Journal of Engineering and Technology for Industrial Applications

ISSN 2447-0228

October 2025

Volume 11/ No 55

Editor-in-Chief: J. C. Leite

www.itegam-jetia.org

ITEGAM - JETIA ISSN 2447-0228 Online
Journal of Engineering and Technology for Industrial Applications (JETIA)

O **ITEGAM-JETIA: Journal of Engineering and Technology for Industrial Applications** is a publication of the Galileo Institute of Technology and Education of the Amazon (ITEGAM), located in the city of Manaus since 2008. JETIA publishes original scientific articles covering all aspects of engineering. Our goal is the dissemination of research original, useful and relevant presenting new knowledge on theoretical or practical aspects of methodologies and methods used in engineering or leading to improvements in professional practice. All the conclusions presented in the articles It should be state-of-the-art and supported by current rigorous analysis and balanced assessment. Public magazine scientific and technological research articles, review articles and case studies.
JETIA will address topics from the following areas of knowledge: Mechanical Engineering, Civil Engineering, Materials and Mineralogy, Geosciences, Environment, Information and Decision Systems, Processes and Energy, Electrical and Automation, Mechatronics, Biotechnology and other Engineering related areas.

Publication Information:
ITEGAM-JETIA (ISSN 2447-0228), (online) is published by Galileo Institute of Technology and Education of the Amazon on a every two months (February, April, June, August, October and December).

Contact information:

Web page: www.itegam-jetia.org
Email: editor@itegam-jetia.org
Galileo Institute of Technology and Education of the Amazon (ITEGAM).

Joaquim Nabuco Avenue, No. 1950. Center. Manaus, Amazonas. Brazil.
Zip Code: 69020-031. Phone: (92) 3584-6145.

Copyright 2014. Galileo Institute of Technology and Education of the Amazon (ITEGAM)
The total or partial reproduction of texts related to articles is allowed, only if the source is properly cited. The concepts and opinions expressed in the articles are the sole responsibility of the authors.

Previous Notice
All statements, methods, instructions and ideas are the sole responsibility of the authors and do not necessarily represent the view of ITEGAM -JETIA. The publisher is not responsible for any damage and / or damage to the use of the contents of this journal. The concepts and opinions expressed in the articles are the sole responsibility of the authors.

Directory

Members of the ITEGAM Editorial Center - Journal of Engineering and Technology for Industrial Applications (ITEGAM-JETIA) of the Galileo Institute of Technology and Education of the Amazon (ITEGAM). Manaus-Amazonas, Brazil.

Jandecy Cabral Leite, CEO and Editorial Editor-in-Chief
Ivan Leandro Rodriguez Rico, Editorial Assistant
Jorge de Almeida Brito Junior, Editorial Assistant

ITEGAM - JETIA ISSN 2447-0228 Online
Journal of Engineering and Technology for Industrial Applications (JETIA)

ITEGAM-JETIA. v.11, n.55. October of 2025. Manaus - Amazonas, Brazil. ISSN 2447-0228 (ONLINE)
https://www.itegam-jetia.org

SUMMARY

STOCKVISION: A MULTI-AGENT AI FRAMEWORK FOR INTELLIGENT STOCK INVESTMENT AND ANALYSIS USING CREWAI **01**
Jitendra Saturwar, Manav Ghetiya, Nehal Kolipaka, Vivek Maurya, Pravin Maurya

EXTENDED KALMAN FILTER ESTIMATOR AND MAMDANI FUZZY LOGIC CONTROLLER FOR SENSORLESS DC MOTOR SPEED CONTROL **11**
Hadjer Abderrezek, Ameur Aissa

AN INTELLIGENT WEB PLATFORM FOR MENSTRUAL TRACKING, HYGIENE, AND COMMUNITY SUPPORT USING MACHINE LEARNING **20**
Payal Tandan, Purnima Nahata, Rovina Dbritto

ASSESSMENT OF SOLAR CHIMNEY POWER PLANT PERFORMANCE UNDER VARIOUS PARAMETERS **29**
Farid Gaci, Mohamed Walid Azizi, Ahcene Akni, Moumtez Bensouici, Djoubeir Debbah

IMPROVING INVERTER EFFICIENCY FOR ELECTRIC VEHICLES: EXPERIMENTAL VALIDATION OF THE NEURAL NETWORK-BASED SHE TECHNIQUE USING RT-LAB **34**
Seyf Eddine Bechekir, Mokhtaria Jbilou, Mostefa Brahami, Fatima Zohra Boudjella, Imen Souhila Bousmaha, Mimouna Oukli, Said Nemmich

INTEGRATING VGG RE-TRAINED FEATURE EXTRACTION WITH MACHINE LEARNING FOR KNEE OSTEOARTHRITIS SEVERITY LEVELS DETECTION USING X-RAY IMAGES **41**
Simeon Yuda Prasetyo, Ghinaa Zain Nabiilah

REDUCING COMPLEXITY IN VEHICLE RECOGNITION: A COMPARISON OF DIMENSIONALITY REDUCTION METHODS FOR ACOUSTIC FEATURES **48**
Yudhi Darmawan, Muladi Muladi, Heru Wahyu Herwanto, Nurrachman Supadmana Muda

IMPROVING THE EFFICIENCY OF THE TECHNOLOGICAL PROCESS OF ELECTROMECHANICAL TREATMENT OF MACHINE COMPONENTS BY MINIMIZING THE NUMBER OF CURRENT-CARRYING CIRCUIT ELEMENTS AND DESIGNING SPECIAL EQUIPMENT **59**
Elena Veremey, Dmitriy Podashev

PARAMETERS ESTIMATION OF THREE-PHASE INDUCTION MOTOR USING THE PARAMETER ESTIMATOR APP **73**
Fares Bettahar, Sabrina Abdeddaim, Omar Charrouf, Achour Betka, Michael Short, Laid Guerrida, Belahcene taha lemdjed

MULTI-STAGE FEATURE SELECTION FOR OPTIMIZING STUDENT DROPOUT PREDICTION **86**
Rabouh Mustapha, Kamel Zouggar, Khelifa Guerraiche

AI FOR POSITIONING ACCURACY ENHANCEMENT IN THE IIOT **96**
Sofiane Latreche, Hocine Bellahsene

LEGALMIND: A FINE-TUNED GEMMA-2-BASED LEGAL ASSISTANT FOR INDIAN JUDICIARY WITH RAG AND EMBEDDING INTEGRATION **105**
Harsh Dilip Pimpale, Aditi Raut, Yash Patil, Gaurav Parpol, Prajwal Yadav, Janhavi Sangoi

INDUSTRIAL CENTRALIZED CONTROL AUTOMATION SYSTEM USING IOT BASED EMBEDDED SYSTEM **118**
Md Mehedi Hassain, Gonga Sutradhar, Md. Nayeem Ibna Harun, Abdur Rahman, Md Ariful Islam

ITEGAM - JETIA
ISSN 2447-0228 Online
Journal of Engineering and Technology for Industrial Applications (JETIA)

RISK QUANTIFICATION IN MANUFACTURING INDUSTRY INVESTMENTS: A STOCHASTIC APPROACH WITH ARTIFICIAL INTELLIGENCE **128**
Rafael Vieira da Silva, Ing., Enzo Morosini Frazzon

DEVELOPMENT OF A VEHICLE SENSOR FAULT DIAGNOSIS DEVICE USING ARTIFICIAL INTELLIGENCE TECHNIQUES TO IMPROVE DIAGNOSTIC ACCURACY **140**
Messaoud Linani, Boubakeur Latreche, Kamal Khalil Toaba, Ali Ben Kholif, Imededdin Djadour, Aboubakr Missaoui

NUMERICAL STUDY OF THE ASYNCHRONOUS EXCITATION EFFECT OF BOUMERDES' EARTHQUAKE ON A BRIDGE BEHAVIOR **154**
Moahmed Guesmi, Nassira Belkheiri, Mohamed Lakhder Guesmi

A NOVEL PRIVACY-PRESERVING DECENTRALIZED FEDERATED LEARNING APPROACH FOR INCREMENTAL HEALTHCARE DATA **161**
Rostom Mennour

ASSESSMENT AND OPTIMIZATION OF ENGINE PERFORMANCE AND EMISSIONS USING ADDITIVE-INFUSED CORN OIL BIODIESEL **174**
Raviteja Surakasi, Makka Jayakrishna, Yenda Srinivasarao, Iván Leandro Rodríguez Rico

ENHANCEMENT OF MECHANICAL AND MICROSTRUCTURAL PROPERTIES OF SIMILAR MILD STEEL WELDS UNDER VIBRATORY PROCESSING **183**
Japhia Sudarsan Nalla, Basant Kumar Palai, Govinda Rao Potnuru

PARAMETRIC OPTIMIZATION AND EXPERIMENTAL VERIFICATION OF FLEXURAL AND IMPACT PERFORMANCE OF FDM PRINTED AUTOMOBILE JIG COMPONENT USING PA12CF **192**
Rahul Patel, Suketu Jani

A FRAMEWORK TO ENHANCE DATA QUALITY IN MASTER DATA MANAGEMENT PROCESS: A HEALTHCARE STUDY **204**
Chafika Benkherourou, Abdelhabib Bourouis

NOVEL INSIGHTS ON THE COMPARATIVE STUDY BETWEEN LSTM AND TRANSFORMER MODELS FOR FINANCIAL TIME SERIES PREDICTION **212**
Rabia Nichani, Laid Gasmi, Salheddine Kabou, Nabil Laiche

A DEEP LEARNING MODEL TO CLASSIFY FOREST FIRES BURNED AREAS USING SENTINEL-2 DATA IN ALGERIA **222**
Ouahab Abdelwhab

PERFORMANCE IMPROVEMENT IN REAL-TIME 3D RENDERING USING OPENGL FRAGMENT SHADER AND UNITY SURFACE SHADER **230**
Eriq Muhammad Adams Jonemaro, Ivan Maulana Agusti, Herman Tolle, Muhammad Aminul Akbar, Aryo Pinandito

INVESTIGATION OF PARTIAL DEMAGNETIZATION FAULTS IN PERMANENT MAGNET SYNCHRONOUS MOTORS BASED ON EFFICIENCY AND LOSS MAPS ADVANCED ANALYSIS **238**
Layachi Chebabhi, Toufik Tayeb Naas, Mohamed Zitouni, Monaem Elmnifi, Abdelkarim Cherhabil

STRUCTURAL ANALYSIS OF UPPER ARM EXOSKELETON FOR BIOMEDICAL APPLICATIONS **250**
Vijayakumar K, Bubesh Kumar D, Sangeetha Krishnamoorthi

MODELING THE ELASTIC BEHAVIOR OF CLAY SOILS REINFORCED WITH RECYCLED RUBBER FIBERS **260**
Melik Bekhiti, Abdelhalim Bensaada, Ali Smaida

ITEGAM - JETIA
ISSN 2447-0228 Online
Journal of Engineering and Technology for Industrial Applications (JETIA)

INVESTIGATION ON PERFORMANCE EVALUATION OF CONTAINERIZED MICROSERVICE FOR EDGE COMPUTING ENVIRONMENT **268**
Pranjit Kakati, Abhijit Bora

COMBINED EFFECT OF GLASS POWDER AND RECYCLED SAND ON FRESH AND HARDENED PROPERTIES OF STEEL FIBER-REINFORCED SELF-COMPACTING CONCRETE **276**
Souhila Fergani, Said Zaouai, Rachid Rabehi, Ali Abbache

COMPARATIVE ANALYSIS OF YOLOV3, MOBILENET-SSD, AND EFFICIENTDET FOR REAL-TIME PERSON DETECTION IN LOW-RESOLUTION IMAGES **293**
JIdham Khalif, Raden Deiny Mardian

Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11 n.55, p. 01-10. July/August., 2025.

DOI: https://doi.org/10.5935/jetia. v11i55.1816

RESEARCH ARTICLE OPEN ACCESS

STOCKVISION: A MULTI-AGENT AI FRAMEWORK FOR INTELLIGENT STOCK INVESTMENT AND ANALYSIS USING CREWAI

Jitendra Saturwar [1], Manav Ghetiya[2], Nehal Kolipaka[3], Vivek Maurya[4], Pravin Maurya[5]

[1,3,4] Department of Computer Engineering, Universal College of Engineering, University of Mumbai, vasai.
[2,5] University of Mumbai, Mumbai.

[1]https://orcid.org/0000-0002-7097-2193, [2]http://orcid.org/0009-0003-1532-3996, [3]http://orcid.org/0000-0002-8722-7181
[4]https://orcid.org/0000-0002-7097-2193, [5]https://orcid.org/0009-0009-3256-3012

Email: jsaturwar@gmail.com, ghetiyamanav@gmail.com, nehalkolipaka73@gmail.com, vivekmaurya1893@gmail.com, pravincoder@gmail.com

ARTICLE INFO

Article History
Received: April 5, 2025
Revised: May 20, 2025
Accepted: September 15, 2025
Published: September 30, 2025

Keywords:
Artificial Intelligence,
Multi-Agent Systems,
Stock Investment Analysis,
Technical Analysis,
Fundamental Analysis

ABSTRACT

This paper introduces an innovative multi-agent platform for stock investment and market analysis that leverages state-of-the-art artificial intelligence techniques alongside real-time financial data processing. Built on a robust Django-based backend emulating a SaaS architecture, the platform integrates CrewAI to orchestrate multiple specialized agents responsible for data acquisition, sentiment analysis, and predictive modeling. By harnessing live market data via the Yahoo Finance API, the system delivers actionable investment recommendations and comprehensive insights—making advanced analytical capabilities accessible even to users with limited financial expertise. Experimental evaluations demonstrate that this multi-agent framework not only enhances analytical precision but also supports a modular, flexible approach to addressing complex decision-making challenges in dynamic market environments. Consequently, the proposed platform represents a significant advancement in automated stock analysis and serves as a robust tool for both academic research and the development of practical investment strategies .

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

AI has significantly transformed financial market analysis by enabling multi-agent systems such as CrewAI to deliver real-time insights through specialized agents that evaluate technical indicators, fundamental metrics, sentiment analysis, and predictive modeling [1]. Our Django-based platform leverages this advanced framework to automate data collection from sources like YahooFinance, thereby processing vast streams of financial data to generate actionable investment recommendations [2]. Designed as a complete SaaS application, the system incorporates robust authentication and authorization mechanisms, comprehensive user profile management, and dynamic report generation to ensure secure, personalized access to market insights [3].

By synthesizing data from diverse financial indicators, our platform empowers investors to make informed decisions in dynamic and volatile market conditions, thereby setting a new standard for automated financial analysis [4]. Experimental evaluations on multiple stocks reveal that the multi-agent architecture not only improves predictive accuracy but also reduces latency in data processing, thereby enhancing overall decision-making efficiency [4]. The seamless integration of CrewAI with Django's scalable backend has demonstrated a significant boost in performance while maintaining stringent security standards [3]. Moreover, user feedback underscores the platform's ability to deliver personalized insights effectively, positioning it as a valuable tool for both novice and professional investors. Future work will focus on incorporating advanced machine learning techniques and additional data sources to further refine its predictive capabilities and adapt to evolving market dynamics [5], [2].

Table 1: Comparative Analysis of Different AI-Driven Stock Market Analysis Approaches.

Algorithm	Method	Strengths	Use Cases
Machine Learning Models	Data-driven pattern recognition	Adapts to changing trends; improves with more data	Predicting stock movements & risk assessment
Deep Learning	Sequential data processing, feature learning	Captures complex dependencies in financial data; robust predictions	High-frequency trading & time-series forecasting
Sentiment Analysis	Natural Language Processing (NLP) on financial news & social media	Detects market sentiment shifts; integrates qualitative insights	News-based trading strategies & risk mitigation

Source: Authors, (2025).

To promote informed investment decisions, the research paper includes functionality to monitor stock performance trends. If a stock in a user's portfolio shows significant volatility or crosses predefined risk thresholds, the system sends an email and an SMS notification to alert the user of potential market movements. The StockVision AI research paper presents a robust and feature-rich solution for real-time stock analysis, reducing investment risks and enhancing decision-making. It combines AI-driven analytics, live market data, and advanced predictive modeling to provide a comprehensive investment management experience for users.

I.1 OBJECTIVES

The objectives of proposed research work are as follows.

- **To Integrate Real-Time Financial Data Processing:** Utilize the Yahoo Finance API to fetch live market data, ensuring that investment recommendations are based on up-to-date and accurate financial information.
- **To Enhance Investment Decision-Making:** Employ advanced machine learning techniques to generate actionable investment insights, assisting users in making data-driven trading decisions with higher precision.
- **To Automate Sentiment Analysis for Market Trends:** Implement Natural Language Processing (NLP) models to analyze financial news, earnings reports, and social media sentiment to predict stock market trends effectively.
- **To Provide a Scalable and Modular SaaS Architecture:** Build a Django-based backend that follows a scalable Software-as-a-Service (SaaS) model, allowing for seamless integration of additional analytical modules as needed.
- **To Improve User Accessibility and Experience:** Design an intuitive and user-friendly interface that enables even non-expert users to interpret stock market data and make informed investment choices.

The overarching objective of this research is to create an intelligent, automated investment analysis platform that democratizes access to sophisticated financial insights, helping both novice and experienced investors navigate dynamic stock markets with confidence.

I.2 LITERATURE REVIEW

According to the paper "Exploration of LLM Multi-Agent Application Implementation Based on LangGraph+CrewAI" [6] by Zhihua Duan and Jialin Wang explores the integration of LangGraph and CrewAI in developing multi-agent systems for complex task management. The study contributes to the growing body of research on large language model (LLM)-powered autonomous agents by demonstrating how these frameworks can enhance multi-agent collaboration, efficiency, and adaptability.

According to the paper "FinVision: A Multi-Agent Framework for Stock Market Prediction" [7] by Sorouralsadat Fatemi and Yuheng Hu presents a novel approach to financial trading using a multi-agent system powered by large language models (LLMs). This research advances previous work in AI-driven stock market prediction by integrating multimodal data analysis and enhancing explainability in decision-making.

According to the paper "An Investment Recommender Multi-agent System in Financial Technology" [8] by Hernández et al. presents a multi-agent system (MAS) designed to enhance investment recommendation processes in financial technology (FinTech). The study integrates artificial intelligence (AI) techniques, virtual organizations (VOs), and hybrid machine learning models to improve the accuracy and efficiency of investment decisions.

According to the paper "Agentic AI Systems Applied to Tasks in Financial Services: Modeling and Model Risk Management Crews" [9] by Okpala, Golgoon, and Kannan explores the application of multi-agent AI systems in financial services, particularly in modeling and model risk management (MRM). This work contributes to the growing research on AI-driven automation by demonstrating how large language model (LLM)-powered agents can collaborate to perform complex financial tasks efficiently.

