinations star-4/re-entrant (SR), honeycomb/star-4 (HS), and re-entrant/honeycomb (RH) and metal-ceramic FGM composite surface layers. The physical parameters of the structure are symbolized by the symbols l_x, l_y, and h. As depicted in Figure 2, h_{face}, h_{aux}, and h_r stand for the thicknesses of the surface layers, auxetic layers, and rim layer, respectively. The total thickness of the plate can be calculated with $h=2h_{aux}+2h_{face}+h_r$. Additionally, h_1, h_2, h_3, h_4, h_5, and h_6 indicate the distance of each plate surface from the center of the sandwich plate ($h_1=-h_r/2-h_{aux}-h_{face}$; $h_2=-h_r/2-h_{aux}$; $h_3=-h_r/2$; $h_4=h_r/2$; $h_5=h_r/2+h_{aux}$ *and* $h_6=h_r/2+h_{aux}+h_{face}$).

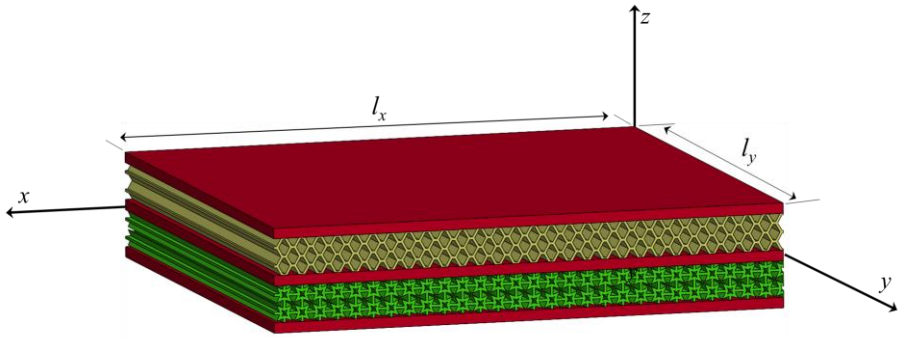


Figure 1: Illustration of a multi-layered hybrid sandwich plate design

The composition of FGM metal-ceramic surface plates varies functionally. The exterior surfaces of FGM plates are composed of pure ceramic (Si_3N_4), while the interior surfaces are composed of pure metal (Ni). The auxetic core layers and rim layer are composed of pure Ni. The material properties of FGM surface plates, which are dependent on the power law index, can be calculated using the method given below (Das et al., 2025; Kardooni et al., 2022):

$$C_{eff}(\mathrm{z}) = \left(C_{Si_3N_4} - C_{Ni}\right)\left(\frac{z - h_r/2 - h_{aux}}{h_{face}}\right)^d + C_{Ni}\,; \qquad h_5 \le z \le h_6 \tag{1}$$

$$C_{eff}(\mathrm{z}) = \left(C_{Si_3N_4} - C_{Ni}\right)\left(-\frac{z + h_r/2 + h_{aux}}{h_{face}}\right)^d + C_{Ni}\,; \quad h_1 \le z \le h_2 \tag{2}$$

in which C_{eff}(z) represents the effective substance characteristics of the upper and lower FGM layers. C_{Si3N4} and C_{Ni} denote the mechanical or thermal characteristics of pure Si_3N_4 and Ni, respectively. Additionally, *d* denotes the power law parameter.

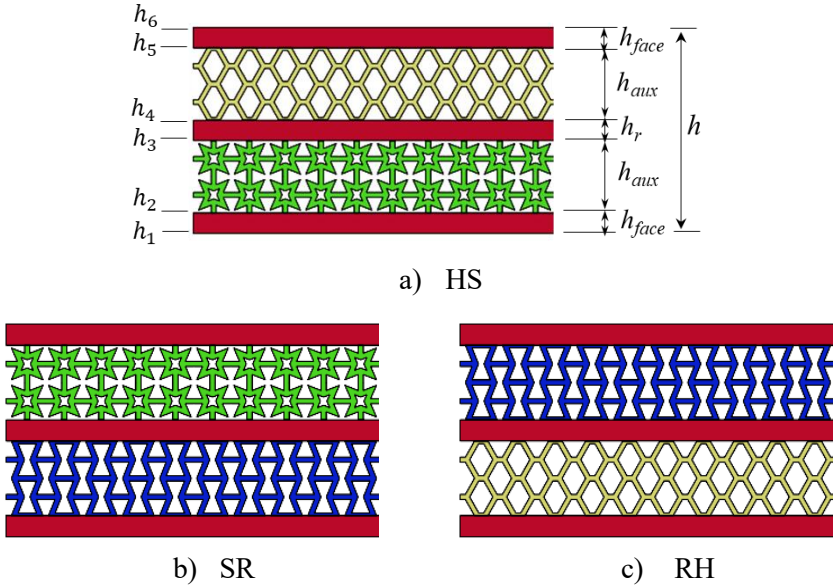


Figure 2: Representation of three different auxetic core layer designs and geometric parameters

The composition of FGM metal-ceramic surface plates varies functionally. The exterior surfaces of FGM plates are composed of pure ceramic (Si_3N_4), while the interior surfaces are composed of pure metal (Ni). The auxetic core layers and rim layer are composed of pure Ni. The material properties of FGM surface plates, which are dependent on the power law index, can be calculated using the method given below (Das et al., 2025; Kardooni et al., 2022):

$$C_{eff}(\mathrm{z}) = \left(C_{Si_3N_4} - C_{Ni}\right)\left(\frac{z - h_r/2 - h_{aux}}{h_{face}}\right)^d + C_{Ni}\,; \qquad h_5 \le z \le h_6 \tag{1}$$

$$C_{eff}(\mathrm{z}) = \left(C_{Si_3N_4} - C_{Ni}\right)\left(-\frac{z + h_r/2 + h_{aux}}{h_{face}}\right)^d + C_{Ni}\,; \quad h_1 \le z \le h_2 \tag{2}$$

Depending on the rise in temperature, the material properties of each plate vary. The material properties of the plates, both mechanical and thermal, can be investigated using the following equation (Reddy & Chin, 1998):

$$C = C_0(C_{-1}T^{-1} + 1 + C_1T + C_2T^2 + C_3T^3) \quad (3)$$

where $\Delta T=T-T_0$ and T_0=300 K. C_0, C_{-1}, C_1, C_2 and C_3 stand for the temperature-dependent material constants.

RESULTS AND DISCUSSION

In this section, the free vibration analysis of sandwich plates with different auxetic cores is examined using FEA with parameters such as auxetic types, aspect ratio, boundary conditions, power-law index, and thermal loads. In the analysis, the geometric parameters of the plate are considered as l_x = 0.1 m, $h = l_x/10$, $h_{face} = 0.1h$, $h_{aux} = 0.35h$, $h_r = 0.1h$. The material properties of Ni, and Si_3N_4 are presented in Table 1.

Table 1: Material properties of nickel and silicon nitride (Reddy & Chin, 1998; Esen et al., 2024; Ozalp & Esen, 2025)

Material	Parameter	E (Pa)	ν	α (K^{-1})	κ (W/mK)	ρ (kg/m^3)
nickel	C_0	223.95x10^9	0.31	9.9209x10^{-6}	58.754	8908
	C_{-1}	0	0	0	0	0
	C_1	-2.794x10^{-4}	0	8.705x10^{-4}	-4.614x10^{-4}	0
	C_2	3.998x10^{-9}	0	0	6.67x10^{-7}	0
	C_3	0	0	0	-1.523x10^{-10}	0
silicon nitride	C_0	348.43x10^9	0.24	5.8723x10^{-6}	13.723	2370
	C_{-1}	0	0	0	0	0
	C_1	-3.07x10^{-4}	0	9.095x10^{-4}	-1.032x10^{-3}	0
	C_2	2.16x10^{-7}	0	0	5.466x10^{-7}	0
	C_3	-8.946x10^{-11}	0	0	-7.876x10^{-11}	0

Free vibration analysis

In this section, the variation in natural frequency values for the first six vibration modes of the multi-layered sandwich plate is examined for three different hybrid auxetic core design (SR, HS, and RH), four different boundary conditions (CCCC, CFFF, CCFF and CFCF), three different aspect ratios (l_y/l_x=0.5, 1, and 2), three different temperature rise (ΔT=0,250,500 K),

and five different power law index (d=0, 1, 5, 10, and 100) through the FEA method.

The analysis of the data in Table 2 shows that the RH core design exhibited the highest natural frequency values in all six vibration modes examined, followed by the HS and SR designs respectively. Specifically, at the fundamental frequency (Mode 1), the RH core reached 1000.2 Hz, creating a difference of approximately 3.27% compared to the SR core. However, this performance difference increases significantly as the number of modes increases. By the sixth mode, the RH core's superiority (7472.4 Hz) over the SR core (6885.8 Hz) reaches approximately 8.52%.

These findings indicate that the combination of the high out-of-plane stiffness of the honeycomb cells and the structural integrity of the re-entrant geometry results in the RH configuration providing the highest global stiffness under the CFFF boundary condition. In conclusion, while all three hybrid auxetic cores deliver effective results, the RH core is the most rigid (stiff) option for reducing resonance risk in high-frequency environments, whereas the SR core offers a more flexible alternative.

Table 2: Variation in natural frequency according to the core type (d=1, l_y/l_x =1, ΔT=0, BC=CFFF)

Core Type	w_n (Hz)					
	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
SR	968.53	2034.6	5074.3	5335.0	5865.9	6885.8
HS	984.01	2093.8	5198.70	5415.10	6053.50	7037.50
RH	1000.2	2172.7	5209.9	5639.5	6475.1	7472.4

Table 3 indicates that the natural frequency values of the plate systematically increase in all modes as the constraints at the boundary conditions increase. The CFFF configuration, which is the most flexible structure, exhibits the lowest frequency values, while the CCCC condition, where all edges are fixed, reaches the highest frequency values. This situation proves that replacing free (F) edges with fixed (C) edges directly increases the dynamic rigidity of the structure.

The approximately eightfold increase in the fundamental mode frequency when moving from the CFFF condition to the CCCC condition clearly demonstrates the dominant effect of boundary conditions on structural stiffness. The values observed in intermediate boundary conditions such as CCFF and CFCF indicate that the vibration characteristics of the

structure can be controlled over a wide range by modifying the support conditions. Consequently, the dynamic response of these sandwich plates with hybrid auxetic cores exhibits a highly sensitive tendency to the boundary conditions selected during the design phase.

Table 3: Variation in natural frequency based on the boundary conditions (RH, d=1, l_y/l_x =1, ΔT=0)

BC	w_n (Hz)					
	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
CFFF	1000.2	2172.7	5209.9	5639.5	2257.60	2534.50
CCFF	1779.6	5540.8	6447.7	10218.0	11635.0	12286.0
CFCF	5681.5	6424.7	9662.0	13251.0	13823.0	14713.0
CCCC	7935.7	13245.0	15409.0	19515.0	20148.0	25247.0

Table 4 reveals that an increase in the aspect ratio correlates with a general rise in the natural frequency values of the plate. Specifically, in the narrow plate structure with l_y/l_x = 0.5, the first mode frequency is 959.57 Hz, whereas when the plate widens to reach an aspect ratio of 2. This value increases approximately four times to reach 3765.40 Hz. This situation proves that the SR hybrid auxetic core structure under the CFFF boundary condition exhibits a dynamically more rigid character as the aspect ratio increases.

Furthermore, as the mode number increases (from the 1st mode to the 6th mode), the frequency values increase as expected; however, the aspect ratio has different levels of effect in each mode. For example, in the case of a square plate (l_y/l_x = 1), the 6th mode frequency exhibits a lower upper limit of 6885.80 Hz compared to other ratios, while the plate with an aspect ratio of 2 reaches the highest vibration resistance at 17385.00 Hz. Consequently, modifying the geometric ratios of the plate during the design phase emerges as a critical tool for managing the structure's resonance risk and ensuring stable operation within specific frequency ranges.

Table 4: Variation in natural frequency according to the aspect ratio (SR, d=1, ΔT=0, BC=CFFF)

l_y/l_x	w_n (Hz)					
	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
0.5	959.57	3220.10	3522.80	5419.50	9437.00	12629.00
1	968.53	2034.60	5074.30	5335.00	5865.90	6885.80
2	3765.40	4742.90	9207.40	13305.20	14608.30	17385.00

Table 5 illustrates that the dynamic features of the multi-layered sandwich plate are directly influenced by variations in the power law index (d). The finite element analysis (FEA) findings indicated that the natural frequency values consistently decreased as the power law index increased throughout all examined vibration modes. The situation is explained by the increase of the d parameter, which decreases the effective bending stiffness of the material, alongside changes in the material composition that decrease the system's dynamic resistance.

A significant academic observation is that the sharpest decrease in frequency values occurs between $d = 0$ and $d = 1$; when d approaches 100, this rate of decline slows down, and the frequency values stabilize at a certain plateau. As expected, as the number of modes increases, the system's wavelength shortens and the frequency values rise significantly. Consequently, in this structural system under CFFF boundary conditions, the power law index emerges as a critical design parameter for controlling the dynamic response and resonance risk of the design. In light of these data, dynamic optimization is possible by selecting an appropriate n value based on the external vibration sources in the environment where the plate will operate.

Table 5: Variation in natural frequency based on the power law index (RH, l_y/l_x=1, ΔT=0, BC=CFFF)

d	w_n (Hz)					
	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
0	1161.20	2556.10	5954.50	6536.50	7534.50	8663.30
1	1000.20	2172.70	5209.90	5639.50	2257.60	2534.50
5	904.34	1944.50	4780.10	5100.20	5838.00	6759.90
10	883.47	1894.80	4688.40	4982.30	5698.30	6604.30
100	861.44	1842.30	4592.20	4856.90	5550.40	6439.60

Table 6 demonstrates that the temperature rise has a direct and significant effect on the dynamic characteristics of the structure. The natural frequency values consistently decreased with increasing temperature in the first six vibration modes. This situation is explained by the fact that high temperatures reduce the elasticity modulus of the material and weaken the structural rigidity by creating thermally induced compressive stresses in the structure. In particular, the decrease in Mode 1 value from 1000.20 Hz to

937.98 Hz proves that the temperature increase directly affects the fundamental vibration mode of the system.

On the other hand, the frequency increases observed in high-frequency modes such as Mode 5 and Mode 6 reveal the complex interaction between the hybrid auxetic core design and the CFFF boundary conditions. This indicates that thermal expansion leads to an increase in structural stress in certain mode shapes. Consequently, the dynamic stability of the design varies with temperature, and it is crucial to assess resonance risks at high temperatures.

Table 6: Variation in natural frequency based on the temperature rise (RH, l_y/l_x=1, d=1, BC=CFFF)

ΔT	w_n (Hz)					
	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
0	1000.20	2172.70	5209.90	5639.50	2257.60	2534.50
250	968.51	2104.20	5033.70	5457.00	6270.80	7229.50
500	937.98	2038.30	4858.50	5278.90	6073.90	6992.00

CONCLUSION

This study presents the design and modal vibration analysis of multi-layered hybrid sandwich plate structures with various auxetic core geometries and FG surface layers, using the finite element method. In the study, the material properties of the surface plates vary with the power law index. Furthermore, since the material properties of the plates are temperature-dependent, the material properties change under the applied thermal load. Significant results are obtained according to detailed vibration analyses performed based on the sandwich plate auxetic core design, plate aspect ratio, temperature rise, boundary condition, and power law index parameters. The RH core design demonstrates the greatest natural frequency values across all six analyzed vibration modes, succeeded by the HS and SR designs, respectively. Although all three hybrid auxetic cores yield effective outcomes, the RH core is the stiffest choice for minimizing resonance potential in high-frequency situations, while the SR core provides a more flexible option. The CFFF arrangement, characterized by its flexibility, demonstrates the lowest frequency values, whilst the CCCC condition, with all edges constrained, attains the greatest frequency values. SR hybrid auxetic core structure under the CFFF boundary condition exhibits a dynamically more rigid character as the aspect ratio increases. The modification of the aspect ratios of the plate appears as an essential instrument for minimizing the resonance risk of the structure and assuring

stable operation within particular frequency ranges. The results demonstrated that the natural frequency values diminished with the rise of the power law index across all analyzed vibration modes. This situation can be attributed to the decrease in the silicon nitride volume fraction and the increase in the nickel volume fraction on the FGM surface plate, along with the increasing *d* parameter. As shown in Table 1, the silicon nitride stiffness properties are much higher than those of nickel. Moreover, the natural frequency levels have diminished with rising temperature across the initial six vibration modes. This phenomenon can be elucidated by the observation that elevated temperatures diminish the elastic modulus of the material, hence compromising structural rigidity through the generation of thermally induced compressive stresses inside the structure. This study, which presents sandwich plate designs and analyses that have not been previously studied in the literature, is expected to contribute to important engineering fields such as airframe structures under high thermal loads in the aviation industry and energy-dissipating shielding systems in the defense industry.

