PIONEER AND INNOVATIVE RESEARCH IN ENGINEERING

All Sciences Academy

PIONEER AND INNOVATIVE RESEARCH IN ENGINEERING

Editor
Asst. Prof. Dr. Umut ÖZKAYA

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

Design: All Sciences Academy Design
Published Date: June 2026
Publisher's Certification Number: 72273
ISBN: 978-625-8993-80-6

www.allsciencesacademy.com
allsciencesacademy@gmail.com

CONTENT

Breast Cancer: Molecular Classification, Tumor Biology, and Histopathological Features

Ercan PULAT[1]

1- Department of Biology, Faculty of Science, Istanbul University, Istanbul 34134, Türkiye.
ercanpulat34@gmail.com ORCID No: 0000-0003-4987-4924

ABSTRACT

Breast cancer comprises biologically diverse malignancies in which morphology, genomic alteration, epigenetic regulation, and the local tissue environment jointly influence clinical course. This chapter links normal mammary organization and multistep tumor development with the molecular events that initiate and sustain malignant progression. Germline susceptibility, acquired mutations, epigenetic changes, non-coding RNA networks, and RNA-level modifications are considered as complementary layers of breast cancer biology. The discussion then reviews intrinsic and emerging phenotypes, including luminal A, luminal B, HER2-positive, triple-negative, claudin-low, interferon-enriched, and molecular apocrine disease, as well as PAM50 classification and multigene prognostic testing. Major pathological categories, from invasive breast carcinoma of no special type and invasive lobular carcinoma to ductal carcinoma in situ and uncommon special patterns, are presented alongside their clinical relevance. Attention is also given to stromal and immune components, particularly cancer-associated fibroblasts, macrophages, extracellular matrix remodeling, and lymphocytic infiltration. Finally, inherited risk, endocrine and reproductive exposures, lifestyle, and environmental determinants are discussed from a prevention perspective. Meaningful risk assessment and treatment selection require the integration of pathological findings with molecular data and patient-specific clinical context.

Keywords - Breast cancer; molecular classification; histopathology; tumor microenvironment; genomics; triple-negative breast cancer; PAM50; risk factors.

1. INTRODUCTION

By 2020, breast cancer had become the most frequently diagnosed malignancy worldwide, exceeding lung cancer in incidence. The GLOBOCAN 2020 estimates reported roughly 2.3 million new diagnoses globally, underscoring its continuing importance as the most common cancer among women (Sung et al., 2021). Earlier detection has improved with advances in imaging and diagnostic practice, although increasing incidence and changing population-level exposures continue to shape disease burden (Arzanova and Mayrovitz, 2022:3). Importantly, breast cancer remains difficult to manage because tumors that share a clinical label can differ markedly in biological behavior and treatment sensitivity.

Rather than representing a uniform disease, breast cancer encompasses entities with distinct tissue architecture, molecular programs, and clinical trajectories. Tumors assigned to the same stage or given a similar histological diagnosis can nevertheless respond very differently to treatment. This variability is the reason contemporary management increasingly uses biologically informed, subtype-directed strategies instead of a uniform treatment approach (Yeo and Guan, 2017:753).

A coherent interpretation of breast cancer subtypes requires more than a description of receptor status. It also requires consideration of normal mammary development, the genetic and epigenetic events that promote transformation, tumor-stroma interactions, molecular taxonomies, and pathological classification. Viewing these components together improves the biological interpretation of tumors and supports risk identification, early detection, and individualized therapeutic planning.

2. DEVELOPMENT OF BREAST CANCER AND TUMOR BIOLOGY

2.1 Normal Breast Anatomy

The breast is a hormonally responsive organ whose structure changes across the reproductive lifespan. Its epithelial compartment, or parenchyma, contains lobules that produce milk and ducts that convey it. These structures are embedded in a supportive stromal framework composed of fat, fibrous tissue, vessels, nerves, and lymphatic channels (Nelson and Bissell, 2006:290).

At the microscopic level, lobules are formed by acini situated on a basement membrane. Acini contain an inner luminal epithelial layer and an outer myoepithelial layer. The luminal compartment is involved in secretory function and commonly expresses estrogen receptor (ER), whereas myoepithelial cells provide contractile support for milk ejection and express basal-associated markers such as p63, smooth muscle actin, calponin, and CK5/14 (Santagata et al., 2014:860). Maintaining this two-layered arrangement is central to normal mammary tissue homeostasis.

The organization of the normal mammary epithelium has informed the concept that tumor subtypes may originate from different cellular states. Neoplasms with luminal differentiation are often ER-positive and, on average, have a more favorable outcome. In contrast, basal-like tumors are frequently triple-negative and generally show more aggressive behavior (Visvader, 2009:2571; Weigelt et al., 2010:199). Histological pattern, molecular profile, and clinical behavior therefore remain closely interrelated rather than independent features of breast cancer (Bertos and Park, 2011:3791).

Mammary development is exposed to hormonal signaling from puberty through menopause, with major additional changes during pregnancy. Earlier menarche, later menopause, nulliparity, and later first childbirth extend cumulative hormonal exposure and are consequently associated with increased breast cancer risk (Sun et al., 2017).

2.2 Breast Carcinogenesis

Malignant transformation in the breast develops through a sequence of cellular and molecular changes rather than a single event. Over time, epithelial cells may acquire sustained proliferative capacity, resistance to growth control, migratory potential, and invasive competence. These changes are consistent with the broader hallmarks-of-cancer framework that applies across tumor types, including breast malignancies (Bertos and Park, 2011; Perou and Borresen-Dale, 2010).

A useful conceptual model divides carcinogenesis into initiation, promotion, and progression. Initial DNA injury or mutation is followed by clonal expansion of altered cells; subsequent genetic and epigenetic events can then generate a biologically more aggressive lesion (Gonzalez-Jimenez et al., 2013:2423). In the breast, this continuum may extend from normal epithelium through hyperplastic and atypical proliferative lesions to carcinoma in situ and ultimately invasive disease (Cichon et al., 2010:390).

Proliferative breast lesions do not all evolve at the same pace, but recognized pathways of progression can move from usual or atypical hyperplasia to ductal carcinoma in situ (DCIS) and invasive carcinoma (Bombonati and Sgroi, 2010:309). Such evolution can span many years and is accompanied by changing mutation profiles and transcriptional programs.

Epigenetic dysregulation accompanies early and later stages of breast tumorigenesis. DNA methylation changes may first include regional or global hypomethylation, whereas promoter hypermethylation can later suppress genes that normally restrain tumor growth (Jovanovic et al., 2010:243; Yang et al., 2001). Accordingly, methylation profiles already distinguish some early breast cancers from normal breast tissue (Yang et al., 2001).

2.3 Genetic and Epigenetic Alterations

Genetic Alterations

Breast cancer genomes contain both inherited variants and somatic alterations acquired during tumor development. PIK3CA mutations are frequent in luminal and HER2-positive disease, whereas TP53 alterations are particularly common in triple-negative breast cancer; GATA3, MAP3K1, and MAP2K4 are also recurrently affected (Lefebvre et al., 2016:7). HER2-positive tumors are driven chiefly by ERBB2 amplification or overexpression. Germline BRCA1 and BRCA2 variants account for a minority of all cases but compromise homologous-recombination repair of DNA double-strand breaks (Roy et al., 2011:68).

Developmental and oncogenic signaling networks, including Wnt, Notch, Hedgehog, TGF-beta, PI3K/AKT, and MAPK/ERK, contribute to breast epithelial biology as well as malignant progression. Chromatin-regulating alterations further emphasize the importance of epigenetic control, while inherited defects in DNA-repair genes such as ATM, CHEK2, and PALB2 modify susceptibility (Garcia-Martinez et al., 2021; Couch et al., 2017; Slavin et al., 2017).

Epigenetic Alterations

Epigenetic mechanisms alter transcriptional output without changing the underlying DNA sequence. In breast cancer, DNA methylation, histone modifications, and chromatin organization can reshape gene expression programs and contribute to both tumor initiation and progression. Global or regional hypomethylation may promote genomic instability and facilitate oncogene activation (Jovanovic et al., 2010:242).

Promoter hypermethylation is a frequent route to functional silencing of tumor-suppressor programs in breast cancer. Such changes can occur early in tumor development and may affect genes involved in cell-cycle control, adhesion, DNA repair, and signal transduction (Jovanovic et al., 2010; Yang et al., 2001).

EZH2, the methyltransferase responsible for H3K27 modification, is often overexpressed in luminal and basal-like breast cancers and can participate in the repression of tumor-suppressive programs. These observations have supported investigation of both histone deacetylase inhibitors and EZH2-directed agents in breast cancer (Holm et al., 2012:501; Yomtoubian et al., 2020).

Non-Coding RNAs and Post-Transcriptional Regulation

Regulatory RNAs add another level of control to breast cancer biology. MicroRNAs, long non-coding RNAs (lncRNAs), and circular RNAs can influence proliferation, invasion, and metastatic competence. Examples include miR-21, miR-155, miR-10b, and miR-145, which have context-dependent oncogenic or tumor-suppressive actions. miR-21 is frequently increased in breast cancer and can repress targets such as PTEN and PDCD4 (Frankel et al., 2007). lncRNAs including HOTAIR, MALAT1, and XIST have also been linked to disease progression and metastatic behavior (Liu et al., 2020:750).

Post-transcriptional regulation is also affected by RNA-binding proteins and alternative-splicing machinery. Increased expression of factors such as SRSF1, SRSF3, and HNRNPA2B1 may redirect splicing toward isoforms that support tumorigenic phenotypes. In particular, SRSF1 overexpression has been associated with breast cancer (Anczukow and Krainer, 2016:1289).

Epitranscriptomic Alterations

Epitranscriptomics refers to reversible chemical marks on RNA molecules. N6-methyladenosine (m6A), the predominant internal modification of mRNA, has attracted attention because altered m6A regulation can influence breast cancer growth and progression. Key regulators under study include METTL3, METTL14, FTO, and ALKBH5 (Wu et al., 2019:326).

2.4 Tumor Microenvironment

Tumor behavior is determined not only by malignant epithelial cells but also by the ecosystem in which they reside. This tumor microenvironment includes cancer-associated fibroblasts (CAFs), tumor-associated macrophages (TAMs), extracellular matrix (ECM), endothelial cells, adipocytes, and multiple immune-cell populations (Kundu et al., 2024:92). Bidirectional communication among these compartments can shape invasion, immune response, and therapeutic resistance.

Cancer-Associated Fibroblasts
CAFs are a major stromal population in breast tumors. Through the release of growth factors, chemokines, IL-6, IL-8, and ECM molecules such as collagen, fibronectin, and hyaluronan, they can modify proliferation, invasion, angiogenesis, and drug response. Their net effects may differ by disease context and CAF state (Zheng and Hao, 2024:1333843).

Tumor-Associated Macrophages
Macrophages within breast tumors can display a spectrum of activation states. M1-like activity is generally inflammatory and anti-tumoral, whereas M2-like polarization is more commonly associated with immune suppression and tumor support. Abundant TAM infiltration has been linked to angiogenesis, invasion, metastasis, and an unfavorable prognosis in breast cancer (Toledo et al., 2024:48).

Extracellular Matrix and Mechanical Pathways
The composition and mechanical properties of the ECM are active determinants of cancer behavior. Altered collagen quantity, fiber organization, and cross-linking can promote tumor progression, while YAP/TAZ, integrin, and FAK signaling connect mechanical cues to cellular phenotypes (Yuan et al., 2024). Dense collagen-rich desmoplasia may additionally limit drug penetration and promote treatment resistance (Winkler et al., 2020).

Immune Microenvironment
Tumor-infiltrating lymphocytes are clinically informative, especially in triple-negative and HER2-positive disease. In early TNBC, greater TIL density has been associated with improved chemotherapy response and survival. It may also indicate a tumor environment more likely to respond to immune-checkpoint blockade (Gruosso et al., 2019:1785; Zagami and Carey, 2022:96).

3. MOLECULAR CLASSIFICATION OF BREAST CANCER

3.1 Luminal A
Luminal A tumors are usually hormone-receptor positive, HER2-negative, and relatively low proliferative. They represent approximately one-half to three-fifths of breast cancers and are generally associated with the most favorable outcome among the major clinical subtypes. Endocrine therapy is central to management, whereas cytotoxic chemotherapy may add limited benefit in appropriately selected low-risk cases (Bernhardt et al., 2016:244).
Gene-expression studies placed luminal A tumors within a differentiated luminal program marked by luminal keratins, GATA3, and FOXA1 and associated with relatively low recurrence risk (Sorlie et al., 2001, 2003). In routine practice, an ER-positive/HER2-negative tumor with low Ki-67 and substantial PR expression can serve as a luminal A-like surrogate. In a 3,169-patient retrospective cohort, the combination of Ki-67 below 14% and PR

expression of at least 20% correlated with favorable disease-free survival (Maisonneuve et al., 2014).

Ki-67 should be interpreted with awareness of technical and interlaboratory variation. Bustreo and colleagues proposed that center-specific median Ki-67 values may improve prognostic use within a St. Gallen-based framework, while later work has emphasized that the often-cited 14% threshold is not universally reproducible (Bustreo et al., 2016; Rewcastle et al., 2024).

PIK3CA alterations are especially common in luminal A disease, with MAP3K1 and GATA3 mutations also occurring regularly. These changes may help sustain ER-positive biology and affect endocrine responsiveness (Guo et al., 2023). Within PAM50/Prosigna assessment, a luminal A tumor with a very low risk-of-recurrence score has a low estimated 10-year distant-recurrence risk, and chemotherapy contributes little in this low-risk setting (Wallden et al., 2015).

3.2 Luminal B

Luminal B cancers retain hormone-receptor expression but are separated from luminal A by greater proliferative activity and, in some cases, HER2 positivity. They comprise roughly 15-20% of breast cancers and usually carry a less favorable prognosis than luminal A disease (Maisonneuve et al., 2014:1).

Relative to luminal A, luminal B tumors show stronger expression of proliferation-related genes and a higher risk of recurrence (Sorlie et al., 2003). Practical Ki-67 cutoffs have shifted across St. Gallen consensus meetings, reflecting uncertainty around a single universal threshold. Microarray-based analyses reported a value near 13.25%, with 14% often used as a workable immunohistochemical approximation (Bustreo et al., 2016; Parker and Perou, 2019).

Adding PR expression to clinicopathological surrogate definitions can move some tumors from the luminal A-like to the luminal B-like category. Although that reassignment can influence adjuvant treatment decisions, it did not uniformly predict worse outcomes in well or moderately differentiated tumors in the Maisonneuve series (Maisonneuve et al., 2014).

TP53 mutations occur more often in luminal B than luminal A tumors (Guo et al., 2023). For hormone receptor-positive/HER2-negative advanced disease, endocrine therapy combined with CDK4/6 inhibition is a key treatment strategy, although treatment selection should be individualized according to disease burden and prior therapy.

PAM50 risk-of-recurrence assessment can help identify luminal B tumors with clinically meaningful distant-recurrence risk and may support adjuvant decision-making. Oncotype DX, MammaPrint, Prosigna, and EndoPredict provide complementary prognostic information for selected ER-positive/HER2-negative tumors and are incorporated into major clinical guidance frameworks (Wallden et al., 2015; Vieira and Schmitt, 2018).

3.3 HER2-Positive Subtype

HER2-positive breast cancer is defined by ERBB2 amplification or HER2 protein overexpression, typically documented by immunohistochemistry (3+) or in situ hybridization. It accounts for about 15-20% of breast cancers. The introduction of trastuzumab and subsequent HER2-directed treatments substantially changed the prognosis of this disease (Guo et al., 2023; Maisonneuve et al., 2014).

HER2-positive disease is molecularly diverse, so receptor status alone does not fully define prognosis or therapeutic sensitivity. HER2-directed treatment includes trastuzumab, pertuzumab, antibody-drug conjugates such as trastuzumab emtansine and trastuzumab deruxtecan, and selected tyrosine-kinase inhibitors. Although outcomes have improved substantially, metastatic relapse can still occur in bone and visceral sites (Exman and Tolaney, 2021; McGuire et al., 2015).

3.4 Triple-Negative Breast Cancer

Triple-negative breast cancer (TNBC) lacks clinically significant ER and PR expression and does not show HER2 amplification or overexpression. In practice, ER/PR staining below 1% and HER2 IHC 0-1+ or negative in situ hybridization are used to define this group. TNBC accounts for roughly 15-20% of breast cancers and is more likely to present with high grade, advanced stage, early metastatic relapse, and BRCA1-associated biology than hormone-receptor positive disease (Lehmann et al., 2011:2750; Zagami and Carey, 2022:95).

TNBC is genomically and transcriptionally diverse. Basal-like expression patterns are common, and many BRCA1-mutated tumors show homologous-recombination defects with marked genomic instability. These features help explain why clinically defined TNBC contains biologically distinct disease states (Lehmann et al., 2011; Zagami and Carey, 2022).

The Lehmann-Pietenpol framework was developed from expression profiling of 587 TNBCs (Lehmann et al., 2011:2750). After subsequent refinement, six stable subtypes were emphasized: basal-like 1 (BL1), basal-like 2 (BL2), immunomodulatory (IM), mesenchymal (M), mesenchymal stem-like (MSL), and luminal androgen receptor (LAR) (Lehmann et al., 2016:17). BL1 is enriched for cell-cycle and DNA-damage-response genes; BL2 for growth-factor signaling and myoepithelial signatures; IM for immune programs; M for motility and epithelial-mesenchymal transition; MSL for stem-like and growth-factor features with lower proliferation; and LAR for luminal transcriptional features driven by androgen-receptor signaling. Independent summaries similarly associate BL1 with high Ki-67 and DNA-repair programs, BL2 with TP53-pathway and myoepithelial features, and IM with strong immune activation (Yin et al., 2020:63).

At a broader level, the TNBC expression classes can be viewed as reflecting several therapeutic hypotheses: proliferative/DNA-repair dominant BL1 disease, growth-factor and metabolic BL2 disease, immune-active IM disease, EMT-rich M/MSL states, and androgen-receptor driven LAR disease. Although the labels differ across reports, the consistent message is that TNBC should not be treated as a single biological entity (Lehmann et al., 2016:17).

Therapeutic sensitivity may vary across TNBC expression states. BL1 models showed particular sensitivity to cisplatin in the Lehmann study. Androgen-receptor targeting with bicalutamide or enzalutamide has been explored in LAR tumors, whereas IM tumors provide a rationale for immune-checkpoint blockade (Lehmann et al., 2016:17; Yin et al., 2020:63).

3.5 Emerging Molecular Subtypes

Beyond the five major intrinsic categories, claudin-low breast cancers exhibit low luminal differentiation and tight-junction gene expression together with epithelial-mesenchymal transition, stemness, and immune/stromal signatures. They often show a triple-negative phenotype; however, claudin-low is increasingly interpreted as a phenotype that can occur across intrinsic subtypes rather than as a separate intrinsic class (Prat et al., 2010; Fougner et al., 2020). Molecular apocrine tumors are typically ER-negative and androgen-receptor positive, sharing biological overlap with the luminal androgen receptor group of TNBC (Zagami and Carey, 2022).

3.6 PAM50 and Modern Genomic Classification

PAM50 measures the expression of 50 genes to assign tumors to luminal A, luminal B, HER2-enriched, basal-like, or normal-like intrinsic categories. It became a cornerstone of integrated genomic classification because it connects a reproducible gene-expression signature to clinically meaningful tumor biology (Perou and Borresen-Dale, 2010). The assay can be performed on formalin-fixed, paraffin-embedded tissue using the nCounter platform and was analytically validated in more than 1,000 ATAC and ABCSG-8 trial samples before U.S. FDA clearance as Prosigna in 2013 (Nielsen et al., 2014; Wallden et al., 2015).

Beyond subtype assignment, Prosigna provides a risk-of-recurrence score that adds prognostic information to conventional factors such as tumor size, nodal status, and proliferation. In selected ER-positive/HER2-negative disease, multigene assays including Prosigna, Oncotype DX, and MammaPrint can support adjuvant systemic-treatment decisions when clinical risk is uncertain (Colomer et al., 2016; Wallden et al., 2015; Vieira and Schmitt, 2018).

Genomic assays have made breast cancer management more biologically individualized, but they provide a measurement from a particular point in time. They may therefore underrepresent intratumoral diversity, clonal evolution, and treatment-related shifts in subtype. Results should be interpreted alongside pathology, clinical risk, and the possibility of dynamic tumor evolution (Guo et al., 2023).

4. HISTOPATHOLOGICAL CLASSIFICATION

4.1 Ductal Carcinomas

Invasive breast carcinoma of no special type (IBC-NST), historically called invasive ductal carcinoma not otherwise specified, is the most common pathological category. The 2019 WHO terminology emphasizes that this diagnosis is applied after tumors with a definable special type, such as tubular, mucinous, or lobular carcinoma, have been excluded. IBC-NST is therefore a heterogeneous group of adenocarcinomas rather than a single uniform morphology (Rakha et al., 2010; Tan et al., 2020; Weigelt et al., 2010).

Histological breast cancer types often carry recognizable cytological and architectural features that correlate with clinical presentation and outcome. Although IBC-NST predominates, special histological types account for a clinically meaningful minority of cases, and WHO classifications recognize numerous distinct subgroups (Tan et al., 2020).

The fifth WHO edition expanded the concept of mixed NST-special type tumors. Lesions containing a 10-90% special-type component can be described in a way that captures both their NST and special-type features (Cserni, 2020; Tan et al., 2020). Morphology remains the basis of pathological classification, but treatment decisions increasingly also rely on luminal, basal-like, and HER2-positive molecular contexts (Cserni, 2020; Tan et al., 2020).

Nottingham histological grade combines tubule or gland formation, nuclear pleomorphism, and mitotic activity. This composite score contributes prognostic information in addition to nodal status and, in some settings, can be more informative than tumor size alone (Rakha et al., 2010).

IBC-NST and invasive lobular carcinoma together make up the large majority of breast cancers. The remainder consists of less common patterns, including medullary, neuroendocrine, tubular, apocrine, metaplastic, mucinous, inflammatory, comedo, adenoid cystic, and micropapillary carcinomas. Prognosis differs across these categories; tubular, mucinous, medullary, and adenoid cystic carcinomas often show more favorable outcomes than many other special types (Bertos and Park, 2011).

Ductal Carcinoma In Situ

DCIS is a non-invasive proliferation of abnormal epithelial cells confined to the ductal-lobular system and is often described as stage 0 breast cancer. It represents approximately 20% of newly diagnosed breast cancers. Architectural patterns include comedo, papillary, solid, cribriform, and micropapillary forms, while nuclear grade is classified as low, intermediate, or high. Higher-grade DCIS is more strongly associated with subsequent invasive carcinoma (Wang et al., 2024).

Conventional DCIS classification based on morphology and cytology has imperfect reproducibility; multicenter evaluations have reported substantial rates of diagnostic reclassification (Pinder and Ellis, 2003). Nuclear grade nevertheless remains clinically relevant, and higher-grade DCIS is more strongly associated with invasive progression risk (Wang et al., 2024).

Whether every DCIS lesion requires immediate treatment remains an important clinical question. The PRECISION report emphasized this uncertainty and noted that calcifications are common among mammographically detected lesions; residual DCIS may also remain after breast-conserving surgery (Van Seijen et al., 2019).

Molecular and clinical variation within DCIS has been described across low-grade, high-grade non-comedo, and high-grade comedo groups. In particular, comedo-pattern DCIS has been linked to greater recurrence risk and a shorter interval to recurrence (Carraro et al., 2013).

Invasive Ductal Carcinoma

IBC-NST constitutes approximately 60-75% of invasive breast cancers. Despite its shared morphological label, it contains ER-positive/HER2-negative, ER-positive/HER2-positive, ER-negative/HER2-positive, and ER-negative/HER2-negative tumors. Gene-expression profiling has shown that IBC-NST spans luminal, basal-like, and HER2-positive molecular classes, whereas most special types are comparatively restricted to one intrinsic subtype, with apocrine carcinoma as an important exception (Weigelt et al., 2008; Rakha et al., 2010).

Within IBC-NST, tumor size and histological grade are significantly associated with nodal involvement in retrospective data (Badowska-Kozakiewicz et al., 2016). Prognostic assessment also integrates ER/PR status, Ki-67, HER2, and TNM stage.

4.2 Lobular Carcinomas

Invasive lobular carcinoma (ILC) is the second most common invasive breast cancer, accounting for approximately 10-15% of cases. It is classically associated with CDH1 inactivation and loss of E-cadherin-mediated adhesion. The resulting growth pattern can be subtle, discohesive, and targetoid, making some ILCs clinically or mammographically difficult to detect (Reed et al., 2015; Djerroudi et al., 2024).

E-cadherin staining in ILC can range from absent or focal to diffuse but abnormal membranous expression. In tumors with retained staining, incomplete membrane labeling or abnormalities in beta-catenin and p120-catenin may still indicate dysfunction of the E-cadherin-catenin complex (Djerroudi et al., 2024; Reed et al., 2015).

ILC may appear prognostically favorable in the short term yet has important late-recurrence behavior and a distinctive metastatic distribution. Peritoneal, gastrointestinal, and gynecological sites are reported more often than in many other breast cancer types (Reed et al., 2021).

4.3 Rare Histological Subtypes

Rare histological patterns, including metaplastic, mucinous, tubular, papillary, adenoid cystic, and medullary-pattern carcinomas, account for a minority of breast cancers but remain clinically important because morphology can inform prognosis, receptor context, and diagnostic interpretation. The WHO fifth edition emphasizes reporting special-type components and recognizes IBC-NST with medullary pattern within a broader TIL-rich spectrum rather than as a separate special type (Bertos and Park, 2011; Cserni, 2020; Tan et al., 2020).

5. RISK FACTORS AND ETIOLOGY

5.1 Genetic Factors

Most breast cancers are sporadic, whereas inherited susceptibility explains a minority of cases. Genetic risk assessment should therefore distinguish high-penetrance syndromic genes from moderate-risk genes and integrate family history with tumor and patient characteristics (Couch et al., 2017; Pashayan et al., 2020).

High-Penetrance Genes

BRCA1 and BRCA2 are autosomal-dominant susceptibility genes responsible for a substantial proportion of hereditary breast cancer. In large panel-testing data, both were classified as definitively associated with breast cancer, with relative-risk estimates of approximately 11.4 and 11.7, respectively (Couch et al., 2017:1190). Their encoded proteins are central to homologous-recombination repair of DNA double-strand breaks (Prakash et al., 2015:1; Roy et al., 2011:68).

The ENVISION consensus quantified markedly different effects across high-penetrance genes. Protein-truncating variants in BRCA1 and BRCA2 were associated with mean relative risks of 11.4 and 11.7, whereas TP53 variants in Li-Fraumeni syndrome carried a much higher estimated risk. CDH1 variants linked to hereditary diffuse gastric cancer and PTEN variants in Cowden syndrome also conferred substantial breast cancer susceptibility (Pashayan et al., 2020).

Moderate-Penetrance Genes

For moderate-penetrance genes, ENVISION reported relative-risk estimates of 2.8 for ATM, 2.1 for BARD1, 3.0 for CHEK2, 2.7 for NBN, 2.6 for NF1, 5.3 for PALB2, and 2.1 for RAD51D, with confidence intervals varying by gene (Pashayan et al., 2020).

PALB2 has particularly important clinical relevance because some estimates approach the risk associated with BRCA2, with an approximately five- to nine-fold increase in breast cancer risk. ATM variants more often show moderate effects, although selected missense variants can confer substantially greater risk (Turnbull et al., 2018).

After BRCA1, BRCA2, and syndromic genes were excluded, panel-testing analyses identified ATM, BARD1, CHEK2, PALB2, and RAD51D as genes with clinically meaningful moderate or high associations with breast cancer (Couch et al., 2017).

Polygenic Risk Scores

Polygenic risk scores may further modify inherited risk. A score based on 86 single-nucleotide variants influenced risk estimates in carriers of pathogenic BRCA1, BRCA2, CHEK2, ATM, and PALB2 variants, demonstrating that monogenic and polygenic factors can be considered together (Gallagher et al., 2020).

5.2 Hormonal Factors

Estrogen signaling is central to the biology of many breast cancers, as reflected by the predominance of ER-positive disease and the major clinical role of endocrine therapy (Guo et al., 2023).

Reproductive Factors

Reproductive histories that extend lifetime exposure to ovarian hormones are associated with higher risk. Examples include menarche before age 12 years, menopause after age 55 years, nulliparity, and first childbirth after age 30 years. Early menarche has been linked to a small but measurable risk increase, and later menopause is likewise associated with greater risk (Sun et al., 2017; Wilkinson and Gathani, 2021).

Hormone Replacement Therapy

Prolonged use of combined estrogen-progestin hormone-replacement therapy is associated with increased postmenopausal breast cancer risk, and duration of use is an important determinant (Collaborative Group on Hormonal Factors in Breast Cancer, 1997).

A pooled reanalysis involving more than 52,000 cases and 108,000 controls found that breast cancer risk increased with hormone-replacement therapy duration. Among current or recently exposed users, the estimated relative risk rose by about 2.3% per year of use and reached 1.35 after at least five years (Collaborative Group on Hormonal Factors in Breast Cancer, 1997).

Long-term follow-up from the Women's Health Initiative distinguished estrogen-only and combined regimens. Among women with prior hysterectomy, conjugated equine estrogen alone was associated with lower breast cancer incidence and mortality than placebo; in women with an intact uterus, combined conjugated equine estrogen plus medroxyprogesterone acetate was associated with higher incidence (Chlebowski et al., 2020).

Oral Contraceptives

The relation between oral contraceptives and breast cancer is complex. Available evidence suggests a modest increase in risk during or near use, with risk returning toward baseline within approximately 10 years after discontinuation (Sun et al., 2017).

5.3 Lifestyle Factors

A considerable proportion of potentially modifiable breast cancer risk is linked to lifestyle. Postmenopausal obesity, alcohol consumption, physical inactivity, and smoking are consistently highlighted as the principal modifiable exposures (Coles et al., 2024).

Obesity

Postmenopausal obesity is particularly associated with ER-positive and PR-positive disease. Estimates suggest that obesity contributes to 8-13% of breast cancer cases in high-income settings and may contribute to a larger proportion among postmenopausal African American women (Coles et al., 2024).

Higher body mass index is consistently associated with increased postmenopausal breast cancer risk; the association is more complex before menopause (Coles et al., 2024).

Alcohol

Alcohol is one of the most consistently documented modifiable breast cancer risk factors. Even lower levels of consumption may contribute at a population level; a European analysis attributed a meaningful fraction of alcohol-related cancers to light-to-moderate drinking, with breast cancer comprising about half of these cases (Coles et al., 2024).

A meta-analysis of 53 epidemiological studies estimated a 32% higher risk at 35-44 g alcohol per day and an approximately 7.1% increase in relative risk for each additional 10 g consumed daily (Sun et al., 2017).

Smoking

The contribution of smoking remains debated, but contemporary systematic-review evidence supports an association of both active and passive smoking with increased breast cancer incidence (He et al., 2022).

Physical Activity

Regular physical activity is associated with lower breast cancer risk and remains an important component of prevention-oriented counseling (Coles et al., 2024).

Diet

Dietary patterns may contribute to risk and outcome. A Western-style diet, high saturated-fat intake, low fiber intake, alcohol use, smoking, and obesity are associated with less favorable breast cancer risk profiles, whereas physical activity and breastfeeding appear protective (Coles et al., 2024).

6. CONCLUSION AND FUTURE PERSPECTIVES

Breast cancer is best understood as a spectrum of biologically distinct diseases rather than a single entity. Histology, molecular phenotype, treatment response, and prognosis vary because tumors arise through different combinations of inherited susceptibility, acquired mutation, epigenetic regulation, hormonal exposure, microenvironmental signaling, and modifiable exposures. A clinically useful model must therefore connect tumor-intrinsic features with the patient and tissue context in which the tumor develops.

Intrinsic-subtype classification has reshaped breast cancer care by linking luminal, HER2-enriched, basal-like, and triple-negative biology to prognosis and treatment selection. Multigene assays such as Prosigna, Oncotype DX, and MammaPrint improve risk stratification and can help tailor adjuvant therapy. Yet these assays do not capture the full temporal complexity of cancer: subclonal populations, therapy-driven selection, and phenotype switching can alter a tumor after the original sample was obtained (Guo et al., 2023).

Pathology remains indispensable because it defines tumor type, grade, architecture, and diagnostic patterns that are not fully reproduced by genomic data alone. IBC-NST, ILC, metaplastic, mucinous, tubular, papillary, and other special-type cancers differ in clinically meaningful ways. Continued refinement of WHO categories illustrates the value of integrating morphology with immunohistochemistry, molecular testing, and clinical information (Tan et al., 2020).

The tumor microenvironment is an additional layer of biological classification. CAFs, macrophages, matrix architecture, vessels, adipocytes, and immune infiltrates affect growth, dissemination, immune escape, and resistance. In particular, TILs in TNBC and HER2-positive cancer show how immune context can provide prognostic and predictive information. Future systems should therefore incorporate immune, stromal, and metabolic features alongside standard pathology and molecular signatures.

Prevention and risk assessment are equally important. High-penetrance genes such as BRCA1, BRCA2, TP53, CDH1, and PTEN, together with moderate-risk genes including ATM, CHEK2, PALB2, BARD1, and RAD51D, guide genetic counseling, surveillance, and risk-reduction strategies. Combining these findings with polygenic scores, family history, endocrine/reproductive factors, lifestyle, and environmental context may further improve individualized prevention. Addressing obesity, alcohol, physical inactivity, smoking, diet, and potentially circadian disruption should remain part of comprehensive breast cancer control.

Future progress will depend on better integration of pathology, genomics, epigenetics, transcriptomics, and microenvironmental information into clinically validated models. Repeated molecular assessment may clarify evolutionary trajectories and mechanisms of resistance. Digital pathology, artificial-intelligence-supported image analysis, and multi-omics approaches may improve diagnostic consistency, identify clinically relevant patterns, and refine treatment selection. The central challenge is to translate this increasing biological complexity into accessible, equitable, and patient-centered care (Tan et al., 2020).

REFERENCES

Anczukow, O., and Krainer, A. R. (2016). Splicing-factor alterations in cancers. RNA, 22, 1285-1301. https://doi.org/10.1261/rna.057919.116

Arzanova, E., and Mayrovitz, H. N. (2022). The epidemiology of breast cancer. In H. N. Mayrovitz, Breast cancer. Exon Publications. https://doi.org/10.36255/exon-publications-breast-cancer-epidemiology

Badowska-Kozakiewicz, A. M., Liszczak, T., et al. (2016). The role of histological grade in invasive ductal carcinoma of no special type. Oncology Letters, 11, 2019-2026. https://doi.org/10.3892/ol.2016.4153

Bernhardt, S. M., Dasari, A., Knipprath-Meszaros, A. A., et al. (2016). Current and emerging treatment strategies for luminal breast cancer. Hormone and Metabolic Research, 48, 241-252. https://doi.org/10.1055/s-0042-100909

Bertos, N. R., and Park, M. (2011). Breast cancer - One term, many entities? Journal of Clinical Investigation, 121, 3789-3796. https://doi.org/10.1172/JCI57118

Bombonati, A., and Sgroi, D. C. (2010). The molecular pathology of breast cancer progression. Journal of Mammary Gland Biology and Neoplasia, 15, 307-319. https://doi.org/10.1007/s10911-010-9188-3

Bustreo, S., Osella-Abate, S., Cassoni, P., et al. (2016). Optimal Ki67 cut-off for luminal breast cancer prognostic evaluation: A large case series study with a long-term follow-up. Breast Cancer Research and Treatment, 157, 363-371. https://doi.org/10.1007/s10549-016-3817-9

Carraro, D. M., Elias, E. A., and Piana de Andrade, V. (2013). Ductal carcinoma in situ of the breast: Morphological and molecular features implicated in progression. BMC Cancer, 13, 256. https://doi.org/10.1186/1471-2407-13-256

Chlebowski, R. T., Anderson, G. L., Aragaki, A. K., et al. (2020). Association of menopausal hormone therapy with breast cancer incidence and mortality during long-term follow-up. JAMA, 324, 369-380. https://doi.org/10.1001/jama.2020.9482

Cichon, M. A., et al. (2010). Microenvironment in breast cancer progression. Cell Cycle, 9, 3906-3912. https://doi.org/10.4161/cc.9.19.13077

Coles, C. E., Earl, H., et al. (2024). Lifestyle and breast cancer: A review. The Breast, 73, 103586. https://doi.org/10.1016/j.breast.2024.103586

Collaborative Group on Hormonal Factors in Breast Cancer. (1997). Breast cancer and hormone replacement therapy: Collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. The Lancet, 350, 1047-1059. https://doi.org/10.1016/S0140-6736(97)08233-0

Colomer, R., Gonzalez-Farre, B., Ballesteros, A. I., et al. (2016). Prosigna gene signature assay: A comprehensive review. The Breast, 30, 82-88. https://doi.org/10.1016/j.breast.2016.08.017

Couch, F. J., Shimelis, H., Hu, C., et al. (2017). Associations between cancer predisposition testing panel genes and breast cancer. JAMA Oncology, 3, 1190-1196. https://doi.org/10.1001/jamaoncol.2017.0420

Cserni, G. (2020). Histological classification of breast cancer: WHO classification 5th edition and beyond. Pathology and Oncology Research, 26, 1439-1445. https://doi.org/10.1007/s12253-020-00830-6

Djerroudi, L., et al. (2024). E-cadherin expression in invasive lobular carcinoma: A retrospective study. Modern Pathology, 37, 100444. https://doi.org/10.1016/j.modpat.2024.100444

Exman, P., and Tolaney, S. M. (2021). Anti-HER2 therapy for metastatic breast cancer. Hematology/Oncology Clinics of North America, 35, 87-107. https://doi.org/10.1016/j.hoc.2020.08.003

Fougner, C., Bergholtz, H., Norum, J. H., and Sorlie, T. (2020). Re-definition of claudin-low as a breast cancer phenotype. Nature Communications, 11, 1787. https://doi.org/10.1038/s41467-020-15574-5

Frankel, L. B., Christoffersen, N. R., Skanderup, A. J., et al. (2007). Programmed cell death 4 is an important functional target of the microRNA miR-21 in breast cancer cells. Journal of Biological Chemistry, 282, 1021-1028. https://doi.org/10.1074/jbc.M611393200

Gallagher, S., Hughes, E., Wagner, M. B., et al. (2020). Association of a polygenic risk score with breast cancer among women carriers of high- and moderate-risk breast cancer genes. JAMA Network Open, 3, e208501. https://doi.org/10.1001/jamanetworkopen.2020.8501

Garcia-Martinez, L., Zhang, Y., Nakata, Y., Chan, H. L., and Morey, L. (2021). Epigenetic mechanisms in breast cancer therapy and resistance. The Journal of Pathology, 254, 1-18. https://doi.org/10.1002/path.5636

Gonzalez-Jimenez, E., Perez-Perez, A., et al. (2013). Breast cancer initiation and progression: A review. Pathology - Research and Practice, 209, 2421-2432. https://doi.org/10.1016/j.prp.2013.06.003

Gruosso, T., Garcia-Mariscal, A., et al. (2019). Immune microenvironment in triple-negative breast cancer. Journal of Clinical Investigation, 129, 1783-1791. https://doi.org/10.1172/JCI124952

Guo, L., Kong, D., Liu, J., et al. (2023). Breast cancer heterogeneity and its implications for treatment. Frontiers in Oncology, 13, 1246719. https://doi.org/10.3389/fonc.2023.1246719

He, Y., Si, Y., Li, X., et al. (2022). Active and passive smoking and breast cancer risk: A systematic review and meta-analysis. Frontiers in Oncology, 12, 728491. https://doi.org/10.3389/fonc.2022.728491

Holm, K., Grabau, D., et al. (2012). Global H3K27me3 changes in breast cancer. Clinical Cancer Research, 18, 501-513. https://doi.org/10.1158/1078-0432.CCR-11-1715

Jovanovic, J., Ronneberg, J. A., Tost, J., and Kristensen, V. (2010). The epigenetics of breast cancer. Molecular Oncology, 4, 242-254. https://doi.org/10.1016/j.molonc.2010.04.002

Kundu, S., et al. (2024). Tumor microenvironment in breast cancer. Oncotarget, 15, 91-97. https://doi.org/10.18632/oncotarget.28500

Lefebvre, C., Bachelot, T., et al. (2016). Mutational profile of metastatic breast cancer. Genome Medicine, 8, 7. https://doi.org/10.1186/s13073-015-0267-3

Lehmann, B. D., Bauer, J. A., Chen, X., et al. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models. Journal of Clinical Investigation, 121, 2750-2767. https://doi.org/10.1172/JCI45014

Lehmann, B. D., Jovanovic, B., Chen, X., et al. (2016). Refinement of triple-negative breast cancer molecular subtypes. PLOS ONE, 11, e0157368. https://doi.org/10.1371/journal.pone.0157368

Liu, Y., et al. (2020). Long noncoding RNAs in breast cancer progression. Journal of Cellular Physiology, 235, 748-760. https://doi.org/10.1002/jcp.29025

Maisonneuve, P., Disalvatore, D., Rotmensz, N., et al. (2014). Proposed new clinicopathological surrogate definitions of luminal A and luminal B. Breast Cancer Research, 16, R65. https://doi.org/10.1186/s13058-014-0465-8

McGuire, K. P., et al. (2015). Patterns of metastasis in HER2-positive breast cancer. Annals of Surgical Oncology, 22, 1478-1484. https://doi.org/10.1245/s10434-014-4135-z

Nelson, C. M., and Bissell, M. J. (2006). Of extracellular matrix, scaffolds, and signaling: Tissue architecture. Annual Review of Cell and Developmental Biology, 22, 287-309. https://doi.org/10.1146/annurev.cellbio.22.010305.104256

Nielsen, T. O., Wallden, B., Schaper, C., et al. (2014). Analytical validation of the PAM50-based Prosigna assay. BMC Cancer, 14, 177. https://doi.org/10.1186/1471-2407-14-177

Parker, J. S., and Perou, C. M. (2019). Tumor heterogeneity: Focus on the intrinsic molecular classification of breast cancer. British Journal of Cancer, 121, 1-9. https://doi.org/10.1038/s41416-019-0521-6

Pashayan, N., Antoniou, A. C., et al. (2020). Personalized early detection and prevention of breast cancer: ENVISION. Nature Reviews Clinical Oncology, 17, 687-705. https://doi.org/10.1038/s41571-020-0388-9

Perou, C. M., and Borresen-Dale, A.-L. (2010). Systems biology and genomics of breast cancer. Cold Spring Harbor Perspectives in Biology, 3, a003293. https://doi.org/10.1101/cshperspect.a003293

Pinder, S. E., and Ellis, I. O. (2003). The classification of ductal carcinoma in situ. Current Diagnostic Pathology, 9, 184-192. https://doi.org/10.1054/cdip.2003.0153

Prakash, R., Zhang, Y., Feng, W., and Jasin, M. (2015). Homologous recombination and human health. Cold Spring Harbor Perspectives in Biology, 7, a016600. https://doi.org/10.1101/cshperspect.a016600

Prat, A., Parker, J. S., Karginova, O., et al. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype. Breast Cancer Research, 12, R68. https://doi.org/10.1186/bcr2635

Rakha, E. A., Reis-Filho, J. S., Baehner, F., et al. (2010). Breast cancer prognostic classification: The role of histological grade. Breast Cancer Research, 12, 207. https://doi.org/10.1186/bcr2647

Reed, A. E. M., Kutasovic, J. R., and Lakhani, S. R. (2015). Invasive lobular carcinoma of the breast. Frontiers in Oncology, 5, 153. https://doi.org/10.3389/fonc.2015.00153

Reed, A. E. M., et al. (2021). Invasive lobular carcinoma: A review. Histopathology, 78, 28-48. https://doi.org/10.1111/his.14230

Rewcastle, E., Skaland, I., Gudlaugsson, E., et al. (2024). The Ki67 dilemma. Breast Cancer Research and Treatment, 207, 167-178. https://doi.org/10.1007/s10549-024-05381-7

Roy, R., Chun, J., and Powell, S. N. (2011). BRCA1 and BRCA2: Different roles in a common pathway of genome protection. Nature Reviews Cancer, 12, 68-78. https://doi.org/10.1038/nrc3181

Santagata, S., Thakkar, A., Ergonul, A., et al. (2014). Taxonomy of breast cancer based on normal cell phenotype. Journal of Clinical Investigation, 124, 859-866. https://doi.org/10.1172/JCI70941

Slavin, T. P., et al. (2017). Genetic testing for breast cancer susceptibility. American Society of Clinical Oncology Educational Book, 37, 44-56. https://doi.org/10.1200/EDBK_175222
Sorlie, T., Perou, C. M., Tibshirani, R., et al. (2001). Gene expression patterns of breast carcinomas. Proceedings of the National Academy of Sciences, 98, 10869-10874. https://doi.org/10.1073/pnas.191368598
Sorlie, T., Tibshirani, R., Parker, J., et al. (2003). Repeated observation of breast tumor subtypes. Proceedings of the National Academy of Sciences, 100, 8418-8423. https://doi.org/10.1073/pnas.0932692100
Sun, Y.-S., Zhao, Z., Yang, Z.-N., et al. (2017). Risk factors and preventions of breast cancer. International Journal of Biological Sciences, 13, 1387-1397. https://doi.org/10.7150/ijbs.21635
Sung, H., Ferlay, J., Siegel, R. L., et al. (2021). Global cancer statistics 2020. CA: A Cancer Journal for Clinicians, 71, 209-249. https://doi.org/10.3322/caac.21660
Tan, P. H., Ellis, I., et al. (2020). The 2019 WHO classification of tumours of the breast. Histopathology, 77, 181-185. https://doi.org/10.1111/his.14095
Toledo, D. A., et al. (2024). Tumor-associated macrophages in breast cancer. Frontiers in Immunology, 15, 478048. https://doi.org/10.3389/fimmu.2024.1354048
Turnbull, C., et al. (2018). Penetrance estimates for breast cancer susceptibility genes. Nature Genetics, 50, 1187-1193. https://doi.org/10.1038/s41588-018-0132-8
Van Seijen, M., Lips, E. H., et al. (2019). Ductal carcinoma in situ: To treat or not to treat. British Journal of Cancer, 121, 285-292. https://doi.org/10.1038/s41416-019-0474-7
Vieira, A. F., and Schmitt, F. C. (2018). An update on breast cancer multigene prognostic tests. Frontiers in Medicine, 5, 248. https://doi.org/10.3389/fmed.2018.00261
Visvader, J. E. (2009). Keeping abreast of the mammary epithelial hierarchy. Genes and Development, 23, 2571-2577. https://doi.org/10.1101/gad.1849509
Wallden, B., Storhoff, J., Nielsen, T., et al. (2015). Development and verification of the PAM50-based Prosigna assay. BMC Medical Genomics, 8, 54. https://doi.org/10.1186/s12920-015-0129-6
Wang, J., Li, B., Luo, M., et al. (2024). Progression from ductal carcinoma in situ to invasive breast cancer. Signal Transduction and Targeted Therapy, 9, 326. https://doi.org/10.1038/s41392-024-02037-6
Weigelt, B., Geyer, F. C., and Reis-Filho, J. S. (2010). Histological types of breast cancer: How special are they? Molecular Oncology, 4, 192-208. https://doi.org/10.1016/j.molonc.2010.04.004
Weigelt, B., Horlings, H. M., et al. (2008). Refinement of breast cancer classification by molecular characterization. Journal of Pathology, 216, 141-150. https://doi.org/10.1002/path.2400
Wilkinson, L., and Gathani, T. (2021). Understanding breast cancer as a global health concern. British Journal of Cancer, 125, 503-510. https://doi.org/10.1038/s41416-021-01464-y
Winkler, J., Abisoye-Ogunniyan, A., Metcalf, K. J., and Werb, Z. (2020). Concepts of extracellular matrix remodelling. Nature Communications, 11, 5120. https://doi.org/10.1038/s41467-020-19393-8
Wu, L., Wu, D., Ning, J., et al. (2019). Changes of N6-methyladenosine modulators promote breast cancer progression. BMC Cancer, 19, 326. https://doi.org/10.1186/s12885-019-5544-9

Yang, X., Yan, L., and Davidson, N. E. DNA methylation in breast cancer. (2001). Endocrine-Related Cancer, 8, 115-127. https://doi.org/10.1677/erc.0.0080115
Yeo, S. K., and Guan, J. L. (2017). Breast cancer: Multiple subtypes within a single tumor. Trends in Cancer, 3, 753-760. https://doi.org/10.1016/j.trecan.2017.09.001
Yin, L., Duan, J.-J., Bian, X.-W., and Yu, S.-C. (2020). Triple-negative breast cancer molecular subtyping. Breast Cancer Research, 22, 63. https://doi.org/10.1186/s13058-020-01296-5
Yomtoubian, S., et al. (2020). Inhibition of EZH2 in breast cancer. Cancer Research, 80, 3145-3156. https://doi.org/10.1158/0008-5472.CAN-19-3527
Yuan, Y., et al. (2024). Extracellular matrix and biomechanics in breast cancer. Matrix Biology, 121, 45-65. https://doi.org/10.1016/j.matbio.2024.05.001
Zagami, P., and Carey, L. A. (2022). Triple negative breast cancer: Pitfalls and progress. npj Breast Cancer, 8, 95. https://doi.org/10.1038/s41523-022-00471-7
Zheng, Y., and Hao, S. (2024). Cancer-associated fibroblasts in breast cancer. Cellular Oncology, 47, 1333843. https://doi.org/10.1007/s13402-024-00952-3

Machine Learning Frameworks for Rheological Optimization of Multimaterial Bioinks in 3D Bioprinting Processes

Fatih Mehmet YILDIZ[1,2]

İlyas KARTAL[1]

1- Department of Metallurgical and Materials Engineering, Faculty of Technology, Marmara University, Turkey
2- Department of Science and Technology, Gokabad Research, Istanbul, Turkey

ABSTRACT

Three-dimensional (3D) bioprinting technology is revolutionizing regenerative medicine by offering the potential to mimic complex tissue and organ structures with precise arrangements of cellular and biological materials. However, the complex non-Newtonian rheological behaviors of bioinks used in extrusion-based biofabrication processes, such as shear-thinning and thixotropic recovery, play a critical role in printability and structural integrity. The transition from simple single-material structures to multi-material and hybrid interfaces, which are essential for clinical applications, presents numerous challenges. Some of these are thermomechanical problems such as co-sintering difficulties and differential shrinkage. Traditional trial-and-error laboratory methods are wasteful and inefficient. In recent years, artificial intelligence (AI) and machine learning (ML) algorithms have been incorporated into biofabrication processes for this reason. This book chapter examines innovative algorithmic frameworks (e.g., HydroThermoMLP and click chemistry optimizations) that successfully predict high-dimensional relationships between polymer formulations and printability outputs. In addition, it addresses the sustainability contributions of closed-loop real-time process control systems that minimize waste by instantly responding to environmental changes during production and automatically calibrating parameters. Finally, it discusses the future potential of Digital Twin ecosystems that combine virtual simulations with physical bio-production cycles and predict tissue development in a computer environment.

Keywords – Machine Learning, 3D-Bioprinting, Bioinks, Optimization, Digital-Twin

INTRODUCTION TO NON-NEWTONIAN DYNAMICS IN EXTRUSION-BASED 3D BIOPRINTING

Three-dimensional (3D) bioprinting technology involves the layer-by-layer printing of living cells, biochemical factors, and supporting biomaterials based on predetermined computer-aided design (CAD). This method has become one of the most promising cornerstones of tissue engineering and regenerative medicine. This innovative manufacturing paradigm offers personalized treatment solutions for critical medical crises such as organ failure and tissue damage. Essentially, this technology utilizes "bioinks" that act as carrier matrices. These bioinks possess properties that can mimic complex anatomical structures and biologically support cellular proliferation, differentiation, and survival functions. Although various bioprinting modalities (inkjet, laser-assisted, stereolithography) exist in the literature, extrusion-based bioprinting (EBB) is by far the most widely preferred in industrial and academic applications due to its capacity to accommodate high cell densities, equipment flexibility, cost-effectiveness, and ability to work with materials across an extremely wide viscosity range [1,2]. However, for extrusion-based 3D printing systems to deliver successful and reproducible results at the clinical level, rheological properties must be managed precisely at the micrometric scale [3].

Hydrogels and biomaterials used for extrusion exhibit non-Newtonian fluid dynamics, unlike simple fluids with constant viscosity such as water or standard oils. This means that the intrinsic friction resistance, or viscosity, of these materials changes in a direct and often nonlinear relationship with the shear stress or shear rate applied to them. The most desirable parameter of this dynamic in bio-production is "shear-thinning" (or pseudoplasticity) behavior. When bioink is pushed through a micro-scale nozzle under high mechanical pressure, the high shear rates to which the system is subjected (typically ranging from 200 to 400 s^{-1} for a standard nozzle with a diameter of 250 μm) temporarily disrupt the structure of the polymeric network. This causes the complex polymer chains to align in the direction of flow. Therefore, it weakens intermolecular interactions, causing a sudden and sharp drop in the viscosity of the material. This property is a crucial advantage for bio-production because it allows the material to flow smoothly and continuously as a filament within the narrow nozzle channel without clogging. It also minimizes cell death rates (apoptosis and necrosis) during printing by reducing harmful mechanical shear stress on the delicate living cells trapped within the matrix. To prevent cell membrane rupture, high fluidity of the bio-ink within the nozzle is necessary [4].

However, the extrusion-based bioprinting process doesn't end with the material exiting the nozzle. The real physical challenge begins immediately after deposition. The moment the bioink is placed on the printing plate or on top of the previous layer, the high shear stress inside the nozzle suddenly approaches zero limitingly. This is a critical point. For the material to maintain its designed 3D shape (shape fidelity and structural integrity) against gravitational forces, surface tension, and its own weight, it must instantly regain its structural integrity, i.e., return to a solid-like state. This rapid recovery process depends on the material's "yield stress" and "thixotropic" recovery kinetics [5]. Yield stress refers to the minimum mechanical stress threshold that must be overcome for the material to transition from a solid state to a liquid-like fluid state and begin to flow. For the bioink to exit the nozzle, the system pressure must absolutely exceed this threshold. However, even after deposition is complete, the material's intrinsic yield stress must rapidly return to a sufficiently high level. This allows for the construction of complex, multi-layered structures reaching several centimeters in height. As a result, the underlying layers do not collapse under the weight of continuously added layers (lattice collapse) or spread horizontally, clogging the microporous architecture (porosity) vital for tissue growth. Loss of porosity prevents the diffusion of oxygen and nutrients needed by the cells and leads to the formation of necrotic nuclei in the interior of the printed tissue [6].