I.3 EXISTING SYSTEMS

Traditional stock market analysis relies heavily on manual research, technical chart analysis, and financial reports, which require significant expertise and time. Investors and traders use tools like Yahoo Finance, Bloomberg, and TradingView, but these platforms primarily provide raw data rather than actionable insights. Additionally, conventional algorithmic trading systems focus on price prediction models, which often fail to incorporate broader market trends, sentiment analysis, and fundamental factors.

While some AI-based stock analysis tools exist, they are often limited in scope. Many rely on basic machine learning models that lack real-time adaptability and multi-agent collaboration. These systems typically do not offer automated investment reports, dynamic stock insights, or interactive decision-making support. As a result, investors still struggle to make well-informed decisions quickly, highlighting the need for a more intelligent, automated, and user-friendly stock analysis solution.

Furthermore, existing AI-powered investment tools often focus on historical data analysis rather than real-time market fluctuations, limiting their effectiveness in fast-paced trading environments. Many platforms do not integrate live financial news, social sentiment analysis, or macroeconomic indicators, which are crucial for understanding sudden market shifts. Without a multi-agent framework that dynamically processes diverse data sources, these systems fail to provide adaptive and context-aware recommendations.

Additionally, most current stock analysis platforms lack comprehensive risk management and portfolio optimization features. Investors must manually adjust their strategies based on fragmented insights from multiple tools. This results in inefficiencies and missed opportunities in volatile markets. The absence of automated, AI-driven decision-making support further complicates investment strategies, making it essential to develop a next-generation AI-driven stock analysis system that offers real-time insights, automated reports, and a seamless user experience.

Moreover, traditional and existing AI-based stock analysis tools often operate in siloed environments, where different aspects of market analysis—such as technical indicators, fundamental analysis, and sentiment tracking—are handled separately rather than in an integrated manner. This fragmentation leads to incomplete insights and forces investors to switch between multiple platforms to make well-rounded decisions. Additionally, many existing systems do not leverage reinforcement learning or adaptive AI models that continuously learn from market patterns, making their predictions static and less reliable in dynamic market conditions. Without a holistic, AI-driven approach that unifies multiple analytical techniques, investors remain at a disadvantage in efficiently navigating the complexities of the stock market.

I.4 LIMITATION ON EXISTING SYSTEM

The limitations of the existing systems are as follows;

- **Dependency on Manual Interpretation:** Most stock analysis tools still require users to interpret results manually, limiting accessibility for non-expert investors.
- **Challenges in Real-Time Data Processing:** Existing AI models struggle with processing and analyzing high-frequency, real-time financial data efficiently.
- **Market Volatility and Accuracy Issues:** Many machine learning models, such as regression analysis and time-series forecasting, face difficulties in predicting market trends accurately across different stock categories.
- **Lack of Fully Automated Decision Support:** While current AI-powered trading systems provide insights, they do not offer **automated** portfolio optimization, risk assessment, or real-time alerts.

I.5 PROBLEM STATEMENT

The stock market is highly volatile and complex, making it challenging for investors to make informed decisions based on real-time data, financial indicators, and market trends. Traditional stock analysis methods often require deep financial expertise and extensive manual effort to interpret data, leading to inefficiencies and potential miscalculations.

Existing AI-driven stock prediction tools either focus solely on price forecasting or provide generic insights without personalized, in-depth analysis. Additionally, most platforms lack an interactive multi-agent system that can dynamically analyze stocks, generate professional investment reports, and assist users with data-driven decision-making.

This research aims to develop Stock Vision AI, a multi-agent stock analysis and investment platform that integrates CrewAI agents with real-time stock market data. The system leverages advanced AI models, including LLMs, financial indicators, and technical analysis, to provide detailed insights, automated Markdown reports, and interactive decision support. By enhancing accuracy, automation, and usability, Stock Guru AI seeks to bridge the gap between traditional financial analysis and AI-powered investment strategies.

II. PROPOSED SYSTEM

The proposed system is designed as a modular, scalable, and secure platform that integrates advanced web technologies with state-of-the-art multi-agent analysis to facilitate real-time stock investment and market analysis. At the highest level, the frontend delivers a dynamic, user-friendly dashboard where investors can view detailed stock trends, analysis reports, and personalized recommendations. This interface communicates with a django-based backend through endpoints, ensuring efficient and secure data exchange.

Within the backend, an API gateway and endpoint manager route incoming HTTP requests to specialized modules. One such module is responsible for real-time data integration, which leverages the YahooFinance API via the yfinance package to fetch current market data. This raw data is then preprocessed and funneled into the core of the system: the multi-agent analysis engine built on CrewAI.

The multi-agent analysis engine consists of a suite of specialized agents—each dedicated to distinct analytical tasks such as technical analysis, fundamental evaluation, sentiment assessment, and predictive modeling. These agents work collaboratively, exchanging information and refining their outputs to provide a comprehensive analysis of market trends. Their individual outputs are aggregated and synthesized in a decision synthesis module, which applies weighting and decision logic to formulate coherent and actionable investment recommendations.

Finally, the response dispatcher formats these insights into structured outputs, which are sent back to the frontend for visualization. The entire system is containerized using Docker, ensuring consistent deployment and scalability across various environments. Secure management of configuration details, including API keys and model parameters, is maintained via environment variables. This holistic architecture not only streamlines complex data processing and analysis but also enhances the robustness and adaptability of the investment decision-making process, positioning it as a cutting-edge solution in the realm of financial market analytics.

III. METHODOLOGY

System Architecture

The Multi-Agent Stock Investment and Analysis Platform is designed using a modular architecture that integrates a Django-based backend. The system leverages CrewAI to facilitate multi-agent interactions, enabling comprehensive stock analysis and investment recommendations. The overall architecture consists of the following key components:

- Frontend : Provides an interactive user interface for stock selection, analysis visualization, and investment recommendations.
- Backend (Django with FastAPI Integration): Serves as the core processing unit, handling API requests, multi-agent interactions, and data retrieval from financial sources.
- Multi-Agent System (CrewAI): Implements distinct agents specializing in stock analysis, investment strategies, and risk assessment.
- Data Sources (Yahoo Finance API): Fetches real-time stock data for analysis.

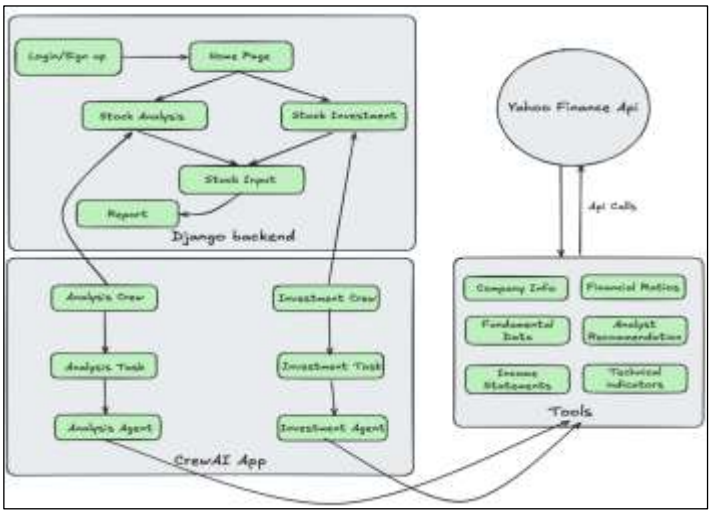


Figure 1: System Architecture.
Source: Authors, (2025).

Multi-Agent Framework

CrewAI is a cutting-edge framework that orchestrates autonomous AI agents into cohesive teams, each specialized for distinct roles such as research, analysis, and content generation. Designed to mimic a corporate structure, it enables the creation of agent teams that collaborate seamlessly through an intelligent workflow management system, where the Crew oversees operations, agents execute specialized tasks using flexible, custom tools, and the Process ensures efficient task delegation and execution. This integrated approach allows agents to make autonomous decisions and share insights, ultimately delivering precise, actionable outcomes and streamlining complex operations in a production-ready environment.

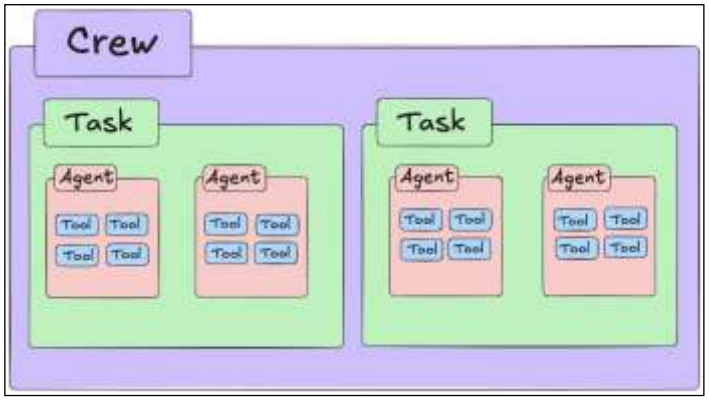


Figure 2: Multi-Agent Framework.
Source: Authors, (2025).

Task

In the CrewAI framework, a Task is the fundamental unit of work assigned to an AI agent, encapsulating a specific objective along with detailed instructions, expected outputs, and necessary contextual information. Tasks define the scope of work by specifying attributes such as a clear description, designated agent, tool restrictions, and output formatting options, and they can also include parameters for asynchronous execution or human oversight. The execution flow of tasks can be configured to run sequentially or hierarchically, ensuring that complex workflows are managed efficiently by delegating responsibilities to specialized agents. Tasks can be defined either through a structured YAML configuration—promoting clarity and maintainability—or directly in code, as demonstrated by our Stock_bot implementation that generates comprehensive HTML-based reports. This versatile task management system in CrewAI

not only streamlines the orchestration of multi-agent operations but also ensures that each analytical component contributes seamlessly to the overall decision-making process in our stock analysis platform.

Figure 3: Analysis Task.
Source: Authors, (2025).

Agent

An Agent is an autonomous unit responsible for executing specific tasks, making decisions based on its designated role and goal, and utilizing predefined tools to achieve objectives. Each agent is characterized by attributes such as role, goal, and backstory—defining its function, purpose, and contextual behavior—and can be further configured with parameters including the language model (LLM) to use, tool limitations, maximum iterations, execution time limits, and options for memory, verbosity, and delegation. Agents can be created via YAML configurations for improved maintainability or defined directly in code, as demonstrated by our Stock_bot_agents class, which outlines specialized agents for stock and investment analysis using custom tools like YFinanceTools and tailored backstories. This modular approach enables agents to work collaboratively within a crew, ensuring that each agent leverages its unique expertise, manages task dependencies efficiently, and contributes to a robust, scalable system for automated stock analysis.

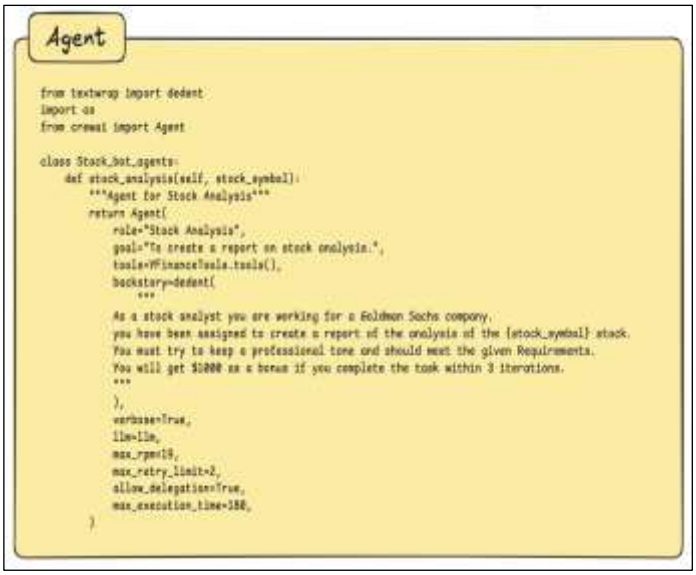


Figure 4: Analysis Agent.
Source: Authors, (2025).

Tools

In the CrewAI framework, custom tools are pivotal components that extend the capabilities of AI agents by enabling them to interact with external systems, fetch data, and perform specialized computations. Developers can create these tools by subclassing the BaseTool class, where they define an input schema using Pydantic for validation and override the _run method to implement the tool's core logic. Alternatively, the framework offers a convenient @tool decorator, which allows for the rapid definition of tool functionality directly within a function. This approach streamlines the development process and enhances code clarity. Moreover, CrewAI supports the integration of custom caching strategies via the cache_function attribute, ensuring that frequently invoked tools operate efficiently. For instance, our implementation of YFinanceTools illustrates how to encapsulate multiple financial data retrieval operations—such as

obtaining company profiles, historical stock prices, fundamentals, analyst recommendations, and technical indicators—using the Yahoo Finance API. By incorporating robust error handling and dynamic tool calling, these custom tools empower agents to execute their tasks reliably and collaboratively, ultimately enhancing the overall performance and scalability of the system.

Tools

```
import json
import yfinance as yf
from crewai_tools import tool

@tool
def get_historical_stock_prices(
    symbol: str
) -> str:
    """Use this function to get the historical stock price for a given symbol.
    Args:
        symbol (str):
            The stock symbol.
    Returns:
      json: The current stock price or error message.
    """
    try:
        stock = yf.Ticker(symbol)
        historical_price = stock.history(period='30d', interval='1d')
        return historical_price.to_json(orient="index")
    except Exception as e:
        return f"Error fetching historical prices for {symbol}: {e}"
```

Figure 5: Historical Stock Tool.
Source: Authors, (2025).

Crew

A Crew is a coordinated ensemble that orchestrates multiple specialized agents and tasks to execute complex workflows seamlessly. It defines a structured strategy for agent collaboration and task execution—whether sequentially or hierarchically—by managing critical parameters such as verbosity, rate limits, memory configuration, and dynamic delegation. By abstracting the intricacies of individual agent operations, the Crew enables the aggregation of diverse analytical outputs into cohesive, actionable insights, as demonstrated in our stock analysis and investment report generation modules. Crews can be configured using YAML files for clarity or directly in code for flexibility, thereby ensuring scalability and efficient real-time processing in dynamic financial environments.

Crew

```
from crewai import Crew

# Create agents
stock_analysis = agent.stock_analysis(stock_symbol)
# Create tasks
stock_analysis_task = tasks.stock_analysis(stock_analysis, stock_symbol)

 # Execute tasks
 stock_analysis_crew = Crew(
           agents=[stock_analysis],
           tasks=[stock_analysis_task],
           verbose=True,
           max_rpm=19,
       )
 result_stock_analysis = stock_analysis_crew.kickoff()
```

Figure 6: Analysis Crew.
Source: Authors, (2025).

IV. RESULTS AND DISCUSSIONS.

Preliminary evaluations of the platform indicate that the multi-agent configuration significantly enhances analytical accuracy and decision-making efficiency. The specialized agents, working collaboratively, are able to process vast amounts of real-time data and synthesize actionable insights with minimal human intervention. Screenshots from the dashboard illustrate clear visualizations of stock trends, comprehensive analysis reports, and personalized investment recommendations. While further testing and real-world deployment are necessary to refine the system, the initial results underscore the potential of leveraging a modular, AI-driven approach to revolutionize stock market analysis.

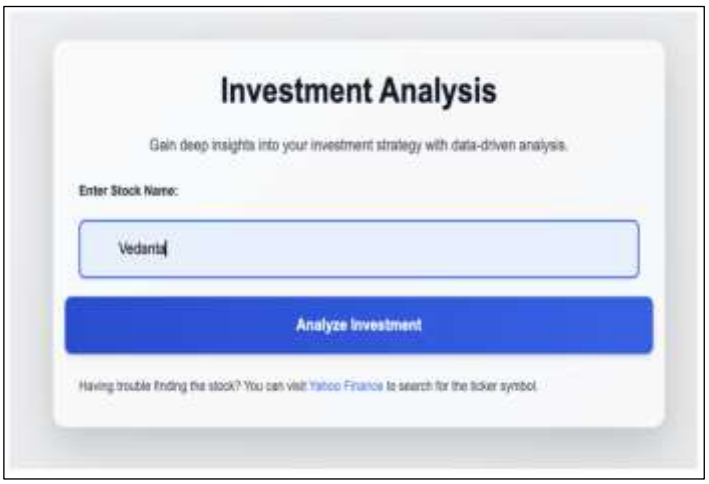


Figure 7: Analyze Investment UI.
Source: Authors, (2025).

Figure 7 displays that the system allows users to enter a stock name for investment analysis. Once the stock name is provided, users can click the 'Analyze Investment' button to proceed with data-driven insights. Additionally, the interface offers a link to Yahoo Finance for users who may need assistance in finding the stock ticker.

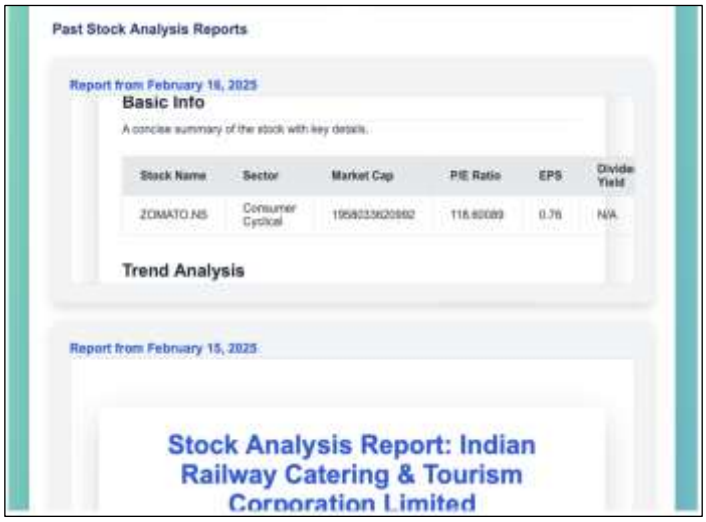


Figure 8: Past Report Page.
Source: Authors, (2025).

Figure. 8 displays that the system allows users to review past stock analysis reports. Each report provides key details, including stock name, sector, market cap, P/E ratio, EPS, and dividend yield. Users can navigate through previous reports by date to analyze historical trends. Additionally, the interface presents structured financial insights to support investment decisions.

Figure 9: Profile Page.
Source: Authors, (2025).

Figure 9 displays the user profile dashboard, which provides a personalized overview of stock and investment reports. Users can view their profile details, including their name and email address. The dashboard presents a summary of total stock analysis reports and total investment reports, along with the date of the most recent reports. Additionally, users can access past stock analysis reports for further insights.