REFERENCE

Aktas, K. G., Pehlivan, F., & Esen, I. (2024). Temperature-dependent thermal buckling and free vibration behavior of smart sandwich nanoplates with auxetic core and magneto-electro-elastic face layers. Mechanics of Time-Dependent Materials, 28(3), 1999–2039. https://doi.org/10.1007/s11043-024-09698-0

Alomarah, A., Masood, S. H., Sbarski, I., Faisal, B., Gao, Z., & Ruan, D. (2020). Compressive properties of 3D printed auxetic structures: Experimental and numerical studies. Virtual and Physical Prototyping, 15(1), 1–21. https://doi.org/10.1080/17452759.2019.1644184

Burlayenko, V. N., & Sadowski, T. (2020). Free vibrations and static analysis of functionally graded sandwich plates with three-dimensional finite elements. Meccanica, 55(4), 815–832. https://doi.org/10.1007/s11012-019-01001-7

Carneiro, V. H., Meireles, J., & Puga, H. (2013). Auxetic materials—A review. Materials Science-Poland, 31(4), 561–571. https://doi.org/10.2478/s13536-013-0140-6

Cho, J. R. (2022). Free Vibration Analysis of Functionally Graded Sandwich Plates with a Homogeneous Core. Applied Sciences, 12(12). https://doi.org/10.3390/app12126054

Daikh, A. A., & Zenkour, A. M. (2019). Free vibration and buckling of porous power-law and sigmoid functionally graded sandwich plates using a simple higher-order shear deformation theory. Materials Research Express, 6(11), 115707. https://doi.org/10.1088/2053-1591/ab48a9

Das, T., Mishra, P. K., & Nath, J. K. (2025). Thermomechanical response of multilayered FGM plates at normal and elevated temperatures. European Journal of Materials, 5(1), 2515036. https://doi.org/10.1080/26889277.2025.2515036

Demirbas, M. D. (2017). Thermal stress analysis of functionally graded plates with temperature-dependent material properties using theory of elasticity. Composites Part B: Engineering, 131, 100–124. https://doi.org/10.1016/j.compositesb.2017.08.005

Dong, P., Gong, Z., Li, Y., Hu, J., Cao, J., & Hou, R. (2025). Tensile behavior analysis and intelligent geometric optimization of concentric re-entrant auxetic metamaterials (CREAMs). Materials Today Communications, 44, 112203. https://doi.org/10.1016/j.mtcomm.2025.112203

Ebrahimi, F., & Rastgo, A. (2008). An analytical study on the free vibration of smart circular thin FGM plate based on classical plate theory. Thin-Walled Structures, 46(12), 1402–1408. https://doi.org/10.1016/j.tws.2008.03.008

Eroğlu, M., Esen, İ., & Koç, M. A. (2025). The effect of an auxetic core layer and symmetric FGM face layers on the 3D wave propagation response of sandwich nanoplates. Archive of Applied Mechanics, 95(3), 69. https://doi.org/10.1007/s00419-025-02772-0

Esen, I., Aktaş, K. G., & Pehlivan, F. (2025). Investigation of the Critical Buckling Temperatures and Free Vibration Response of Porous Functionally Graded Magneto-Electro-Elastic Sandwich Higher-Order Nanoplates with Temperature-Dependent Material Properties. Journal of Vibration Engineering & Technologies, 13(2), 162. https://doi.org/10.1007/s42417-024-01542-6

Esen, I., Garip, Z. S., & Eren, E. (2024). The effects of the foam and FGM distributions on thermomechanical buckling response of sandwich plates. Acta Mechanica, 235(2), 1319–1343. https://doi.org/10.1007/s00707-023-03808-8

Farshbaf, S., Dialami, N., Cervera, M., Bakhshan, H., & Navid Chakherlou, T. (2025). Enhancing the mechanical performance of additively manufactured auxetic structures through design modifications: Experimental and numerical analysis. Progress in Additive Manufacturing, 10(7), 4143–4164. https://doi.org/10.1007/s40964-025-01062-z

Garg, A., Chalak, H. D., Belarbi, M.-O., Chakrabarti, A., & Houari, M.-S.-A. (2021). Finite Element-based Free Vibration Analysis of Power-Law, Exponential and Sigmoidal Functionally Graded Sandwich Beams. Journal of The Institution of Engineers (India): Series C, 102(5), 1167–1201. https://doi.org/10.1007/s40032-021-00740-5

Guo, C., Liu, T., Bin, Q., Wang, Q., & Wang, A. (2022). Free vibration analysis of coupled structures of laminated composite conical, cylindrical and spherical shells based on the spectral-Tchebychev technique. Composite Structures, 281, 114965. https://doi.org/10.1016/j.compstruct.2021.114965

Han, X., Cai, H., Sun, J., Wei, Z., Huang, Y., Wang, A., Han, X., Cai, H., Sun, J., Wei, Z., Huang, Y., & Wang, A. (2022). Numerical Studies on Failure Mechanisms of All-Composite Sandwich Structure with Honeycomb Core under Compression and Impact Loading Conditions. Polymers, 14(19). https://doi.org/10.3390/polym14194047

Jang, W. C., & Roh, H. D. (2025). Sandwich Composites Manufacturing: A Review of Materials, Methods, Applications and Challenges. International Journal of Precision Engineering and Manufacturing, 26(8), 2093–2109. https://doi.org/10.1007/s12541-025-01285-8

Jin, X., Wang, Z., Ning, J., Xiao, G., Liu, E., & Shu, X. (2016). Dynamic response of sandwich structures with graded auxetic honeycomb cores under blast loading. Composites Part B: Engineering, 106, 206–217. https://doi.org/10.1016/j.compositesb.2016.09.037

Kamble, Z. (2024). Advanced structural and multi-functional sandwich composites with prismatic and foam cores: A review. Polymer Composites, 45(18), 16355–16382. https://doi.org/10.1002/pc.27849

Kardooni, M. R., Shishesaz, M., Mosalmani, R., Kardooni, M. R., Shishesaz, M., & Mosalmani, R. (2022). Three-Dimensional Thermo-Mechanical Elastic Analysis of Functionally Graded Five Layers Composite Sandwich Plate on Winkler Foundations. Journal of Composites Science, 6(12). https://doi.org/10.3390/jcs6120372

Kihm, F., Langelier, A., & Munson, K. (2018). Influence of the modal damping on the estimated fatigue life. Procedia Engineering, 213, 270–281. https://doi.org/10.1016/j.proeng.2018.02.028

Mahesh, V. (2022). Nonlinear free vibration of multifunctional sandwich plates with auxetic core and magneto-electro-elastic facesheets of different micro-topological textures: FE approach. Mechanics of Advanced Materials and Structures, 29(27), 6266–6287. https://doi.org/10.1080/15376494.2021.1974619

Mashat, D. S., Carrera, E., Zenkour, A. M., Al Khateeb, S. A., & Filippi, M. (2014). Free vibration of FGM layered beams by various theories and finite elements. Composites Part B: Engineering, 59, 269–278. https://doi.org/10.1016/j.compositesb.2013.12.008

Najafi, M., Ahmadi, H., & Liaghat, G. (2021). Experimental investigation on energy absorption of auxetic structures. Materials Today: Proceedings, 34, 350–355. https://doi.org/10.1016/j.matpr.2020.06.075

Najafi, M., Ahmadi, H., & Liaghat, G. (2022). Investigation on the flexural properties of sandwich beams with auxetic core. Journal of the Brazilian Society of Mechanical Sciences and Engineering, 44(2), 61. https://doi.org/10.1007/s40430-022-03368-3

Nilsson, E., & Nilsson, A. C. (2002). PREDICTION AND MEASUREMENT OF SOME DYNAMIC PROPERTIES OF SANDWICH STRUCTURES WITH HONEYCOMB AND FOAM CORES. Journal of Sound and Vibration, 251(3), 409–430. https://doi.org/10.1006/jsvi.2001.4007

Ozalp, A. F., & Esen, I. (2025). Wave dispersion analysis with FGM face layers and GPL-reinforced metallic foam core thermally loaded sandwich nanoplate using NSGT. Archive of Applied Mechanics, 95(9), 217. https://doi.org/10.1007/s00419-025-02919-z

Palomba, G., Epasto, G., & Crupi, V. (2022). Lightweight sandwich structures for marine applications: A review. Mechanics of Advanced Materials and Structures, 29(26), 4839–4864. https://doi.org/10.1080/15376494.2021.1941448

Parhi, J. D., & Roy, T. (2025). Nonlinear Vibration And Energy Absorption Responses of Auxetic Sandwich Nanocomposite Structures. Journal of Vibration Engineering & Technologies, 13(2), 164. https://doi.org/10.1007/s42417-024-01667-8

Patel, M. (2025). A comprehensive review of functionally graded materials and their ballistic impact performance: Current status and future challenges. Next Materials, 8, 100704. https://doi.org/10.1016/j.nxmate.2025.100704

Pelayo, F., Skafte, A., Aenlle, M. L., & Brincker, R. (2015). Modal Analysis Based Stress Estimation for Structural Elements Subjected to Operational Dynamic Loadings. Experimental Mechanics, 55(9), 1791–1802. https://doi.org/10.1007/s11340-015-0073-6

Pham, Q.-H., Tran, V. K., & Tran, T. T. (2023). Vibration characteristics of sandwich plates with an auxetic honeycomb core and laminated three-phase skin layers under blast load. Defence Technology, 24, 148–163. https://doi.org/10.1016/j.dt.2022.10.002

Prawoto, Y. (2012). Seeing auxetic materials from the mechanics point of view: A structural review on the negative Poisson's ratio. Computational Materials Science, 58, 140–153. https://doi.org/10.1016/j.commatsci.2012.02.012

Prusty, J. K., Sahu, D. P., & Mohanty, S. C. (2025). Experimental and numerical investigation of free vibration behaviour of sandwich plates with natural-fibre-based laminated composite cores. Nondestructive Testing and Evaluation, 40(1), 246–270. https://doi.org/10.1080/10589759.2024.2319247

Radaković, A., Čukanović, D., Bogdanović, G., Blagojević, M., Stojanović, B., Dragović, D., Manić, N., Radaković, A., Čukanović, D., Bogdanović, G., Blagojević, M., Stojanović, B., Dragović, D., & Manić, N. (2020). Thermal Buckling and Free Vibration Analysis of Functionally Graded Plate Resting on an Elastic Foundation According to High Order Shear Deformation Theory Based on New Shape Function. Applied Sciences, 10(12). https://doi.org/10.3390/app10124190

Reddy, J. N., & Chin, C. D. (1998). Thermomechanical Analysis of Functionally Graded Cylinders and Plates. Journal of Thermal Stresses, 21(6), 593–626. https://doi.org/10.1080/01495739808956165

Ren, J., Cao, Q., Ren, J., & Cao, Q. (2021). Assignment of Natural Frequencies and Mode Shapes Based on FRFs. Aerospace, 8(9). https://doi.org/10.3390/aerospace8090262

Sah, S. K., & Ghosh, A. (2021). Free vibration and buckling analysis of FGM plates using inverse trigonometric shear deformation theory. Aircraft Engineering and Aerospace Technology, 93(2), 298–310. https://doi.org/10.1108/AEAT-01-2020-0001

Sebaq, M. H., & Liu, Z. (2024). Energy absorption and vibration mitigation performances of novel 2D auxetic metamaterials. International Journal of Computational Materials Science and Engineering, 13(02), 2350022. https://doi.org/10.1142/S2047684123500227

Tarlochan, F. (2021). Sandwich Structures for Energy Absorption Applications: A Review. Materials, 14(16). https://doi.org/10.3390/ma14164731

Thang, P. T., Kim, H., Kim, C., Jang, H., Kim, T., & Kim, J. (2024). Free vibration analysis of barrel-shaped sandwich shells with auxetic honeycomb core using modified thick shell theory. Aerospace Science and Technology, 145, 108861. https://doi.org/10.1016/j.ast.2023.108861

Zhang, K., Zhao, P., Zhao, C., Hong, F., & Deng, Z. (2020). Study on the mechanism of band gap and directional wave propagation of the auxetic chiral lattices. Composite Structures, 238, 111952. https://doi.org/10.1016/j.compstruct.2020.111952

Recycling Methods of Waste Lithium-Ion Batteries

Seher POLAT[1]

1- Associate Professor Dr. Sakarya University Engineering Faculty Industrial Engineering Department aseher@sakarya.edu.tr ORCID No: 0000-0001-6023-2901

ABSTRACT

Lithium-ion batteries have become strategically important due to their widespread use in electric vehicles, portable electronic devices, and renewable energy storage systems. However, the increasing volume of end-of-life lithium-ion batteries poses significant environmental risks because of the valuable and toxic metals they contain. Therefore, the development of sustainable recycling methods that enable efficient metal recovery without harming the environment is of critical importance.

In this study, an environmentally friendly leaching method was developed for the recovery of valuable metals from end-of-life lithium-ion batteries, and an artificial intelligence–based (İlter and Kırelli, 2025) prediction model was implemented to enhance the predictability and controllability of the recycling process. Instead of conventional inorganic acids, eco-friendly leaching approaches based on organic acids were employed to reduce environmental impact. Experimental parameters affecting recycling efficiency, such as temperature, processing time, molarity, and acid type, were considered as input variables in the proposed model.

The prediction system was developed using the Support Vector Regression (SVR) algorithm implemented in the Python programming language and trained on an experimental dataset consisting of 1,000 records. Model performance was evaluated using the coefficient of determination (R^2) and mean squared error (MSE) metrics, and the results demonstrated that the model could predict recycling efficiency with high accuracy. Furthermore, a paired t-test conducted between the SVR predictions and the actual experimental data yielded a p-value of 0.8258, indicating no statistically significant difference between the predicted and observed results.

The model enables reliable efficiency prediction without extensive experiments, reducing costs, accelerating development, minimizing impacts, and supporting sustainable industrial recycling.

Keywords – Lithium-ion batteries; environmentally friendly recycling; machine learning; support vector regression (SVR); efficiency prediction

1.INTRODUCTION

The rapid growth of electric vehicles, portable electronic devices, and renewable energy storage systems has led to an unprecedented increase in the global demand for lithium-ion batteries (Harper et al., 2019; Zeng et al., 2014). Owing to their high energy density, long cycle life, and favorable performance characteristics, lithium-ion batteries have become the dominant energy storage technology across multiple sectors (Gaines, 2018). However, the accelerated deployment of these batteries has simultaneously resulted in a growing volume of end-of-life lithium-ion batteries, posing serious environmental, economic, and sustainability-related challenges (Xu et.al., 2020).

End-of-life lithium-ion batteries contain a wide range of valuable metals, such as lithium, cobalt, nickel, and manganese, which are considered critical raw materials due to their limited natural reserves and high economic value (Zeng et al., 2014; European Commission, 2020). At the same time, these batteries also include hazardous components that may cause severe environmental pollution and human health risks if improperly disposed of. Landfilling or uncontrolled treatment of spent lithium-ion batteries can lead to soil and water contamination, greenhouse gas emissions, and irreversible ecological damage (Harper et al., 2019; UNEP, 2021). Consequently, the development of effective and environmentally friendly recycling technologies has become a strategic priority within the framework of the circular economy and sustainable production (Gaines, 2018).

Conventional recycling methods for lithium-ion batteries primarily include pyrometallurgical and hydrometallurgical processes. While these methods are capable of recovering valuable metals, they often suffer from high energy consumption, complex operational requirements, and significant environmental burdens (Ordoñez et al., 2016; Xu et al., 2020). In particular, the use of strong inorganic acids in traditional hydrometallurgical processes raises concerns regarding secondary pollution, hazardous waste generation, and occupational safety (Zeng et al., 2014). In response to these limitations, recent research has increasingly focused on the development of eco-friendly recycling approaches, such as organic acid leaching, which offer lower toxicity, reduced environmental impact, and improved process sustainability (Li et al., 2013; Gaines, 2018).

Despite advances in environmentally friendly recycling techniques, optimizing recycling efficiency remains a complex and resource-intensive task. Recycling performance is influenced by multiple interacting parameters, including temperature, processing time, acid type, and solution concentration. Determining optimal operating conditions through conventional experimental approaches typically requires a large number of physical experiments, leading to increased costs, extended development times, and higher material and energy consumption. Therefore, there is a growing need for intelligent decision-support tools capable of predicting recycling outcomes and guiding process optimization with minimal experimental effort (Kusiak, 2018).

In recent years, artificial intelligence (AI) (Yılmaz Yalçıner & Gelen Mert, 2021; Korkusuz Polat, 2022; Memika & Korkusuz Polat, 2023) and machine learning (ML) techniques have emerged as powerful tools for modeling complex, nonlinear systems in industrial and environmental applications. Machine learning algorithms can learn hidden patterns from experimental data and provide accurate predictions without requiring explicit physical or chemical models (Zhang et al., 2022). In the context of recycling systems, AI-based models offer significant potential for predicting process efficiency, reducing experimental workloads, and supporting sustainable process design. Among various machine learning techniques, Support Vector Regression (SVR) has gained attention due to its robustness, high generalization capability, and effectiveness in handling nonlinear relationships with limited data (Smola & Schölkopf, 2004).

Although several studies have explored machine learning applications in materials processing and waste management, the integration of AI-based prediction models with environmentally friendly lithium-ion battery recycling methods remains relatively limited (Zhang et al., 2022; Xu et al., 2020). In particular, there is a lack of comprehensive studies that combine organic acid–based leaching processes with data-driven prediction frameworks aimed at reducing experimental intensity and enhancing industrial applicability.

In this context, the present study aims to develop an environmentally friendly recycling approach for end-of-life lithium-ion batteries and to support this process with an artificial intelligence–based prediction model. An eco-friendly leaching method using organic acids is employed to recover

valuable metals, while a Support Vector Regression model implemented in Python is developed to predict recycling efficiency based on key experimental parameters. By enabling accurate efficiency prediction without extensive physical experimentation, the proposed approach contributes to cost reduction, process acceleration, and environmental impact mitigation. Overall, this study seeks to bridge the gap between sustainable recycling technologies and digitalization, offering a practical framework for intelligent and environmentally responsible battery recycling systems.

1.1.Applications and Recycling Methods of Lithium-Ion Batteries

Lithium-ion batteries have become one of the most widely used energy storage technologies due to their high energy density, long cycle life, low self-discharge rate, and lightweight structure. These advantages have enabled their extensive adoption across a wide range of sectors, making lithium-ion batteries a critical component of modern energy systems.

1.1.1.Applications of Lithium-Ion Batteries

The most common application of lithium-ion batteries is in portable electronic devices, including smartphones, laptops, tablets, cameras, and wearable technologies. Their compact size and high energy storage capacity make them particularly suitable for such devices. In addition, lithium-ion batteries play a central role in electric and hybrid vehicles, where they are essential for reducing dependence on fossil fuels and lowering greenhouse gas emissions. The rapid growth of the electric vehicle market has significantly increased the demand for lithium-ion battery technologies.

Lithium-ion batteries are also widely used in renewable energy storage systems. They provide effective solutions for balancing the intermittent nature of renewable energy sources such as solar and wind power, both in grid-connected and off-grid energy storage applications. Furthermore, lithium-ion batteries are increasingly employed in industrial automation systems, medical devices, aerospace technologies, and defense applications, where reliability and high performance are critical requirements.

1.1.2. Recycling Methods of Lithium-Ion Batteries

The expansion of lithium-ion battery applications has led to a substantial increase in the number of end-of-life batteries, highlighting the importance of effective recycling strategies. Spent lithium-ion batteries contain valuable and critical metals, including lithium, cobalt, nickel, and manganese, which can be recovered and reused, thereby reducing the demand for primary raw materials.

Recycling methods for lithium-ion batteries are generally classified into mechanical, pyrometallurgical, hydrometallurgical, and direct recycling approaches. Mechanical recycling involves dismantling, crushing, and physical separation processes and is typically used as a pretreatment step. Pyrometallurgical methods rely on high-temperature processing to recover metals; however, they are associated with high energy consumption and significant environmental emissions.

Hydrometallurgical recycling methods are based on the dissolution of metals using acidic solutions and are known for their high recovery efficiencies. Nevertheless, the use of strong inorganic acids raises concerns related to environmental pollution and operational safety. As a result, environmentally friendly hydrometallurgical approaches employing organic acids have gained increasing attention in recent years. Direct recycling methods aim to preserve the chemical structure of battery components for reuse and represent a promising, energy-efficient alternative.