Another factor that complicates extrusion dynamics is the viscoelastic nature of the material. Soft polymeric hydrogels exhibit a simultaneous combination of both viscous (energy-absorbing like a liquid) and elastic (energy-storing like a solid) properties. These properties are typically measured in rheological analyses using the storage modulus (G^', representing the elastic behavior of the material) and loss modulus (G^", representing the viscous behavior of the material) parameters [7]. These form the fundamental equations that define the printability window of the material. This viscoelastic structure can lead to unpredictable effects on the extrusion line during printing. For example, hydrogels with very high elasticity can suddenly release their intrinsic elastic energy when passing through a narrow nozzle channel and into a pressureless environment. This phenomenon is described by a dimensional deformation called "die swell" (or Barus effect) [8]. This situation leads to the printed filament diameter being much wider and more irregular than the nozzle's inner diameter, completely destroying the targeted micrometric printing resolution. On the other hand, materials that are not sufficiently elastic liquefy and disperse uncontrollably [9].

In traditional polymer chemistry and tissue engineering laboratories, various studies are conducted to achieve this critical rheological equilibrium. These studies evaluate key aspects such as polymer chain concentrations (e.g., alginate, gelatin, methacrylate hyaluronic acid - HAMA ratios), molecular weights, crosslinker densities, and additives that regulate system fluidity (e.g., laponite, nano-cellulose, or synthetic rheology modifiers). Unfortunately, this traditional optimization approach relies on researchers' physical intuition and countless trial-and-error cycles. It is an extremely inefficient and laborious experimental methodology that can take months or even longer, leading to the waste of large amounts of valuable cell culture and expensive biosynthetic materials. Moreover, when transitioning from single-material systems to more realistic multi-material tissue structures, traditional experimental approaches yield negative results due to the nonlinear nature of interpolymer physicochemical interactions. In the literature, there is a huge research and optimization gap in the development of complex bioinks that are "multi-material" and "stimulus-responsive" to environmental changes such as temperature, pH, and light. This is precisely where artificial intelligence (AI) and machine learning (ML) algorithms, capable of detecting hidden patterns in massive datasets, simultaneously optimizing multivariate systems, performing simulations if necessary, and modeling nonlinear physical relationships, can offer indispensable solutions for polymer science and bioprinting processes [3,10].

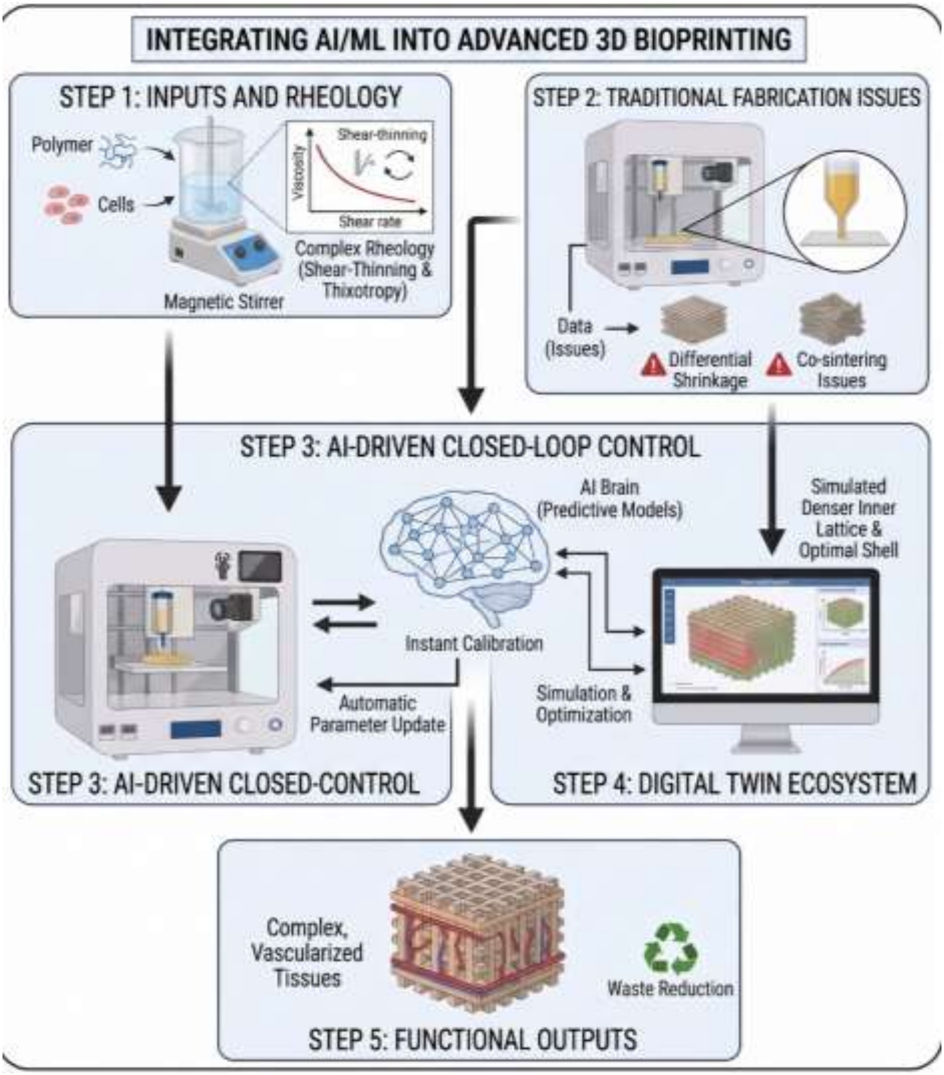


Fig. 1. Integrating AI/ML Into Advanced 3D-Bioprinting

DATA-DRIVEN PARADIGMS IN SOFT MATERIALS SCIENCE

Traditional approaches to soft materials science, particularly in hybrid systems using multiple biopolymers, suffer from a lack of mathematical formulas for expressing complex, nonlinear physicochemical interactions and phase behaviors. Researchers focusing on interdisciplinary studies have made a radical shift towards data-driven paradigms centered on machine learning (ML) and artificial intelligence (AI) [11]. ML algorithms do not rely on predefined rigid physics equations. Instead, they are trained on multidimensional datasets obtained from experimental rheology tests, structural analyses, and cellular experiments, identifying hidden mathematical patterns that the human mind or classical equations cannot detect. This feature allows for the highly accurate prediction of microstructural formations, viscosity profiles, mechanical strengths, and thermodynamic properties of bioinks [12]. This not only predicts the properties of an existing material but also, using an "inverse design" philosophy, performs simulations with the available datasets. This allows the system to algorithmically determine which chemicals should be mixed in what proportions to achieve the predetermined target mechanical properties [13].

In soft materials engineering, various machine learning methods can provide mathematical solutions to solve rheological problems or predict specific parameters. One of the most important studies in this area is the design of synthetic hydrogels that mimic extracellular matrix (ECM) structures. For example, an innovative artificial intelligence tool created by Cadamuro et al. (2025) for the design of biopolymers has played a role in the construction of multi-component polymer networks based on "click chemistry" reactions (specifically strain-assisted azide-alkyne cycloaddition and Diels-Alder reactions) [8]. In these complex formulations consisting of basic ECM components such as gelatin, hyaluronic acid, and polyethylene glycol (PEG) crosslinkers, researchers generally resort to traditional methods. Polymers are mixed and the results are tested. Instead, the algorithm can be directly given the targeted rheological outputs (desired storage modulus G^' and loss modulus G^" values) as input. The developed Artificial Neural Network (ANN) model can calculate the optimum gelatin-hyaluronic acid ratios, degrees of functionalization (derivative) and environmental cross-linking conditions required to achieve these mechanistic goals using inverse prediction [14]. This data-driven process eliminates the need for hundreds of expensive and laborious iterative trials that might take months in a standard polymer lab. These calculations are compressed into minutes using computer simulations, paving the way for the production of customized bio-inks for highly specific applications such as tumor microenvironment modeling.

Table 1. Machine Learning Methodology and Application Areas

Machine Learning Method	Application Area / Type of Bioink	Optimization Goal / Model Output
Multilayer Sensor (MLP) [15]	HAMA/GelMA Hybrid Hydrogels (HydroThermoMLP model)	Viscosity estimation, shear stress profiling, thermodynamic consistency
Artificial Neural Networks (ANN) [16]	Alginate, Gelatin and Laponite Based Nano-composites	Zero shear viscosity (ZSV), yield stress (YS) and thixotropic recovery time estimation
Physics-Inspired Reverse Design [17]	Gelatin and Hyaluronic Acid (Click Chemistry Interfaces)	Ideal polymer/crosslinker mix ratios to achieve pre-targeted G' and G'' modules
Decision Trees and Clustering [18]	Shear-Thinning Natural Hydrogels	Rheological viscosity estimation, balancing cellular viability with skeletal structural accuracy
Extreme Gradient Enhancement (XGB) [19]	Konjac Glucomannan (KGM) and Polysaccharide Colloids	Rapid formulation screening and flow resistance prediction for printable biopolymer gels

One of the main criticisms faced by machine learning-based material design is the risk that data-driven predictions may not always conform to universal physical and thermodynamic laws in the real world. A model based entirely on statistical optimization might suggest a polymer blend that looks perfect on paper but is physicochemically impossible. To overcome this problem, "physics-informed" machine learning architectures are being developed [13,17]. One of the most current and powerful examples of this approach is the "HydroThermoMLP" AI framework, specifically developed to model the viscosity and rheological shear stress of hybrid blends of hyaluronic acid methacrylate (HAMA) and gelatin methacrylate (GelMA) hydrogels, widely used in 3D bioprinting [20]. This framework is not limited to a traditional Multilayer Perceptron (MLP) model; it has also recorded the characteristics of the viscosity dynamics and thermodynamic behavior of multi-component blends. It directly integrates the "Redlich-Kister (R-K) polynomial," a proven mathematical representation, into the main architecture and learning process of the artificial neural network. The shared loss function in the model's training phase is structured according to the strict thermodynamic constraints set by the R-K polynomial [21]. Thanks to this integration, the model can be trained with very limited laboratory data, eliminating the need for thousands of data points required by deep learning algorithms.

Furthermore, the synthetic outputs produced by the model do not contradict the laws of thermodynamics. Achieving statistical accuracy rates entirely equivalent to or superior to the popular and powerful Random Forest algorithm, HydroThermoMLP can produce safe and generalizable results in biomaterial formulation design without deviating from fundamental thermodynamic principles.
Similar data-driven strategies are also finding applications in other complex polymer systems of regenerative medicine. Support Vector Machines (SVM) and classification algorithms, Extreme Gradient Boosting (XGBoost) methods, and Decision Trees are far outperforming traditional empirical formulas and statistical models in areas such as viscosity prediction of polysaccharide colloids, storage modulus control in polyacrylamide hydrogels cross-linked with organic agents, and elasticity modulus estimation in inorganic-organic hybrid gels (e.g., calcium silicate hydrate networks) [21].

ALGORITHMIC OPTIMIZATION OF BIOINK PRINTABILITY

Determining the chemical and rheological properties of material formulations in a virtual environment using machine learning models is a key factor in the success of the bioprinting process. Material science alone handles half of this [15]. The other half of the equation is solved by how this optimized smart material interacts with the machine's electromechanical hardware and environmental factors after it is physically loaded into the bioprinter cartridge. The concept of "printability," frequently cited in the literature, is not solely related to the viscosity of the material; rather, other parameters are involved. These include the intrinsic rheology of the bio-ink and the device's printing parameters-extrusion pressure (pneumatic or mechanical), micro-nozzle inner diameter, print head horizontal movement speed (print speed/feedrate), vertical printing distance between the nozzle and the build plate, and build plate heating/cooling temperatures.
Artificial intelligence approaches, found in the literature and going beyond traditional mathematical equations (e.g., Navier-Stokes or traditional Cross models), play a critical role in classifying the impact of these parameters on printability. Among these algorithmic approaches, Bayesian Optimization (BO) is heavily preferred in bioprinting engineering, particularly for extracting the maximum statistical information from limited, costly, and time-consuming cellular experimental data [22]. Based on probability theory, BO, unlike traditional Grid Search or random search algorithms, uses the results of previous trials without blindly traversing the search space. This creates a surrogate model that "intelligently" guides which parameters should be used in the next experiment. In pioneering studies on complex hybrid bio-inks based on HAMA and GelMA, BO has managed to determine the material's ability to flow into a smooth filament with a much smaller

number of laboratory experiments (sometimes as few as one-tenth) compared to traditional laboratory methods [20].
Another powerful algorithmic approach used in this process optimization is Multi-Response Optimization (MRO) techniques [23]. The biggest obstacle to "perfect" printability in EBB systems is the complexity of the optimization goals. A typical dilemma faced by researchers in bioprinting is this: If live cells are used, the viscosity and extrusion pressure of the bioink must be kept very low to prevent damage to the live cells in the matrix from high pressure and shear forces (to ensure high cell viability). However, a material with a low viscosity, like water, cannot maintain its shape after exiting the nozzle. The printed scaffold also turns into a puddle. Increasing the density of the formulation (e.g., yield strength) to improve shape accuracy, on the other hand, increases the pressure requirement, breaking the cells inside. The MRO framework uses a mathematical strategy called Desirability Function Analysis (DFA) to reconcile these conflicting goals. In this technique, all input variables in the system (e.g., biopolymer concentrations in the formulation, additives, temperature, and pressure) are converted into surface response graphs using Central Composite Design (CCD) experimental models [24]. DFA determines the mechanically robust side of the system while also ensuring that the applied extrusion stress is weak enough not to rupture the cell membranes. In this way, machine learning algorithmically presents the designer with the ideal equilibrium point where conflicting biological and engineering requirements work together optimally.

ADDRESSING THE COMPLEXITY OF MULTI-MATERIAL AND HYBRID INTERFACES

While machine learning algorithms have achieved remarkable success with single-material inks, the ultimate goal of modern tissue engineering is not to print uniform, plastic-like masses of polymers. The real aim is to create natural human tissues (e.g., osteochondral tissues, heart muscle interwoven with vascular networks, or skin layers containing nerve networks), structures where mechanical strengths, cellular phenotypes, and biochemical signals vary at millimeter and even micrometric scales. Furthermore, the goal is to create complex, anisotropic, multi-material structures where cells, rigid collagen fibers, soft elastin tissues, capillary blood vessels, and mineralized calcium regions are interwoven in a perfect hierarchy [25]. To produce clinically implantable functional tissue counterparts for patients, researchers are compelled to move beyond a single material and towards hybrid 3D bioprinting, where multiple different bioinks, or even inorganic components such as biocompatible ceramics or electrically conductive polymers, are simultaneously and coordinately integrated.

However, the transition to multi-material production faces enormous rheological, thermodynamic, and mechanical hurdles that transcend the limitations of traditional materials science. Printing multiple materials together on the same platform (e.g., simultaneously processing a rigid thermoplastic scaffold to support mechanical loads with soft, cell-loaded hydrogels to create tissue formation, either in successive layers) can present challenges at the material interface. For instance, weak molecular bonding (interfacial delamination) can occur between heterogeneous materials [26]. Traditional theoretical models and experimental formulas cannot predict the highly nonlinear, chaotic nature of the interactions occurring at these multi-material interfaces.

In this context, the most significant problem hindering multi-material deposition in production is the "differential shrinkage" problem. When two different biomaterials in contact with each other and forming the same structure are exposed to environmental influences after the printing process (e.g., ultraviolet-UV cross-linking reactions, chemical coagulation baths, solvent evaporation, or thermal baking/sintering processes applied at high temperatures), they tend to physically shrink or expand [27]. However, each material has a very different solids loading ratio, liquid-holding capacity, and, most importantly, a coefficient of thermal expansion (CTE). Therefore, these materials shrink or expand at different rates and speeds at the interfaces. This thermomechanical incompatibility, frequently encountered in polymer, glass, and bioceramics, leads to severe internal and residual stresses in the printed composite part, warping, delamination of layers, or permanent cracking in fine microstructural details. Differential shrinkage is not just a mechanical failure. It is also a destructive factor that eliminates the optical transparency, biological insulation, and cell guidance capabilities of the structure [28].

Artificial intelligence models and advanced computerized systems have the potential to transform this differential shrinkage problem from a mere production obstacle into a revolutionary, programmable mechanical advantage. Furthermore, these models are revolutionizing the process of turning it into a practical advantage. On the one hand, machine learning (ML)-based compensation algorithms can dynamically optimize nozzle movement speed, build-up density in multiple material passes, active and passive cooling times, and bed temperatures during printing by learning from historical data about thermomechanical shrinkage and stress evolution rates during printing and curing of different materials. This allows for the development of pre-stress strategies to prevent deformation [29].

A rising star in the literature, 4D Bioprinting, utilizing user-directed stimulus-responsive bioinks, has developed highly innovative strategies. One of these strategies is to build biologically mobile, intelligent robotic hydrogels and tissues that "shape morph" in the time dimension (4th dimension) by utilizing these shrinkage differences. In this approach, machine learning algorithms calculate the behavioral models of intelligent hydrogels that shrink rapidly in response to environmental temperature changes, such as NIPAAm, and "Janus hydrogels" with asymmetric layers composed of inorganic binders such as rGO (reduced graphene oxide) or laponite, by combining them with complex Finite Element Analysis (FEA) simulations [30]. When factors such as near-infrared (NIR) light, a specific body temperature (37°C), or a physiological pH change are applied, the directional differential shrinkage is calculated, precisely predicting it during the design phase by the artificial intelligence. In the laboratory, the initially 2-dimensional (planar) structure, which is very easy to produce, spontaneously folds, much like biological origami, transforming into complex 3-dimensional tissue forms such as helical, conical, tubular (blood vessel-like), or lumenal. Without ML (Material Mapping), calculating the final form of this complex deformation mechanism is nearly impossible.

Another chronic problem encountered in the management of multi-material technology is the thermal incompatibility of inorganic bioceramics used in hard bone tissue regeneration with living cell-loaded organic polymers. In traditional ceramic production, the material needs to be sintered at high temperatures exceeding 1000°C to gain strength and density. In this case, these temperatures instantly destroy hydrogels, proteins, and of course, living cells. To overcome this major problem, the innovative "Cold Sintering Process" (CSP) developed in materials science has been integrated into bio-manufacturing technology with machine learning-driven multi-material mapping. CSP technology; It utilizes a carefully optimized combination of transient aqueous environments, dissolution-precipitation reactions, and external mechanical pressures at the interfaces [31,32]. Therefore, it can reduce the sintering temperatures of biocompatible ceramics such as calcium phosphate or zirconia (ZrO_2) to levels below 300°C, or even to room temperatures (25°C-150°C). These low thermal conditions allow brittle biopolymers to withstand degradation. It also enables the non-degradable, single-step co-printing of rigid ceramic bone grafts with flexible, organic cell-carrying hydrogels. These studies have paved the way for a new generation of hybrid tissue scaffolds.

REAL-TIME PROCESS CONTROL AND THE FUTURE OF SUSTAINABLE BIOPRINTING

Having perfectly optimized, ideal multi-material formulations and appropriate machine initial parameters is crucial. However, while this is an important stage in the challenging journey of bio-manufacturing, it is not the end of the process. Bioink rheology is not a static, unchanging concept. Rather, instantaneous humidity fluctuations in the laboratory environment where the bioprinter is located, micro-variations in ambient temperature, and spontaneous chemical changes that the polymers in the bioink syringe undergo as the process progresses are important. Examples include time-dependent cross-linking reactions, enzyme activations, or localized water evaporation. These continuously change the physical printability of the material throughout the hours-long production process. Traditional "open-loop" 3D bioprinting systems can be described as hardware-blind systems [33]. That is, after the user sends G-Code (toolpath and speed commands) to the machine, the printer mechanically executes these commands but cannot react to the quality of the material exiting the nozzle or changes in the environment. This situation, where sudden changes in material fluidity lead to nozzle clogging (extrusion discontinuity), layer slippage, and incomplete fillings (voids), ultimately results in defective prints, massive live cell death, and the waste of highly valuable clinical biomaterials.

To rescue bioprinting processes from this static blindness, machine learning-powered "closed-loop" real-time process control architectures are revolutionizing the biofabrication industry. These intelligent systems collect a continuous stream of data from a wide variety of sensors, including high-resolution macro cameras directly integrated into the printing hardware, laser optical scanners, acoustic emission sensors, and micro-pressure transducers in extruder syringes [34]. They then utilize computer vision systems equipped with Convolutional Neural Networks (CNNs) to process this data simultaneously. The CNN model analyzes the physical width of the fluid filament exiting the nozzle each second, the droplet size (in EHD inkjet printing), whether the material is stacking on top of the previous layer, or whether there is stringiness/collapse in the structure, all within microseconds (in real time). This allows for verification against the target CAD model [35].

The transformative aspect of the process begins at this analysis stage. The computer vision algorithm detects a sudden drop in ambient temperature in the laboratory or a sudden increase in the viscosity of the material due to local polymerization in the syringe. As a result, if the print line width is narrower than it should be or under-extrusion, the ML-based control loop instantly sends new commands to the bioprinter's motherboard without requiring human intervention. The autonomous system can slow down the motor feed rate in milliseconds. Alternatively, it can increase the temperature of the heaters around the nozzle or increase pneumatic air pressure to accelerate extrusion. As a result, it recalibrates the filament width and shape integrity to the targeted values [36].

This AI-powered adaptive and autonomous printing approach also forms the technological basis of "sustainable bioprinting," which has gained significant importance in biomedical research in recent years. Traditional manual calibration methods cause enormous trial-and-error costs, material waste, and unnecessary energy consumption. Furthermore, AI-integrated anomaly detection models can automatically halt the printing operation by immediately classifying the process as an "anomalous condition" when they detect irreparable damage to the bio-scaffold's structural integrity (shape fidelity) or when they predict the nozzle may damage the machine [33,35,36]. This prevents the waste of materials and time that would otherwise occur by continuing to build a defective structure, unlike open-loop systems.

FUTURE PERSPECTIVES: DIGITAL TWIN FRAMEWORKS IN BIOMANUFACTURING

The ultimate and most futuristic evolution of all these data-driven theoretical models and advanced sensor technologies' real-time closed-loop algorithms is described in the literature as "Digital Twin" (DT) bioprinting methods [37]. The digital twin concept is an interactive virtual copy of the bioprinter production system, the material used, or the biological tissue to be printed, managed virtually by artificial intelligence, and continuously fed and evolved with real-time live sensor data from the physical system, rather than being static. This is essentially based on the principle of developing virtual data simultaneously with real data [38]. Traditionally used in Industry 4.0 and the aerospace sector, this concept has been transferred to the field of biofabrication by An et al. (2021) and described as the "missing link" that will close the gap between cellular resolution and system engineering in the bioprinting ecosystem [11,39]. The aim of this framework is to place clinical production entirely in silico using massive amounts of Big Data. Therefore, creating a seamless bridge between virtual simulations and physical bioprinting cycles is the future of this field. The modern biomanufacturing architecture based on digital twins offers an extremely comprehensive and seamless cycle that goes beyond mere machine control:

Big Data Compilation and Virtual Shadow Generation

The fundamental point here is data mining. Large databases are created using complex "omics" data from high-resolution clinical imaging scans (MRI, CT, ultrasonography), histological tissue slices, immunohistochemical analyses, genetic sequencing (genomics), and patient profiles. Machine learning algorithms analyze this data. Subsequently, they segment medical images at the single-cell level and construct a flawless digital twin (virtual shadow) of the relevant anatomical organ (e.g., a patient-specific heart valve or liver lobe) with spatial accuracy, material biology, and mechanical topography. This model serves as a benchmark for the biomaterial to be produced [40]. Therefore, it is very different from a static 3D drawing or CAD file; it is an autonomous simulation reality subject entirely to mathematical rules, biophysical laws, and thermodynamic laws [41].

In Silico Optimization and Cell-Oriented Slicing

Long before any materials are used and a single wet experiment is performed in the lab, all printing scenarios are simulated in a risk-free virtual "sandbox." The simulation analyzes which printing parameters and nozzle speeds the extruder should operate at. It also predicts how different bioink formulations will cause shear stresses under nozzle pressures and how much damage these stresses will cause to cell membranes (cell viability) in the tissue. In bioprinting, layer-by-layer slicing of the model is far more critical than in standard 3D plastic printing. In the digital twin architecture, slicing is performed automatically and safely by ML, taking into account "intact cell-based" constraints so as not to damage the diameter, volume, and integrity of the cells [37,41].

Real-Time Bi-directional Synchronization Loop

The true power of the DT concept kicks in the moment the 3D printing process begins after simulation optimization is complete. Continuous sensor flows collected from the machine and printing environment (computer vision cameras, micro-extrusion pressure readers, thermal radiation cameras, and humidity monitors) are synchronized instantly to the digital twin model via wireless networks or industrial cloud systems. A micrometric geometric shift or rheological fluidity deviation occurring in the physical process from the design (model) is immediately detected by the digital twin. Then, the necessary compensation algorithms are calculated. Within milliseconds, the digital twin sends feedback commands to the physical machine, and the extruder pressure, nozzle height, or motor feedrate is automatically recalibrated. Case studies in the literature (e.g., the 2024 study by Ahmed et al.) have reported a massive structural improvement of 16.8% in the strut fidelity of polycaprolactone (PCL) scaffolds using this simultaneous digital twin cycle [37,42].

Post-Printing Tissue Maturation and Clinical Prediction

The most revolutionary medical feature of the digital twin framework is not only the mechanical control of the printer. It is also necessary to monitor and simulate the time period during which the generated living scaffold is transferred from the bioprinter to a "bioreactor" where tissue maturation takes place, as well as its long-term clinical performance. Nutrient flow, oxygen diffusion gradient, electrical stimulation, and fluid perfusion rates within the bioreactor are monitored by hardware sensors. Using this data, the digital twin can predict, with high confidence, cellular division (proliferation) in patient-specific tissue scaffolds, new collagen accumulation in the tissue, and even long-term in vivo (in-body) tissue integration or biological degeneration that would unfold over months, months in advance, via Artificial Neural Networks (ANNs). In neurological DT modeling, tumor radiotherapy planning, or liver cell degeneration research, these models have been found to successfully and accurately map, in silico, how real cells will respond to exogenous therapies (e.g., specific chemotherapy agents or antioxidants) applied after suppression, without the need for a laboratory environment [43]. These bioinformatics processes are currently performed using massive databases such as GDSC, TCGA, or CCLE [44].

Table 2. Digital Twin Frameworks in Biomanufacturing as Future Perspectives

Digital Twin (DT) Stages	Implemented Artificial Intelligence Strategy	Basic Bio-manufacturing Function and Output
Preprocessing and Data Compilation [41]	Data Mining and Semantic Segmentation	Extraction of single-cell precision 3D tissue maps (Virtual Shadows) from MRI and histology data.
In Silico Optimization [37,41]	Simulated ML, Interactive Optimization	Toolpath and slicing calculations to prevent cell loss in a risk-free virtual environment before printing.
Bi-directional Synchronization [37,42]	Real-Time Convolutional Neural Networks (CNN)	Instant adaptation of bio-printer hardware based on sensor flows; pressure and speed correction.
Tissue Maturation [43]	Predictive ML	Analysis of cell growth in bioreactors, projection of long-term mechanical degradation/recovery, drug testing.

CONCLUSIONS

In conclusion, the ultimate culmination of this vision, Machine Learning frameworks for rheological optimization of multimaterial bioinks in 3D bioprinting processes, offer cyber ecosystems that optimize the manufacturing process from the pre-clinical design phase to the bioreactor tissue maturation phase after printing, through a continuous virtual loop. As the current biofabrication literature clearly indicates, the transformation of a wide variety of live cell cultures and stimulus-responsive hybrid materials into reliable standard tissue grafts in clinical settings depends entirely on training these autonomous machine learning architectures with a continuous data stream. Simultaneously, this also depends on the industrial maturation of next-generation intelligent biofabrication frameworks based on high algorithmic predictive intelligence.

ACKNOWLEDGEMENTS

The author would like to express their sincere gratitude to Assoc. Prof. Dr. Ilyas Kartal for his invaluable guidance, continuous support, and constructive feedback during the preparation of this book chapter. This study was carried out within the scope of the graduate seminar course at the Department of Metallurgy and Materials Engineering, Institute of Pure and Applied Sciences, Marmara University.

REFERENCE

[1] I.A. Duceac, J.J. Janačopíková, J. Elango, C. Zamora-Ledezma, Rheological, Structural, and Biological Trade-Offs in Bioink Design for 3D Bioprinting, Gels 2025, Vol. 11, Page 659 11 (2025) 659. https://doi.org/10.3390/GELS11080659.

[2] S. Freeman, S. Calabro, R. Williams, S. Jin, K. Ye, Bioink Formulation and Machine Learning-Empowered Bioprinting Optimization, Front. Bioeng. Biotechnol. 10 (2022) 913579. https://doi.org/10.3389/FBIOE.2022.913579/FULL.

[3] S.M. Limon, R. Sarah, A. Habib, Integrating Decision Trees and Clustering for Efficient Optimization of Bioink Rheology and 3D Bioprinted Construct Microenvironments, J. Manuf. Sci. Eng. 147 (2025) 091003. https://doi.org/10.1115/1.4068429.

[4] J. Yu, D. Yao, L. Wang, M. Xu, Machine Learning in Predicting and Optimizing Polymer Printability for 3D Bioprinting, Polymers (Basel). 17 (2025). https://doi.org/10.3390/POLYM17131873.

[5] M.E. Cooke, D.H. Rosenzweig, The rheology of direct and suspended extrusion bioprinting, APL Bioeng. 5 (2021) 011502. https://doi.org/10.1063/5.0031475.

[6] I. Qavi, S. Halder, G. Tan, Modeling and Optimizing Rheology and Printability of Alginate-Gelatin Laponite Bioink with Multi-Response Optimization (Mro) and Artificial Neural Networks (Ann), (2023). https://doi.org/10.2139/SSRN.4584982.

[7] R. Sarah, K. Schimmelpfennig, R. Rohauer, C.L. Lewis, S.M. Limon, A. Habib, Characterization and Machine Learning-Driven Property Prediction of a Novel Hybrid Hydrogel Bioink Considering Extrusion-Based 3D Bioprinting, Gels 11 (2025) 45. https://doi.org/10.3390/GELS11010045/S1.

[8] F. Cadamuro, M. Piazzoni, E. Gamba, B. Sonzogni, F. Previdi, F. Nicotra, A. Ferramosca, L. Russo, Artificial Intelligence tool for prediction of ECM mimics hydrogel formulations via click chemistry, Biomaterials Advances 175 (2025). https://doi.org/10.1016/J.BIOADV.2025.214323.

[9] N.A. Dudukovic, L.L. Wong, D.T. Nguyen, J.F. Destino, T.D. Yee, F.J. Ryerson, T. Suratwala, E.B. Duoss, R. Dylla-Spears, Predicting Nanoparticle Suspension Viscoelasticity for Multimaterial 3D Printing of Silica-Titania Glass, ACS Appl. Nano Mater. 1 (2018) 4038–4044.

https://doi.org/10.1021/ACSANM.8B00821/SUPPL_FILE/AN8B00821_SI_003.AVI.
[10] N. Lagneau, P. Tournier, B. Halgand, F. Loll, Y. Maugars, J. Guicheux, C. Le Visage, V. Delplace, Click and bioorthogonal hyaluronic acid hydrogels as an ultra-tunable platform for the investigation of cell-material interactions, Bioact. Mater. 24 (2023) 438. https://doi.org/10.1016/J.BIOACTMAT.2022.12.022.
[11] J. An, C.K. Chua, V. Mironov, Application of Machine Learning in 3D Bioprinting: Focus on Development of Big Data and Digital Twin, Int. J. Bioprint. 7 (2021) 1–6. https://doi.org/10.18063/IJB.V7I1.342.
[12] J. Sun, K. Yao, J. An, L. Jing, K. Huang, D. Huang, Machine learning and 3D bioprinting, Int. J. Bioprint. 9 (2023) 717. https://doi.org/10.18063/IJB.717.
[13] Z. Zhang, Y. Wang, W. Wang, Machine Learning in Gel-Based Additive Manufacturing: From Material Design to Process Optimization, Gels 11 (2025). https://doi.org/10.3390/GELS11080582.
[14] I.C. Karaoglu, A.O. Kebabci, S. Kizilel, Optimization of Gelatin Methacryloyl Hydrogel Properties through an Artificial Neural Network Model, ACS Appl. Mater. Interfaces 15 (2023) 44796–44808. https://doi.org/10.1021/ACSAMI.3C12207.
[15] S. Zhou, D. Song, L. Pu, W. Xu, Recent Advances in Machine Learning Assisted Hydrogel Flexible Sensing, Z. Anorg. Allg. Chem. 650 (2024) e202400051. https://doi.org/10.1002/ZAAC.202400051;WGROUP:STRING:PUBLICATION.
[16] D. Ege, A.R. Boccaccini, Investigating the Effect of Processing and Material Parameters of Alginate Dialdehyde-Gelatin (ADA-GEL)-Based Hydrogels on Stiffness by XGB Machine Learning Model, Bioengineering 11 (2024) 415. https://doi.org/10.3390/BIOENGINEERING11050415/S1.
[17] Z. Zhu, R. Wang, N. Li, Z. Pan, K. Miao, S. Huang, Q. Li, D. Li, Physics-informed machine learning for mapping of composition and mechanical properties in 316L/AlCrFeNi multi-material alloys fabricated by directed energy deposition, Virtual Phys. Prototyp. 21 (2026). https://doi.org/10.1080/17452759.2026.2648350.
[18] S. Xu, X. Chen, S. Wang, Z. Chen, P. Pan, Q. Huang, Integrating machine learning for the optimization of polyacrylamide/alginate hydrogel, Regen. Biomater. 11 (2024). https://doi.org/10.1093/RB/RBAE109.
[19] X. Liu, L. Hu, S. Chen, Y. Ran, J. Pang, S. Wu, Machine learning analysis for the rheological mechanism of polysaccharide colloids, J. Mol. Liq. 424 (2025) 127093. https://doi.org/10.1016/J.MOLLIQ.2025.127093.
[20] B. Deng, S. Chen, F.L. Lasaosa, X. Xue, C. Xuan, H. Mao, Y. Cui, Z. Gu, M. Doblare, Predicting rheological properties of HAMA/GelMA hybrid hydrogels via machine learning, J. Mech. Behav. Biomed. Mater. 168 (2025) 107005. https://doi.org/10.1016/J.JMBBM.2025.107005.
[21] A. Gayathri, T. Venugopal, K. Venkatramanan, Redlich-Kister Coefficients on the Analysis of Physico-Chemical Characteristics of Functional Polymers, Mater. Today Proc. 17 (2019) 2083–2087. https://doi.org/10.1016/J.MATPR.2019.06.257.
[22] B. Tan, S. Kuang, X. Li, al -, W. Long Ng, C. Vyas, B. Huang, J. Min Lee, M. Zhou, D. Mohammadrezaei, L. Podina, J. De Silva, M. Kohandel, Cell viability prediction and optimization in extrusion-based bioprinting via neural network-

based Bayesian optimization models, Biofabrication 16 (2024) 025016. https://doi.org/10.1088/1758-5090/AD17CF.
[23] I. Qavi, S. Halder, G. Tan, Optimization of printability of bioinks with multi-response optimization (MRO) and artificial neural networks (ANN), Progress in Additive Manufacturing 2024 10:5 10 (2024) 3573–3598. https://doi.org/10.1007/S40964-024-00828-1.
[24] P. Karande, S.K. Gauri, S. Chakraborty, Applications of utility concept and desirability function for materials selection, Mater. Des. 45 (2013) 349–358. https://doi.org/10.1016/J.MATDES.2012.08.067.
[25] N. Betancourt, X. Chen, Review of extrusion-based multi-material bioprinting processes, Bioprinting 25 (2022) e00189. https://doi.org/10.1016/J.BPRINT.2021.E00189.
[26] Y. Wu, H. Su, M. Li, H. Xing, Digital light processing-based multi-material bioprinting: Processes, applications, and perspectives, J. Biomed. Mater. Res. A 111 (2023) 527–542. https://doi.org/10.1002/JBM.A.37473;CTYPE:STRING:JOURNAL.
[27] A.A. Al-Tamimi, M. Tlija, M.H. Abidi, A. Anis, A.E.E. Abd Elgawad, Material Extrusion of Multi-Polymer Structures Utilizing Design and Shrinkage Behaviors: A Design of Experiment Study, Polymers 2023, Vol. 15, Page 2683 15 (2023) 2683. https://doi.org/10.3390/POLYM15122683.
[28] S.B. Ghantasala, G. Singh, J.M. Missiaen, D. Bouvard, Shrinkage and deformation of material extrusion 3D printed parts during sintering: Numerical simulation and experimental validation, Acta Mater. 282 (2025) 120518. https://doi.org/10.1016/J.ACTAMAT.2024.120518.
[29] X. Hua, Y. Cao, B. Liu, X. Yang, H. Xu, L. Li, J. Wu, Prediction of Thermomechanical Behavior of Wood–Plastic Composites Using Machine Learning Models: Emphasis on Extreme Learning Machine, Polymers 2025, Vol. 17, Page 1852 17 (2025) 1852. https://doi.org/10.3390/POLYM17131852.
[30] W. Guo, M. Mirzaei, L. Nie, Janus Hydrogels: Design, Properties, and Applications, Gels 2025, Vol. 11, Page 717 11 (2025) 717. https://doi.org/10.3390/GELS11090717.
[31] S. Grasso, M. Biesuz, L. Zoli, G. Taveri, A.I. Duff, D. Ke, A. Jiang, M.J. Reece, A review of cold sintering processes, Advances in Applied Ceramics 119 (2020) 115–143. https://doi.org/10.1080/17436753.2019.1706825;ISSUE:ISSUE:DOI.
[32] H. Guo, T.J.M. Bayer, J. Guo, A. Baker, C.A. Randall, Current progress and perspectives of applying cold sintering process to ZrO2-based ceramics, Scr. Mater. 136 (2017) 141–148. https://doi.org/10.1016/J.SCRIPTAMAT.2017.02.004.
[33] J. Arduengo, N. Hascoet, F. Chinesta, J.Y. Hascoet, Open-Loop Control System for High Precision Extrusion-Based Bioprinting Through Machine Learning Modeling, Journal of Machine Engineering 24 (2024) 103–117. https://doi.org/10.36897/JME/186044.
[34] D. Ke, AI-driven closed-loop extrusion-based bioprinting system: from dynamic regulation to medical translation, Bioprinting 54 (2026) e00459. https://doi.org/10.1016/J.BPRINT.2025.E00459.
[35] J. Arduengo, N. Hascoët, … F.C.-J. of M., undefined 2025, Closed-loop control of extrusion-based bioprinting through real-time computer vision, Yadda.Icm.Edu.PlJ Arduengo, N Hascoët, F Chinesta, JY HascoetJournal of

Machine Engineering, 2025•yadda.Icm.Edu.Pl (n.d.). https://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-b0a1785a-92fd-452d-93ea-2fb7015e942f (accessed May 2, 2026).
[36] D. Kelly, V. Sergis, L. Ventura i Blanco, K. Mason, A.C. Daly, Autonomous Control of Extrusion Bioprinting Using Convolutional Neural Networks, Adv. Funct. Mater. 35 (2025) 2424553. https://doi.org/10.1002/ADFM.202424553;JOURNAL:JOURNAL:10990712;PAGE:STRING:ARTICLE/CHAPTER.
[37] N. Ahmed, L.K. Narayanan, R. Shirwaiker, A framework for digital twin integration in biofabrication and a scaffold 3D bioplotting case study, Manuf. Lett. 41 (2024) 1182–1191. https://doi.org/10.1016/J.MFGLET.2024.09.144.
[38] V. Jokanović, M. Živković, Applications of Digital Twins in Personalized Medicine and Healthcare, Digital Twins and Simulation Technology: Concepts and Applications (2025) 95–115. https://doi.org/10.1201/9781003582489-8/APPLICATIONS-DIGITAL-TWINS-PERSONALIZED-MEDICINE-HEALTHCARE-JOKANOVI.
[39] L. Wang, Y. Jiang, J. Zhu, K. Lin, X. Ma, J. He, B. Lu, Digital twin for in – space manufacturing: a comprehensive review from framework to future, Virtual Phys. Prototyp. 21 (2026). https://doi.org/10.1080/17452759.2026.2639147.
[40] J. Möller, R. Pörtner, Digital Twins for Tissue Culture Techniques—Concepts, Expectations, and State of the Art, Processes 2021, Vol. 9, Page 447 9 (2021) 447. https://doi.org/10.3390/PR9030447.
[41] V. Kumar, A.K. Yadav, T. Ahmad, P.K. Srivastava, A. Ahmadian, Emerging Challenges and Future Directions in Multiscale Modelling for Integration of Biology and Materials Design, Smart Materials Engineering (2026) 187–212. https://doi.org/10.1007/978-3-032-09540-4_10/SAVE-RESEARCH.
[42] J. Xie, A. Saluja, A. Rahimizadeh, K. Fayazbakhsh, Development of automated feature extraction and convolutional neural network optimization for real-time warping monitoring in 3D printing, Int. J. Comput. Integr. Manuf. 35 (2022) 813–830. https://doi.org/10.1080/0951192X.2022.2025621.
[43] W.L. Ng, A. Chan, Y.S. Ong, C.K. Chua, Deep learning for fabrication and maturation of 3D bioprinted tissues and organs, Virtual Phys. Prototyp. 15 (2020) 340–358. https://doi.org/10.1080/17452759.2020.1771741.
[44] Z. Yao, Q. Zhang, GCLM-CDR: A graph contrastive learning method with multi-omics for cancer drug response prediction, Proceedings - 2024 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2024 (2024) 1302–1307. https://doi.org/10.1109/BIBM62325.2024.10822189.

Characterization Techniques for Spinodal Alloys

Fehim FINDIK[1,2,c]

1- Prof. Dr. Metallurgy and Materials Engineering Department, Faculty of Technology, Sakarya Applied Sciences University, Sakarya, Turkey. ORCID No: 0000-0003-2537-1951
2: Prof. Dr. Istanbul Aydın University, Mechanical Eng Dept, Engineering Faculty, Florya Halit Aydın Campus, Istanbul, Turkey
C: Corresponding author; e-mail: findik@subu.edu.tr

ABSTRACT

The unique physical and mechanical properties of spinodal alloys—arising from their characteristic, spontaneously modulated microstructures—necessitate a multi-faceted characterization approach that spans from the atomic to the macroscopic scale. At the most fundamental level, atom probe tomography (APT) provides unparalleled, three-dimensional atomic-scale mapping of compositional fluctuations, enabling direct observation of the early stages of spinodal decomposition with sub-nanometer resolution. Complementing APT, transmission electron microscopy (TEM), particularly in combination with energy-filtered and high-angle annular dark-field imaging, visualizes the periodic strain contrasts and morphology of the modulated structure, while in-situ heating stages allow for the study of coarsening kinetics. Small-angle X-ray and neutron scattering (SAXS/SANS) offer statistically robust, quantitative measurements of the characteristic wavelength, amplitude, and the kinetics of decomposition across macroscopic volumes. Further structural insights are provided by X-ray and neutron diffraction, which reveal lattice parameter variations and the presence of coherency strains essential to the alloy's strengthening mechanism. Finally, to connect this nanoscale architecture to performance, localized mechanical testing methods such as nanoindentation and micropillar compression are employed to directly correlate the refined microstructure with the exceptional strength and stability characteristic of these materials. In the current work, experimental techniques are criticized for the sake of understanding for spinodal alloys. Then the microstructural and mechanical properties of some Cu-Ni based spinodal alloys are studied via XRD, SEM, TEM as well as hardness tests, respectively. It is concluded that wavelength is obtained as a constant value in the early stages of ageing and then it is increased gradually in the later stages of ageing for both Cu-Ni base spinodal alloys.

Keywords – Spinodal alloys, characterization, XRD, SEM-TEM, hardness

INTRODUCTION

Spinodal alloys represent a unique and technologically significant class of materials whose remarkable properties derive not from a multiphase equilibrium structure, but from a specific, non-classical mode of phase transformation known as spinodal decomposition. In contrast to the more common nucleation-and-growth mechanism, which requires a thermodynamic barrier to initiate the formation of discrete precipitate particles, spinodal decomposition occurs spontaneously within a miscibility gap. When a homogeneous solid solution is rapidly quenched from a single-phase region into the spinodal region of the phase diagram—defined by the condition $\partial^2G/\partial c^2<0$, where G is the Gibbs free energy and c*c* is the composition—the alloy becomes unstable against infinitesimally small compositional fluctuations. These fluctuations do not require activation energy; instead, they grow exponentially with time through a process of "uphill diffusion," where atoms diffuse *towards* regions of their own species, against the concentration gradient.

This process results in a characteristic, interconnected, and highly periodic microstructure. The hallmark of a spinodal structure is a diffuse interface with a sinusoidal compositional variation, which evolves from a fine, coherent modulation to a coarser, more defined structure as decomposition proceeds. The defining length scale of this microstructure—the spinodal wavelength (λ)—is a function of the diffusion mobility and the thermodynamic driving force, and it dictates the alloy's ultimate physical and mechanical properties. Because the transformation occurs via a continuous, rather than a discrete, path, the resulting microstructure is free of the nucleation barriers and interfacial energies that characterize conventional precipitation-hardened alloys, often leading to superior strength, thermal stability, and unique magnetic or electrical properties.

The practical history of spinodal alloys predates their theoretical understanding. As early as the 1940s and 1950s, metallurgists observed unusual hardening behavior in alloys such as Cu-Ni-Fe and Alnico (Al-Ni-Co). The Alnico family of permanent magnets, developed in the 1930s, exhibited exceptionally high coercivity, which was later attributed to a finely modulated, interconnected two-phase structure—a classic example of a spinodal alloy, though not recognized as such at the time.

The theoretical breakthrough by Cahn and Hilliard in the late 1950s provided the framework to interpret these observations. Their work transformed the field from empirical alloy development to a science-based discipline. The 1960s and 1970s saw the deliberate design of spinodal alloys, most notably the "Spinodal" copper-nickel-tin (Cu-Ni-Sn) alloys developed by Bell Telephone Laboratories, which demonstrated exceptional strength and conductivity. The advent of advanced characterization tools, particularly transmission electron microscopy (TEM) and, later, atom probe tomography (APT), was crucial in validating the Cahn-Hilliard theory by providing direct, real-space images of the predicted sinusoidal modulations and interconnected morphologies.

The spinodal decomposition phenomenon has been harnessed across a diverse range of metallic, ceramic, and even polymeric systems. Key alloy families include copper, iron, nickel, alnico alloys as well as high entropy systems:

- **Copper-Based Alloys:** Cu-Ni-Sn (e.g., C72900) and Cu-Ni-Fe are quintessential spinodal systems. Cu-Ni-Sn alloys are renowned for their combination of high strength (comparable to beryllium copper), good corrosion resistance, and excellent electrical conductivity. They are a safer alternative to beryllium-containing alloys and are used extensively in connectors, springs, and bushings.
- **Iron-Based Alloys:** Fe-Cr (ferrite in stainless steels) and Fe-Cr-Co are important spinodal systems. Fe-Cr separation is critical to the embrittlement of duplex stainless steels at elevated temperatures but is also harnessed to create high-coercivity permanent magnets in the Fe-Cr-Co system, which offer a ductile alternative to Alnico.
- **Nickel-Based Alloys:** Ni-Al, Ni-Ti, and Ni-Cr-Al (used in superalloys) undergo spinodal decomposition that contributes to the formation of the coherent γ' (Ni3Al) precipitates, providing exceptional high-temperature creep resistance in turbine blades.
- **Alnico Alloys:** (Al-Ni-Co) The prototype spinodal magnet. Its complex decomposition into Fe-Co-rich ($\alpha 1$) and Ni-Al-rich ($\alpha 2$) phases creates a finely modulated structure that is the source of its high coercivity, making it indispensable in applications requiring stable magnetic fields, such as sensors, motors, and traveling-wave tubes.
- **Emerging Systems:** The principles of spinodal decomposition are increasingly applied in high-entropy alloys (HEAs) to create hierarchical, stable microstructures and in ceramic systems such as zirconia-based oxides for solid oxide fuel cells, where spinodal decomposition is used to create interconnected ionic conduction pathways.

The unique microstructural features of spinodal alloys—namely, the absence of discrete interfaces, the interconnected morphology, and the coherency strain fields—impart a suite of properties that are exploited in high-performance applications in aerospace, electronics, magnetic, biomedical and marine areas:

- **Aerospace and Energy:** Nickel-based superalloys used in gas turbine engines rely on spinodal decomposition to form coherent precipitate networks that retain strength at temperatures exceeding 1000°C. The exceptional oxidation resistance and creep strength of these materials are directly linked to the stability of the spinodal decomposed microstructure.
- **Electronics and Connectors:** Cu-Ni-Sn spinodal alloys are the material of choice for high-reliability electrical connectors in automotive, aerospace, and consumer electronics. They offer a unique combination of high yield strength, stress relaxation

resistance, and conductivity, essential for maintaining contact force over long periods.

- **Magnetic Applications:** Alnico and Fe-Cr-Co spinodal magnets dominate applications requiring high magnetic induction and temperature stability, such as in electric motors for aerospace actuators, magnetic sensors, and high-end loudspeakers. The spinodal structure allows for a fine tuning of magnetic properties through controlled heat treatments.
- **Biomedical and Marine:** The high strength, galling resistance, and corrosion resistance of spinodal Cu-Ni-Sn alloys make them suitable for marine hardware, subsea connectors, and certain biomedical instruments where non-magnetic properties are also desirable.

The aim of this study is to define the spinodal, mention about the history, various spinodal systems and application areas. Also, the experimental methods to study microstructural investigation as well as mechanical properties of spinodal alloys are criticized. Finally, to detect and compare the spinodal structures of some Cu-Ni based spinodal alloys are discussed.

EXPERIMENTAL TECHNIQUES

Many experimental techniques are used in the characterization of spinodal alloys. Some of these techniques are used in microstructure examination, while others are used in determining mechanical properties. XRD, SEM, TEM, ultrasound, atom probe, and DSC systems are used in microstructure examination [Findik, 1992:85]. Hardness measurement techniques provide information about the mechanical properties of spinodal alloys. Some of these techniques will be summarized here.

XRD

The formation of spinodally modulated structures has been associated with anomalous diffraction phenomena. Daniel and Lipson [Cahn, 1961:795] explained how the fundamental Bragg reflections in X-ray diffraction (XRD) patterns can be surrounded by sidebands arising from microstructural regularity.

The wavelength calculated from this equation, known as the Daniel-Lipson equations, is the perpendicular orientation of the planes (hkl). Due to the <100> directionality of modulated structures in cubic crystals, sidebands around {200} are commonly analyzed. It is reported that there is usually good agreement between the modulation wavelengths calculated from this equation and those recorded from microscopic examination [Cahn, 1962:907], In fact, some researchers have been able to estimate the strain amplitude from the integrated area of the sidebands [Cahn and Hilliard, 1971:151], [Lefevre et al, 1978:577]. However, validation of this approach through alternative characterization methods is lacking. Therefore, sideband intensity is only taken as a qualitative indicator of amplitude.

The sideband method is particularly effective in the range of ~5-15 nm wavelengths [Meijering, 1957:257]. For very short wavelengths specific to the early stages of decomposition, even with other techniques, sidebands are generally not observed even in the distribution of the clustering of controlled elements [Baik and Yon, 1990:1525]. This is probably due to the low amplitudes and the presence of a broad wavelength spectrum before selective amplification of a characteristic wavelength, which makes the sidebands too weak and diffuse to be detected. Alternatively, small-angle X-ray scattering (SAXS) is widely used to extend the early stages of spinodal technology [Davis, 1977:92]. Its regular variation in scattering power directly affects the regional regions in the area, allowing for the possibilities of amplitude distribution of the wavelength spectrum. In systems where the separating elements, such as Fe-Cr and Au-Pt, have similar atomic numbers, the X-ray scattering contrast is minimal. In such cases, small-angle neutron scattering (SANS) is often preferred due to the non-monotonic variation of neutron-bound scattering lengths with atomic number [Tsubaino et al., 1991:2851].

Sideband analysis can also be challenging for long-wave body modulus structures, as Bragg's peaks may be too close for reliable analysis of sidebands, but more advanced diffractometers have enabled more convenient decomposition of these peaks [Selcuk, 1989:105]. These advances also reveal that sideband distributions are often asymmetric [Meijering, 1957:257]. He et al. [Kunze and Wincierz, 1965:421] showed that this can be explained by the superposition of lattice parameter variation and scattering power variation. However, strong density asymmetry has also been reported in systems with minimal scattering power variation between separating elements. In such cases, an asymmetric composition profile has been suggested as the source of the observed asymmetry [Böhm, 1961:564], which may occur in alloys with asymmetric compositions. This is supported by the observation of [Rao et al., 1986:3759] that the sideband asymmetry in Cu-Ni-Cr alloys is strongest in asymmetric compositions. Some researchers have also reported asymmetry in the position of sidebands, particularly in Cu-Ni-Sn [Lefevre et al, 1978:577], [Fournelle, 1979:1135] and Cu-Ti [Frebel and Schenk, 1979:230] alloys.

SEM

Microstructural characterization at a resolution of 10–100 nm can be performed using scanning electron microscopy (SEM). This allows observation of features such as grain boundaries, twin structures, and sufficiently large precipitates. In spinodal alloys, samples for SEM are prepared following standard metallographic preparation techniques. Final polishing can be performed with a 0.1 μm diamond suspension. Microstructural examination can be performed using a Zeiss Gemini 300 microscope operating at an acceleration voltage of 20 kV. This energy is sufficient to produce characteristic X-rays for the relevant elements when performing compositional analyses. These analyses can be performed via energy dispersive X-ray spectroscopy (EDX) using an Oxford Instruments X-MaxN 50 solid-state detector attached to the same instrument.

TEM

Transmission electron microscopy (TEM) is used to image finer microstructural features that cannot be resolved using SEM. Electron-transmitting specimens can be prepared using one of two methods: double-jet electropolishing and focused ion beam (FIB) milling. The former can be used to prepare Cu-15Ni-8Sn (wt.%) specimens for ex situ study of their microstructural evolution. Discs approximately 3 mm in diameter and approximately 0.2 mm thick can be electropolished at -30 °C using a Tenupol-5 apparatus and a 10% volume nitric acid solution in methanol. The latter can be used to produce Au-Pt-Pd alloy specimens to minimize the loss of precious metals. During this process, a small lamella is removed from the polished surface. The specimens can then be thinned to a thickness of approximately 100 nm using an ion source.

Microstructural characterization can be performed using an FEI Tecnai Osiris microscope operating at 200 kV and equipped with four FEI Super-X EDX detectors. Samples are first oriented along an appropriate region axis, and selected area diffraction patterns (SADPs) are acquired to enable phase identification. Morphological aspects of the microstructure can be examined by scanning transmission electron microscopy (STEM) with images obtained using both bright-field (BF) and high-angle ring dark-field (HAADF) signals. Additionally, chemical variations can be assessed by obtaining STEM-EDX elemental distribution maps.