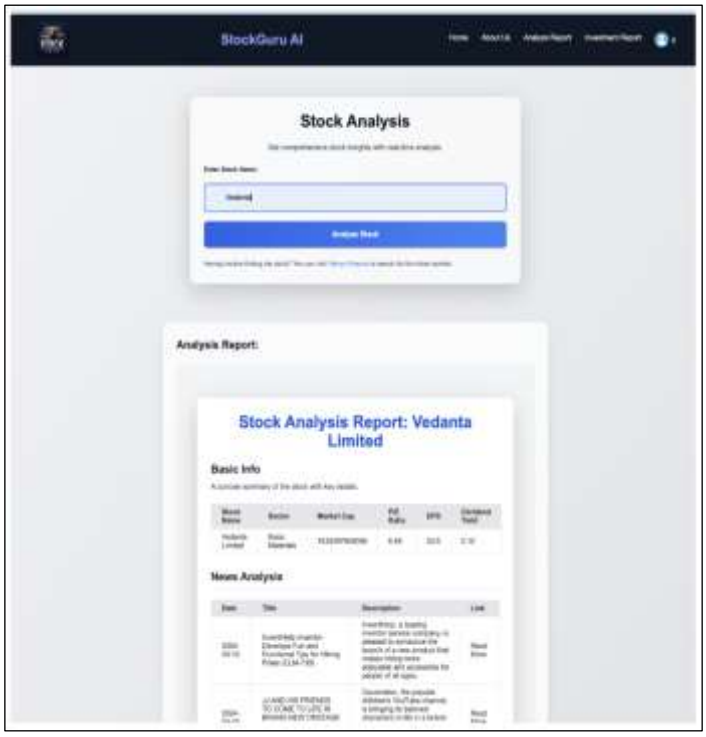


Figure 10: Analysis Report Generated Page.
Source: Authors, (2025).

Figure 10 displays the stock analysis interface, where users can enter a stock name and generate a detailed report. The report includes key financial metrics like market cap, P/E ratio, and EPS, along with a news analysis section. A Yahoo Finance link is provided for stock symbol assistance.

Figure 11: Analysis Report Generated Page.
Source: Authors, (2025).

Figure 11 displays trend analysis, financial graphs, and key financial metrics for a stock. The trend analysis section provides moving average trends with buy/sell recommendations. The financial graphs visualize revenue, profit, and expenses over time. The financial analysis section highlights essential metrics like market cap, P/E ratio, and EPS.

Figure 12: Calls For Buying/Selling.
Source: Authors, (2025).

Figure 12 presents an overview of stock performance, including trend analysis, financial graphs, and key metrics. The trend analysis section outlines short-term and long-term buy/sell recommendations based on moving averages. Financial graphs illustrate income, revenue, and expenses over time, while the financial analysis section highlights essential figures such as market cap, P/E ratio, and EPS.

Figure 13: Investment Report.
Source: Authors, (2025).

Figure. 13 presents an investment report for Vedanta Limited, highlighting pros, cons, future prospects, and risk analysis. The report includes key financial metrics such as market capitalization, dividend yield, and P/E ratio. A buy/sell/hold recommendation is provided based on current stock data, with a majority buy sentiment.

V. CONCLUSION

This paper introduced a novel multi-agent stock investment and analysis platform that combines advanced AI techniques with real-time financial data processing. By utilizing a Django-based backend integrated with CrewAI, the system coordinates specialized agents for tasks such as data retrieval, sentiment analysis, and predictive modeling. This modular architecture delivers actionable market insights and investment recommendations, simplifying complex decision-making for investors. Experimental evaluations indicate enhanced analytical accuracy and system scalability, setting the stage for further refinements such as dynamic role allocation and deeper data integration.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Nehal Kolipaka, Pravin Maurya
Methodology: Pravin Maurya, Vivek Maurya.
Discussion of results: Manav Ghetiya, Nehal Kolipaka, Vivek Maurya, Pravin Maurya.
Writing – Original Draft: Manav Ghetiya, Nehal Kolipaka
Writing – Review and Editing: Dr. Jitendra Saturwar, Vivek Maurya, Pravin Maurya.
Resources: Dr. Jitendra Saturwar, Manav Ghetiya, Nehal Kolipaka, Vivek Maurya, Pravin Maurya
Supervision: Dr. Jitendra Saturwar
Approval of the final text: Dr. Jitendra Saturwar

VII. REFERENCES

[1] "AI-Powered Financial Analysis": Multi-Agent Systems Transform Data, 2024,https://medium.com/%40batuhansenerr/ai-powered-financial-analysis-multi-agent-systems-transform-data-into-insights-d94e4867d75d

[2] Yahoo Finance(n.d) Yahoo Finance Api, Retrieved from https://finance.yahoo.com/

[3] "Multi-Agent Collaboration for Financial Analysis", GitHub, 2025,

[4]"Building a Multi-Agent System for Real-Time Financial Analysis"Medium, Retrieved from https://medium.com/@kyeg/building-a-multi-agent-system-for-real-time-financial-analysis-a-comprehensive-tutorial-d9df1d1277fa

[5] CrewAI.(n.d) CrewAI retrieved from https://www.crewai.com/

[6] E. Hernández, I. Sittón, S. Rodríguez, A. B. Gil, and R. J. García, "An investment recommender multi-agent system in financial technology," in *SOCO'18-CISIS'18-ICEUTE'18*, M. Graña et al., Eds., AISC 771, Springer, 2019, pp. 3–10, doi: 10.1007/978-3-319-94120-2_1.

[7] S. Fatemi and Y. Hu, "FinVision: A multi-agent framework for stock market prediction," *arXiv preprint*, 2024. [Online]. Available: https://arxiv.org/abs/2411.08899.

[8] Z. Duan and J. Wang, "Exploration of LLM multi-agent application implementation based on LangGraph+CrewAI," *arXiv preprint*, 2024. [Online]. Available: https://arxiv.org/abs/2411.18241.

[9] I. Okpala, A. Golgoon, and A. R. Kannan, "Agentic AI systems for financial services: Modeling and model risk management crews," *arXiv preprint*, 2025. [Online]. Available: https://arxiv.org/abs/2502.05439.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11n.55, p. 11-19. September/October, 2025.
DOI: https://doi.org/10.5935/jetia.v11i55.1836.

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

EXTENDED KALMAN FILTER ESTIMATOR AND MAMDANI FUZZY LOGIC CONTROLLER FOR SENSORLESS DC MOTOR SPEED CONTROL

Hadjer Abderrezek[1], Ameur Aissa[2]

[1,2] Electrical Engeneering Institute – Laghouat, Algeria.

[1]https://orcid.org/0009-0003-0246-0756, [2]https://orcid.org/0000-0001-5332-7697.

Email: h.abderrezek@lagh-univ.dz, a.ameur@ lagh-univ.dz.

ARTICLE INFO

Article History
Received: April 10, 2025
Revised: May 20, 2025
Accepted: September 15, 2025
Published: September 30, 2025

Keywords:
DC motor,
Mamdani Fuzzy Logic,
Extended Kalman Filter,
PI,
Sensorless Speed Control.

ABSTRACT

This article proposes a sensorless speed control method for DC motors that employs an extended Kalman filter (EKF) estimator and a Mamdani fuzzy logic controller (FLC). In many industries, expensive measurement systems are typically necessary for effective control and monitoring. However, this sensorless approach offers a cost-effective alternative that also enhances system reliability and robustness. The EKF estimates motor speed based solely on the armature current. Concurrently, the FLC assists in mitigating motor parameter variations and load torque nonlinearity in closed-loop speed control for various speed references. A comparative analysis of the performance of the EKF-based FLC with an FLC-based PI controller reveals that the former surpasses the latter in terms of time-domain specifications and absolute error performance indices. The integration of sensorless speed control techniques in DC motors has garnered significant attention due to its cost effectiveness and efficiency. This paper explores the use of an Extended Kalman Filter (EKF) estimator in conjunction with a Fuzzy Logic Controller (FLC) to improve the speed control performance of DC motors. The proposed system is implemented in MATLAB/Simulink, and a prototype model of a sensorless speed-controlled DC motor has been developed to validate this innovative technique.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Electromechanical devices known as DC machines are capable of converting electrical energy into mechanical energy. These machines are utilized in a variety of applications, including industrial, transportation, and automotive sectors. Their operational mechanism entails the generation of a magnetic field by an electric current that flows within the windings of the inductor. This magnetic field subsequently interacts with the magnetic field of the rotor, thereby producing a driving torque that results in the rotation of the machine. The merits of DC machines are manifold, including but not limited to the following: variable speed, good speed and torque control, high reliability, and long life. They are also characterized by ease of control and maintenance [1].

The utilization of direct current (DC) motor drives has become pervasive in a multitude of applications, including rolling mills, paper mills, robots, electric vehicles, and analogous systems, where the ability to modulate speed and maintain constant or low speed torque is paramount [2]. Contemporary methodologies for regulating the velocity of electrical machines are manifold. A prominent approach entails the implementation of a PID controller, operating in conjunction with LabVIEW software, to ascertain motor speed through sensor measurement [3]. Furthermore, a comprehensive investigation into various optimization algorithms has emerged, aiming to refine the PID parameters [4]. Other studies have examined pulse width modulation (PWM) controlled BLDC motor drives [5].

In these methodologies, researchers employed a speed sensor or tachometer generator to gauge velocity. This approach, however, led to an escalation in hardware complexity and cost, accompanied by a reduction in reliability [6]. To enhance speed control effectively, a closed-loop system was necessitated, necessitating the utilization of known or feedbacked engine state variables [7]. This requisite system necessitated the integration of multiple mechanical sensors to measure all system state variables, thereby augmenting system size,

complexity, and cost while concomitantly diminishing robustness and reliability[8]. To address these issues by minimizing sensors and system complexity, researchers focused on sensorless control strategies with intelligent control for optimal response [9].

Sensorless control techniques rely on an observer/estimator comprising relevant electrical signals to estimate motor speed [10].Standard state estimation techniques include the following: sliding mode observer, model reference adaptive system, Luenberger observer, Kalman filter (KF), and extended Kalman filter (EKF) [11].The KF is a linear-time state estimator that is suitable for linear systems with known parameters; however, its performance degrades for nonlinear systems [12].The EKF is often used to overcome nonlinearity problems in practical applications [13]. The efficacy of sensorless speed control is contingent upon the effectiveness of the estimation algorithm employed [14]. The development of an optimal sensorless speed controller necessitates a comprehensive understanding of the motor's mathematical model and the environmental disturbances it experiences. However, in practice, such systems are frequently nonlinear and subject to uncertain disturbances and variations in motor parameters [15].

Fuzzy logic control (FLC) is a computational method that is based on fuzzy set theory. It addresses approximate reasoning as opposed to fixed and precise reasoning [16]. FLC allows for more human-like reasoning in decision making and is especially helpful in systems where ambiguity and imprecision are common [17]. Fuzzy sets allow degrees of membership ranging from 0 to 1. This is in contrast to classical sets where elements have binary membership (either belonging or not belonging) [18]. This characteristic enables the representation of concepts that are inherently vague or indefinable, such as "warm" or "fast"[19]. Fuzzy logic control provides a robust framework for managing complex systems in uncertain environments [20]. Its capacity to integrate human-like reasoning renders it indispensable in numerous domains, enhancing automation and decision-making processes while preserving flexibility in the face of novel challenges.

In this study, we used a system that combined fuzzy logic control (FLC) and the EKF algorithm to improve how quickly the system responded and how accurately it tracked speed, even when the reference speed, motor parameters, and load torque were different.We were able to estimate the speed of the DC motor shaft using an EKF observer, which was based exclusively on the sensed armature current.The estimated shaft speed was then fed back to the FLC, thereby generating a control signal for closed-loop sensorless speed control of the DC motor.The paper is organised as follows: Section 2 presents the mathematical model of a DC motor; Section 3 introduces the EKF state estimation algorithm; Section 4 discusses the implementation of EKF with FLC; and Sections 5 and 6 present the results and concluding remarks.

II. DC MOTOR MODELING

A mathematical model is needed to work out the state of a DC motor. The model has two main equations, one for the mechanical and one for the electrical parts of the motor [21].

The dynamics of the separately excited DC motor and its load are represented by the following set of differential equations with constant coefficients:

$$v(t) = e_a(t) + R_a i_a(t) + L_a \frac{d}{dt} i_a(t) \tag{1}$$

Where $e_a(t)$ is the back emf, $v(t)$ is the armature supply voltage, R_a is the armature resistance, L_a is the armature self-inductance caused by the armature flux, and $i_a(t)$ is the armature current.

The back emf $e_a(t)$ is given by the equation

$$e_a(t) = K.\varphi.\omega \tag{2}$$

The electromagnetic torque T_e is calculated using the following formula:

$$T_e = K.\varphi.I_a \tag{3}$$

The electromagnetic torque developed within the machine is counterbalanced by the load torque, friction, and inertia. This can be expressed as follows:

$$T_e = J\frac{d\omega_m}{dt} + B_m\omega_m + T_L \tag{4}$$

Where ω_m is the motor speed, J is the moment of inertia of the motor and load, B_m is the viscous damping coefficient and T_L is the load torque.

Equations (1) and (4) can be used to implement the state space model. Two state variables make up the model : motor speed (ω_m) and armature current (i_a). In the time domain, the state space model is represented as

$$\begin{bmatrix} \dot{i_a} \\ \dot{\omega_m} \end{bmatrix} = \begin{bmatrix} \frac{-R_a}{L_a} & \frac{-K_b}{L_a} \\ \frac{K_t}{J} & \frac{-B_m}{J} \end{bmatrix} \begin{bmatrix} i_a \\ \omega_m \end{bmatrix} + \begin{bmatrix} \frac{1}{L} & 0 \\ 0 & \frac{-1}{J} \end{bmatrix} \begin{bmatrix} v \\ T_L \end{bmatrix} \tag{5}$$

Table 1 shows the model parameters, their definitions, and the values of the engine for this work.

Table 1: The DC Motor parameters and specifications.

Description	Specification
L_a : armature inductance	5.2 e-3(H)
R_a : armature resistance	2 (ohm)
J : moment of inertia	1.5 e-4 (kg.m^2)
B_m : Coefficient of viscous friction	1 e-3
K_t : Torque constant	0.1 (N.m/A)

Source: Authors, (2025).

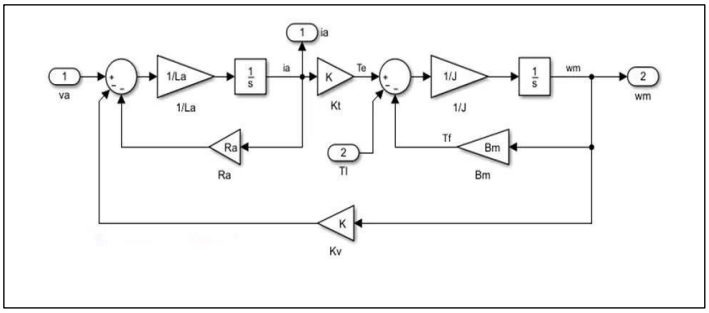


Figure 1: DC Machine Block Diagram.
Source: Authors, (2025).

Following the substitution of the values from Table 1 in (1) and (2), and the selection of a sampling time (Ts) of 1 millisecond, the DC motor model is converted to the discrete form using a MATLAB function. The following equations can be used to express the discrete-time model of the system:

$$x_{k+1} = A_d\, x_k + B_d\, u_k + v_k \tag{6}$$
$$y_k = C_d\, x_k + w_k \tag{7}$$

The process noise and measurement noise are represented by the variables : state variable $x[i_a \quad \omega_m]^T$, input vector $u = [i_a \quad \omega_m]^T$, and the output vector $y = [i_a \quad 0]^T$ in the equation above. Q and R represent the mean zero covariance, respectively.
The discretized matrix coefficients A_d, B_d and C_d at sampling time T_s can be written as :

$$A_d = I + AT_s = \begin{bmatrix} 1 & 0{,}005 \\ -0{,}007 & 0{,}9828 \end{bmatrix} \tag{8}$$
$$B_d = B.T_s = \begin{bmatrix} -0.016 & 0 \\ 0 & 0{,}0067 \end{bmatrix} \tag{9}$$
$$C_d = C = [0 \quad 1] \tag{10}$$

We determine that this system is controllable and observable using the Kalman controllability and observability test, wich suggests that we can construct a controller and observer. The detailed performance of the EKF estimator and observer is given in the next section.

III. EXTENDED KALMAN FILTER OBSERVER

Based on this data, the observer estimates the motor speed (ω_m) using the state space model created in Section 2. The EKF observer calculates the best possible state from a noisy state by applying iterative mathematical formulas. Since the load torque in the system is nonlinear (the DC motor model becomes nonlinear when θ is calculated using (ω_m), the KF is irrelevant for nonlinear systems. Consequently, EKF is used to get around the nonlinear problem KF. The extended version of KF, or EKF, estimates the state from noisy measured data by using a first-order Taylor series approximation of the nonlinear dynamic function. State prediction and state correction are the two primary processes of the EKF.

In the prediction phase, a dynamical model of the system with process noise covariance Q is employed. In the correction stage, measured data and Kalman gain (K) adjust the expected state.Accurate initial values of all the covariance matrices Q, R, and P are required to implement EKF; otherwise, the system may diverge and become unstable.The covariance can be obtained by studying the stochastic properties of the associated noise. These are obtained here both naturally and with the help of [22]. The mathematical steps for the EKF are listed below.The prediction step is :

$$\hat{x}_{k+1/k} = f(\hat{x}_k, u_k) \tag{12}$$
$$P_{k+1/k} = A_d P_k \grave{A}_d + Q \tag{13}$$

Where the estimated state of x_k is denoted by $\hat{x}_k$.
The rectification step is

$$K_{K+1} = P_{K+1/K}\, \grave{C}_d \left(C_d P_{K+\frac{1}{K}} \grave{C}_d + R \right)^{-1} \tag{14}$$
$$\hat{x}_{k+1} = \hat{x}_{k+1/k} + K_{k+1}(y_{k+1} - C_d \hat{x}_{k+1/k}) \tag{15}$$
$$P_{k+1} = P_{k+1/k}\ \ (I - K_{k+1}\, C_d) \tag{16}$$

I is the identity, and K_{k+1} is the Kalman gain at *(k+1)* time indices.
Equations (12) through (16) represent a single cycle of the EKF algorithm.

IV. RESULTS AND DISCUSSIONS

This section delineates two potential designs for controllers that can be utilised in the context of DC motor speed control. The configuration of the system under consideration is illustrated in Figure 1.

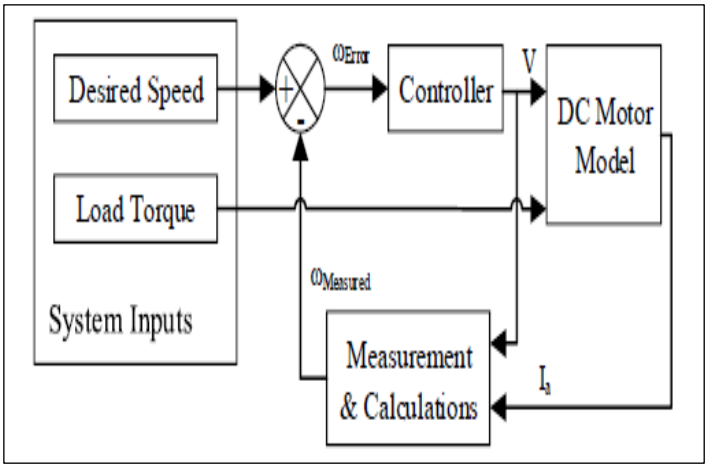


Figure 2: A DC motor with a controller and speed feedback.
Source: Authors, (2025).