Overall, the widespread use of lithium-ion batteries underscores the necessity of developing sustainable, efficient, and environmentally friendly recycling methods to support circular economy and sustainable production practices.

1.1.3. Artificial Intelligence–Based Prediction and Recycling Optimization

Recycling processes are inherently complex due to heterogeneous material structures, variable operational conditions, and significant uncertainties. These characteristics limit the effectiveness of conventional deterministic approaches and necessitate the development of advanced data-driven methods. In this context, artificial intelligence (AI) and machine learning techniques have become essential tools for establishing decision-

support systems in waste management and recycling applications (Kusiak, 2018; Zhang et al., 2022).

In lithium-ion battery recycling, AI-based models are increasingly employed to predict recoverable metal quantities, estimate waste generation, classify end-of-life batteries, and optimize recycling processes with high accuracy (Harper et al., 2019; Li et al., 2013). Algorithms such as deep learning models, decision trees, artificial neural networks, and support vector machines are widely applied in material separation systems and in predicting key process parameters. The integration of these methods enhances process efficiency, reduces energy consumption, and minimizes environmental impacts (Zhang et al., 2022; Gaines, 2018).

From an industrial engineering perspective, AI-supported prediction systems enable more effective management of production planning, resource allocation, and scheduling in recycling facilities. By improving decision-making under uncertainty, these systems provide strategic advantages in terms of sustainability, cost reduction, and environmental performance. Consequently, recycling operations can be better aligned with circular economy principles and sustainable manufacturing objectives (Chopra & Meindl, 2019; Kusiak, 2018).

Recent studies conducted in Türkiye also demonstrate the growing impact of AI-based solutions in waste management. Deep learning and transfer learning approaches have been successfully applied to waste classification, achieving high accuracy rates and enabling automated identification in recycling facilities (Çavuşoğlu & Badem, 2024). Furthermore, AI-supported optimization of waste collection routes has been shown to significantly reduce energy consumption and operational costs. In industrial battery recycling applications, the integration of automation and AI technologies has further improved metal purification efficiency and component separation performance (Harper et al., 2019; Xu et al., 2020).

The recovery of valuable metals such as lithium, cobalt, nickel, and manganese represents not only an environmental necessity but also a strategic economic opportunity due to their critical importance in modern energy technologies (Harper et al., 2019; Zeng et al., 2014). Recycling methods developed for this purpose are generally classified into pyrometallurgical, hydrometallurgical, and direct recycling approaches (Ordoñez et al., 2016; Xu et al., 2020). While pyrometallurgical and

hydrometallurgical methods are widely applied, they are often associated with high energy consumption and environmental burdens. In contrast, direct recycling methods aim to preserve the chemical structure of battery components, offering lower energy consumption and enhanced environmental sustainability (Xu et al., 2020; Zeng et al., 2014).

Overall, AI-based prediction and optimization approaches play a critical role in improving the efficiency, sustainability, and industrial applicability of lithium-ion battery recycling systems by reducing experimental effort, lowering operational costs, and supporting environmentally responsible waste management strategies (Zhang et al., 2022; Kusiak, 2018).

1.2.Widespread Adoption of Lithium-Ion Batteries

The rapid expansion of lithium-ion batteries has been driven primarily by their extensive use in electric vehicles (EVs), renewable energy storage systems, and portable electronic devices. Owing to their high energy density, long cycle life, and favorable performance characteristics, lithium-ion batteries have become the dominant energy storage technology in modern applications. In particular, the global transition toward electric mobility and renewable energy integration has significantly accelerated lithium-ion battery production and deployment (Harper et al., 2019; Xu et al., 2020). As a result, the volume of batteries reaching end-of-life is increasing at an unprecedented rate, raising critical concerns regarding waste management and resource sustainability.

1.3.Environmental Risks of Waste Lithium-Ion Batteries

End-of-life lithium-ion batteries contain hazardous and toxic substances, including heavy metals and flammable electrolytes, which pose serious risks to the environment and human health if not properly managed. Improper disposal practices such as landfilling or uncontrolled dismantling may lead to soil and groundwater contamination, air pollution, and fire hazards. Moreover, the loss of valuable metals such as lithium, cobalt, nickel, and manganese contributes to resource depletion and increases dependency on primary raw material extraction (Zeng et al., 2014; Gaines, 2018).

1.4.Importance of Recycling for Sustainability

Recycling lithium-ion batteries is a key component of sustainable waste management and circular economy strategies. Efficient recycling processes enable the recovery of critical metals, reduce environmental pollution, and decrease the ecological footprint associated with mining and material processing. From a sustainability perspective, battery recycling supports resource efficiency, energy savings, and greenhouse gas emission reduction, while simultaneously enhancing supply chain resilience for critical materials (Ordoñez et al., 2016; Harper et al., 2019).

1.5.Role of Artificial Intelligence in Recycling Processes

Despite advances in recycling technologies, lithium-ion battery recycling remains a complex and multi-parameter process. Recycling efficiency is influenced by numerous interacting variables, such as temperature, processing time, chemical composition, and operational conditions. In this context, artificial intelligence (AI) and machine learning techniques have emerged as powerful tools for modeling, prediction, and optimization of recycling processes. AI-based models enable accurate prediction of recycling performance, reduction of experimental effort, and improved process control, thereby enhancing operational efficiency and environmental performance (Kusiak, 2018; Zhang et al., 2022).

1.6.Aim and Original Contribution of the Chapter

The primary aim of this chapter is to provide a comprehensive analysis of environmentally friendly lithium-ion battery recycling methods and to highlight the role of artificial intelligence–based prediction systems in optimizing recycling processes. The chapter’s original contribution lies in the integration of sustainable recycling approaches with AI-driven modeling frameworks, offering a data-driven perspective that reduces experimental intensity and enhances industrial applicability. By bridging sustainable materials management and digitalization, this chapter contributes to the development of intelligent and environmentally responsible battery recycling systems.

2. Applications and Recycling Methods of Lithium-Ion Batteries

Lithium-ion batteries (LIBs) have become the dominant energy storage technology of the modern era due to their high energy density, long cycle life, low self-discharge rate, and relatively lightweight structure. These advantages have enabled their widespread adoption across various sectors, significantly increasing global LIB production and, consequently, the volume of end-of-life batteries requiring proper management (Harper et al., 2019; Xu et al., 2020).

2.1 Applications of Lithium-Ion Batteries

One of the most common application areas of lithium-ion batteries is portable electronic devices, including smartphones, laptops, tablets, cameras, and wearable technologies. Their compact size and high power capacity make them particularly suitable for consumer electronics that require long operating times and fast charging capabilities.

Another major application area is electric and hybrid vehicles, where lithium-ion batteries play a critical role in reducing greenhouse gas emissions and dependence on fossil fuels. The rapid expansion of the electric vehicle market has significantly increased the demand for high-capacity lithium-ion battery systems, making battery sustainability and recycling strategic priorities (Harper et al., 2019).

Lithium-ion batteries are also widely used in renewable energy storage systems. They provide effective solutions for balancing the intermittent nature of renewable energy sources such as solar and wind power, both in grid-connected and off-grid energy systems. Additionally, LIBs are increasingly employed in industrial automation, medical devices, aerospace technologies, and defense applications, where reliability, safety, and performance are essential requirements (Xu et al., 2020).

2.2 Recycling Methods of Lithium-Ion Batteries

The rapid growth in lithium-ion battery applications has resulted in a substantial increase in end-of-life batteries, raising concerns related to environmental pollution, resource depletion, and waste management. Spent LIBs contain valuable and critical metals such as lithium, cobalt, nickel, and manganese, which can be recovered and reused through appropriate recycling processes (Zeng et al., 2014; Gaines, 2018).

Lithium-ion battery recycling methods are generally classified into mechanical, pyrometallurgical, hydrometallurgical, and direct recycling approaches.

Mechanical recycling involves disassembly, crushing, and physical separation of battery components. This method is typically used as a pretreatment step to separate active materials from casings and current collectors. Although mechanical processes are relatively simple, they are insufficient on their own for recovering high-purity metals (Ordoñez et al., 2016).

Pyrometallurgical recycling methods rely on high-temperature processes to recover metals through smelting. While these methods are industrially mature and capable of handling mixed battery waste, they suffer from high energy consumption, greenhouse gas emissions, and limited recovery of lithium and other light elements (Harper et al., 2019).

Hydrometallurgical recycling methods use chemical leaching agents to dissolve metals into liquid phases, enabling high recovery efficiencies. However, traditional hydrometallurgical processes often employ strong inorganic acids, which pose environmental and safety concerns. To address these issues, environmentally friendly hydrometallurgical approaches based on organic acids have gained increasing attention in recent years (Xu et al., 2020; Zhang et al., 2022).

Direct recycling methods aim to preserve the chemical structure of battery components, allowing them to be reused with minimal processing. These approaches offer lower energy consumption and reduced environmental impact, making them a promising alternative for sustainable battery recycling systems, although their industrial implementation remains limited (Zeng et al., 2014).

Overall, the widespread use of lithium-ion batteries highlights the urgent need for efficient, environmentally friendly, and sustainable recycling methods to support circular economy principles and long-term resource security.

3. ARTIFICIAL INTELLIGENCE IN LITHIUM-ION BATTERY RECYCLING: A LITERATURE REVIEW

The increasing complexity of lithium-ion battery (LIB) recycling processes has motivated the adoption of artificial intelligence (AI) and machine learning (ML) techniques to support data-driven decision making. LIB recycling involves nonlinear relationships among process parameters, heterogeneous material compositions, and uncertain operating conditions, which limit the effectiveness of conventional analytical and empirical approaches. In this context, AI-based models provide powerful tools for modeling, prediction, classification, and optimization across different stages of battery recycling systems (Kusiak, 2018; Zhang et al., 2022).

3.1 AI Applications in Battery Sorting and Classification

One of the earliest and most widely studied applications of AI in LIB recycling is battery sorting and classification. Accurate identification of battery types, chemistries, and states of health is essential for ensuring safe handling and selecting appropriate recycling routes. Computer vision and deep learning techniques, particularly convolutional neural networks (CNNs), have been successfully applied to automate battery recognition and waste classification tasks. These approaches significantly improve sorting accuracy while reducing labor intensity and operational risks (Zhang et al., 2022; Çavuşoğlu & Badem, 2024).

Recent studies demonstrate that deep learning models outperform traditional machine learning algorithms in classifying battery waste streams. Transfer learning architectures such as EfficientNet, MobileNet, and YOLO-based models have achieved high classification accuracy in real-time waste management applications, enabling automated preprocessing in recycling facilities (Xu et al., 2020).

3.2 AI-Based Process Modeling and Parameter Prediction

Beyond classification tasks, AI techniques are increasingly used to model and predict key recycling process parameters. Hydrometallurgical and environmentally friendly leaching processes, in particular, are influenced by multiple interacting variables, including temperature, reaction time, reagent concentration, and solvent type. Machine learning models such as artificial

neural networks (ANNs), support vector regression (SVR), random forest (RF), and gradient boosting algorithms have been applied to predict metal recovery efficiency and optimize process conditions (Li et al., 2013; Gaines, 2018).

Among these techniques, SVR has attracted considerable attention due to its robustness, high generalization capability, and effectiveness in handling nonlinear systems with limited experimental data. Studies report that SVR-based models can accurately estimate metal recovery yields while significantly reducing the number of physical experiments required, thereby lowering costs and environmental impacts (Zhang et al., 2022).

3.3 AI for Optimization and Decision Support in Recycling Systems

AI-driven optimization plays a crucial role in improving the overall performance of LIB recycling systems. Optimization tasks include maximizing metal recovery rates, minimizing energy consumption, reducing emissions, and improving economic feasibility. Hybrid approaches that integrate machine learning with optimization algorithms have been proposed to support multi-objective decision making in recycling operations (Kusiak, 2018).

From an industrial engineering perspective, AI-based decision support systems enhance production planning, resource allocation, and scheduling in recycling facilities. By enabling predictive insights and scenario analysis, these systems support sustainable manufacturing strategies and circular economy goals (Chopra & Meindl, 2019).

3.4 Research Gaps and Motivation for the Present Study

Despite significant progress in AI applications for LIB recycling, several research gaps remain. Most existing studies focus on either process modeling or waste classification, with limited integration of environmentally friendly recycling methods and AI-based prediction frameworks. In particular, the combined use of organic acid–based leaching processes and machine learning models for efficiency prediction is still underexplored in the literature. Furthermore, many studies rely on limited datasets or lack rigorous statistical validation of prediction results.

Motivated by these gaps, the present study aims to contribute to the literature by integrating an environmentally friendly leaching approach with an AI-based prediction model. By employing support vector regression and

statistically validating model performance, this work seeks to advance data-driven, sustainable, and industrially applicable lithium-ion battery recycling systems.

4. METHODOLOGY: AI-BASED PREDICTION MODEL

This section describes the methodological framework adopted to develop an artificial intelligence–based prediction model for estimating recycling efficiency in environmentally friendly lithium-ion battery recycling processes. The proposed methodology integrates experimental data with machine learning techniques to reduce experimental effort, optimize process parameters, and enhance predictive accuracy.

4.1 Overview of the Methodological Framework

The methodological approach consists of four main stages: (i) data collection and preprocessing, (ii) selection of input and output variables, (iii) development of the machine learning model, and (iv) model evaluation and statistical validation. A Support Vector Regression (SVR) algorithm was selected due to its strong generalization capability and effectiveness in modeling nonlinear relationships under limited data conditions.

4.2 Data Set and Experimental Parameters

The dataset used in this study consists of 1,000 experimental records obtained from environmentally friendly leaching experiments conducted on end-of-life lithium-ion batteries. The experiments focus on organic acid–based hydrometallurgical recycling processes aimed at minimizing environmental impact.The input variables considered in the model include:

1. Temperature (°C)
2. Processing time (min)
3. Molarity (mol/L)
4. Type of organic acid

These parameters were selected based on their significant influence on metal recovery efficiency, as reported in the literature (Xu et al., 2020; Zhang et al., 2022). The output variable of the model is recycling efficiency (%), representing the proportion of valuable metals successfully recovered during the recycling process.

Prior to model training, the dataset was normalized to eliminate scale differences among variables and improve model convergence.

4.3 Support Vector Regression Model

Support Vector Regression is a supervised machine learning technique derived from Support Vector Machines (SVMs) and is particularly suitable for regression problems involving nonlinear relationships. SVR aims to find an optimal regression function by minimizing prediction error within a predefined tolerance margin (ε) while maximizing model generalization.

In this study, the SVR model was implemented using the Python programming language and relevant machine learning libraries. A radial basis function (RBF) kernel was employed due to its effectiveness in capturing nonlinear patterns in experimental data. Key hyperparameters, including the penalty parameter (C), kernel coefficient (γ), and epsilon (ε), were optimized through cross-validation to achieve the best predictive performance.

4.4 Model Training and Validation

The dataset was divided into training and testing subsets to evaluate the generalization performance of the model. The SVR model was trained using the training dataset and subsequently tested on unseen data to assess prediction accuracy. Model performance was evaluated using widely accepted regression metrics:

1. Coefficient of determination (R^2)
2. Mean squared error (MSE)

These metrics provide quantitative insight into the model's ability to explain variance in the data and minimize prediction errors.

4.5 Statistical Analysis and Model Verification

In addition to conventional performance metrics, a paired t-test was conducted to statistically compare the predicted recycling efficiency values with the corresponding experimental results. The null hypothesis assumed no significant difference between predicted and observed values. The resulting p-value ($p = 0.8258$) indicates that the null hypothesis cannot be rejected, confirming that the SVR model predictions are statistically consistent with experimental data.

4.6 Methodological Contribution

The proposed methodology enables accurate prediction of recycling efficiency without the need for extensive physical experimentation. By integrating environmentally friendly recycling techniques with artificial intelligence–based prediction models, the methodology contributes to cost reduction, process acceleration, and environmental impact mitigation. Furthermore, the framework offers a scalable and adaptable approach that can be extended to other recycling systems and sustainable manufacturing applications.

5. APPLICATION

5.1 Determination of Experimental Parameters (Leaching Time, Temperature, Acid Type, and Molarity)

The experimental parameters used in this study—namely leaching time, temperature, acid type, and molarity—were determined using two complementary approaches: an extensive literature review and preliminary laboratory experiments. In the first stage, previous academic studies focusing on the recycling of waste lithium-ion batteries were systematically reviewed, with particular emphasis on environmentally friendly organic acid–based leaching methods. Frequently reported temperature ranges, acid concentrations, and leaching durations were identified from the literature, and initial parameter intervals were defined accordingly.

Subsequently, the practical applicability of these parameter ranges was evaluated through preliminary laboratory experiments. These pre-tests were conducted to verify process stability, reaction feasibility, and operational safety under laboratory conditions. Based on the outcomes of these experiments, the initial parameter ranges were refined and validated. As a result, the experimental values employed throughout the study were established on a reliable basis by integrating scientific evidence from the literature with empirical observations obtained from laboratory practice.

5.2 Collection and Organization of Experimental Data

The data obtained from the experimental studies were initially transferred to a Microsoft Excel environment, where preliminary data organization was performed. Missing, inconsistent, or duplicate observations

were manually examined and removed to ensure data integrity. Following this step, the cleaned dataset was imported into the Python programming environment for further processing.

Using the Pandas library, column names were standardized, data types were converted where necessary, and the dataset was structured into a consistent format suitable for analysis. All observations were consolidated into a single dataset, and filtering and sorting operations were applied to prepare the data for subsequent analytical and modeling stages.

Table 2 presents a sample of the raw experimental dataset prior to digitization. At this stage, categorical variables such as acid type are represented in textual form, while numerical parameters reflect the measured experimental conditions.

Table 2. Raw Data Cross-Section Before Digitization.

Acid Type	Temperature (°C)	Molarity (M)	Time (min)	Efficiency (%)
Citric Acid	50	0.25	60	76.2
Acetic Acid	70	0.5	90	81.7

To enable computational processing and compatibility with machine learning algorithms, categorical variables were transformed into numerical representations. This digitization step ensured that all variables could be mathematically processed during model training.

Table 3 illustrates the dataset after the digitization process, where acid types were encoded numerically.

Table 3. Data Cross-Section After Digitization

Acid Type	Temperature (°C)	Molarity (M)	Time (min)	Efficiency (%)
0	50	0.25	60.0	76.2
1	70	0.5	90.0	81.7

5.3 Data Preprocessing for Analysis (Cleaning, Encoding, and Scaling)

Prior to data analysis and model development, the experimental dataset was further processed using the Pandas library in Python. During this stage, missing or erroneous observations were identified and removed to prevent potential bias in the modeling results. Subsequently, categorical

variables—such as acid type—were converted into numerical representations using appropriate encoding techniques to ensure compatibility with machine learning algorithms.