In all samples examined in this study, compositional modulation can be expected to occur primarily along the elastically soft <100> directions. Therefore, EDX data can be obtained by viewing downwards from the [001] region axis so that the [100] and [010] directions are located in the image plane.

ULTRASOUND

Various techniques have been used to study spinodal alloys, in addition to microscopy and diffraction-based methods. Resonant ultrasound spectroscopy (RUS) is commonly used to investigate the variation of elastic properties with microstructural conditions. This technique is used to measure the normal frequencies of vibration of an object, which depend on the geometry, density, symmetry, and elastic properties of the sample. If the geometry, density, and symmetry are known, the elastic properties can be obtained by analyzing the RUS data. Various sample geometries are possible, but the most commonly used is the rectangular prism. In a previous study, rectangular prisms with nominal dimensions of $2.5 \times 3.5 \times 4.5$ mm^3 were prepared by electrical discharge machining (EDM). The surfaces were ground using successively finer SiC sandpaper to a final surface quality of 4000 grit to remove the surface layer affected by machining. To minimize errors in RUS measurements, it is important to ensure that opposite faces remain parallel and that the corners are sharp. The final sample mass and dimensions were measured using a balance and digital caliper, respectively [Hogg, 2025:136].

A detailed description of the RUS technique is provided by Migliori and Sarrao [Migliori, 1997:1]. In the RUS system, the sample was held gently at its corners between two piezoelectric transducers. One transducer was progressively scanned in the ultrasonic frequency range. The second transducer measured the amplitude of the sample displacement at each frequency, with peaks occurring in the sample's natural mechanical resonances. Spectra were collected using ACE-RUS008 electronics from Alamo Creek Engineering. To ensure that all resonance peaks were detected, data was collected by holding each sample between three different sets of corners.

ATOM PROBE

Atomic probe tomography (APT) can be applied to detect nanoscale compositional fluctuations beyond the resolution of STEM-EDX. In a previous study, sharp tips were prepared from polished surfaces using a Zeiss electron microscope, according to the approach reported by Thompson et al. [Thompson et al., 2007:131]. APT measurements were performed using a Cameca LEAP 4000X HR. The laser pulse mode was applied with a pulse rate of 125 kHz and a pulse energy of 50 pJ. The base of each sample was kept at ~50 K (~-223 °C), and the target detection rate was adjusted so that an average of 5 ions were detected per 1000 pulses. 3D reconstructions of the data were performed using the IVAS 3.6.8 software package.

DSC

Differential scanning calorimetry (DSC) can be applied to detect thermal signals associated with phase transitions. In a previous study [Hogg, 2025:110], the measured transition temperatures (e.g., dissolution point, solidification point) provided significant constraints for the thermal processing of the alloy. Disc-shaped specimens with a diameter of ~5 mm and a thickness of ~1 mm was prepared by electro-discharge machining (EDM). To remove the remelted layer, all surfaces were successively sanded with finer SiC sandpaper to a grit size of 4000 grit. Calorimetric data were obtained using a Netzsch 404 instrument under argon flowing at a flow rate of 50 mL min^{-1}. Temperature was measured with specific thermocouples and calibrated with pure elements.

RESULTS

In this part, the results of microstructural characterization and mechanical properties (especially hardness measurements) of Cu-Ni-Cr and Cu-Ni-Sn alloys are discussed respectively.

CU-NI-CR ALLOYS

In this study, four alloys (Cu-Ni-2.5Cr, Cu-Ni-5Cr, Cu-Ni-10Cr, and Cu-Ni-15Cr) are investigated.

MICROSTRUCTURE

As-cast micrographs of Cu-Ni based spinodal alloys (2.5Cr, 5Cr, 10Cr and 15Cr alloys) are seen as dendritic structures in Figure 1.

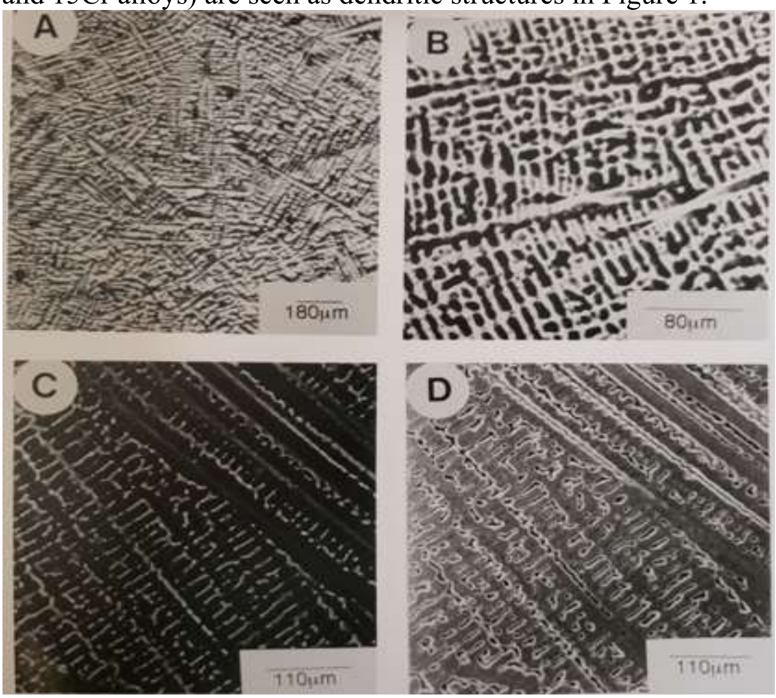


Fig 1. As -cast structures of Cu-Ni based spinodal alloys.

AFTER AGING MICROSTRUCTURE EARLY STAGES OG AGING

Various evidence of spinodal decomposition products was detected in TEM image of rapidly cooled Cu-based samples. In the early stages of aging, the spinodal decomposition inspected by XRD and TEM were: (i) sidebands in XRD peaks, (ii) spinodal structure in electron micrographs (Figure 2), (iii) preliminary aging related with constant wavelength, [Findik and Flower, 1992:197], [Findik, 2012:131].

The volume ratios of the phases at the aging temperature were strongminded from the coupling lines generated by X-ray EDS analysis of the phases. For example, the volume ratio of the Cu-free phase was 0.12 (wt.%) for the Cu-30Ni-5Cr alloy and 0.46 (wt.%) for the Cu-45Ni-10Cr alloy at 800°C [Findik 2002:337], [Findik, 2012:131].

In all Cu-based alloys, the particles were small, periodic, and cubic throughout the preliminary stages of aging, and λ persisted constant. No heterogeneous precipitation or precipitation-free regions were detected alongside grain boundaries. The particles clustered together and coalesced to form elongated periodic rod arrays in the <100> directions in both alloys as aging progressed. After prolonged aging periods at high temperatures, interfacial dislocations established in the particles (Figure 3b and 3c), and grain boundary precipitation was detected (Figure 3d). This is evidence of loss of coherence and is associated with the decrease in hardness in the aging curves. After prolonged aging, the coarsened structure confined platelets [Zhao and Notis, 1998:4203], [Findik, 2012:131].

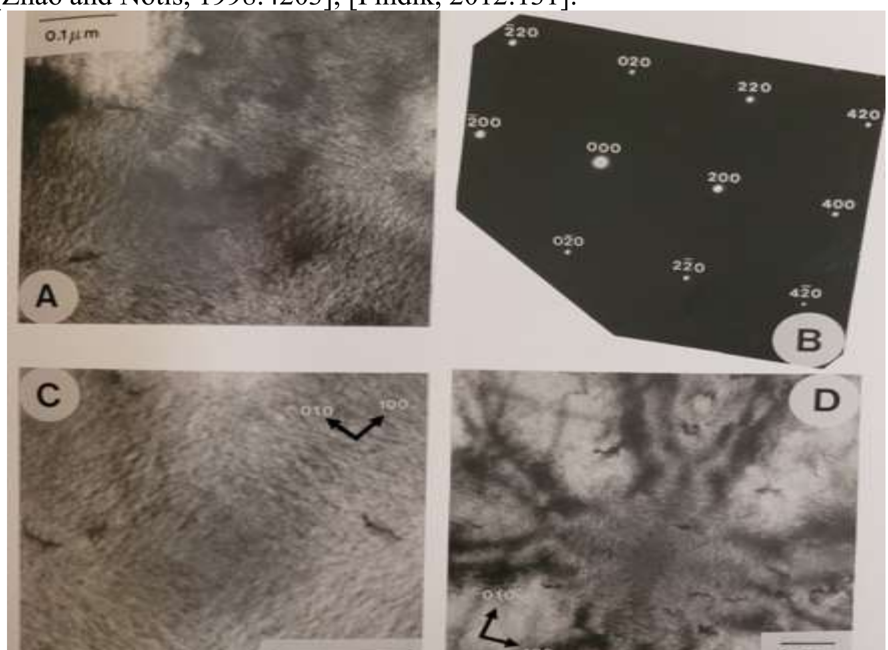


Fig. 2. Early stages of aging for spinodal alloys.

LATER STAGES OF AGING

In copper-based alloys, protracted aging, especially at higher temperatures, led to the formation of small spherical particles of a third phase (Figure 3). Diffraction showed that these particles were bcc, and EDS analysis determined that they were very rich in chromium [Findik, 2012:131].

In the Cu-45Ni-10Cr alloy, the microstructure of the decomposition products did not change excessively with aging, and the cubic form was preserved even at high aging temperatures. Instead, at 700 and 800°C, the cubic particle preparation was irregular in spreading (e.g., Figure 3), and no side bands were observed in the XRD patterns [Zhao and Notis, 1998:4203].

In contrast, the initially produced cubics in the Cu-45Ni-10Cr alloy transformed into rods aligned in the (100) direction (Figure 3a) and finally into plates (Figure 3b). This morphological change progression is consistent with previous reports [Laughlin and Cahn, 1975:329].

The coarsening of the decomposition products was monitored by measuring the wavelength of the periodic structure as a function of aging temperature and time. The decomposition wavelength was determined either directly from magnified microstructures (large L) or by calculating the sidebands in the XRD patterns via the Daniel-Lipson equation [Rao et al., 1991:1485], [Findik, 2012:131].

The progression of the wavelength with time and temperature can be detected in Figure 5 for both alloys. At lower temperatures (300-500°C), these values are not included in Figure 5a because the measurement of wavelengths is not easy. Furthermore, in the Cu-45Ni-15Cr alloy, the wavelengths at 300 and 400°C were very similar, and therefore only the values at 400°C. The coarse grain performance determined from the log L-log(t - to) graphs of the Cu-45Ni-15Cr alloy is according to equation (2) when the exponent $n = 3$. The fit to the data is less good for other n values. However, the Cu-30Ni-5Cr alloy did not obey this roughening law, especially at low temperatures (600°C) [Baker et al, 2009:886], [Findik, 2012:131]. The association amid the average particle size p and the wavelength L is exposed in Figure 4.

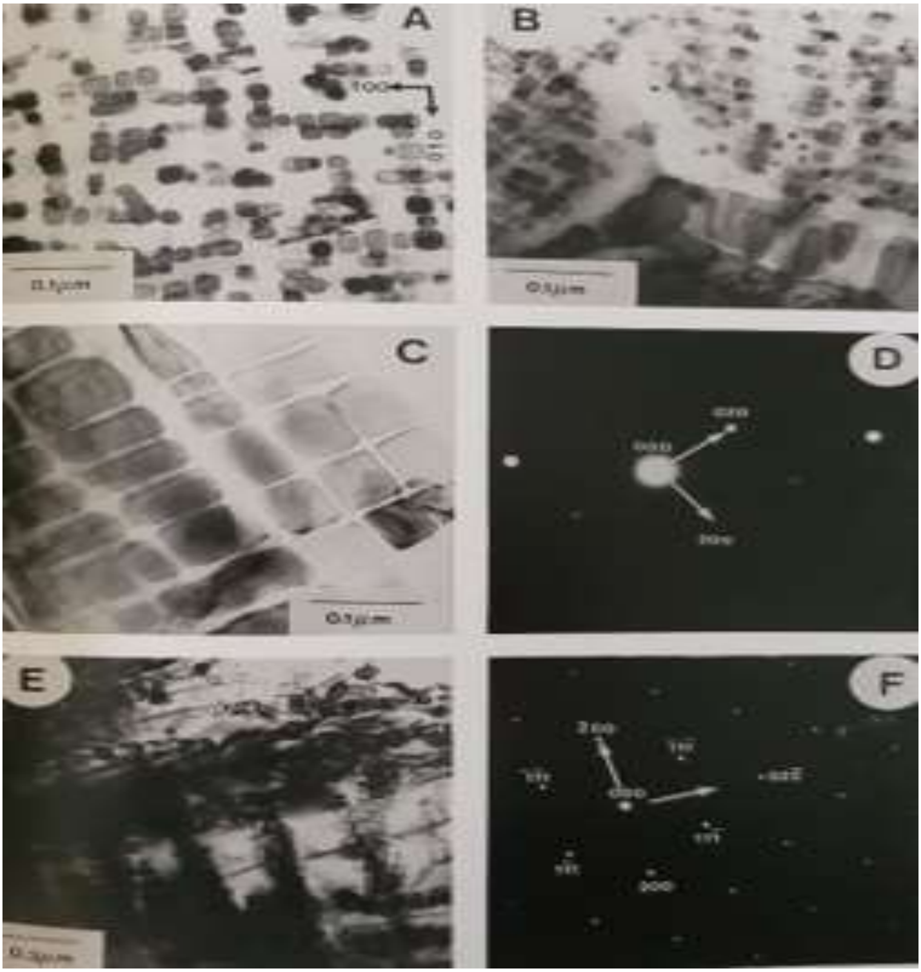


Fig. 3. Later stages of aging for spinodal alloys.

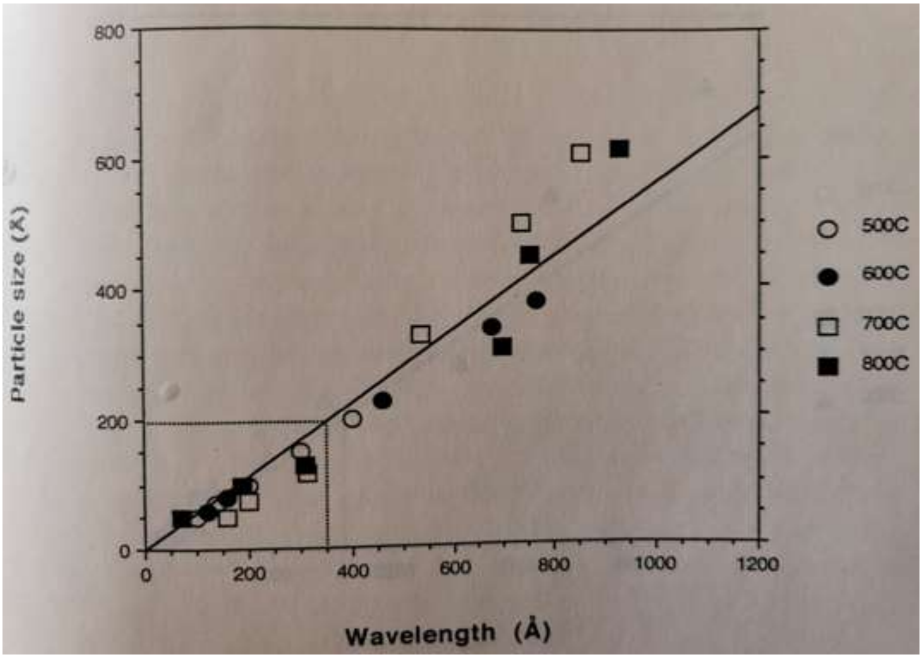


Fig. 4. Cu-5Ni-2.5Cr alloy, particle size vs wavelength.

HARDNESS

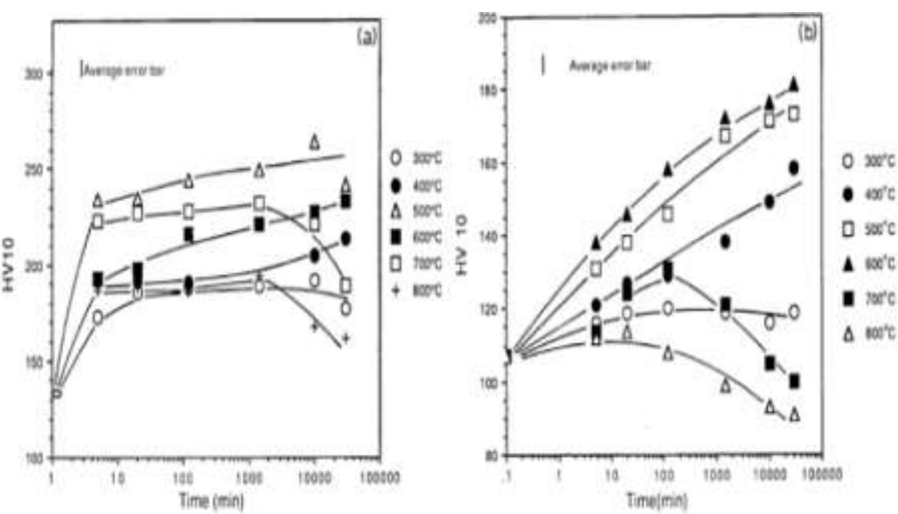


Fig. 5. Hardness and aging time for a) 10Cr and b) 2.5Cr alloy [Findik, 1992:212].

As shown in Figure 5, both Cu-45Ni-10Cr (10Cr) and Cu-30Ni-2.5Cr (2.5Cr) alloys were aged for various durations (5 minutes to 30,000 minutes) between 300 and 800°C. Basically, in the early stages of aging for both alloys, the hardness of the samples increased continuously due to the coherent nature of spinodal decomposition, reaching its highest levels, and then decreased in the later stages of aging due to coarsening of the particles in the structure. However, this coarsening was not observed in the 10Cr alloy even at 30,000 minutes between 300-600°C. Similarly, a similar situation was obtained in the 2.5Cr alloy between 300-600°C, meaning the particles did not coarsen. The highest hardness in the 10Cr alloy was obtained at 260 HV after aging at 500 °C for 10,000 minutes. In the 2.5Cr alloy, the maximum hardness was obtained at 180 HV after aging at 600 °C for 30,000 minutes.

XRD

When spinodal decomposition occurs, edge bands are formed on both sides of the XRD diffraction patterns. In this study, edge bands were observed in all alloys examined (2.5Cr, 5Cr, 10Cr, and 15Cr) in XRD [Findik, 1992:225]. Here, only the edge bands of the 2.5Cr alloy are given in Figure 6. These edge bands were observed at angles of 50° and 52° with values of 2θ. For this, the samples were scanned in XRD between 47° and 54° at a speed of 0.25° min^{-1}, and edge bands were obtained at 200 Bragg peaks. These outcomes are along with preceding studies [Rao et al, 1991:1485].

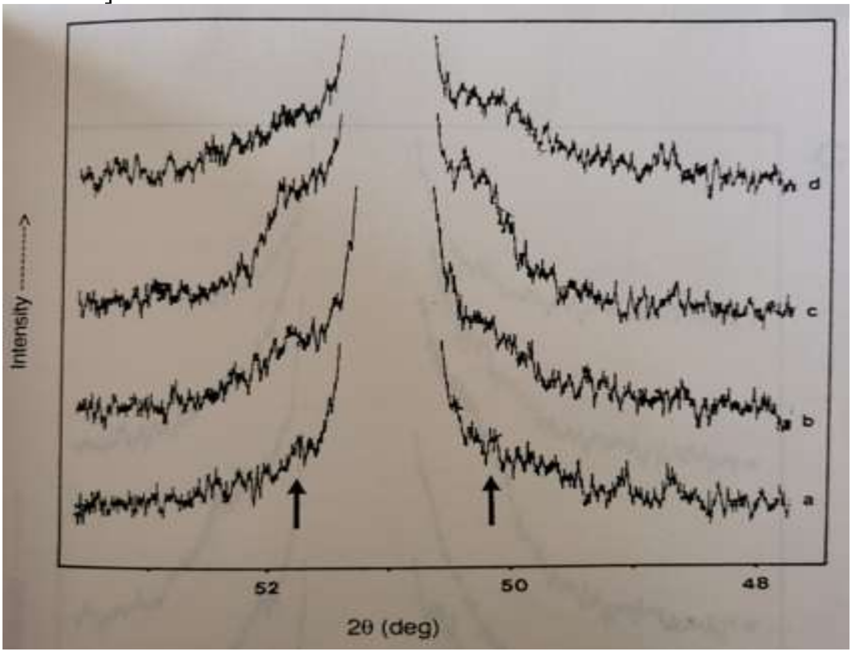


Fig 6. Side-bands in XRD patterns for Cu-30 Ni-2.5Cr spinodal alloy in (200) planes: a) aged 2 h at 300 ºC, b) aged 1 week at 400 ºC, c) aged 1 week at 500 ºC, d) aged 20 minutes at 600 ºC.

CONCLUSIONS

Comprehensive characterization and analysis of Cu-Ni based spinodal alloys leads to the following key conclusions:

a. Spinodal decomposition in Cu-Ni based alloys occurs spontaneously via "uphill diffusion," forming a characteristic, highly periodic, and interconnected microstructure without the need for an activation energy barrier.

b. The modulation wavelength (λ) remains constant in the early stages of aging. As aging progresses, especially at higher temperatures, the wavelength gradually increases as the structure thickens.

c. In Cu-Ni-Cr systems, decomposition products initially appear as small, periodic cubic structures. With prolonged aging, these cubic structures coalesce into long rod arrays aligned in the <100> directions and eventually transform into plates or a chromium-rich third bcc phase.

REFERENCES

Findik, F., *Precipitation reactions in some Cu-NiCr alloys*, PhD Thesis, Imperial College, 1992, England.

Cahn, J. W. (1961). On spinodal decomposition. *Acta Metallurgica*, 9(9), 795–801.

Cahn, J. W. (1962). Coherent Fluctuations and Nucleation in Isotropic Solids. *Acta Metallurgica*, 10(10), 907–913.

Cahn, J. W., & Hilliard, J. E. (1971). *Acta Metallurgica*, 19(2), 151-161.

Lefevre, B. G., D'Annessa, A. T., & Kalish, D. (1978). Spinodal decomposition in a Cu-9Ni-6Sn alloy. *Metallurgical Transactions A*, 9(4), 577–586.

Meijering, J. L. (1957). Calculation of the nickel-chromium-copper phase diagram from binary data, *Acta Met*., 5 (5) 257-264.

Baik, Y. J. Yoon., & D. N. (1990). The migration of chemically induced liquid film and grain boundaries in Mo-Ni, *Acta Met.,* 38(8), 1525-1534.

Davis, R., The deformation of Ni3Al based alloys, PhD Thesis, Imperial College, 1977, England.

Tsubaino, H., Nozato, R., & Yamammoto, A. (1991). Discontinuous precipitation in Al-Zn alloys, *Journal of Materials Science*, 26(11) 2851-2856.

Selcuk, A., Precipitations in some Cu-Be alloys, PhD Thesis, Imperial College, 1989, England.

Kunze, G., & Wincierz, K. (1965). On the determination of lattice distortions and particle sizes from X-ray intensity measurements, *Z. Metallkunde*, 56 (7) 421-430.

Böhm, H. (1961). Zur Ausscheidungshärtung von Aluminium-Magnesium-Legierungen, *Z. Metallkunde*, 52, 564-571.

Rao, P. P. (1986). The effect of ageing on the microstructure and properties of a Cu-15Ni-8Sn alloy, *J. of Mater. Sci.,* 21(11) 3759-3766.

Fournelle, R. A. (1979). The genesis of cellular precipitation in copper-9at.% magnesium alloys, *Acta Met,* 27 (7), 1135-1145.

Frebel, M., & Schenk, J. (1979). Die diskontinuierliche Fällung in Kupfer-Indium-Mischkristallen, *Z. Metallkunde,* 70(4), 230-240.

Hogg, J. M. Novel approaches for the characterization of spinodal alloys, PhD Thesis, Cambridge University, 2025, England.

Migliori, A. & Jarrao, J. (1997). Resonant Ultrasound Spectroscopy: Applications to Physics, Material Measurements and Nondestructive Evaluation. New York: John Wiley and Sons Inc.

Thompson, K., Lawrence, D., Larson, D. J., Olson, J. D., Kelly, T. F., & Gorman, B. (2007). In situ site-specific specimen preparation for atom probe tomography, *Ultramicroscopy*, 107(2-3)131-139.

Findik, F., & Flower H. M. (1992). Microstructure and Hardness Development in Cu-30Ni-2.5Cr and Cu 45Ni-10Cr Spinodal Alloys, *Materials Science and Technology*, 8(3) 197-205.

Findik, F., & Flower H. M. (1993) Morphological Changes and Hardness Evolution in Cu-30Ni-5Cr & Cu-45Ni 54 15Cr Spinodal Alloys, *Mater Sci & Techn*, 9(5) 408-416.

Findik, F. (2012). Improvements in spinodal alloys from past to present, *Mater & Design,* 42(1) 131-146.

Findik, F. (2002). Modulated Structures in Cu-32Ni-3Cr and Cu-46Ni-17Cr Alloys, *Canadian Metallurg Quart,* 41(3) 337-347.

Zhao, J. C. & Notis, M. R. (1998). Spinodal decomposition, ordering transformation, and discontinuous precipitation in a Cu–15Ni–8Sn alloy, *Acta Materialia*, 46(12) 4203–4218.

Laughlin, D. E. & Cahn, J. W. (1975). Spinodal decomposition in age hardening copper titanium alloys, *Acta Metallurgica,* 23(3) 329–339.

Rao, P. P., Agrawal, B. K. & Rao, A. M. (1991). Comparative study of spinodal decomposition in symmetric and asymmetric Cu-Ni-Cr alloys, *J Mater Sci,* 26(6) 1485–1496.

Baker, I., Zheng, R. K., Saxey, D. W., Kuwano, S., Wittmann, M. W. et al., (2009). Microstructural evolution of spinodally formed Fe35Ni15Mn25Al25', *Intermetallics*, 17(11) 886–893.

Waste Management and Decontamination Strategies in CBRN Incidents

Osman KARABAL[1]

Gökhan NUR[2]

1- M.Sc. Student.; İskenderun Technical University, Department of Chemical, Biological, Radiological and Nuclear Threats Management. cm.osmankarabal@gmail.com ORCID No: 0009-0000-7003-5661

2- Prof. Dr.; İskenderun Technical University, Faculty of Engineering and Natural Sciences, Department of Biomedical Engineering. gokhan.nur@iste.edu.tr ORCID No: 0000-0002-5861-8538

ABSTRACT

Chemical, biological, radiological, and nuclear (CBRN) threats encompass multidimensional risk scenarios that directly endanger public health and safety, including industrial accidents, terrorist attacks, and environmental releases. This study aims to examine, from an academic perspective, the classification and management of hazardous wastes generated as a result of CBRN incidents, as well as the decontamination strategies employed in these processes. Within the scope of the study, technical aspects such as the physiological properties of chemical warfare agents, the diversity of biological agents, and the ionizing effects of radioactive elements are addressed. In the waste management process, critical strategies are emphasized, including the reduction of waste at its source (minimization), the categorization of waste according to its type, and the application of subsurface or deep geological disposal methods for radioactive waste, taking into account the concept of half-life. The decontamination process is defined as a life-saving intervention achieved by removing the agent from the environment or neutralizing it through structural degradation. In this context, the operational efficiency of commonly used agents, such as hypochlorite solutions and adsorbent powders, is analyzed. Furthermore, the significance of investigations conducted at CBRN sites is evaluated from the perspectives of forensic science and environmental security, particularly within the framework of chain-of-custody procedures. In conclusion, preparedness against CBRN threats is found to depend on compliance with national and international standards, continuous personnel training, and the integration of advanced technological decontamination systems.

Keywords – CBRN, Hazardous Waste Management, Decontamination Radiological Waste, Biological Risk, Environmental Scurity

INTRODUCTION

Hazardous wastes are substances in solid, liquid, or gaseous forms that may pose lethal or significant risks to humans and the environment. Such wastes may include toxic chemicals, flammable or radioactive materials, and originate from a wide range of sources, including industrial waste from chemical plants or nuclear reactors, agricultural residues, pesticides, fertilizers, medical waste, as well as hazardous household waste such as toxic paints or solvents. Chemical, biological, radiological, and nuclear (CBRN) wastes must be managed and, where possible, recycled safely without causing harm to the environment. Concentration or stabilization of toxic chemicals, radioactive materials, and biological agents prior to final disposal is considered one of the most critical stages of hazardous waste management (Padarev, 2018). Especially chemical, biological, radiological and nuclear (CBRN) wastes are high-risk wastes and need unique methods for storage, transportation, treatment and disposal. Such wastes should be managed following national and international safety standards to minimize unintentional discharges, environmental contamination and consequent public health concerns (Akduman Yiğit, 2024).

The growing global attention to CBRN occurrences has been largely contributed by the rapid industrialization, technical progress, greater utilization of hazardous materials and the appearance of asymmetric threats, e.g. terrorism, throughout the last decades. Major industrial accidents, nuclear disasters and biological epidemics have shown that poor waste management and insufficient decontamination capacity can have serious environmental, economic and societal effects. There are several examples that can be considered hazardous to waste control systems and sustainable decontamination strategies. These include the Chernobyl nuclear disaster, the Fukushima Daiichi accident, and the COVID-19 pandemic (IAEA, 2022; WHO, 2020). CBRN occurrences are inherently different from traditional hazardous material accidents due to the complexity of contaminants, their persistence in the environment, the possibility for mass exposure and the need for multidisciplinary emergency response systems. Effective management of CBRN waste involves coordinated implementation of waste minimization, segregation, transportation, storage, decontamination and final disposal processes with simultaneous assurance of environmental protection, occupational safety and public health security (Kumar et al., 2010). In addition, contemporary CBRN management approaches are placing even greater emphasis on ecologically sustainable technology and quick mobile decontamination systems as well as internationally standardized operational standards. There are several studies on individual aspects of hazardous waste management or decontamination technologies, but the literature is lacking sufficiently integrated review studies, where chemical, biological, radiological and nuclear waste management processes, comparative decontamination approaches and international operational practices are considered together. Therefore, further comprehensive review of existing waste management strategies, decontamination techniques and environmental safety issues is still needed.

In this regard, the present research attempts to explore the classification, treatment and disposal strategies of hazardous wastes created during CBRN incidents and evaluate existing decontamination approaches from environmental, operational and public health perspectives. In addition, the paper examines global practices, technical issues and new trends in CBRN waste management to provide a holistic academic foundation for future research and emergency preparedness planning.

1. Methodology of Literature Review

The present review study was performed by a comprehensive literature search on the waste management and decontamination procedures in chemical, biological, radiological and nuclear (CBRN) accidents. Relevant publications were located by searching major scientific databases such as Web of Science, Scopus, PubMed, and Google Scholar. Literature search was conducted for research published between 2010 and 2026. The search employed keywords such as “CBRN waste management”, “decontamination strategies”, “chemical warfare agents”, “biological hazards”, “radiological waste”, “hazardous waste disposal” and “environmental security”. Priority was given to peer-reviewed journal papers, worldwide recommendations, institutional reports and review studies published in English. They reviewed studies on operating decontamination systems, hazardous waste classification, radiological waste disposal and international response protocols. We removed non-scientific articles, duplicate records and research with insufficient methodological details. The material reviewed was comparatively assessed on the basis of waste management systems, decontamination effectiveness, operational application, environmental impacts, and compliance with international standards. “The results were compiled into a broad review of the state of the art in CBRN waste management and decontamination.

2. Classification of CBRN Waste

Chemical, Biological, Radiological and Nuclear (CBRN) threats are not just restricted to military operations, but also include multidimensional risk scenarios such as terrorist attacks, industrial mishaps, environmental releases and public health catastrophes. In such cases, the primary goal is to safeguard human health and the environment but also to identify the cause, impact and probable perpetrators of the event in a forensically provable way. So, forensic procedures in case of CBRN accidents are different from the normal crime scene investigation protocols, requiring high level of biosafety, multidisciplinary teamwork and modern technical skills (Balbudhe et al., 2025). Chemical Warfare Agent are generally classified in three major groups, depending upon their features and effects. They are classified as solid, liquid, gas and aerosol according to their physical qualities. Based on the physiological effects they are separated into two subgroups: deadly (toxic) agents and incapacitating agents. Lethal chemical warfare agents include nerve agents (sarin, tabun, soman, cyclosarin), blister agents (sulfur mustard, nitrogen mustard, lewisite, phosgene oxime), choking agents (chlorine, phosgene, diphosgene, chloropicrin), and blood agents (hydrogen sulfide, cyanogen chloride, potassium cyanide, sodium cyanide, arsine, carbon monoxide). Incapacitating chemical agents are divided into two categories, incapacitating agents (e.g., quinuclidinyl benzilate, fentanyl derivatives, lysergic acid diethylamide, etc.) and riot control agents (e.g., chlorobenzylidene malononitrile, chloroacetophenone). According to volatility, chemical agents are divided into non-persistent (volatile) – they cause effects for less than 30 min, and persistent-they continue to act for more than 30 min. Biological warfare agents, on the other hand, are used deliberately to contaminate water, food, animals and humans to trigger disease outbreaks and mass poisoning. The concept of bioterrorism and the approaches to prevention and response have therefore emerged.

Biological warfare agents are typically categorized into three main groups based on their origin: bacteria-based agents, virus-based agents and toxin (poison)-based agents. Radioactivity radiation is classified in two types, ionizing radiation and non-ionizing radiation. From this point of view the radiation that poses a radiological threat is the ionizing radiation. The formation mechanisms and interactions of ionizing radiation types, such as alpha particles, beta particles, neutrons, cosmic rays, X-rays and gamma rays, with matter pose major threats to living organisms and the environment (Karayünlü, 2025). Radiological and nuclear waste cause serious damage to human health and to the environment. The prominent instances of the damaging impacts of radiation release include the Chernobyl accident and Fukushima Daiichi nuclear disaster. An accident at the Chernobyl Nuclear Power Plant, in 1986, killed people and made vast areas of land unusable. The 2011 Fukushima tragedy also resulted in massive evacuations and long-lasting environmental pollution. Radiation has very serious impacts on the human health include increasing the risk of cancer and altering the equilibrium of the ecosystem (IAEA, 2009). CBRN waste can trigger catastrophic events either through deliberate action or negligence. Thus, tight standards in waste management can be implemented, inspections can be regularized, educational and awareness campaigns can be promoted and methods can be updated according to environmental conditions. Otherwise, public health, environmental security and societal stability will be severely threatened (WHO, 2016; NRC, 2014).

Table 1: General Characteristics of CBRN Waste Types [Adapted and synthesized from (IAEA, 2022; WHO, 2020; Kumar et al., 2010; OPCW, 2023)].

Waste Type	Main Risk	Environmental Persistence	Transmission Potential	Preferred Management Strategy
Chemical	Toxicity	Moderate–High	Low	Neutralization and isolation
Biological	Infection	Low–Moderate	Very High	Disinfection and containment
Radiological	Ionizing radiation	High	Low	Shielding and controlled storage
Nuclear	Long-term radiation	Very High	Low	Deep geological disposal

2.1. Chemical and Biological Waste Management

Safe management of chemical wastes in accordance with the procedures and principles provided for in the current legislation is a legal obligation. The waste management process includes a sequence of integrated steps such as the waste prevention, the minimization at the source, the reuse where feasible, the classification according to the type and characteristics of the waste, the separate collection and accumulation of each category, the temporary storage in designated facilities, the transportation, the recycling and recovery, the final disposal, the monitoring, control and inspection of potential adverse effects arising from disposal. Waste minimization can be achieved by reducing the quantity of waste generated or by recovering it at the source. Excessive storage of chemical substances not only leads to the increase of waste volume, but also to the increase of the risk of occupational accidents and diseases of occupational pathology. In order to minimize waste effectively, chemical substances that generate less waste should be used, and equipment, production methods and processes that are aimed at minimizing waste generation should be preferred (Kavaklı Vatansever and Ateş, 2018). The generation of biological waste has increased significantly in recent years, thus presenting considerable challenges for its management and disposal. Biological waste is not treated properly in a small fraction of cases. Untreated biological waste is a serious threat to the environment and public health as it can attract pathogenic organisms and other vectors. Recalcitrant biological wastes are converted into value-added products such as syngas, pyro-gas, biochar and biofuels through advanced and complex processes such as pyrolysis, hydrothermal carbonization and transesterification (Saravanan et al., 2023). CBRN wastes may also be classified on the basis of their origin and toxicological characteristics, persistence in the environment, ability to be transmitted and operational response requirements. Long-lived radionuclides and persistent chemical agents can represent environmental hazards for a long time and require long-term monitoring and specialized disposal systems. However, some biological agents may be transmitted rapidly through human or animal populations, even if they do not persist long in the environment, and require rapid containment and biosecurity measures. The classification of CBRN waste is operationally important for the selection of appropriate decontamination procedures, personal protective equipment (PPE) requirements, transportation protocols and ultimate disposal strategies. International organisations such as the World Health Organization (WHO), International Atomic Energy Agency (IAEA) and Organisation for the Prohibition of Chemical Weapons (OPCW) highlight the importance of risk-based classification systems to promote emergency preparedness and environmental protection (IAEA, 2022; OPCW, 2023).

2.2. Management of Radiological Waste

Radioactive wastes are those materials that contain radioactive elements (radionuclides) able to cause damage to biological systems by exposure to ionizing radiation and represent serious environmental risks. Radioactive elements can be generated by nuclear engineering processes in addition to radionuclides that are found in nature. Radionuclides are generally large, unstable atoms that attempt to become stable by releasing atomic energy in the form of radiation. Radioactivity is emitted as positively charged alpha particles, negatively charged beta particles, gamma rays and neutrons. Radioactive decay Radioactive decay happens when radionuclides reach zero radioactivity. The large unstable atoms break up into smaller atoms. Eventually all radioactive substances decay into non-radioactive elements. The progressive decay of radioactivity is expressed by the half-life concept (Solgun et al., 2022). Radioactive waste can be found in many different forms. Spent nuclear fuel, contaminated materials from hospitals, uranium mining residues, scrap metals from nuclear decommissioning, and legacy wastes are examples of common solid radioactive wastes. Liquid radioactive wastes are mostly process effluents from activities like spent fuel reprocessing (SFR), uranium milling and rare earth processing, where chemicals are used to extract particular materials. Other liquid radioactive wastes include sludge from the cleaning and maintenance of reactor coolant systems, leaks from reactor pipelines, water from spent fuel pools and some types of nuclear medical waste such as liquid excreta (Yun et al., 2013). Radioactive waste management is generally categorized into two main groups: wastes for which management strategies have been clearly defined, and wastes for which no universally accepted permanent solution has yet been established. The first category consists of the most voluminous wastes with the lowest levels of radioactivity, for which final disposal facilities are already available. This group covers around 98% of the radioactive waste, for example, low-level waste (LLW), very low-level waste (VLLW), and very short-lived waste (VSLW). In some national regulatory systems, short-lived intermediate-level waste (ILW) is also included in this group. Management of these wastes generally involves procedures such as initial treatment, packaging, and disposal in near-surface facilities, with a depth of up to about 50 meters depending on the concentration of radionuclides. Such methods are used all over the world (Solgun et al., 2022). The second group mainly consists of the most radioactive wastes, which constitute less than 2% of the total volume of waste. This group includes high-level waste (HLW) and long-lived intermediate-level waste (ILW), for which management strategies are still under study and for which a widely accepted permanent disposal solution has not yet been fully implemented (Council, 1999). In some national regulatory regimes, short-lived intermediate-level waste (ILW) is also included in this group. In general, these wastes are managed by processes such as initial treatment, packaging and disposal in near-surface facilities at depths of up to about 50 metres depending on the concentration of radionuclides. These approaches are widely implemented globally (Solgun et al., 2022). The second category mainly contains the most radioactive wastes (less than 2% of the total volume of waste). This group is made up of high level waste (HLW) and long lived intermediate level waste (ILW) whose management strategies are still under research and for which a widely accepted permanent disposal solution has not yet been fully implemented (Council, 1999).

3. Decontamination Strategies

Decontamination is one of the most critical operational stages in the management of chemical, biological, radiological, and nuclear (CBRN) incidents. The purpose of decontamination is to lower the concentrations of hazardous substances to levels that do not threaten human and environmental health. Effective decontamination procedures are vital, especially since emergencies can lead to numerous infections (Kumar et al., 2010; Karayünlü, 2025). The efficiency of decontamination strategies is highly dependent on the physical and chemical nature of the contaminating agent, environmental conditions, duration of exposure and type of contaminated surface. In general, chemical agents need to be rapidly neutralized or physically removed from surfaces. Disinfection procedures are generally required for biological contamination and are directed toward pathogenic microorganisms. Decontamination in radiological and nuclear accidents mainly aims at eliminating radioactive particles and avoiding further spread in the environment (Baş ve Karakurt, 2014). The present methods of decontamination are commonly classified into three types: physical, chemical and biological. The physical decontamination procedures are washing, adsorption, isolation and mechanical removal to reduce surface contamination. Chemical decontamination involves the use of reactive materials that can hydrolyze or oxidize hazardous compounds. Common agents include sodium hypochlorite, hydrogen peroxide based formulations, reactive sorbents and alcohol based disinfectants Biological decontamination methods are based on the application of sterilization procedures or antimicrobial compounds to kill pathogenic agents (Kumar et al., 2010; Baş ve Karakurt, 2024). Rapid decontamination of skin and equipment is a life-saving intervention in a chemical incident. Long-lasting nerve and blister agents can remain on surfaces for a long time, exposing people to a second dose. Working availability and high oxidizing power result in the frequent application of hypochlorite based solutions. Also, typical adsorbent materials used for emergency dry decontamination procedures are fuller’s earth, reactive powders and activated carbon based systems. Recent studies have also shown the increasing importance of environmentally sustainable decontamination agents with reduced corrosive effects and lesser ecological toxicity (Baş ve Karakurt, 2024). Biological decontamination is highly dependent on infection control protocols, especially for highly transmissible pathogens. Health care workers should follow universal precautions in biological events like wearing personal protective equipment (PPE), proper disposal of contaminated waste, surface sterilization and isolation techniques. The COVID-19 pandemic has highlighted the significance of the fast management of biological waste, disinfection possibility, and systems to control healthcare-associated infections worldwide to prevent contamination on a mass scale (European Parliament, 2021). Radiological and nuclear decontamination procedures differ significantly from chemical and biological decontamination procedures, due to the longevity of the radioactive particles and the long-term effects of radiation exposure on the environment. In such events the decontamination is mainly aimed at reducing the rate of radiation dose by removing contaminated materials, shielding, isolating and storing waste in a controlled manner. Common techniques include high pressure washing, chemical fixation, foam based decontamination systems, ion-exchange technologies and surface stripping procedures. To prevent environmental risks from radionuclides, especially long-lived ones, they can be buried very deep in the soil and contained (Malizia et al., 2022). Mobile decontamination platforms, automated spraying systems, sensor-assisted detection technologies and unmanned aerial vehicles are common. This has significantly increased the operational capability of decontamination systems (JPEO-CBRND, 2026). In addition, international organizations and emergency management agencies note that effective decontamination operations depend on multidisciplinary coordination, ongoing training of personnel, standardized operational procedures, and integration of principles of environmental sustainability. Modern CBRN preparedness policies are focusing more and more on long-term environmental recovery, waste minimization and resilient public health infrastructures besides immediate contamination control (European Parliament, 2021; Malizia et al., 2022).

Table 2: Commonly Used Decontamination Agents for Chemical Warfare Agents. Adapted and recreated from the relevant source (Erkekoğlu ve Koçer-Gümüşel, 2018).

Decontamination Agent / Method	Main Function	Operational Notes
Cold water	Initial physical removal of contaminants	Hot water should not be used; scrubbing must be avoided
Soapy, detergent-based, soda-based, or bicarbonate-containing water	Surface cleaning and dilution of contaminants	Commonly used for primary decontamination procedures
Hypochlorite solution (0.5%)	Hydrolysis of nerve agents and oxidation of mustard agents	Household bleach diluted tenfold; applied for 10–15 min followed by thorough rinsing
Fuller's earth	Adsorption of hazardous agents	Common dry decontamination material
Adsorbent powder (magnesium oxide + chlorinated lime)	Chemical adsorption and neutralization	Rapid emergency decontamination support
M291 kit (AMBERGARD XE-555 resin)	Dry adsorption-based decontamination	Portable military/general dry decontamination system
Talc, kaolin, clay, or wheat flour	Emergency adsorption when standard materials are unavailable	Alternative field-expedient materials
2.5% sodium thiosulfate solution	Specific neutralization of mustard agents	Used for targeted chemical agent decontamination
Liquid oil, kerosene, paraffin, or alcohol-based solvents	Dissolution/removal of oily contaminants	Requires careful handling because of flammability and residue risks

4. Environmental and Legal Dimensions

Chemical, biological, radiological and nuclear (CBRN) incidents are considered to be highly complex situations with major environmental and forensic implications in a range of contexts such as natural disasters, industrial accidents, terrorist activities and wartime situations. Such incidents require a multidisciplinary approach not just from emergency response teams, but also from forensic science experts, environmental specialists, public health authorities and legal investigators. Field investigations in the CBRN environment are distinctive from routine crime scene investigations with regard to technical skills, interagency coordination, and safety measures (Özkan et al., 2026). One of the main environmental issues in CBRN disasters is the long-term presence of hazardous pollutants in soil, water resources, air and ecological systems. Chemical agents and radionuclides can remain in the environment for long periods of time, resulting to chronic hazardous exposure, loss of biodiversity, pollution of agriculture and disturbance of the balance of the ecosystem. Similarly, secondary transmission mechanisms of biological contamination may be influential on human and animal populations. Environmental cleanliness and effective waste management are vital (Malizia et al., 2022). The legal framework on CBRN accidents is rather complex. Careful thought is predicated on environmental law, public health legislation and international security treaties. Operational standards and legal frameworks for CBRN have been established by the International Atomic Energy Agency (IAEA), the Organization for the Prohibition of Chemical Weapons (OPCW), the World Health Organization (WHO), and the United Nations Office for Disarmament Affairs (UNODA). These regulations are crucial for the protection of the environment, the integrity of evidence, and public safety during emergency response activities (IAEA, 2022; OPCW, 2023). Unlike a conventional criminal investigation, CBRN related cases involve the deployment of special personal protective equipment (PPE), contamination monitoring systems and decontamination methods to avoid secondary contamination and to maintain the correctness of the forensic results. However, the integrity of the evidence is very essential and can have a harmful influence on the forensic processes and environmental evaluations (Özkan et al., 2026). Effective CBRN crisis management requires the ability to prepare for the environment, coordinate legal issues and communicate the risk quickly, as shown by recent international experiences. The current trends include integrated environmental monitoring systems, international data sharing mechanisms, mobile analytical laboratories and sustainable remediation technologies. In this sense, the reinforcement of legal infrastructures, the improvement of institutional coordination, and the expansion of multidisciplinary training capacities continue to be core priorities to reduce the environmental and social impacts of upcoming CBRN incidents (Malizia et al., 2022).

Figure 1: Evidence Management Process Flow in CBRN Incidents (Özkan et al., 2026).

5. Comparative Evaluation of Decontamination Methods

Decontamination procedures employed in chemical, biological, radiological and nuclear (CBRN) events range significantly depending on the physicochemical properties of the hazardous agents, their environmental persistence, exposure pathways, operational conditions and response goals. Thus, the choice of a suitable decontamination approach is a key factor for the operational success, safety of the environment and fatality reduction. Recent advances in CBRN response tactics have moved decontamination practices from a standardized "one-size-fits-all" application to more flexible based on risk assessment and incident-specific conditions (Haslam et al., 2024). The basic purpose of chemical decontamination methods is to neutralize the hazardous agents in the shortest possible time or to physically remove them from the contaminated surfaces, technical equipment and exposed people. Sodium hypochlorite is a strong oxidizer and is commonly utilized because of its great hydrolytic and oxidative capacity against nerve and blister agents. However, the hypochlorite based systems may have corrosive effects on the sensitive equipments and may generate environmentally dangerous by-products under certain operating situations. Hence, current studies show the growing necessity of less corrosive and ecologically friendly decontamination formulations (Baş ve Karakurt, 2024). Dry decontamination methods (Fuller's earth, activated carbon-based absorbents, reactive sorbent powders and resin-based systems) offer significant operational advantages in emergency response scenarios. These approaches are especially useful in mass casualty events because to their rapid applicability, mobility and reduced water dependence. However, dry decontamination techniques may be less effective for persistent liquid chemical agents that have penetrated porous materials or biological tissue (Kumar et al., 2010). Biological decontamination techniques are very different from chemical decontamination processes since the main goal is the destruction of pathogenic microorganisms and the disruption of transmission. Disinfection processes based on alcohol-based disinfectants, hydrogen peroxide vapor systems, chlorine compounds, UV irradiation and sterilizing technologies are extensively used in healthcare and emergency response contexts. The COVID-19 pandemic proved that biological decontamination success depends not only on disinfectant activity, but also on the health care infrastructure, waste segregation capacity, ventilation systems and compliance with infection-control measures (European Parliament, 2021). Radiological and nuclear cleaning procedures create further operational issues as radioactive particles may stay in environmental systems for lengthy durations. Radiological decontamination differs from chemical neutralization techniques in that the goal is to minimize radiation dose rates by physically removing, containing, shielding, and controlling the isolation of waste. Surface washing, foam-based systems, abrasive removal techniques, chemical fixing and ion-exchange technologies are often employed to treat contaminated infrastructure and equipment. However, the high operational costs, the creation of secondary radioactive waste and the need for long-term environmental monitoring are still key limits for radiological decontamination systems (Malizia et al., 2022).

Comparison between the various decontamination methods does not show any one of them to be fully effective in all operating settings. Integrated and multi-stage decontamination methods are often more successful for complex CBRN occurrences. Operational decision-making should therefore take into account elements such as agent persistence, toxicity level, environmental circumstances, infrastructure sensitivity, responder safety, waste generation potential and ecological sustainability. Recent technological developments are increasingly in support of automated systems, portable mobile platforms, unmanned response technologies and environmentally compatible decontamination agents, to improve operational efficiency and reduce secondary environmental impacts (Haslam et al., 2024; Malizia et al., 2022). In recent years, sustainable decontamination procedures have become a major research concern. Conventional decontamination techniques can create substantial amounts of secondary hazardous waste and consume large amounts of water and chemical resources. Thus, new approaches tend to be more and more oriented towards green decontamination technologies, biodegradable reactive agents, low-water systems and recyclable absorbents. These ecologically friendly measures not only enhance operating safety but also help to long-term ecological protection and environmental resilience in the post-incident recovery procedures (Baş ve Karakurt, 2024).

Table 3: Comparative Evaluation of Common Decontamination Methods (Adapted and synthesized from (Kumar et al., 2010; Haslam et al., 2024; Baş ve Karakurt, 2024; Malizia et al., 2022).

Method	Main Application Area	Advantages	Limitations
Sodium hypochlorite solutions	Chemical agents	Strong oxidizing capacity, rapid action	Corrosive effects and secondary chemical residues
Fuller’s earth / absorbent powders	Chemical contamination	Rapid deployment and low water requirement	Reduced effectiveness on persistent agents
Hydrogen peroxide vapor systems	Biological contamination	Broad antimicrobial activity	Equipment dependency and operational cost
UV irradiation systems	Biological decontamination	Non-chemical sterilization	Limited penetration capacity
Foam-based radiological decontamination	Radiological contamination	Surface adherence and particle removal	Secondary radioactive waste generation
Ion-exchange technologies	Radioactive liquid waste	Effective radionuclide separation	High maintenance and disposal costs

6. International Regulations and Operational Applications

The rise of chemical, biological, radiological and nuclear (CBRN) threats has fueled the development of holistic international standards and operational frameworks to mitigate the environmental damage, safeguard public health and improve emergency response capabilities. Modern CBRN preparedness approaches have progressed beyond the military defence paradigm and are being increasingly integrated into civilian disaster management, environmental protection policies, healthcare systems and critical infrastructure security planning (Malizia et al., 2026). Operational standards and response protocols for CBRN incidents are developed by international organisations (International Atomic Energy Agency (IAEA), Organisation for the Prohibition of Chemical Weapons (OPCW), World Health Organization (WHO), European Union Civil Protection Mechanism (UCPM), NATO, etc.) that are responsible for a CBRN incident response. The above organizations are especially good when it comes to decontamination processes and environmental remediation. IAEA safety standards (IAEA, 2022) are mainly considered in the field of radiation waste and protection, while OPCW standards (OPCW, 2023) are used to identify and dispose of chemical agents. Recent operational experience around the world has underscored the need for integrated and flexible response systems in large scale emergencies. The Fukushima Daiichi nuclear accident underscored the need for coordinated planning for evacuation, separation of radioactive waste and environmental remediation beyond immediate crisis response. Recently, the need for advanced radiation epidemiology models, environmental monitoring technologies and decision-support systems to increase operational readiness in radiological emergencies has been emphasized (Quaranta et al., 2024; Yoneyama et al., 2024). In chemical incidents, operational decontamination protocols are increasingly focusing on rapid casualty management, portable response systems and environmentally sustainable decontamination technologies. The current European and NATO-oriented operational approaches define the decontamination activities in phases of immediate, operational and thorough decontamination according to the degree of contamination and operational objectives. Immediate decontamination includes life-saving measures and mitigation of acute exposures while operational decontamination is focused on continuity of operations and reducing the risk of secondary contamination. Thorough decontamination procedures (eNOTICE-2, 2024) restore long-term operational safety and environmental recovery.

Biological emergencies have also had a significant impact on international arrangements for CBRN preparedness. The COVID-19 pandemic exposed major weaknesses in the medical waste management systems, hospital preparedness, personal protective equipment logistics and biological decontamination capacity of many countries. Thus, international agencies are increasingly focusing on integrated healthcare preparedness, resilient supply chains, rapid isolation capacity, and infection-control protocols as elements of modern biological emergency response systems (European Parliament, 2021). Furthermore, recent studies have emphasized the need for inclusive emergency management practices to consider vulnerable populations, people with disabilities, and electronic assistive devices during mass decontamination processes (Dauvrin et al., 2025). Technological innovation has also had an important role in the operational management of CBRN. We are seeing an increasing integration of artificial intelligence decision support systems, real-time environmental monitoring technologies, autonomous robotic platforms and smart sensor networks into modern decontamination and hazard detection systems. These technologies increase operational efficiency, reduce responder exposure risks and improve situational awareness during complex incidents. Recently, the use of intelligent process-control systems and automated decision-support mechanisms has been shown to considerably speed up and improve CBRN response operations (Taşcioglu et al., 2026). Moreover, sustainability and environmental resilience are increasingly integrated into the international operational frameworks for post-incident recovery strategies. Conventional decontamination methods can produce a large amount of secondary hazardous waste, consume a large amount of water resources, and bring additional environmental loads. Modern regulatory approaches therefore encourage the development of environmentally sustainable decontamination technologies, biodegradable reactive compounds, recyclable absorbent materials and low-resource operational systems. These approaches contribute not only to strengthening emergency response capacity, but also to supporting long-term environmental sustainability and more resilient disaster management policies (Baş ve Karakurt, 2024). International regulations and operational applications demonstrate that successful CBRN management requires multidisciplinary coordination, evidence-based planning, technological integration, and harmonized international cooperation mechanisms. Key priorities for improving global resilience to future CBRN threats are the constant improvement of operational standards, the increase of multinational training programmes and the reinforcement of environmental security infrastructures.