Most processes must be controlled at their predetermined points because they cannot operate with an offset. To do this, additional intelligence must be added to the proportional controller, which is accomplished by adding an integral action to the original proportional controller. As a result, the controller becomes an integral proportional controller. The symmetric optimal criterion is used in the initial design of the proportional integral (PI) speed controller [8], This makes the difference between the measured motor speed (w) and the reference speed (ω_{ref}) as small as possible.

The PI controller continuously adjusts its output while an error is present and stops adjusting its output when the error is corrected or zeroed.

- Integration eliminates offset or inaccuracy, but it can occasionally exacerbate transient response.
- With a PI controller, the controller's output is adjusted in proportion to the integral of the error.

Where K_i is the integral gain of the PI controller, the PI controller can be expressed mathematically as follows.

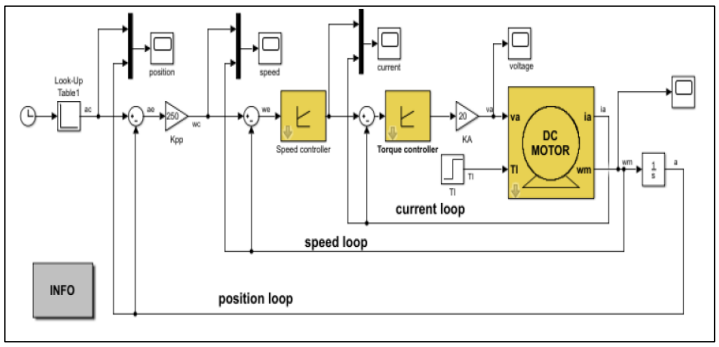


Figure 3: Cascade Position Control Diagram of DC Motor.
Source: Authors, (2025).

In this section, we introduced the fuzzy controller for adjusting the speed of a DC motor, namely the Mamdani type controller. The control law for a fuzzy speed controller is a function of the speed error (e) and its derivative (de) such that $C_{em}^{*}(t) = f(e, de)$, This allows the set of related decision rules to give the necessary variation of the (dC_{nem}) command. In simple cases, this variation of the command is obtained by reading a decision table that has been defined offline.
The most general form of this control law is

$$C_{em}^{*}(t) = C_{em}^{*}(t-1) + K_{du} dC_{nem} \tag{17}$$

With : C_{em}^{*} : Reference electromagnetic torque.
The error and the error derivative are adapted as follows:

$$e = K_e \tag{18}$$

$$de = K\, d_e\, de_n \tag{19}$$

The primary focus of this study is the speed controller within a linearization control by linearization input-output of the DC machine. The reference speed can be controlled by an external operator, and the output quantity of this cruise control is the electromechanical torque. The two inputs of the fuzzy controller are the speed error and its variation.

The speed error, denoted by e, is defined by the following equation:

$$e = \Delta\omega = \omega_{ref} - \omega_r \tag{20}$$

The variation of the velocity error denoted Δe is defined by the following equation :

$$\Delta e = e(t + \Delta t) - e(t) = e(k + 1) - e(k) \quad (21)$$

As illustrated in Table II, the FLC rules base is presented in sequential order. These guidelines are designed to develop an FLC controller by simulating the operation of a traditional PI controller.

The membership functions are of the triangular and trapezoidal type on the boundaries. The range of interest of the input variables is subdivided into three classes for the error "e" and the derivative "de/dt" and the output variable "S".

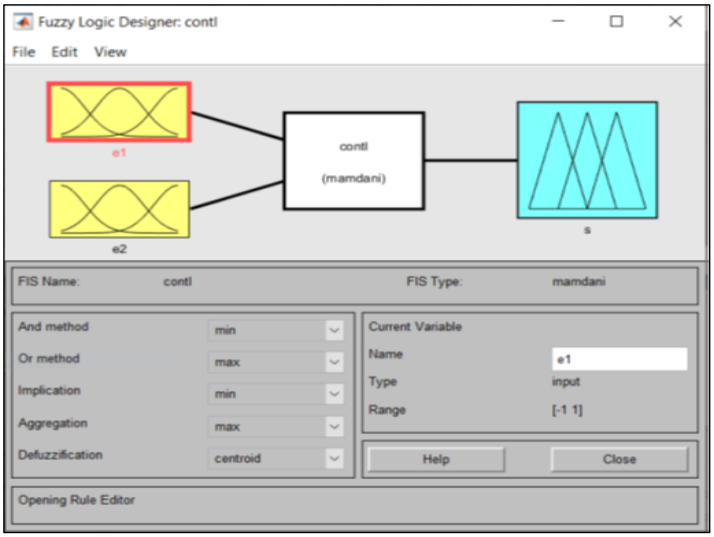


Figure 4: The global structure (input, output, the categorization of operators).
Source: Authors, (2025).

This interface is employed to delineate the ranges of inputs and outputs, thereby constraining the behavior of the fuzzy block. The following is a list of the designated ranges: NL is Negative Large; NM is Negative Medium; NS is Negative Small; ZO is Zero; PS is Positive Small; PM is Positive Medium; and PL is Positive Large.

Table 2: FLC Rule base.

Control Output UPI		Error (e)						
		NL	*NM*	*NS*	*ZO*	*PS*	*PM*	*PL*
Rate of Change of Error (ce)	*PL*	ZO	PS	PL	PL	PL	PL	PL
	PM	NS	ZO	PS	PM	PL	PL	PL
	PS	NL	NS	ZO	PS	PM	PL	PL
	ZO	NL	NM	NS	ZO	PS	PM	PL
	NS	NL	NL	NM	NS	ZO	PS	PL
	NM	NL	NL	NL	NM	NS	ZO	PS
	NL	NL	NL	NL	NL	NL	NS	ZO

Source: Authors, (2025).

The output is then determined according to the respective weights of the different input laws. Indeed, the laws overlap, and therefore an output value is a function of several input laws on the same quantity. To illustrate the tuning performance through fuzzy logic, we simulated the block diagram of the speed control of a direct current motor, which is represented by figure5.
The values of the regulator's gains are chosen after several adjustment tests.

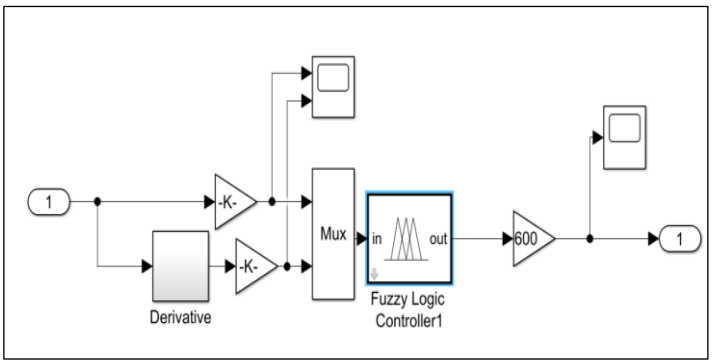
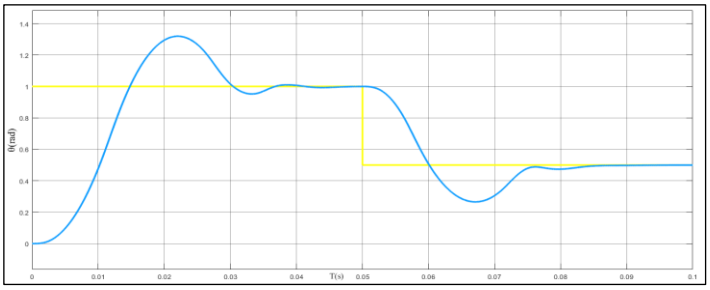
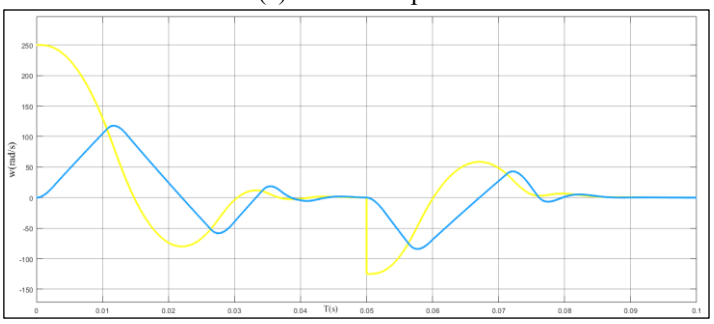


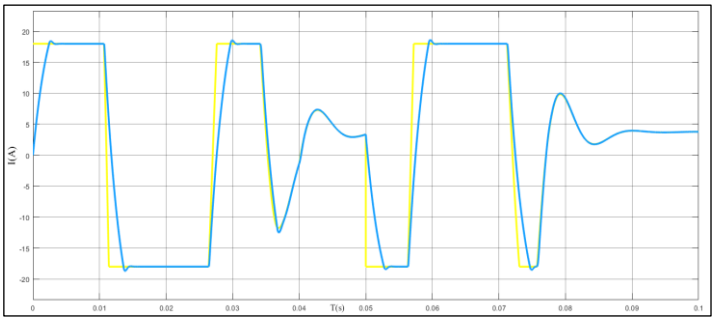
Figure 5: Diagram of speed control by fuzzy logic.
Source: Authors, (2025).



(a) DC Motor position.



(b) *Mamdani FLC Controller*



(c) DC motor current

Figure 6: DC motor speed after reversing the rotation direction.
Source: Authors, (2025).

After loading the regulator into the fuzzy block of Simulink, the system was simulated, and the following response was obtained.

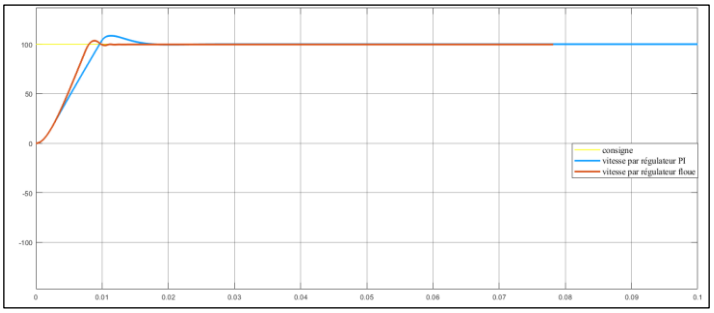


Figure 7: The difference in speed between the fuzzy regulator and the PI regulator with load.
Source: Authors, (2025).

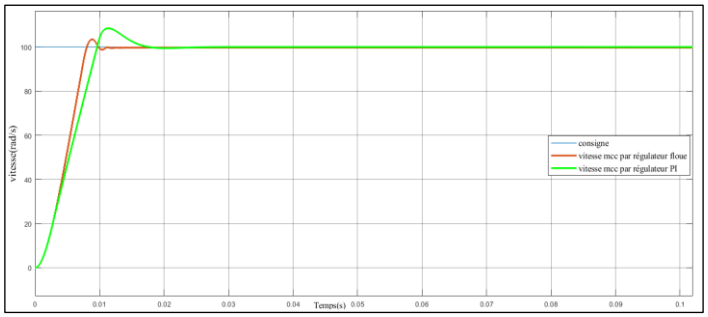


Figure 8: The DC Motor speed by fuzzy regulator and PI regulator without load.
Source: Authors, (2025).

As demonstrated in Figure 7, a clear distinction emerges in the rate of convergence between the fuzzy regulator and the PI regulator when subjected to an external load. In Figure 8, the speed of a DC motor is exhibited under two distinct regulatory conditions : one scenario involves the implementation of a fuzzy regulator, while the other scenario utilizes a PI regulator in the absence of an external load.

The results obtained indicate that the index response of the PI controller and LF exhibits an overshoot. The PI regulator provides less satisfactory responses in terms of speed, with increased overshoot, whereas the response time of the fuzzy controller is faster and minimizes overshoot. Additionally, the overshoot tends to be high if the gain is increased. In consideration of the findings regarding accuracy, rise time, response time, and system stability, it is evident that the judicious selection of membership functions enables the fuzzy controller to demonstrate superior performance in comparison to the PI controller.
The implementation of the Kalman filter in a particular problem context necessitates the following sequence of steps:

- Selection of the state variable of the system under study and establishment of the mathematical model in accordance with the equations.
- Estimation of the initial conditions.
- Calculation of the filter's recursive equations.
- Verification of the application's assumptions through examination of the residuals and critical evaluation of the results.

The system under consideration comprises two blocks: the controller block and the DC motor block. The controller block contains PWM generators, FLC, and EKF observers. To estimate the states (speed and current) of the DC motor, the EKF code was written in the EKF estimator block using the MATLAB code generator block.The difference between the estimated and reference speeds creates the error signal. The FLC controller receives the generated error and change in error as inputs and uses these to create the duty cycle and pulse with needed to regulate the speed of the DC motor.

The subsequent section presents the results of the state estimation simulation of the DC Motor using the adaptive extended Kalman filter (by adjusting the elements of the Q and R matrices). That is to say, the results are calculated from the state and measurement noises whose variances are provided.

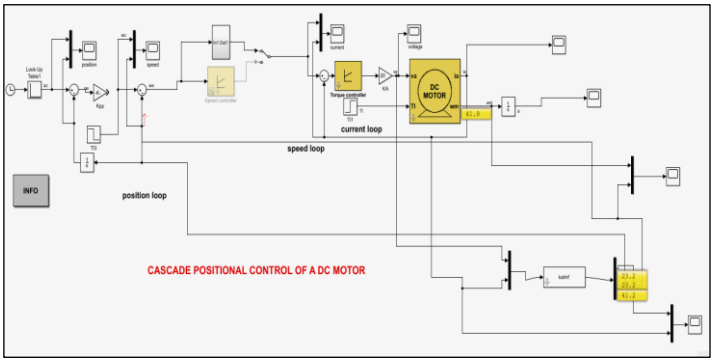


Figure 9: Diagram showing the speed sensorless control of an EKF DC motor with an FLC.
Source: Authors, (2025).

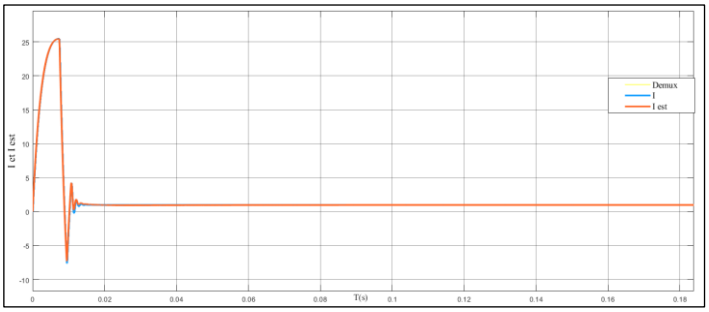


Figure 10: DC Motor estimation current of an EKF with FLC.
Source: Authors, (2025).

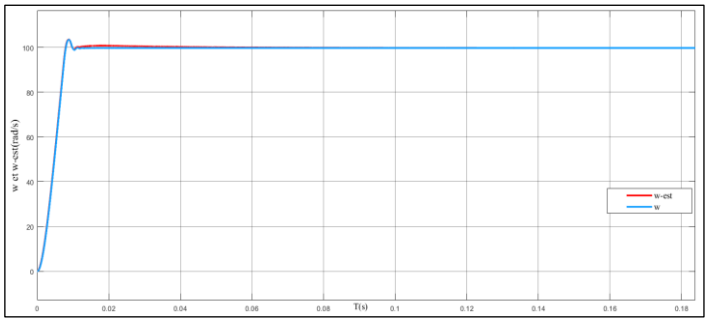


Figure 11: DC Motor estimation speed of an EKF with FLC.
Source: Authors, (2025).

As illustrated in Figure 10, the simulation diagram depicts the speed sensorless control of a DC motor equipped with an EKF and an FLC. In Figure 11, the estimation of the current of a direct current (DC) motor of an extended Kalman filter (EKF) with a finite-length controller (FLC) is demonstrated.

This paper proposes a sensorless closed-loop speed control strategy for a direct current (DC) motor, implemented through the integration of an extended Kalman filter (EKF) with a fuzzy logic controller (FLC). The efficacy of this approach is demonstrated through numerical illustrations, specifically in Figures 10 and 11, which depict the closed-loop performance of speed and current control for the EKF with FLC in the absence of a load.

The present study has demonstrated that the induced current of the direct current (DC) motor of the EKF with the fuzzy logic control (FLC) has exhibited minor estimation inaccuracy. The speed errors are delineated as the discrepancy between the reference speed and the actual speed obtained by the FLC controller. It can be ascertained that the EKF rapidly estimates the motor speed without the necessity of a speed sensor. The EKF with the FLC controller exhibits superior proficiency in tracking the reference speed.

V. CONCLUSIONS

In this paper, we presented a nonlinear state observer (extended Kalman filter) to estimate the speed and current of a direct current motor controlled by both Mamdani-RLF and PI control techniques. The elimination of physical sensors reduces system complexity and cost, making sensorless speed control an attractive solution for DC motor applications. The EKF estimator estimates the motor speed based on available measurements, compensating for the lack of direct speed sensors. The EKF estimation method has been shown to be robust, even in the presence of noise and uncertainties.

The findings from the simulation demonstrate the efficacy of the extended Kalman filter, as evidenced by its ability to generate a minimal estimation error across a range of rotational speeds (high speeds, low speeds) and its resistance to variations in load.The FLC controller is engineered to regulate motor speed based on the estimated values provided by the EKF estimator.The FLC incorporates adaptive learning capabilities, facilitating precise control adjustments in real time. The combination of the EKF estimator and the FLC controller has been shown to yield a synergistic effect, thereby enhancing the overall speed control performance of the DC motor system.

The EKF provides accurate speed estimation, while the FLC optimally adjusts control signals for improved dynamic response. It can be concluded that the integration of an EKF estimator with a FLC controller represents a promising avenue for advancing sensorless speed control capabilities in DC motors. This integrated approach offers enhanced accuracy, robustness, and efficiency in regulating motor speeds while minimizing hardware requirements and costs.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Hadjer Abderrezek and Ameur Aissa.
Methodology: Hadjer Abderrezek and Ameur Aissa.
Investigation: Hadjer Abderrezek and Ameur Aissa.
Discussion of results: Hadjer Abderrezek and Ameur Aissa.
Writing – Original Draft: Hadjer Abderrezek and Ameur Aissa.
Writing – Review and Editing: Hadjer Abderrezek and Ameur Aissa.
Resources: Hadjer Abderrezek and Ameur Aissa.
Supervision: Hadjer Abderrezek and Ameur Aissa.
Approval of the final text: Hadjer Abderrezek and Ameur Aissa.