Following the encoding process, the dataset was standardized to eliminate potential imbalances caused by differences in variable scales. This step is particularly important for distance-based algorithms such as Support Vector Regression (SVR), where variables with larger numeric ranges can disproportionately influence model performance. Table 4 presents the value ranges of key variables prior to the application of the scaling process. Before applying the StandardScaler transformation, the dataset retained its original numerical scale, as illustrated in Table 5.

Table 4. Value Ranges Used for Feature Scaling

Variable	Value Range
Temperature (°C)	25–90 °C
Molarity (M)	0.1–1.0 mol/L
Time (min)	15–120 dk

Table 5. Data Cross-Section Without StandardScaler Application

Acid Type	Temperature (°C)	Molarity (M)	Time (min)
0.0	50.0	0.25	60.0
1.0	70.0	0.5	90.0

After applying the StandardScaler technique, all variables were transformed to have a mean value of zero and a standard deviation of one. This transformation ensures that each variable contributes equally to the learning process, thereby improving numerical stability and convergence during model training.

Table 6. Data Cross-Section with StandardScaler Applied

Acid Type	Temperature (°C)	Molarity (M)	Time (min)
-1.0	-1.0	-1.0	-1.0
1.0	1.0	1.0	1.0

5.4 Model Training and Data Partitioning

During the training of the Support Vector Regression (SVR) model, the dataset was divided into training and testing subsets to evaluate the model's generalization capability on unseen data. This partitioning was performed using the train_test_split function in Python, with 50% of the

dataset allocated for model training and the remaining 50% reserved for testing purposes.

The training dataset was used to construct the SVR model, while the testing dataset was employed to assess predictive performance. This approach ensured that the model was not only fitted to the training data but also evaluated on previously unseen observations, providing a reliable assessment of its generalization ability.Table 7 presents a representative subset of the training dataset used for model development.

Table 7. Cross-Section of Training Data

Acid Type	Temperature (°C)	Molarity (M)	Time (min)	Acid Type
0.0	50.0	0.25	60.0	76.2
0.0	55.0	0.4	70.0	77.3
1.0	80.0	0.6	100.0	84.1
1.0	65.0	0.3	80.0	79.5

Similarly, Table 8 presents a subset of the test dataset used to evaluate the predictive performance of the trained model.

Table 8. Cross-Section of Test Data

Acid Type	Temperature (°C)	Molarity (M)	Time (min)	Acid Type
0.0	45.0	0.2	50.0	74.8
0.0	60.0	0.35	75.0	78.5
1.0	70.0	0.5	90.0	81.7
1.0	75.0	0.55	95.0	82.9

5.4. Development of the Recycling Efficiency Prediction Model

Due to the nonlinear and complex relationships between process parameters and recycling efficiency, a Support Vector Regression (SVR)–based modeling approach was selected for predicting recycling performance in this study. Variables such as temperature, processing time, molarity, and acid type interact in a highly nonlinear manner, making conventional statistical and linear modeling techniques insufficient for accurate prediction. SVR was therefore selected as the most suitable method due to its strong generalization capability and effectiveness in handling nonlinear relationships. The algorithm is capable of learning complex input–output mappings and producing reliable predictions even when the dataset size is limited. For these reasons, SVR was adopted as the core predictive model in this study to estimate recycling efficiency with high accuracy.

5.5. Hyperparameter Optimization of the SVR Model

After establishing the SVR model structure, hyperparameter optimization was performed to enhance predictive performance. Three key hyperparameters—namely the penalty parameter (C), the epsilon-insensitive loss parameter (ε), and the kernel function—were systematically optimized.

The optimization process was conducted using the GridSearchCV method implemented in the Scikit-learn library. This approach evaluates multiple combinations of hyperparameters using cross-validation, allowing the identification of the optimal configuration that maximizes predictive performance while minimizing overfitting.

Through this optimization process, the selected values of C, ε, and kernel function significantly improved the model's learning capacity and generalization ability. As a result, the optimized SVR model demonstrated stable performance across both training and testing datasets.

5.6. Model Performance Evaluation

The performance of the trained SVR model was evaluated using commonly adopted regression metrics, including Mean Squared Error (MSE) and the coefficient of determination (R^2). These metrics provide quantitative insight into prediction accuracy and the model's ability to explain variance in the observed data. Table 9 presents the performance results obtained from the test dataset.

Table 9. Model Performance Metrics

Metrik Adı	Değer
Ortalama Kare Hata (MSE)	5.21
R-kare (R^2) Skoru	0.98

The results indicate that the SVR model achieved a high level of predictive accuracy, with an R^2 value of 0.98, demonstrating a strong agreement between predicted and observed recycling efficiency values. The relatively low MSE further confirms the robustness and reliability of the proposed model.

Overall, these results indicate that the developed SVR-based prediction model is highly effective for estimating recycling efficiency and can serve as a reliable decision-support tool for optimizing environmentally friendly recycling processes.

6.ANALYSIS

6.1 T-Test Analysis

To evaluate the predictive performance of the developed SVR model, a statistical comparison was conducted between the predicted recycling efficiency values and the experimentally obtained efficiency values. Since each predicted value corresponds directly to an experimentally observed value, the analysis was performed on two dependent datasets. For this purpose, a paired t-test was employed, as it is the most appropriate statistical method for determining whether the mean difference between two related samples is statistically significant.

The paired t-test evaluates whether the observed differences between paired observations occur randomly or represent a systematic deviation. In this study, it was used to assess how closely the SVR model predictions matched the experimental measurements, thereby providing a quantitative evaluation of model reliability and predictive consistency. Table 10 presents the comparison between actual and predicted recycling efficiency values obtained from the SVR model.

Table 10. Actual and Predicted Recycling Efficiency Values

Acid Type	Temperature (°C)	Molarity (M)	Time (min)	Actual Productivity (%)	Estimated Productivity (%)
0	45	0,16	60	36,86	40,37
1	95	1,43	120	83,3	84,1
0	65	0,66	45	55,88	53,3
1	60	1,89	30	51,99	51,11
1	90	1	45	76,13	75,13
1	30	1,74	15	24,7	25,17
1	45	1,67	30	40,32	39,35
1	25	0,54	105	29,4	28,59
1	25	0,25	105	28,65	28,71
0	30	1,42	105	29,45	29,68
1	100	0,75	30	81,48	81,31
0	95	1,48	120	81,66	83,23
0	25	0,22	120	27,06	27,46
0	75	0,7	45	61,68	60,85
0	45	1,13	60	38,04	39,51
0	70	1,6	120	62,99	63,86
1	65	1,29	120	61,52	60,14
1	50	1,78	105	48,26	48,51
1	75	0,54	30	63,58	61,67
0	25	0,15	30	23,64	21,35
1	85	1,75	60	69,09	72,45

Table 11. Paired t-Test Results

İstatistik	Değer
t-değeri	0,2220
p-değeri	0,8258

The paired t-test results indicate a t-value of 0.2220 and a p-value of 0.8258. Since the p-value is significantly higher than the commonly accepted significance level of 0.05, there is no statistically significant difference between the predicted and experimentally observed recycling efficiency values. This finding confirms that the SVR model provides reliable and accurate predictions.

6.2. SWOT Analysis

The SWOT analysis presented in Table X provides a comprehensive overview of the internal and external factors influencing the proposed AI-based recycling framework. The results indicate that the model offers significant strengths, particularly in terms of predicting recycling efficiency, reducing operational costs, and supporting environmentally sustainable production through the use of organic acids. At the same time, certain limitations remain, such as dependency on data quality and the need for technical expertise during implementation. From an opportunity perspective, the increasing global emphasis on sustainable recycling practices and the expansion of industrial-scale recycling infrastructures create favorable conditions for the adoption of the proposed approach. However, potential challenges—including regulatory uncertainties, limited digital infrastructure in some industrial environments, and fluctuations in chemical supply and costs—may influence large-scale implementation. Overall, the SWOT analysis highlights that while certain risks and constraints exist, the proposed framework demonstrates strong potential for supporting intelligent, sustainable, and scalable recycling systems.

Table 12. SWOT Analysis Results

Strengths	Weaknesses	Opportunities	Threats
Enables prediction of recycling efficiency, reducing cost and time.	Model performance depends heavily on data quality.	High applicability in industrial-scale recycling systems.	Regulatory changes may affect the use of eco-friendly chemicals.
Uses environmentally friendly organic acids.	Technical expertise may be required for industrial implementation.	Growing recycling market increases commercialization potential.	Limited AI infrastructure in some industries.
Reduces chemical consumption by minimizing experimental trials.	Model validity is limited to the training data range.	Can be adapted to different waste types.	Resistance to technological change in traditional recycling facilities.
Supports digitalization and sustainable production.	Availability of organic acids may vary regionally.	Enhances sustainability-driven innovation.	Fluctuations in chemical costs may affect economic feasibility.

7. CONCLUSION

This study aimed to develop an environmentally friendly and data-driven framework for predicting the recycling efficiency of waste lithium-ion batteries using an artificial intelligence–based approach. Considering the environmental risks, high operational costs, and time-intensive nature of conventional experimental recycling methods, the proposed Support Vector Regression (SVR)–based model offers an effective alternative for predicting recycling efficiency without the need for extensive physical experimentation.

The results obtained from the experimental analysis demonstrate that the proposed model achieved a high predictive performance, with an R^2 value of 0.98 and a mean squared error (MSE) of 5.21. Furthermore, the paired t-test conducted between the experimentally measured and predicted efficiency values yielded a p-value of 0.8258, indicating no statistically significant difference between the two datasets. This finding confirms the reliability, robustness, and consistency of the proposed SVR-based prediction model in representing real experimental behavior.

One of the major contributions of this study is the integration of environmentally friendly organic acid–based leaching processes with artificial intelligence–driven modeling. By significantly reducing the need for repeated experimental trials, the proposed approach minimizes chemical

consumption, energy usage, and operational costs while enhancing process efficiency. In this context, the developed model not only supports environmentally sustainable practices but also contributes to safer and more resource-efficient recycling operations.

From an industrial perspective, the proposed framework provides decision-makers with a powerful tool for predicting recycling performance under different operating conditions prior to implementation. This capability allows manufacturers to optimize process parameters, reduce trial-and-error experimentation, and improve overall production planning. Consequently, the model supports both economic competitiveness and sustainability goals within the circular economy framework.

Despite its promising results, this study has certain limitations. The predictive performance of the model is inherently dependent on the quality and range of the experimental dataset. Therefore, extrapolation beyond the investigated parameter ranges may reduce prediction accuracy. In addition, the current model focuses on specific organic acids and operating conditions; future studies may expand the dataset to include different battery chemistries, alternative leaching agents, and large-scale industrial conditions.

Future research should also explore the integration of advanced machine learning techniques such as ensemble learning, deep neural networks, and hybrid optimization algorithms to further enhance predictive accuracy. Moreover, combining artificial intelligence with life cycle assessment (LCA) approaches could provide a more comprehensive evaluation of environmental impacts, supporting sustainable decision-making in battery recycling systems.

Table 13. Comparison Between Conventional Methods and the SVR-Based Model

Criteria	Conventional Methods	SVR-Based Model
Experimental Duration	High (separate experiments required for each parameter)	Low (rapid prediction using trained model)
Chemical Consumption	High (trial-and-error approach)	Low (reduced experimental dependency)
Cost	High (materials, energy, labor)	Low (optimized process reduces costs)
Prediction Capability	Not available	High (predictive modeling capability)
Environmental Impact	High (increased waste and energy use)	Low (controlled and eco-friendly process)
Operational Complexity	Requires extensive physical experimentation	Digital and simulation-based implementation
Process Optimization	Time-consuming and resource-intensive	Fast and efficient parameter optimization
Repeatability	Low due to experimental variability	High due to model consistency

In conclusion, this study presents a reliable, scalable, and environmentally responsible framework for lithium-ion battery recycling. By integrating artificial intelligence with sustainable process design, the proposed approach offers significant potential for both academic research and industrial implementation, contributing to the advancement of circular economy principles and green manufacturing practices.In conclusion, this study presents a robust, environmentally responsible, and digitally supported approach to lithium-ion battery recycling. By integrating artificial intelligence with sustainable processing techniques, the proposed framework offers a scalable and adaptable solution with high potential for real-world industrial applications.

REFERENCE

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018

Çavuşoğlu, B. B., & Badem, H. (2025). Transfer öğrenme kullanılarak geri dönüşüm atıklarının sınıflandırılması. Niğde Ömer Halisdemir Üni. Müh. Bil. Dergisi, 14(1), 141–151. https://doi.org/10.28948/ngumuh.1504730

Gaines, L. (2018). Lithium-ion battery recycling processes: Research towards a sustainable course. Sustainable Materials and Technologies, 17, e00068.

Harper,G.,Sommerville, R.,Kendrick, E.,Driscoll, L.,Slater,P.,Stolkin, R.,Walton, A., Christensen,P., Heidrich, O., Lambert, S., Abbott, A.,Ryder, K.,Gaines, L., & Anderson,P. (2019). Recycling lithium-ion batteries from electric vehicles.Nature,575(7781),75–86.https://doi.org/10.1038/s41586-019-1682-5

Helms, M. M., & Nixon, J. (2010). Exploring SWOT analysis – Where are we now? A review of academic research from the last decade. Journal of Strategy and Management, 3(3), 215–251.

İlter, P., & Kırelli, Y. (2025). Decision support systems based on machine learning for detection of liver diseases. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 25(4), 827–837. https://doi.org/10.35414/akufemubid.1580440

Korkusuz Polat, T. (2022). Forecasting of production and scrap amounts using artificial neural networks. Emerging Materials Research, 11(3), 345–355.

Li, L., Dunn, J. B., Zhang, X., Gaines, L., Chen, R., Wu, F., & Amine, K. (2013).Recovery of metals from spent lithium-ion batteries with organic acids as leaching reagents and environmental assessment. Journal of Power Sources, 233, 180–189.

Memika, T., & Korkusuz Polat, T. (2023). Internet of Things supported airport boarding system and evaluation with fuzzy logic. Intelligent Automation & Soft Computing, 35(3), 2687–2702.

Ordoñez, J., Gago, E. J., & Girard, A. (2016). Processes and technologies for the recycling and recovery of spent lithium-ion batteries. Renewable and Sustainable Energy Reviews, 60, 195–205.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Pregowska, A., Osial, M., Piotrowski, T., & Kurczewski, K. (2022). The application of artificial intelligence in effective battery waste management. AI, 7(6), 81.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14, 199–222.

Student. (1908). The probable error of a mean. Biometrika, 6(1), 1–25.

Urbańska, W. (2020). Recovery of Co, Li, and Ni from spent Li-ion batteries by organic leaching. Minerals, 10(6), 555. https://doi.org/10.3390/min10060555

Yılmaz Yalçıner, A., & Gelen Mert, M. B. (2021). Estimating the occupancy rate of an accommodation business using artificial neural networks. Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 47, 209–218.

Zeng, X., Li, J., & Singh, N. (2014). Recycling of spent lithium-ion battery: A critical review. Critical Reviews in Environmental Science and Technology, 44(10), 1129–1165. https://doi.org/10.1080/10643389.2013.763578

An Intelligent Vision-Based Occupational Health and Safety Monitoring System Using Deep Learning

Seher POLAT[1]

1- Associate Professor Dr. Sakarya University Engineering Faculty Industrial Engineering Department aseher@sakarya.edu.tr ORCID No: 0000-0001-6023-2901

ABSTRACT

Occupational Health and Safety (OHS) is a critical discipline aimed at protecting the physical and mental well-being of employees while ensuring sustainable organizational performance. However, traditional OHS practices are largely based on manual monitoring processes, which limit efficiency, accuracy, and timely risk detection. With the rapid advancement of digitalization, artificial intelligence (Memika and KorkusuzPolat, 2023; Yılmaz Yalçıner and Gelen Mert, 2021; KorkusuzPolat, 2022, İlter and Kırelli, 2025), and image processing technologies, new opportunities have emerged to enhance occupational safety through intelligent and proactive monitoring systems.

The purpose of this study is to develop an image processing–based occupational health and safety monitoring system supported by deep learning algorithms capable of real-time risk detection. The proposed system enables early identification of potential hazards before accidents occur, monitors the proper use of personal protective equipment, and detects unsafe behaviors in industrial environments. By minimizing human error and enabling real-time intervention, the system significantly contributes to the improvement of workplace safety performance.

Furthermore, the proposed framework integrates data-driven analysis and reporting mechanisms that support informed decision-making and continuous improvement processes. The system not only enhances transparency in safety management but also provides a scalable and adaptable solution for different industrial settings. Through automated monitoring and intelligent risk assessment, organizations can proactively manage safety risks rather than relying on reactive approaches.

In conclusion, this study presents a practical and effective solution for improving occupational health and safety practices through the integration of artificial intelligence and image processing technologies. The developed system contributes to the prevention of occupational accidents, supports the establishment of a strong safety culture, and promotes sustainable and resilient working environments in modern industrial settings.

Keywords – Occupational Health and Safety, Digitalization, Artificial Intelligence, Deep Learning, Risk Analysis, Continuous Improvement

INTRODUCTION

Occupational Health and Safety (OHS) plays a critical role in protecting employees' physical and psychological well-being while ensuring operational continuity and productivity in industrial environments. Despite continuous improvements in safety regulations and management practices, occupational accidents remain a major global concern, particularly in high-risk sectors such as construction, manufacturing, and heavy industry. According to international reports, a significant proportion of workplace accidents are still caused by unsafe behaviors, inadequate supervision, and delayed risk detection, highlighting the limitations of traditional safety management approaches.

In many industrial settings, OHS practices are largely based on manual inspections, periodic audits, and human observation. Although these conventional methods contribute to compliance with safety standards, they often fail to provide continuous and real-time monitoring of dynamic work environments. As a result, hazardous situations such as non-compliance with personal protective equipment (PPE), unauthorized access to restricted areas, and unsafe operational behaviors may remain undetected until an incident occurs. This reactive nature of traditional safety practices significantly increases the likelihood of occupational accidents and limits the effectiveness of preventive interventions.

Moreover, manual inspection-based systems heavily depend on human attention, experience, and situational awareness. Factors such as fatigue, workload, and environmental complexity can reduce the reliability of human observation, leading to overlooked hazards and delayed responses. In modern industrial environments characterized by high operational speed and complexity, such limitations pose serious challenges to maintaining continuous and objective safety monitoring. Consequently, there is a growing need for intelligent systems capable of providing consistent, real-time, and unbiased safety supervision.