DISCUSSION

Chemical, biological, radiological, and nuclear (CBRN) incidents represent some of the most complex emergencies affecting environmental safety, public health, and national security. CBRN scenarios, unlike traditional hazardous material incidents, require highly specialized response systems because of the diversity of hazardous agents, their persistence in the environment, and the potential for mass casualties and long-term ecological impacts. This review reveals that effective CBRN waste management and decontamination strategies require not only technical intervention capacities, but also integrated multidisciplinary coordination between health care systems, environmental agencies, forensic experts, emergency responders and legal authorities. A consistent theme throughout the literature stresses the absence of a single decontamination strategy suitable for all CBRN scenarios. Chemical, biological, radiological, and nuclear contaminants vary widely in their toxicity, environmental behavior, potential for transmission, and operational response requirements. Therefore, decontamination techniques should be selected in accordance with the specific conditions of the incident, the characteristics of the contamination and the environmental risk profile. In complex incidents combined decontamination systems that include physical removal, chemical neutralisation, biological disinfection and radiological containment are normally more effective than single method approaches (Haslam et al., 2024). In the comparative analysis, it is also found that operational effectiveness is strongly affected by logistical readiness, rapid deployment capability and personnel training. In large-scale incidents there could be an overload of the existing emergency infrastructures, with the absence of adequate mobile decontamination systems, hazardous waste storage facilities and trained personnel. The recent international experiences, especially the COVID-19 pandemic and radiological emergency preparedness exercises in Europe and Asia, have demonstrated that the healthcare capacity, waste segregation systems and institutional coordination are essential to reduce secondary contamination and environmental damage (Yoneyama et al., 2024). Another important issue raised in the literature is the environmental sustainability of the current decontamination techniques. The traditional decontamination techniques usually involve significant secondary hazardous waste, high water consumption and could cause additional chemical pollution to environment. Consequently, recent research has paid more attention to environmentally sustainable decontamination systems including biodegradable reactive compounds, recyclable absorbent materials, low-water consumption technologies and green remediation strategies (Baş ve Karakurt, 2024). Operational planning for long-term recovery following radiological and nuclear incidents (where environmental contamination can be long-lasting, on the order of decades) has become especially important. The review also highlights that international regulatory harmonisation continues to be a major challenge for global CBRN preparedness. The International Atomic Energy Agency (IAEA), Organisation for the Prohibition of Chemical Weapons (OPCW), World Health Organization (WHO) and NATO provide operational guidance and technical frameworks, but implementation capacities vary widely across countries. Adverse effects on operational efficiency in the event of multinational or cross-border emergencies may stem from differences in technological infrastructure, emergency response resources, legal frameworks and institutional coordination mechanisms. In the world, greater international cooperation against CBRN threats is essential. This collaboration is mostly limited to joint exercises and sharing of information (Malizia et al., 2026). A further key dimension of modern CBRN management is technological innovation. Decision support systems based on artificial intelligence, autonomous robotic platforms, smart environmental monitoring networks and advanced sensor technologies are more and more integrated in emergency response operations. These systems can contribute to faster hazard detection, reduce exposure of responders, improve resource allocation and support real time operational decision making. Technological advancements provide important benefits, but the high operating costs of emergency response systems create serious challenges for sustainability in developing countries (Taşcioglu et al., 2026). Maintaining the integrity of evidence in contaminated environments remains one of the biggest operational challenges in forensic and legal settings. The use of chain-of-custody procedures in a hazardous environment requires very specific response procedures, continuous monitoring of contamination and effective multidisciplinary coordination in order to ensure personnel security and legal validity of forensic investigations. Failure to preserve evidence or control contamination can compromise criminal investigation, environmental analysis, and judicial accountability processes (Özkan et al., 2026). Recent scientific literature indicates that future CBRN preparedness policies should pay more attention to integrated environmental management approaches, sustainable decontamination solutions, advanced operational coordination systems and continuous multidisciplinary education and training programs. Building stronger international operational partnerships, fostering environmentally sustainable recovery planning and increasing institutional resilience will be critical to reduce long-term environmental and societal impacts of future CBRN events.

CONCLUSION

CBRN incidents should not be considered solely within the framework of defence or military threats but as complex emergency situations that may have a global impact on the environment, society, the economy and public health. The effects of CBRN incidents are not limited to the initial contamination site, but can continue over a long time by means of environmental and biological processes (Malizia et al., 2026), as shown by industrial accidents, terrorist acts, environmental spills, epidemics and nuclear disasters. CBRN preparedness capacity is thus today a key element of environmental security and policies related to disaster management. This review shows that the CBRN waste management and decontamination processes are very different from the classical hazardous waste management systems, as a result of the high risk, persistent and complex nature of the agents. An effective response to CBRN incidents requires a holistic approach to operational planning, encompassing hazardous waste classification, rapid decontamination practices, environmental monitoring activities, preservation of forensic evidence, and long-term remediation efforts. CBRN incidents need interagency cooperation (health organizations, environmental organizations, legal authorities, emergency response teams) to minimize environmental and human harm (Haslam et al., 2024). Comparative assessments of decontamination methods show that no single decontamination method provides sufficient effectiveness against all incident scenarios. Hence, it is understood that the combination of physical, chemical, biological and radiological decontamination techniques in incident based and integrated strategies produces more effective results. The results also indicate that due to the environmental burden of traditional decontamination practices, sustainable decontamination technologies, eco-friendly reactive compounds, and low resource demanding operating systems are becoming increasingly important (Baş ve Karakurt, 2024).

This study is significant as it shows how international cooperation can be utilized to reduce CBRN threats. This area is contributed to by many institutions (International Atomic Energy Agency-IAEA, Organization for the Prohibition of Chemical Weapons-OPCW, World Health Organization-WHO, NATO) through the standards, legislation and cooperation they have developed (IAEA, 2022; OPCW, 2023). Differences in technological infrastructure and institutional capacity between countries remain a problem, all these positive approaches notwithstanding. Next generation technologies like AI-powered decision systems, autonomous robotic platforms, mobile decontamination units and real-time environmental monitoring technologies provide significant advantages to improve operational efficiency and ensure the safety of response personnel. The assumption that access to hazardous materials will be easier in the future is considered a factor that could facilitate the emergence of CBRN threats. Therefore, it is crucial to anticipate and respond to risks in advance, to provide continuous training for personnel, and to reduce the extent of the danger created by this gap through drills (Taşcioglu et al., 2026). In conclusion, CBRN waste management needs to be considered a fundamental environmental and public health policy issue to protect sustainable environmental safety and social stability. Strengthening international cooperation, enhancing multidisciplinary operational coordination, promoting environmentally sound decontamination technologies and continuously updating regulatory frameworks to reflect advances in technology will be decisive to mitigate the environmental and societal impacts of future CBRN incidents.

REFERENCE

Akduman Yiğit, O. (2024). Sürdürülebilirlik Odaklı Bütünleşik Atık ve Afet Yönetimi. Gazi Kitabevi, Bölüm 12, ISBN: 978-625-365-785-7, pp: 315-344.

Baş, K., & Karakurt, S. (2024). Decontamination solutions and techniques for chemical, biological, radiological and nuclear (CBRN) incidents. The Journal of Defense Science, 20(1), 29-48. https://doi.org/10.17134/khosbd.1360355

Balbudhe, M. S., Mahajan, A. A., Dama, S. M., & Bhowate, R. R. (2025). Crime scene reconstruction: A scoping review. Egyptian Journal of Forensic Sciences, 15, 79. https://doi.org/10.1186/s41935-025-00495-5

Council NR. (1999). Disposition of high-level radioactive waste through geological isolation: development, current status, and technical and policy challenges. The National Academies Press, Washington, DC.

Dauvrin, M., Vybornova, O., & Gala J. L. (2025). Inclusive emergency management after CBRN incidents: adapting decontamination for individuals with disabilities and assistive devices. Front. Disaster Emerg. Med. 3:1614121. doi: 10.3389/femer.2025.1614121

eNOTICE-2. (2024). Decontamination of assistive devices in CBRN incidents. European Union Civil Protection Knowledge Network. https:// https://civil-protection-knowledge-network.europa.eu/media/decontamination-assistive-devices-0

European Parliament. (2021). EU preparedness and responses to chemical, biological, radiological and nuclear (CBRN) threats. Directorate-General for External Policies of the Union. https://www.europarl.europa.eu/thinktank/en/document/EXPO_STU(2021)653645. https://doi.org/10.2861/857738

Haslam, J. D., Russell, P., Hill, S., Emmett, S. R., & Blain, P. G. (2022). Chemical, biological, radiological, and nuclear mass casualty medicine: a review of lessons from the Salisbury and Amesbury Novichok nerve agent incidents. Br J Anaesth. 128(2):e200-e205. https://doi.org/ 10.1016/j.bja.2021.10.008

International Atomic Energy Agency (IAEA). (2022). Preparedness and response for a nuclear or radiological emergency (IAEA Safety Standards Series No. GSR Part 7). International Atomic Energy Agency. https://www.iaea.org/publications

International Atomic Energy Agency (IAEA). (2009). Classification of radioactive waste (General Safety Guide No. GSG-1). IAEA. https://www.iaea.org/publications/8154/classification-of-radioactive-waste

Kavaklı Vatansever, B., & Ateş, A. (2018). Ar-Ge kuruluşunda kimyasal maddelerin sınıflandırılması, etiketlenmesi, ambalajlanması, depolanması, taşınması ve oluşan kimyasal atıkların bertarafı. Sakarya Üniversitesi Fen Bilimleri Dergisi, 22 (2), 159-173. https://doi.org/10.16984/saufenbilder.292664

Kumar, V., Goel, R., Chawla, R., Silambarasan, M., & Sharma, R. K. (2010). Chemical, biological, radiological, and nuclear decontamination: Recent trends and future perspective. Journal of Pharmacy & Bioallied Sciences, 2(3), 220-238. https://doi.org/10.4103/0975-7406.68505

Karayünlü, G. (2025). Kimyasal Biyolojik Radyoaktif Nükleer Tehlikelere Yönelik Yerel Yönetimlerin Görev ve Hazırlıklarının Değerlendirilmesi. The Bulletin of Legal Medicine, 30(1), 45-57. https://doi.org/10.17986/blm.1739

Organisation for the Prohibition of Chemical Weapons (OPCW). (2023). Guidelines for chemical incident response and decontamination operations. OPCW. https://www.opcw.org

Özkan, A., Sakar, A. U., Özkan, V., & Nur, G. (2026). KBRN Olaylarında Delil Toplama ve Numune Yönetimi Süreçleri Üzerine Literatür İncelemesi. Caucasian Journal of Science, 12(2), 139-171. https://doi.org/10.48138/cjo.1830192

NRC (Nuclear Regulatory Commission). (2014). Low-Level Waste Disposal Regulations, Guidance, and Communications. Retrieved from https://www.nrc.gov

Malizia, A., D'Arienzo, M., Contessa, G. M., D'Errico, F., Souza Lalic, S., Duschek, F., Vasiliou, V., Hooker, A. M., & Gaudio, P. (2026). CBRNE events: prevention, mitigation, consequences and recovery. Eur. Phys. J. Plus 141, 137 (2026). https://doi.org/10.1140/epjp/s13360-026-07405-1

Malizia, A., Chatterjee, P. & D'Arienzo, M. New technologies for detection, protection, decontamination, and developments of the decision support systems in case of CBRNe events: editorial. Eur. Phys. J. Plus 137, 1192 (2022). https://doi.org/10.1140/epjp/s13360-022-03408-w

Padarev, N. I. (2018). Best practices in CBRN waste management in military operations. International Scientific Journal Security & Future, 2(3), 130-133.

Erkekoğlu, P., & Koçer-Gümüşel, B. (2018). Kimyasal Savaş Ajanları: Tarihçeleri, Toksisiteleri, Saptanmaları ve Hazırlıklı Olma. Hacettepe University Journal of the Faculty of Pharmacy, 38(1), 24-38.

Saravanan, A., Karishma, S., Senthil, K. P., & Rangasamy, G. (2023). A Review on Regeneration of Biowaste into Bio-Products and Bioenergy: Life Cycle Assessment and Circular Economy. Fuel, Vol. 338, 127221. https://doi.org/10.1016/j.fuel.2022.127221

Solgun, E., Baran, H., & Ünvar, S. (2022). Nükleer Enerji Sistemlerinde Radyoaktif Atıklar Yönetimi ve Atık Depolama Yöntemleri. In book: Farkli Yaklaşimlarla Enerji Kaynaklari, Orient Yayınları, pp.145-162.

WHO (World Health Organization) (2020a). Water, sanitation, hygiene, and waste management for SARS-CoV-2, the virus that causes COVID-19. https://www.who.int/publications

WHO (World Health Organization) (2016). Radiation: Protecting Health in Nuclear Emergencies. https://www.cdc.gov/radiation-emergencies/ safety/index.html (son erişim tarihi 3.11.2024).

Joint Program Executive Office for Chemical, Biological, Radiological and Nuclear Defense (JPEO-CBRND). (2026). Portable decontamination system reduces logistical burden for CBRN operations. https://www.jpeocbrnd.osd.mil

Taşcioglu, F., Konukbay, A., Karabacak, H., & Koluman, A. (2026). Advancements in Process Control Methodologies for CBRN Decontamination Techniques: A Review. Resilience, 9(2), 357-370. https://doi.org/10.32569/resilience.1530724

Quaranta, R., Ludovici, G. M., Romano, L., Manenti, G., & Malizia, A. (2024). A rapid radiation epidemiology tool for the analysis of the propagation of radiation following a radiological dispersal device explosion. Eur. Phys. J. Plus 139, 1110 (2024). https://doi.org/10.1140/epjp/s13360-024-05905-6

Yoneyama, T., Koyama, T., Miyanoue, K., Goto, H., & Nishioka, M. (2024). Utilization and development of evacuation time estimates in wide-area evacuation planning for nuclear disasters in Japan. Eur. Phys. J. Plus 139, 784 (2024). https://doi.org/10.1140/epjp/s13360-024-05587-0

Yun, J. I., Jeong, Y. H., & Kim, J. H. (2013). Republic of Korea: experience of radioactive waste (raw) management and contaminated site clean-up. Editor(s): William E. Lee, Michael I. Ojovan, Carol M. Jantzen, In Woodhead Publishing Series in Energy, Radioactive Waste Management and Contaminated Site Clean-Up, Woodhead Publishing, 673-696. https://doi.org/10.1533/9780857097446.2.673

Methanol-Fueled Internal Combustion Engines: Clean Combustion, Low Carbon, and Sustainable Future Perspectives

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

Environmental impacts originating from fossil fuels, greenhouse gas emissions, and energy supply security concerns have increasingly highlighted the importance of using low-carbon alternative fuels in internal combustion engines. In this context, methanol has emerged as a promising fuel option for both diesel and gasoline engines due to its high oxygen content, high octane number, wide flammability limits, high latent heat of vaporization, and potential for renewable production. This review comprehensively evaluates the use of methanol in internal combustion engines in terms of fuel properties, application methods, combustion characteristics, performance, emissions, knock behavior, after-treatment requirements, and sustainability aspects. In diesel engines, methanol is applied through methanol–diesel blends, port fuel injection, methanol/diesel dual-fuel combustion, pilot diesel ignition, and dual direct injection systems. In these applications, methanol generally demonstrates the potential to reduce soot, smoke, particulate matter, and NOx emissions; however, its low cetane number, combustion instability, and increased HC, CO, unburned methanol, and formaldehyde emissions constitute significant limitations. In gasoline engines, methanol is utilized as a neat fuel, in methanol–gasoline blends, and through strategies such as direct injection, lean combustion, high compression ratios, and EGR/air dilution. In these engines, methanol can improve thermal efficiency, provide more stable combustion, and reduce certain emissions owing to its high knock resistance and fast-burning characteristics. Nevertheless, increased fuel consumption associated with its low lower heating value, cold-start difficulties, unregulated emissions, and material compatibility issues remain major challenges to be addressed. Overall, when combined with appropriate engine calibration, injection strategies, and blend ratios, methanol can be considered a strong low-carbon transitional fuel and a long-term sustainable alternative fuel for existing internal combustion engine technologies.

Keywords – Methanol, internal combustion engine, diesel engine, gasoline engine, combustion, sustainable fuels.

INTRODUCTION

Internal combustion engines continue to remain one of the fundamental components of the global energy system in transportation, agriculture, marine applications, generator systems, and heavy-duty operations. Although electric vehicles are rapidly becoming widespread, particularly in passenger transportation, the complete phase-out of internal combustion engines in the short and medium term is not expected due to the large size of the existing vehicle fleet, the high energy density requirements of heavy-duty and long-range transportation, the widespread fuel-refueling infrastructure, and the limitations of battery-based solutions in certain sectors. Therefore, not only the advancement of engine technologies but also the reduction of the carbon intensity of the fuels used is of critical importance for achieving sustainable transportation goals. The significant contribution of the transportation sector to global CO_2 emissions has increased interest in low-carbon, oxygenated, and renewable alternative fuels that can replace or be used alongside conventional fossil-based gasoline and diesel fuels (International Energy Agency, 2026).

In the search for alternative fuels, alcohol-based fuels have gained prominence due to their high oxygen content, cleaner combustion tendencies, potential for production from renewable resources, and adaptability to existing engine technologies at certain blending ratios. Among these fuels, methanol stands out as a particularly attractive option because of its simple molecular structure, ease of storage in liquid form, broad range of production pathways, and distinctive effects on engine combustion characteristics. Although methanol can be produced from fossil resources, it can also be synthesized from biomass, waste streams, renewable hydrogen, and captured CO_2. This makes methanol not only a fuel additive but also a potential renewable energy carrier that can be integrated into the carbon cycle (International Renewable Energy Agency & Methanol Institute, 2021, IEA Advanced Motor Fuels TCP, 2026).

The primary reason for the growing interest in methanol for internal combustion engine applications lies in the direct influence of its physical and chemical properties on the combustion process. Methanol's high octane number, high latent heat of vaporization, and oxygenated structure significantly enhance knock resistance, particularly in spark-ignition engines, thereby enabling advantages such as higher compression ratios, advanced ignition timing, greater charge cooling, and potentially improved thermal efficiency. The high octane rating of neat methanol makes it a highly knock-resistant fuel for gasoline engines, whereas its low lower heating value requires a higher fuel flow rate to achieve the same power output (IEA Advanced Motor Fuels TCP, 2026, Verhelst et al., 2019). Nevertheless, its high heat of vaporization and oxygen content can provide substantial benefits in terms of efficiency and emissions when combined with appropriate engine calibration.

However, the use of methanol in engines is not without challenges. Its low volumetric energy density may result in reduced driving range, its very low cetane number causes difficulties in autoignition in diesel engines, its complete miscibility with water can create fuel stability problems, and its corrosive nature may lead to compatibility issues with fuel system materials. In addition, under incomplete combustion conditions, the formation of unburned methanol, formaldehyde, and other oxygenated intermediate species constitutes an important environmental and toxicological concern (IEA Advanced Motor Fuels TCP, 2026, Verhelst et al., 2019). Therefore, the effective utilization of methanol in engines is not limited to determining blend ratios alone, but must also be evaluated together with multidimensional parameters such as injection strategy, ignition timing, compression ratio, exhaust gas recirculation, catalytic after-treatment, and material compatibility.

In gasoline, or spark-ignition (SI), engines, methanol is generally used in low-percentage methanol–gasoline blends, although some studies have investigated high-blend ratios and even neat methanol applications. Low-level blends are more easily compatible with existing engines and fuel infrastructures, whereas high methanol concentrations require more extensive engine optimization. Methanol's high octane number suppresses knock tendency, thereby enabling increased compression ratios and more aggressive ignition strategies. Furthermore, its high latent heat of vaporization produces a cooling effect in the intake air, which can improve cylinder charging and extend the knock limit, particularly in turbocharged engines (Verhelst et al., 2019, Wei et al., 2008). These characteristics position methanol not only as an emission-reducing additive but also as a strategic fuel for the development of high-efficiency spark-ignition engines.

In diesel, or compression-ignition (CI), engines, the use of methanol presents a more complex scenario. The main reason is methanol's extremely low cetane number, which prevents reliable autoignition under conventional diesel engine conditions. Therefore, methanol is generally applied in diesel engines through diesel–methanol blends, the use of additives or co-solvents, intake manifold injection, fumigation techniques, pilot diesel ignition dual-fuel operation, or dual direct injection systems (Yao et al., 2017, Yin et al., 2023a). In recent years, premixed compression ignition, low-temperature combustion, and dual-fuel combustion strategies have also become important research areas for methanol–diesel systems. In these applications, diesel fuel typically acts as the pilot fuel initiating ignition, while methanol serves as the low-carbon, oxygenated primary energy carrier supplied to the cylinder. In this way, reductions in soot and particulate matter formation, combustion temperature control, and improvement of the NO_x–PM trade-off are targeted.

However, combustion instability at low loads, increases in CO and HC emissions, ignition delay, and unburned methanol emissions remain major challenges to be addressed (Yao et al., 2017, Yin et al., 2023a).

From the perspective of environmental sustainability and net-zero emission targets, the use of methanol should be evaluated not only in terms of tailpipe emissions but also through its entire life-cycle impacts, from production to end use. Since methanol combustion still releases CO_2 from the exhaust, methanol cannot be directly classified as a “zero tailpipe carbon emission” fuel. However, when produced using biomass, waste resources, green hydrogen generated from renewable electricity, and captured CO_2, methanol can play a significant role in low-carbon or net-zero fuel strategies by enabling carbon recycling within the energy system (Sollai et al., 2023, Güleroğlu & Yumurtacı, 2025). In this respect, renewable methanol can function as a transition fuel for reducing the carbon intensity of existing internal combustion engine infrastructures without requiring their complete replacement. Moreover, the absence of carbon–carbon bonds in the methanol molecule and its inherent oxygen content contribute to the reduction of soot, particulate matter, and certain incomplete combustion products in both diesel and gasoline engines (Verhelst et al., 2019, Yao et al., 2017). Nevertheless, the environmental benefits of methanol strongly depend on its production pathway. Fossil-derived methanol provides only limited greenhouse gas reduction benefits, whereas green methanol produced using renewable energy offers substantially lower life-cycle environmental impacts (Sollai et al., 2023, Güleroğlu & Yumurtacı, 2025, Pan et al., 2015). Consequently, the use of methanol in internal combustion engines should be assessed not only in terms of performance and exhaust emissions, but also from the perspectives of renewable fuel production, carbon capture and utilization technologies, energy storage, and the circular carbon economy.

This study aims to evaluate the use of methanol in internal combustion engines not merely as a general alternative fuel option, but through a holistic perspective considering engine type, fuel application method, combustion characteristics, performance outcomes, and emission behavior. Although numerous studies in the literature have separately investigated the effects of methanol in gasoline and diesel engines, many of these studies focus on specific blend ratios, limited operating conditions, or isolated emission parameters. The distinctive contribution of this review lies in its comparative evaluation of methanol utilization pathways in both spark-ignition and compression-ignition engines; its systematic analysis of the relationships between fuel physicochemical properties and combustion, performance, and emission characteristics; and its identification of research gaps related to future engine optimization, renewable methanol production, fuel infrastructure, and emission control strategies. In this way, the study provides a comprehensive assessment of methanol's potential as a transitional fuel, a high-efficiency engine fuel, and a renewable carbon carrier within existing internal combustion engine technologies.

MATERIALS AND METHOD

Methanol Application Methods in Diesel Engines

One of the most fundamental methods of methanol application in diesel engines is the use of methanol–diesel fuel blends. In this approach, methanol is mixed with diesel fuel at specific volumetric or mass ratios and supplied to the engine as a single blended fuel. The prepared fuel mixture is delivered through the existing diesel fuel system and injected into the cylinder via the conventional diesel injection system. While the basic fuel system of the engine is largely preserved, only the chemical composition of the fuel is altered. Due to the limited solubility of methanol in diesel fuel, some studies employ co-alcohol additives such as n-pentanol to prevent phase separation. This method is widely preferred for investigating the effects of different methanol blending ratios on combustion characteristics, engine performance, and exhaust emissions (Lapuerta et al., 2008).

Another widely used approach for methanol utilization in diesel engines is the port-injection dual-fuel system. In this method, methanol is introduced into the intake port or intake manifold as the primary fuel, while diesel fuel is directly injected into the cylinder through the conventional diesel injector. As the methanol–air mixture enters the cylinder, diesel fuel injected near the end of the compression stroke acts as a pilot fuel to initiate combustion. In the literature, these systems are described by various terms such as "dual-fuel operation," "pilot diesel ignition methanol combustion," or "diesel/methanol compound combustion systems," although they are all based on the same fundamental operating principle. In such systems, the methanol energy fraction, pilot diesel quantity, injection timing, and excess air ratio are the main control parameters (Ma et al., 2026, Zhu et al., 2023, Hassan & Al-Abboodi, 2025).

Another advanced application method involves dual direct injection methanol/diesel systems. In this configuration, separate injectors are employed for methanol and diesel fuel, and both fuels are directly injected into the cylinder. Diesel fuel serves as the high-reactivity pilot fuel, while methanol acts as the low-reactivity primary fuel. These systems allow independent control of methanol injection timing, diesel injection timing, injection pressure, injection duration, and energy substitution ratio. As a result, different reactivity zones can be established inside the cylinder, enabling optimization of the combustion process and more precise control over engine performance and emission characteristics (Dong et al., 2020, Karvounis & Theotokatos, 2025).

Methanol applications are also increasingly utilized in marine and heavy-duty engines. In these engines, the applied systems are generally based on high-pressure direct methanol injection combined with pilot diesel ignition. Methanol is used as the primary fuel, while a small quantity of diesel fuel is injected into the cylinder to initiate ignition. In some advanced applications, triple direct injection systems employing multiple pilot diesel injectors together with a methanol injector are preferred. In addition, certain marine engines are converted into spark-ignition configurations, where methanol is introduced through intake port fuel injection (PFI-SI systems). These applications are particularly important for reducing carbon emissions and expanding the use of alternative fuels in large-displacement engines (Guo et al., 2025, MAN Energy Solutions, 2021, Sun et al., 2025).

Methanol Application Methods in Gasoline Engines

The most common method of methanol utilization in gasoline engines is the use of methanol–gasoline blends. In this approach, methanol is mixed with gasoline at specified volumetric ratios to obtain a single-phase fuel blend. Different methanol blending ratios such as M5, M10, M15, M20, M45, and M85 are widely reported in the literature. The prepared fuel blend is supplied to the engine through the existing port fuel injection system or carburetor system. In these applications, the fundamental mechanical structure of the engine is generally not modified; instead, operational parameters such as air–fuel ratio, ignition timing, and compression ratio are adjusted to investigate engine performance, combustion characteristics, and exhaust emissions (You, 1983, Elfasakhany, 2014).

Another major application method in gasoline engines is the use of neat methanol fuel. In this configuration, the engine operates exclusively on methanol. Due to methanol's lower heating value and its stoichiometric air–fuel ratio being different from gasoline, the fuel system must be modified accordingly. In carbureted engines, the fuel passage area is enlarged, while in fuel-injected engines the injection duration, injected fuel quantity, and electronic control parameters are recalibrated for methanol operation.

Combustion is initiated by a spark plug, and the engine operates according to the conventional spark-ignition principle. Moreover, many studies exploit methanol's high knock resistance by operating the engine at higher compression ratios in order to improve thermal efficiency (Turner et al., 2018, Reitz & Duraisamy, 2015, Boggavarapu et al., 2025a).

Another important method is the application of direct methanol injection in gasoline engines. In this method, methanol is injected directly into the cylinder rather than into the intake port. Dedicated direct injection injectors are employed, or the existing injection system is modified to accommodate methanol injection. In some configurations, gasoline is supplied through the intake port while methanol is directly injected into the cylinder, thereby enabling dual-fuel spark-ignition operation. In these systems, injection pressure, injection timing, injection duration, methanol ratio, and ignition timing constitute the principal control parameters. Methanol injection may occur during either the intake stroke or the compression stroke, with the aim of controlling in-cylinder temperature and suppressing knock tendency (Shu et al., 2024).

Another widely investigated application in methanol-fueled spark-ignition engines involves lean-burn and diluted combustion conditions. In this method, the engine operates with an excess air ratio greater than stoichiometric conditions or employs exhaust gas recirculation (EGR). In lean combustion studies, increasing the lambda value aims to reduce fuel consumption and emissions, whereas EGR applications seek to lower combustion temperatures and reduce NO_x emissions. In these systems, air–fuel ratio, EGR rate, ignition timing, injection parameters, and engine load are simultaneously controlled. Particularly in marine-type PFI-SI methanol engines, ignition timing and air–fuel ratio are optimized to prevent knock formation (Feng et al., 2013).

As illustrated in this classification, the method of methanol application varies considerably depending on engine type. In diesel engines, methanol implementation generally relies on pilot diesel ignition, dual-fuel combustion, and direct injection systems, whereas in spark-ignition engines methanol is utilized directly as the primary fuel, in gasoline blends, through direct injection, or under diluted combustion strategies. Therefore, evaluation of methanol applications should consider not only the type of fuel used, but also the fuel delivery method, injection location, ignition system, and control parameters together.

RESULTS AND DISCUSSION

Diesel Engines

In the study by Li et al. (2021), a single-channel diesel/methanol dual-fuel injector for direct methanol injection with pilot diesel operation in diesel engines was numerically investigated. The effects of residual methanol volume, pilot diesel quantity, nozzle diameter, and injection pressure on combustion, fuel economy, knock tendency, and emissions were evaluated. The results indicated that increasing the residual methanol amount delayed diesel ignition and increased the tendency for knock. The optimum pilot diesel ratio was determined to be approximately 10% of the total fuel energy. In addition, the combination of a 0.36 mm nozzle diameter and 27.4 MPa injection pressure provided the best overall performance and produced results close to those of a dual-channel injector. However, the single-channel injector was found to be more prone to knock, emphasizing that injection timing must be carefully optimized.

In the study by Wang et al. (2022), diesel–methanol diffusion combustion was investigated in a high-speed light-duty diesel engine using a dual direct injection system. Conventional diesel combustion and diesel–methanol diffusion combustion were compared, and the effects of different methanol substitution ratios on combustion, emissions, and energy distribution were analysed. The results showed that the methanol substitution ratio could be increased up to 87% under full-load conditions. Compared with conventional diesel combustion, diesel–methanol diffusion combustion significantly reduced NOx and soot emissions, with soot emissions approaching nearly zero at high load. Although CO and THC emissions increased at low load, these adverse effects diminished with increasing load, resulting in improved emission characteristics at full load. Furthermore, indicated thermal efficiency increased by approximately 10% at full load, while CO_2 emissions decreased by about 19%. Energy distribution analysis revealed that methanol utilization reduced heat transfer and exhaust losses, thereby contributing to higher efficiency.

In the study by Kumar et al. (2022), methanol injection/fumigation through the intake manifold was experimentally investigated in a single-cylinder diesel engine. The methanol energy substitution ratio was maintained at 50% (M50) under all operating conditions, and the engine was tested at 1500 rpm under different loads and two intake air temperatures (40 °C and 80 °C). The results demonstrated that heating the intake air reduced the cooling effect caused by methanol evaporation and improved combustion performance, particularly under the M50_80 °C condition, where in-cylinder pressure and brake thermal efficiency increased. Methanol fumigation generally reduced NO and smoke/soot emissions compared with neat diesel fuel; however, HC and CO emissions increased because of lower in-cylinder temperatures and incomplete combustion. Nevertheless, higher intake air temperature decreased HC and CO emissions, contributing to a more stable and efficient implementation of methanol fumigation. The study concluded that methanol is a feasible alternative capable of significantly reducing diesel consumption in diesel engines without major hardware modifications, although intake air temperature must be carefully optimized.

In the study by Tao et al. (2022), the effects of pilot diesel injection strategies on combustion and emissions in a methanol–diesel dual-fuel engine were experimentally investigated. Experiments were conducted on a single-cylinder diesel engine equipped with intake-port methanol injection and a common-rail diesel injection system. Pilot injection timing, pilot diesel mass ratio, and methanol substitution ratio were evaluated. The results showed that pilot injection advanced ignition timing and increased the heat release rate. Advancing the pilot injection timing increased ignition delay while shortening combustion duration, leading to varying trends in thermal efficiency. Increasing the pilot diesel ratio raised the maximum in-cylinder pressure and heat release rate, while ignition timing and CA50 shifted earlier. In terms of emissions, increasing the pilot diesel amount increased NOx emissions but reduced smoke emissions. When the methanol substitution ratio was increased from 20% to 40%, the combined use of advanced pilot injection timing, higher pilot diesel ratio, and exhaust gas recirculation (EGR) improved indicated thermal efficiency by 1.9 percentage points, while reducing NOx and smoke emissions by approximately 88% and 61%, respectively. The study concluded that, at high methanol substitution ratios, the combination of early pilot injection, increased pilot diesel ratio, and EGR constitutes a more suitable strategy for improving combustion performance and emission control.

In the study by Cung et al. (2024), renewable diesel–methanol dual-fuel combustion (RMDF) was experimentally investigated in a heavy-duty single-cylinder engine. Methanol was injected through the intake port, while renewable diesel was directly injected as the pilot/ignition fuel. The effects of methanol substitution ratio on combustion characteristics, thermal efficiency, and NOx, soot, CO, and unburned hydrocarbon emissions were evaluated under different engine loads. The results indicated that the methanol substitution ratio could reach up to 83.4% at low load and approximately 49.4% at high load due to pressure rise rate limitations.

Increasing the methanol ratio generally improved thermal efficiency and significantly reduced NOx emissions. In addition, methanol utilization reduced heat transfer losses and showed a tendency to decrease engine-out CO_2 emissions. However, at high methanol ratios, especially under low-load conditions, ignition delay increased and CO and unburned hydrocarbon emissions rose. Under high-load conditions, increasing the methanol ratio elevated the maximum pressure rise rate, which became the primary factor limiting operation. The study emphasized that RMDF combustion offers considerable potential for cleaner and more efficient engine operation; however, injection and combustion strategies must be optimized to control combustion losses and excessive pressure rise rates at high loads.

In the study by Shi et al. (2026), a methanol–diesel dual-fuel ICCI combustion mode with a high methanol energy ratio under low-load conditions was experimentally investigated. The effects of different methanol and diesel injection timings, as well as single and multiple injection strategies, on combustion, performance, and emissions were evaluated. The results showed that, with appropriate methanol injection timing, the indicated thermal efficiency exceeded 43% at methanol energy ratios of 71–84%, while the coefficient of variation under 6 bar IMEP conditions decreased to approximately 1.5%. Delaying methanol injection reduced CO emissions without causing significant changes in NOx and HC emissions. In addition, methanol, formaldehyde, and formic acid emissions were found to decrease, while most emitted particles were in the nucleation mode and below 200 nm in size. Among the investigated strategies, the combination of single diesel injection with double methanol injection was considered the most promising at low loads because it achieved similar thermal efficiency with lower cyclic variation. However, strategies involving methanol injection only during the intake stroke or double injection of both diesel and methanol were considered disadvantageous due to lower performance and higher emissions.

In the study by Huang et al. (2026), the combined effects of methanol substitution ratio, intake air temperature, and excess air ratio on combustion, performance, and emissions were experimentally investigated in a diesel–methanol dual-fuel engine operating under low-load conditions. Experiments were conducted at 25% load, followed by multi-objective optimization of BSFC, NOx, PN, THC, and CO emissions using a combination of a BP artificial neural network and the NSGA-III algorithm. The results indicated that increasing the methanol substitution ratio prolonged ignition delay while reducing the maximum heat release rate and average in-cylinder temperature. Consequently, NOx and particulate emissions decreased, whereas BSFC, THC, and CO emissions increased because of incomplete combustion at low load. Increasing the intake air temperature reduced THC and PN emissions, although its effect on CO emissions varied depending on operating conditions. Increasing the excess air ratio reduced BSFC, NOx, and PN emissions but increased THC emissions. Optimization results showed that, when BSFC and THC were simultaneously improved, BSFC decreased by 1.69%, THC by 2.75%, PN by 18.22%, and CO by 7.27%, while NOx increased by only 0.75%. The study emphasized that intake temperature, methanol ratio, and excess air ratio must be jointly optimized to ensure the feasibility of diesel–methanol dual-fuel engines under low-load conditions.

In the study by Valera et al. (2026), the performance, combustion, cyclic variation, and emission characteristics of methanol–diesel blends were experimentally investigated in an off-road compression ignition diesel engine. Due to the phase separation issue of methanol in diesel fuel, iso-butanol was used as a co-solvent in the blends, while 1-dodecanol was additionally used at higher methanol concentrations. MBD5 and MBD10 fuels were tested at a constant speed of 1200 rpm under different load conditions and compared with neat diesel fuel. The results showed that both MBD5 and MBD10 generally produced higher in-cylinder pressure than diesel fuel; they exhibited similar heat release rates at low and medium loads but higher heat release rates at high loads. Brake thermal efficiency was comparable to or higher than that of diesel, while brake specific energy consumption was similar or slightly lower. In terms of cyclic variations, methanol–diesel blends demonstrated more stable operation than diesel fuel, particularly at low and medium loads. Regarding emissions, methanol blends generally produced lower CO emissions, NOx emissions comparable to diesel, and slightly higher HC emissions. The study concluded that, with the use of appropriate co-solvents, phase-stable methanol–diesel blends can be implemented in existing off-road diesel engines without major hardware modifications.

In the study by Li et al. (2026a), methanol–diesel dual-fuel DMCC combustion with a 90% methanol substitution ratio was numerically investigated to evaluate the high methanol utilization potential in small- and medium-bore marine diesel engines. Using a 1D–3D CFD model validated by experimental data, the effects of diesel injection timing and EGR ratio on combustion, knock tendency, and emissions were analyzed. The results indicated that excessively advanced diesel injection disturbed the balance between premixed and diffusion combustion, thereby increasing the heat release rate and negatively affecting efficiency. Delayed injection increased knock tendency; however, the application of 10–20% EGR effectively suppressed this tendency. The use of EGR reduced NOx emissions by approximately 30–60% and CO emissions by about 10%. On the other hand, increasing the EGR ratio increased soot formation because of oxygen deficiency and lower combustion temperatures. PCA analysis further identified the relationships between "load–energy output–NOx" and "injection timing–combustion phase–soot," demonstrating that this approach provides a useful calibration framework for combustion and emission optimization in high-methanol marine engines.

In the study by Li et al. (2025), two different mixture formation strategies for methanol utilization in diesel engines—methanol intake injection (MII) and direct injection of methanol/F-T diesel blends (MMDI)—were comparatively investigated. Experiments were conducted at 2400 rpm and full-load conditions, while 3D CFD simulations were used to analyze in-cylinder methanol distribution, OH radical formation, and soot precursor species. The results showed that increasing the methanol ratio increased both the maximum in-cylinder pressure and the pressure rise rate in both methods. At a 30% methanol ratio, the maximum in-cylinder pressure increased by 1.32 MPa in the MMDI method and by 2.27 MPa in the MII method. The MII approach produced an earlier and more homogeneous methanol atmosphere, allowing a larger amount of methanol to participate in combustion and increasing OH radical concentration. However, the pressure rise rate became the limiting factor in the MII method, whereas the coefficient of variation was the primary limiting factor in the MMDI method. The study proposed maximum methanol blending ratios of 23% for MMDI and 67% for MII, emphasizing that intake port injection is more suitable for high methanol ratios.

In the study by Wang et al. (2026), high-pressure direct injection (HPDI) methanol–diesel dual-fuel combustion for marine and heavy-duty engines was experimentally investigated. The study was conducted in a constant-volume combustion chamber using natural flame luminosity imaging, CH chemiluminescence imaging, two-colour pyrometry, and soot KL analysis to evaluate the effects of pilot diesel quantity, diesel–methanol injection interval, and methanol injection duration on ignition, flame development, temperature, and soot formation. The results showed that methanol ignited in the high-temperature regions generated by the pilot diesel flame, particularly starting from the spray periphery. Increasing the pilot diesel quantity improved ignition stability and flame temperature but could also increase initial soot formation because of locally rich regions. Increasing the interval between diesel and methanol injections reduced the cooling effect of methanol, resulting in more stable ignition, higher CH intensity, higher flame temperature, and lower soot formation. In contrast, excessively long methanol injection durations increased the evaporative cooling effect, thereby reducing CH intensity and combustion intensity. The study emphasized that proper adjustment of pilot diesel quantity and injection timing is critical for achieving more stable, cleaner, and more efficient combustion in HPDI methanol–diesel systems.

In the study by Yao et al. (2008), a diesel/methanol compound combustion system (DMCC) for methanol utilization in diesel engines was introduced and experimentally investigated in a naturally aspirated four-cylinder diesel engine. In the DMCC system, methanol was injected through the intake port to form a homogeneous air–methanol mixture, while diesel fuel acted as the pilot fuel to ignite the mixture. Experiments were conducted under different load and speed conditions, both with and without an oxidation catalyst. The results indicated that DMCC application increased ignition delay, reduced in-cylinder temperature, and decreased diffusion combustion while promoting more intense premixed combustion. As a result, smoke/soot and NOx emissions were significantly reduced; the maximum smoke reduction reached approximately 50% with DMCC alone and about 80% when combined with an oxidation catalyst. However, CO and HC emissions increased because of lower in-cylinder temperatures and incomplete combustion. According to the ECE R49 13-mode test results, DMCC alone reduced NOx emissions by approximately 40% but increased HC, CO, and PM emissions. In contrast, the combined use of DMCC and an oxidation catalyst reduced HC, CO, NOx, and PM emissions simultaneously compared with conventional diesel operation. The study concluded that the combination of intake-port methanol injection, diesel pilot ignition, and oxidation catalyst application constitutes an effective emission reduction strategy for methanol-fueled diesel engines.

In the study by Liu et al. (2025), a methanol/diesel dual-fuel operating mode integrated with a DOC + CDPF + SCR aftertreatment system was experimentally investigated in an off-road diesel engine under full-load conditions. The engine was tested under conventional diesel operation and dual-fuel operation with a 30% methanol substitution ratio over a speed range of 1000–2400 rpm. The results showed that fuel economy improved at low and medium engine speeds in dual-fuel mode, with the maximum brake thermal efficiency reaching approximately 40.6%. Methanol/diesel dual-fuel operation reduced NOx and PM emissions by up to 53% and 57%, respectively, owing to lower-temperature and more homogeneous combustion. However, due to methanol's high latent heat of vaporization, low combustion temperature, and incomplete combustion tendency, HC, CO, unburned methanol, and formaldehyde emissions increased significantly. Evaluation of the aftertreatment system showed that the DOC achieved conversion efficiencies above 95% for CO and HC in dual-fuel mode, while the CDPF achieved PM trapping efficiencies above 98%. Although the SCR system effectively reduced NOx emissions, its conversion efficiency decreased sharply with increasing engine speed, dropping from 95.6% at 2000 rpm to 37.1% at 2400 rpm. In addition, the SCR system reduced NOx emissions while increasing N_2O emissions. The study emphasized that maintaining the low NOx and PM advantages of methanol/diesel dual-fuel engines requires improved SCR efficiency at high engine speeds as well as effective control of unregulated emissions such as formaldehyde and unburned methanol.

In the study by Yilmaz (2021), standard diesel fuel and ternary biodiesel–methanol–diesel (BMD) and biodiesel–ethanol–diesel (BED) fuel blends were comparatively tested in the same diesel engine. The tested blends included B45M10D45, B40M20D40, B45E10D45, and B40E20D40, and brake specific fuel consumption, exhaust gas temperature, and CO, NO, and HC emissions were evaluated. The results indicated that all alcohol-containing blends produced higher BSFC than standard diesel fuel, and fuel consumption increased with increasing alcohol content. No significant difference was observed in exhaust gas temperature. Increasing alcohol content generally reduced NO emissions while increasing CO and HC emissions. Methanol-containing blends were found to be more advantageous than ethanol blends in reducing CO and HC emissions, whereas ethanol blends were more effective in reducing NO emissions. The study emphasized that the effects of alcohol-added biodiesel–diesel blends depend not only on the oxygen content of the fuel but also on the cooling effect of the alcohol, blending ratio, and engine operating conditions [40].

In the study by Chen et al. (2019), the combustion and emission characteristics of diesel/n-pentanol/methanol ternary fuel blends were experimentally investigated in a common-rail diesel engine. Neat diesel, D80P20, D70P20M10, and D70P15M15 fuels were tested, and cylinder pressure, heat release rate, combustion temperature, NOx, soot, and ultrafine particle emissions were evaluated under different load conditions. The results showed that n-pentanol enabled methanol to dissolve stably and homogeneously in diesel fuel, thereby improving the feasibility of methanol blend utilization. Increasing the methanol ratio prolonged ignition delay, shortened combustion duration, and intensified heat release. At low and partial loads, the blended fuels generated higher peak heat release rates and combustion temperatures. In terms of emissions, methanol addition significantly reduced soot and ultrafine particle emissions, with the D70P15M15 fuel producing the lowest soot and particulate emissions. However, NOx emissions increased because of the higher combustion temperature. It was also reported that delaying injection timing, particularly for the D70P15M15 fuel, enabled simultaneous reductions in NOx and particulate emissions. The study concluded that n-pentanol-assisted diesel/methanol blends offer strong potential for reducing soot and ultrafine particles, although the NOx increase must be controlled through injection strategies.

In the study by Yin et al. (2023b), the operating range of a dual direct injection engine fueled with methanol and diesel was experimentally investigated. Two independent injection systems were used, and the methanol energy substitution ratio (ESR), engine load, combustion stability, indicated thermal efficiency, and emissions were evaluated. The results showed that the methanol ESR reached a maximum of 65.1%; however, at high loads this ratio was limited by harsh/knocking combustion. The best overall operating point was achieved at approximately 52.4% ESR, where the highest indicated thermal efficiency reached 43.4%. At low loads, methanol's high latent heat of vaporization and low ignition tendency weakened combustion and reduced efficiency. In contrast, at medium and high loads, increasing the methanol ratio improved combustion rate and cyclic stability while reducing NOx emissions. However, HC and CO emissions in dual direct injection mode remained higher than those in neat diesel operation. The study highlighted that methanol/diesel dual direct injection systems offer strong potential for achieving high efficiency and low NOx emissions, particularly under medium-to-high load conditions, although harsh combustion and HC–CO emissions must be controlled at high methanol ratios.

In the study by Yang et al. (2025), a novel triple direct injection (TDI) combustion organization for methanol/diesel dual-fuel application in large-bore marine engines was numerically investigated. In the proposed system, the methanol injector was positioned at the center, while two diesel pilot injectors were placed on either side of the methanol injector, and this configuration was compared with a conventional dual direct injection (DDI) system. The objective of the study was to address issues such as limited cylinder head space, cold start difficulties, irregular methanol ignition, and combustion center displacement. The results showed that the TDI system achieved approximately 95% of the IMEP of the original diesel engine in neat diesel mode and outperformed the conventional DDI system. In dual-fuel mode, the TDI system provided higher power output and thermal efficiency than the DDI system because of more homogeneous methanol ignition and the coexistence of premixed and diffusion combustion. Delaying pilot diesel injection timing and reducing injection pressure increased the methanol premixed ratio, thereby improving IMEP and indicated thermal efficiency while reducing HC and soot emissions. However, the high methanol premixed ratio also increased NOx emissions and knock tendency because of higher local temperatures. The optimum pilot diesel injection condition was reported to be approximately −15.5°CA ATDC injection timing and 32 MPa injection pressure, providing a balanced compromise among high performance, low HC emissions, and acceptable knock tendency.

In the study by Yin et al. (2025), the effects of methanol and diesel injection timings on the maximum methanol energy substitution ratio, combustion, performance, and emissions were experimentally investigated in a diesel/methanol dual direct injection (DMDDI) engine. Methanol and diesel were directly injected into the cylinder using separate high-pressure injection systems, while methanol injection timing and diesel main injection timing were adjusted to control combustion phasing and fuel reactivity stratification. The results showed that, with appropriate methanol injection timing, the maximum methanol energy substitution ratio could reach up to 96.0%. Delaying methanol injection and advancing diesel injection created a suitable fuel concentration and reactivity gradient within the cylinder, thereby shortening combustion duration and increasing indicated thermal efficiency. The highest indicated thermal efficiency of 41.5% was achieved at a methanol injection timing of −60 °CA aTDC and a diesel main injection timing of −14 °CA aTDC. In terms of emissions, advancing methanol injection and delaying diesel injection reduced NOx emissions but increased HC, CO, and soot emissions. In contrast, under the −60 °CA methanol and −14 °CA diesel injection condition, HC, CO, and soot emissions reached their minimum levels while NOx emissions became the highest. The study concluded that dual direct injection technology is highly effective for achieving high methanol substitution ratios and controlled combustion in methanol/diesel engines, although careful optimization is required to balance high efficiency and NOx emissions.

The addition of methanol in diesel engines leads to various effects on combustion, performance, and emissions characteristics. A general comparison of methanol applications in diesel engines is presented in Table 1.

Table 1. General comparison of methanol applications in diesel engines

Application method	Key advantage	Key limitation	Overall result
Methanol–diesel mixture	Lower soot and particulate formation tendency	Phase separation, low cetane number, combustion instability	Additive ratio and co-solvent selection
Port fuel methanol injection	Homogeneous mixture, high methanol substitution ratio	Increase in HC, CO, and unburned methanol emissions	Suitable for high methanol ratios, but aftertreatment is required
Diesel pilot + methanol dual-fuel	Low NOx and soot emissions, cleaner combustion	Sensitive to pilot quantity and injection timing	One of the most common and feasible diesel–methanol strategies
Dual direct injection	Combustion phasing and reactivity stratification can be controlled	System complexity and NOx–efficiency trade-off	Strong potential for high methanol substitution ratio
Marine engine HPDI/TDI applications	High power output, low-carbon marine fuel potential	Ignition stability and emission control	Applicable to large engines, but requires optimization

Gasoline Engines

In the study by Balki et al. (2021), the optimal operating parameters of methanol–gasoline blends in a small-capacity spark-ignition engine were experimentally and statistically investigated. In the experiments, M5, M10, M15, and M20 methanol–gasoline blends were tested at different compression ratios (7:1, 8:1, and 9:1) and two load conditions (10 and 20 Nm). The results indicated that increasing the methanol content generally improved combustion due to higher oxygen content and higher laminar flame speed, leading to increased brake thermal efficiency (BTE) and reduced CO and UHC emissions. The highest BTE of 34.53% was achieved with M20 fuel at 20 Nm load and a compression ratio of 9:1. Brake specific fuel consumption (BSFC), however, increased with increasing methanol content due to its lower lower heating value. In terms of emissions, M20 fuel was found to be more advantageous, particularly at a compression ratio of 9:1, in reducing CO and UHC emissions. NOx emissions increased at low methanol ratios due to the oxygen effect, but tended to decrease at higher methanol contents because of lower flame temperature. ANOVA results showed that fuel type, compression ratio, and engine load had statistically significant effects on performance and emissions. In addition, multi-objective optimization using grey relational analysis demonstrated that optimal operating conditions could be selected depending on performance and emission priorities. The study concluded that methanol–gasoline blends can be considered a sustainable fuel alternative for small SI engines, and that the selection of optimal compression ratio and blend ratio is critical for balancing performance and emissions .

In the study by Boggavarapu et al. (2025b), the usability of high-methanol-content fuels in a small, air-cooled, port fuel-injected spark-ignition engine was experimentally investigated. Gasoline, M85 (85% methanol + 15% gasoline), and M100 (pure methanol) fuels were compared under full load, partial load, and WMTC motorcycle driving cycle conditions. The results showed that M85 and M100 fuels increased engine torque and power by approximately 0.3–0.5 Nm compared with gasoline, with the highest performance achieved using M100. Due to methanol's high laminar flame speed and lower combustion temperature, brake thermal efficiency increased by approximately 2–3% for M85 and M100. However, because of methanol's low lower heating value, brake specific fuel consumption increased significantly, by about 80% for M85 and 110% for M100 compared to gasoline. In terms of emissions, pre-catalyst CO, UHC, and NOx emissions were significantly reduced with methanol fuels. However, due to lower exhaust gas temperatures, catalyst efficiency decreased, resulting in similar post-catalyst UHC and NOx emissions compared to gasoline. Additionally, increasing the compression ratio from 9.7 to 10.7 to exploit methanol's high octane number resulted in approximately 22% increases in torque and power and about a 7% improvement in brake thermal efficiency. The study concluded that high-methanol-content fuels offer strong potential in small SI engines in terms of performance and efficiency; however, high fuel consumption, catalyst thermal management, and methanol-specific engine calibration must be carefully addressed.

In the study by Gong et al. (2011), the effects of ambient temperature on ignition behavior and formaldehyde and unburned methanol emissions during cold start in spark-ignition methanol and LPG-assisted methanol engines with electronically controlled port fuel injection were experimentally investigated. A single-cycle fuel injection strategy was applied, and ignition behavior was evaluated at different ambient temperatures. The results showed that when ambient temperature decreased from 301 K to 289 K, the minimum methanol injection quantity required for reliable ignition increased by approximately 86%. Below 289 K, stable engine operation with methanol alone was not possible without auxiliary measures, mainly due to poor vaporization and insufficient formation of a combustible mixture. The addition of LPG as a supplementary fuel improved ignition reliability at low ambient temperatures. As ambient temperature increased, in-cylinder peak pressure and maximum pressure rise rate increased, while unburned methanol emissions decreased significantly. In contrast, formaldehyde emissions increased due to more active combustion. The study concluded that cold-start issues in methanol engines are primarily caused by poor vaporization characteristics, and that auxiliary fuels such as LPG can effectively improve cold-start performance.

In the study by Geng and Yao (2015), formaldehyde (HCHO) emissions from methanol–gasoline blends in a spark-ignition engine were comparatively investigated using different measurement methods. M0, M15, and M45 fuels were tested, and formaldehyde emissions were measured under identical engine conditions using FTIR, HPLC, and GC methods. The results showed that formaldehyde emissions increased significantly with increasing methanol content. In addition, formaldehyde emissions decreased from low to medium loads, but increased again from medium to high loads. Among the measurement methods, FTIR produced higher formaldehyde values than HPLC and GC. In contrast, HPLC and GC results were very close to each other. These differences were attributed to variations in measurement principles, interference effects from water and other exhaust components, and calibration differences. The study concluded that FTIR is more suitable for real-time and on-board measurements, whereas HPLC and GC provide higher separation quality and better repeatability and reliability. Therefore, careful selection of measurement methods is essential for accurate evaluation of unregulated emissions such as formaldehyde in methanol-fueled engines.