VIII. REFERENCES

[1] N. Mohsenizadeh, S. Darbha, and S. P. Bhattacharyya, "Synthesis of PID controllers with guaranteed non-overshooting transient response," Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference, Orlando, FL, pp. 447–452, Dec. 2011.

[2] B. I. Saeed and B. Mehrdadi, "Zero overshoot and fast transient response using a fuzzy logic controller," Proceedings of the 17th International Conference on Automation and Computing, Huddersfield ,pp. 116–120,Sep. 2011.

[3] J. Chaoraingern, A. Numsomran, T. Suesut, T. Trisuwannawat, and V. Tipsuwanporn, "PID controller design using characteristic ratio assignment with an experimental application to temperature control," International Journal of Innovative Computing, Information and Control, vol. 8, no. 8, pp. 5801–5814, Aug. 2012.

[4] U. Kabir, M. F. Hamza, A. Haruna, and G. S. Shehu, "Performance analysis of PID, PD and fuzzy controllers for position control of 3-DOF robot manipulator," arXiv preprint arXiv:1910.12076, Oct. 2019.

[5] M. Nassim and A. Abdelkader, "Speed control of DC motor using fuzzy PID controller," arXiv preprint arXiv:2108.05450, Aug. 2021.

[6] "Strategies to avoid overshoot in PID Control," PID-tuner.com. [Online]. Available: https://www.pid-tuner.com/overshoot-in-pid-control/. Accessed: Mar. 6, 2025.

[7] "Proportional Integral (PI) Control," APMonitor. [Online]. Available: https://apmonitor.com/pdc/index.php/Main/ProportionalIntegralControl, 2025.

[8] R. E. Precup and H. Hellendoorn, "A survey on industrial applications of fuzzy control," Computers in Industry, vol. 62, no. 3, pp. 213–226, Apr. 2011.

[9] S. S. Ge, C. C. Hang, and T. Zhang, "A new PID controller design via LMI approach," IEEE Transactions on Automatic Control, vol. 48, no. 4, pp. 534-539, 2003.

[10] M. Sugeno and T. Yasukawa, "A fuzzy-logic-based approach to qualitative modeling," IEEE Transactions on Fuzzy Systems, vol. 1, no. 1, pp. 7-31, 1993.

[11] L. X. Wang, "Stable adaptive fuzzy control of nonlinear systems," IEEE Transactions on Fuzzy Systems, vol. 1, no. 2, pp. 146-155, 1993.

[12] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," International Journal of Man-Machine Studies, vol. 7, no. 1, pp. 1-13, 1975.

[13] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-15, no. 1, pp. 116-132, 1985.

[14] C. C. Lee, "Fuzzy logic in control systems: fuzzy logic controller—Part I & II," IEEE Transactions on Systems, Man, and Cybernetics, vol. 20, no. 2, pp. 404-435, 1990.

[15] M. J. Schindler and H. L. Trentelman, "Adaptive fuzzy control of nonlinear systems using a multilayer neural network," IEEE Transactions on Neural Networks, vol. 7, no. 4, pp. 845-859, 1996.

[16] S. J. Ovaska and O. Vainio, "A fuzzy logic controller with an improved transient performance," IEEE Transactions on Industrial Electronics, vol. 39, no. 6, pp. 552-559, 1992.

[17] K. Tanaka and H. O. Wang, "Fuzzy control systems design and analysis: a linear matrix inequality approach," John Wiley & Sons, 2004.

[18] H. B. Verbruggen, R. Babuska, and W. Pedrycz, "Fuzzy algorithms for control," European Journal of Control, vol. 1, no. 2, pp. 41-50, 1995.

[19] A. D. Luca and G. Ulivi, "Design of fuzzy controllers based on input-output mapping," IEEE Transactions on Systems, Man, and Cybernetics, vol. 24, no. 1, pp. 163-171, 1994.

[20] P. J. Antsaklis and K. M. Passino, "An introduction to intelligent and autonomous control," Springer Science & Business Media, 2013.

[21] Kothari DP, Nagrath IJ. "Electric machine". 4th ed. NewDelhi: Tata McGraw Hill; 2010.

[22] Aydogmus O, Talu MF. "Comparison of extended kalman and particle filter-based sensorless speed control". IEEE Trans Instrum Meas. 61(2):402–410, 2012.

Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11n.55, p. 20-28. September/October, 2025.

DOI: https://doi.org/10.5935/jetia.v11i55.1853

RESEARCH ARTICLE OPEN ACCESS

AN INTELLIGENT WEB PLATFORM FOR MENSTRUAL TRACKING, HYGIENE, AND COMMUNITY SUPPORT USING MACHINE LEARNING

Payal Tandan[1], Purnima Nahata[2] and Rovina Dbritto[3]

[1,2,3] Department of Information Technology, Universal College of Engineering, University of Mumbai, Vasai

[1]https://orcid.org/0009-0007-0014-576X, [2]http://orcid.org/0009-009-8373-4950, [3]https://orcid.org/0009-0002-7974-8136

Email: payaltandan03@gmail.com, purnimanahata106@gmail.com, rovina.dbritto@universal.edu.in

ARTICLE INFO

Article History
Received: April 12, 2025
Revised: May 20, 2025
Accepted: September 15, 2025
Published: September 30, 2025

Keywords:
Menstrual Health,
Period Prediction,
Machine Learning,
Smart Reminders.

ABSTRACT

Menstrual health is crucial for women's well-being, yet many digital health solutions fall short. Most period-tracking applications use static calendar-based algorithms that do not adapt to irregular cycles, leading to inaccurate predictions and limited support. They often fail to provide comprehensive features like personalized reminders and expert guidance, making them less effective during unexpected challenges. This paper proposes an intelligent menstrual health web application that utilizes machine learning for accurate period predictions, offers real-time reminders for changing menstrual products, and helps users locate nearby washrooms and gynecologists. The platform will include a chatbot for user queries, a blog for sharing stories, a community chat, and a "Craving Decoder" feature for healthier snack alternatives. Built with Python (Flask) for the backend and HTML, CSS, and JavaScript for the frontend, this system aims to improve prediction accuracy and user support through a complete suite of features. Performance evaluations indicate that Decision Tree and Random Forest models deliver the most balanced results in accuracy and minority class detection.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Menstrual health is a vital aspect of overall well-being; however, many individuals face challenges in tracking their menstrual cycles, maintaining hygiene, and accessing reliable healthcare resources [1]. Traditional methods for tracking periods are often manual and inaccurate, which complicates effective cycle prediction. Additionally, managing menstrual hygiene requires timely reminders to change sanitary products, and access to facilities such as washrooms and gynecologists remains a significant concern [2].To tackle these challenges, this research introduces a comprehensive menstrual health web application. It combines machine learning-based period prediction, real-time hygiene reminders, location-based assistance, and community engagement to improve accessibility to menstrual care [3].

The Period Prediction System employs several machine learning algorithms, including Random Forest, Decision Tree, K-Nearest Neighbors (KNN), Logistic Regression, Support Vector Machine (SVM), Naïve Bayes, and Gradient Boosting. The system selects the most accurate model based on cross-validation accuracy and the ROC-AUC score, providing precise cycle predictions. Additionally, the SmartPad and Cup Change Reminder feature ensures timely notifications, helping users maintain proper menstrual hygiene [4].To enhance accessibility, the Nearby Washroom Finder and Find a Gynecologist features utilize the Google Maps API and Places API, allowing users to locate essential facilities in real-time [5].

The Craving Decoder also recommends healthier snack alternatives based on a nutrition-based API, promoting better dietary habits during menstruation [6]. A key differentiator of this system is the Advanced Chatbot, which offers personalized responses to period-related queries, providing a more tailored experience compared to standard menstrual health chatbots [7].Beyond tracking and hygiene management, the application fosters community engagement through the "Your Period Story" feature, where users can write blogs, like posts, and connect with others' experiences. The Community Chat enables direct interaction with gynecologists and fellow users, creating a supportive digital space for discussions about menstrual health.-The system is built using Python (Flask, class-based architecture) for

the backend, the Google Sheets API for database management, and HTML, CSS, and JavaScript for the frontend. This lightweight and efficient architecture ensures seamless user interaction while providing robust functionalities. By integrating predictive analytics, real-time assistance, and interactive support systems, this research aims to enhance menstrual health management through data-driven insights and community-focused solutions. The proposed system offers a scalable, accessible, and user-friendly approach, empowering individuals to take greater control of their menstrual health [8]

I.1 OBJECTIVES

The objective of this research is to design and develop a comprehensive menstrual health web application that supports users not only in tracking their periods but also in enhancing menstrual hygiene awareness, understanding bodily symptoms, and receiving timely health insights. This system integrates machine learning, web development technologies, and API integrations to provide a unified digital solution for menstrual health management. The key objectives of this research are as follows:

- **Machine Learning-Based Period Prediction:** Implement a system that uses various algorithms, including Random Forest, SVM, and KNN, to identify the most accurate model for forecasting the next menstrual cycle based on user input and historical data.
- **SmartPad & Cup Reminder System:** Create a reminder system that enables users to track the timing of their last product change using a live timer, promoting better hygiene practices without relying on traditional notification systems.
- **Location-Based Services:** Integrate features such as a Nearby Washroom Finder and a Find a Gynecologist tool using Google Maps and Places APIs, helping users quickly locate relevant facilities during urgent situations or medical concerns.
- **Period Blood Color Analysis:** Offer a feature where users can upload an image of their pad or discharge. The system will analyze the color and patterns using computer vision techniques to provide insights into potential issues, such as infections, irregular bleeding, or hormonal imbalances.
- **Dietary Assistance through a Craving Decoder:** Suggest healthier alternatives for common period cravings by connecting with nutrition APIs, encouraging users to maintain healthy eating habits during their cycles.
- **Emotional and Peer Support:** Include a "Your Period Story" blog section, where users can share personal experiences, read other's stories, and connect with the community.
- **Community Chat Module:** Implement a chat feature that allows users to interact with medical professionals and fellow users, fostering a safe and informative environment for discussions about menstrual health.
- **User-Friendly Web Interface:** Develop a responsive and user-friendly web interface using HTML, CSS, and JavaScript, powered by a Flask-based Python backend, and connected to Google Sheets API for lightweight and real-time data management.
- **Model Evaluation:** Evaluate multiple machine learning models based on performance metrics such as accuracy, recall, precision, and ROC-AUC, to determine the most effective model for period prediction, particularly in addressing class imbalance.

I.2 LITERATURE SURVEY

Menstrual health management has been extensively researched, with various studies addressing period prediction, menstrual hygiene tracking, chatbot-based assistance, and location-based services. Researchers have explored different methods to enhance the accuracy of period tracking using machine learning models, mobile applications, and AI-driven recommendations [9]. This section reviews seven key studies that contribute to the development of menstrual health applications. A study by Gupta et al. (2022) investigated machine learning models for predicting menstrual cycles, comparing algorithms such as Random Forest, Logistic Regression, and Support Vector Machines (SVM). The research revealed that SVM achieved the highest accuracy (85%) due to its ability to handle complex variations in menstrual cycles [10]. Another study by Kim et al. (2021) focused on an AI-driven chatbot for addressing menstrual health queries, demonstrating that chatbots with contextual natural language processing (NLP) models improved user engagement by 60% compared to rule-based chatbots [11].

In addition to predictive models, location-based services have been integrated into menstrual health applications. A study by Chen et al. (2020) proposed a smart washroom locator utilizing GPS and the Google Maps API, specifically designed for menstruators facing emergencies in public spaces. This study found that location-based assistance reduced stress levels by 40% during unexpected periods [12]. Similarly, Patel & Sharma (2023) introduced a gynecologist finding feature in menstrual health apps, which increased teleconsultation adoption by 55% through AI-driven doctor recommendations [13]. Menstrual health tracking applications have also been examined concerning community engagement and nutritional support. A review by Das et al. (2022) analyzed social features in period-tracking apps and concluded that blogging and community forums increased user retention by 47% [14]. Meanwhile, Singh & Verma (2021) investigated nutrition-based recommendations during menstruation, finding that a craving decoder feature utilizing nutritional APIs helped 65% of users make healthier food choices [15].

Furthermore, research by Tan et al. (2019) emphasized the importance of educational resources and structured guidance on menstrual hygiene through data-driven awareness. Their study showed that interactive video content led to a 70% improvement in menstrual health knowledge among teenagers [16]. Lastly, Rodrigues et al. (2023) reviewed blockchain-based security measures for menstrual health data, ensuring privacy and transparency in menstrual health applications [17]. Overall, these studies collectively highlight the role of machine learning, AI-driven chatbots, location-based services, community features, nutritional guidance, and blockchain security in enhancing menstrual health applications. By integrating insights from this body of research, the proposed system aims to provide a comprehensive, user-centric menstrual health solution that includes predictive analytics, smart reminders, location assistance, and interactive features [18].

I.3 PROBLEM STATEMENT

Menstrual health is a crucial yet often overlooked aspect of life for menstruators, particularly in digital health technology. Current period tracking apps primarily use basic calendar logic, which does not cater to those with irregular cycles or health issues. They lack

personalized insights and fail to help users find nearby restrooms or understand unusual symptoms, such as changes in period blood color. Furthermore, these apps do not focus on early identification of potential menstrual health issues or acknowledge cravings and emotional swings. There is also a lack of community support for users to share experiences without fear of judgment. Our project addresses these gaps with a comprehensive menstrual health web application. It employs machine learning for accurate period predictions and includes features like a SmartPad timer tracker for hygiene, a Period Blood Color Analysis tool for health concerns, and a location service to find nearby restrooms and gynecologists. Additionally, it offers a Craving Decoder for healthy snack options and fosters emotional expression through blogs and chat communities. In essence, while current apps only track periods, our system aims to provide the full support that users need personalized, intelligent, and caring.

II. PROPOSED MODEL

The primary objective of this innovative platform is to establish a cohesive and user-friendly space dedicated to menstrual health management and education. By merging cutting-edge features with a straightforward interface, the platform is designed to empower individuals to take charge of their health and well-being throughout their menstrual cycles. The key objectives of the platform are outlined below:

Comprehensive Health Management: The platform is equipped with a range of essential features, such as an accurate period prediction tool that estimates cycle dates, timely reminders for changing sanitary products to prevent leaks and discomfort, and a geo-locator that identifies the nearest washrooms. These functionalities work together to ensure users can effectively manage their menstrual health with confidence and ease.

Personalized Health Insights: By utilizing intricate machine learning algorithms, the system analyzes individual menstrual patterns and behaviors to forecast future cycles. This results in customized insights that allow users to comprehend their bodily rhythms better and plan personal and professional activities around their cycles, ultimately enhancing their lifestyle.

Security and Privacy: Prioritizing user confidentiality, the

platform employs robust security measures, including advanced cryptographic algorithms that encrypt passwords. This ensures that sensitive personal data is safeguarded against unauthorized access, providing users with peace of mind as they navigate their menstrual health.

Educational Resources: To further enhance users' understanding of menstrual health, the platform offers an extensive library of resources, including informative videos, in-depth study materials, and expert opinions. These resources empower users to make well-informed choices regarding their menstrual care and overall health.

Community Engagement: The platform features a dedicated section, "Your Period Story," along with a Community Chat feature that encourages users to connect with one another. This fosters a supportive community where users can share their experiences, offer encouragement, and seek guidance in a safe and welcoming environment.

Healthier Lifestyle Support: The “Craving Decoder” is a unique feature that provides users with nutritious alternatives to common cravings experienced during menstruation. By suggesting healthier snack options, this feature promotes better eating habits and overall wellness during the menstrual cycle.

Access to Medical Professionals: The “Find a Gynecologist feature is designed to assist users in easily locating certified medical professionals in their locality. This ensures that users have access to qualified guidance and support whenever they need it, reinforcing their health management.

User-Friendly Interface: With a commitment to simplicity and accessibility, the platform is intuitively designed to accommodate users of all ages and tech-savviness levels. This focus on user experience enables effortless navigation, allowing individuals to access various features and resources to enhance their menstrual health seamlessly.

Empowerment Through Technology: Utilizing state-of-the-art technologies such as machine learning, interactive chatbots, and a dynamic user interface, the platform delivers a high-quality, personalized experience. This technological integration serves to educate, inform, and empower users, equipping them with the tools they need to take control of their menstrual health effectively. menstrual health.

III. METHODOLOGY

The system is a web-based platform designed to improve user engagement and health awareness through a set of features.

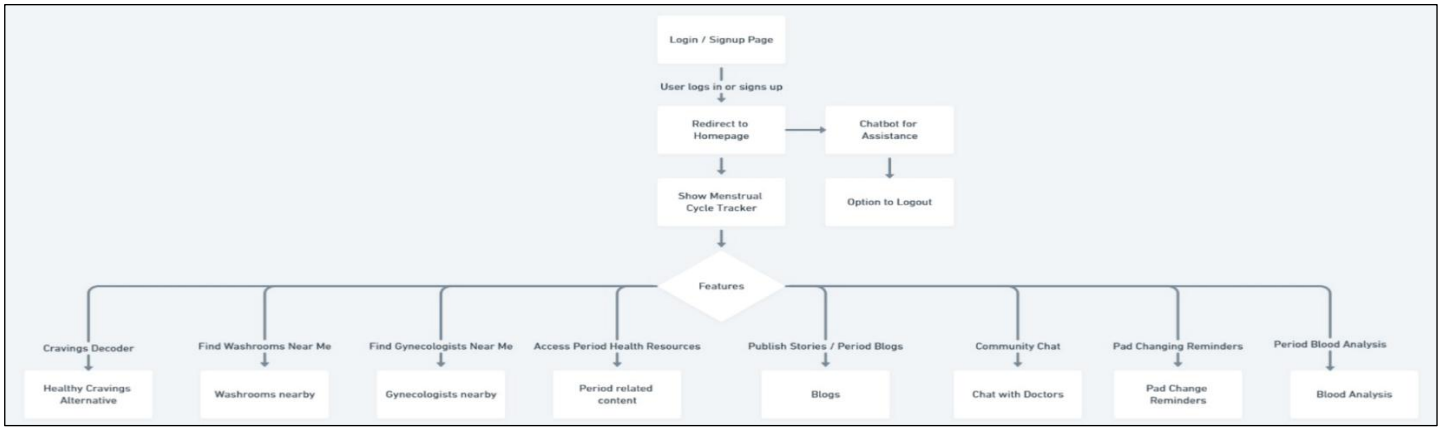


Figure 1: Feature Workflow System.
Source: Authors, (2025).

As shown in Fig. 1, It starts at the Login/Signup Page, where users can create or access accounts. After logging in, users are directed to the Homepage, which includes the Menstrual Cycle Tracker.

Key features include:

- Chatbot for Assistance: Provides interactive support and guidance.
- Secure Logout Option: Ensures privacy and session control.