Recent advances in digitalization, artificial intelligence (AI), and computer vision technologies have created new opportunities to overcome these challenges. Vision-based monitoring systems, supported by deep learning algorithms, enable the automated detection of unsafe behaviors, improper use of personal protective equipment, and hazardous interactions between workers and machinery. These technologies allow safety risks to be identified at an early stage, enabling proactive interventions before accidents occur. Furthermore, data-driven monitoring systems support evidence-based decision-making and contribute to the establishment of a proactive safety culture.

Despite these technological advancements, many existing solutions remain limited in scope, often focusing solely on detection accuracy without integrating structured risk assessment or decision-support mechanisms. This

gap restricts their practical applicability in complex industrial environments, where safety management requires not only detection but also prioritization, interpretation, and continuous improvement.

In this context, the present study aims to develop an image processing–based occupational health and safety monitoring system supported by deep learning and structured risk assessment techniques. The proposed system is designed to provide real-time hazard detection, minimize human dependency, and enhance proactive safety management. By integrating intelligent visual analysis with risk-based decision support, this study seeks to contribute to the development of more resilient, efficient, and sustainable occupational safety systems suitable for modern industrial environments.

2. LITERATURE REVIEW

2.1 Occupational Health and Safety in Industrial Environments

Occupational Health and Safety (OHS) plays a fundamental role in protecting employees from work-related injuries and illnesses while ensuring sustainable industrial productivity. According to the International Labour Organization (ILO, 2021), millions of occupational accidents occur annually worldwide, resulting in significant economic and social losses. Traditional OHS systems primarily rely on manual inspections, safety audits, and compliance-based procedures. Although these methods have contributed to improved safety standards, they are often reactive, time-consuming, and highly dependent on human judgment (Kumar & Ghadge, 2019).

Previous studies indicate that human-related factors such as fatigue, distraction, and cognitive overload significantly increase the probability of unsafe behaviors in industrial settings (Reason, 1997). Consequently, conventional safety management systems struggle to detect dynamic hazards in real time, particularly in complex and fast-paced working environments. These limitations have motivated researchers to explore advanced technological solutions capable of enhancing situational awareness and proactive risk prevention.

2.2 Digitalization and Smart Safety Systems

The emergence of Industry 4.0 has significantly transformed traditional industrial processes by integrating cyber–physical systems, artificial intelligence (AI), and data-driven decision-making tools. Within this paradigm, smart safety systems have gained prominence as essential components of modern occupational safety management (Lasi et al., 2014). These systems enable continuous monitoring, real-time data processing, and predictive risk assessment, thereby improving both operational efficiency and safety performance.

Digital transformation allows organizations to shift from reactive safety management to proactive and predictive strategies. Studies have shown that intelligent monitoring systems can significantly reduce workplace accidents by identifying hazardous patterns before incidents occur (Zhou et al., 2020). However, the effectiveness of such systems depends heavily on the accuracy of data acquisition and the robustness of analytical algorithms.

2.3 Image Processing and Computer Vision in OHS

Image processing and computer vision technologies have become increasingly prominent in occupational safety research due to their non-intrusive nature and high adaptability. Vision-based systems are capable of detecting unsafe worker behaviors, monitoring personal protective equipment (PPE) usage, and identifying hazardous zones in real time (Fang et al., 2018).

Recent advancements in deep learning, particularly convolutional neural networks (CNNs), have significantly improved the accuracy of object detection and behavior recognition tasks. Algorithms such as YOLO, Faster R-CNN, and SSD have been widely adopted for real-time safety monitoring applications (Redmon et al., 2016; Ren et al., 2015). These models enable continuous surveillance without disrupting workflow, making them highly suitable for industrial environments.

Despite their advantages, vision-based safety systems face challenges related to lighting variability, occlusions, camera positioning, and generalization across different work environments (Zhang et al., 2021). These limitations highlight the need for robust and adaptable system architectures.

2.4 Integration of Risk Assessment and Decision Support Systems

While image processing technologies are effective for hazard detection, their integration with structured risk assessment frameworks is essential for comprehensive safety management. Traditional methods such as Failure Modes and Effects Analysis (FMEA), Hazard Identification and Risk Assessment (HIRA), and Fault Tree Analysis (FTA) provide systematic approaches for evaluating risk severity, occurrence, and detectability (Stamatis, 2003).

Recent studies emphasize the importance of combining AI-based monitoring with risk assessment methodologies to enhance interpretability and decision-making (Abdelgawad & Fayek, 2018). By linking detected hazards to quantified risk levels, organizations can prioritize interventions more effectively and allocate resources efficiently. Moreover, data-driven decision support systems facilitate continuous improvement by enabling trend analysis and performance evaluation.

However, existing studies often address detection and risk evaluation separately, limiting their practical applicability in complex industrial environments. There remains a significant research gap in developing integrated frameworks that combine real-time visual monitoring with structured risk assessment and continuous improvement mechanisms.

2.5 Research Gap and Motivation

Despite substantial progress in intelligent safety systems, several challenges persist. Most existing solutions lack scalability across different industrial contexts and fail to provide holistic integration between perception, analysis, and decision-making processes. Additionally, limited attention has been given to transforming real-time visual data into actionable insights that support long-term safety strategies.

To address these gaps, this study proposes an integrated image processing–based occupational health and safety monitoring system supported by deep learning and risk assessment methodologies. The proposed framework aims to enhance situational awareness, reduce human error, and promote proactive safety management. By bridging the gap between intelligent perception and structured decision-making, this research contributes to the development of resilient and sustainable occupational safety systems.

3. METHODOLOGY

3.1 Steps Followed for Problem Solving

In this study, a systematic and structured methodology was adopted to develop an image processing–based occupational health and safety (OHS) monitoring system. The proposed approach aims to address the limitations of traditional safety monitoring methods by enabling real-time detection of hazardous situations and supporting proactive decision-making. The solution process was designed in sequential stages, starting from problem identification and literature review to model development and performance evaluation.

The methodological framework ensures both scientific rigor and practical applicability by integrating advanced image processing techniques with deep learning–based modeling and performance evaluation metrics. The main stages followed in the proposed solution are summarized in Table 1.

Table 1. Stages of the Proposed Solution Approach

Stage	Description	Tools / Methods Used
1	Literature review and problem definition	Google Scholar, academic databases
2	Image data collection	Roboflow
3	Data preprocessing and annotation	Roboflow, LabelImg
4	Model development and training	YOLOv8, Python, Jetson Nano
5	Performance evaluation	Precision, Recall, F1-score, mAP

3.2 Literature Review and Problem Definition

In the first stage, an extensive review of existing studies related to occupational health and safety, image processing, and artificial intelligence–based monitoring systems was conducted. The objective was to identify current limitations, technological gaps, and research trends within the field. Previous studies indicate that most conventional safety systems rely heavily on manual supervision and post-incident analysis, which limits their effectiveness in dynamic industrial environments.

Based on this analysis, the core problem was defined as the lack of real-time, automated, and intelligent monitoring mechanisms capable of detecting unsafe behaviors and hazardous conditions. This problem definition formed the foundation for the development of the proposed system.

3.3 Image Data Collection

Image data were collected to represent various operational scenarios commonly observed in industrial environments. The dataset included images reflecting both safe and unsafe working conditions, such as proper and improper use of personal protective equipment (PPE), presence in restricted areas, and unsafe worker behaviors.

The Roboflow platform was utilized to organize, manage, and standardize the collected image dataset. This platform facilitated dataset versioning and ensured consistency across training and testing phases.

3.4 Data Preprocessing and Annotation

To improve model robustness and generalization capability, several preprocessing operations were applied to the collected images. These included resizing, normalization, noise reduction, and data augmentation techniques such as rotation, scaling, and flipping.

Following preprocessing, images were manually annotated using the LabelImg tool. Bounding boxes were created to label objects of interest, including workers, PPE components, and hazardous zones. Accurate annotation was essential for improving detection performance and minimizing classification errors during model training.

3.5 Model Development and Training

The model development phase employed the YOLOv8 architecture due to its high accuracy and real-time processing capability. YOLOv8 was selected for its efficiency in object detection tasks and suitability for deployment in real-time monitoring systems.

Model training was conducted using Python, and computational acceleration was achieved through the use of the NVIDIA Jetson Nano platform. The training process included hyperparameter tuning and iterative optimization to enhance detection accuracy and reduce false predictions. The trained model was then evaluated using unseen test data to assess its generalization capability.

3.6 Performance Evaluation

The performance of the proposed system was evaluated using widely accepted metrics, including Precision, Recall, F1-Score, and mean Average Precision (mAP). These metrics provide a comprehensive assessment of classification accuracy, detection reliability, and overall model effectiveness.

The evaluation results demonstrate that the proposed system can accurately identify unsafe conditions and support real-time safety monitoring. The findings confirm the feasibility of deploying the system in real-world industrial environments.

3.7 Summary

In summary, the proposed methodology presents a structured and systematic approach for developing an intelligent occupational health and safety monitoring system. By integrating image processing, deep learning, and performance evaluation techniques, the system effectively addresses the limitations of traditional safety practices. The methodology provides a solid foundation for developing scalable, reliable, and real-time safety monitoring solutions applicable to various industrial sectors.

4. APPLICATION

This section presents the practical implementation of the proposed image processing–based occupational health and safety (OHS) monitoring system. The application process follows a structured workflow consisting of five main stages: (i) literature review and problem definition, (ii) image data acquisition, (iii) data preprocessing and annotation, (iv) model development and training, and (v) performance evaluation. Each stage was carefully designed to ensure methodological rigor, system reliability, and real-world applicability.

Stage 1: Literature Review and Problem Definition

In this stage, existing applications of image processing technologies in the field of occupational health and safety (OHS) were comprehensively

reviewed. To identify current trends and research gaps, an extensive literature search was conducted using various academic databases. The review revealed that a large proportion of existing OHS systems still rely heavily on manual inspection processes, while the adoption of intelligent and automated solutions has only recently begun to increase with advancements in artificial intelligence technologies.

In particular, studies have demonstrated that object detection–based algorithms play a significant role in monitoring the use of personal protective equipment (PPE), thereby contributing to the reduction of occupational accidents. These studies, drawn from both national and international literature, were comparatively analyzed in terms of their hardware infrastructure, software architectures, and detection accuracy. The findings indicate that although notable progress has been achieved, many existing solutions remain limited in terms of real-time applicability and large-scale industrial deployment.

Furthermore, an examination of studies conducted in Türkiye reveals that image processing–based OHS systems are still in an early stage of adoption and are predominantly confined to theoretical or experimental implementations. This highlights a critical gap between academic research and practical industrial applications. Consequently, there is a clear need for developing a robust, scalable, and practically applicable monitoring system that can be deployed in real industrial environments.

Based on these observations, the core research problem of this study is defined as follows:

"How can an intelligent system be designed to automatically detect the use of personal protective equipment and identify hazardous situations in order to enhance occupational safety in working environments?"

The insights obtained from the literature review played a crucial role in shaping the design decisions of the proposed system. Specifically, they guided the selection of the technical infrastructure, hardware compatibility requirements, and software architecture. Moreover, the findings informed subsequent stages of the study by providing a solid foundation for developing an effective, reliable, and application-oriented occupational safety monitoring system.

2. Stage: Image Data Collection

In the system development process, image data were obtained from publicly available datasets rather than direct on-site camera recordings. This approach was preferred to ensure ethical compliance, data accessibility, and reproducibility of the study. In this context, the Construction Site Safety dataset provided through the Roboflow platform was utilized as the primary data source. This dataset is widely used in occupational safety research and contains diverse visual scenarios representative of real-world industrial environments.

The dataset consists of hundreds of images captured from various angles and operational conditions, reflecting common safety-related situations encountered in industrial workplaces. These images include scenarios involving correct and incorrect use of personal protective equipment (PPE), worker–machine interactions, and different behavioral patterns that may pose potential safety risks. Such diversity is essential for training a robust and generalizable deep learning model capable of adapting to complex and dynamic environments.

Moreover, the dataset incorporates images collected under varying illumination conditions, including daylight, low-light, and partially shaded environments. This variability enhances the model's robustness against lighting-related challenges, which are frequently encountered in real-world industrial settings. In addition, samples from different sectors—such as construction sites and manufacturing facilities—were included to increase environmental diversity and improve the model's generalization capability.

All collected images were carefully reviewed to ensure consistency and relevance to the research objectives. The dataset was then organized and standardized within the Roboflow platform, enabling efficient dataset management and preprocessing in subsequent stages. To ensure data integrity and experimental reproducibility, the dataset was stored locally and used exclusively in an offline environment throughout the study.

Overall, this stage provided a comprehensive and diverse dataset that forms the foundation for accurate model training and evaluation. The richness and variability of the collected visual data play a crucial role in enabling the proposed system to effectively detect safety violations and adapt to different operational contexts. Figure 3. Sample outputs obtained from the proposed image-based occupational safety monitoring system. The model successfully detects safety helmets worn by workers and assigns confidence scores (e.g., 0.94). Bounding boxes indicate detected objects under real industrial working conditions.

Figure 3. Detection of Personal Protective Equipment Using the Proposed System

The dataset consists of unlabeled images belonging to four different classes. These images were transferred to the preprocessing stage to be reorganized and prepared for the subsequent annotation process in accordance with the project requirements. During this phase, particular attention was given to ensuring data quality and consistency to support effective model training. Low-quality or irrelevant samples were filtered out in order to enhance the reliability of the dataset.

The original version of the dataset was transferred to a local server environment and used entirely in an offline setting throughout the study. This approach ensured data security, eliminated dependency on external network connections, and enabled full control over the data processing workflow. Moreover, maintaining the dataset in a local environment contributed to compliance with data protection and ethical research considerations.

Following the preprocessing stage, the dataset was structured to support efficient training and testing of the proposed model. The prepared image set provided sufficient diversity in terms of visual content, environmental conditions, and object representation. As a result, the dataset effectively supported the development of a robust and generalizable image processing–based system.

At the end of this stage, a high-quality and well-structured dataset was obtained, forming a solid foundation for the subsequent training and evaluation phases. This preparation phase played a critical role in enhancing the overall accuracy, reliability, and generalization capability of the proposed occupational safety monitoring system.

3. Stage: Data Preprocessing and Annotation

At this stage, preprocessing and annotation procedures were performed to prepare the image dataset for effective model training. The images used in this study were obtained from the Roboflow platform and reorganized according to the specific requirements of the proposed system. Prior to model training, all images were carefully examined to ensure data quality and consistency.

Initially, a quality assessment process was conducted to evaluate image resolution, clarity, and illumination conditions. Low-quality or unsuitable images were excluded to prevent potential degradation of model performance. Subsequently, various data augmentation techniques were applied to enhance the diversity and robustness of the dataset. These techniques included image rotation, horizontal and vertical flipping, brightness and contrast adjustment, and the introduction of controlled noise. Such transformations are essential for improving the model’s ability to generalize across varying environmental conditions and unseen scenarios.

Data augmentation played a critical role in increasing the variability of the training dataset, thereby reducing overfitting and enhancing the

adaptability of the deep learning model. By exposing the model to diverse visual patterns, the system became more resilient to changes in lighting conditions, object orientation, and background complexity.

Following the preprocessing stage, the dataset was prepared for the annotation process. All images were formatted to meet the input requirements of the selected deep learning architecture, ensuring consistency and compatibility throughout the training pipeline. This structured preparation phase significantly contributed to improving the overall accuracy, robustness, and generalization capability of the proposed image-based occupational safety monitoring system.

Figure 1. Original Image

Figure 2. Image After Augmentation

For the annotation process, the LabelImg tool was employed to manually label the visual data. Each image was annotated according to predefined object classes, including “Hardhat”, “No-Hardhat”, “Excavator”, and “Person”. Bounding boxes were drawn around the relevant objects to precisely define their spatial locations within the images.

Particular attention was given to accurately labeling safety-related conditions, especially cases involving missing personal protective equipment such as safety helmets. This careful annotation process was essential to ensure reliable learning and accurate classification during model training. By explicitly distinguishing between compliant and non-compliant safety behaviors, the dataset was structured to support robust detection performance.

All annotations were saved in a format compatible with the YOLO architecture, ensuring seamless integration into the training pipeline. This meticulous labeling process significantly contributed to improving the

model's ability to detect safety violations and enhanced the overall reliability of the proposed occupational safety monitoring system.

Following the annotation process, all labeled data were organized in accordance with the YOLOv8 format, where annotation files in .txt format were paired with their corresponding .jpg images. This structure ensured compatibility with the training pipeline and enabled efficient data loading during model development.

Subsequently, the annotated dataset was divided into three subsets: training (73%), validation (16%), and testing (11%). This partitioning strategy was adopted to ensure reliable evaluation of the model's generalization capability and to prevent overfitting. The training set was used to learn feature representations, the validation set to fine-tune model parameters, and the test set to assess the final performance of the trained model on previously unseen data.

At the end of this stage, a high-quality, well-balanced, and diversified labeled dataset was obtained, providing a solid foundation for the subsequent model development phase. The structured separation of data ensured that the evaluation process accurately reflected real-world performance conditions.

To prepare the images for deep learning, a preprocessing pipeline was applied prior to model training. Images were resized to match the input dimensions required by the network architecture, and color channels were adjusted to ensure compatibility with the model's input format. Additionally, pixel values were normalized and converted into multi-dimensional tensor representations, enabling efficient numerical computation during training. When necessary, dimensional alignment operations were performed to maintain consistency across all input samples.

These preprocessing steps ensured that the input data were properly formatted and optimized for effective learning, thereby improving convergence stability and overall model performance.

4. Stage: Model Development and Training

After the annotated image dataset was prepared, the next stage focused on developing a model capable of detecting missing personal protective equipment (PPE) and identifying unsafe behaviors in industrial environments. For this purpose, the YOLOv8 (You Only Look Once) object detection algorithm was employed due to its high detection accuracy and real-time processing capability, which are essential for safety-critical applications.

Model development was conducted using the Python programming language, and the training environment was configured through the Ultralytics framework. The annotated dataset was defined using a .yaml configuration file, enabling seamless integration with the YOLOv8

architecture. Model training was initiated using the “detect” task supported by YOLOv8, allowing the system to learn object localization and classification simultaneously.

To accelerate convergence and improve generalization performance, a transfer learning approach was adopted. A pre-trained YOLOv8 model (yolov8l.pt) was fine-tuned using the project-specific dataset. This strategy significantly reduced training time while enhancing detection accuracy, particularly for safety-related object classes. The model was trained to recognize both compliant and non-compliant safety conditions, including the presence of safety helmets (“Hardhat”) and their absence (“No-Hardhat”).