In the study by Balki et al. (2016), optimal operating parameters for gasoline, ethanol, and methanol fuels in a single-cylinder spark-ignition engine were determined using the Taguchi experimental design and ANOVA method. Engine speed, compression ratio, and ignition timing were selected as control parameters, while brake thermal efficiency (BTE) and brake-specific hydrocarbon (BSHC) and brake-specific NO emissions (BSNO) were considered performance and emission responses. The results showed that the optimal engine speed for all fuels was 2400 rpm and the optimal compression ratio was 9.0. The optimal ignition timing was 26° CA for gasoline and 20° CA for ethanol and methanol. The Taguchi method reduced the number of experiments by approximately 89%, and validation tests were within a 95% confidence interval. Under optimized conditions, the highest BTE of 32.5% was achieved with methanol. In addition, BSFC was reduced by an average of 15.64% compared to baseline conditions. Regarding emissions, methanol operation resulted in the lowest BSHC, BSCO, and BSNO emissions among the tested fuels. The study demonstrated that methanol can provide superior performance and emission characteristics under optimized conditions and highlighted the effectiveness of the Taguchi–ANOVA approach for engine optimization.

In the study by Li et al. (2019), the effect of exhaust gas recirculation (EGR) on knock behavior in a high-compression-ratio spark-ignition methanol engine was numerically investigated. A SAGE combustion model coupled with a detailed methanol chemical kinetic mechanism consisting of 46 species and 247 elementary reactions was used. Simulations were conducted for a port-fuel-injected methanol engine with a compression ratio of 17.5 under different EGR rates. The results showed that increasing the EGR rate reduced in-cylinder temperature and pressure, prolonged ignition delay, decreased heat release rate, and delayed knock onset. Knock intensity was significantly suppressed at high EGR levels. Without EGR, the knock-prone region was located on the intake side of the combustion chamber, while increasing EGR shifted the knock tendency toward the exhaust side. It was also found that CO_2, H_2O, and N_2 in the EGR mixture reduced knock tendency through their high specific heat capacity effects. Chemical species analysis indicated that OH radicals were strongly correlated with temperature and knock duration, H_2O_2 with flame propagation, and HCO with knock onset, while CH_2O, H_2O_2, CO, and OH exhibited higher reaction rates during knocking conditions.

In the study by Ning et al. (2020), lean-burn direct-injection spark-ignition combustion in a high-compression-ratio engine converted from a diesel engine was experimentally investigated. The engine was operated at 1500 rpm, full load, and a compression ratio of 15.0 under different excess air ratios (λ = 1.2, 1.3, and 1.4) and injection timings. The results showed that increasing the lean mixture ratio improved indicated thermal efficiency but reduced IMEP. In addition, increasing λ reduced THC, CO, CO_2, and NOx emissions. The best operating point was achieved at an injection timing of −240° CA ATDC, where IMEP and ITE were maximized and cyclic variation was minimized. At this condition, THC and CO emissions were minimized, while CO_2 and NOx increased due to more complete combustion. The engine produced 25.3% higher torque than the original diesel configuration at the same speed, with a maximum torque of 27.6 Nm. No knocking was observed in all tests, and both maximum pressure rise rate and COVIMEP remained within safe limits. The study demonstrated that converting diesel engines to methanol DISI operation offers strong potential for high efficiency, low emissions, and knock-free operation.

In the study by Güdden et al. (2021), a large-bore high-speed diesel engine with approximately 5 L cylinder displacement was converted into a port-fuel-injected spark-ignition methanol engine and experimentally investigated. The effects of relative air–fuel ratio, exhaust backpressure, and engine load on performance and emissions (NOx, CO, HC, unburned methanol, and formaldehyde) were analyzed. The results showed that performance benefits observed in small SI methanol engines could also be achieved in large-bore high-speed engines. Pre-ignition was not a major issue, while knock was the main load-limiting factor. Projections indicated that a full-scale PFI-SI methanol engine could achieve approximately 44% brake thermal efficiency, which is higher than that of comparable natural gas engines. It was also demonstrated that NOx emissions could be kept below IMO Tier III limits (<2 g/kWh) without aftertreatment. However, unburned methanol and especially formaldehyde emissions (~1 g/kWh) require oxidation catalyst aftertreatment. The study concluded that methanol offers high efficiency and low NOx potential in large marine engines, but control of formaldehyde and unburned methanol emissions is essential.

In the study by Wouters et al. (2023), the applicability of methanol as an alternative fuel in spark-ignition engines was comprehensively evaluated in terms of spray characteristics, engine performance, emissions, and numerical analysis. Methanol was compared with RON95 E10 gasoline. First, spray behavior was investigated in a low-pressure optical chamber, followed by experiments in a single-cylinder SI research engine using the same injector. The results showed that, at identical injection duration, there were no significant differences between methanol and gasoline in terms of spray penetration and cone angle. However, due to methanol's lower energy density, a longer injection duration was required under real engine conditions, leading to an increase in spray penetration of up to 64.5%. Engine tests indicated that methanol provided 7–24% higher net indicated efficiency compared to gasoline, depending on load conditions, and increased the maximum engine load by 9 bar IMEP. Under lean-burn conditions, methanol achieved a net indicated efficiency above 45%, and the lean combustion limit was extended by 0.1 λ compared to gasoline. Zero-dimensional and one-dimensional analyses showed lower heat losses and shorter combustion duration for methanol. Three-dimensional CFD results revealed that, although greater wall film formation occurred initially in methanol operation, it decreased to levels comparable to gasoline toward the end of compression. The study concluded that methanol offers high efficiency and low-emission potential in SI engines due to its high knock resistance, high latent heat of vaporization, and fast combustion characteristics.

In the study by Yu et al. (2024), the knock suppression effect of direct methanol injection in a port-fuel-injected gasoline engine operating under high-load conditions was experimentally investigated. Methanol was injected directly into the cylinder using a modified low-flow injector, while gasoline was supplied through the intake port. The effects of methanol injection timing and methanol fraction on knock intensity, knock probability, and end-gas auto-ignition timing were evaluated. The results showed that the most effective knock suppression was achieved when methanol was injected near bottom dead center. Very early injection weakened the cooling and stratification effects of methanol, whereas very late injection increased knock tendency due to reduced mixture homogeneity near compression end conditions. At low methanol fractions, knock intensity increased in some cases due to methanol's high laminar flame speed; however, increasing methanol fraction gradually suppressed knock. Beyond a certain methanol fraction, further increases had limited additional knock suppression benefits. The study concluded that direct methanol injection is an effective strategy for reducing high-load knock in gasoline SI engines, but its effectiveness is highly sensitive to injection timing and methanol fraction.

In the study by Liu et al. (2024), gasoline–ammonia dual-fuel combustion was investigated in a high-compression-ratio four-cylinder spark-ignition engine using both experimental and chemical kinetic analyses. Ammonia was considered a carbon-free sustainable fuel, and its effects on knock suppression, combustion phasing, thermal efficiency, and nitrogen-based emissions were analyzed. The results showed that ammonia addition significantly suppressed knock, expanded the allowable ignition advance range, and enabled combustion phasing optimization. Consequently, indicated thermal efficiency increased by up to approximately 4.7% under high-load conditions. However, at low loads, the low flame speed and low reactivity of ammonia negatively affected combustion, limiting efficiency benefits. In terms of emissions, while NOx increased with ignition advance in pure gasoline operation, ammonia addition reduced NOx emissions. Chemical kinetic analysis indicated that NOx formation in ammonia combustion is closely related to reactive radicals such as OH, H, and O, and that high-pressure three-body reactions consume these radicals, thereby reducing NOx formation. It was also reported that unburned NH_3 emissions mainly originate from crevice volumes and adsorption/desorption processes in oil films. Although not directly focused on methanol, the study provides a relevant comparative reference for carbon-neutral fuels in SI engines regarding knock suppression, combustion phasing optimization, and NOx chemistry.

In the study by Gandolfo et al. (2024), combustion, knock, and emission characteristics of methanol, ethanol, hydrous ethanol, and ethanol–water blends were experimentally compared in a high-compression-ratio (14.8:1) direct-injection spark-ignition engine. The main objective was to evaluate whether methanol and wet ethanol 80 (80% ethanol + 20% water) could be used interchangeably in SI engines. The results showed that methanol, compared with ethanol and ethanol–water blends, enabled more advanced ignition timing and higher efficiency at high loads due to its higher flame speed and superior vaporization characteristics. Although wet ethanol 80 and methanol exhibited similar theoretical cooling potential, methanol provided stronger charge cooling in intake-stroke injection and exhibited superior knock resistance due to its higher volatility. Load sweep results indicated that methanol achieved the highest net fuel conversion efficiency, reaching approximately 40.7% at 16 bar IMEP, compared with 39.2% for wet ethanol 80. Emissions analysis showed that methanol produced lower unburned HC and generally lower NOx tendencies. The study concluded that methanol is more advantageous than ethanol–water blends in high-compression SI engines, particularly under high-load, knock-limited conditions.

In the study by Singh et al. (2024), the effects of dibutyl ether (DBE)–methanol blends on knock behavior in a spark-ignition engine were experimentally and numerically investigated. To enhance the low reactivity of methanol, 0–40% volumetric DBE was added, and experiments were conducted under lean, stoichiometric steady-state, and transient stoichiometric conditions. The results showed that under lean conditions, increasing DBE content increased mixture reactivity, requiring combustion phasing retardation to maintain the same knock intensity. Under stoichiometric steady-state operation, at low DBE ratios a different knock phenomenon—termed deflagration-based knock—was observed, characterized by pressure oscillations without end-gas auto-ignition. At higher DBE levels (30–40%), the knock mechanism transitioned to conventional end-gas auto-ignition knock. Under transient conditions, deflagration-based knock dominated across all DBE–methanol blends. CFD analysis indicated that this phenomenon was caused by increased heat release in high-pressure regions near the flame front, creating a feedback mechanism that amplified pressure oscillations. The study concluded that knock in fast-burning renewable fuels such as methanol cannot always be explained by classical auto-ignition theory, and conventional ignition delay strategies may be insufficient for certain knock regimes.

In the study by Esposito et al. (2025), combustion, performance, and emission characteristics of methanol in a direct-injection spark-ignition single-cylinder research engine were investigated using both experimental data and a predictive 0D/1D modeling approach. A fractal flame propagation-based quasi-dimensional combustion model was developed for methanol, while phenomenological and simplified kinetic models were used for HC, CO, and NOx emissions. The model was validated under various engine loads, excess air ratios, EGR rates, and ignition timing conditions. The results showed that the model accurately predicted methanol combustion rates, pressure traces, and combustion duration variations. The average deviation was approximately 0.74% for IMEP and 2.3% for peak cylinder pressure. For emissions, the average deviations were 11.7% for unburned HC, 17.7% for NOx, and 16.7% for CO. Experimental trends indicated that dilution via air or EGR significantly reduced NOx emissions but increased unburned HC due to wall flame quenching and incomplete oxidation. Compared with gasoline, methanol exhibited lower intake pressure requirements and faster heat release due to its higher latent heat of vaporization and faster combustion characteristics. The study provided a robust modeling framework for optimization and calibration of methanol-fueled SI engines and methanol–gasoline blends.

In the study by Wang et al. (2024), knock control strategies in a port-fuel-injected methanol-fueled heavy-duty marine spark-ignition engine were experimentally investigated. Three main knock suppression strategies were compared under different load conditions: retarding ignition timing, increasing excess air ratio (leaning the mixture), and delaying methanol injection timing. Experiments were conducted at 1200 rpm under low load (MAP = 73 kPa) and high load (MAP = 139 kPa). The results showed that all strategies effectively reduced knock at low load. In particular, delaying injection timing enhanced charge cooling by allowing partial in-cylinder vaporization of methanol. At high load, ignition retard and mixture leaning were more effective, while injection delay was less effective due to deteriorated mixture homogeneity caused by long injection duration. It was also observed that reducing knock intensity generally led to lower IMEP, indicating a trade-off between knock suppression and power output. At low load, injection timing and air–fuel ratio optimization minimized power losses, whereas ignition retardation was the most effective strategy at high load. Air dilution was identified as the strategy with the smallest efficiency penalty.

In the study by Wu et al. (2016), lean-burn characteristics of methanol and gasoline fuels under idle conditions were experimentally compared in a port-fuel-injected spark-ignition engine. The engine was operated at 800 rpm with excess air ratios of λ = 1.0, 1.2, and 1.4. The results showed that methanol outperformed gasoline under lean idle conditions. Due to methanol's higher laminar flame speed and wider flammability limits, combustion was accelerated, and flame development and propagation durations were shortened. This improved combustion stability, particularly under lean conditions, significantly reducing COVIMEP compared to gasoline. With methanol, indicated thermal efficiency increased with increasing λ and reached 24.7% at λ = 1.4, whereas gasoline showed only limited improvement. In terms of emissions, methanol reduced HC, CO, and NOx emissions due to more homogeneous mixture formation, oxygen content, and lower combustion temperature. The study concluded that methanol is a more advantageous fuel than gasoline for SI engines under idle and lean-burn conditions in terms of combustion stability, efficiency, and emissions.

In the study by Balki and Sayin (2014), the effects of compression ratio on performance, combustion, and emissions were experimentally investigated in a single-cylinder spark-ignition engine fueled with methanol, ethanol, and unleaded gasoline. Experiments were conducted at compression ratios of 8.0:1, 8.5:1, 9.0:1, and 9.5:1 at a constant engine speed of 2400 rpm under full-load conditions. The results showed that both methanol and ethanol increased brake mean effective pressure, brake thermal efficiency, and volumetric efficiency compared with gasoline. Methanol was the most advantageous fuel due to its high oxygen content, high latent heat of vaporization, and rapid flame propagation. The highest BTE values were 29.73% for gasoline, 30.22% for ethanol, and 30.47% for methanol. Brake specific fuel consumption increased for alcohol fuels due to their lower heating value, but decreased as compression ratio increased. Combustion analysis showed earlier and higher peak cylinder pressures for alcohol fuels, with methanol reaching approximately 4994 kPa. In terms of emissions, methanol and ethanol reduced UHC, CO, and NOx emissions compared to gasoline, with methanol achieving the lowest NOx levels. The study concluded that methanol is a strong alternative fuel for high-compression SI engines, although higher fuel consumption and knock or instability risks at elevated compression ratios must be considered.

The addition of methanol in gasoline engines leads to various effects on combustion, performance, and emissions characteristics. A general comparison of methanol applications in gasoline engines is presented in Table 2.

Table 2. General comparison of methanol applications in gasoline engines

Application method	Key advantage	Key limitation	Overall result
Pure methanol	High octane number, fast combustion, low NOx potential	High fuel consumption, cold-start issues	Suitable for high compression ratio SI engines
Methanol–gasoline blends	Easier to apply in existing engines	Performance variation depending on blend ratio	Valuable as a transition fuel
Direct methanol injection	Strong charge cooling, knock suppression	Requires spray optimization and wall wetting control	Advantageous under high-load conditions
Lean combustion	Increased thermal efficiency, low NOx emissions	Combustion stability limit	Suitable due to methanol's wide flammability range
EGR/air dilution	Knock and NOx control	Increase in HC and CO emissions and slower combustion	High efficiency can be achieved through calibration
Marine-type PFI-SI methanol engine	Low-carbon fuel use in large engines	Knock control and power loss trade-off	Requires load- and strategy-dependent control

RESULTS

Methanol is widely regarded as one of the most promising alternative fuels capable of replacing fossil fuels in internal combustion engines. Owing to its high oxygen content, high octane number, wide flammability limits, high latent heat of vaporization, and rapid laminar flame speed, methanol improves combustion quality while significantly reducing knock tendency. These properties enable advanced ignition timing, faster heat release, and higher thermal efficiency, particularly in high-compression-ratio engines and diluted combustion strategies. However, its relatively low lower heating value remains a major limitation, leading to increased fuel consumption. In addition, challenges such as cold-start difficulties, as well as elevated emissions of formaldehyde, unburned methanol, and in some cases HC and CO, must be considered. Nevertheless, the possibility of producing methanol from renewable sources or via CO_2 utilization and green hydrogen pathways positions it not only as an engine fuel but also as an important energy carrier in the transition toward sustainable energy systems.

In diesel engines, methanol is typically applied through direct blending, port fuel injection, dual-fuel operation with pilot diesel ignition, dual direct injection systems, and high-pressure direct injection strategies developed for marine applications. Due to its low cetane number, methanol cannot be readily auto-ignited under compression ignition conditions; therefore, it is commonly used in combination with diesel as a pilot or ignition-improving fuel. The literature indicates that diesel–methanol dual-fuel systems significantly reduce NOx, soot, and particulate emissions, particularly at high methanol substitution ratios, resulting in cleaner combustion. However, lower combustion temperatures, high latent heat of vaporization, and incomplete oxidation may lead to increased CO, HC, unburned methanol, and formaldehyde emissions. Consequently, the effectiveness of methanol utilization in diesel engines strongly depends on the optimization of injection timing, pilot diesel quantity, EGR rate, methanol substitution ratio, and aftertreatment systems. Overall, methanol in diesel engines is considered a strong low-carbon fuel that improves the NOx–soot trade-off, although combustion stability and unregulated emissions remain critical challenges.

In spark-ignition (SI) engines, methanol provides significant advantages due to its high octane number and strong charge-cooling effect, particularly in terms of knock suppression, increased compression ratio potential, lean-burn operation, and high-load performance. Studies on neat methanol and methanol–gasoline blends show improvements in thermal efficiency, shorter combustion duration, reduced cyclic variability, and, under many operating conditions, lower HC, CO, and NOx emissions. At idle and lean conditions, methanol exhibits more stable combustion compared to gasoline, while in high-compression engines it enables higher in-cylinder pressures, faster heat release, and improved efficiency. However, in some operating regimes, its rapid combustion characteristics may induce pressure oscillations that differ from conventional end-gas auto-ignition knock, associated instead with flame propagation-driven instability. Therefore, methanol utilization in SI engines requires careful optimization of compression ratio, ignition timing, injection strategy, air–fuel equivalence ratio, and EGR/dilution levels.

Overall, methanol stands out as a highly promising sustainable fuel for both diesel and gasoline engines. In diesel engines, its primary benefits are associated with reductions in soot, particulate matter, and NOx emissions, whereas in SI engines, its high octane number, high latent heat of vaporization, and fast flame propagation enable improved knock resistance, higher compression ratios, and enhanced efficiency. Nevertheless, common limitations across both engine types include low energy density, increased fuel consumption, cold-start difficulties, material compatibility issues, and the formation of unburned methanol and formaldehyde. Therefore, the successful large-scale application of methanol as a transportation fuel requires not only fuel substitution but also integrated optimization of engine calibration, injection systems, combustion strategies, aftertreatment technologies, and fuel production pathways. In this respect, methanol represents a transitional and potentially long-term sustainable fuel option capable of contributing to carbon reduction without necessitating the complete abandonment of existing internal combustion engine technologies.

REFERENCE

Balki, M. K., ve Sayin, C. (2014). The effect of compression ratio on the performance, emissions and combustion of an SI (spark ignition) engine fueled with pure ethanol, methanol and unleaded gasoline. Energy, 71, 194-201.

Balki, M. K., Sayin, C., ve Sarıkaya, M. (2016). Optimization of the operating parameters based on Taguchi method in an SI engine used pure gasoline, ethanol and methanol. Fuel, 180, 630-637.

Balki, M. K., Temur, M., Erdoğan, S., Sarıkaya, M., ve Sayin, C. (2021). The determination of the best operating parameters for a small SI engine fueled with methanol gasoline blends. Sustainable Materials and Technologies, 30, e00340.

Boggavarapu, P., Ashwin, B. M., Athreya, V. A., Raviteja, S., Himabindu, M., ve Ravikrishna, R. V. (2025). Performance and emission characteristics of methanol and methanol-blended gasoline fuels in a high compression ratio SI engine. Applied Thermal Engineering, 281, 128703.

Boggavarapu, P., Mallikarjun, A. B., Mendu, H., Ramraju, V. N., Krishnamachari, S., Pandey, S. K., ve Ravikrishna, R. V. (2025). Performance and emission measurement of spark ignition engines operating with 100% methanol fuel. Heat Transfer Engineering, 46(9-10), 839-851.

Chen, H., Su, X., He, J., ve Xie, B. (2019). Investigation on combustion and emission characteristics of a common rail diesel engine fueled with diesel/n-pentanol/methanol blends. Energy, 167, 297-311.

Cung, K. D., Wallace, J., Kalaskar, V., Smith III, E. M., Briggs, T., ve Bitsis Jr., D. C. (2024). Experimental study on engine and emissions performance of renewable diesel methanol dual fuel (RMDF) combustion. Fuel, 357, 129664.

Dong, Y., Kaario, O., Hassan, G., Ranta, O., ve Larmi, M. (2020). High-pressure direct injection of methanol and pilot diesel: A non-premixed dual-fuel engine concept. Fuel, 277, 117932.

Elfasakhany, A. (2014). Experimental study on emissions and performance of an internal combustion engine fueled with gasoline and gasoline/n-butanol blends. Energy Conversion and Management, 88, 277-283.

Esposito, S., Malfi, E., De Felice, M., Golc, D., Beeckmann, J., Pitsch, H., ve De Bellis, V. (2025). Methanol fuelling of a spark-ignition engine: Experiments and 0D/1D predictive modelling for combustion, performance, and emissions. Fuel, 393, 134657.

Feng, H., Zheng, Z., Yao, M., Cheng, G., Wang, M., ve Wang, X. (2013). Effects of exhaust gas recirculation on low temperature combustion using wide distillation range diesel. Energy, 51, 291-296.

Gandolfo, J., Lawler, B., ve Gainey, B. (2024). Experimental study of high compression ratio spark ignition with ethanol, ethanol-water blends, and methanol. Fuel, 375, 132528.

Geng, P., ve Yao, C. (2015). Comparative study on measurements of formaldehyde emission of methanol/gasoline fueled SI engine. Fuel, 148, 9-15.

Gong, C.-M., Li, J., Li, J.-K., Li, W.-X., Gao, Q., ve Liu, X.-J. (2011). Effects of ambient temperature on firing behavior and unregulated emissions of spark-ignition methanol and liquefied petroleum gas/methanol engines during cold start. Fuel, 90(1), 19-25.

Guo, H., Başhan, V., Yu, C., Bolat, F., Demirel, H., ve Tian, X. (2025). Effect of methanol injection timing on performance of marine diesel engines and emission reduction. Journal of Marine Science and Engineering, 13, 949.

Güdden, A., Pischinger, S., Geiger, J., Heuser, B., ve Müther, M. (2021). An experimental study on methanol as a fuel in large bore high speed engine applications – Port fuel injected spark ignited combustion. Fuel, 303, 121292.

Hassan, Q. H., ve Al-Abboodi, H. (2025). Experimental investigation of the impact of methanol-diesel blends on diesel engine emissions and performance. Combustion Engines, 201(2), 150-157.

Huang, F., Xu, J., Wan, M., ve Lei, J. (2026). Experimental investigation and optimization of performance and emissions reduction in a diesel/methanol dual-fuel engine under low load condition. Fuel, 406, 137022.

Güleroğlu, H., ve Yumurtacı, Z. (2025). Life cycle assessment of green methanol production based on multi-seasonal modeling of hybrid renewable energy and storage systems. Sustainability, 17(2), 624.

International Energy Agency. (2026). Transport. Paris, France: International Energy Agency.

International Renewable Energy Agency, ve Methanol Institute. (2021). Innovation outlook: Renewable methanol. Abu Dhabi, United Arab Emirates: International Renewable Energy Agency.

IEA Advanced Motor Fuels TCP. (2026). Fuel properties: Methanol.

Karvounis, P., ve Theotokatos, G. (2025). Performance improvement and emissions reduction of methanol fuelled marine dual-fuel engine with variable compression ratio. Fuel Processing Technology, 272, 108208.

Kumar, D., Sonawane, U., Chandra, K., ve Agarwal, A. K. (2022). Experimental investigations of methanol fumigation via port fuel injection in preheated intake air in a single cylinder dual-fuel diesel engine. Fuel, 324, 124340.

Lapuerta, M., Armas, O., ve Rodríguez-Fernández, J. (2008). Effect of biodiesel fuels on diesel engine emissions. Progress in Energy and Combustion Science, 34(2), 198-223.

Li, R., Liu, Q., Li, X., Yang, D., Yue, H., ve Meng, Y. (2025). The influence of methanol mixture formation mode on diesel engine combustion process. Applied Thermal Engineering, 278, 127279.

Li, X., Zhen, X., Wang, Y., Liu, D., ve Tian, Z. (2019). The knock study of high compression ratio SI engine fueled with methanol in combination with different EGR rates. Fuel, 257, 116098.

Li, Z., Wang, Y., Yin, Z., Gao, Z., Wang, Y., ve Zhen, X. (2021). Parametric study of a single-channel diesel/methanol dual-fuel injector on a diesel engine fueled with directly injected methanol and pilot diesel. Fuel, 302, 121156.

Li, Z., Liu, Y., Wang, Y., Zhang, Y., ve Gao, H. (2026). Research on premixed combustion parameterization of methanol/diesel dual fuel engine based on high methanol substitution rate. Fuel, 415, 138346.

Liu, S., Lin, Z., Qi, Y., Lu, G., Wang, B., Li, L., ve Wang, Z. (2024). Combustion and emission characteristics of a gasoline/ammonia fueled SI engine and chemical kinetic analysis of NOx emissions. Fuel, 367, 131516.

Liu, Y., Liu, J., Ji, Q., Wang, X., ve Ao, C. (2025). Experimental study on performance and emission characteristics of non-road methanol/diesel dual-fuel engine with after-treatment system at full loads. Journal of Environmental Management, 376, 124486.

Ma, B., Wang, W., Guo, S., Zhan, Q., ve Yao, C. (2026). A comparative study on combustion and emissions between diesel and diesel methanol dual fuel modes under different loads. Fuel, 407(Part C), 137572.

MAN Energy Solutions. (2021). ME-LGIM two-stroke dual-fuel methanol engine. Copenhagen, Denmark: MAN Energy Solutions Technical Report.

Ning, L., Duan, Q., Liu, B., ve Zeng, K. (2020). Experimental assessment of lean-burn characteristics for a modified diesel engine operated in methanol direct injection spark ignition (DISI) mode at full throttle condition. Fuel, 279, 118455.

Pan, W., Yao, C., Han, G., Wei, H., ve Wang, Q. (2015). The impact of intake air temperature on performance and exhaust emissions of a diesel methanol dual fuel engine. Fuel, 162, 101-110.

Reitz, R. D., ve Duraisamy, G. (2015). Review of high efficiency and clean reactivity controlled compression ignition (RCCI) combustion in internal combustion engines. Progress in Energy and Combustion Science, 46, 12-71.

Shi, Z., Qian, Y., Mi, S., ve Lu, X. (2026). Application of different injection strategies under methanol/diesel dual-fuel intelligent charge compression ignition (ICCI) mode with high methanol energy ratio at low engine loads. Fuel, 405, 136561.

Shu, M., Liu, Z., Wu, F., ve diğerleri. (2024). Experimental study on the combustion and emission characteristics of methanol/gasoline fuels in direct injection Miller cycle gasoline engines. International Journal of Automotive Technology, 25, 1517-1527.

Singh, E., Abboud, R., Strickland, T., Kim, N., Lopez Pintor, D., ve Sjöberg, M. (2024). Effect of dibutyl ether-methanol blend ratios on deflagration-based and autoignition-based knock in spark-ignition engines. Fuel, 376, 132670.

Sollai, S., Porcu, A., Tola, V., Ferrara, F., ve Pettinau, A. (2023). Renewable methanol production from green hydrogen and captured CO_2: A techno-economic assessment. Journal of CO_2 Utilization, 68, 102345.

Sun, Z., Lian, Z., Ma, J., Wang, C., Li, W., ve Pan, J. (2025). Combustion and emissions optimization of diesel-methanol dual-fuel engine: Emphasis on valve phasing and injection parameters. Processes, 13(4), 1183.

Tao, W., Sun, T., Guo, W., Lu, K., Shi, L., ve Lin, H. (2022). The effect of diesel pilot injection strategy on combustion and emission characteristic of diesel/methanol dual fuel engine. Fuel, 324, 124653.

Turner, J. W., Lewis, A. G., Akehurst, S., ve diğerleri. (2018). Alcohol fuels for spark-ignition engines: Performance, efficiency and emission effects at mid to high blend rates for binary mixtures and pure components. Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 232(1), 36-56.

Valera, H., Chauhan, S. S., Biswas, S., Cedík, J., Pexa, M., ve Agarwal, A. K. (2026). Alcohol-diesel blends fueled off-road compression ignition engine development: Part 1 – Assessing the influence of methanol-diesel blends on performance, combustion, and emissions. Fuel, 407, 137025.

Verhelst, S., Turner, J. W. G., Sileghem, L., ve Vancoillie, J. (2019). Methanol as a fuel for internal combustion engines. Progress in Energy and Combustion Science, 70, 43-88.

Wang, X., Xia, T., Li, H., Dong, Q., Qiao, Y., ve Cheng, Q. (2026). Combustion characteristics of a methanol-diesel dual fuel engine with high-pressure direct injection. Applied Thermal Engineering, 282, 128748.

Wang, Y., Long, W., Dong, P., Tian, H., Wang, Y., Xie, C., Tang, Y., Lu, M., ve Zhang, W. (2024). Experimental investigation of knock control criterion considering power output loss for a PFI SI methanol marine engine. Energy, 303, 132007.

Wang, Y., Xiao, G., Li, B., Tian, H., Leng, X., Wang, Y., Dong, D., ve Long, W. (2022). Study on the performance of diesel-methanol diffusion combustion with dual-direct injection system on a high-speed light-duty engine. Fuel, 317, 123414.

Wei, Y., Liu, S., Li, H., Yang, R., Liu, J., ve Wang, Y. (2008). Effects of methanol/gasoline blends on a spark ignition engine performance and emissions. Energy & Fuels, 22(2), 1254-1259.

Wouters, C., Burkardt, P., Steeger, F., Fleischmann, M., ve Pischinger, S. (2023). Comprehensive assessment of methanol as an alternative fuel for spark-ignition engines. Fuel, 340, 127627.

Wu, B., Wang, L., Shen, X., Yan, R., ve Dong, P. (2016). Comparison of lean burn characteristics of an SI engine fueled with methanol and gasoline under idle condition. Applied Thermal Engineering, 95, 264-270.

Yang, Y., Long, W., Qian, Y., Dong, P., Wang, Y., ve Cai, Z. (2025). Spatiotemporal regulation of methanol premixed and diffusion co-combustion for marine methanol/diesel dual fuel engine. Energy, 315, 134349.

Yao, C., Cheung, C. S., Cheng, C., Wang, Y., Chan, T. L., ve Lee, S. C. (2008). Effect of diesel/methanol compound combustion on diesel engine combustion and emissions. Energy Conversion and Management, 49(6), 1696-1704.

Yao, C., Pan, W., ve Yao, A. (2017). Methanol fumigation in compression-ignition engines: A critical review of recent academic and technological developments. Fuel, 209, 713-732.

Yilmaz, N. (2012). Comparative analysis of biodiesel-ethanol-diesel and biodiesel-methanol-diesel blends in a diesel engine. Energy, 40(1), 210-213.

Yin, X., Ren, X., Wang, J., Duan, H., Hu, E., ve Zeng, K. (2025). Influence of methanol and diesel injection timings on the maximum methanol energy substitution ratio and performance of diesel/methanol dual-direct injection engine. Energy, 318, 134762.

Yin, X., Xu, L., Duan, H., Wang, Y., Wang, X., Zeng, K., ve Wang, Y. (2023). In-depth comparison of methanol port and direct injection strategies in a methanol/diesel dual fuel engine. Fuel Processing Technology, 241, 107607.

Yin, X., Yue, G., Liu, J., Duan, H., Duan, Q., Kou, H., Wang, Y., Yang, B., ve Zeng, K. (2023). Investigation into the operating range of a dual-direct injection engine fueled with methanol and diesel. Energy, 267, 126625.

You, B. C. (1983). Performance of a spark ignition engine fueled with methanol or methanol-gasoline blends. SAE Preprints, 830898.

Yu, H., Su, Y., Shen, B., Zhang, Y., Wang, B., ve Zhou, Y. (2024). Effect of direct methanol injection based on low-flow injector on knock of gasoline engine under heavy load condition. Fuel, 357, 129899.

Zhu, J., Wang, Z., Li, R., Liu, S., ve Hua, Y. (2023). Experimental study and prediction model of combustion stability and combustion mode variation of burning methanol/biodiesel blends for diesel engines. Fuel, 335, 127038.

GoogLeNet-Based Recognition of Banana Leaf Disease and Damage Patterns

Serhat KÜÇÜKDERMENCİ[1]

1- Assoc. Prof. Dr.; Balıkesir University, Faculty of Engineering, Department of Electrical and Electronics Engineering. kucukdermenci@balikesir.edu.tr, ORCID No: 0000-0002-6421-7773

ABSTRACT

Foliar disorders in banana plants reduce the effective photosynthetic surface and can signal serious production risks before yield loss becomes visually obvious at field scale. This study develops an image-based prototype for recognizing banana leaf disease and damage patterns with a pretrained GoogLeNet network in MATLAB. The local image archive includes five balanced folders: Banana Skipper Damage, Black and Yellow Sigatoka, Chewing insect damage on banana leaf, Healthy Banana leaf, and Panama Wilt Disease. Each folder contains 845 images, producing 4,225 images in total. The MATLAB script imports these folders with imageDatastore, resizes the images to the GoogLeNet input size, replaces the final learnable and classifier layers, and trains the adapted model with SGDM. Because the command splitEachLabel(Dataset, 7.0) assigns seven images per category to the training subset, the reported run is interpreted as a small-sample transfer learning demonstration rather than a full benchmark. With 35 training images and 4,190 validation images, the experiment reached 65.30% final validation accuracy after 100 epochs. The result shows that GoogLeNet features can capture useful banana-leaf cues, while the training curves also reveal the need for stronger data splitting, augmentation, and external testing before field use.

Keywords –banana leaf disorders, GoogLeNet, transfer learning, MATLAB, imageDatastore, agricultural image analysis

INTRODUCTION

The banana leaf is one of the earliest visual surfaces where disease, insect activity, nutrient stress, and environmental injury become visible. Changes in color, texture, tissue continuity, and local necrosis can appear before the farmer has a quantitative measure of the plant's physiological condition. In practice, these visual signals are judged under variable light, mixed backgrounds, and different leaf ages. Such variation makes manual scouting valuable but inconsistent, especially when many plants must be checked in a short time.

For banana cultivation, the distinction between foliar disease and insect injury is operationally important. Sigatoka symptoms, Panama Wilt related discoloration, chewing damage, and skipper damage can all disturb the leaf surface, but they do not imply the same management response. A classifier that separates these patterns from healthy leaf tissue can support faster screening, provided that the model is trained and evaluated with transparent assumptions.

Deep learning is attractive for this task because convolutional networks learn visual descriptors from the image itself. Instead of manually defining color thresholds or texture formulas, a CNN can build layered representations from edges, veins, lesions, holes, and broader discoloration. Plant-disease literature has therefore moved strongly toward CNN-based classification, detection, and severity analysis, particularly when leaf images are available in class-labeled folders (Mohanty et al., 2016; Sladojevic et al., 2016; Ramcharan et al., 2017; Barbedo, 2018; Ferentinos, 2018; Liu & Wang, 2021; Pacal et al., 2024).

Training a deep CNN from the beginning is usually unnecessary for an exploratory agricultural dataset. Transfer learning offers a shorter route: a network trained on a broad image collection is reused as a feature extractor, and only the task-specific output layers are replaced. This approach is especially useful when the available data are organized and balanced but the training subset for a given prototype run is small. Studies on plant leaf images and pretrained CNNs support this practical strategy (Brahimi et al., 2017; Fuentes et al., 2017; Rangarajan et al., 2018; Too et al., 2019; Zhang et al., 2020).

MATLAB provides an integrated environment for numerical computing, data visualization, simulation, control design, signal processing, image processing, machine learning, and deep learning. Its Deep Learning Toolbox and imageDatastore structure make it possible to move from folder-based image archives to training-progress plots with relatively little code. The same MATLAB/Simulink ecosystem is also widely used in engineering prototyping, embedded control, observer design, real-time interfaces, and educational implementations. Erdem Ilten's studies illustrate this engineering use of MATLAB/Simulink across DC motor, PMSM, induction motor, Arduino, Raspberry Pi, DSP, and fractional-order control applications (Ilten, 2024e, 2025a, 2025b, 2025c, 2025d, 2025f; Ilten, Karaduman, & Demirtas, 2022a, 2022b; Ilten, Ozdemir, & Demirtas, 2023a, 2023b, 2024, 2025a, 2025b). Kucukdermenci's work similarly reflects simulation, PID control, sensor modeling, LabVIEW, Proteus, and MATLAB/Simulink-oriented engineering computation (Kucukdermenci, 2022a, 2022b, 2025a, 2025b, 2025c, 2025d, 2025e, 2025f, 2025g, 2025h).

This study documents a banana leaf classification experiment produced from a local MATLAB workflow. The aim is to describe the dataset structure, the GoogLeNet adaptation, the actual split implied by the code, and the meaning of the observed training-progress output. The work is positioned as a reproducible prototype for agricultural image analysis, not as a finished disease advisory system (Kamilaris & Prenafeta-Boldú, 2018; Zhang et al., 2020).

Literature Context

Agricultural computer vision studies often begin with a class prediction problem: an image is assigned to a disease, pest, stress, or healthy category. This framing is simple enough for rapid prototyping, yet it is sensitive to image capture conditions, class definitions, and the split between training and validation data. Reviews and application papers show that CNNs remain central in this field because they extract multi-level visual features and can be transferred between crop problems (Kamilaris & Prenafeta-Boldú, 2018; Hassan et al., 2022; Arsenovic et al., 2019; Picon et al., 2019; Wang et al., 2016).

Banana leaf imagery has particular challenges. Some symptoms are localized, such as holes or feeding marks; others appear as broader yellowing, dark streaking, or wilt-related tissue deterioration. CNN-based studies on leaf disease recognition suggest that pretrained features can separate many of these visual patterns, but they also warn that validation accuracy alone is incomplete without class-level error analysis (Sladojevic et al., 2016; Ashqar & Abu-Naser, 2018; Fuentes et al., 2017).

MATERIALS AND COMPUTATIONAL PROCEDURE

Image Archive and Class Structure

The local image archive was arranged exactly as MATLAB expects for folder-derived labels. Each disease or damage category was placed in a separate subfolder, and imageDatastore was called with IncludeSubfolders set to true and LabelSource set to foldernames. This organization produced five label groups and 4,225 images in total. The archive is balanced at folder level because every class contains 845 images.

The five folders correspond to Banana Skipper Damage, Black and Yellow Sigatoka, Chewing insect damage on banana leaf, Healthy Banana leaf, and Panama Wilt Disease. The folder names act as the labels used by the training script, so no separate spreadsheet or annotation file is required. This simplicity is useful for reproducibility, although the biological reliability of the labels still depends on how the original image folders were curated.

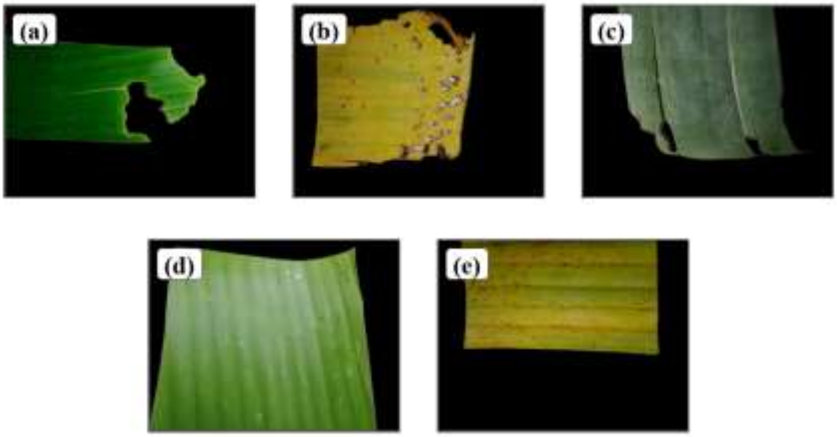


Figure 1: Overview of the five banana leaf image categories used in the experiment: (a) Banana Skipper Damage, (b) Black and Yellow Sigatoka, (c) Chewing insect damage, (d) Healthy Banana leaf, and (e) Panama Wilt Disease.

Class-specific sample panels were included to make the visual range inside each folder easier to inspect. The panels also show why a classifier must handle both localized tissue loss and wider color or texture changes.

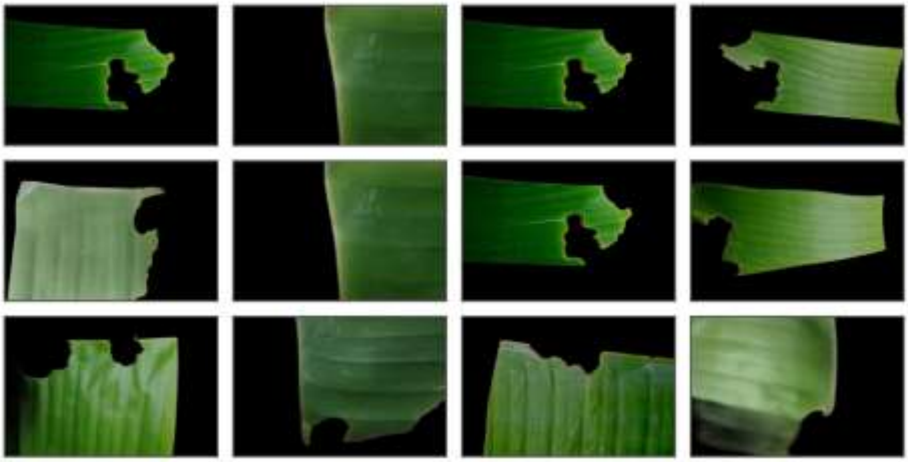


Figure 2: Banana Skipper Damage image examples.

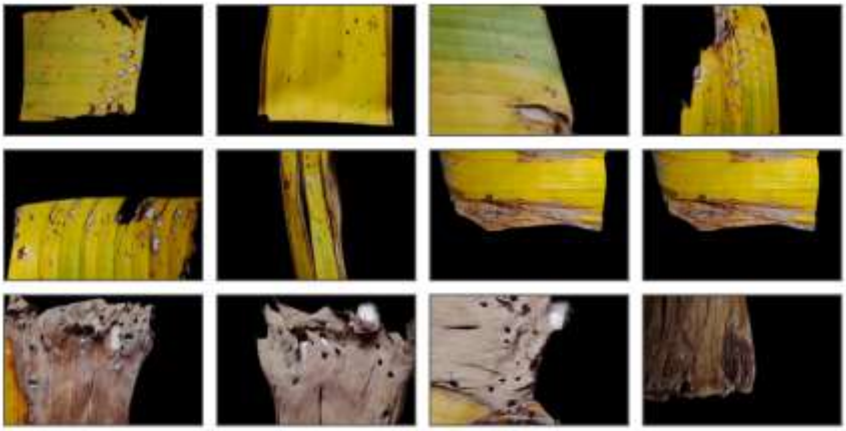


Figure 3: Black and Yellow Sigatoka image examples.

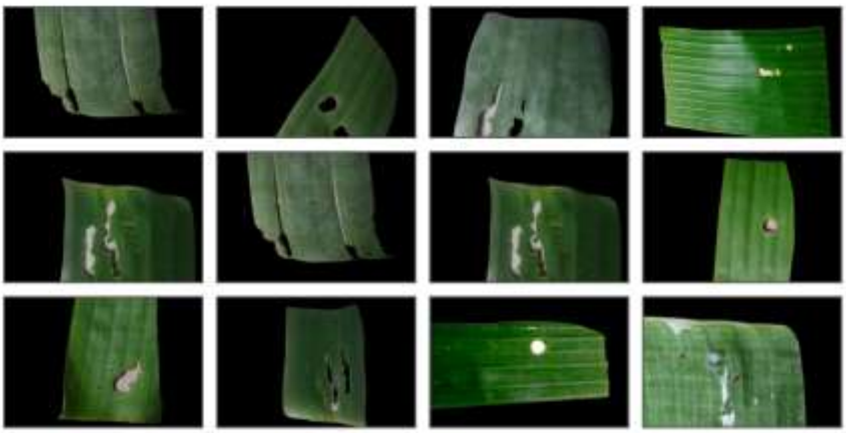


Figure 4: Chewing insect damage image examples.

Figure 5: Healthy Banana leaf image examples.

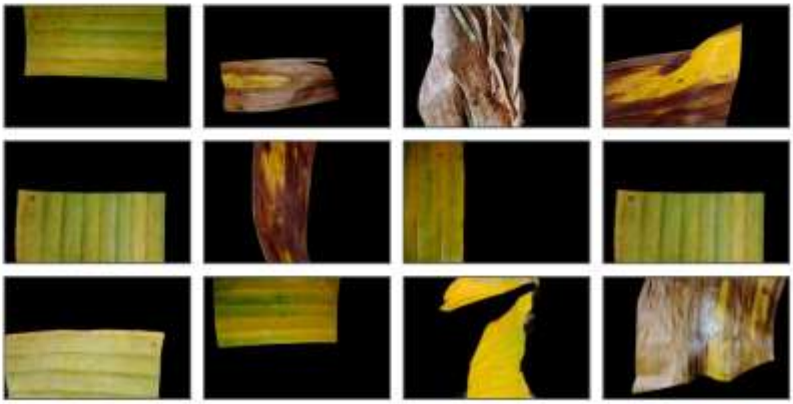


Figure 6: Panama Wilt Disease image examples.

GoogLeNet Reconfiguration

GoogLeNet was used as the pretrained visual backbone. Its inception modules process information at multiple filter scales inside the same network stage, which is useful when foliar symptoms vary from small spots and holes to larger yellow or necrotic regions. In this study, GoogLeNet is not proposed as a novel architecture; it is used as a stable baseline for testing whether the local banana leaf archive can be handled through a standard transfer-learning routine (Ferentinos, 2018; Too et al., 2019; Zhang et al., 2020).

The script reads the input size from the first GoogLeNet layer and uses augmentedImageDatastore to resize training and validation images. The original final learnable layer and classification layer are then replaced with a fully connected layer sized to the five leaf classes and a new classification layer. In the MATLAB code, these layers are named Banana Leaf Disease Learner and Banana Leaf Disease Classifier.

This setup keeps the pretrained feature extractor largely intact while changing the decision end of the network. The design is intentionally economical: it tests the usefulness of existing visual filters before moving toward heavier fine-tuning or alternative architectures.

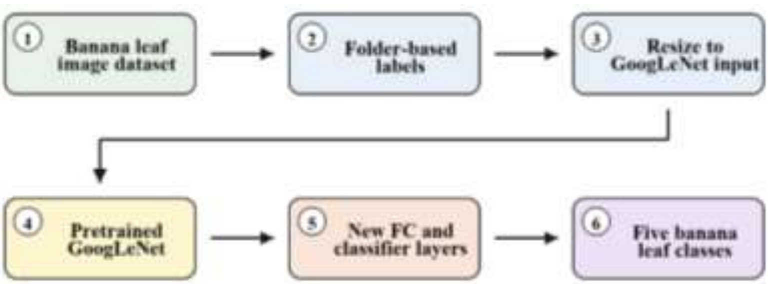


Figure 7: Computational flow of the GoogLeNet transfer-learning experiment for banana leaf images, from folder-based image input to final class prediction.

Table 1: MATLAB training options and network adaptation settings

Parameter	Value
Method	Stochastic Gradient Descent with Momentum (SGDM)
Momentum	Default SGDM momentum
Initial Learn Rate	3.0000e-04
Max Epochs	100
Learn Rate Schedule	Constant / none
Mini-batch size	5
Shuffle	Every epoch
Validation Data	augmentedImageDatastore
Validation Frequency	7 iterations
Execution Environment	Auto; single CPU in final run

Split Configuration and Run Setup

The split command is central to interpreting the result. The script uses splitEachLabel(Dataset, 7.0), which selects seven files from each label for the first output. It should not be read as a 70% training split. With five categories, the run therefore uses 35 training images and leaves 4,190 images for validation. The training-progress window is consistent with this reading because a mini-batch size of 5 produces 7 iterations per epoch.

This split turns the run into a deliberately narrow training demonstration. It is useful for showing that the pipeline executes and that pretrained features can separate some visual structure, but it is not a fair final benchmark for banana leaf diagnosis. A later experiment should use a randomized proportional split, for example splitEachLabel(Dataset, 0.7, 'randomized'), and should reserve an independent test set.

Table 2: Effective training and validation image counts produced by the script

Class	Training images	Validation images
Banana Skipper Damage	7	838
Black and Yellow Sigatoka	7	838
Chewing insect damage on banana leaf	7	838
Healthy Banana leaf	7	838
Panama Wilt Disease	7	838
Total	35	4,190

Reading the Training Output

The available evidence for this run is the MATLAB training-progress artifact. It reports the evolution of training accuracy, validation accuracy, training loss, validation loss, epoch count, iteration count, elapsed time, validation frequency, hardware resource, and learning-rate setting. Since the folder does not include a confusion matrix or saved class-wise predictions, precision, recall, F1-score, and per-class error patterns are outside the scope of the present document.

Validation accuracy is treated as the main numerical indicator. In a balanced five-class setting, a naive random decision would average about 20% accuracy. The reported value must therefore be compared with this baseline, but it must also be read together with the unusually small training subset and the behavior of the loss curves.

PROCEDURE TRAINING OBSERVATIONS

The run reached the planned endpoint of 100 epochs. MATLAB reported a final validation accuracy of 65.30%, 700 total iterations, 7 iterations per epoch, and an elapsed time of 186 minutes and 37 seconds.

Table 3: Training-progress values reported for the observed run

Result item	Observed value
Final validation accuracy	65.30%
Training completion status	Maximum epochs completed
Epochs	100 of 100
Iterations	700 of 700
Iterations per epoch	7
Elapsed time	186 min 37 sec
Hardware resource	Single CPU
Learning rate schedule	Constant

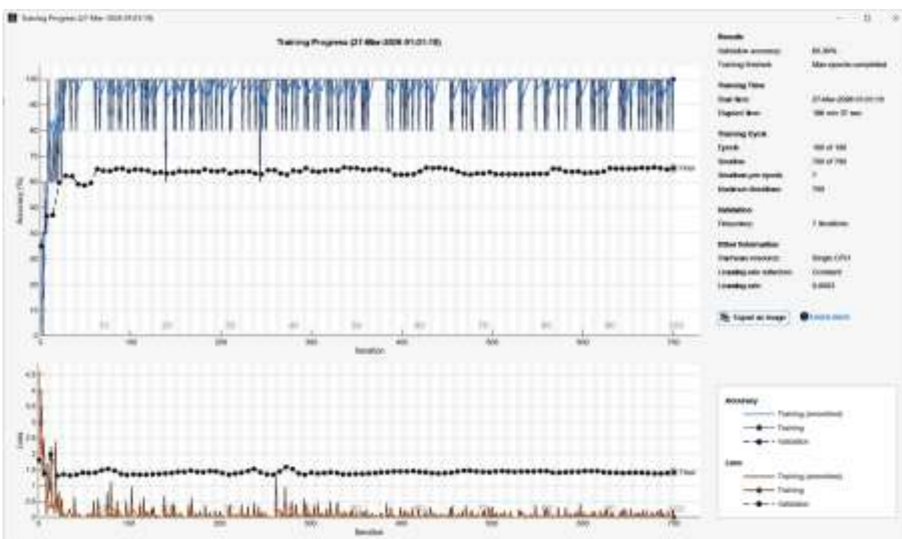


Figure 8: MATLAB training-progress window for the banana leaf GoogLeNet run.

The accuracy plot shows rapid adaptation to the tiny training subset. Training accuracy climbs quickly and frequently approaches the top of the scale, which is plausible when only 35 images are used for fitting. Validation accuracy follows a more restrained path and settles around the mid-60% range, indicating that the network learned some transferable leaf cues but did not generalize in a fully stable way.

The loss plot gives the same warning from another angle. Training loss becomes very small, whereas validation loss remains much larger and fluctuates after the early part of training. This gap suggests that the adapted classifier has memorized the selected training examples more strongly than it has learned the broader class variation present in the validation images.

Even with this limitation, the result is not trivial. The final 65.30% validation accuracy is well above the 20% chance-level reference for five balanced classes. The prototype therefore indicates that pretrained GoogLeNet features contain useful information for banana leaf symptom recognition, while also exposing the limits of a seven-image-per-class training design.

INTERPRETATION AND PRACTICAL IMPLICATIONS

The experiment demonstrates a low-barrier workflow for moving from a labeled banana leaf folder archive to a trained CNN classifier in MATLAB. Its value lies in procedural clarity: image folders become labels, images are resized automatically, the final GoogLeNet layers are exchanged, and the training process is visualized without building a network from scratch.

The numerical result should be presented as exploratory. A 65.30% validation accuracy is encouraging for a prototype with only 35 training images, but it is not adequate evidence for a field-ready diagnostic tool. The validation set is much larger and more diverse than the training subset, so the model is being asked to generalize from a very narrow sample of each class.

Confusion between categories is also likely. Sigatoka, Panama Wilt symptoms, skipper injury, and chewing damage can share yellowing, dark marks, ruptured tissue, or irregular leaf edges. Healthy leaves may include minor marks or illumination artifacts that visually resemble early stress. A confusion matrix is therefore necessary in the next run to identify which labels are actually separable and which ones require more data or a different modeling strategy.

The balanced folder archive is a strong starting asset. A more defensible experiment would use a randomized 70/30 or 80/20 split, keep a separate test subset, and document image provenance. That change would transform the current proof of execution into a more meaningful performance evaluation.

Augmentation should also be used more deliberately. In the present script, augmentedImageDatastore mainly serves resizing. For banana leaf images, rotations, reflections, brightness changes, scale variation, and mild contrast shifts could help the model handle camera orientation and field lighting. Similar plant-disease studies report that augmentation and fine-tuning choices can strongly affect generalization (Too et al., 2019; Zhang et al., 2020).

A methodological point is worth emphasizing: the code, not the intended split, determines what was actually tested. Reporting splitEachLabel(Dataset, 7.0) as if it meant 70/30 would misrepresent the run. The present study therefore keeps the seven-image-per-class training condition explicit.

Compared with a generic leaf-deficiency example, this study centers on banana-specific foliar categories and a larger balanced archive. The MATLAB routine remains simple, but the diagnostic meaning is different because the labels mix disease, insect damage, and healthy tissue.

Before practical deployment, the model would need testing with field images collected under different lighting, backgrounds, growth stages, and camera devices. It would also need confidence reporting and an uncertain-case option, because a forced label can be misleading when symptoms are mixed or image quality is poor.