Upon login, users can access the following modules:

- Cravings Decoder: Suggests healthy alternatives for common menstrual cravings.
- Find Washrooms Near Me: Uses geolocation to locate nearby washrooms.
- Find Gynecologists Near Me: Lists nearby gynecologists for timely consultations.
- Access Period Health Resources: Offers medically reviewed videos, articles, and infographics on menstrual hygiene.
- Publish Stories / Period Blogs: Allows users to share experiences anonymously or publicly.
- Community Chat: Provides a secure environment for discussions with doctors and other users.
- Pad Changing Reminders: Sends reminders for changing menstrual products.
- Period Blood Analysis: Accepts user input on blood flow and characteristics for menstrual health insights.

User data is securely stored in a MySQL database, with passwords protected by cryptographic hash functions. Public content like blog entries and chat messages is visible to other users, while private health records remain encrypted and secure.
This design makes the system user-friendly, informative, and privacy-focused, offering tools for health tracking and community-driven support.

IV. RESULTS AND DISCUSSIONS

The system was designed to provide a comprehensive and user-friendly solution for managing and understanding menstrual health. This section showcases the visual interface of the web application through demonstrations of key features. Each screenshot highlights a core module of the system, illustrating its functionality, user interaction, and contribution to the overall user experience. From secure login to intelligent cycle tracking and health analysis, the system aims to empower users by enhancing accessibility, awareness, and timely support.

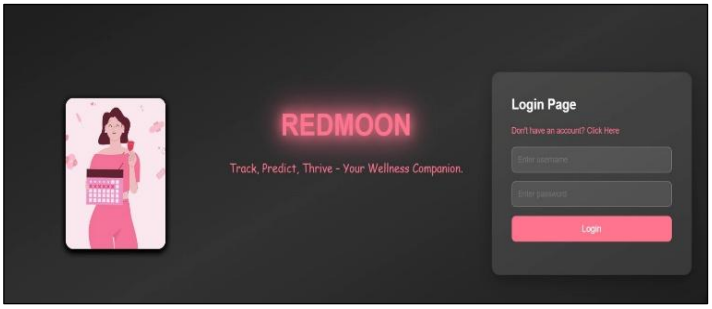


Figure 2: Login & Registration Page.
Source: Authors, (2025).

Figure. 2 shows this login and registration interface provides secure access to the platform. New users can sign up by providing essential information, while returning users can log in for a personalized experience.

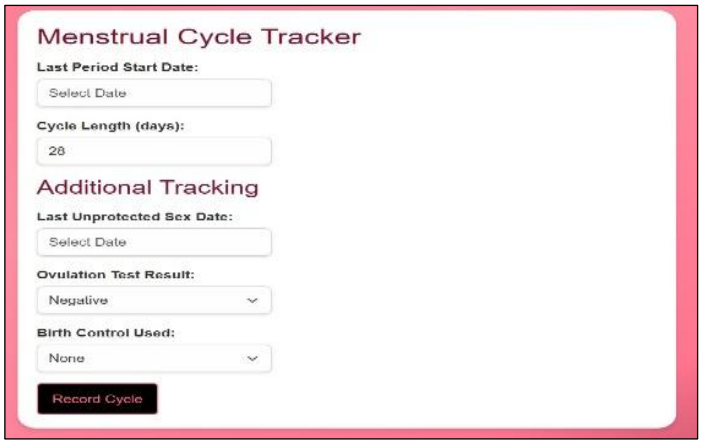


Figure 3: Menstrual Cycle Tracker
Source: Authors, (2025).

Figure. 3 shows. This feature predicts upcoming periods based on inputs such as the start date of the last period, cycle length, and ovulation test results. It offers a smart and adaptive way to monitor menstrual cycles and manage fertility awareness.

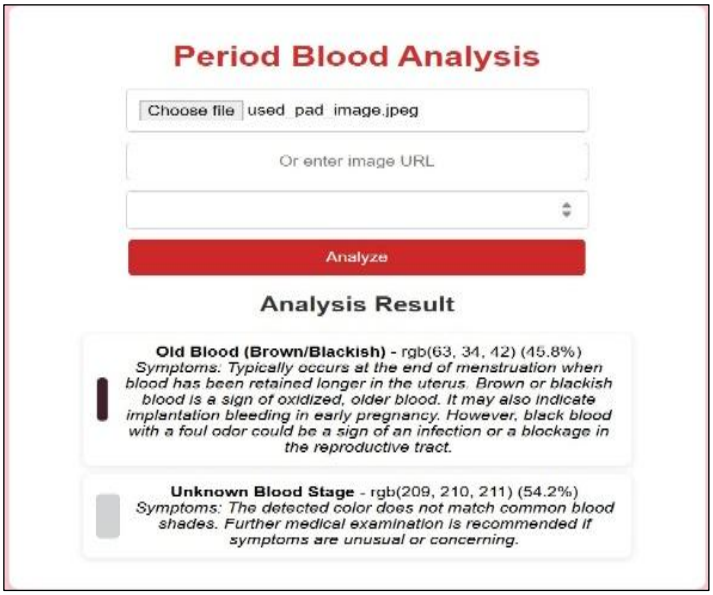


Figure 4: Period Blood Analysis.
Source: Authors, (2025).

Figure. 4 shows this feature allows users to input observations about their menstrual blood, color. It provides insights and health indicators based on medical research and guidelines.

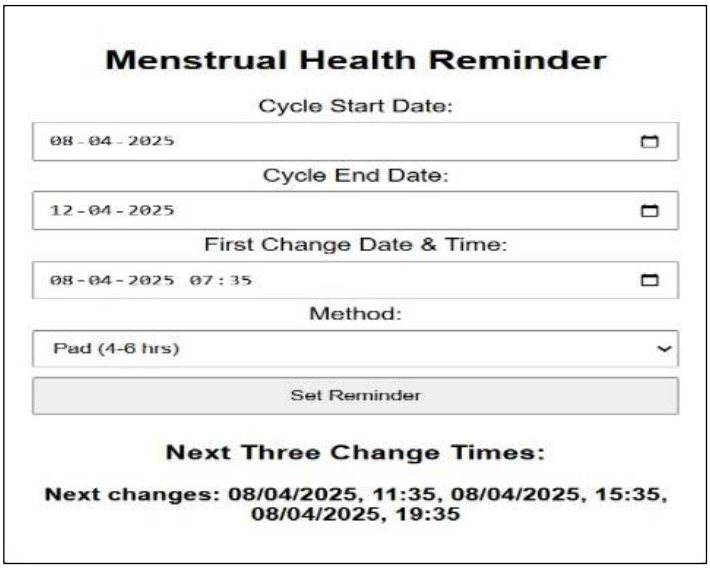


Figure 5: Pad/Cup Change Reminder
Source: Authors, (2025).

Figure. 5 shows an automated reminder system alerts users when it's time to change their pad, tampon, or menstrual cup. This feature ensures better hygiene and helps prevent health risks associated with prolonged use.

Figure 6: Craving Decoder
Source: Authors, (2025).

Figure. 6 shows the Craving Decoder suggests healthier snack alternatives based on user cravings during their cycle. This feature promotes better nutritional choices and helps reduce unnecessary binge eating.

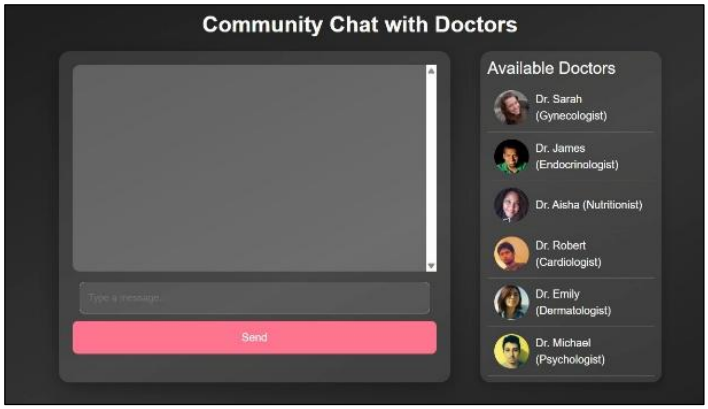


Figure 7:Community Chat with Doctors
Source: Authors, (2025).

Figure. 7 shows an interactive chat space where users can ask questions related to menstrual health and receive guidance from certified medical professionals, or engage in conversations with peers.

The system effectively combines various features to meet user needs, prioritizing accessibility, accuracy, and engagement. Each module underwent thorough testing for functionality and ease of use, ensuring a smooth and intuitive experience for users. The screenshots and outputs provided above showcase the practical application and effectiveness of the proposed solution. These results confirm the system's ability to tackle key challenges and provide reliable support to users.

V. MODEL PERFORMANCE EVALUATION

The evaluation of various machine learning models on the dataset reveals significant insights. Most models achieved high accuracy, with Logistic Regression, Support Vector Machine (SVM), Naïve Bayes, AdaBoost, and MLP Classifier all reaching an accuracy of 84.50%. However, a concerning pattern emerged in the classification reports—these models struggled to correctly classify the minority class (label 1), as evidenced by their 0% recall for this category. This indicates that they are biased toward predicting the majority class (0), leading to misleadingly high accuracy but poor performance in handling class imbalance.

Table 1: Model-Wise Evaluation.

Model	Accuracy(%)	Recall (Class 1)	ROC-AUC	Execution Time
Logistic Regression	84.50	0.00	0.47	0.35
Support Vector Machine	84.50	0.00	0.51	1.04
Naïve Bayes	84.50	0.00	0.46	0.12
AdaBoost	84.50	0.00	0.50	1.73
MLP Classifier	84.50	0.00	0.55	3.30
Gradient Boosting	84.00	0.00	0.49	2.44
K-Nearest Neighbors	84.00	6.00	0.55	0.36
Random Forest	81.00	10.00	0.53	4.65
Decision Tree	77.50	19.00	0.56	0.13
Perceptron	15.50	100.00	0.50	0.11

Source: Authors, (2025).

Random Forest and K-Nearest Neighbors (KNN) performed better in distinguishing between classes. Random Forest achieved an accuracy of 81.00%, and although its recall for class 1 was still low (10%), it was better than that of other models. KNN performed even better, with an accuracy of 84.00% and a slightly higher recall than most classifiers (6% for class 1). The Decision Tree lagged slightly behind with 77.50% accuracy but demonstrated better recall compared to many other models.

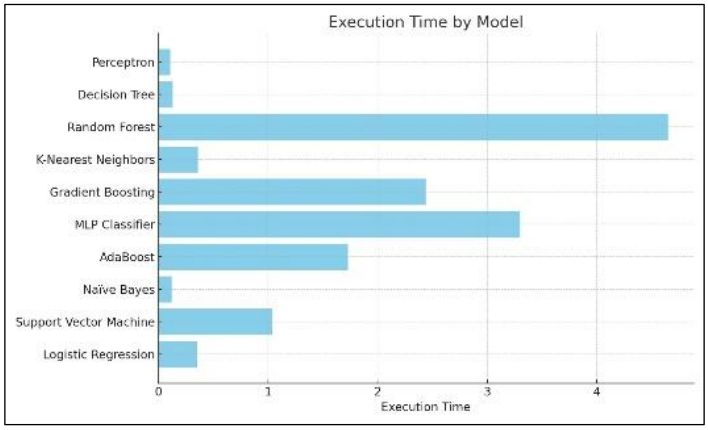


Figure 8: Execution Time Analysis.
Source: Authors, (2025).

This comparison displays the execution time of each algorithm. The Perceptron and Naïve Bayes algorithms were the quickest, making them suitable for lightweight applications, while the MLP Classifier and Random Forest had longer execution times due to their complexity.

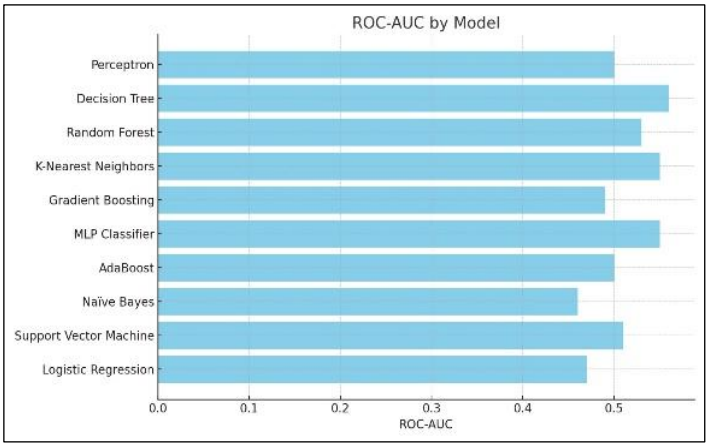


Figure 9: ROC-AUC Score Chart.
Source: Authors, (2025).

The ROC-AUC scores provide insights into the trade-off between sensitivity and specificity. The K-Nearest Neighbors and Random Forest models obtained relatively higher ROC-AUC values, indicating a more balanced classification capability.

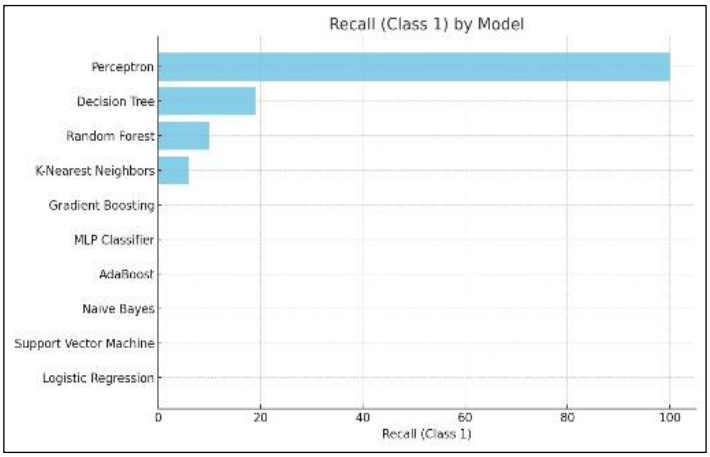


Figure 10: Recall Comparison Across Algorithms.
Source: Authors, (2025).

The recall graph focuses on each model's ability to identify the minority class (Class 1). The Decision Tree exhibited the highest recall rate of 19%, demonstrating its effectiveness in detecting instances of the minority class, although its overall accuracy was lower.

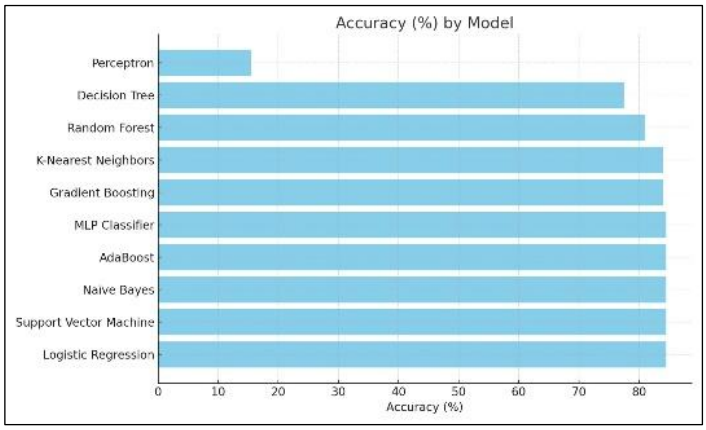


Figure 11: Accuracy Comparison of ML Models.
Source: Authors, (2025).

This graph illustrates the accuracy performance of all the implemented models. Logistic Regression, SVM, and Naïve Bayes achieved the highest accuracy of 84.5%. However, their inability to detect the minority class highlights a limitation in handling imbalanced data.

The ROC-AUC scores, which measure a model's ability to differentiate between classes, were generally low across the models. The highest ROC-AUC score was 0.56 for the Decision Tree, followed by 0.55 for KNN and the MLP Classifier. The low ROC-AUC

scores further confirm that most models struggled to identify the minority class correctly. The Perceptron, in particular, performed poorly, achieving an accuracy of only 15.50%, making it the worst-performing model.

While Logistic Regression, SVM, Naïve Bayes, AdaBoost, and MLP Classifier demonstrated the highest accuracy (84.50%), their inability to correctly classify the minority class renders them unsuitable for imbalanced classification tasks. Random Forest and KNN provided a better balance between accuracy and recall. The Decision Tree had the best recall for class 1 (19%), making it the most effective at identifying the minority class, albeit with lower overall accuracy. If the goal is to maximize overall accuracy, Logistic Regression or KNN would be good choices. However, if the priority is to identify instances of the minority class, the Decision Tree or Random Forest would be preferable.

VI. CONCLUSIONS

This research presents a menstrual health management system that uses machine learning to predict menstrual cycles while prioritizing user well-being through supportive and educational features. The system is designed to be personalized, privacy-focused, and user-friendly, enhancing awareness and management of menstrual health.

We utilized various machine learning algorithms, including Logistic Regression, Decision Tree, Naïve Bayes, Support Vector Machine (SVM), Random Forest, and K-Nearest Neighbors (KNN), to predict menstrual cycles. Each model was evaluated based on accuracy, recall, ROC-AUC, and execution time. The Decision Tree model performed particularly well in recall, making it especially effective for predicting early or irregular periods.

The web application includes essential features such as a smart pad/cup change reminder, a craving decoder, a cycle tracker, a community chat with medical professionals, and educational resources. A unique period story blog allows users to share their experiences, fostering a supportive community while ensuring user privacy through secure design.

Screenshots included in the results section demonstrate the integration of the user interface, backend, and data analysis outputs. The application is modular and scalable, designed for future enhancements, which may include more advanced machine learning models and mobile app support.

Overall, this menstrual health system effectively combines technology with health awareness, aiming to break taboos and provide a reliable, interactive tool for menstrual health tracking and education. This research lays the foundation for future innovations in femtech and broader reproductive health.

VII. AUTHOR'S CONTRIBUTION

Conceptualization: Payal Tandan, Purnima Nahata and Rovina Dbritto.
Methodology: Payal Tandan, Purnima Nahata.
Investigation: Payal Tandan, Purnima Nahata.
Discussion of results: Payal Tandan, Purnima Nahata and Rovina Dbritto.
Writing – Original Draft: Payal Tandan, Purnima Nahata.
Writing – Review and Editing: Payal Tandan, Purnima Nahata.
Resources: Payal Tandan, Purnima Nahata.
Supervision: Rovina Dbritto.
Approval of the final text: Payal Tandan, Purnima Nahata and Rovina Dbritto.

VIII. REFERENCES

[1]. Hennegan, Julie et al., "Menstrual health: a definition for policy, practice, and research," Sexual and Reproductive Health Matters, vol. 29, no. 1, pp. 31-38, 2021. Available: https://doi.org/10.1080/26410397.2021.1911618 [Accessed Apr 11, 2025].

[2]. Yu, Jia-Le et al., "Tracking of menstrual cycles and prediction of the fertile window via measurements of basal body temperature and heart rate as well as machine-learning algorithms," Reproductive Biology and Endocrinology, vol. 20, no. 1, 2022. Available: https://doi.org/10.1186/s12958-022-00993-4 [Accessed Apr 11, 2025].