Training was performed on the NVIDIA Jetson Nano platform, selected for its low power consumption and adequate computational capability for real-time inference tasks. This hardware configuration ensured that the developed system could be deployed in real-world industrial environments without requiring high-end computing infrastructure.

The dataset was divided into training (73%), validation (16%), and testing (11%) subsets to enable comprehensive performance evaluation. Throughout the training process, key performance metrics—including precision, recall, and mean Average Precision (mAP)—were continuously monitored to assess detection accuracy and model robustness. The results demonstrated that the trained model achieved high accuracy and reliable performance, effectively supporting real-time detection of safety-related events.

Overall, the training process confirmed that the proposed model is capable of operating efficiently in dynamic industrial environments while maintaining high detection accuracy, making it suitable for deployment in occupational safety monitoring applications.

The outputs generated by the trained model were stored for subsequent processing and prepared for integration into a Flask-based user interface. This integration enabled real-time analysis of incoming visual data obtained from the operational environment. As a result, the system became capable of detecting and interpreting potentially hazardous situations directly from live image streams, thereby supporting timely and effective decision-making.

To enable seamless deployment, the trained YOLOv5 model was incorporated into the application pipeline using a structured initialization process. Initially, the directory containing the trained model was defined within the system environment. Subsequently, the learned model weights (best.pt) were loaded, and essential parameters such as input image size and confidence thresholds were configured accordingly. The DetectMultiBackend structure was utilized to ensure compatibility with different hardware platforms, allowing the model to operate efficiently on either CPU or GPU architectures.

Prior to inference, a warm-up procedure was applied to the model to stabilize computational performance and reduce initial latency. This step ensured that the system could deliver consistent real-time predictions once deployed. Through this integration, the proposed framework achieved a seamless transition from model training to real-time operational deployment, enabling continuous monitoring and rapid detection of safety-related risks in industrial environments.

5. Stage: Performance Evaluation

Following the training phase, the performance of the proposed system was evaluated using several quantitative metrics to assess its accuracy, reliability, and robustness. This evaluation aimed to measure how effectively the trained YOLOv8 model performs in real-world operational conditions, particularly in detecting safety-related risks in industrial environments.

The primary performance metrics used in this evaluation include Precision, Recall, F1-score, and mean Average Precision (mAP). These metrics provide a comprehensive assessment of the model's detection capabilities and overall effectiveness.

Precision (92.7%) indicates the proportion of correctly identified positive detections among all detected instances. The high precision value demonstrates that the model produces a low number of false positives, particularly in identifying safety violations such as missing safety helmets.

Recall (77.4%) represents the proportion of actual positive cases that were correctly detected by the model. Although a small number of hazardous situations were not detected, the overall recall rate indicates a satisfactory sensitivity level for practical safety monitoring applications.

F1-score (84.3%), which represents the harmonic mean of precision and recall, reflects a balanced trade-off between detection accuracy and completeness. This result confirms that the model maintains consistent performance across different operational scenarios.

Mean Average Precision (mAP = 84.1%) was calculated across multiple Intersection over Union (IoU) thresholds. This metric demonstrates the model's ability to accurately localize objects and maintain consistent detection performance across varying classes and scene conditions.

During performance evaluation, the test dataset—comprising previously unseen images—was utilized to prevent overfitting and to objectively assess the generalization capability of the trained model. The results indicate that the model performs reliably under diverse environmental conditions.

In addition, the accuracy of the bounding boxes generated by the model was evaluated using the Intersection over Union (IoU) metric. The average IoU value was approximately 50%, indicating that the predicted bounding boxes were well-aligned with the actual object locations in most

cases. This level of spatial accuracy confirms the system's effectiveness in detecting and localizing safety-related elements within the scene.

Overall, the obtained performance metrics demonstrate that the proposed model is capable of accurately identifying hazardous situations in real-time. The results provide a strong foundation for subsequent risk scoring and prioritization processes, contributing to the development of an intelligent and reliable occupational safety monitoring system.

5. PERFORMANCE EVALUATION

Following the training phase, the performance of the proposed model was systematically evaluated to assess its accuracy, robustness, and reliability in real-world occupational safety scenarios. The evaluation focused on measuring how effectively the trained YOLOv8 model could detect safety violations—particularly the absence of personal protective equipment (PPE)—under realistic operational conditions.

To achieve this, several widely accepted performance metrics were employed, including **Precision**, **Recall**, **F1-score**, and **mean Average Precision (mAP)**. These metrics collectively provide a comprehensive assessment of classification accuracy, detection consistency, and localization performance.

5.1 Precision

Precision measures the proportion of correctly identified positive predictions among all predicted positives. In the context of occupational safety monitoring, this metric reflects how accurately the system identifies unsafe conditions—such as workers not wearing protective helmets—without generating excessive false alarms.

$$\text{Precision} = TP/(FP+TP)$$ where:

TP (True Positives) represents correctly detected unsafe conditions,

FP (False Positives) represents incorrectly detected unsafe conditions.

In the conducted experiments, the model achieved a **precision of 92.7%**, indicating that the majority of detected safety violations were correctly classified. For instance, among 43 detections labeled as "no helmet," 40 were true positives, while only 3 were false positives. This result demonstrates that the system maintains a low false alarm rate, which is critical for minimizing unnecessary interventions and maintaining operational efficiency.

5.2 Recall

Recall evaluates the model's ability to identify all actual unsafe cases present in the dataset. It reflects how effectively the system detects true safety violations without missing critical events.

$$Recall=TP/(FN+TP)$$ where:

FN (False Negatives) refers to actual unsafe instances that the model failed to detect.

The model achieved a **recall value of 77.4%**, indicating that approximately three-quarters of unsafe situations were correctly identified. While a small portion of risky cases remained undetected, the overall recall performance demonstrates a satisfactory level of sensitivity for real-world occupational safety monitoring. Nevertheless, reducing false negatives remains a critical objective, as missed detections may pose significant safety risks.

5.3 F1-Score

The F1-score represents the harmonic mean of precision and recall, providing a balanced measure of model performance when both false positives and false negatives are critical factors.

$$F1=2\times((Precision+Recall)/(Precision\times Recall))$$

Using the obtained precision (92.7%) and recall (77.4%) values, the resulting **F1-score was calculated as 84.3%**. This indicates that the model maintains a strong balance between accuracy and completeness, making it suitable for safety-critical applications where both reliability and sensitivity are essential.

5.4 Intersection over Union (IoU)

Intersection over Union (IoU) is a key metric used to evaluate the spatial accuracy of object localization. It measures the overlap between the predicted bounding box and the ground truth bounding box:

$$IoU= AreaIntersection / AreaUnion$$

In this study, IoU values were used to assess how accurately the model localized detected objects, such as workers and protective equipment. For example, given a ground truth bounding box area of 40 px^2, a predicted bounding box area of 50 px^2, and an intersection area of 30 px^2, the IoU value is calculated as:

$$IoU=30/60=0.50$$

An IoU threshold of 0.5 was considered acceptable for correct detection, consistent with common practices in object detection research. The model achieved an average IoU of approximately 50%, indicating that

the predicted bounding boxes were well-aligned with actual object locations in most cases.

5.5 Mean Average Precision (mAP)

Mean Average Precision (mAP) is a comprehensive performance metric widely used in object detection tasks. It represents the mean of average precision values computed across different object classes and IoU thresholds. $mAP=(1/n)\sum APi$ where n represents the number of evaluated classes or IoU thresholds, and AP_n denotes the average precision for each class.

In this study, mAP was calculated across multiple IoU thresholds (e.g., IoU > 0.50, 0.60, and 0.70). The resulting **mAP value of 84.1%** demonstrates that the model performs consistently well across different detection thresholds and environmental conditions. This result confirms the robustness and reliability of the proposed system for real-world occupational safety monitoring.

5.6 Overall Evaluation

The performance evaluation results indicate that the proposed model is highly effective in detecting safety-related risks within industrial environments. The combination of high precision, satisfactory recall, and strong mAP performance demonstrates the system's capability to operate reliably under diverse conditions.

Moreover, these performance outcomes provide a solid foundation for subsequent risk assessment and prioritization processes. The model's ability to accurately detect unsafe situations supports proactive safety management and contributes to the development of intelligent, data-driven occupational health and safety systems.

6. CONCLUSION

This study presented the development of an image processing–based occupational health and safety monitoring system designed to enhance worker safety and reduce the risk of occupational accidents in industrial environments. By integrating artificial intelligence and deep learning techniques, the proposed system enables real-time detection of unsafe conditions and non-compliance with safety regulations, thereby supporting proactive and data-driven safety management practices.

The developed system demonstrates strong potential in improving workplace safety by continuously monitoring operational environments and identifying risk factors that may lead to accidents. Through real-time alert

mechanisms and data-driven analysis, the system facilitates rapid intervention and supports decision-making processes aimed at preventing hazardous situations. In addition, the integration of intelligent monitoring contributes to optimizing operational workflows while maintaining a high level of occupational safety.

The experimental results indicate that the proposed approach achieves reliable performance in detecting safety-related risks under varying environmental conditions. The use of deep learning–based object detection enables accurate identification of personal protective equipment usage and unsafe behaviors, thereby enhancing the effectiveness of safety supervision. The obtained performance metrics confirm the feasibility and robustness of the proposed system for real-world applications.

Looking forward, the system can be further enhanced through the integration of advanced artificial intelligence models, more sophisticated image processing techniques, and adaptive learning mechanisms. Improvements in hardware performance and sensor technologies may also increase detection accuracy and system responsiveness. Furthermore, the adaptability of the proposed framework allows it to be extended to different industrial domains, contributing to the establishment of standardized and scalable occupational safety solutions.

In conclusion, the proposed image processing–based safety monitoring system represents a significant step toward intelligent and sustainable occupational health and safety practices. By combining modern artificial intelligence techniques with practical industrial applications, this study contributes to the advancement of safer, more efficient, and technology-driven working environments.

REFERENCE

Abdelgawad, M., & Fayek, A. R. (2018). Risk management in the construction industry using combined fuzzy FMEA and fuzzy AHP. Journal of Construction Engineering and Management, 144(1), 04017090. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001400

Fang, Q., Li, H., Luo, X., Ding, L., Luo, H., Rose, T. M., & An, W. (2018). Detecting non-hardhat-use by a deep learning method from far-field surveillance videos. Automation in Construction, 85, 1–9. https://doi.org/10.1016/j.autcon.2017.09.018

International Labour Organization (ILO). (2021). World employment and social outlook: Trends 2021. ILO Publications. https://www.ilo.org

İlter, P., & Kırelli, Y. (2025). Decision support systems based on machine learning for detection of liver diseases. Afyon Kocatepe Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 25(4), 827–837. https://doi.org/10.35414/akufemubid.1580440

KorkusuzPolat T. (2022), "Forecasting of production and scrap amounts using artificial neural networks". Emerging Materials Research, Vol. 11 No. 3 pp. 345–355, doi: https://doi.org/10.1680/jemmr.22.00036

Kumar, S., & Ghadge, A. (2019). Risk management in manufacturing: A literature review. International Journal of Production Research, 57(15–16), 4810–4830. https://doi.org/10.1080/00207543.2018.1518606

Lasi, H., Fettke, P., Kemper, H. G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. Business & Information Systems Engineering, 6(4), 239–242. https://doi.org/10.1007/s12599-014-0334-4

Memika, T., KorkusuzPolat, T. (2023). Internet of Things Supported Airport Boarding System and Evaluation with Fuzzy. Intelligent Automation & Soft Computing, 35(3), 2687–2702. https://doi.org/10.32604/iasc.2023.026955

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 779–788. https://doi.org/10.1109/CVPR.2016.91

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards real-time object detection with region proposal networks. Advances in Neural Information Processing Systems (NIPS), 28, 91–99.

Sahin, S., Ocak, M. A., & Uluyol, C. (2009). A systematic evaluation of preservice teachers' opinions on learning objects. International Journal of Human Sciences, 6(2), 723–736.

Stamatis, D. H. (2003). Failure mode and effect analysis: FMEA from theory to execution. ASQ Quality Press.

Yılmaz Yalçıner, A. ve Gelen Mert, M. B. (2021). "Estimating The Occupancy Rate of an Accommodation Business Using Artificial Neural Networks" Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, Sayı 47, Denizli, ss. 209-218.

Zhang, Y., Li, S., Liu, Q., & Wang, Y. (2021). Vision-based safety monitoring system for construction sites using deep learning. Safety Science, 136, 105155. https://doi.org/10.1016/j.ssci.2020.105155

Zhou, C., Ding, L., Luo, H., & Luo, X. (2020). Real-time detection of unsafe behaviors in construction sites using deep learning. Automation in Construction, 113, 103118. https://doi.org/10.1016/j.autcon.2020.103118

Innovative Solutions for Minimizing Municipal and Industrial Sludge: Towards Sustainable Wastewater Management

Mirac Nur CINER[*1]

Serdar AYDIN[2]

H. Kurtulus OZCAN[3]

Emine ELMASLAR ÖZBAŞ[4]

1- Res. Asst..; [1]Istanbul University-Cerrahpaşa, Engineering Faculty, Environmental Engineering Department. mirac.ciner@iuc.edu.tr ORCID No: 0000-0002-9920-928X
2- Prof. Dr.; [2]Istanbul University-Cerrahpaşa, Engineering Faculty, Environmental Engineering Department. saydin@iuc.edu.tr ORCID No: 0000-0002-0626-5224
3- Prof. Dr.; [3]Istanbul University-Cerrahpaşa, Engineering Faculty, Environmental Engineering Department. hkozcan@iuc.edu.tr ORCID No: 0000-0002-9810-3985
4- Prof. Dr.; [4]Istanbul University-Cerrahpaşa, Engineering Faculty, Environmental Engineering Department. elmaslar@iuc.edu.tr ORCID No: 0000-0001-9065-6684

ABSTRACT

Wastewater treatment processes generate substantial amounts of sludge containing a variety of components, including organic matter, nutrients, heavy metals, and pathogenic microorganisms. Municipal treatment sludge, characterized by its high organic content, has potential for reuse in biological processes, whereas industrial sludge often contains more complex chemical pollutants, such as heavy metals and toxic organic compounds. Globally, increasing urbanization and industrialization have led to significant annual increases in sludge production, making sustainable sludge management a critical objective. Improper sludge management results in severe environmental, economic, and logistical challenges, including ecosystem degradation, public health risks, and greenhouse gas emissions. This study aims to evaluate innovative methods for minimizing sludge production and to develop sustainable management strategies. Based on a literature review, the research highlights the importance of integrated treatment approaches tailored to the characteristics of different sludge sources. For municipal treatment sludge, biological and thermal processes such as thermophilic membrane reactors, liquid-phase bioconversion, and thermal hydrolysis are prominent. In contrast, industrial sludge shows significant potential for technologies such as vitrification, pyrolysis, and optimized coagulation-flocculation methods. The study underscores the necessity of employing innovative technologies to reduce the environmental impacts and enhance the economic efficiency of sludge management, emphasizing the importance of adopting a management approach aligned with global sustainability goals.

Keywords – Sewage Sludge Management, Municipal Sludge, Industrial Sludge, Sustainable Waste Management, Environmental Sustainability, Innovative Treatment Technologies

INTRODUCTION

Industrial and municipal treatment sludges are by-products of wastewater treatment processes, produced in substantial quantities, and contain various components such as organic matter, nutrients, heavy metals, and pathogenic microorganisms. Municipal treatment sludge is typically characterized by its high organic matter content, making it potentially beneficial for various biological processes (Chen & Chen, 2013). In contrast, industrial sludge exhibits a wide range of chemical diversity depending on its source and contains more complex components, such as heavy metals and toxic organic compounds (Mantis et al., 2005). The formation of treatment sludge varies based on the type of treatment processes employed in wastewater treatment plants (WWTPs), the efficiency of these processes, the

plant's capacity, and the physicochemical and biological properties of incoming wastewater. Furthermore, increasing industrialization, urbanization, and population growth not only result in higher wastewater production but also significantly elevate the annual quantities of both industrial and municipal sludge. This trend is particularly pronounced in large-scale WWTPs, where production volumes are increasing at a more intensive rate (Peccia & Westerhoff, 2015).

According to 2022 data from the Turkish Statistical Institute (TURKSTAT), WWTPs operated by municipalities in Turkey produced an annual total of 348.000 tons of sludge on a dry matter basis (TURKSTAT, 2022). Globally, for example, sludge production in China increased by 13% between year 2007 and 2013, reaching approximately 6.25 million tons. However, this amount is considered low compared to other developed countries, and dewatering followed by landfilling is the most commonly used disposal method (Yang et al., 2015). In the United States, approximately 9.5 million tons of dry treatment sludge were generated in 2011 from about 16.600 WWTPs (Hasanbeigi et al., 2012). A study conducted by Wiśniowska et al. (2019) compared sludge management regulations among major global and EU member countries. It identified population density, land availability, disposal costs, and social acceptance as key factors influencing sludge management. The study highlighted that sludge management regulations in developed countries are more comprehensive, emphasizing the need to minimize waste production or reuse sludge. Furthermore, landfilling is less preferred in developed countries, whereas developing countries primarily rely on landfilling and agricultural usage. In the EU-15 countries, agricultural usage and composting are the most common disposal methods for treatment sludge, while landfilling remains the dominant method in the EU-12 countries. Across the EU, the distribution of sludge disposal methods is as follows: 32% for agricultural use, 25% for landfilling, 13% for composting, and 13% for incineration (Kelessidis & Stasinakis, 2012). These data underline the critical importance of sustainable environmental policies in the management and disposal of wastewater treatment sludge.

Treatment sludge can provide environmental benefits and economic contributions through its utilization in various fields. These applications can be categorized into five main areas (Figure 1). The first involves open-field applications, offering potential use in agricultural lands, forests, degraded areas (e.g., coal and mining sites), and recreational spaces such as parks. The second category pertains to the utilization of dried sludge. Dried sludge can be employed as agricultural fertilizer and soil conditioner or utilized as an additional energy source by incineration for heat and electricity production, as

auxiliary fuel in cement production, or as a supplementary energy resource in coal-fired power plants. Third, the production of biogas from treatment sludge presents significant potential, with further processing enabling the production of methanol from the biogas. The fourth category encompasses energy conversion of sludge, often combined with other waste materials, through methods such as gasification and pyrolysis. Finally, innovative methods, including wet oxidation, hydrolysis, hydrothermal oxidation, and microwave-high-pressure processing, are emerging as notable approaches for sludge utilization. These methods optimize sludge management processes and contribute to sustainable waste management strategies (Kelessidis & Stasinakis, 2012; SPD, 2023).