Study Boundaries

The training subset is extremely small because splitEachLabel(Dataset, 7.0) selects seven images from each class. This makes the run useful as a workflow demonstration but weak as a final accuracy estimate. No class-level performance table was exported. Without a confusion matrix, the document cannot identify which banana leaf categories are most often confused. Image acquisition details are not available in the local materials. Camera type, lighting, background, leaf age, and symptom confirmation procedures should be documented in a stronger study. The trained model was not evaluated on an external field dataset. External testing is required before the workflow can support claims about real-world diagnostic reliability.

CONCLUSION

This study converted a local MATLAB script into a documented banana leaf image classification study. The dataset contains five balanced categories with 845 images per category, and the GoogLeNet adaptation replaces the final learnable and classification layers for the banana leaf task.

The observed run used 35 training images and 4,190 validation images because of the splitEachLabel(Dataset, 7.0) command. Under this restrictive training condition, the model reached 65.30% final validation accuracy, which is above the five-class chance baseline but still far from a complete diagnostic validation.

Future work should repeat the experiment with a randomized proportional split, export class-level metrics, use targeted augmentation, compare GoogLeNet with alternative lightweight CNNs, and test the trained model on independent banana leaf images. These steps would turn the present MATLAB prototype into a stronger agricultural image-analysis study.

REFERENCES

Arsenovic, M., Karanovic, M., Sladojevic, S., Anderla, A., & Stefanovic, D. (2019). Solving current limitations of deep learning based approaches for plant disease detection. Symmetry, 11(7), 939. https://doi.org/10.3390/sym11070939

Ashqar, B. A. M., & Abu-Naser, S. S. (2018). Image-based tomato leaves diseases detection using deep learning. International Journal of Academic Engineering Research, 2(10), 10-16.

Barbedo, J. G. A. (2018). Factors influencing the use of deep learning for plant disease recognition. Biosystems Engineering, 172, 84-91. https://doi.org/10.1016/j.biosystemseng.2018.05.013

Brahimi, M., Boukhalfa, K., & Moussaoui, A. (2017). Deep learning for tomato diseases: Classification and symptoms visualization. Applied Artificial Intelligence, 31(4), 299-315. https://doi.org/10.1080/08839514.2017.1315516

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. Computers and Electronics in Agriculture, 145, 311-318. https://doi.org/10.1016/j.compag.2018.01.009

Fuentes, A., Yoon, S., Kim, S. C., & Park, D. S. (2017). A robust deep-learning-based detector for real-time tomato plant diseases and pests recognition. Sensors, 17(9), 2022. https://doi.org/10.3390/s17092022

Hassan, S. M., Amitab, K., Jasinski, M., Leonowicz, Z., Jasinska, E., Novak, T., & Maji, A. K. (2022). A survey on different plant diseases detection using machine learning techniques. Electronics, 11(17), 2641. https://doi.org/10.3390/electronics11172641

Ilten, E. (2024e). Super-twisting sliding mode control implementation for a DC motor with Raspberry Pi and Simulink external mode. 3rd International Conference on Scientific and Innovative Studies, Konya, Türkiye.

Ilten, E. (2025a). 2-DOF PID position control application for DC motor with Arduino and Simulink external mode. 4th International Conference on Trends in Advanced Research, Türkiye.

Ilten, E. (2025b). New concepts and advanced studies in engineering - Basic applications in Simulink environment for Raspberry Pi Pico W. All Sciences Academy.

Ilten, E. (2025c). DC motor fractional order PID position control with Arduino Nano 33 IoT and Simulink external mode. 6th International Conference on Engineering, Natural and Social Sciences, Türkiye.

Ilten, E. (2025d). Extended EMF observer application for a PMSM with using DSP and Simulink. 2nd International Conference on Modern and Advanced Research, Türkiye.

Ilten, E. (2025f). Implementation of tilt integral derivative (TID) position controller for DC motor system with using Simulink and Arduino Nano 33 IoT. 8th International Conference on Applied Engineering and Natural Sciences, Türkiye.

Ilten, E., Karaduman, H., & Demirtas, M. (2022a). Conformable fractional order controller design and optimization for sensorless control of induction motor. COMPEL - The International Journal for Computation and Mathematics in Electrical and Electronic Engineering, 41(6), 1904-1922. https://doi.org/10.1108/COMPEL-09-2021-0334

Ilten, E., Karaduman, H., & Demirtas, M. (2022b). Conformable fractional order controller design and implementation for per-phase voltage regulation of three-phase SEIG under unbalanced load. Electric Power Components and Systems, 50(17), 1134-1151. https://doi.org/10.1080/15325008.2022.2139433

Ilten, E., Ozdemir, A. A., & Demirtas, M. (2023a). Design of induction motor speed observer based on long short-term memory. Neural Computing and Applications, 35, 19941-19954. https://doi.org/10.1007/s00521-022-07458-0

Ilten, E., Ozdemir, A. A., & Demirtas, M. (2023b). Fractional order weighted mixed sensitivity-based robust controller design and application for a nonlinear system. Fractal and Fractional, 7(10), 769. https://doi.org/10.3390/fractalfract7100769

Ilten, E., Ozdemir, A. A., & Demirtas, M. (2024). Application of a novel synergetic observer for PMSM in electrical vehicle. Electrical Engineering. https://doi.org/10.1007/s00202-024-02297-9

Ilten, E., Ozdemir, A. A., & Demirtas, M. (2025a). Conformable fractional-order extended EMF observer design and implementation for PMSM sensorless control system. Electrical Engineering. https://doi.org/10.1007/s00202-025-03171-y

Ilten, E., Ozdemir, A. A., & Demirtas, M. (2025b). Classification of fractional-order chaotic systems using deep learning methods. The European Physical Journal Special Topics. https://doi.org/10.1140/epjs/s11734-025-01604-0

Kamilaris, A., & Prenafeta-Boldú, F. X. (2018). Deep learning in agriculture: A survey. Computers and Electronics in Agriculture, 147, 70-90. https://doi.org/10.1016/j.compag.2018.02.016

Kucukdermenci, S. (2022a). Analysis of field-free region formed by parametric positioning of a magnet pair for targeted magnetic hyperthermia. Kuwait Journal of Science, 49(4). https://doi.org/10.48129/kjs.16097

Kucukdermenci, S. (2022b). Investigation of field free region formed by dual Halbach array for focused magnetic hyperthermia. Journal of Electrical Engineering, 73(2).

Kucukdermenci, S. (2025a). FEM-based numerical analysis of a shielded interdigitated capacitive humidity sensor. International Journal of Advanced Natural Sciences and Engineering Researches, 9(6).

Kucukdermenci, S. (2025b). Modeling and simulation of DC motor speed PID control using Outseal with LabVIEW integration. International Journal of Advanced Natural Sciences and Engineering Researches, 9(10).

Kucukdermenci, S. (2025c). Simulation of longitudinal vehicle dynamics under FTP-75 drive cycle using MATLAB/Simulink. International Journal of Advanced Natural Sciences and Engineering Researches, 9(9).

Kucukdermenci, S. (2025d). Implementation of deadband-enhanced PID position control using LabVIEW and Arduino for DC motor systems. International engineering conference proceeding.

Kucukdermenci, S. (2025e). PID control of DC motor position and speed in a LabVIEW-Arduino framework. 6th International Conference on Engineering and Applied Natural Sciences, Türkiye.

Kucukdermenci, S. (2025f). Real-time PID-based DC motor position control interface using LabVIEW and Proteus co-simulation. 1st International Conference on Pioneer and Academic Research, Türkiye.

Kucukdermenci, S. (2025g). Design of a hybrid PID-feedforward control system for smart greenhouse. International engineering conference proceeding.
Kucukdermenci, S. (2025h). Real-time speed control system: PID tuning and visualization with Proteus and LabVIEW. International engineering conference proceeding.
Liu, J., & Wang, X. (2021). Plant diseases and pests detection based on deep learning: A review. Plant Methods, 17, 22. https://doi.org/10.1186/s13007-021-00722-9
Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. Frontiers in Plant Science, 7, 1419. https://doi.org/10.3389/fpls.2016.01419
Pacal, I., Kunduracioglu, I., Alma, M. H., Deveci, M., Kadry, S., Nedoma, J., Slany, V., & Martinek, R. (2024). A systematic review of deep learning techniques for plant diseases. Artificial Intelligence Review, 57, Article 304. https://doi.org/10.1007/s10462-024-10944-7
Picon, A., Seitz, M., Alvarez-Gila, A., Mohnke, P., Ortiz-Barredo, A., & Echazarra, J. (2019). Plant disease identification from individual lesions and spots using deep learning. Biosystems Engineering, 180, 96-107. https://doi.org/10.1016/j.biosystemseng.2019.02.002
Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., & Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. Frontiers in Plant Science, 8, 1852. https://doi.org/10.3389/fpls.2017.01852
Rangarajan, A. K., Purushothaman, R., & Ramesh, A. (2018). Tomato crop disease classification using pre-trained deep learning algorithm. Procedia Computer Science, 133, 1040-1047. https://doi.org/10.1016/j.procs.2018.07.070
Sladojevic, S., Arsenovic, M., Anderla, A., Culibrk, D., & Stefanovic, D. (2016). Deep neural networks based recognition of plant diseases by leaf image classification. Computational Intelligence and Neuroscience, 2016, 3289801. https://doi.org/10.1155/2016/3289801
Too, E. C., Yujian, L., Njuki, S., & Yingchun, L. (2019). A comparative study of fine-tuning deep learning models for plant disease identification. Computers and Electronics in Agriculture, 161, 272-279. https://doi.org/10.1016/j.compag.2018.03.032
Wang, G., Sun, Y., & Wang, J. (2016). Machine learning for plant disease incidence and severity measurements from leaf images. 2016 15th IEEE International Conference on Machine Learning and Applications, 158-163. https://doi.org/10.1109/ICMLA.2016.0034
Zhang, K., Wu, Q., Liu, A., & Meng, X. (2020). Using deep transfer learning for image-based plant disease identification. Computers and Electronics in Agriculture, 173, 105393. https://doi.org/10.1016/j.compag.2020.105393

Low-Pressure Casting System: Design and Manufacturing

Tansel TUNÇAY[1*]
Abdulsefa CİN[2]
Badegül TUNÇAY[3]

[1]Karabük University, Faculty of Engineering, Metallurgy and Materials Engineering Department, Balıklarkayası, Karabuk, Turkey, tanseltuncay@karabuk.edu.tr
[2]Karabük University, Technology Faculty, Manufacturing Engineering Department, Balıklarkayası, Karabuk, Turkey, sefacin@gmail.com
[3]Sivas Science and Technology University, Faculty of Aviation and Space Sciences, Aircraft Engineering Department, Sivas,Turkey,btuncay@sivas.edu.tr
[*]tanseltuncay@karabuk.edu.tr

ABSTRACT

Conventional casting methods are often insufficient to satisfy the growing expectations for casting quality, especially in applications requiring specific mechanical properties and dimensional accuracy. In this context, technological advancements and evolving engineering requirements have led to the development of specialized casting methods. In particular, specialized casting methods have been developed to overcome limitations related to high mechanical properties, surface quality, dimensional accuracy, and mass production In this study, a low-pressure casting machine (LPCM), which has become indispensable in recent years for the development and production of Al and Mg alloys, was designed and manufactured for experimental applications. The design and manufacturing processes were successfully completed through the studies carried out. In addition, empirical equations describing the relationship between liquid metal velocity and the applied pressure, which can be readily used in practice, are summarized.

Keywords – Casting, low-pressure casting, design.

I. INTRODUCTION

Casting is one of the most cost-effective manufacturing methods, particularly for components with complex geometries. Using the casting process, components ranging from a few grams to several hundred tons can be manufactured. The quality of cast components is directly influenced by two fundamental physical phenomena occurring during the casting process. The first is the transfer of the liquid metal through the gating system into the mold cavity, and the second is the solidification of the liquid metal.

These two physical phenomena, both individually and through their interaction with each other, determine the microstructure and properties of the cast material [1–6]. Gravity casting (ca. 4000–3000 BC) is the oldest known casting method. Specialized casting methods were developed to overcome the limitations of conventional gravity casting. The evolution of specialized casting methods, in chronological order, can be outlined as follows. *Investment Casting*: A wax pattern with the desired geometry is first produced and then coated with a slurry containing refractory materials of appropriate AFS grain size to form the mold. It is the casting method that provides the highest surface quality and the lowest dimensional tolerances. Its origins date back to approximately 3500–3000 BC, and it was widely used for the production of ancient sculptures and jewellery. It is the oldest known 'precision' casting method used to produce artistic objects. Today, this process is widely employed in the manufacture of aircraft turbine blades. *Permanent Mold Casting:* A casting process in which liquid metal is poured into reusable steel or cast-iron moulds under the influence of gravity.

Permanent mold casting originated around 1000 BC during the Bronze Age, when stone or bronze moulds were used to produce metal weapons and tools. *Centrifugal Casting:* A casting process in which liquid metal is forced against the mold walls and solidifies under the action of centrifugal force generated by the rapid rotation of a sand or metal mold. This method is particularly suitable to produce pipes, rings, and cylindrical components. The process was patented by Anthony Eckhardt in 1809; however, its widespread industrial application did not occur until the late nineteenth century. *Die Casting:* The process begins with the injection of liquid metal into a metallic mold under high pressure, typically ranging from 40 to 150 MPa depending on the metal density, section thickness, and component weight. It enables very high production rates and is commonly preferred for aluminium and zinc alloys. The process emerged from the need for the rapid production of printing type. The first patented machine was developed in 1849. Die casting processes are classified into two categories according to the chamber system used for the molten metal. *Hot-Chamber Die Casting:* This process is generally used for low-melting-point metals such as zinc (Zamak) and magnesium alloys. The applied pressures are typically in the range of 70–150 bar (7–15 MPa). *Cold-Chamber Die Casting*: This process is particularly used for aluminium casting, and the applied pressures are significantly higher than those used in hot-chamber die casting, typically ranging from 500 to 1500 bar. *Shell Moulding:* In this process, the metallic pattern is first heated to a temperature between 200°C and 300°C. Subsequently, resin-coated sand is applied onto the heated pattern to form a rigid shell mold. The sand grain size is selected according to the desired surface quality, whereas the shell thickness is determined based on the component weight and section thickness. Shell Moulding (1940s): This process was developed by the German engineer Johannes Croning during World War II and is also known as the Croning Process. *Continuous Casting:* A process in which molten metal is continuously solidified into a strip, bar, or billet by passing it through a water-cooled copper mold. Continuous casting forms the foundation of modern steel production. Although its basic principles were established in the nineteenth century, continuous casting became a commercially viable and widely adopted technology in the 1950s, revolutionizing steel manufacturing. *Vacuum casting* methods were developed to improve casting quality by minimizing the formation of oxides and bifilms generated by liquid metal movement during mold filling. These techniques utilize vacuum conditions to reduce flow-induced defects, thereby enhancing the cleanliness and overall quality of cast components This method utilizes vacuum to prevent the formation of pores that develop as a result of the diffusion of insoluble gases in the molten metal into bifilms. Vacuum and Computer-Aided Modern Casting Methods (1970s and beyond): These are advanced casting techniques developed for the aerospace industry that minimize gas porosity and incorporate simulation-assisted process design.

LPDC is a casting process in which molten metal is transferred into a metallic mold using argon or nitrogen gas, typically at pressures ranging from 0.3 to 0.5 bar, thereby ensuring laminar flow conditions. During most LPDC processes, the molten metal delivered into the mold cavity is maintained under gas pressure until solidification is completed. LPDC is widely used in the casting of automotive components requiring high integrity, such as automobile wheels and cylinder heads, as well as in the aerospace industry. [5,7,8]

The transfer of molten metal into the mold cavity offers a few advantages, including ease of automation. Although the cycle time is generally longer than that of gravity-fed permanent mold casting systems, a single operator can typically supervise multiple LPDC units simultaneously. In addition, like most counter-gravity casting systems, LPDC provides a metal yield of approximately 80–95%, compared with systems that require large feeders due to defects associated with gating systems in gravity casting methods, where the metal yield typically ranges from 50% to 75%. As in gravity casting methods, the use of sand cores does not present a problem in LPDC, provided that the pressure applied during molten metal transfer is properly controlled. Otherwise, excessive pressure—sometimes applied to prevent the opening of bifilm oxide layers that can lead to microporosity—may cause sintering between the sand cores and the molten metal. Obviously, if the molten metal is of high quality, there is no need to apply excessive pressure to suppress porosity.

In this study, a Low Pressure Die Casting (LPDC) system with manual and constant pressure control was designed and manufactured for the production and development of oxidation-sensitive alloys. Furthermore, the study aimed to improve the understanding of the effects of the molten metal volume, the applied pressure, and selected process parameters during molten metal transfer and casting within the LPDC system.

II. DESIGN AND MANUFACTURING

In the LPDC machine designed and manufactured in this study, the outer shell was fabricated by bending stainless steel sheets to form the external casing. A supporting frame was fabricated from 40 × 60 mm steel profiles with a wall thickness of 2 mm to enable the assembly of the equipment. The side sections were fixed using mounted bearings to facilitate assembly and maintenance operations in the event of potential malfunctions. In the design of the LPDC machine, a crucible and its flange were fabricated from 310-grade stainless steel for use during the melting process. The crucible had a diameter of 74 mm and a height of 250 mm. To prevent deformation and leakage caused by the high temperatures in the upper section of the melting unit, an ST52 steel flange was welded to the stainless-steel wall structure. Thus, the stability and dimensional accuracy of both the stainless-steel crucible and the melting unit were ensured during the melting process. The front (a) and top (b) views of the LPDC machine designed and manufactured in this study are presented in Figure 1. In the LPDC machine, two separate heating systems were incorporated to facilitate molten metal preparation and to raise the permanent mold to the desired temperature. Molten metal preparation was achieved using a 4-kW heating element with a thickness of 1.5 mm, while the mold heating unit was equipped with six cartridge-type heaters, each having a diameter of 8 mm and a length of 130 mm, providing a total heating power of 3.75 kW. The heating element brick of the melting unit was fabricated to the required dimensions using castable refractory material and was prepared for operation through a 24-hour heat-treatment process. The molten metal preparation capacity of the system is approximately 1 kg of aluminium per hour, while the metallic mold can be heated to 200°C within approximately 20–25 minutes. Temperature control of the melting unit was achieved using a stainless-steel-sheathed thermocouple with a diameter of 3 mm and a wire diameter of 0.5 mm. The mold heating unit was equipped with two stainless-steel-sheathed thermocouples of the same dimensions to monitor and control the temperature of the metallic moulds. A dual-stage pressure regulator was employed to reduce the high pressure of the argon gas supplied from the gas cylinder to an appropriate level and to deliver the gas to the molten metal surface at a constant pressure and flow rate, thereby enabling controlled molten metal movement. The gas regulator allowed the argon gas inlet pressure to be adjusted within a range of 0–230 bar, while the outlet pressure could be controlled between 0 and 1 bar in increments of 0.1 bar. This enabled the delivery of a stable argon pressure to the surface of the molten metal. To control the transfer of molten metal, a graphite runner with an outer diameter of 20 mm and an inner diameter of 10 mm was machined and installed vertically at the center of the melting crucible. The runner was mounted through the center of the crucible lid, and a high-temperature liquid sealant was applied to prevent gas leakage. The graphite vertical runner, which enabled the transfer of molten metal from the melting crucible to the mold cavity, was coated with boron nitride (BN) to prevent contamination of the molten metal. Grooves were machined into the ST52 component mounted on the outer wall of the melting furnace, as well as on the lower surface of the cover to which the metallic mold and the vertical runner were attached. Subsequently, sealing was achieved by installing Viton O-rings within these grooves, owing to their ability to maintain stability at elevated temperatures.

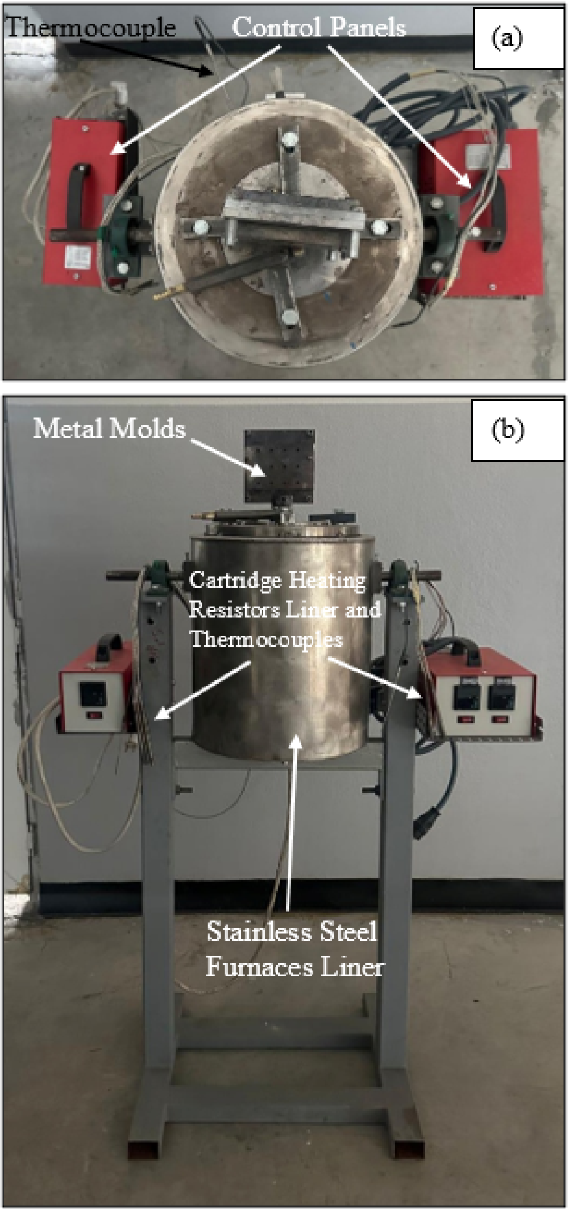


Figure 1. General front (a) and top (b) views of the Low Pressure Die Casting (LPDC) machine designed and fabricated in this study.

The dual-diaphragm gas regulator, which enables the movement of the liquid metal and allows precise adjustment of the gas pressure, is shown in Figure 2.

Figure 2. Dual-stage diaphragm gas regulator used for controlling molten metal movement and adjusting the applied gas pressure.

To evaluate the quality of the molten metal, metallic moulds were fabricated from ST52 steel. Steel fixtures were mounted on the outer surfaces of the moulds to accommodate cartridge heaters, and the assembly was secured using bolt–nut fasteners. Figure 3 presents photographs of the top view (a), front view (b), and mold cavity (c) of the metallic moulds.

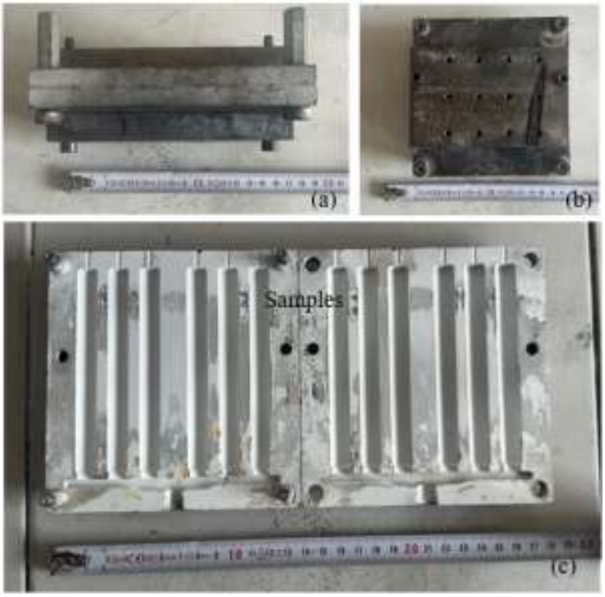


Figure 3. Top view (a), front view (b), and mold cavity view (c) of the metallic mold.

As shown in Figure 3, the inner surface of the melting crucible was coated with BN spray to reduce diffusion between the molten metal and the metallic mold and to extend the service life of the mold.

III. Results and Discussion

A. *Design and Fabrication*

Within the scope of this study, a simple LPDC machine was designed and manufactured for experimental use. Approximately 500 g of aluminium was melted in the liquid metal melting unit within about 40 minutes. The metal mold heating system was able to reach a temperature of 200 °C in approximately 10 minutes. A leak-tightness test was conducted on the melting furnace operating at 700 °C under pressures of 0.1, 0.2, and 0.4 bar. The system pressure was continuously monitored for approximately 8 hours using a manometer integrated into the system. No argon gas leakage was detected during the testing period. A simple low-pressure casting machine was successfully designed and manufactured for experimental use.

B. Calculations Used in the Low-Pressure Casting Machine Study

The equations used in the design and fabrication of the low-pressure casting machine developed for experimental applications were derived based on Bernoulli's principle, the continuity equation, and the laws of conservation of mass and energy. The mathematical equations required for calculating the gas pressure necessary to achieve the desired liquid metal velocity, depending on the cross-sectional area of the vertical runner or nozzle inlet, were derived according to the schematic shown in Figure 4. Initially, a 15 mm layer of liquid metal was retained at the bottom of the melting crucible as a safety reserve (250 − 235 = 15 mm). Depending on the weight of the liquid metal charged into the melting crucible, the liquid metal height in both the crucible and the vertical runner must first be determined.

To determine the liquid metal height (H, m), M is the mass of the liquid metal (kg), D is the diameter of the melting crucible (70 mm according to Figure 3) (m), D_1 is the outer diameter of the vertical runner (20 mm according to Figure 3) (m), D_2 is the inner diameter of the vertical runner (10 mm according to Figure 3) (m), K is the height of the safety liquid metal reserve located below the vertical runner and at the bottom of the melting crucible (15 mm according to Figure 3) (m), and ρ is the density of the liquid metal (kg/m³). By simplifying these parameters, Equation (1) can be derived as follows [9,10].

$$H = K + \frac{\left(\frac{4M}{\rho\pi}\right) - D^2 K}{D^2 - D_1^2 + D_2^2} \qquad \text{Equation (1)}$$

Using this equation, the height of the liquid metal in the melting crucible can be determined once the charged metal has completely melted. Subsequently, the distance between the liquid metal surface in the crucible and the top of the vertical runner can be readily calculated. This height is denoted by H_l.

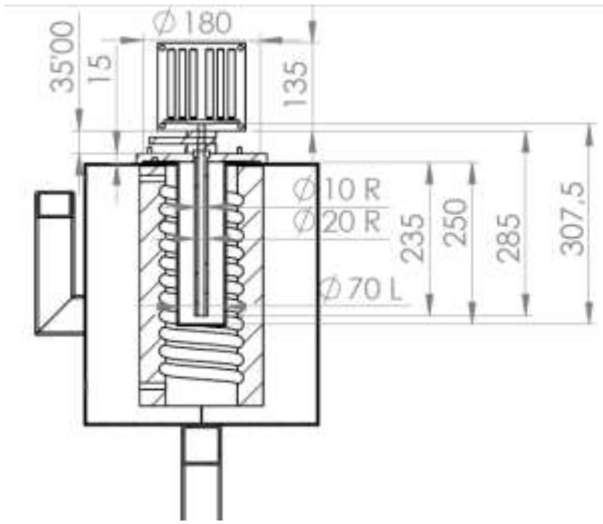


Figure 4. Dimensions used in the derivation of the equations for the low-pressure casting machine

The general form of Bernoulli's equation in fluid mechanics is given in Equation (2).

$$\frac{P_1}{\rho g} + \frac{V_1}{2g} + Z_1 = \frac{P_2}{\rho g} + \frac{V_2}{2g} + Z_2 + H_l \quad \text{(Equation 2)}$$

In Equation (2), P_1represents the pressure required to transfer the liquid metal, while P_2 denotes the pressure at the metal mold cavity or at the upper outlet of the graphite vertical runner. Here, $g$is the acceleration due to gravity, and $\rho$is the density of the liquid metal. The terms $Z_1$and Z_2represent the elevations of the liquid metal measured relative to a specified reference point. In the low-pressure casting process, the liquid metal inside the melting crucible is assumed to be stationary; therefore, the velocity $V_1$is taken as zero. The parameter H_l represents the vertical distance between the free surface of the liquid metal in the melting crucible and the entrance of the metal mold (or the upper end of the graphite vertical runner). If the liquid metal velocity at the metal mold inlet (or at the upper end of the graphite vertical runner), denoted by V_2, is represented as V, the equation required to determine the velocity $V$is given in Equation (3). Here, ΔP represents the pressure difference between $P_1$and P_2.

$$v = \sqrt{\frac{2(\Delta P - \rho g H_b)}{\rho}} \quad \text{(Equation 3)}$$

If the pressure required for the liquid metal to reach the top of the graphite vertical runner at a specified velocity $V$is to be determined, it can be calculated using Equation (4). In these equations, pressure losses resulting from friction within the melting crucible, at the graphite vertical runner inlet, and along the runner walls have been neglected. If the discharge coefficient is incorporated into the analysis, Equation (3) is modified to Equation (5), and Equation (4) becomes Equation (6) [11,12].

$$\Delta P = P = \rho g H_b + \frac{\rho v^2}{2} \quad \text{(Equation4)}$$

$$v = C_d \sqrt{\frac{2(\Delta P - \rho g H_b)}{\rho}} \quad \text{(Equation 5)}$$

$$\Delta P = \rho g H_b + \frac{\rho v^2}{2C_d^2} \quad \text{(Equation 6)}$$

IV. CONCLUSION

The following findings were obtained regarding the design and manufacturing challenges encountered during this study.

- Although no significant issues were identified during the design stage of the low-pressure casting machine, several problems were encountered during its fabrication and testing. These issues were subsequently resolved through a series of design revisions and modifications. During the melting furnace leak-tightness tests, thermal distortion and sheet metal warping occurred at the upper rim of the stainless-steel outer shell surrounding the melting crucible. These deformations resulted in gas leakage under pressure. To mitigate this problem, heat dissipation was improved and a “+”-shaped opening was introduced in the circular crucible housing. Although repeated tests showed a reduction in the degree of warping, argon gas leakage persisted.
- To eliminate argon gas leakage, a 25 mm thick steel reinforcement plate was added to the outer stainless-steel wall of the melting furnace to improve dimensional stability. In addition, a Viton O-ring was incorporated around the melting crucible to minimize heat loss and ensure dimensional accuracy, thereby improving the sealing performance of the system.

- Equations were developed to calculate the gas pressure required to transfer the liquid metal into the mold cavity at a desired velocity.
- Equations were developed to calculate the velocity of the liquid metal as a function of the applied gas pressure.

ACKNOWLEDGMENT

The authors would like to thank the Scientific Research Projects Coordination Unit of Karabük University for its support and contributions to this study under Project No. KBÜBAP-22-YL-154.

REFERENCES

[1] J. Campbell, The 10 Rules for good castings, in: Complet. Cast. Handb., Elsevier, 2011: pp. 605–737. https://doi.org/10.1016/b978-1-85617-809-9.10010-6.

[2] K.R. Ravi, R.M. Pillai, K.R. Amaranathan, B.C. Pai, M. Chakraborty, Fluidity of aluminum alloys and composites: A review, J. Alloys Compd. 456 (2008) 201–210. https://doi.org/10.1016/j.jallcom.2007.02.038.

[3] D.M. Stefanescu, Science and engineering of casting solidification: Third edition, 2015. https://doi.org/10.1007/978-3-319-15693-4.

[4] T.L. Sutton, The basic principles of fluid dynamics applied to running systems of castings, (2007).

[5] A.E.W. Jarfors, S. Seifeddine, Metal casting, 2015. https://doi.org/10.1007/978-1-4471-4670-4_81.

[6] M.C. Flemings, Solidification Process, McGraw-Hill, New York, 1974.

[7] J. Campbell, Complete Casting Metal Casting Processes, Metallurgy, Techniques and Design, First edit, Butterworth-Heinemann, Oxford, 2011.

[8] J. Campbell, Casting, 2003.

[9] J. Campbell, Castings Practice, Cast. Pract. 10 Rules Cast. (2004). https://doi.org/10.1016/b978-0-7506-4791-5.x5000-2.

[10] J. Gilbert Kaufman and Elwin L.Rooy, Aluminum Alloy Castings Properties,Processes, and Applications, 2004. https://doi.org/10.1016/j.matdes.2009.05.042.

[11] A.P.R. Bruce R. Munson.,Theodore H. Okiishi.,Wade W. Huebsch., Fundamentals of Fluid Mechanics, Seventh Ed, John Wiley & Sons, Inc., Jefferson City, 2013.

[12] J.C.L.M. Manhattan, Philip J. Pritchard., John W. Mitchell., FOX AND MCDONALD’S Introduction to Fluid Mechanics, 9th Editio, New Delhi, 2014.

Data-Driven Evaluation of Sustainability Performance in Plastic Injection Manufacturing Processes

Sakarie İbrahim AHMED*

Department of Industrial Engineering, Graduate School, Karabük University, Karabük, Türkiye
Zakiibro280@gmail.com

ABSTRACT

Sustainability performance is a critical criterion for enhancing manufacturing systems, particularly in processes where energy consumption, product quality, material efficiency, and environmental impact are interrelated. Although plastic injection molding is extensively utilized in polymer manufacturing, prior research frequently assesses sustainability using isolated indicators such as energy consumption, quality emissions, or CO_2 emissions. The present study proposes a data-driven framework for evaluating sustainability performance in plastic injection manufacturing by integrating process parameters, machine data, sensor outputs, quality indicators, and derived sustainability metrics.

An open experimental plastic injection molding dataset from the B2SHARE platform is utilized. This dataset comprises machine setting parameters, Euromap77 machine data, viscometer pressure data, repeated production cycles, and quality-related outputs. Descriptive statistics, correlation analysis, and comparative tests are employed to examine the relationships between process conditions and sustainability indicators. Additionally, energy-related performance and estimated CO_2 impact are evaluated.

The results indicate that sustainability performance cannot be explained solely by final product quality. Process stability, repeatability, and cycle-to-cycle variation also influence resource efficiency and sustainable production outcomes. This study contributes by integrating energy, quality, material loss, and environmental considerations into a unified analytical framework, thereby facilitating the identification of critical areas for improvement.

Keywords – Plastic injection molding, sustainability performance, data-driven analysis, process stability, energy efficiency, CO_2 impact.

I. INTRODUCTION

Sustainability has become a central concern in modern manufacturing systems, as industries are increasingly required to improve productivity while reducing energy consumption, material waste, and environmental impacts. In this context, sustainable engineering focuses not only on economic efficiency but also on the responsible use of resources throughout production processes. Plastic injection molding is one of the most widely used polymer manufacturing processes due to its suitability for mass production, dimensional accuracy, and the ability to produce complex parts. However, the process is also sensitive to machine settings, material properties, cycle conditions, and quality variation, which can directly influence energy use, material efficiency, and environmental performance.[1], [2].

Previous studies have shown that process parameters such as cylinder temperature, injection pressure, cooling time, injection speed, and cycle time can significantly affect both energy consumption and product quality in injection molding [2], [9]. Energy-oriented research has mainly focused on reducing electricity consumption and improving specific energy performance, while quality-oriented studies have generally examined defect reduction, dimensional accuracy, and process optimization [5], [6], [13]. In parallel, life cycle assessment studies have evaluated the environmental impacts of plastic products, including material selection, recycling potential, mold production, and CO_2-related effects [3], [4], [7]. These studies make important contributions; however, many examine energy, quality, material loss, and environmental impact as separate dimensions rather than interconnected components of sustainability performance.

Advancements in data-driven manufacturing, artificial intelligence, and process monitoring have enabled more comprehensive evaluation of injection molding processes. Machine data, sensor outputs, repeated-cycle measurements, and quality indicators facilitate the assessment of process stability and the identification of conditions conducive to sustainable production [8], [10], [15]. However, existing literature lacks studies that integrate process parameters, energy-related indicators, quality consistency, material efficiency, and CO_2 impact within a unified analytical framework. Addressing this gap is critical, as unstable process conditions can increase quality defects, reprocessing requirements, material waste, and indirect environmental impacts.

Despite significant contributions from previous studies on plastic injection molding, the literature remains fragmented. Energy-focused research typically addresses electricity consumption and process parameters, whereas quality-focused studies primarily investigate defect reduction and product consistency. Environmental studies often assess CO_2 emissions and life cycle impacts at a macro level yet seldom link these impacts to cycle-level process behavior. Consequently, few studies have integrated energy-related performance, quality consistency, material loss, process stability, and estimated CO_2 impact within a single data-driven framework.

To address this gap, the present study introduces a holistic data-driven evaluation framework for assessing sustainability performance in plastic injection manufacturing processes. In contrast to prior research that examines energy, quality, material loss, or environmental impact in isolation, this study evaluates these dimensions collectively using machine-setting parameters, Euromap77 data, sensor outputs, repeated production cycles, and quality indicators. The main contribution of the study is to demonstrate how process stability, cycle-to-cycle variation, energy-related indicators, and estimated CO_2 impact can be jointly analyzed to identify critical process conditions and opportunities for improvement for more sustainable plastic injection manufacturing.

II. MATERIALS AND METHOD

A. Dataset and Research Design

This study adopts a data-driven quantitative research design to evaluate sustainability performance in plastic injection manufacturing processes. Sustainability performance in manufacturing encompasses not only environmental impact but also energy efficiency, product quality, material efficiency, and process stability [1], [2]. Therefore, this study evaluates plastic injection molding performance using energy-related indicators, estimated CO_2 impact, quality consistency, and cycle-to-cycle variation.

An open experimental plastic injection molding dataset obtained from the B2SHARE research data platform was used in this study [17]. The dataset includes 68 production cycles across 17 machine-setting combinations, with 4 repeated cycles per experimental condition. It contains machine-setting parameters, Euromap77 machine data, viscometer pressure data, repeated-cycle observations, and quality-related outputs. The use of repeated cycles is important because process repeatability and stability directly affect product quality, scrap formation, energy use, and sustainable production performance [5], [6].

B. Variables and Sustainability Indicators

The variables were classified into process parameters, machine outputs, sensor-based process data, quality indicators, and derived sustainability indicators. Process parameters include machine settings such as cylinder temperature and injection speed. Machine outputs include cycle time, plastification time, injection time, heating energy, main drive energy, total energy, maximum power, and maximum machine pressure. Sensor-based data, especially viscometer pressure measurements, were used to represent process behavior and stability.

Sustainability performance was evaluated through energy-related performance, estimated CO_2 impact, process stability, quality consistency, and cycle-to-cycle variation. Energy consumption is one of the main sustainability indicators in injection molding because machine settings and thermal conditions can strongly influence electricity use [2], [8], [10]. The estimated CO_2 impact was calculated using electricity-related energy values and an emission factor, as is commonly used in environmental and life-cycle-based manufacturing assessments [3], [7], [12].

C. Data Preprocessing and Analysis Method

Before analysis, the dataset was organized by experiment ID and repeated-cycle information. The repeated observations were checked to ensure that each machine setting condition had corresponding cycle-level records. Numerical variables were examined using descriptive statistics, and energy-related variables were converted into comparable units where necessary. Total energy values were converted from Wh to kWh, and the estimated CO_2 impact was calculated based on electricity consumption.

The analysis was conducted in three stages. First, descriptive statistics were used to summarize the main process and sustainability-related variables, including mean, standard deviation, minimum, and maximum values. Second, correlation analysis was applied to examine the relationships between process parameters and energy-related sustainability indicators. Third, comparative statistical tests, including analysis of variance, were used to determine whether groups with different cylinder temperatures and injection speeds produced significant differences in energy-related performance. This methodological approach enables the identification of critical process conditions affecting sustainability performance in plastic injection molding.

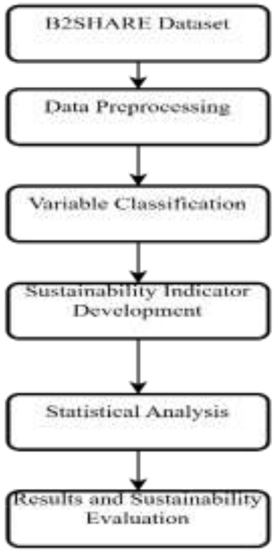


Fig. 1. Research workflow of the proposed data-driven sustainability evaluation framework

III. RESULTS AND DISCUSSION

Results should be clear and concise. The most important features and trends in the results should be described but should not interpreted in detail.

A. Descriptive Results

The descriptive analysis was conducted to characterize the general behavior of the plastic injection molding process and its sustainability-related performance indicators. The results indicate that the average total energy consumption was 45.549 Wh per cycle, equivalent to approximately 0.046 kWh per cycle. The estimated average CO_2 impact was 0.017 kg CO_2 per cycle. Total energy consumption ranged from 41.000 Wh to 49.300 Wh, demonstrating that variations in process settings produced measurable differences in energy-related sustainability performance.

The average cycle time was 41.262 s, while the average plastification time and injection time were 12.895 s and 1.464 s, respectively. The average heating energy was 16.613 Wh, whereas the average main drive energy was 22.193 Wh. These findings indicate that energy consumption in plastic injection molding is not only related to production time, but also strongly associated with thermal conditions and machine-level operating behavior. Therefore, sustainability performance should be evaluated through a combined perspective that includes energy use, process parameters, machine behavior, and environmental impact. Figure 2 should show the average total energy values for each cylinder temperature level. The expected trend is an increase in total energy consumption as cylinder temperature increases. The average total energy rises from 42.125 Wh at 220°C to 47.945 Wh at 260°C, showing that cylinder temperature is a critical factor affecting energy performance.

Table I. Descriptive Statistics of Main Process and Sustainability Variables

Variable	Mean	Std. Dev.	Min.	Max.
Total energy (Wh)	45.549	2.405	41.000	49.300
Electricity (kWh)	0.046	0.002	0.041	0.049
Estimated CO_2 (kg/cycle)	0.017	0.001	0.015	0.018
Cycle time (s)	41.262	0.393	40.630	41.920
Plastification time (s)	12.895	1.095	8.940	13.950
Injection time (s)	1.464	0.372	1.060	2.000
Heating energy (Wh)	16.613	3.428	10.700	21.600
Main drive energy (Wh)	22.193	1.297	20.000	24.500
Maximum power (kW)	7.722	2.799	4.027	14.189
Maximum machine pressure (bar)	1018.971	169.006	809.000	1418.000
Cylinder temperature (°C)	242.353	14.874	220.000	260.000

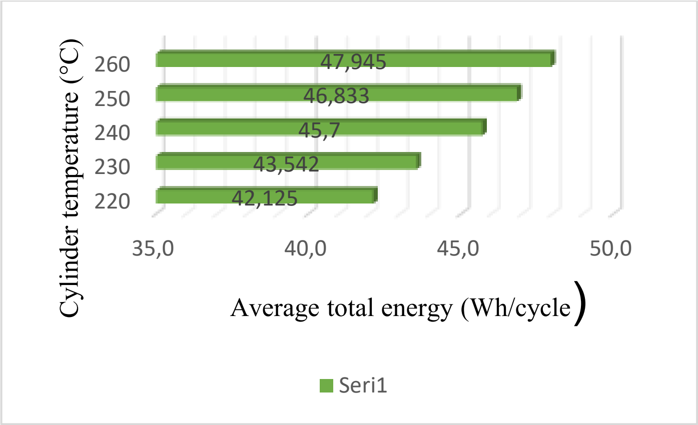


Fig. 2: Average total energy consumption by cylinder temperature

B. Relationship Between Process Parameters and Energy Performance

A correlation analysis identified the process variables most strongly associated with total energy consumption. The strongest positive relationship was observed between total energy and heating energy, with a correlation coefficient of $r = 0.988$ and $p < 0.001$. These results indicate that heating energy is the primary factor influencing total energy consumption in the analyzed injection molding cycles.

A very strong positive relationship was also identified between total energy and cylinder temperature ($r = 0.907$, $p < 0.001$). This finding demonstrates that higher cylinder temperature settings result in increased energy consumption, thereby elevating the estimated environmental impact. In contrast, total energy exhibited strong negative relationships with main drive energy and maximum machine pressure, with correlation coefficients of $r = -0.897$ and $r = -0.866$, respectively. This shows that cycle time, plastification time, injection time, and average injection speed had weak and statistically insignificant relationships with total energy. This indicates that thermal parameters are more influential than time-based variables in explaining energy variation in this dataset. Figure 3 is presented as a scatter plot or line chart. It should show that total energy consumption increases with cylinder temperature. This figure supports both the correlation analysis and the comparative results. Figure 4 should be presented as a scatter plot or a line chart. It should show that total energy consumption increases with cylinder temperature. This figure supports both the correlation analysis and the comparative results.

Table II. Correlation Results Between Process Variables and Total Energy

Variable	Correlation with Total Energy	p-value	Interpretation
Heating energy	0.988	<0.001	Very strong positive
Cylinder temperature	0.907	<0.001	Very strong positive
Main drive energy	-0.897	<0.001	Very strong negative
Maximum machine pressure	-0.866	<0.001	Very strong negative
Maximum power	-0.460	<0.001	Moderate negative
Plastification time	0.144	0.242	Weak, not significant
Cycle time	-0.136	0.268	Weak, not significant
Injection time	0.038	0.759	Weak, not significant
Average injection speed	-0.020	0.873	Weak, not significant

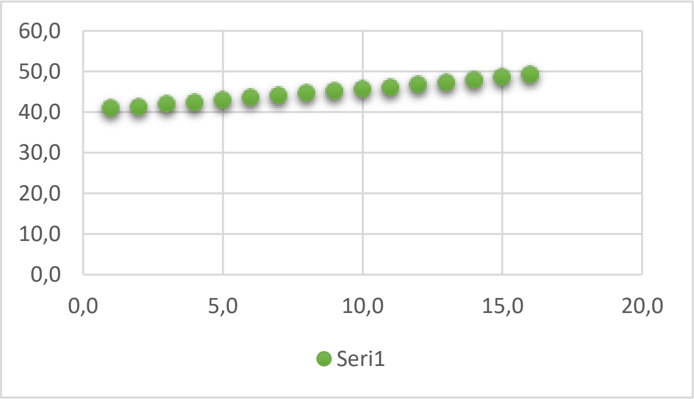


Fig. 3: Relationship between heating energy and total energy

.

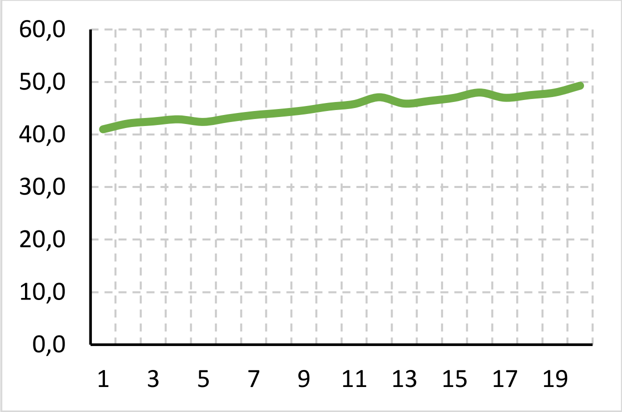


Fig. 4. Relationship between cylinder temperature and total energy.

C. Comparative Analysis of Process Settings

A comparative analysis was conducted to determine whether different process-setting groups showed statistically significant differences in total energy consumption. The results indicated that injection speed did not have a statistically significant effect on total energy consumption. The ANOVA result for injection speed was F = 0.334 and p = 0.718. The average total energy values were similar across the injection speed groups: 45.560 Wh at speed 50, 45.786 Wh at speed 75, and 45.205 Wh at speed 100. These findings suggest that injection speed was not a critical factor influencing energy-related sustainability under the experimental conditions analyzed.

Conversely, cylinder temperature demonstrated a statistically significant effect on total energy consumption. The ANOVA result for cylinder temperature was F = 80.131 and $p < 0.001$. The average total energy consumption increased from 42.125 Wh at 220°C to 47.945 Wh at 260°C. These results indicate that cylinder temperature is a key process parameter influencing energy performance in plastic injection. From a sustainability perspective, this finding is significant because increased energy consumption correlates with a higher estimated CO_2 impact. Consequently, optimizing cylinder temperature can directly reduce energy consumption, minimize environmental impact, and enhance sustainability performance.

D. Sustainability Discussion

The results indicate that sustainability performance in plastic injection molding cannot be assessed solely based on final product quality. Energy consumption, process stability, thermal conditions, cycle-to-cycle variation, and estimated CO_2 impact must be evaluated collectively. The observed strong relationship between heating energy, cylinder temperature, and total energy consumption confirms that thermal control represents the most critical area for improvement in the analyzed injection molding process.

Another key finding is that process parameters do not uniformly affect sustainability performance. Cylinder temperature significantly influenced energy consumption, whereas injection speed did not demonstrate a statistically significant effect. Therefore, improvement actions should be prioritized based on the actual impact of each process parameter. In practice, manufacturers are advised to focus on controlling cylinder temperature, minimizing unnecessary heating demand, enhancing thermal stability, and monitoring repeated-cycle behavior.

Overall, the findings validate the proposed data-driven approach for evaluating sustainability. Integrating process parameters, machine data, energy-related indicators, estimated CO_2 impact, and process stability yields a more comprehensive understanding of sustainability performance in plastic injection manufacturing. The results further demonstrate that data-driven analysis can identify critical process conditions and inform more sustainable production decisions.

IV. CONCLUSION

This study presented a data-driven evaluation of sustainability performance in plastic injection manufacturing processes. Using an open experimental injection molding dataset, process parameters, machine-level outputs, energy-related indicators, repeated cycle information, and estimated CO_2 impact were analyzed together. The main objective was to identify the process conditions that most strongly influence energy-related sustainability performance.

The results showed that the average total energy consumption was 45.549 Wh per cycle, while the estimated average CO_2 impact was 0.017 kg CO_2 per cycle. Correlation analysis revealed that heating energy had the strongest positive relationship with total energy consumption ($r = 0.988$, $p < 0.001$). In addition, cylinder temperature showed a very strong positive relationship with total energy consumption ($r = 0.907$, $p < 0.001$). The comparative analysis confirmed that cylinder temperature had a statistically significant effect on total energy consumption ($F = 80.131$, $p < 0.001$), whereas injection speed did not have a significant effect ($F = 0.334$, $p = 0.718$).

These findings indicate that thermal conditions are the most critical factors affecting sustainability performance in the analyzed plastic injection molding process. Therefore, reducing unnecessary heating demand, optimizing cylinder temperature, and maintaining stable process conditions can directly contribute to lower energy consumption and reduced estimated CO_2 impact.

The main contribution of this study is the integration of process parameters, machine data, energy-related indicators, process stability, and estimated environmental impact within a single data-driven sustainability evaluation framework. However, the study is limited to an open experimental dataset with a limited number of production cycles. Future studies should apply the proposed framework to real industrial production data and include direct measurements of energy consumption, CO_2 emissions, material waste, and product defects.

ACKNOWLEDGMENT

The author gratefully acknowledges Dr. Tuğrul Bayraktar for his academic guidance, valuable suggestions, and constructive feedback throughout the preparation of this conference paper. The author also thanks the providers of the open experimental plastic injection molding dataset available on the B2SHARE research data platform, which supported the empirical analysis conducted in this study.

REFERENCES

[1] J. B. Tranter, P. Refalo, and A. Rochman, "Towards sustainable injection molding of ABS plastic products," Journal of Manufacturing Processes, vol. 29, pp. 399–406, 2017.

[2] I. Meekers, P. Refalo, and A. Rochman, "Analysis of process parameters affecting energy consumption in plastic injection moulding," Procedia CIRP, vol. 69, pp. 342–347.

[3] B. Agarski, I. Budak, D. Vukelic, M. Hodolic, J. Hodolic, and D. Kozak, "Evaluation of the environmental impact of plastic cap production, packaging, and disposal," Journal of Environmental Management, vol. 245, pp. 55–65, 2019

[4] C. Vassallo, A. Rochman, and P. Refalo, "The impact of polymer selection and recycling on the sustainability of injection moulded parts," Procedia CIRP, vol. 90, pp. 504–509, 2020.

[5] J. Huang, Y. Li, X. Li, Y. Ding, F. Hong, and S. Peng, "Energy consumption prediction of injection molding process based on rolling learning informer model," Polymers, vol. 16, no. 21, p. 3097, 2024.

[6] C.-Y. Lin, J. Gim, D. Shotwell, M.-T. Lin, J.-H. Liu, and L.-S. Turng, "Explainable artificial intelligence and multi-stage transfer learning for injection molding quality prediction," Journal of Intelligent Manufacturing, vol. 36, pp. 3587–3606, 2025.

[7] OECD, Global Plastics Outlook: Economic Drivers, Environmental Impacts and Policy Options. Paris, France: OECD Publishing, 2022,10.1787

[8] EUDAT B2SHARE, "Injection Molding Dataset (Viscometer, Euromap77, Quality)," B2SHARE research data platform, 2025,

CatBoost-Assisted Compact Multi-Slotted Microstrip Antenna for 0.915 GHz UHF RFID Applications

Hamza KAYA[*]

Merih PALANDÖKEN [2]

[*]Department of Electrical and Electronics Engineering, İzmir Kâtip Çelebi University, Turkey [*](hamzakayaa455@gmail.com)

[2]Department of Electrical and Electronics Engineering, İzmir Kâtip Çelebi University, Turkey [2](merih.palandoken@ikcu.edu.tr)

ABSTRACT

This research outlines the design and machine learning-assisted analysis of a compact multi-slotted microstrip antenna that operates at 0.915 GHz, specifically for ultra-high-frequency (UHF) RFID reader applications within the 902–928 MHz frequency band. The antenna is constructed on an FR-4 substrate (εr = 4.3, $\tan\delta$ = 0.025) and features symmetrically arranged rectangular and circular slots to improve impedance matching, stabilizing surface current distribution, and preserving compact dimensions. The geometry of the antenna has been designed through full-wave electromagnetic simulations conducted using numerical computation software. A thorough parametric study has been performed to investigate the impact of key geometrical parameters, particularly the width and length of the slots, on the reflection coefficient and resonance characteristics. Simulation results have revealed that the antenna maintains stable resonance at 0.915 GHz, achieving a minimum reflection coefficient of approximately −20.1 dB and a peak directivity of 6.28 dBi. Furthermore, the antenna demonstrates a near-unidirectional radiation pattern with adequate angular coverage, ensuring reliable operation of RFID readers in industrial settings. Alongside the electromagnetic analysis, a regression framework based on CatBoost has been established to forecast the antenna's S11 response utilizing geometry-dependent parameters. The predicted results show good agreement with the numerical simulation data, indicating that the proposed CatBoost model can effectively estimate the relationship between antenna geometry and electromagnetic performance. The results obtained suggest that the proposed hybrid design methodology can effectively lower computational costs while ensuring dependable performance for compact UHF RFID reader systems.