[3]. Campo Woytuk, Nadia et al., "Touching and Being in Touch with the Menstruating Body," Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1-14, 2020. Available: https://doi.org/10.1145/3313831.3376471 [Accessed Apr 11, 2025].

[4]. Li, Yumei, Zeng, Hong, and Fu, Jing, "Preovulatory progesterone levels are the top indicator for ovulation prediction based on machine learning model evaluation: a retrospective study," Journal of Ovarian Research, vol. 17, no. 1, 2024. Available: https://doi.org/10.1186/s13048-024-01495-0 [Accessed Apr 11, 2025].

[5]. Vijendra Babu, D et al., "RFID and GSM based Automatic Dispensing and Monitoring of Sanitary Napkin Vending Machine for Menstrual Hygiene among Women," IOP Conference Series: Materials Science and Engineering, vol. 993, no. 1, pp. 012074, 2020. Available: https://doi.org/10.1088/1757-899x/993/1/012074 [Accessed Apr 11, 2025].

[6]. Ghazzawi, Hadeel A. et al., "Menstrual cycle symptoms are associated with nutrient intake: Results from network analysis from an online survey," Women's Health, vol. 19, 2023. Available: https://doi.org/10.1177/17455057231185624 [Accessed Apr 11, 2025].

[7]. Rahman, Rifat et al., "AdolescentBot: Understanding Opportunities for Chatbots in Combating Adolescent Sexual and Reproductive Health Problems in Bangladesh," Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1-15, 2021. Available: https://doi.org/10.1145/3411764.3445694 [Accessed Apr 11, 2025].

[8]. Panda, Nishisipa et al., "Menstrual health and hygiene amongst adolescent girls and women of reproductive age: a study of practices and predictors, Odisha, India," BMC Women's Health, vol. 24, no. 1, 2024. Available: https://doi.org/10.1186/s12905-024-02894-7 [Accessed Apr 11, 2025].

[9]. Schantz, Joelle S., Fernandez, Claudia S. P., and Jukic, Anne Marie Z., "Menstrual Cycle Tracking Applications and the Potential for Epidemiological Research: a Comprehensive Review of the Literature," Current Epidemiology Reports, vol. 8, no. 1, pp. 9-19, 2021. Available: https://doi.org/10.1007/s40471-020-00260-3 [Accessed Apr 11, 2025].

[10]. Rego, Rosana C. B., "Predictive Modeling of Menstrual Cycle Length: A Time Series Forecasting Approach,", 2023. Available: https://doi.org/10.48550/ARXIV.2308.07927 [Accessed Apr 11, 2025].

[11]. Kim, Soomin et al., "Moderator Chatbot for Deliberative Discussion," Proceedings of the ACM on Human-Computer Interaction, vol. 5, no. CSCW1, pp. 1-26, 2021. Available: https://doi.org/10.1145/3449161 [Accessed Apr 11, 2025].

[12]. Wang, Jenhung et al., "Who needs automotive on-board navigation systems? Predicting operational performance from spatial anxiety and gender differences," Transportation Planning and Technology, vol. 43, no. 6, pp. 539-552, 2020. Available: https://doi.org/10.1080/03081060.2020.1780706 [Accessed Apr 11, 2025].

[13]. De, Prasenjit, and Pradhan, Manas Ranjan, "Effectiveness of mobile technology and utilization of maternal and neonatal healthcare in low and middle-income countries (LMICs): a systematic review," BMC Women's Health, vol. 23, no. 1, 2023. Available: https://doi.org/10.1186/s12905-023-02825-y [Accessed Apr 11, 2025].

[14]. Paripoorani, Deborah et al., "A systematic review of menopause apps with an emphasis on osteoporosis," BMC Women's Health, vol. 23, no. 1, 2023. Available: https://doi.org/10.1186/s12905-023-02612-9 [Accessed Apr 11, 2025].

[15]. Ramakrishnan, Ritu et al., "Application of AI in Nutrition," arXiv, 2023. Available: https://doi.org/10.48550/ARXIV.2312.11569 [Accessed Apr 11, 2025].

[16]. Sahajananda, H, and Anuradha, RV, "A Study on Menstrual Problems among High School Girls Studying in Rural Field Practice Area of a Tertiary Care Hospita1," The Journal of Medical Sciences, vol. 5, no. 3, pp. 70-72, 2019. Available: https://doi.org/10.5005/jp-journals-10045-00127 [Accessed Apr 11, 2025].

[17]. Veiga, Rodrigo, Endler, Markus, and de Paiva, Valeria, "Extracting Blockchain Concepts from Text," 2023. Available: https://doi.org/10.48550/ARXIV.2305.10408 [Accessed Apr 11, 2025].

[18]. Delanerolle, Gayathri et al., "Artificial intelligence: A rapid case for advancement in the personalization of Gynaecology/Obstetric and Mental Health care," Women's Health, vol. 17, 2021. Available: https://doi.org/10.1177/17455065211018111 [Accessed Apr 11, 2025].

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11n.55, p. 29-47. September/October, 2025.
DOI: https://doi.org/10.5935/jetia.v11i55.1854

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

ASSESSMENT OF SOLAR CHIMNEY POWER PLANT PERFORMANCE UNDER VARIOUS PARAMETERS

Farid Gaci[1], Mohamed Walid Azizi[2], Ahcene Akni[3], Moumtez Bensouici[4], Djoubeir Debbah[5]

[1, 3] Department of Transportation Engineering, Frères Mentouri University -Constantine 1, Algeria.
[2, 4, 5] Department of Mechanical Engineering and Electromechanical, Abdelhafid Boussouf University- Mila, Algeria.

[1] https://orcid.org/0009-0000-9487-4929, [2] http://orcid.org/ 0000-0003-1066-740X
[3] https://orcid.org/0000-0002-3180-4218, [4] http://orcid.org/ 0000-0002-5855-1775
[5] http://orcid.org 0009-0003-8797-2952

Email: farid.gaci@umc.edu.dz, medwalid.azizi@centre-univ-mila.dz, akni.ahcene@umc.edu.dz, moumtez.bensouici@umc.edu.dz, debbah.dj@centre-univ-mila.dz

ARTICLE INFO

Article History
Received: April 12, 2025
Revised: May 20, 2025
Accepted: September 15, 2025
Published: September 30, 2025

Keywords:
Solar chimney,
Turbulence model,
Natural convection,
Power,
Radiation.

ABSTRACT

This numerical study primarily seeks to illustrate that the power output of a solar chimney is determined by the equilibrium between a geometry designed to enhance thermal draft and a turbulence model that faithfully captures the physics of the flow. Several turbulence models were used to evaluate their ability to simulate flows and heat transfer within the solar chimney. A Discrete Ordinates approach was adopted to solve the radiative transfer equation and to refine performance predictions and guide the optimization of geometric and thermal parameters. The most appropriate RNG k-ε turbulence model highlighted notable differences in the distribution of dynamic and thermal fields compared to other turbulence models used. For example, the maximum variation compared to the standard k-ε model reached 0.47 m/s for velocity, 6.55 m²/s² for temperature and 15.07 K for turbulent kinetic energy. Additionally, the study highlighted the critical impact of geometric parameters on solar chimney performance. Increasing the height of the solar chimney by 25% to 50%, for example, could boost power output by 26 % to 52 %. These results highlight key areas for improvement, paving the way for a more efficient design of this type of energy system.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Renewable energies play a key role in the fight against climate change, offering a clean and sustainable alternative to conventional energy sources. Among these technologies, the solar chimney stands out as an innovative solution for electricity production. It works by capturing the sun's thermal energy to generate an upward airflow, which drives a turbine connected to an electrical generator. Experimental scientific studies have been conducted in this context. These are often accompanied by other, less expensive, complementary approaches based on numerical simulation, which allow the modeling of airflows and heat transfers in solar chimney power plants (SCPPs) with the aim of reducing electricity production costs and thus enhancing the efficiency and viability of this technology for large-scale applications, particularly in sunny regions.

Research on solar chimneys has been developed through various studies, ranging from experimentation with small prototypes to more sophisticated approaches, integrating numerical simulations to improve the design and performance of solar chimney power plants.

Schlaich's investigations [1] focused on the first solar chimney prototype built in Manzanares in 1982. Since then, much work has been done by scientists and researchers on the Solar Chimney Power Plant (SCPP). In their experimental study, N. Pasumarthi and S. A. Sherif [2] examine a solar chimney, making two improvements to the collector aimed at enhancing the temperature and air flow to maximize the power output. Das P and Chandramohan VP [3], [4] have developed a compact Solar Chimney Power Plant in India, which includes a chimney measuring 6 meters in height, with a radius of 0.3 m and an inclination of 30 degrees, as well as a radius of

1.75 m. Additionally, a parametric study was performed where the diverging angles were adjusted between 1° and 5 °. The findings revealed that the chimney achieved its best performance at a diverging angle of 2 °.

Ahmed et al. [5] conducted a review of hybrid solar chimney systems integrated with photovoltaic panels, solar ponds, or geothermal energy, assessing their efficiency gains and suggesting avenues for enhancement. A double-pass canopy was designed by Nasraoui et al. [6], resulting in a 28% improvement in collector efficiency. Murena et al [7] evaluated the performance of a pilot solar chimney, tested over a period of seven months in southern Italy, providing new insights, particularly on the impact of ambient humidity. Kumar,et al. [8] conducted a review that analyzes the heat demand across various sectors, addresses the challenges of integrating solar energy (such as sunlight availability, costs, and compatibility), and suggests strategies to enhance this energy transition. Ayadi et al. [9] studied by simulation adopting the k-ε turbulence model four turbine diameters to analyze their impact on the local airflow characteristics in the SCPP, using ANSYS Fluent 17. In addition, they highlighted the significant impact of the collector roof height on the performance of the SCPP. In their study, [10] quantitatively evaluated the effect of various types of collectors on the performance of chimney solar power plants.

The integration of a second roof in the collector section helps to increase the temperature as well as the airflow velocity, which improves the efficiency of the system. Fallah and Valipour [11] conducted a study that demonstrated that artificial roughness applied to the bottom surface of the collector increases heat transfer to the air while decreasing the airflow velocity. The system exhibited superior performance when the roughness was concentrated around the collector opening. Kebabsa et al. [12] investigated the impact of collector entry shape on the performance of solar chimney power plants. Their research revealed that a solar chimney featuring a collector distance of 0.8 meters and a slope of 9.1 degrees produces 16 % more electricity compared to a flat collector design. Guo et al. [13] used a three-dimensional numerical approach incorporating radiation models, demonstrating that this model is essential to avoid overestimating the absorbed energy. Furthermore, the ambient temperature mainly influences the air velocity without significantly affecting the energy production of the SCPP. Mebarki et al. [14] employed computational fluid dynamics to investigate how the geometric scale factor influences the implementation of the SUT in urban environments. Hassan et al. [15] conducted a parametric three-dimensional computational fluid dynamics (CFD) analysis to investigate how the angles of the collector slope and chimney divergence affect the performance of the Manzanares prototype. The numerical simulations, which were validated against experimental data, employed the RNG k-ε turbulence model.

Maia et al. [16] performed a numerical analysis of the turbulent flow inside a solar chimney by adopting the standard k-ε model. An elaborate model was implemented to examine the impact of the main geometric parameters on airflow behavior. Mohamad et al. [17] conducted a study using 2D numerical analysis to assess the performance of a solar tower power plant by varying geometric and environmental parameters. The findings indicate that a negatively inclined collector and an increased divergent angle enhance both air velocity and power output. Awan et al. [18] conducted a study that aims to design, optimize and evaluate the performance of a solar chimney power plant, with optimization of design parameters such as tower height and thermal storage. Torabi et al. [19] examined how geometric parameters and boundary conditions influence the performance of the SUT, finding that both the angle of the chimney divergence and the rate of radiation significantly affect power generation.

Furthermore, it is noted that many studies in computational fluid dynamics (CFD) have favored a single turbulence model without exploring other alternatives. Thus, the research of Siyang M. Jirena et al. [20] and Younes Noorollahi et al. [21] relies exclusively on the standard k-ε model. Conversely, Ehsan Gholamalizadeh et al. [22] performed a three-dimensional simulation using the RNG k-ε turbulence closure, while S. Montelpare et al. [23] opted for the Realizable variant of the k-ε model. The literature reveals that the majority of researchers examine various geometric configurations using a single turbulence model without performing comparative tests with other models. The synergistic combination of solar chimney power plants (SCPPs) with other technological applications is emerging as a multifunctional strategy with high potential. Integrating the system with brackish water desalination processes (Zuo et al. [24]) would allow the use of residual heat to produce drinking water. At the same time, its coupling with agricultural drying [25] would offer an energy-efficient solution for food preservation, while its use in urban air purification [26] could capture polluting particles via the upward airflows generated by the chimney. From a climate perspective [27], the ability of SCPPs to sequester CO_2 and reduce the carbon footprint of buildings, thanks to low-energy hybrid heating and ventilation systems strengthen their role in the fight against global warming [28]. This versatility, combining clean energy production and ecosystems services, positions SCPPs as key technological platforms for a holistic energy transition.

The objective of this numerical study is to analyze the performance of a solar chimney (SCPP) by coupling accurate modeling of turbulent flows with a systematic optimization of its geometric parameters. Through numerical simulations, the emphasis is placed on the choice of an appropriate and non-random turbulence model in order to faithfully capture the thermal and dynamic phenomena within the system. This rigorous approach makes it possible to identify and quantify the impact of geometric parameters (chimney height and radius, inclination angle, etc.) on overall energy efficiency while determining the optimal configurations to maximize thermal draft and energy production..

II. NUMERICAL MEHOD

II.1 MODEL CONFIGURATION

The structure studied is a solar chimney similar to that of the Manzanares power plant [1], with a spatial organization that reflects its fundamental characteristics. Although its main dimensions correspond to those of the original model — height and radius of the chimney, and height and radius of the collector, as illustrated in Figure 1 — a notable modification was made: the collector, initially horizontal, was inclined at an angle of 1.01°. This inclination creates a distinct structural dynamic, reinforced by the conical junction connecting the ceiling to the central tower. This junction, in the form of a nozzle, aims to maximize the thermal draft by promoting a progressive acceleration of the hot air flow while minimizing pressure losses.

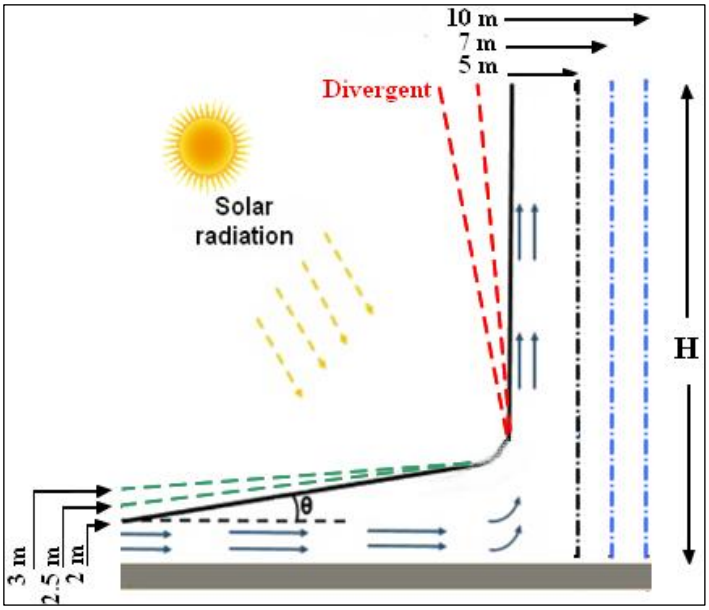


Figure 1: Solar chimney with sloped collector.
Source: Authors, (2025).

The structure preserves dimensional consistency with the reference model, while incorporating functional modifications. The chimney heights (H) range from 200 to 300 m, with diverging outlet radii of 5 to 10 m corresponding to respective inclinations of 0 to 1.43 degrees, as shown in red. The chimney radii (R_{ch}), shown in blue, vary from 5 to 10 m. The collector inlet height (H_{col}), marked in green, is between 2 and 3 m while maintaining a constant radius for the collector (R).

II.2. NUMERICAL MODELING

The airflow inside the solar chimney is characterized using the Cartesian coordinate system, considering the Boussinesq principle. According to this principle in heat transfer, it is assumed that the density of a fluid varies only with temperature, and not with pressure or other variables. This implies that the density variation of the fluid due to temperature difference is negligible compared to its mean values to simplify the fundamental equations of natural convection that occurs when temperature gradients are present. This simplification greatly simplifies the mathematical equations used to describe fluid flow, making the analysis more manageable. The governing equations governing the steady turbulent flow of an incompressible Newtonian fluid within the considered solar setup are as stated below:

a - Continuity equation

The continuity equation translating the principle of conservation of mass is:

$$\frac{\partial u_i}{\partial x_i} = 0 \tag{1}$$

b - Momentum equations

The momentum conservation equations are written as follows for i = 1, 2, and 3:

$$u_j \frac{\partial u_i}{\partial x_j} = -\frac{1}{\rho}\frac{\partial p}{\partial x_i} + \frac{1}{\rho}\frac{\partial}{\partial x_j}\left(\mu \frac{\partial u_i}{\partial x_j}\right) \tag{2}$$

c - Energy equation

The energy equation neglecting viscous dissipation is:

$$\frac{\partial (u_j T)}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\alpha \frac{\partial T}{\partial x_j}\right) \tag{3}$$

The Navier-Stokes equations being non-linear, the averaging operation reveals new terms which reflect the influence of turbulence on the average flow. The system to be solved then has more unknowns than equations; hence the need to use closure models to close it. Generally, closure methods are classified based on the number of additional equations to solve. All Reynolds-Averaged Navier Stokes (RANS) models aim to determine the components of the Reynolds tensor, or Reynolds tensions to determine the average pressure and velocity. A commonly employed approach involves the application of the Boussinesq hypothesis to establish a connection between the Reynolds stresses and the mean velocity gradients. The RANS models used in our study are the first order Standard k−ε , RNG k-ε, Realizable k-ε models which use algebraic relationships to directly model the Reynolds tensor and the second order RSM (Reynolds Stress Model) which solves transport equations on Reynolds stresses.