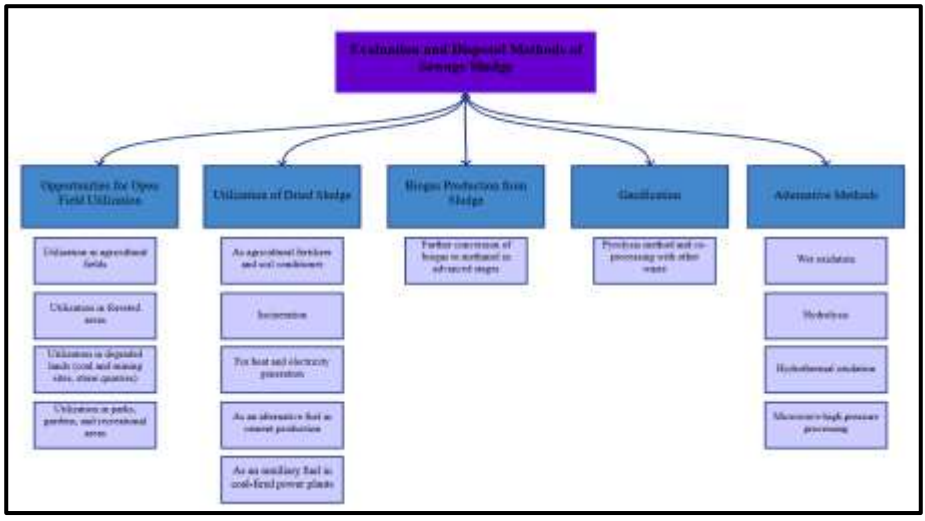


Figure 1. Evaluation of Disposal Methods of Sewage Sludge

Different management methods for treatment sludge can provide environmental benefits while also posing various challenges and risks. Each of these methods presents advantages and disadvantages depending on the intended use, geographical conditions, and economic feasibility. For instance, open-field applications can improve the physical and chemical properties of soil in agricultural areas but may lead to adverse effects such as the spread of heavy metals and pathogens into the environment. Similarly, biogas production offers energy savings but requires high initial investment costs.

The mismanagement of treatment sludge can lead to multifaceted and severe environmental, economic, and logistical problems. From an environmental perspective, heavy metals, organic pollutants, pathogens, and

microorganisms present in sludge can infiltrate soil, surface waters, and groundwater, disrupting natural ecosystems and adversely affecting aquatic life. This may result in ecosystem imbalances, reduced agricultural productivity, and a decline in biodiversity. Furthermore, the introduction of these harmful substances into the food chain poses threats to both animal and human health. The accumulation of heavy metals in soil and water can lead to toxic effects, while the presence of pathogens may cause outbreaks of infectious diseases (Cieślik et al., 2015). Additionally, improper disposal methods can lead to the decomposition of sludge, releasing methane and other greenhouse gases, which negatively impact climate change. Therefore, an effective management strategy is essential to minimize the environmental impacts of sludge. Economically, mitigating such pollution requires costly remediation processes, and the inability to recover resources from sludge reduces economic efficiency (Murray et al., 2008). Logistically, particularly in large urban areas, the transportation, storage, and disposal of sludge can pose significant challenges due to inadequate infrastructure. To address these issues, the implementation of sustainable management strategies and technologies is imperative.

This study aims to provide a comprehensive evaluation by focusing on the selection of treatment methods that target minimal sludge production for both industrial and domestic sources. Based on a literature review, the research emphasizes the importance of adopting sustainable and integrated treatment approaches tailored to the characteristics of sludge produced from different sources. For municipal treatment sludge, biological and thermal processes such as thermophilic membrane reactors, liquid-phase bioconversion, and thermal hydrolysis (TH) are highlighted. In contrast, vitrification, pyrolysis, and optimized coagulation-flocculation methods emerge as prominent options for industrial sludge. The study underscores the necessity of employing innovative technologies in treatment processes to achieve both environmental and economic sustainability.

CASE STUDIES ON SLUDGE MINIMIZATION IN MUNICIPAL WASTEWATER: A LITERATURE REVIEW

Table 1 summarizes the most common management methods for treatment sludge along with their respective advantages and disadvantages.

Table 1. Overview of Sewage Sludge Management Methods: Categories, Methods, Advantages, Disadvantages

Category	Methods	Advantages	Disadvantages	References
Use in Agriculture and Soil Reclamation	Stabilization in Drying Beds	Removes pathogenic bacteria; prepares sludge for use as fertilizer or soil amendment; lower investment and operational costs compared to thermal processes.	Untreated sludge may cause physical changes in the soil; drying is essential.	Uggetti et al., 2009
	Stabilization Using Earthworms	Reduces organic carbon content; increases bioavailable phosphorus; lowers carbon-to-nitrogen ratio, enhancing fertilizer quality; reduces heavy metal content.	Presence of heavy metals in industrial wastewater may limit its use; requires controlled conditions.	Suthar, 2010
	Anaerobic Stabilization with Biogas Recovery	Produces methane-rich biogas; reduces organic carbon content; renewable energy generation; contributes to reduction CO_2 emission suitable for agriculture and soil reclamation.	Requires optimal conditions for methane fermentation; limited by economic and geographic constraints.	Mills et al., 2014; Xu et al., 2014; Lag-Brotons et al., 2014
Thermal Processing	Drying	Reduces water content; prepares sludge for use as fuel or construction material; reduces operating costs with biodrying.	May require additional energy; low-calorific products require further processing for practical applications.	Winkler et al., 2013
	Conventional Incineration	Significantly reduces sludge volume; allow for cutting operating costs; generates energy; heat energy can be used for heating or clinker production.	Requires pre-drying (18–35% dry solids); requires management of ash and exhaust gases.	Samolada & Zabaniotou, 2014; Donatello & Cheeseman, 2013; Winkler et al., 2013

Category	Methods	Advantages	Disadvantages	References
	Co-Incineration	Enhances energy efficiency by co-burning sludge with fuels; contains bioavailable phosphorus for agricultural uses.	Additional pollutants may enter final products; requires chemical and ecotoxicological analysis of ash.	Donatello & Cheeseman, 2013; Liu et al., 2011
Use in the Construction Industry	Cementing	Immobilizes environmental contaminants; ash (ash cement properties depend on the reactants) is solidified with cement for use in building or road materials; the optimal ash ratios were 10% for cement substitution and 2% for sand substitution; reduces storage volume.	Sometimes strength of material prepared in such way is unsatisfying	Wu et al., 2011; Chen et al., 2013; Donatello & Cheeseman, 2013
	Vitrification	Stabilizes heavy metals at high temperatures; converts toxic organics into inorganics; inorganic compounds and pose no threat to the environment	High energy demand increases cost; requires comprehensive analysis to ensure long-term environmental safety.	Bernardo & Dal Maschio, 2011
Raw Materials Recovery	Pyrolysis	Produces bio-oil and adsorbents; A material processed this way features a large surface area; creates economically viable products; avoids oxygen combustion.	Requires thorough analysis to ensure product safety and environmental compliance.	Spisona et al., 2011; Samolada & Zabaniotou, 2014
	Phosphorus Recovery	Recovers valuable phosphorus for agricultural and industrial use; adaptable to various extraction methods.	Presence of heavy metals may limit recovery for agriculture; high energy and cost requirements.	Donatello et al., 2010; Cordell et al., 2009

Category	Methods	Advantages	Disadvantages	References
	Rare Earth Metal Recovery	Enables recovery of valuable metals; adds economic value as strategic raw materials; suitable for industrial applications.	Requires high energy costs; demands specialized technology and infrastructure.	Osaka & Jung, 2007

Collivignarelli et al. (2017) conducted a study aimed at reducing sludge production from municipal WWTPs using thermophilic membrane reactor (TMR) technology. Conducted on a pilot-scale facility, the study optimized operational parameters such as aeration and hydraulic retention time (HRT) to achieve volatile suspended solids (VSSs) and chemical oxygen demand (COD) removal rates exceeding 80%. Optimization was performed through various trials to determine the durations of aerated and non-aerated cycles. The best results were achieved with an HRT below 15 days, combined with 2-hour aeration and 6-hour non-aeration cycles. These cycles regulated microbial metabolism to adapt to both aerobic and anaerobic conditions, thereby maximizing sludge reduction. Under thermophilic conditions, these aeration strategies promoted cellular disintegration, accelerated biological oxidation processes, and enhanced the solubility of organic matter. Ammonia produced during the process was recovered as fertilizer through stripping followed by absorption. Additionally, the high biological degradability of the permeate allowed its use as a carbon source in a denitrification unit. Consequently, this technology emerged as a solution that limits sludge production through biological oxidation and cellular lysis while supporting both economic and environmental sustainability. The experiments, conducted on sludge from municipal wastewater, demonstrated similar or superior results compared to previous studies on industrial sludge (Collivignarelli et al. 2017).

The study by Malhotra and Garg (2019) utilized TH and wet oxidation (WO) processes to enhance the solubility of waste activated sludge (WAS) and improve sludge management. The research was conducted using secondary wastewater sludge obtained from a municipal WWTP in Navi Mumbai, India. Operational parameters such as temperature, duration, pH, and oxidation coefficient were optimized using the Box-Behnken design. During these processes, the sludge was treated at temperatures ranging from 140 to 180°C, for durations of 2 to 5 hours, and within a pH range of 3.3 to 9.3. TH and WO processes facilitated the breakdown of cell walls, enhancing the solubility of organic matter. Optimal conditions were achieved at a temperature of 180°C, a 5-hour processing time, pH 3.3, and an oxidation coefficient of 0.5. Under these conditions, a total chemical oxygen demand (tCOD) reduction of 58%

and a VSSs solubility of 52% were observed. The study demonstrated multiple benefits of thermal processes. Increased sludge solubility released macromolecules such as proteins, carbohydrates, and volatile fatty acids (VFAs), which showed significant potential for boosting biogas production in anaerobic digestion processes. Additionally, the use of acidic pH reduced the color intensity of the solution and improved filtration properties. Ultimately, the process resulted in a reduction in sludge volume and a decrease in disposal costs. TH and wet oxidation processes emerged as effective methods for environmentally and economically sustainable sludge management. These methods not only improved sludge solubility but also enhanced disposal and energy recovery processes (Malhotra & Garg, 2019).

The study by Rahman et al. (2016) investigated the reduction of sludge from municipal wastewater treatment plants (MWWTPs) using the Liquid State Bioconversion (LSB) process. This process utilized a mixed fungal inoculum comprising Aspergillus niger and Penicillium corylophilum. The study was conducted in a pilot-scale bioreactor, and parameters such as hydraulic retention time (HRT) and substrate input concentration (S_0) were optimized using Response Surface Methodology (RSM). The optimal conditions were determined as an HRT of 3.62 days and an S_0 of 10.12 g/L. Under these conditions, a COD removal efficiency of 97% was achieved. Sludge reduction was facilitated by the mixed fungal inoculum dissolving organic matter and making biosolids more biodegradable. At the end of the process, a sludge volume reduction of up to 70% was achieved, and the water content of the sludge decreased, improving solid-liquid separation. Additionally, filtration resistance was reduced, enhancing process efficiency. Optimizing HRT shortened the process duration, reducing economic and operational costs. Furthermore, the method contributed significantly to environmental sustainability, enabling sludge management to be conducted in an environmentally friendly manner. These findings demonstrate that the LSB method offers an effective solution for reducing sludge generated by MWWTPs. The study highlights both the economic and environmental benefits of this approach (Rahman et al., 2016).

The study by Chen et al. (2015) applied a multi-objective optimization method to the "Step A^2/O Activated Sludge Process" (SA^2/OCM), developed to improve active sludge processes in MWWTPs. This system combines the advantages of Sequencing Batch Reactors (SBR) and conventional A^2/O processes, offering an eight-stage reactor configuration for the biological removal of nitrogen, phosphorus, and carbon. The study optimized parameters such as effluent quality (EQ) and operating costs (OC) while evaluating system performance under dynamic influent conditions. During the

optimization, an open-loop strategy (OLO) was compared with three closed-loop strategies (TC_{NO3}, TC_{NH4}, TC_{PO4}). Operational parameters such as oxygen transfer rate (K_La_i), sludge return ratio (R_{Qr}), and sludge retention time (SRT) were optimized to enhance efficiency. Notably, the T_{CNO3} strategy, using a time controller based on nitrate concentration, enabled automatic adjustment of reaction times according to dynamic influent conditions. The innovative approaches demonstrated that closed-loop strategies reduced waste sludge production, thereby lowering sludge disposal costs. Specifically, the OLO strategy reduced OC by 24.13%. The TC_{NO3} strategy improved nitrogen and phosphorus removal and enhanced EQ, particularly under variable influent conditions. Additionally, the optimized process conditions significantly decreased operating costs. This study presents an economically and environmentally sustainable solution for active sludge management in MWWTPs, emphasizing that the TC_{NO3} strategy effectively improved process performance and reduced sludge production under variable conditions (Chen et al., 2015).

Nazif et al. (2016) evaluated the performance of the Oxic-Settling-Anoxic (OSA) process combined with thermal and mechanical treatments in reducing sludge generated from MWWTPs. The OSA system was optimized to facilitate sludge cell disintegration and accelerate endogenous oxidation processes under low oxidation-reduction potential (ORP) conditions. For thermal treatments, temperatures of 50°C, 70°C, and 90°C were tested, while mechanical treatments were applied at various voltage levels (80V, 120V, and 150V). This combination effectively reduced sludge by promoting floc disintegration and cell lysis. The results indicated that the OSA system alone reduced sludge production by 24%. Integrating thermal treatment increased sludge reduction to up to 46%, while mechanical treatment alone achieved a 34% reduction. Optimal conditions were identified as 90°C and 80V, under which both sludge reduction was maximized, and EQ was maintained. Additionally, methane gas produced during the thermal process was utilized as an energy source, improving energy efficiency. The system achieved a COD removal efficiency of 88%, with thermal and mechanical treatments having no adverse effects on this efficiency. These findings demonstrate that the OSA system, when combined with thermal and mechanical treatments, is an effective method for sludge reduction. Furthermore, this innovative approach offers a highly applicable solution for MWWTPs, meeting economic and environmental sustainability criteria (Nazif et al., 2016).

The study by Beck et al. (2005) focused on optimizing activated sludge processes (ASPs) in WWTPs to address issues caused by high organic loads during grape harvesting seasons. The research was conducted at the

Beblenheim WWTP in the Alsace region, where municipal wastewater is mixed with high organic content wastewaters from vineyards. This facility, with a capacity of 20,000 population equivalents, experiences efficiency losses in its treatment processes due to seasonal fluctuations in organic loads from vineyards. The study investigated the reconfiguration of biological stages and the effects of various process modifications. Simulations evaluated scenarios such as merging two aeration tanks, adding a primary sedimentation tank, and integrating a buffer tank into the treatment line. The buffer tank proved particularly effective in balancing the high organic loads during the grape harvest period and contributed to the preliminary treatment of organic matter. Dividing the aeration tanks and adding a primary sedimentation tank reduced sludge volume and decreased the ratio of settleable solids. These optimizations resulted in up to a 50% improvement in BOD_5 and COD removal. Additionally, simulation results demonstrated that optimized aeration strategies reduced energy consumption and enhanced the overall efficiency of the facility. The study highlights that the use of buffer tanks and stepwise biological systems offers an effective solution for managing organic loads and reducing sludge in ASPs. This approach provides a sustainable solution, especially for facilities dealing with seasonal organic load fluctuations (Beck et al., 2005).

The study by Iticescu et al. (2015) aimed to reduce the environmental impact and promote the sustainable use of sludge generated by the MWWTP in Galati, Romania. The research evaluated the use of treatment sludge mixed with metallurgical coke as an energy source and raw material in iron agglomeration processes. This approach sought to optimize energy recovery from sludge and reduce coal consumption in metallurgical processes. The chemical composition and calorific value of the sludge were analyzed. Two types of sludge were used: anaerobically stabilized fermented sludge and non-fermented sludge. The sludge underwent mechanical drying and thermal treatment, followed by mixing with metallurgical coke at a 10% ratio. These mixtures were employed as fuel in iron ore sintering processes. The analysis revealed that the sludge had a high carbon content (46.02 wt%), demonstrating significant potential for energy recovery. The results showed that using sludge in metallurgical processes significantly reduced sludge volume and minimized the amount of sludge requiring disposal. Mixtures prepared with a 10% sludge ratio optimized energy recovery by reducing coal consumption in sintering processes. This method also provided economic benefits and demonstrated that repurposing sludge as a raw material contributed to environmental sustainability. The study concluded that municipal sludge can be effectively utilized in the metallurgical industry, offering a viable solution for sludge reduction and energy recovery (Iticescu et al., 2015).

The study by Andreottola and Foladori (2006) examined the effectiveness of advanced physicochemical methods, such as sonolysis (ultrasonic disintegration) and alkaline thermolysis (a combination of high pH and temperature), in minimizing excess sludge production in MWWTPs. The sludge used in the study was obtained from a WWTP in the Trento Nord region of Italy and evaluated in activated sludge bioreactors. The research revealed that these methods enhanced sludge solubilization, thereby improving its biological degradation potential. The sonolysis method used ultrasonic waves at a frequency of 20 kHz to break down biological flocs, which increased the release of colloidal and soluble COD fractions. The alkaline thermolysis method, applied at a pH of 12 and temperatures of 60–80°C, facilitated the solubilization of 38% of the total COD into a dissolved form. Both methods promoted the dissolution of sludge particles, the breakdown of cell walls, and microbial lysis, enhancing sludge biodegradability. The results demonstrated that sonolysis could reduce sludge production by up to 30% and could be applied with moderate specific energy, optimizing energy costs. Similarly, the alkaline thermolysis method achieved comparable sludge reduction even under low-cost conditions. The increase in dissolved COD also indicated significant potential for enhanced biogas production during anaerobic digestion processes. These physicochemical treatments supported economic and environmental sustainability by reducing the costs associated with sludge disposal, including storage, transportation, and drying. The study highlights that sonolysis and alkaline thermolysis offer effective solutions for reducing municipal sludge while supporting biological processes and providing both economic and environmental benefits (Andreottola & Foladori, 2006).