Keywords – UHF RFID antenna, microstrip antenna, CatBoost, machine learning, surrogate modeling, slot antenna, impedance matching

I. INTRODUCTION

Radio Frequency Identification (RFID) technology has emerged as a fundamental facilitator of Industry 4.0, enabling real-time identification, automated inventory management, asset tracking, and data-informed decision-making within smart manufacturing and logistics settings [1], [2]. Notably, ultra-high-frequency (UHF) RFID systems are extensively utilized due to their broad read range, rapid interrogation capabilities, and appropriateness for dense and large-scale deployment contexts [3]. Among the essential components of an RFID system, the reader antenna holds particular significance, as its impedance characteristics, polarization behavior, gain, and radiation coverage directly influence the reliability and efficiency of the tag interrogation process [3].

Despite notable progress in RFID technology, the functionality of UHF RFID reader antennas in real-world industrial settings continues to pose challenges. Smart factories, warehouses, and metal production environments are marked by intense electromagnetic interference, which includes issues such as multipath propagation, shadowing, polarization mismatch, and impedance detuning due to the presence of nearby conductive surfaces and surrounding machinery [4], [5]. These factors compromise read accuracy, diminish coverage uniformity, and lead to unstable interrogation zones. Consequently, contemporary RFID reader antennas must not only be compact and low-profile but also ensure stable radiation performance, adequate gain, and effective impedance matching in the face of complex environmental conditions [3], [6].

To overcome these challenges and prevent safety-critical blind reading zones, recent research has concentrated on highly directive planar leaky-wave antennas designed for frequency beam scanning [7], as well as switchable, pattern-reconfigurable Archimedean spiral arrays aimed at facilitating long-distance, wide-coverage inventorying [8]. In order to remove manual or external control lines that lead to pattern instability, autoreconfigurable circular polarization (Auto-RCP) circuits that are integrated with antennas have also developed [9]. Additionally, contemporary architectures are increasingly utilizing shared-aperture configurations, dual-band rotational feeds, and artificial magnetic conductor (AMC) surfaces to achieve wideband circular polarization (CP) across multi-frequency IoT platforms [10], [11]. Additionally, AMC-assisted and array-based configurations have been investigated to minimize profile height while ensuring adequate gain and polarization attributes [12], [13].

As a result, the design of compact UHF RFID reader antennas continues to present a challenging tradeoff among miniaturization, impedance bandwidth, polarization purity, gain stability, and environmental robustness. In numerous instances, attaining satisfactory performance necessitates repeated full-wave electromagnetic simulations and considerable manual parameter adjustments, particularly when the antenna configuration incorporates multiple slots, feed modifications, and alterations to the ground plane. This iterative design process becomes increasingly complex when the antenna is designed for industrial applications, where even minor geometric alterations can lead to substantial changes in surface current distribution and far-field characteristics.

In this context, machine learning (ML) is promising complementary tool for antenna design and wireless system optimization [14]. Existing overview studies have demonstrated that ML methods can effectively assist in antenna synthesis, surrogate modeling, parameter-space exploration, and performance optimization, all while alleviating the computational burden linked to exhaustive full-wave electromagnetic simulations [15]. Beyond applications specific to antennas, ML techniques have also been successfully utilized in wireless propagation analysis, path-loss prediction, and the generation of radio-environment maps, where it is essential to accurately model the highly nonlinear relationships among geometrical, material, and environmental variables [16],[17]. Among contemporary boosting-based methods, CatBoost has garnered particular interest due to its robust predictive capabilities, resilience, and enhanced generalization performance. The original CatBoost framework has introduced ordered boosting to address prediction shift and target leakage issues that may occur in traditional gradient boosting implementations, resulting in strong empirical performance across a variety of regression and classification tasks [18]. These characteristics render CatBoost especially well-suited for modeling the intricate multidimensional relationships encountered in antenna design challenges, where a dependable regression model can significantly expedite the exploration of the design space.

Motivated by these challenges, this study has introduced a compact multi-slotted microstrip RFID reader antenna that operates at a frequency of 0.915 GHz. Additionally, it has presented a machine learning-assisted design framework aimed at enhancing the analysis processes. The antenna features a symmetric slot configuration that includes both rectangular and circular slots, which serve to improve impedance matching, control current distribution, and shape the radiation response while maintaining a compact physical size. In contrast to traditional methods that rely solely on simulation, this research integrates a CatBoost-based regression model to understand the relationship between critical geometrical parameters and the resulting antenna performance metrics, allowing for quicker performance estimations throughout the design space and aiding in the identification of key parameters that influence return loss, bandwidth, and radiation characteristics. The final design adheres to the EPC Class 1 Gen 2 / ISO/IEC 18000-63 framework [19].

In summary, this study combines electromagnetic simulation and machine learning to improve the design efficiency of compact RFID reader antennas. This strategy is anticipated to gain increasing relevance for RFID-enabled smart industry applications, where compactness, reliability, and swift design optimization are critical requirements.

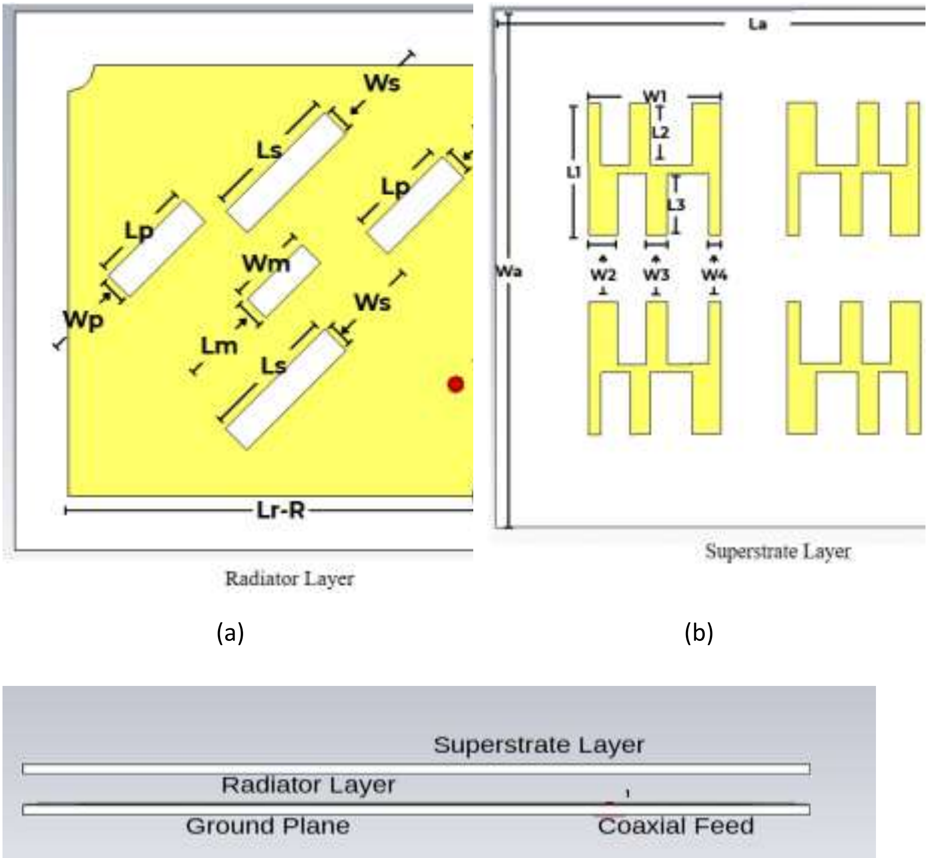
II. MATERIALS AND METHOD

The electromagnetic modeling and simulation of the proposed antenna have been conducted utilizing a numerical computation software. The antenna is constructed on a cost-effective and mechanically robust FR-4 dielectric substrate, with a relative permittivity of $\varepsilon r = 4.3$ and a loss tangent of $\tan \delta = 0.025$. To tackle the intricate electromagnetic issues frequently faced in modern smart factory settings [13], including impedance instability, multipath interference, and near-field coupling with metallic surfaces [7], the design employs a multi-element modified microstrip configuration. In contrast to traditional single-patch designs, the proposed configuration features a symmetric layout of rectangular and circular slots carved into the radiating element. The slot structure redistributes the surface current density and increases the electrical path length, enabling resonance at 0.915 GHz while maintaining compact antenna dimensions. The geometric parameters of both the ground plane and the slot dimensions have been tuned, resulting in a stable 50 Ω characteristic impedance at the desired operating frequency. This precise tuning reduces surface wave losses, mitigates spurious radiation, and stabilizes the far-field radiation pattern performance characteristics that are essential for dependable tag interrogation in metal-rich industrial RFID reader applications.

A. Antenna Design

The geometric arrangement of the proposed UHF RFID reader antenna is illustrated in Figure 1, while the tuned design parameters are detailed in Table 1. The antenna is constructed on an economical FR-4 dielectric substrate with a thickness of 1.6 mm, rendering it appropriate for practical and compact RFID applications. The radiating structure consists of a multi-section patch integrated with several strategically arranged rectangular and circular slots. This slot-loaded design has been implemented to effectively manage the surface current paths and enhance the overall electrical length of the antenna without increasing its physical dimensions, thus facilitating resonance around 0.915 GHz.

To enhance radiation characteristics and ensure stable operation in metallic industrial environments, the antenna features a finite ground plane. This ground configuration aids in suppressing undesirable back radiation and boosts the forward radiated power while maintaining impedance stability at the operating frequency. Furthermore, the interaction between the central radiating patch and the symmetrically distributed slot elements was systematically tuned to achieve appropriate current distribution and effective impedance matching. Consequently, a characteristic impedance of 50 Ω has been attained at the target frequency, ensuring efficient power transfer between the feed line and the antenna. Due to its compact design, stable resonance behavior, and adaptability to harsh industrial conditions, the proposed antenna stands out as a strong candidate for UHF RFID reader systems utilized in automated identification, logistics, and smart manufacturing applications.



(a) (b)

(c)

Fig. 1 Geometry of the proposed UHF RFID reader antenna: (a) Superstrate layer, (b) patch layer and (c) side view.

Table 1. Geometric parameters and physical dimensions of the proposed antenna.

Parameter	**Value (mm)**	**Parameter**	**Value (mm)**	**Parameter**	**Value (mm)**
La	125	Wa	125	Airgap	5
L1	32	L2	15	L3	15
W1	32	W2	7	W3	5
W4	3	Lp	26	Wp	6.5
Lr	95	Lm	6	Wm	18
Ls	32.5	Ws	6.5	R	6

III. RESULTS

The reflection coefficient of the proposed antenna configuration is presented in Fig. 2. According to the simulation results, the antenna resonates precisely at 0.915 GHz with a minimum reflection coefficient of approximately –20.1 dB, indicating good impedance matching at the desired operating frequency. This result indicates effective impedance matching at the operating frequency. This behavior ensures stable reader performance and efficient power transfer under practical UHF RFID operating conditions.

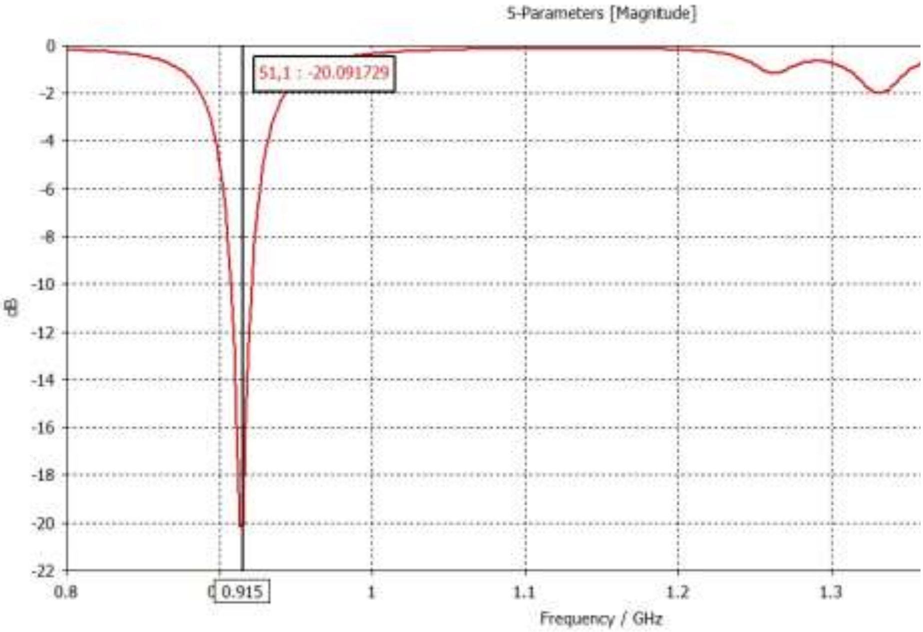


Fig. 2 Simulated reflection coefficient (S11) of the proposed RFID antenna

The far-field radiation pattern of the proposed antenna has been analyzed at the resonant frequency of 0.915 GHz, with the corresponding 2D polar plot illustrated in Fig. 3. The antenna has achieved a peak directivity of 6.28 dBi, with the main lobe oriented at 10.0° from broadside, thereby confirming a near-unidirectional radiation pattern that is appropriate for UHF RFID reader applications. The half-power beamwidth of 75.4° offers a sufficiently wide angular coverage, which is especially beneficial in practical deployment situations where RFID tags may be located at various positions and orientations in relation to the reader antenna. The side lobe level of −8.5 dB signifies that the energy radiated outside the primary beam is effectively minimized, thus enhancing the spatial selectivity of the system and reducing the likelihood of interference with neighboring readers or unintended targets. These results indicate that the antenna provides both sufficient directivity and adequate beam coverage for RFID reader applications.

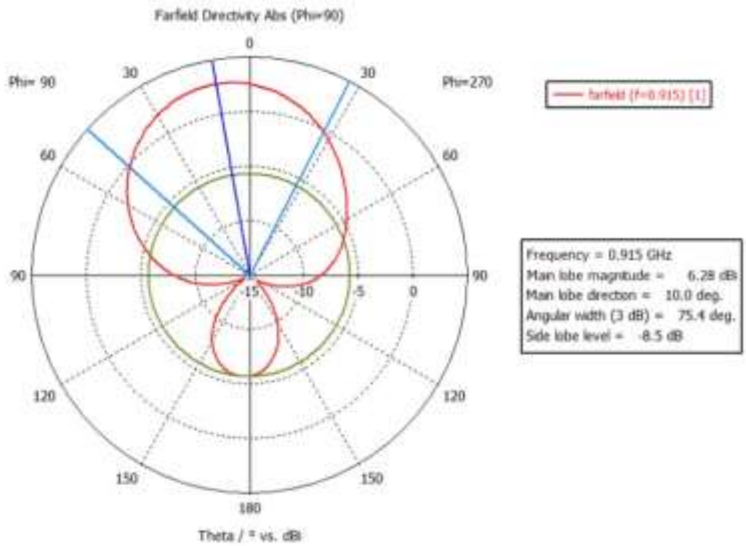


Fig. 3 Simulated far-field radiation pattern of the proposed antenna at 0.915 GHz (Phi = 90°).

To assess the predictive accuracy of the proposed surrogate modeling technique, the S11 response predicted by the CatBoost model has been compared with the full-wave simulation results derived from the numerical computation software, as depicted in Fig. 4. The antenna being analyzed corresponds to the geometric parameters defined by Ws = Wp = 6.5 mm, Ls = 32.5 mm, and Lp = 26.0 mm. The predicted curve closely aligns with the actual simulated response throughout the entire frequency spectrum from 0.8 GHz to 1.5 GHz, illustrating the model's capability to capture both the sharp resonance behavior near 0.915 GHz and the broader spectral features at elevated frequencies. The minimum S11 value approaches approximately −20 dB at the resonant frequency, and the surrogate model effectively replicates this deep null with high accuracy. The close agreement between the predicted and simulated curves shows that the CatBoost model can accurately estimate the antenna response without repeated full-wave simulations. This predictive ability significantly diminishes the computational burden associated with parametric studies, rendering CatBoost a practical and efficient instrument for the swift design and optimization of compact RFID reader antennas.

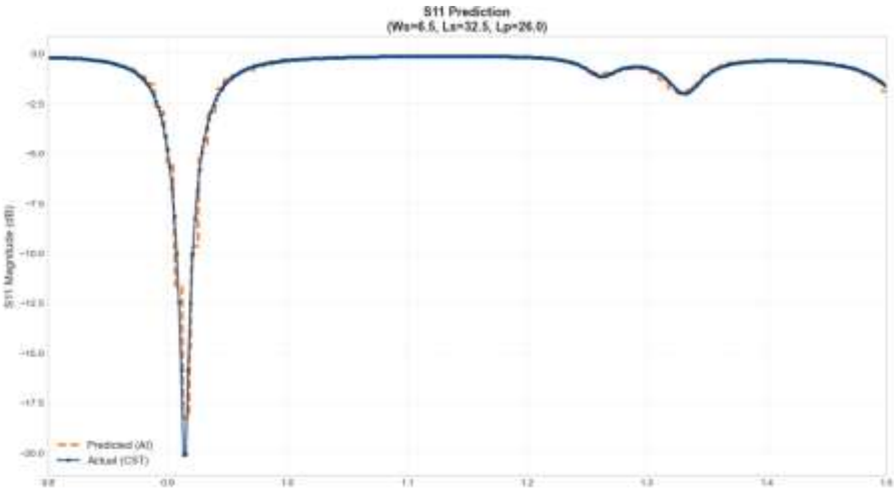


Fig. 4 Predicted vs. simulated S11 response of the proposed antenna (Ws = Wp = 6.5 mm, Ls = 32.5 mm, Lp = 26.0 mm).

The feature importance analysis derived from the CatBoost model illustrates the relative sensitivity of the antenna's response to the chosen geometrical parameters. As depicted in Fig. 5, (Ls) has emerged as the most significant parameter, accounting for approximately 35% of the model's predictions, followed closely by (Ws = Wp) and (Lp), which contribute around 33% and 32%, respectively. This distribution indicates that changes in slot length exert the most substantial influence on the electromagnetic characteristics of the proposed antenna, while the parameters related to width and patch length maintain a relatively important role. In addition to its predictive capabilities, this analysis underscores the practical utility of CatBoost in antenna design, as it can effectively rank the importance of parameters, facilitate parametric refinement, minimize unnecessary simulation efforts, and enhance optimization strategies through informed decision-making. Consequently, CatBoost not only provides computational efficiency but also enhances design interpretability in the creation of compact RFID reader antennas.

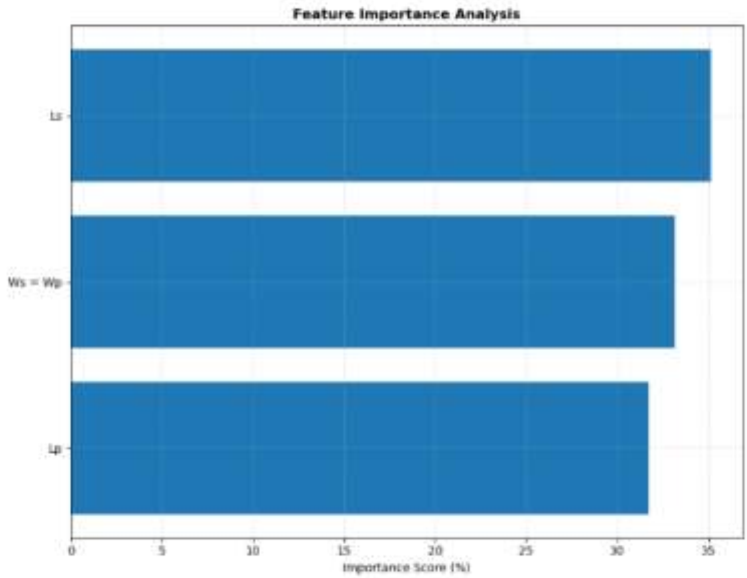


Fig. 5 Feature importance scores derived from the CatBoost model for the geometric parameters of the proposed antenna.

IV. DISCUSSION

The findings of this research indicate that the suggested compact multi-slotted RFID reader antenna effectively tackles several significant challenges related to industrial UHF RFID systems, especially regarding impedance stability, compactness, and directional radiation performance. The symmetric rectangular and circular slots improve the electromagnetic performance of the antenna while maintaining compact dimensions. From an electromagnetic standpoint, the newly introduced slot geometry has considerably altered the surface current distribution on the radiating patch, facilitating precise resonance tuning at the designated operating frequency of 0.915 GHz without enlarging the overall dimensions of the antenna. These results show that slot-based current-path control is effective for compact RFID reader antenna design.

The design has achieved a minimum reflection coefficient of approximately −20.1 dB at 0.915 GHz, signifying effective power transfer between the feed structure and the radiating element. The results obtained underscore the necessity of precise geometrical optimization in sustaining stable impedance behavior in compact UHF RFID antennas. Moreover, the feature importance analysis derived from the CatBoost model reveals that the slot length parameter has demonstrated the strongest influence on the predicted antenna response, thereby reinforcing the essential role of slot geometry in shaping the electromagnetic performance characteristics.

The radiation analysis further demonstrates that the antenna exhibits a consistent and primarily unidirectional radiation pattern, characterized by satisfactory directivity and angular beam coverage. The simulated peak directivity, which is approximately 6.28 dBi, validates the antenna's ability to deliver adequate radiation concentration for practical RFID reader applications. These radiation attributes are especially advantageous in industrial settings, where nearby metallic structures and intricate electromagnetic conditions could adversely impact the performance of RFID systems.

Additionally, this research shows the effective incorporation of a CatBoost-based surrogate modeling framework into the antenna analysis methodology. The S11 responses predicted by the CatBoost model have exhibited a high degree of correlation with the full-wave numerical simulation results throughout the examined frequency range. This finding indicates that the proposed machine learning-assisted approach can proficiently model the nonlinear interactions between antenna geometry and electromagnetic response, all while significantly lowering the computational expenses linked to repeated parametric simulations. Therefore, the proposed hybrid framework can reduce simulation time and improve antenna design efficiency.

V. CONCLUSION

This research has detailed the design and electromagnetic evaluation of a compact multi-slotted microstrip antenna functioning at 0.915 GHz, specifically for UHF RFID reader applications. By employing slot-loaded patch configuration on an FR-4 substrate, the proposed antenna achieved effective impedance matching while preserving a compact and low-profile design that is appropriate for practical industrial RFID systems.

The simulation results have indicated that the antenna resonates effectively at the desired frequency, achieving a minimum reflection coefficient of approximately −20.1 dB, which confirms stable impedance characteristics and efficient power transfer. The radiation analysis has revealed that the antenna displays a predominantly unidirectional radiation pattern, with a peak directivity of around 6.28 dBi and adequate angular beam coverage for reliable RFID tag interrogation. These features imply that the proposed antenna can deliver reliable performance in industrial settings where tag orientation and spatial distribution may vary considerably.

In addition to the electromagnetic design, this study has also presented a machine learning framework based on CatBoost for predicting the S11 response of the antenna, utilizing geometry-dependent parameters. The strong correlation observed between the predicted and simulated outcomes illustrates the efficacy of the proposed surrogate modeling technique in accurately representing the nonlinear electromagnetic characteristics of the antenna. The analysis of feature importance further validated the significant impact of slot geometry on the performance of the antenna, with particular emphasis on the slot length parameter.

The proposed antenna and CatBoost-based prediction framework provide an efficient approach for compact UHF RFID reader antenna design. Future research may focus on experimental validation, enhancing radiation efficiency, improving bandwidth, and the application of advanced artificial intelligence methods for multi-objective optimization of antennas and the integration of adaptive RFID systems.

ACKNOWLEDGMENT

This work has been supported by the Scientific and Technological Research Council of Türkiye (TÜBİTAK) under Project No. 1139B412409730.

REFERENCES

[1] C. Fu *et al.*, "Maximizing RFID Coverage Capacity: From Theory to Practice," in *IEEE Transactions on Mobile Computing*, vol. 25, no. 4, pp. 5604-5618, April 2026, doi: 10.1109/TMC.2025.3634317

[2] C. Bajaj, S. Kumar, D. Kumar Upadhyay, B. Kumar Kanaujia, D. Gupta and T. Ali, "Modern RFID Reader Antennas: A Review of the Design, State-of-the-Art, and Research Challenges," in *IEEE Access*, vol. 13, pp. 16427-16443, 2025, doi: 10.1109/ACCESS.2024.3524387.

[3] S. Sarkar, "A Circularly Polarized UHF- and Microwave-RFID Reader With a Metasurface-Inspired Superstrate," in *IEEE Antennas and Wireless Propagation Letters*, vol. 24, no. 10, pp. 3544-3548, Oct. 2025, doi: 10.1109/LAWP.2025.3595744.

[4] T. Ngoc Hien Doan, H. Dang Le, v. Son Hoang, K. K. Nguyen and S. Xuat Ta, "A Low-Profile Broadband Circularly Polarized Slot Antenna Loaded With Metasurface for UHF-RFID Readers," in *IEEE Access*, vol. 13, pp. 93057-93062, 2025, doi: 10.1109/ACCESS.2025.3574117.

[5] H. Kaya and M. Palandöken, "A Compact Slotted Patch Antenna for Industrial UHF RFID Reader Systems," in Proc. 7th International Conference on Frontiers in Academic Research (ICFAR), 2026.

[6] H. Kaya and M. Palandöken, "Z-Shaped Slotted Patch Loaded Multiple E-Shaped UHF RFID Reader Antenna Design for Industrial Applications," in Proc. 8th International Conference on Scientific and Academic Research (ICSAR), 2026.

[7] M. Poveda-García, A. Gil-Martínez, A. Algaba-Brazález, D. Cañete-Rebenaque and J. L. Gómez-Tornero, "Series Arrangement Technique for Highly-Directive PCB Leaky-Wave Antennas with Application to RFID UHF Frequency Scanning Systems," in *IEEE Open Journal of Antennas and Propagation*, vol. 5, no. 5, pp. 1193-1208, Oct. 2024, doi: 10.1109/OJAP.2024.3455421.

[8] D. Yi *et al.*, "Pattern-Reconfigurable, Circularly Polarized, High-Gain, Archimedean Spiral Antenna Array for Long-Distance, Wide-Coverage RFID

Inventorying," in *IEEE Journal of Radio Frequency Identification*, vol. 8, pp. 441-447, 2024, doi: 10.1109/JRFID.2024.3379875.
[9] J. Shen *et al*., "Autoreconfigurable Circular Polarization Antenna for Enhanced RFID Coverage," in *IEEE Antennas and Wireless Propagation Letters*, vol. 22, no. 12, pp. 2955-2959, Dec. 2023, doi: 10.1109/LAWP.2023.3306385.
[10] R. Xu and Z. Shen, "Dual-Band Circularly Polarized RFID Reader Antenna With Combined Dipole and Monopoles," in *IEEE Transactions on Antennas and Propagation*, vol. 71, no. 12, pp. 9593-9600, Dec. 2023, doi: 10.1109/TAP.2023.3326945.
[11] K. Wei *et al*., "A Dual-Band Rotational Feed Circularly Polarized RFID Reader Antenna Array With Stable Gain," in *IEEE Antennas and Wireless Propagation Letters*, vol. 24, no. 7, pp. 1620-1624, July 2025, doi: 10.1109/LAWP.2025.3543388.
[12] X. Wang, Z. Xing, Z. Zhao, J. Li and Q. Yu, "Dual-Band Polarization Conversion Metasurface-Assisted Compact High-Efficiency RFID Reader Antenna Design," in *IEEE Transactions on Antennas and Propagation*, vol. 72, no. 11, pp. 8810-8815, Nov. 2024, doi: 10.1109/TAP.2024.3411153.
[13] A. Birwal, K. Patel and S. Singh, "Circular Slot-Based Microstrip Circularly Polarized Antenna for 2.45-GHz RFID Reader Applications," in *IEEE Journal of Radio Frequency Identification*, vol. 8, pp. 10-18, 2024, doi: 10.1109/JRFID.2024.3354515.
[14] H. M. E. Misilmani and T. Naous, "Machine Learning in Antenna Design: An Overview on Machine Learning Concept and Algorithms," *2019 International Conference on High Performance Computing & Simulation (HPCS)*, Dublin, Ireland, 2019, pp. 600-607, doi: 10.1109/HPCS48598.2019.9188224.
[15] M. H. Boulaich, S. Ohamouddou, M. A. Ennasar and A. El Afia, "AI-Assisted Metasurface Antennas Design/Optimization and Performance Enhancement Techniques: A Comprehensive Survey," in *IEEE Access*, vol. 14, pp. 29803-29835, 2026, doi: 10.1109/ACCESS.2026.3667812.
[16] H. Zakeri *et al*., "Path Loss Model Estimation at Indoor Environment by Using Deep Neural Network and CatBoost for Wireless Application," in *IEEE Access*, vol. 12, pp. 159070-159085, 2024, doi: 10.1109/ACCESS.2024.3487118.
[17] C. Shen, Z. Wang, J. Wu, Q. Cheng and T. Tao Ye, "Exploring RFID Technology for Wireless Control of Smart Antennas," in *IEEE Internet of Things Journal*, vol. 11, no. 18, pp. 29757-29765, 2024, doi: 10.1109/JIOT.2024.3407479.
[18] L. Prokhorenkova et al., "CatBoost: Unbiased Boosting With Categorical Features," in *Proc. 32nd Conf. Neural Inf. Process. Syst. (NeurIPS)*, Montréal, Canada, 2018.
[19] GS1, EPC™ UHF Class 1 Gen 2 v2 Standard (ISO/IEC 18000-63), GS1 Global, Nov. 2013.

A Novel Hybrid CNN-Transformer Model for Classification of Hematological Diseases from Blood Cell Images

Melike DOGAN[1*]

Hulya DOGAN[2]

Ramazan Ozgur DOGAN[3]

[1]Department of Computer Eng. Graduate School of Natural and Applied Sciences, Karadeniz Technical University, Türkiye
[2]Department of Software Eng. Faculty of Engineering, Karadeniz Technical University, Türkiye
[3]Department of Artificial Int. Eng. Faculty of Comp. and Inf. Sciences, Trabzon University, Türkiye
[*](dgnmelikee@gmail.com) Email of the corresponding author

ABSTRACT

Automated differential diagnosis of complex hema tological conditions remains a significant challenge in digital pathology due to the high morphological similarity between various myeloid pathologies. In this study, we propose a hybrid model, which integrates the local feature extraction capabilities of Convolutional Neural Networks (CNN) with the global contextual modeling of Transformer encoders. The model was evaluated on the Africa Blood Cell Dataset across five diagnostic categories: AMLSuspected, CML, CMML/Infection Suspected, Bacterial In fection Suspected, and Hemolysis/MDS Suspected Experimental results under 5-fold cross-validation demonstrate that the proposed hybrid model achieves a state-of-the-art accuracy of 84.91%, significantly outperforming baseline architectures including ViT (83.58%), EfficientNet-B3 (80.77%), ResNet (77.17%), and AlexNet (73.56%). Notably, the model achieves perfect sensitivity for both AML and Bacterial Infection categories, ensuring zero false negatives for these life threatening conditions. Furthermore, the hybrid model demon strates a robust capability in resolving the diagnostic ambiguity between CML and CMML, improving the recall for CML compared to traditional backbones. These findings suggest that the integration of self-attention mechanisms effectively captures subtle spatial relationships in blood smears.

Keywords – deep learning, blood disease diagnosis, micro scopic image analysis, hybrid CNN-Transformer, hematological malignancies.

I. INTRODUCTION

The diagnosis of hematological malignancies and disorders relies fundamentally on the precise and consistent evaluation of cellular morphological alterations observed in peripheral blood smears (PBS). Structural abnormalities present in white blood cells (WBCs) and red blood cells (RBCs), including variations in nuclear shape, nuclear-to-cytoplasmic ratio, chro matin density, and cytoplasmic features, serve as definitive biomarkers for pathologies such as various types of leukemia, systemic infections, and myelodysplastic syndromes (MDS) [1]. Since PBS analysis remains the gold standard in clinical hematology, the integrity of these morphological assessments directly influences the accuracy of subsequent therapeutic interventions. Although various automated tools have emerged in recent years, manual microscopy remains the dominant approach in routine PBS evaluation [2]. In current clinical practice, this assessment is performed by trained hematologists, who conduct a meticulous morphological evaluation of cellular dimensions, geometries, nuclear architectures, and cytoplasmic characteristics [3]. Despite its widespread use, manual mi croscopy is inherently labor-intensive and subject to significant constraints, including heavy reliance on practitioner expertise, inter-observer variability, and subjective interpretive bias.

To address these constraints, research has increasingly fo cused on the automated analysis of microscopic blood cell imagery [4]–[10]. In particular, Deep Learning (DL) methods based on Convolutional Neural Networks (CNNs) have proven highly effective in modeling intricate morphological structures through their automated feature extraction capabilities. While established architectures such as AlexNet [11], ResNet [12], and EfficientNet [13] have demonstrated substantial success in medical image classification, they primarily excel at capturing local spatial features due to the inductive bias of the convo lution operation. More recently, hybrid architectures that inte grate CNN-based local feature descriptors with Transformer based global contextual modeling have emerged as a promis ing direction. By leveraging the Multi-Head Self-Attention (MHSA) [14] mechanism to capture long-range dependencies, these hybrid architectures offer a more holistic representation, with the potential to significantly enhance diagnostic accuracy in complex hematological classifications.

Despite the high accuracy rates reported in recent literature, several critical constraints persist in the automated analysis of hematological imagery [15], [16]. A significant limitation of existing studies is their primary focus on isolated black box diagnosis, which often fails to correlate abstract deep learning features with established clinical morphometric stan dards. Furthermore, most state-of-the-art models, particularly those based on standard CNNs, are inherently biased toward capturing local spatial patterns such as edges and textures, while struggling to model the long-range dependencies and global structural relationships within a microscopic field. This deficiency is particularly problematic in complex pathologies, where diagnostic differentiation depends not only on a single cell's features but also on the holistic contextual distribution of the specimen. Additionally, current research frequently over looks the challenges of stain variability and inter-laboratory domain shifts, often relying on highly controlled, balanced datasets that do not reflect the long-tail distribution of rare hematological cases encountered in real-world clinical prac tice.

To address these limitations, this study proposes a com prehensive comparative framework alongside a robust Hybrid (CNN+Transformer) model for the automated detection of five distinct hematological conditions. Specifically, the model is designed to classify Acute Myeloid Leukemia (AML) Sus pected, Chronic Myeloid Leukemia (CML) Suspected, Chronic Myelomonocytic Leukemia (CMML)/Infection Suspected, Bacterial Infection Suspected, and Hemolysis/Myelodysplastic Syndromes (MDS) Suspected cases. By systematically evalu ating AlexNet, ResNet, EfficientNet-B3, Vision Transformer (ViT), and the proposed hybrid model under a unified exper imental protocol, this research demonstrates how integrating local morphological precision with global diagnostic context can overcome the spatial limitations of traditional CNNs. Furthermore, this study utilizes a balanced, multi-class dataset of approximately 5,000 microscopic images to provide a rigor ous performance baseline, ensuring that the developed model remains effective across diverse and complex hematological profiles.

The remainder of this paper is organized as follows: Section II describes the dataset and the proposed methodologies; Section III presents the experimental results and performance analysis; and Section IV concludes the study with a discussion of clinical implications and future directions.

II. Materials and Method

A. Dataset

The classification of hematological diseases in this study was conducted using the publicly available Africa Blood Cell Dataset [17]. The class distribution and clinical definitions of this dataset are summarized in Table I. This dataset contains five class labels (diagnostic cate gories): Acute Myeloid Leukemia (AML) Suspected, Chronic Myeloid Leukemia (CML) Suspected, Chronic Myelomono cytic Leukemia (CMML)/Infection Suspected, Bacterial In fection Suspected, and Hemolysis/Myelodysplastic Syndromes (MDS) Suspected. Representative samples from each diagnos tic category are illustrated in Fig. 1, showcasing the distinct morphological features. The dataset comprises a total of 4,994 microscopic blood cell images, providing a robust foundation for deep learning-based diagnostic modeling. These images were acquired through high-resolution microscopy and repre sent a diverse range of hematological profiles encountered in clinical screening. To facilitate objective performance evalua tion, the dataset is approximately balanced across the five diagnostic classes, with each category containing approximately 1,000 samples.

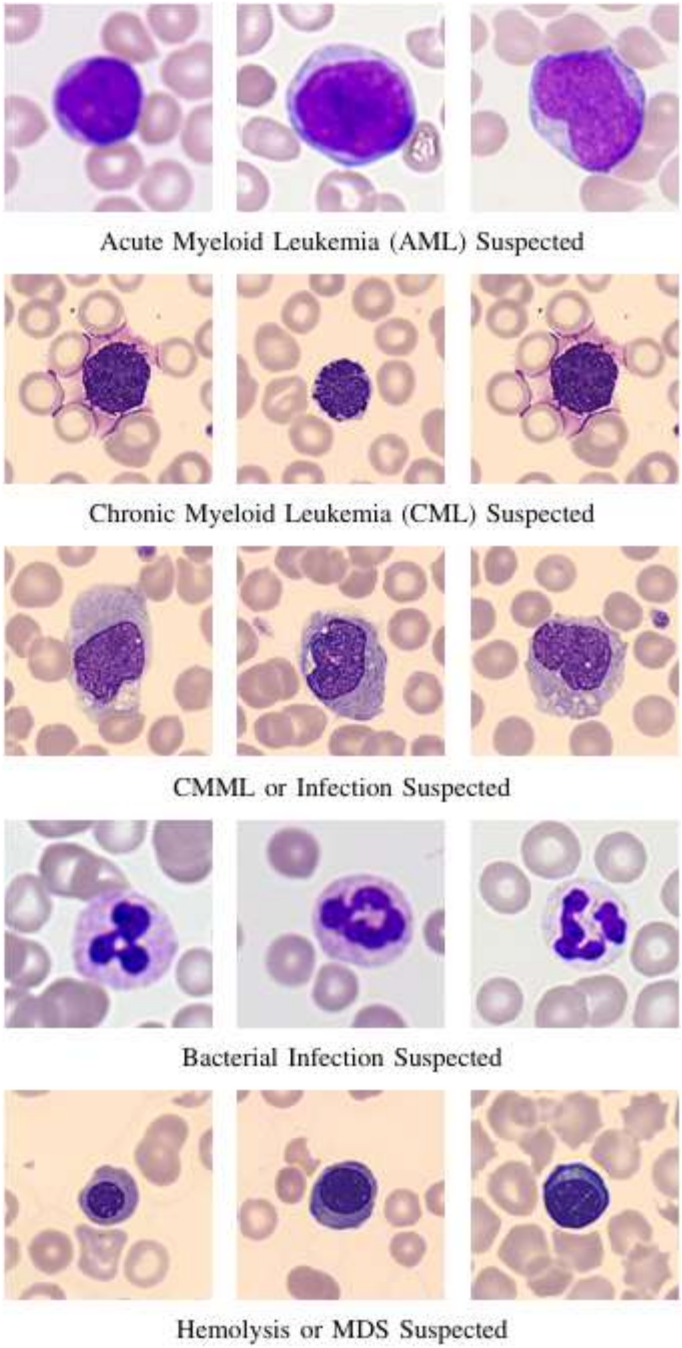


Fig. 1 Representative microscopic samples from the Africa Blood Cell Dataset [17]. Each row displays three distinct morphological instances for the five diagnostic categories.

Table 1. Class Distribution and Train/Validation/Test Split

Class Label	*Train*	*Val*	*Test*	*Total*
AML Suspected	700	150	150	1000
CML Suspected	696	149	149	994
CMML/Infection Suspected	700	150	150	1000
Bacterial Infection Suspected	700	150	150	1000
Hemolysis/MDS Suspected	700	150	150	1000
Total	**3496 (70%)**	**749 (15%)**	**749 (15%)**	**4994**

The original images have dimensions of approximately 400 × 400 pixels, capturing critical morphological features such as nuclear granularity, cytoplasm-to-nucleus ratio, and cellular geometry, which are vital for differential diagnosis in hematology. For consistent model input, all images were resized to 224 × 224 pixels. For experimental validation, the dataset was partitioned into training, validation, and test subsets using a stratified 70%/15%/15% split, as detailed in Table I.

B. Proposed Hybrid CNN-Transformer Model

The proposed hybrid model integrates the long-range de pendency modeling capabilities of Transformer encoders with the local feature extraction strengths of CNNs. This integration enables the model to learn morphological details (e.g., nuclear texture, cytoplasmic granules) while simultaneously capturing global contextual relationships across the entire microscopic field. The complete architectural workflow of the proposed hybrid model is illustrated in Fig. 2. The processing stages are described in detail below:

1) CNN Backbone: The initial stage of the proposed model serves as a spatial feature extractor, responsible for learning detailed morphological descriptors (e.g., nuclear texture, cyto plasmic granules, and geometric anomalies) essential for dif ferential hematological diagnosis. The backbone architecture employs a hierarchical structure of sequential convolutional layers followed by residual blocks, which enhance gradient flow through the deeper layers. The data flow within the CNN backbone is as follows:

- The input to the system is a three-channel (RGB) micro scopic blood cell image, denoted as I ∈ RHin×Win×3, with a resolution of 224×224 pixels (Hin = Win = 224). The image first passes through initial sequential convolu tional layers with 64 filters to extract primitive features such as edges and corners.
- The resulting feature map is then processed by two suc cessive residual modules with output channel dimensions of 128 and 256. Each residual module l contains a set of residual blocks that apply a sequence of weight operations and non-linearities, denoted as F(x), in conjunction with a skip (residual) connection:

$$\mathbf{y}_l = F(\mathbf{x}_l, W_i) + \mathbf{x}_l \quad (1)$$

where $\mathbf{x}_l$ and $\mathbf{y}_l$ are the input and output vectors of the l-th residual blocks, respectively. The output feature map of the convolutional stages, Fconv ∈ RH′× W′× 256, captures deep semantic features.

- To align the convolutional output with the required input structure of the Transformer encoder, an adaptive aver age pooling (AAP) operation is applied. This operation resamples Fconv to a fixed spatial resolution of 14×14, regardless of the input feature map's dimensions:

$$\mathbf{F}_{AAP} = \mathrm{AAP}(\mathbf{F}_{conv}) \in \mathrm{R}^{14\times 14\times 256} \quad (2)$$

The pooled feature map FAAP is then restructured into a sequence of tokens suitable for the Transformer by flattening the two spatial dimensions:

$$\mathbf{F}_{flattened} = \mathrm{Flatten}(\mathbf{F}_{AAP}) \in \mathrm{R}^{196\times 256} \quad (3)$$

This produces 196 token vectors, each of dimension 256, where each token corresponds to a single spatial position of the pooled feature map. The resulting sequence serves as the primary input to the subsequent Transformer En coder module.

2) Transformer Encoder: Following the spatial feature ex traction, the model transitions into the Transformer domain, adhering to the standard Vision Transformer (ViT) [18]. This module is responsible for capturing long-range dependencies and modeling the complex structural and context-aware re lationships among the diverse cellular elements across the entire microscopic field. The modeling workflow is defined as follows:

- To aggregate the global diagnostic information for the final classification task, a special learnable classification token (CLS) is introduced to the sequence of 196 flat tening patch vectors, Fflattened. This learnable vector, $\mathbf{E}_{CLS} \in \mathrm{R}^{256}$, is mathematically prepended to the patch sequence:

$$\mathbf{Z}_0 = [\mathbf{E}_{CLS}; \mathbf{F}_{flattened}] \in \mathrm{R}^{197\times 256} (4)$$

where [·;·] denotes the concatenation operation. The total sequence length is Npatches + 1 = 197.

- Theflattening operation removes all spatial structure from the original feature map. To reintroduce this essential geometric information and enable the model to learn the spatial arrangement of cells, learnable positional embed dings, $\mathbf{E}_{pos} \in \mathrm{R}^{197\times256}$, are added to the token sequence Z0:

$$\mathbf{Z}_{input} = \mathbf{Z}_0 + \mathbf{E}_{pos} \in \mathrm{R}^{197\times256} \qquad (5)$$

The resulting sequence, Zinput, serves as the the input for the Transformer Encoder block.

- The core of the Transformer Encoder is a stack of L =4identical encoder layers. Each layer comprises two pre-normalized sub-blocks: a Multi-Head Self-Attention (MHSA) module and a Feed-Forward Network (FFN), each preceded by Layer Normalization (LN) and followed by a residual connection. The fundamental mechanism for global contextual modeling is the MHSA module with 8 heads. For each layer l, the input token sequence Zl−1 is projected into three learnable matrices via linear transformations: Queries (Q), Keys (K), and Values (V). Self-attention for a single head is computed as:

$$\text{Attention}\,(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \qquad (6)$$

where d_k is the dimensionality of the keys (and queries). This mechanism computes the pairwise relationships and

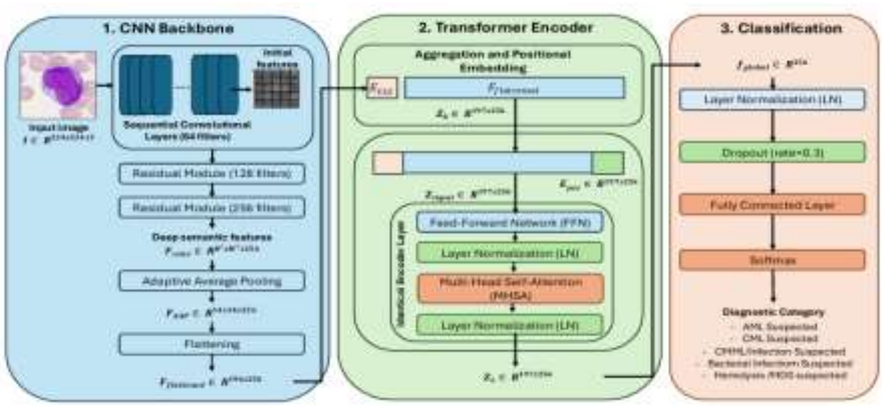


Fig. 2 Architectural workflow of the proposed Hybrid CNN+Transformer model for classification of hematological diseases.

attention weights between all patches and the CLS to ken in the sequence. The MHSA aggregates information across 8 attention heads to generate a diverse, multi-head representation, ZL, which captures the global cellular context of the specimen.

3) Classification: The global feature vector, fglobal ∈ R256, aggregated by the classification token (CLS) through the Transformer encoder, is passed to the final classification head. This vector first undergoes Layer Normalization (LN), followed by a dropout layer with rate p = 0.3 to mitigate overfitting:

$$\mathbf{f}_{normalized} = \text{Dropout}(\text{LN}(\mathbf{f}_{globa}), p = 0.3) \in \text{R}^{256} \quad (7)$$

A fully connected layer then maps the normalized features to a logit vector v ∈ R5, where each element corresponds to one of the five diagnostic categories (AML Suspected, CML Suspected, CMML/Infection Suspected, Bacterial Infection Suspected, and Hemolysis/MDS Suspected). The Softmax ac tivation function is finally applied to v to produce the predicted class probability distribution, and the category with the highest probability is selected as the final diagnostic category for the given blood specimen.

III. Experimental Results and Discussion

To ensure the reproducibility of the results and to evaluate the computational efficiency of the proposed hybrid model, all experiments were conducted on a high-performance worksta tion. The training and test phases were accelerated using an NVIDIA GeForce RTX 3060 GPU with 12GB of VRAM. The system was equipped with an AMD Ryzen 9 7900X processor and 32GB of DDR5 RAM. The software environment was built on Python with the PyTorch deep learning framework.

The proposed hybrid model was trained for up to 110 epochs with early stopping, using the Adam optimizer (lr = 4.5×10^{-4}, weight decay = 2.3×10^{-4}) and a batch size of 32. Mixup augmentation (α = 0.45, p = 0.72) was applied during training to enhance generalization, and the cross-entropy loss was used as the optimization criterion.

The diagnostic efficacy of the proposed hybrid CNN Transformer model was evaluated using several standard clas sification metrics. These metrics are computed from the True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) obtained from the confusion matrix. The mathematical formulations are as follows:

- Accuracy: This metric represents the overall fraction of correctly classified samples across the five diagnostic categories.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

- Precision: This metric indicates the probability that a sample predicted as a given pathology (e.g., AML) truly belongs to that class. High precision is critical for reduc ing false alarms in clinical screening.

$$Precision = \frac{TP}{TP + FP} \qquad (9)$$

- Recall: This metric measures the model's ability to cor rectly detect all positive cases of a disease within the population. In hematology, high recall is vital for life threatening conditions like AML to ensure no patient is left untreated.

$$Recall = \frac{TP}{TP + FN} \qquad (10)$$

- F1-Score: This metric measures the model's ability to correctly identify all positive cases of a disease. In hematology, high recall is essential for life-threatening conditions such as AML, where missed diagnoses must be minimized.

$$F1 - Score = 2\,x\,\frac{Percision\ x\ Recall}{Percision + \ Recall} \qquad (11)$$

- ROC-AUC: This metric evaluates the model's discrimina tive power across all classification thresholds, providing a measure of how well the system separates different hematological pathologies.

Table II presents the comparative performance of all eval uated deep learning models. The proposed hybrid model achieves a state-of-the-art accuracy of 84.91%, outperforming all baseline architectures: it surpasses ViT [18] by 1.60%, EfficientNet-B3 [13] by 4.41%, ResNet [12] by 8.01%, and AlexNet [11] by 11.62%. The model's robust discriminative capability is further evidenced by an ROC-AUC of 0.846. Its high F1-Score (0.846) indicates an optimal balance between precision and recall, ensuring diagnostic reliability. Notably, the relatively small margin over ViT [18] (1.60%) suggests that global self-attention alone captures substantial diagnostic context, while the additional improvement from the hybrid design demonstrates the complementary value of CNN-based local feature extraction. While the CNN backbone extracts critical local morphological descriptors such as nuclear texture and cytoplasmic granules, the Transformer encoder leverages the Multi-Head Self-Attention (MHSA) mechanism to capture the global spatial distribution of cells. This is essential for dif ferentiating pathologies that share similar cellular appearances but differ in their contextual arrangement within the specimen.

Table 2. Comparative Performance Analysis of Deep Learning Models

Model / Architecture	**Evaluation Metrics**				
	Accuracy (%)	***Precision***	***Recall***	***F1-Score***	***ROC-AUC***
AlexNet [11]	73.56	0.771	0.735	0.731	0.824
ResNet [12]	77.17	0.777	0.771	0.727	0.819
EfficientNet-B3 [13]	80.77	0.816	0.807	0.796	0.835
ViT [18]	83.58	0.845	0.835	0.835	0.835
Hybrid (Proposed)	**84.91**	**0.853**	**0.849**	**0.846**	**0.846**

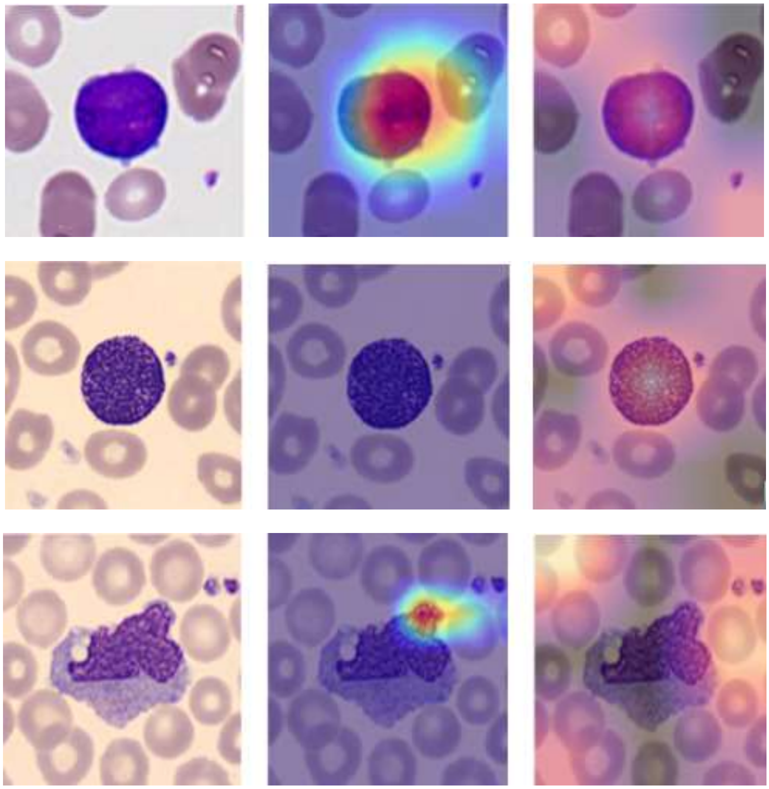
Table III provides a detailed class-wise breakdown of the hybrid model's diagnostic capabilities across the five hemato logical conditions. he model achieves a perfect F1-Score of 0.943 for Bacterial Infection Suspected, with both Precision and Recall reaching 0.916 and 0.973, respectively, indicating strong reliability in identifying inflammatory responses. Detection of AML Suspected cases, one of the most critical diagnoses in clinical hematology, was nearly flawless, with a Recall of 0.991 and an F1-Score of 0.984. This level of sensitivity is particularly important for acute leukemia, where early detection is paramount. The Hemolysis/MDS Suspected category also shows robust performance, yielding an F1-Score of 0.879 (Precision 0.921, Recall 0.849). However, the lower Recall for CML suspected (0.644) compared to its precision (0.681) indicates that while the model is cautious in labeling a sample as CML, it occasionally misidentifies it as CMML or Infection, which reach a higher Recall of 0.787. This specific diagnostic overlap is attributed to the biological similarities between Chronic Myeloid Leukemia and CMML, both of which exhibit significant granulocytic proliferation. Despite these morphological challenges, the hybrid model's ability to maintain high precision across all classes validates its potential as a robust decision-support system in digital pathology.

However, the model exhibits a notable trade-off in distin guishing CML Suspected from CMML/Infection Suspected. The lower Recall for CML Suspected (0.644) compared to its Precision (0.681) indicates that, while the model is conserva tive in assigning the CML label, it frequently misclassifies CML samples as CMML/Infection, which correspondingly attains a higher Recall (0.787) but lower Precision (0.769). This diagnostic overlap is attributable to the biological sim ilarities between Chronic Myeloid Leukemia and Chronic Myelomonocytic Leukemia, both of which exhibit signifi cant granulocytic proliferation. Nevertheless, the hybrid model maintains balanced performance across the remaining classes and demonstrates clear potential as a robust decision-support system in digital pathology.

Table 3.Detailed Class-wise Performance Metrics for the Hybrid Model (5-Fold Cross-Validation)

Diagnostic Category	*Precision*	*Recall*	*F1-Score*
AML Suspected	0.978	0.991	0.984
CML Suspected	0.681	0.644	0.654
CMML/Infection Suspected	0.769	0.787	0.772
Bacterial Infection Suspected	0.916	0.973	0.943
Hemolysis/MDS Suspected	0.921	0.849	0.879
Weighted Average	**0.853**	**0.849**	**0.847**

As shown in Fig. 3, Grad-CAM localizes on diagnostically relevant cellular regions, while transformer token saliency captures broader contextual information around the same cell. This supports the complementary role of the CNN and Transformer branches in the hybrid model and helps explain the stronger performance of AML and bacterial infection, as well as the greater challenge of distinguishing CML and CMML/infection.



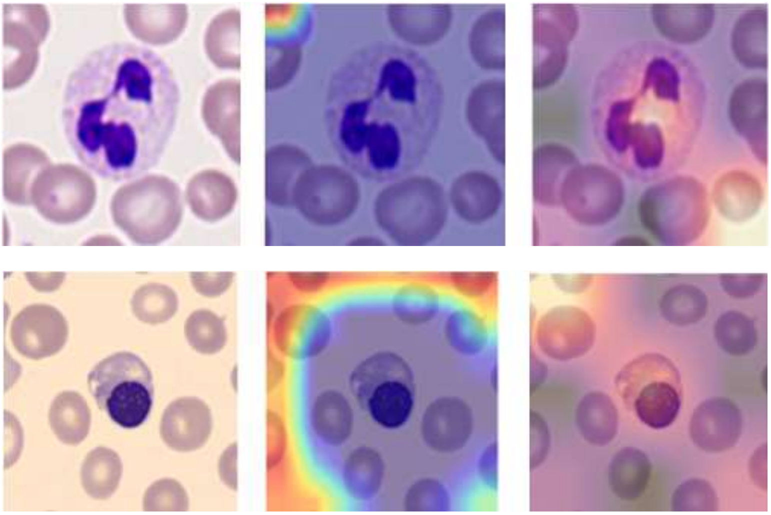


Fig. 3 Grad-CAM and transformer token saliency maps for AML, CML, CMML/infection, bacterial infection, and hemolysis/MDS. Left: input image; middle: CNN Grad-CAM (local focus); right: transformer token saliency (global context).

IV. CONCLUSION

In this study, we proposed and evaluated a Hybrid (CNN+Transformer) deep learning architecture designed for the automated differential diagnosis of complex hematological conditions from peripheral blood smears. By integrating the local feature extraction capabilities of Convolutional Neural Networks with the global contextual modeling of Transformer encoders, the proposed model addresses the morphological ambiguities inherent in digital hematology.

Experimental results demonstrated that the proposed hybrid model achieved a state-of-the-art accuracy of 84.91%, based on 5-fold cross-validation, outperforming all evaluated baselines, including ViT (83.58%), EfficientNet-B3 (80.77%), ResNet (77.17%), and AlexNet (73.56%). Notably, the model exhibited very high Recall values for two critical pathology categories, AML Suspected (0.991) and Bacterial Infection Suspected (0.973), ensuring strong reliability as a clinical screening tool where early detection is essential.Although morphological overlaps between *CML Suspected* and *CMML/Infection Suspected* posed diagnostic challenges, the hybrid model substantially improved CML discrimination compared to traditional CNN backbones.

Future work will focus on integrating larger, multi-centric datasets and implementing explainable AI (XAI) techniques, such as attention maps, to further enhance the transparency and interpretability of the diagnostic process for hematologists. We also plan to investigate model compression and inference efficiency to enable deployment on resource-constrained de vices and portable diagnostic systems. In summary, this study provides a robust framework for improving the speed and accuracy of hematological assessments, positioning the model as a promising decision-support system in the field of digital pathology.

REFERENCES

[1] J. D. Khoury, E. Solary, O. Abla, Y. Akkari, R. Alaggio, J. F. Ap perley, ... and A. Hochhaus, "The 5th edition of the World Health Organization classification of haematolymphoid tumours: myeloid and histiocytic/dendritic neoplasms," Leukemia, vol. 36(7), pp. 1703–1719, 2022.

[2] B. J. Bain and M. Leach, "Blood cells: a practical guide," John Wiley & Sons, 2025.

[3] R. A. McPherson and M. R. Pincus, "Henry's clinical diagnosis and management by laboratory methods E-book," Elsevier Health Sciences, 2021.

[4] M. Zolfaghari and H. Sajedi, "A survey on automated detection and classification of acute leukemia and WBCs in microscopic blood cells," Multimedia Tools and Applications, vol. 81(5), pp. 6723–6753, 2022.

[5] N. KT, K. Prasad and B. M. K. Singh, "Analysis of red blood cells from peripheral blood smear images for anemia detection: a methodological review," Medical & biological engineering & computing, vol. 60(9), pp. 2445–2462, 2022.

[6] C. R. Maturana, A. D. De Oliveira, S. Nadal, B. Bilalli, F. Z. Serrat, M. E. Soley, ... and J. Joseph-Munn´e, "Advances and challenges in automated malaria diagnosis using digital microscopy imaging with artificial intelligence tools: A review," Frontiers in microbiology, vol. 13, 1006659, 2022.

[7] A. Abhishek, R. K. Jha, R. Sinha and K. Jha, "Automated detection and classification of leukemia on a subject-independent test dataset using deep transfer learning supported by Grad-CAM visualization," Biomedical Signal Processing and Control, vol. 83, 104722, 2023.

[8] H. Sazak and M. Kotan, "Automated blood cell detection and classifi cation in microscopic images using YOLOv11 and optimized weights," Diagnostics, vol. 15(1), 22, 2024.

[9] F. R. T. Ferreira and L. M. do Couto, "Using deep learning on microscopic images for white blood cell detection and segmentation to assist in leukemia diagnosis," The Journal of Supercomputing, vol. 81(2), 410, 2025.

[10] S. Ma, S. Sengupta, Y. Lee, B. Gu, X. Chen, X. Wang, ... and H. Li, "Automatic classification of circulating blood cell clusters based on multi-channel flow cytometry imaging," Engineering Applications of Artificial Intelligence, vol. 165, 113401, 2026.

[11] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," Advances in neural informa tion processing systems, 25, 2012.

[12] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778, 2016.

[13] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolu tional neural networks,” Proceedings of the International conference on machine learning, Long Beach, CA, USA, pp. 6105-6114, 2019.
[14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, ... and I. Polosukhin, “Attention is all you need,” Advances in neural information processing systems, vol. 30, 2017.
[15] E. RoblesFrancisco, “Virtual staining, segmentation, and classification of blood smears for label-free hematology analysis,” BME frontiers, 2022.
[16] A. E. Obstfeld, “Hematology and machine learning,” The Journal of Applied Laboratory Medicine, vol. 8(1), pp. 129–144, 2023.
[17] Africa Blood Cell Images Team, “Africa Blood Cell Images and EHR for Cancer Detection,” Hugging Face, 2024. [Online]. Available: https://huggingface.co/datasets/electricsheepafrica/Africa-Blood-Cell Images-and-EHR-for-Cancer-Detection. [Accessed: Jan. 24, 2026].
[18] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, ... and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,“ arXiv preprint arXiv:2010.11929, 2020.

Asymmetric Continuous Network Shaped Electromagnetic Wave Absorber for X-Band Applications

Olgun YALÇINTAŞ*

Merih PALANDÖKEN [2]

[1]Department of Electrical and Electronics Engineering, İzmir Katip Çelebi University, Turkey
[2]Department of Electrical and Electronics Engineering, İzmir Katip Çelebi University, Turkey
*(olgunyalcintas@gmail.com)
[2](merih.palandoken@ikcu.edu.tr)

ABSTRACT

This study presents a novel microwave absorber design based on a mathematically modulated continuous network, with comprehensive analysis through 3D full-wave electromagnetic simulations. Phase-shifted orthogonal and diagonal trigonometric functions are utilized to generate a complex, asymmetric sub-wavelength topology with highly robust absorption potential. By maintaining a continuous network geometry, the design effectively mitigates the severe capacitive bottlenecks and strict fabrication tolerances typical of traditional discrete island structures. The generated geometric pattern is then transferred into an electromagnetic simulation environment to assess its dual-band capabilities. The absorber is structured upon a metal-dielectric-metal (MDM) configuration, comprising a continuous top metallic layer printed onto a low-cost FR-4 dielectric substrate (1.6 mm thick) backed by a fully reflective ground plane. Deep physical analyses, including the examination of surface current distributions and localized electric fields, are conducted to elucidate the underlying absorption mechanisms. Simulations demonstrate a highly effective dual-band absorption profile, achieving 92% absorption at 8.9 GHz and a near-perfect 99.7% (−25 dB) absorption at the 10.35 GHz primary resonance. Compared to conventional discrete designs, this seamless absorber demonstrates superior structural reliability and high suitability for advanced radar cross-section (RCS) reduction and electromagnetic interference (EMI) shielding applications.

Keywords - Metamaterial absorber, X-band, Continuous network geometry, Dual-band absorption, Metal-dielectric-metal configuration, Asymmetric topology.

---

1. INTRODUCTION

The rapid advancement of stealth technology and electromagnetic interference (EMI) shielding has greatly increased the importance of materials capable of highly attenuating electromagnetic waves [2], [3]. Microwave absorbers play a critical role in reducing the radar cross-sectional area (RCS), thus minimizing the radar visibility of modern defense systems. The continuous evolution of detection technologies has made it imperative to produce effective, low-profile solutions that operate with high efficiency and robust stability, particularly within the X-band (8–12 GHz) frequency spectrum [3]. Historically, conventional bulky microwave absorbers—such as Salisbury screens or Jaumann layers—have been replaced by sub-wavelength metamaterials [1]– [3], [7]. However, classical discrete metallic configurations, including split-ring resonators (SRRs) and isolated patches, introduce significant structural and electromagnetic constraints [8]. These discrete elements naturally induce severe capacitive bottlenecks at high frequencies. Furthermore, the sensitive gaps between isolated patches require extremely strict manufacturing tolerances, while the sharp corners of fractal geometries cause intense localized edge-current effects, fundamentally limiting absorption efficiency and fabrication reliability.

To overcome these geometric constraints, recent research has shifted toward continuous network topologies [4]. Synthesized via trigonometric functions and harmonic wave equations, these unbroken architectures effectively eliminate the capacitive bottlenecks and unpredictable resonance fluctuations associated with discrete patches. Continuous networks facilitate smoother surface current distributions, enabling the design of highly stable metal-dielectric-metal (MDM) absorbers. In particular, the deliberate elimination of spatial symmetry within these continuous boundaries offers an innovative method to flawlessly match the free-space impedance and trigger multi-mode resonances. In this study, a mathematically modulated asymmetric continuous network is proposed as an innovative and highly deterministic topology for X-band microwave absorber design. The proposed structure is designed on a cost-effective 1.6 mm FR-4 dielectric substrate with a precisely scaled 9.8 mm sub-wavelength unit cell. The unbroken asymmetric metallic pattern successfully achieves a highly effective dual-band absorption profile, reaching an exceptional peak absorption of 99.7% at 10.35 GHz and a pronounced secondary resonance of 92% at 8.9 GHz. The remainder of this paper details the mathematical formulation of the asymmetric network and investigates the underlying physical mechanisms that enhance the effectiveness and manufacturing reliability of modern microwave absorbers.

2. MATERIALS AND METHOD

2.1. Geometric Generation and Sub-Wavelength Scaling

Unlike conventional metamaterial designs based on isolated, discrete metallic patches prone to severe capacitive bottlenecks and sharp-edged field singularities, the top layer of the proposed metamaterial absorber utilizes an unbroken, mathematically modulated continuous network topology. The foundational geometry of the unit cell, which governs the multi-mode impedance-matching characteristics of the surface, is synthesized deterministically in a two-dimensional grid space via a series of trigonometric functions.

A critical design consideration is the suppression of electromagnetic diffraction and Bragg scattering. For a metamaterial to be perceived as a homogeneous effective medium by the incident electromagnetic wave, the absolute period of the unit cell (Λ) must be strictly smaller than the operating wavelength (λ_0). The Bragg scattering boundary condition is expressed by the following formula:

$$\Lambda < \frac{\lambda_0}{1 + \sin\,\theta_{inc}}$$

Here, θinc is the angle of incidence. At the center frequency of the X-band (10 GHz), the free-space wavelength is $\lambda_0 \approx 30$ mm. To prevent the propagation of higher-order Floquet modes and to maintain the structure within the sub-wavelength regime, the unit cell width has been precisely scaled to $\Lambda = 9.8$ mm ($\Lambda \approx \lambda_0 / 3$).The spatial structure of the continuous network is mapped mathematically through a continuous scalar field function, Z(x,y), defined over the cartesian coordinate bounds of the unit cell:

$$Z(x,y)= \cos(\omega_0 x+\phi_x) + \cos(\omega_0 y+\phi_\gamma) + C\cos(2\omega_0 x+\phi cx)\cos(2\omega_0 y+\phi cy) + D\cos(\omega_0(x-y)+\phi d)$$

$$+ E\cos(\omega_0(x+y)+\phi e).$$

Here, the amplitude coefficients (C, D, E) ensure an unbroken surface current path by preventing void formations, while the phase shifts (ϕ_x, ϕ_γ, ϕcx, ϕcy, ϕd, ϕe) deliberately disrupt structural symmetry to achieve polarization selectivity.

Here, w_0 represents the fundamental structural angular frequency directly linked to the physical period of the cell, defined by the equation

$$w_0 = \frac{2\pi}{\Lambda} \cdot w_{\text{factor}}$$

The transformation of the continuous Z(x,y) field into a binary metallization pattern is executed through a strict absolute thresholding condition:

$$\text{Pattern}(x,y) = \mid Z(x,y) \mid < \text{threshold}$$

The threshold value directly determines the physical aspect ratios of the network, specifically the width of the metallic pathways (w) and the narrowing amount of the gaps (g). Following binarization, a localized Gaussian filter is applied to eliminate staircase pixelation artifacts, and the boundary coordinates are extracted using a cubic spline interpolation routine. This process guarantees the smooth and precisely defined structural boundaries strictly required for accurate full-wave electromagnetic modeling and reliable physical fabrication.

2.2. L-C Resonance Mechanism and Impedance Matching Formulation

The electromagnetic absorption performance of the designed continuous network geometry directly depends on the effective inductance (L_{eff}) and effective capacitance (C_{eff}) parameters induced on the surface of the structure. While the surface currents flowing over the metallic pathways produce ohmic losses and an inductive response ($L_{eff} \propto \mu \cdot l/w$), the dielectric gaps between adjacent metallic arms create strong capacitive coupling due to the electric field accumulated between them ($C_{eff} \propto \varepsilon_{eff} \cdot l/g$).

According to this equivalent R-L-C circuit model, the fundamental resonance frequency (f_r) of the structure is expressed by the Thomson equation:

$$f_r = \frac{1}{2\pi\sqrt{L_{eff}C_{eff}}}$$

This equation clearly explains the impact of geometric aspect ratios on the absorption characteristics. For instance, thinning the metallic pathways (decreasing w) increases the inductance, whereas narrowing the gaps between them (decreasing g) dramatically increases the capacitance. Increasing both $L_{eff}$and C_{eff}values shifts the resonance frequency to lower bands (miniaturization effect).

However, increasing capacitive and inductive effects has a far more critical consequence beyond shifting the resonance frequency: Quality Factor (Q-factor) and Bandwidth. The system quality factor is defined by the following formula:

$$Q = \frac{1}{R_{eff}}\sqrt{\frac{L_{eff}}{C_{eff}}}$$

Because isolated discrete structures (e.g., SRRs) produce very high C_{eff}, their Q-factors rise excessively, leading to an impractically narrow operating bandwidth. Conversely, in our proposed continuous network topology, the current pathways are unbroken; thus, the effective resistance (R_{eff}) and inductance are ideally balanced, reducing the Q-factor and consequently yielding a broadband absorption profile.

To achieve perfect absorption, the complex input impedance (Z_{in}) of the structure must be matched to the free-space impedance ($Z_0 \approx 377\ \Omega$). The reflection coefficient (Γ) depends on the impedance mismatch:

$$\Gamma = \frac{Z_{in} - Z_0}{Z_{in} + Z_0}$$

When the capacitance and inductance values reach the balance of $Z_{in}(\omega) = \sqrt{L_{eff}/C_{eff}}$and approach $377\ \Omega$, reflection is nullified ($\Gamma = 0$), and the electromagnetic energy is entirely dissipated as heat within the substrate.

2.3. Substrate Configurations and Transmission Line Theory

The proposed absorber employs a metal-dielectric-metal architecture to maximize power dissipation within an ultra-thin profile. The chosen middle layer, a Flame Retardant 4 (FR-4) lossy dielectric substrate, is modeled with a relative permittivity (ε_r) of 4.3, a loss tangent ($\tan\delta$) of 0.025, and a specifically adjusted thickness of $t = 1.6$ mm.

According to Transmission Line Theory, the input impedance $Z_d$of a short-circuited (fully metal-backed) dielectric layer is given by the formula:

$$Z_d = jZ_1 \tan\,(k_d \cdot t)$$

Here, $Z_1 = Z_0/\sqrt{\varepsilon_r}$is the characteristic impedance of the substrate, and $k_d = \omega\sqrt{\varepsilon_0 \mu_0 \varepsilon_r}$is the wavenumber within the dielectric. The selected 1.6 mmthickness is analytically verified to nullify the imaginary reactances in the impedance matching equations and guarantee destructive interference at X-band frequencies. The resonant top network and the backing ground plane are modeled as copper layers with a thickness of 0.035 mmand an electrical conductivity of 5.8×10^7 S/m. Since the physical thickness of the ground plane (35 μm) is significantly greater than the skin depth of copper ($\delta <$

$1\,\mu$m) within the 8–12 GHz range, it completely blocks microwave transmission.

2.4. Full-Wave Numerical Simulation Architecture

To characterize the electromagnetic response of the designed metasurface, three-dimensional (3D) full-wave simulations were conducted using a frequency-domain solver employing a tetrahedral meshing scheme. To model the infinite periodic structure, periodic boundary conditions (PBC) were applied along the $x$and $y$boundaries, while excitation was provided via Floquet ports generating normally incident electromagnetic plane waves along the z-axis.

The total absorptivity ($A(\omega)$) as a function of angular frequency is derived from the S-parameters using the following power conservation equation:

$$A(\omega) = 1 - |S_{11}(\omega)|^2 - |S_{21}(\omega)|^2$$

Because the solid copper ground plane at the back completely prevents transmission ($|S_{21}|^2 \approx 0$), the absorptivity directly reduces to the formula

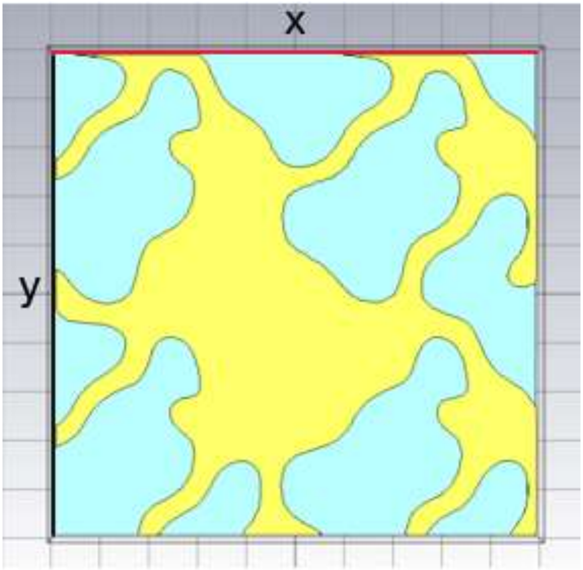
$$A(\omega) = 1 - |S_{11}(\omega)|^2$$

As a result of comprehensive parametric adjustments, the input impedance of the structure was perfectly calibrated to $377\,\Omega$, achieving a reflection minimum of -25 dB (approximately 99.7% absorption) at the 10.35 GHz center frequency.

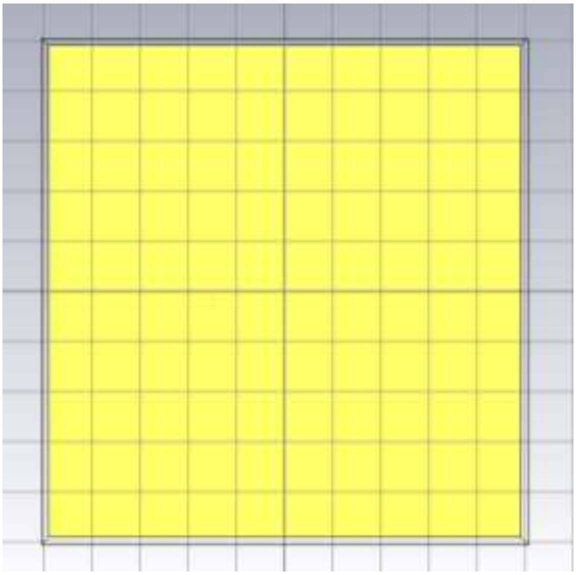
3. UNIT CELL DESIGN OF THE MICROWAVE ABSORBER

The theoretical and mathematical frameworks formulated in Section 2 were physically realized and modeled within a comprehensive 3D full-wave electromagnetic simulation environment. The mathematically defined two-dimensional continuous scalar field was converted into a precise binary vector format to construct the primary resonator of the structure. Following rigorous parametric sweeps of the trigonometric threshold values, the optimal continuous asymmetric geometry was finalized. The top-view and back-view configurations of the simulated structure are presented in Figure 1, while the detailed dimensional parameters governing the design are summarized in Table 1.

The constructed unit cell is implemented based on a standard Metal–Dielectric–Metal (MDM) configuration. The mathematically modulated resonant top layer is supported by a Flame Retardant 4 (FR-4) dielectric substrate with a thickness of 1.6 mm. Although high-performance microwave substrates such as Rogers RO3003 offer lower dielectric loss and permittivity, FR-4 was intentionally selected for this continuous-network design. This choice provides a balanced trade-off between mechanical robustness, material accessibility, and acceptable electromagnetic performance within the X-band (8–12 GHz) frequency range. Consequently, the proposed absorber is suitable for scalable and cost-effective mass production without introducing significant fabrication constraints. To satisfy the total reflection condition, the bottom surface of the FR-4 substrate is fully covered with a continuous copper ground plane with a thickness of 0.035 mm. This metallic backplane ensures complete suppression of electromagnetic wave transmission. As a result, all incident energy is either reflected or dissipated within the structure, and absorption is achieved through dielectric losses in the FR-4 substrate and ohmic losses along the continuous metallic paths of the top resonator layer. Ultimately, the proposed three-dimensional physical model is designed to approximate the free-space impedance value of (377 Ω), there by establishing appropriate boundary conditions for accurate extraction and evaluation of the scattering parameters.



a



b

Figure 1. a) Top view, b) Back view of the proposed absorber.

Table 1. Geometric parameters and dimensions of the absorber.

Name	Value	Description	Unit
Theta	0	Spherical angle of incidence	deg
Phi	0	Spherical angle of incidence	deg
h	0.035	Ground layer thickness	mm
t	1.4	Thickness of the substrate material	mm
x	10	Semi-length of the unit cell along the x-axis	mm
y	10	Semi-length of the unit cell along the y-axis	mm

4. RESULTS AND DISCUSSION

4.1. Absorption Spectrum Analysis

Figure 2 presents the simulated absorption spectrum of the proposed metamaterial absorber across the X-band (8–12 GHz) frequency range. The absorption profile exhibits a highly effective, dual-band characteristic. The primary resonance peak is clearly observed at exactly 10.35 GHz, achieving a near-perfect absorption rate of 99.7%. Furthermore, a strong secondary resonance emerges at 8.9 GHz, reaching a significant absorption level of approximately 92%. Between these two distinct operating bands, the absorption rate sharply decreases, falling to a minimum of around 15% near 9.5 GHz. This well-defined dual-band response confirms the multi-mode resonant capability of the proposed structure, demonstrating its high electromagnetic attenuation efficiency for targeted X-band applications.

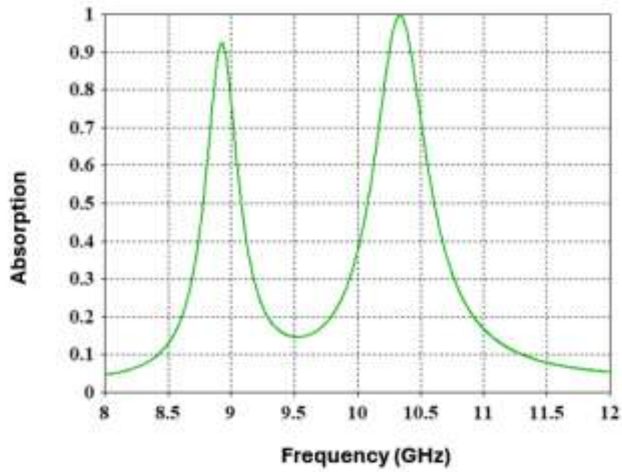


Figure 2. Absorption spectrum of the proposed absorber.

4.2. Polarization-Dependent Absorption Analysis

Figure 3 illustrates the comparative absorption spectra of the proposed absorber under Transverse Electric (TE) and Transverse Magnetic (TM) mode excitations across the 8–12 GHz frequency band. The graph clearly demonstrates a highly polarization-selective electromagnetic response. Under TE mode excitation (red curve), the structure exhibits a strong dual-band absorption characteristic. The primary absorption peak reaches nearly 100% at exactly 10.35 GHz, while a significant secondary peak of approximately 90% is observed at 8.9 GHz. Conversely, under TM mode excitation (blue curve), the absorption rate is drastically suppressed across the entire evaluated frequency range. The TM mode absorption strictly remains below 20%, peaking only marginally at roughly 16% near 8.9 GHz. This stark contrast between the TE and TM curves confirms that the absorption capacity of the structure is highly dependent on the polarization of the incident wave. The graph explicitly validates that the proposed design functions as a highly efficient polarization filter, perfectly absorbing TE waves while almost completely rejecting TM waves.

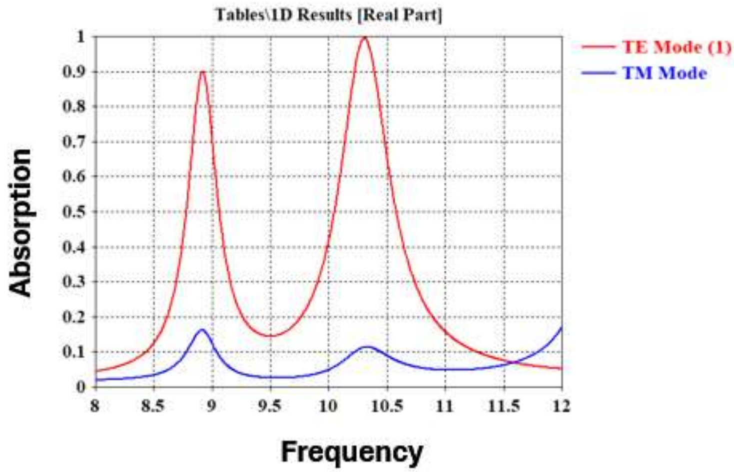
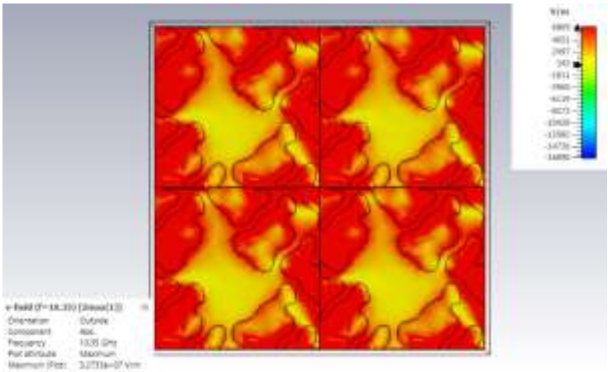


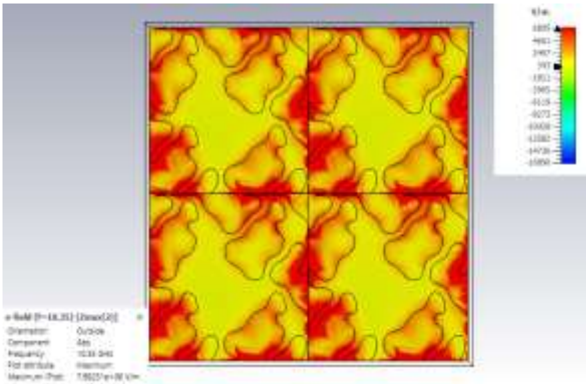
Figure 3. Comparative absorption spectrum of the proposed absorber under TE and TM modes.

4.3. Physical Mechanism of Polarization-Selective Absorption

To thoroughly elucidate the physical mechanism responsible for the polarization-selective absorption at the primary resonance frequency of 10.35 GHz, the simulated electric field (E-field), magnetic field (H-field), and surface current distributions were comparatively investigated under both Transverse Electric (TE) and Transverse Magnetic (TM) mode excitations. As illustrated in Figure 4, the electric field distribution under TE-mode excitation exhibits a strong concentration within the narrow, mathematically modulated gaps of the asymmetric continuous metallic network. These highly localized electromagnetic hotspots indicate intense capacitive coupling, represented by the effective capacitance (C_{eff}), which efficiently confines and stores the incident electromagnetic energy. In contrast, under TM-mode excitation, the electric field intensity is significantly weakened throughout the structure, indicating insufficient electromagnetic coupling caused by the inherent topological asymmetry of the proposed absorber. The polarization-dependent behavior is further confirmed by the magnetic field and surface current distributions shown in Figures 5 and 6. Under TE-mode excitation, strong and continuous surface currents propagate along the interconnected metallic pathways, producing a substantial inductive response (L_{eff}) and corresponding magnetic field enhancement. The synergistic interaction between the effective capacitance generated by the narrow gaps and the effective inductance provided by the continuous current paths establishes a pronounced LC resonance at 10.35 GHz. As a result, the incident electromagnetic energy is effectively trapped and dissipated through conductor losses in the copper layer as well as dielectric losses within the FR-4 substrate. Conversely, under TM-mode excitation, only weak surface currents are induced, resulting in negligible magnetic field localization and a substantially suppressed resonant response. The absence of a strong LC resonance leads to severe impedance mismatch between the absorber and free space, causing the majority of the incident TM-polarized wave to be reflected rather than absorbed. Therefore, the field and current distributions provide clear physical evidence that the proposed asymmetric continuous-network metasurface selectively supports resonance for TE-polarized waves while suppressing resonance under TM polarization. This mechanism enables near-perfect absorption for TE waves and high reflection for TM waves, there by producing the observed polarization-selective absorption characteristic.

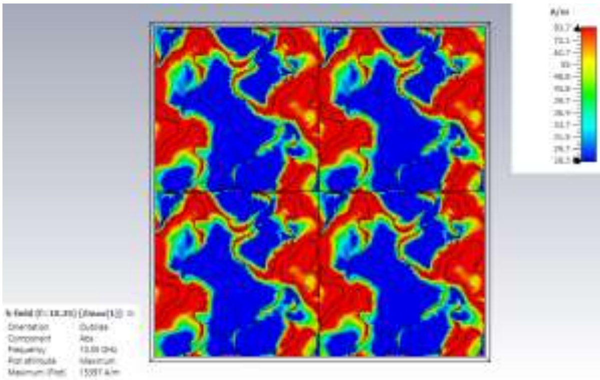


a

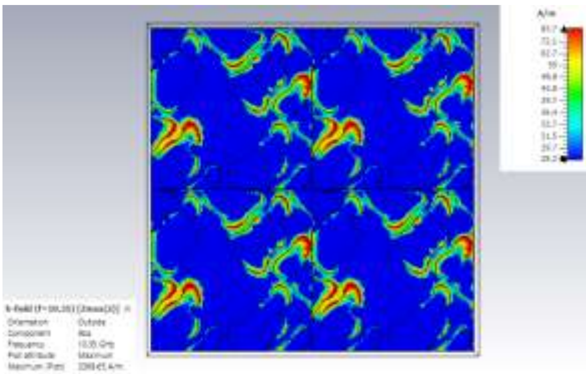


b

Figure 4. Simulated electric field distributions of the proposed absorber: (a) under TE mode excitation, and (b) under TM mode excitation.

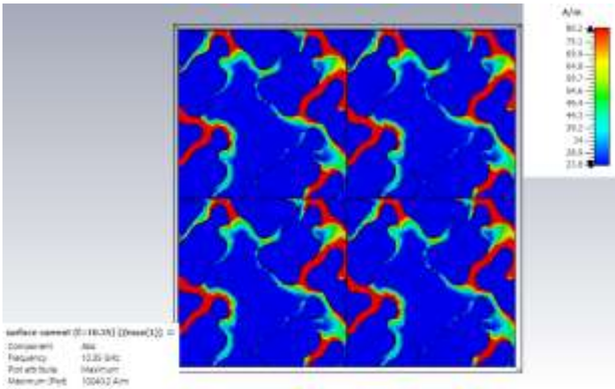


a

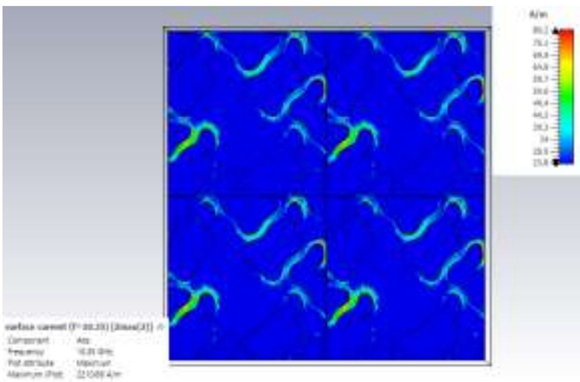


b

Figure 5. Simulated magnetic field distributions of the proposed absorber: (a) under TE mode excitation, and (b) under TM mode excitation.



a



b

Figure 6. Simulated surface current distributions of the proposed absorber: (a) under TE mode excitation, and (b) under TM mode excitation.

5. CONCLUSIONS

This study presented the design and numerical investigation of a novel polarization-selective microwave absorber based on a mathematically modulated asymmetric continuous-network metasurface. By utilizing trigonometric spatial modulation functions, a continuous subwavelength metallic topology was realized on a low-cost 1.6 mm thick FR-4 substrate backed by a continuous copper ground plane. Unlike conventional patch-based absorbers, the proposed continuous architecture minimizes structural discontinuities and promotes uninterrupted current propagation, thereby enhancing electromagnetic energy confinement and dissipation. Comprehensive three-dimensional full-wave electromagnetic simulations demonstrated that the proposed absorber exhibits a pronounced dual-band absorption response within the X-band frequency range (8–12 GHz) under Transverse Electric (TE) polarization. The optimized configuration achieved a peak absorptivity of 99.7% at 10.35 GHz, accompanied by a secondary absorption resonance of approximately 90% at 8.9 GHz. The observed absorption performance originates from the combined effect of strong capacitive coupling generated within the modulated gap regions and the enhanced inductive response provided by the continuous metallic pathways. The interaction of these mechanisms establishes an efficient LC resonant system capable of trapping incident electromagnetic energy and dissipating it through conductor and dielectric losses. In addition, the intentional geometrical asymmetry introduced a significant polarization-dependent response. While TE-polarized waves strongly excited the resonant modes of the structure, TM-polarized waves exhibited considerably weaker coupling, resulting in substantially lower absorption levels throughout the operating frequency range. This polarization-selective behavior demonstrates the effectiveness of the proposed topology in tailoring electromagnetic responses through controlled structural asymmetry. Overall, the results indicate that mathematically generated asymmetric continuous-network metasurfaces represent a promising approach for the development of compact, scalable, and polarization-selective microwave absorbers. The proposed design methodology offers valuable insights for future absorber engineering and may find potential applications in radar cross-section (RCS) reduction, electromagnetic interference (EMI) mitigation, polarization filtering, and advanced stealth-related electromagnetic systems.

REFERENCES

[1] Kaur, K. P., Upadhyaya, T. K., & Palandoken, M. (2017). Dual-band polarization-insensitive 1-)etamaterial inspired microwave absorber for LTE-band applications. Progress In Electromagnetics Research C, 77, 91-100. [2] Kaur, K. P., Upadhyaya, T., & Palandoken, M. (2018). Dual-Band Compact Metamaterial-Inspired Absorber with Wide Incidence Angle and Polarization Insensitivity for GSM and ISM Band Applications. Radioengineering, 27(4).

[3] Kaur, K. P., Upadhyaya, T., & Palandoken, M. (2018, April). Ultrathin wideband polarization independent compact metamaterial microwave absorber. In 2018 28th International Conference Radioelektronika (RADIOELEKTRONIKA) (pp. 1-6). IEEE.

[4] B. Çolak, M. Oral, M. Bakir, et al., "Chaotic system based ultra-wideband microwave absorber designed by resistive ink modeling," Journal of Computational Electronics, vol. 24, no. 6, p. 205, Oct. 2025, doi: 10.1007/s10825-025-02442-y.

[5] O. Yalçıntaş and M. Palandöken, "Superformula-Based Star-Shaped Electromagnetic Wave Absorber," Department of Electrical and Electronics Engineering, İzmir Katip Çelebi University,

İzmir,Türkiye.[Online].Available:https://www.researchgate.net/publication/405449702_Superformula-Based_Star-Shaped_Electromagnetic_Wave_Absorber. Accessed: Jun. 22, 2026.
[6] O. Yalçıntaş and M. Palandöken, "Complex-Structured Superformula-Based Electromagnetic Wave Absorber," Department of Electrical and Electronics Engineering, İzmir Katip Çelebi University, İzmir, Türkiye. [Online]. Available: https://www.researchgate.net/publication/405449981_Complex-Structured_Superformula-Based_Electromagnetic_Wave_Absorber. Accessed: Jun. 22, 2026.
[7]. Shahsavaripour, A., Badiei, M. H., Kalhor, A. & Yousefi, L. (2025). Design of thin, wideband electromagnetic absorbers with polarization and angle insensitivity using deep learning. Scientific Reports, 15, 9426.
[8]. Huang, Q., Xie, W., Alqahtany, F. Z., Cao, T., Mersal, G. A. M., & Toktarbay, Z. (2025). Study on thin-layer broadband metamaterial absorber based on composite multi-opening ring pattern of magnetic dielectric layers. Advanced Composites and Hybrid Materials, 8, 180.
[9]. Tuong, P. V., Lam, V. D., Park, J. W., Choi, E. H., Nikitov, S. A., & Lee, Y. P. (2011). Perfect-absorber metamaterial based on a flower-shaped structure. Journal of Physics D: Applied Physics,44(47),475401.

Numerical Investigation with Mathematical Modeling of Electro-Mechanical Governor Systems Using Differential Equations

Gül OKULMUŞ[1]

,

,

[1]Mechanical Engineering, İstanbul Medeniyet University, Turkey *(gulokulmuus@gmail.com)

ABSTRACT

Electromechanical governor systems are widely used in diesel engines to maintain stable rotational speed under varying load conditions and ensure reliable dynamic performance. This study presents a comprehensive mathematical and numerical investigation of an electromechanical governor system designed for diesel engine speed regulation. A dynamic mathematical model is developed based on Newton's law of rotational motion by considering important system parameters such as moment of inertia, damping coefficient, engine torque, load torque, and governor gain. The developed model is transformed into the Laplace domain to obtain the system transfer function and evaluate stability characteristics. A closed-loop feedback control structure is employed to analyze the transient response behavior and disturbance rejection capability of the system. MATLAB-based numerical simulations are conducted to investigate important performance criteria including settling time, rise time, overshoot, and steady-state error under both normal operating conditions and sudden load disturbances. In addition, parameter sensitivity analysis is performed to examine the influence of inertia, damping, and governor gain on system stability and dynamic response characteristics. Simulation results demonstrate that properly tuned governor parameters significantly improve speed regulation performance, reduce oscillatory behavior, and enhance overall system robustness. The agreement between theoretical analysis and numerical simulation results validates the accuracy of the proposed mathematical model. Overall, this study contributes to the understanding, analysis, and optimization of stable and efficient electromechanical governor systems for modern diesel engine applications.

Keywords – Diesel engine, electromechanical governor system, control systems, dynamic modeling, feedback control, stability analysis, MATLAB simulation

1. Introduction (System Description)

Electromechanical governor systems play a crucial role in maintaining stable rotational speed and ensuring reliable operation of diesel engines under varying load conditions. Accurate speed regulation is essential for improving fuel efficiency, operational stability, and overall engine performance in industrial, marine, and automotive applications [1]–[3]. The primary objective of a governor system is to continuously compare the actual engine speed with a predefined reference value and compensate for any deviations through an appropriate feedback control mechanism [4]. The increasing complexity of modern engineering systems has created a growing demand for accurate mathematical modeling, advanced computational algorithms, and intelligent engineering techniques for the analysis and design of control systems [18], [21]. In this context, simulation-based mathematical modeling has become an efficient and economical approach for evaluating both static and dynamic characteristics of engineering systems before practical implementation [19]. Modern computational tools enable engineers to investigate system performance, optimize controller parameters, and predict system behavior under different operating conditions with high accuracy.

In modern control engineering, feedback-based electromechanical regulators are preferred because of their fast dynamic response, robustness against disturbances, and ability to maintain stable operation under changing environmental and loading conditions [5], [6]. Governor systems are therefore considered essential components in diesel engine applications where precise speed regulation and transient performance are required. The dynamic behavior of electromechanical governor systems can be analyzed using classical rotational motion equations derived from Newton's law of rotation [7]. In such systems, important physical parameters include angular velocity (ω), engine torque (T_m), load torque (T_L), moment of inertia (J), damping coefficient (B), and governor gain (K_g). These variables collectively determine the transient response characteristics, system stability, and disturbance rejection capability of the control mechanism [8].

Mathematical modeling and numerical simulation techniques are extensively employed to evaluate governor system performance before experimental implementation [9]. MATLAB and Simulink provide powerful environments for analyzing system dynamics, stability characteristics, and transient response under various operating conditions [10]. Furthermore, optimization of PID governor parameters has been widely investigated to improve dynamic response, minimize overshoot, and reduce steady-state error in diesel engine control systems [20]. Previous studies have consistently demonstrated that appropriate controller parameter tuning significantly enhances speed regulation performance while improving system stability [11], [12]. This study presents a comprehensive mathematical and numerical investigation of an electromechanical governor system for diesel engine speed regulation. A dynamic mathematical model is developed using differential equations and transformed into the Laplace domain to obtain the system transfer function and evaluate stability characteristics. MATLAB-based simulations are performed to analyze transient response, disturbance rejection capability, and parameter sensitivity. The results contribute to the development of stable, efficient, and reliable electromechanical governor systems while providing valuable guidance for future optimization and control design in modern diesel engine applications [13],[15].

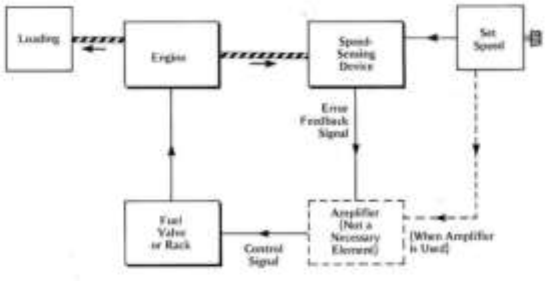


Fig. 1. Closed-loop control structure of the electro-mechanical governor system. The reference speed is compared with the actual speed, and the error signal is processed by the governor to adjust the motor torque.

2. Literature Review

Several studies in the literature have investigated the modeling, simulation, and optimization of governor systems used in diesel engines and power generation applications. Malkhede et al. [20] studied the optimization of PID governor parameters for diesel engine systems and demonstrated that properly tuned controller gains significantly improve transient response characteristics and reduce steady-state errors. Tsegaye and Fante [17] analyzed hydro governor control systems of synchronous machines using MATLAB/Simulink and emphasized the importance of accurate governor modeling for maintaining system stability under varying operating conditions.

Gagan Singh and Chauhan [18] proposed a simplified hydraulic governor-turbine model for stable operation of power systems and reported that mathematical modeling plays a critical role in improving operational reliability. Zhang et al. [19] conducted a simulation-based study of hydraulic turbine governor systems and concluded that MATLAB/Simulink environments provide efficient tools for investigating dynamic behavior, disturbance rejection capability, and control system performance. Previous studies collectively indicate that governor parameter tuning, feedback control structure, and accurate mathematical modeling are essential for achieving stable and efficient speed regulation performance in electromechanical and hydraulic governor systems. These studies demonstrate that mathematical modeling and PID-based governor control approaches remain highly important for improving dynamic stability, disturbance rejection capability, and transient response performance in modern diesel engine applications.

3. Materials and Methods

3.1 Mathematical Modeling

The system dynamics are derived from Newton's law of rotational motion:

The rotational dynamics equation is commonly employed in electromechanical system analysis and control applications [8].

$$J\frac{d\omega}{dt} + B\omega = T_m - T_L \quad (1)$$

The governor control law is defined as:

$$T_m = K_g(\omega_{ref} - \omega) \quad (2)$$

where:

J = moment of inertia of the rotating system (kg·m²)

ω = angular velocity of the engine shaft (rad/s)

$d\omega/dt$ = angular acceleration (rad/s²)

B = viscous damping coefficient (N·m·s/rad)

T_m = motor or engine torque generated by the governor system (N·m)

T_L = external load torque applied to the system (N·m)

K_g = governor gain coefficient

ω_{ref} = reference angular velocity or desired rotational speed (rad/s)
$\Omega(s)$ = Laplace transform of the angular velocity

$\Omega_{ref}(s)$ = Laplace transform of the reference speed signal
s = Laplace domain complex frequency variable

3.2 Laplace Domain Analysis

Applying the Laplace transform (zero initial conditions):

$$Js\Omega(s) + B\Omega(s) = K_g(\Omega_{ref}(s) - \Omega(s)) \quad (3)$$

The transfer function becomes:

$$\frac{\Omega(s)}{\Omega_{ref}(s)} = \frac{K_g}{Js+B+K_g} \quad (4)$$

3.3 Stability Analysis

The characteristic equation is:

$$Js + B + K_g = 0 \quad (5)$$

The system pole is:

$$s = -\frac{B+K_g}{J} \quad (6)$$

Since all parameters are positive, the system is stable.

3.4 Numerical Simulation

Simulations are performed using MATLAB with the following parameters:

- $J = 10$ kg·m²
- $B = 5$ N·m·s/rad
- $K_g = 20$

MATLAB/Simulink is widely used for dynamic system analysis and control system design because of its advanced numerical computation capabilities [10].

The MATLAB simulation code used in this study is provided in Appendix A.

Two scenarios are analyzed:

- Step input response

- Load disturbance at $t = 3$ seconds

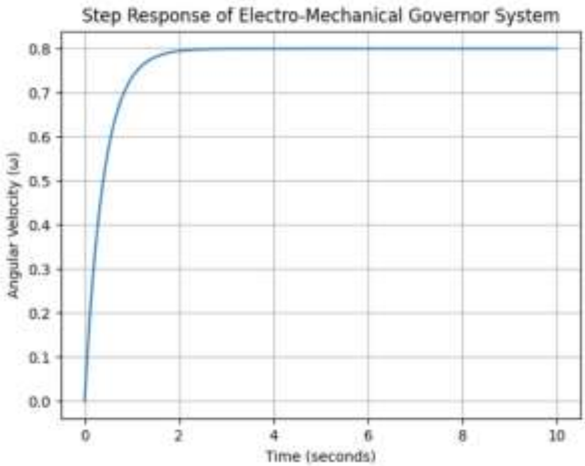


Fig. 2. Step response of the system. The system reaches steady-state in approximately 3 seconds with minimal overshoot, indicating stable and efficient performance.

The obtained response characteristics indicate that the system satisfies classical transient response performance criteria such as acceptable settling time and low overshoot [11].

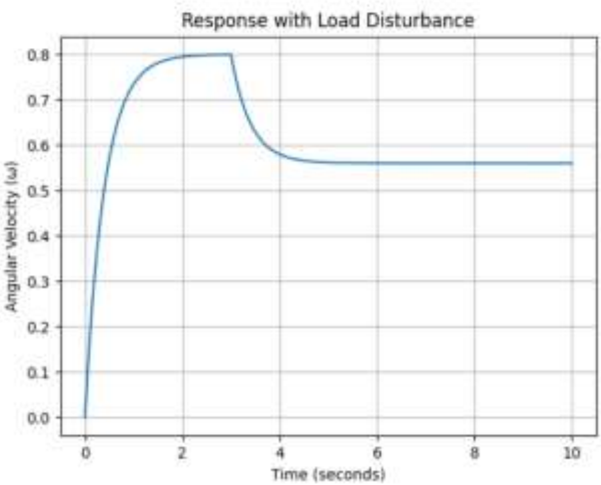


Fig. 3. System response under load disturbance. A disturbance is applied at 3 seconds, and the system successfully compensates and returns to steady-state.

Disturbance rejection capability is one of the most important indicators of governor system robustness and operational reliability [12]

3.5 Parameter Sensitivity Analysis

The effects of system parameters are summarized as follows:

- Increasing inertia (J) slows down the system response
- Increasing damping (B) reduces oscillations
- Increasing governor gain (K_g) speeds up the response but may increase overshoot

Sensitivity analysis demonstrates that system inertia directly affects response speed, while damping contributes significantly to oscillation suppression. Furthermore, excessive governor gain may lead to aggressive system behavior and instability tendencies [13].

4. Results and Discussion

The simulation results demonstrate that the proposed electromechanical governor system effectively maintains rotational speed stability under both nominal operating conditions and external load disturbances. The obtained step response characteristics indicate that the system reaches steady-state operation in approximately 3 seconds with low overshoot and negligible steady-state error, which are considered desirable transient response properties in classical control engineering applications [11], [13]. The response curves presented in Figures 2 and 3 confirm that the closed-loop control structure provides stable and well-damped dynamic behavior throughout the simulation process.Furthermore, the disturbance analysis shows that the governor system successfully compensates for sudden load variations applied at different operating intervals. Following the disturbance input, the system rapidly returns to the desired reference speed without exhibiting excessive oscillatory behavior. This demonstrates the robustness and disturbance rejection capability of the controller, which are essential performance indicators for electromechanical speed regulation systems used in diesel engine applications [12], [14].
Parameter sensitivity analysis also reveals that the dynamic response of the system strongly depends on inertia, damping coefficient, and governor gain values. Increasing the inertia parameter slows the transient response, whereas higher damping values effectively suppress oscillations and improve system stability. In contrast, excessive governor gain may produce faster responses but can also increase overshoot and reduce stability margins [6], [8]. These findings are consistent with previous studies reported in the

control systems literature and validate the theoretical assumptions used in the mathematical modeling process [14], [15].

Overall, both theoretical analysis and MATLAB-based numerical simulations confirm that the proposed governor model provides reliable speed regulation performance, stable transient characteristics, and satisfactory disturbance compensation capability for modern diesel engine control applications.

5. Conclusion

This study presented a comprehensive mathematical and numerical investigation of an electromechanical governor system used for diesel engine speed regulation applications. A dynamic system model was developed using differential equations derived from Newton's law of rotational motion, and the obtained equations were transformed into the Laplace domain to evaluate system stability and transfer function characteristics [1], [7]. The developed closed-loop control structure enabled detailed investigation of transient response behavior and disturbance rejection performance under varying operating conditions.

MATLAB-based simulations demonstrated that the proposed governor system achieves fast settling time, low overshoot, and negligible steady-state error while maintaining stable operation during sudden load disturbances. The simulation findings confirmed that appropriate tuning of governor gain, damping coefficient, and inertia parameters significantly improves overall system stability and dynamic response performance [10], [11]. In addition, the consistency between theoretical stability analysis and numerical simulation results validates the accuracy and reliability of the proposed mathematical model. The parameter sensitivity analysis further showed that system inertia primarily affects response speed, while damping contributes to oscillation suppression and improved transient behavior. Although higher governor gain values increase response speed, excessive gain may negatively influence stability margins and produce undesirable oscillatory effects [6], [13]. These observations are in agreement with previously published studies related to electromechanical control systems and diesel engine governor applications [14].

By integrating mathematical modeling, Laplace transform analysis, feedback control principles, and numerical simulations, this study provides a systematic framework for analyzing and optimizing electromechanical governor systems. Overall, the findings contribute to the development of efficient, stable, and reliable control strategies for modern diesel engine technologies. Future studies may focus on nonlinear system modeling, adaptive and intelligent control approaches, and experimental validation using real-time diesel engine platforms to further improve governor system performance and operational robustness [12], [15].

References

1. K. Ogata, Modern Control Engineering, 5th ed. Upper Saddle River, NJ, USA: Prentice Hall, 2010.
2. R. C. Dorf and R. H. Bishop, Modern Control Systems, 12th ed. Boston, MA, USA: Pearson, 2011.
3. G. F. Franklin, J. D. Powell, and A. Emami-Naeini, Feedback Control of Dynamic Systems, 7th ed. Pearson, 2015.
4. MathWorks, “MATLAB Documentation,” 2024.
5. IEEE Xplore Digital Library, “Governor System Studies,” Accessed: 2024.
6. N. S. Nise, Control Systems Engineering, 8th ed. Wiley, 2019.
7. W. Bolton, Mechatronics: Electronic Control Systems in Mechanical Engineering, 7th ed. Pearson, 2020.
8. B. C. Kuo and F. Golnaraghi, Automatic Control Systems, 9th ed. Wiley, 2014.
9. W. J. Palm III, System Dynamics, 3rd ed. McGraw-Hill, 2016.
10. MathWorks, “Simulink Documentation,” 2024.
11. K. J. Åström and R. M. Murray, *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton, NJ, USA: Princeton University Press, 2021.
12. H. K. Khalil, *Nonlinear Control Systems*, 3rd ed. Upper Saddle River, NJ, USA: Pearson, 2015.
13. G. F. Franklin, J. D. Powell, and M. L. Workman, *Digital Control of Dynamic Systems*, 3rd ed. Boston, MA, USA: Pearson, 2019.
14. *IEEE Transactions on Control Systems Technology*, various articles on governor system studies.
15. *SAE Technical Papers*, various papers on diesel engine governor control systems.
16. G. Janevska and S. Panovska, “Aspects on Modeling of Hydro-Turbine Speed Governor,” International Journal of Scientific & Engineering Research, vol. 10, no. 6, 2019.
17. T. Shewit and K. A. Fante, “Hydro governor control of synchronous machines: MATLAB/SIMULINK based analysis,” International Research Journal of Engineering and Technology, vol. 3, no. 8, 2018.
18. G. Singh and D. S. Chauhan, “Simplified modeling of hydraulic governor-turbine for stable operation under operating conditions,” Elixir Electrical Engineering, vol. 39, pp. 5030–5032, 2011.
19. H. Zhang, J. Liu, and Y. Liang, “Modeling and Simulation Study of Hydraulic Turbine Governor Based on SIMULINK,” 2022 International Conference on Automation, Robotics and Computer Engineering (ICARCE), Wuhan, China, 2022, pp. 1–5.
20. D. N. Malkhede, H. C. Dhariwal, and M. C. Joshi, “On optimization of the PID governor for diesel engine,” Mathematical Modelling and Analysis, vol. 7, no. 1, pp. 135–150, 2002.

21. O. Boydak and N. Bouhmala, “Guest Editorial: Advanced Algorithms and Techniques for Miscellaneous Engineering Applications Such As Discrete Networks, Machinery and Computing Platforms,” Journal of Internet Technology, vol. 19, no. 3, pp. 905–907, May 2018.