The study by Zhu and Chen (2011) developed a method to reduce nitric oxide (NO) and nitrous oxide (N_2O) emissions using anaerobic-aerobic biological processes, with alkaline fermentation liquid derived from MWWTP sludge serving as a carbon source. The use of fermentation liquid was compared to traditional acetic acid as a carbon source. The sludge used in the research was obtained from a MWWTP in Shanghai, China. In this method, the fermentation liquid was processed at pH 10 and enriched with short-chain fatty acids (SCFAs). The use of fermentation liquid reduced NO and N_2O formation during the denitrification process. The study demonstrated that propionic acid and copper ions (Cu^{2+}) in the fermentation liquid enhanced the activity of the nitrous oxide reductase enzyme, playing a crucial role in reducing nitrogen oxide emissions. Experimental results showed that fermentation liquid produced 68.7% less N_2O and 50% less NO compared to acetic acid. The findings revealed that using fermentation liquid as a carbon source not only reduced sludge disposal but also promoted waste reuse and improved process performance by enhancing total nitrogen (TN) and total

phosphorus (TP) removal efficiencies (77.5% TN, 98.9% TP removal). Additionally, significant reductions in greenhouse gas emissions, such as NO and N_2O, were achieved. This innovative approach also reduced energy consumption, offering substantial economic and environmental sustainability benefits. This study demonstrates that utilizing municipal sludge as fermentation liquid is an effective solution for sludge reduction and greenhouse gas emissions control (Zhu & Chen, 2011).

The study by Jiang et al. (2021) focused on reducing the volume of sludge generated from MWWTPs through thermochemical methods while recovering energy. The research comprehensively examined various thermochemical processes, including combustion, pyrolysis, gasification, and hydrothermal carbonization, and analyzed municipal sludge from urban WWTPs in China. The primary objective of the study was to develop innovative methods that enhance economic efficiency while ensuring environmental sustainability in sludge management. Thermochemical methods were found to reduce sludge volume rapidly and effectively. Combustion, particularly mono-combustion, reduced sludge volume by up to 70%, significantly lowering storage requirements. Gasification emerged as a cost-effective method with minimal environmental risks, facilitating renewable energy production through syngas generation. Hydrothermal carbonization converted the organic components of sludge into carbon-rich fuels, providing economic and environmental benefits. Additionally, optimization techniques such as atomization and high-pressure air application enhanced process efficiency. The results demonstrated that thermochemical methods not only addressed sludge disposal but also reduced environmental harm through energy recovery and heavy metal removal. These methods minimized the impacts of toxic components while increasing the economic value of by-products, thus reducing overall costs. In conclusion, the study highlighted that thermochemical methods offer effective and sustainable solutions for volume reduction and energy recovery in municipal sludge management (Jiang et al., 2021).

The study by Mannina et al. (2020) introduces a multi-objective optimization model aimed at enhancing the efficiency of sludge management in membrane bioreactors (MBRs). The research was conducted on sludge derived from a mixture of municipal wastewater. Specifically, the optimization of operational parameters, such as sludge retention time (SRT), resulted in a reduction in sludge production. This reduction was attributed to the optimized SRT, which minimized cell growth and sludge formation. Additionally, the application of optimal operational parameters achieved a 48% reduction in operational costs and a 10% decrease in direct greenhouse

gas emissions. The study demonstrates that MBRs, when operated with carefully optimized parameters, offer an effective solution for sustainable sludge management and reducing the environmental footprint (Mannina et al., 2020).

The study by Chu et al. (2009) investigated the effectiveness of ozonation technology in sludge reduction and evaluated its impact on sludge management. The research was conducted on sludge generated from the treatment of municipal and industrial wastewater. Ozonation effectively reduced sludge production by promoting sludge lysis and subsequent cryptic growth. Ozone dosage was optimized between 0.03 and 0.05 g O_3/g total suspended solids (TSSs), achieving both sludge reduction and cost-effectiveness. This method enhanced the biodegradability of treated sludge and improved its settling properties. The contribution of ozonation to sludge reduction was attributed to the disintegration of cells, leading to the removal of inorganic and inert particles and resulting in lower sludge volumes. However, appropriate sludge return rates must be determined to prevent the long-term accumulation of inorganic particles in the process. The study demonstrated that ozonation is an effective sludge management strategy for facilities facing high sludge disposal costs or operational issues such as foaming or volume expansion (Chu et al., 2009).

CASE STUDIES ON SLUDGE MINIMIZATION IN INDUSTRIAL WASTEWATER: A LITERATURE REVIEW

The study by Kılıç et al. (2011) focused on an oxidative method for recovering chromium from tannery sludge and reducing its environmental impact. An advanced oxidation process using hydrogen peroxide was applied to convert chromium from its Cr(III) form to Cr(VI), which was subsequently reduced back to the less toxic Cr(III) form for recovery. This method not only facilitated the reuse of chromium but also reduced the toxic properties of the sludge. The recovery process minimized the environmental impact of the sludge, decreasing the amount of waste requiring disposal. The study utilized sludge from tannery facilities with a chromium concentration of approximately 8041 mg/kg. The total chromium removal efficiency was reported to be 70%. Additionally, the environmental benefits of producing chromium sulfate from the recovered chromium highlighted the system as an economically and environmentally sustainable solution. The study emphasized that scaling up the laboratory-scale results to an industrial level could optimize chemical usage and further reduce environmental impacts. Therefore, chromium recovery from tannery sludge emerges as a method offering both environmental and economic advantages (Kiiıç et al., 2011).

The study by Kusiak and Wei (2013) developed a multi-objective model to optimize the ASP and applied it to enhance treatment efficiency while reducing energy consumption. This research focused on sludge generated from the treatment of industrial wastewater, aiming to optimize dissolved oxygen (DO) levels to balance energy savings and EQ. The study employed a multilayer perceptron (MLP)-based artificial neural network model to analyze data and implemented a Multi-Objective Particle Swarm Optimization (MOPSO) algorithm. Through this approach, optimizing the air flow rate achieved a 15% reduction in energy consumption while effectively controlling TSSs and carbonaceous biochemical oxygen demand (CBOD) levels. Sludge reduction was achieved by appropriately adjusting DO levels and optimizing the aeration process. These optimizations enhanced the efficiency of microbial metabolism and minimized unnecessary sludge production. The study demonstrated that optimizing activated sludge systems provides a sustainable and cost-effective method for industrial waste management, delivering both energy savings and improved environmental performance (Kusiak & Wei, 2013).

The study by Wang et al. (2021) addressed the reduction of wastewater sludge generated from industrial printed circuit board (PCB) manufacturing facilities through optimized chemical dosing in the coagulation-flocculation process. The research employed data-driven modeling techniques, such as genetic algorithms (GA) and particle swarm optimization (PSO), to address heavy metal contamination (particularly copper) in PCB production. The coagulation-flocculation process utilized chemicals such as sodium sulfide (Na_2S) and ferric sulfate ($FeSO_4$), with optimal chemical dosages determined through modeling. Multiple regression analysis was employed to examine the relationship between copper removal efficiency and sludge levels. The findings revealed copper removal efficiencies exceeding 90%, while optimized chemical dosages minimized sludge volume, reducing disposal costs. Specifically, optimization of chemical dosages achieved a 10% reduction in treatment costs. The optimization model maintained sludge levels within the targeted range, enabling more efficient solid-liquid separation during treatment. This approach also enhanced energy efficiency and contributed to environmental sustainability. The study demonstrated that data-driven optimization of coagulation-flocculation processes offers an effective sludge reduction strategy for industrial facilities, such as PCB manufacturing (Wang et al., 2021).

CONCLUSIONS

Studies on the management of municipal and industrial wastewater treatment sludge have demonstrated that advanced treatment technologies and optimization strategies provide significant achievements in sludge reduction. Methods such as TMBRs, TH, and LSB employed in WWTPs have substantially reduced sludge volume. These technologies enhanced the solubility of organic matter and supported nutrient recovery by improving microbial activity and optimizing operational parameters. Additionally, innovative approaches like A^2/O processes and ozonation effectively minimized excess sludge production, improved energy efficiency, and maintained EQ. These methods offer effective solutions that meet economic and environmental sustainability criteria.

Studies on the reduction of industrial sludge have highlighted the environmental and economic benefits of innovative approaches. Oxidative chromium recovery methods have not only mitigated the toxic effects of heavy metals but also facilitated their reuse, making waste management more sustainable. Optimization of ASPs has reduced energy consumption and prevented unnecessary sludge production, thereby enhancing process efficiency. Additionally, optimized chemical dosing in coagulation-flocculation processes has achieved high heavy metal removal efficiency while significantly reducing sludge volume and disposal costs. Overall, data-driven modeling, thermochemical methods, and optimized physicochemical processes contribute significantly to reducing sludge volume, supporting energy recovery, and promoting environmental sustainability in industrial sludge management. These methods provide effective solutions for minimizing environmental impacts and enhancing economic efficiency in industrial waste management.

These studies highlight the critical importance of tailoring sludge management methods to the specific characteristics of wastewater in municipal and industrial contexts. The integrated application of biological, chemical, and thermochemical methods not only reduces sludge volume but also provides additional benefits such as biogas production, nutrient recovery, and energy savings. Innovative technologies and systematic optimization techniques enable process parameters to adapt to dynamic conditions, facilitating the achievement of both economic and environmental sustainability goals. Such integrated and optimized approaches in sludge management contribute to long-term sustainable development by enhancing energy efficiency and resource recovery potential.

Future research should focus on enhancing the scalability of these solutions for industrial applications, developing hybrid technologies adaptable to various wastewater types, and conducting more detailed assessments of the environmental impacts of these methods. This approach will enable the achievement of more effective and sustainable outcomes on a global scale in the management of municipal and industrial wastewater sludge.

REFERENCES

Andreottola, G., & Foladori, P. (2006). A review and assessment of emerging technologies for the minimization of excess sludge production in wastewater treatment plants. *Journal of Environmental Science and Health Part A*, *41*(9), 1853-1872.

Beck, C., Prades, G., & Sadowski, A. G. (2005). Activated sludge wastewater treatment plants optimisation to face pollution overloads during grape harvest periods. *Water Science and Technology*, *51*(1), 81-88.

Bernardo, E., & Dal Maschio, R. (2011). Glass–ceramics from vitrified sewage sludge pyrolysis residues and recycled glasses. *Waste Management*, *31*(11), 2245-2252.

Chen, M., Blanc, D., Gautier, M., Mehu, J., & Gourdon, R. (2013). Environmental and technical assessments of the potential utilization of sewage sludge ashes (SSAs) as secondary raw materials in construction. *Waste management*, *33*(5), 1268-1275.

Chen, S., & Chen, B. (2013). Net energy production and emissions mitigation of domestic wastewater treatment system: A comparison of different biogas–sludge use alternatives. *Bioresource Technology*, *144*, 296-303.

Chen, W., Lu, X., & Yao, C. (2015). Optimal strategies evaluated by multi-objective optimization method for improving the performance of a novel cycle operating activated sludge process. *Chemical Engineering Journal*, *260*, 492-502.

Chu, L., Yan, S., Xing, X. H., Sun, X., & Jurcik, B. (2009). Progress and perspectives of sludge ozonation as a powerful pretreatment method for minimization of excess sludge production. *Water research*, *43*(7), 1811-1822.

Cieślik, B. M., Namieśnik, J., & Konieczka, P. (2015). Review of sewage sludge management: standards, regulations and analytical methods. *Journal of cleaner production*, *90*, 1-15.

Collivignarelli, M. C., Abbà, A., Castagnola, F., & Bertanza, G. (2017). Minimization of municipal sewage sludge by means of a thermophilic membrane bioreactor with intermittent aeration. *Journal of Cleaner Production*, *143*, 369-376.

Cordell, D., Drangert, J. O., & White, S. (2009). The story of phosphorus: global food security and food for thought. *Global environmental change*, *19*(2), 292-305.

Donatello, S., & Cheeseman, C. R. (2013). Recycling and recovery routes for incinerated sewage sludge ash (ISSA): A review. *Waste Management*, *33*(11), 2328-2340.

Donatello, S., Tong, D., & Cheeseman, C. R. (2010). Production of technical grade phosphoric acid from incinerator sewage sludge ash (ISSA). *Waste management*, *30*(8-9), 1634-1642.

Hasanbeigi, A., Lu, H., Williams, C., & Price, L. (2012). *International best practices for pre-processing and co-processing municipal solid waste and sewage sludge in the cement industry*. Ernest Orlando Lawrence Berkeley National Laboratory.

Iticescu, C., Georgescu, L. P., Gurau, G., Murarescu, M., Dima, D., Murariu, G., & Gheorghies, C. (2015). Methods to reduce environmental impact of municipal waste water sewage sludge. *Environmental Engineering & Management Journal (EEMJ)*, *14*(10).

Jiang, G., Xu, D., Hao, B., Liu, L., Wang, S., & Wu, Z. (2021). Thermochemical methods for the treatment of municipal sludge. *Journal of Cleaner Production*, *311*, 127811.

Jung, C. H., & Osako, M. (2007). Thermodynamic behavior of rare metals in the melting process of municipal solid waste (MSW) incineration residues. *Chemosphere*, *69*(2), 279-288.

Kelessidis, A., & Stasinakis, A. S. (2012). Comparative study of the methods used for treatment and final disposal of sewage sludge in European countries. *Waste management*, *32*(6), 1186-1195.

Kiliç, E., Puig, R., Baquero, G., Font, J., Colak, S., & Gürler, D. (2011). Environmental optimization of chromium recovery from tannery sludge using a life cycle assessment approach. *Journal of hazardous materials*, *192*(1), 393-401.

Kusiak, A., & Wei, X. (2013). Optimization of the activated sludge process. *Journal of Energy Engineering*, *139*(1), 12-17.

Lag-Brotons, A., Gómez, I., Navarro-Pedreño, J., Mayoral, A. M., & Curt, M. D. (2014). Sewage sludge compost use in bioenergy production–a case study on the effects on Cynara cardunculus L energy crop. *Journal of cleaner production*, *79*, 32-40.

Liu, Q., Jiang, P., Zhao, J., Zhang, B., Bian, H., & Qian, G. (2011). Life cycle assessment of an industrial symbiosis based on energy recovery from dried sludge and used oil. *Journal of Cleaner Production*, *19*(15), 1700-1708.

Malhotra, M., & Garg, A. (2019). Performance of non-catalytic thermal hydrolysis and wet oxidation for sewage sludge degradation under moderate operating conditions. *Journal of environmental management*, *238*, 72-83.

Mannina, G., Ni, B. J., Rebouças, T. F., Cosenza, A., & Olsson, G. (2020). Minimizing membrane bioreactor environmental footprint by multiple objective optimization. *Bioresource technology*, *302*, 122824.

Mantis, I., Voutsa, D., & Samara, C. (2005). Assessment of the environmental hazard from municipal and industrial wastewater treatment sludge by employing chemical and biological methods. *Ecotoxicology and environmental safety*, *62*(3), 397-407.

Mills, N., Pearce, P., Farrow, J., Thorpe, R. B., & Kirkby, N. F. (2014). Environmental & economic life cycle assessment of current & future sewage sludge to energy technologies. *Waste management*, *34*(1), 185-195.

Murray, A., Horvath, A., & Nelson, K. L. (2008). Hybrid life-cycle environmental and cost inventory of sewage sludge treatment and end-use scenarios: a case study from China.

Nazif, S., Mehrdadi, N., Zare, S., & Mosavari, S. (2016). Improvement and prediction of OSA system performance in sludge reduction through integration with thermal and mechanical treatment. *Water Science and Technology*, *74*(9), 2087-2096.

Peccia, J., & Westerhoff, P. (2015). We should expect more out of our sewage sludge.

Rahman, R. A., Molla, A. H., Barghash, H. F., & Fakhru'l-Razi, A. (2016). Optimization of process parameters for pilot-scale liquid-state bioconversion of sewage sludge by mixed fungal inoculation. *Environmental technology*, *37*(1), 1-15.

Samolada, M. C., & Zabaniotou, A. A. (2014). Comparative assessment of municipal sewage sludge incineration, gasification and pyrolysis for a sustainable sludge-to-energy management in Greece. *Waste management*, *34*(2), 411-420.

Spinosa, L., Ayol, A., Baudez, J.-C., Canziani, R., Jenicek, P., Leonard, A., Rulkens, W., Xu, G., & Van Dijk, L. (2011). Sustainable and innovative solutions for sewage sludge management. *Water*, *3*(2), 702-717.

Su Politikaları Derneği (SPD). (2023, May 13). *SPD: Atıksu arıtma çamurları ekonomiye kazandırılmalı! (in Turkish)* Retrieved from https://supolitikalaridernegi.org/2023/05/13/spd-atiksu-aritma-camurlari-ekonomiye-kazandirilmali/

Suthar, S. (2010). Pilot-scale vermireactors for sewage sludge stabilization and metal remediation process: comparison with small-scale vermireactors. *Ecological engineering*, *36*(5), 703-712.

Turkish Statistical Institute (TURKSTAT). (2022). Atık İstatistikleri 2022 *(in Turkish)*. Retrieved from https://data.tuik.gov.tr/Bulten/Index?p=Atik-Istatistikleri-2022-49570

Uggetti, E., Llorens, E., Pedescoll, A., Ferrer, I., Castellnou, R., & García, J. (2009). Sludge dewatering and stabilization in drying reed beds: characterization of three full-scale systems in Catalonia, Spain. *Bioresource Technology*, *100*(17), 3882-3890.

Wang, K. J., Wang, P. S., & Nguyen, H. P. (2021). A data-driven optimization model for coagulant dosage decision in industrial wastewater treatment. *Computers & Chemical Engineering*, *152*, 107383.

Winkler, M. K., Bennenbroek, M. H., Horstink, F. H., Van Loosdrecht, M. C. M., & Van de Pol, G. J. (2013). The biodrying concept: An innovative technology creating energy from sewage sludge. *Bioresource Technology*, *147*, 124-129.

Wiśniowska, E., Grobelak, A., Kokot, P., & Kacprzak, M. (2019). Sludge legislation-comparison between different countries. In *Industrial and municipal sludge* (pp. 201-224). Butterworth-Heinemann.

Wu, K., Shi, H., & Guo, X. (2011). Utilization of municipal solid waste incineration fly ash for sulfoaluminate cement clinker production. *Waste Management*, *31*(9-10), 2001-2008.

Xu, C., Chen, W., & Hong, J. (2014). Life-cycle environmental and economic assessment of sewage sludge treatment in China. *Journal of Cleaner Production*, *67*, 79-87.

Yang, G., Zhang, G., & Wang, H. (2015). Current state of sludge production, management, treatment and disposal in China. *Water research*, *78*, 60-73.

Zhu, X., & Chen, Y. (2011). Reduction of N_2O and NO generation in anaerobic–aerobic (low dissolved oxygen) biological wastewater treatment process by using sludge alkaline fermentation liquid. *Environmental science & technology*, *45*(6), 2137-2143.