II.2.1 STANDARD k-ε MODEL

The standard k-ε model uses the eddy viscosity hypothesis to model the Reynolds stress tensor. It consists of two equations that provide a comprehensive overview of turbulence through the use of two supplementary transport equations,, one for the turbulence kinetic energy (k) and one for the turbulence dissipation rate (ε). The two equations are obtained from the following transport equations:

$$\frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_i}(\rho k u_i) = \frac{\partial}{\partial x_i}\left[\left(\mu + \frac{\mu_k}{\sigma_k}\right)\right] + G_k + G_b - \rho\varepsilon - Y_M + S_k \quad (4)$$

$$\frac{\partial}{\partial t}(\rho\varepsilon) + \frac{\partial}{\partial x_i}(\rho\varepsilon u_i) = \frac{\partial}{\partial x_i}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right) + \frac{\partial\varepsilon}{\partial x_j}\right] + C_{1\varepsilon}\frac{\varepsilon}{k}(G_k + C_{3\varepsilon}G_b) - C_{2\varepsilon}\rho\frac{\varepsilon^2}{k} + S_\varepsilon \quad (5)$$

G_k is a measure of the production of k, illustrating the transfer of energy from mean motion to turbulence, under the influence of Reynolds stresses. G_b indicates the contribution of buoyancy to the creation or destruction of k, related to gravity and thermal gradients. S_k and S_ε are source terms. Y_M refers to the dissipation of k due to expansion, which can be considered negligible in an incompressible context. The turbulent Prandtl numbers σ_k and σ_ε govern the diffusion of k and ε. The coefficients $C_{1\varepsilon}$ and $C_{2\varepsilon}$ control the production and destruction of ε, while $C_{3\varepsilon}$ adjusts the effect of buoyancy on ε. The turbulent viscosity μ_t is calculated as follows:

$$\mu_t = \rho\, C_\mu \frac{k^2}{\varepsilon} \quad (6)$$

$$G_k = \mu_t S^2 \quad (7)$$

$$S = \sqrt{2S_{ij}S_{ij}} \quad (8)$$

$$S_{ij} = \frac{1}{2}\left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j}\right) \quad (9)$$

The constants of this model are:

$\sigma_k = 1;\ \sigma_\varepsilon = 1.3;\ C_{1\varepsilon}$=1.44; $C_{2\varepsilon}$=1.92; C_μ=0.09

II.2.2 RNG k-ε MODEL

The RNG k-ε turbulence model is an improved version of the standard k-ε model, analytically developed from the instantaneous Navier-Stokes equations through the application of the renormalization group (RNG) technique. This mathematical approach allows for the effects of unresolved turbulent scales to be taken into account via a systematic averaging process, unlike the standard model whose constants are empirically calibrated.

$$\frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_i}(\rho k u_i) = \frac{\partial}{\partial x_j}\left(\alpha_k \mu_{eff}\frac{\partial k}{\partial x_j}\right) + G_k + G_b - \rho\varepsilon - Y_M + S_k \quad (10)$$

$$\frac{\partial}{\partial t}(\rho\varepsilon) + \frac{\partial}{\partial x_i}(\rho\varepsilon u_i) = \frac{\partial}{\partial x_j}\left(\alpha_\varepsilon \mu_{eff}\frac{\partial\varepsilon}{\partial x_j}\right) + C_{1\varepsilon}\frac{\varepsilon}{k}(G_k + C_{3\varepsilon}G_b) - C_{2\varepsilon}\rho\frac{\varepsilon^2}{k} - R_\varepsilon + S_\varepsilon \quad (11)$$

The constants identified in the model are not the same as those in the standard k-ε model:

$\alpha_k = \alpha_\varepsilon \simeq 1.393$; $C_{1\varepsilon}$=1.42 ; $C_{2\varepsilon}$=1.68.

II.2.3 REALIZABLE k-ε MODEL

The realizable k-ε turbulence model is an evolution of the standard k-ε model, designed to improve accuracy in complex flows by ensuring physically realizable turbulent stresses, i.e., that the normal turbulence stresses remain positive. The transport equations are:

$$\frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_j}(\rho k u_j) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_j}\right] + G_k + G_b - \rho\varepsilon - Y_M + S_k \quad (12)$$

$$\frac{\partial}{\partial t}(\rho\varepsilon) + \frac{\partial}{\partial x_j}(\rho\varepsilon u_j) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial\varepsilon}{\partial x_j}\right] + \rho C_1 S\varepsilon - \rho C_2\frac{\varepsilon^2}{k+\sqrt{\upsilon\varepsilon}} + C_{1\varepsilon}\frac{\varepsilon}{k}C_{3\varepsilon}G_b + S_\varepsilon \quad (13)$$

where:

$$C_1 = \max\left[0.43, \frac{\eta}{\eta+5}\right] \quad (14)$$

$$\eta = S\frac{k}{\varepsilon} \quad (15)$$

In order to ensure optimal performance of this turbulence model for certain canonical flows (Channel Flow, free jet, etc.), the constants C_2, σ_k and σ_ε have been determined. The values of these constants are as follows:

$C_{1\varepsilon} = 1.44;\ C_2 = 1.9\ ;\ \sigma_\varepsilon = 1.2\ ;\ \sigma_k = 1$

Buoyancy influences ε, which is defined by the constant $C_{3\varepsilon}$. In Fluent, this constant is not provided, but is determined by the following formula:

$$C_{3\varepsilon} = tanh\left|\frac{v}{u}\right| \quad (16)$$

The flow velocity is decomposed into two components: v, parallel to the gravitational vector, and u, perpendicular to it. Therefore, $C_{3\varepsilon}$ takes the value 1 for buoyant shear layers whose main flow follows the direction of gravity. Conversely, $C_{3\varepsilon}$ becomes zero for buoyant shear layers perpendicular to the gravitational vector.

II.2.4 RSM TURBULENCE MODEL

The RSM model takes a different approach within RANS approach. Instead of relying on the eddy viscosity assumption, RSM solves transport equations for each individual component of the Reynolds Stress Tensor directly. The modeled transport equations for k and ε in the RSM model are:

$$\frac{\partial}{\partial t}(\rho k) + \frac{\partial}{\partial x_i}(\rho k u_i) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_k}\right)\frac{\partial k}{\partial x_j}\right] + \frac{1}{2}(G_{ii} + P_{ii}) - \rho\varepsilon(1 + 2M_t^2) + S_k \quad (17)$$

$$\frac{\partial}{\partial t}(\rho \varepsilon) + \frac{\partial}{\partial x_i}(\rho \varepsilon \mu_i) = \frac{\partial}{\partial x_j}\left[\left(\mu + \frac{\mu_t}{\sigma_\varepsilon}\right)\frac{\partial \varepsilon}{\partial x_j}\right] C_{\varepsilon 1}\frac{1}{2}[Pii + C_{\varepsilon 3}G_{ii}]\frac{\varepsilon}{k} - C_{\varepsilon 2}\rho\frac{\varepsilon^2}{k} + S_\varepsilon \quad (18)$$

Where, $\sigma_k = 0.82$, $\sigma_\varepsilon = 1.0$, $C_{\varepsilon 1} = 1.44$, $C_{\varepsilon 2} = 1.92$, $C_{\varepsilon 3}$ is evaluated as a function of the local flow direction relative to the gravitational vector, and S_k and S_ε are a user-defined source terms. M_t is the turbulent Mach number. G_{ii} is the production terms due to buoyancy. P_{ii} is the stress production.

III. POWER GENERATION

The performance of a chimney solar power plant relies on optimal synergy between its key components, forming an interdependent system where each element plays a critical role. The solar collector, by converting radiation into heat, generates an upward flow of hot air. This thermal energy is then amplified by the chimney effect, which accelerates natural convection thanks to the difference in density between the hot air and the atmosphere. The turbine, strategically positioned at the base of the chimney, finally transforms this kinetic energy into electricity. A failure or undersizing of one of these elements leads to a drop in energy production. Thus, system optimization requires precise modeling of the physical parameters of each component: collector surface area, model or height of the chimney, and electrical power generated by the turbine. The performance of a chimney solar power plant depends in a complex way on the efficiency of each component, which can be expressed by the mathematical formulas specific to each of them.

III.1 THE SOLAR COLLECTOR

The efficiency of the collector is of utmost importance in chimney solar power plants as it directly impacts the overall performance and electricity generation capacity of these systems. The efficiency of collector is given by:

$$\eta_{coll} = \frac{\dot{Q}}{A_{coll}.G} \quad (19)$$

The generated heat $\dot{Q}$ and the mass flow rate $\dot{m}$ are provided as follows:

$$\dot{Q} = \dot{m}.C_p.\Delta T \quad (20)$$

$$\dot{m} = \rho_{coll}.V_c.A_c \quad (21)$$

Where, ΔT is the temperature difference between the heated air under the collector and the ambient air (K).
By substituting in Equation 19, $\dot{Q}$ and $\dot{m}$ by their expressions, we obtain the relation:

$$\eta_{coll} = \frac{\rho_{coll}.V_c.A_c.C_p.\Delta T}{A_{coll}.G} \quad (22)$$

III.2 THE SOLAR CHIMNEY

The efficiency of the chimney is expressed according to Ref [1] by:

$$\eta_{sc} = \frac{g.H_{sc}}{C_p.T_0} \quad (23)$$

The calculation of the pressure difference generated by the chimney is performed as follows:

$$\Delta p_{tot} = \rho_{coll}.g.H_{sc}\frac{\Delta T}{T_0} \quad (24)$$

III.3 THE TURBINE

The Solar chimney system turbine belongs to a pressure-based wind turbine. The front and rear air velocities are almost the same, but the pressure changes significantly. The output power is given by:

$$P = x\sqrt{1-x}.\eta_{coll}.\eta_{wt}.\left(\frac{g}{C_p\,T_0}\right).H_{sc}.A_{coll}.G \quad (25)$$

The maximum output power of the system is achieved when the turbine pressure drop factor (x) is set to 2/3. Meanwhile, in this study, the turbine efficiency (η_{wt}) was established to 80 %.

IV. BOUNDARY CONDITIONS

The geometric model of the chimney with sloped roof considered in our study presents a notable similarity to the scenario of the SCPP of Manzanares. The boundary conditions for the simulation of the solar collector are specified according to the highest solar radiation in the Constantine region which is approximately 960 W [29]. The collector inlet and chimney outlet are characterized by their

respective inlet and outlet pressure limits, which are set at 0 Pa, while axisymmetric conditions were implemented at the center of the device. The chimney wall is considered an adiabatic boundary. The environmental temperature is set at 298 K. The bottom of the energy storage layer is set as a constant temperature boundary, with the temperature set at 300 K. Finally, the 1.01° inclined collector roof takes into account the convection limits with a coefficient set at 9.5 W/(m^{2}K), suitable for low ambient air velocities. At a temperature of 25 °C, Table 1 shows the thermophysical properties of air. The boundary conditions for different locations are shown in detail in Table 1.

Table 1: Thermophysical properties of air at 25 °C.

Fluid	ρ (kg/m^3)	k (W/m.K)	μ (kg/m.s)	C_p (J/kg.K)
Air	1.184	0.0255	1.849 10^{-5}	1007

Source: [20].

The governing equations for flow and energy (Equations (1)–(10)) are solved numerically under the specified boundary conditions shown in Table 2, using the commercial code Ansys Fluent, which employs a finite volume method.

Table 2. Boundary conditions.

Surface	Type	Value
Collector roof	Wall	Mixed; T_{ext} = 300 K, h_s = 9.5 W/m^{2}K,
wall surface	Wall	Adiabatic; $qch = 0\ W/m^2$
Ground	Wall	G = 960 W/m^2
Inlet of collector	Pressure-inlet	T_{ext} = 302 K, p_{gauge} = 0 Pa
Outlet of chimney	Pressure-outlet	T_{ext} = 300 K, p_{gauge} = 0 Pa

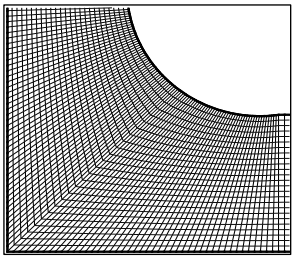
Source: Authors, (2025).

The geometry is designed and meshed with Gambit software 2.4.6. To jointly solve the continuity, momentum, and pressure equations, the SIMPLE (Semi-Implicit Method for Pressure-Linked Equations) algorithm, based on pressure-velocity coupling, is used. The convective terms of the momentum and energy transport equations are discretized using a power law scheme. The discrete ordinate (DO) model was chosen to solve the radiative transfer equation (RTE) in the modeling of radiation heat transfer phenomena. To maintain numerical stability, under-relaxation factors are applied during the iteration of the discretized equations. The convergence criterion is set at 10^{-5} for the mass and momentum equations, and at 10^{-6} for the energy equation.

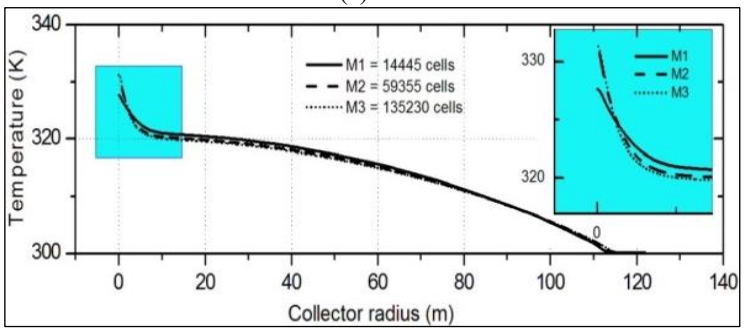
V. RESULTS AND DISCUSSION

V.1 ANALYSIS OF MESH SENSITIVITY

The precision of the calculation and the overall accuracy of numerical simulation results are influenced by the type and size of the grid. Conducting grid-independent analysis allows for the determination of a grid density that is appropriate for the specific problem, thereby ensuring the accuracy of the results. To reduce the mesh requirements, wall laws are implemented to connect turbulent variables, such as k and ε, with wall conditions. All the turbulence models mentioned earlier are intended for use with (y+) values greater than 30, indicating that the first mesh cell must be positioned within the logarithmic region of the boundary layer, in our case (y+ > 177). A variety of uniform quadrilateral structured mesh generated (Figure 2(a)), were tested. Three mesh grids were evaluated as shown in Figure 2(b) to determine the optimal trade-off between independence results, and calculation time.



(a)



(b)

Figure 2: (a) Grid mesh and (b) mesh sensitivity analysis.
Source: Authors, (2025).

The tests carried out reveal a significant difference of the temperature curves obtained along the median radius of the collector between the M1 and M2 mesh configurations, while the M2 and M3 configurations are in perfect agreement with a minimal deviation. This demonstrates that the solution obtained is not significantly affected by the mesh resolution after the implementation of the M2 configuration, i.e. 59355 cells, indicating that it is truly representative of the physical system being simulated. Therefore, the M2 configuration was used for all simulations performed in this study.

V.2 CODE VALIDATION

Table 3 presents a comparison of our numerical results with those of Schlaich [1] for the same geometric parameters of his Manzanares solar chimney. The parameter analyzed here is the power developed by the chimney, evaluated through several distinct turbulence models. Among these models, the RNG k-ε turbulence model stands out for a better agreement with the reference data, displaying a minimum relative error of 1.32, which underlines its accuracy in predicting the dynamic behaviors of the air flow in the solar chimney. Conversely, the Standard k-ε turbulence model records the most significant deviation, with a significantly higher relative error value, indicating a poorer agreement with the expected results. These deviations highlight the critical impact of the choice of the model on the reliability of the simulations, especially for applications requiring high accuracy in terms of the electrical power of the solar chimney station.

Table 3: Comparisons between our numerical procedure and the experimental results of.

Turbulence model	H (m)	G (W/m^2)	P (kW)	P (kW) Error Relat
Standard k-ε	200	1000	52,87	5,73156
Realizable k-ε	200	1000	51,13	2,26635
RNG k-ε	200	1000	49,34	1,31928
RSM	200	1000	49,05	1,9
Experimental (Schlaich)	200	1000	50	

Source: [1]

V.3 TURBULENCE MODEL EFFECT

Figure 3 illustrates the velocity contours obtained under the specified boundary conditions (G = 960 W/m^2, H = 200 m, H_{col}= 2 m, R_{ch} = 5 m), revealing a broadly similar spatial distribution among the different turbulence models. The maximum velocity is systematically located near the curved wall of the solar chimney, highlighting a concentration of dynamic effects in this area.

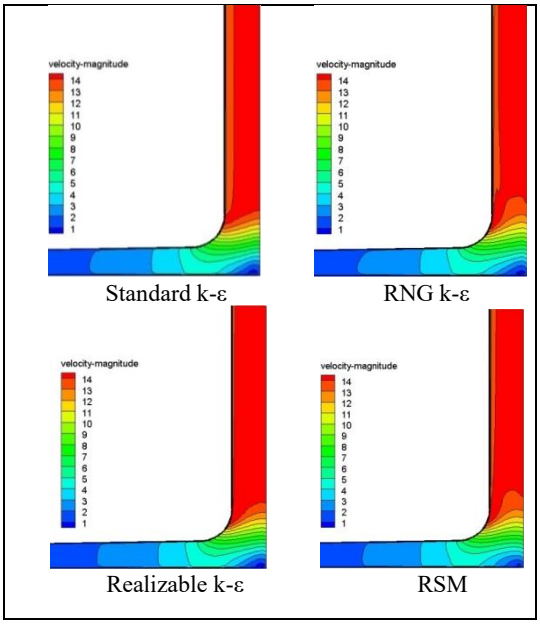
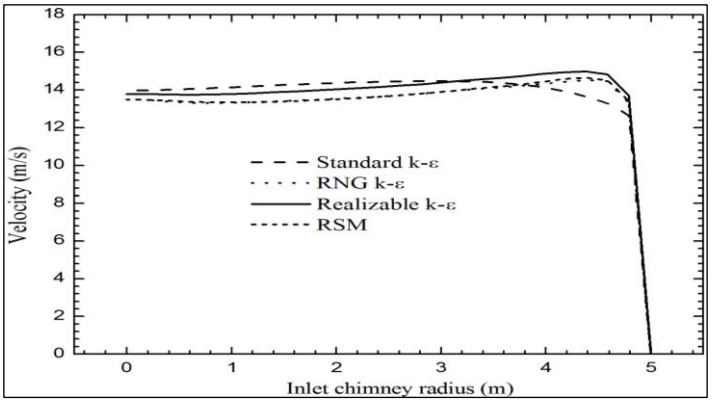
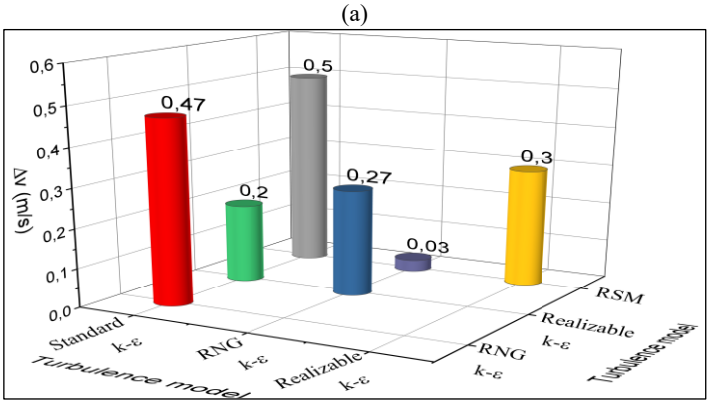


Figure 3: Velocity contours (H_{col}= 2 m, R_{ch} = 5m).
Source: Authors, (2025).

However, when examining the velocity profiles at the tower inlet, Figure 4(a) reveals a notable disparity in the velocity distribution near the wall. Moreover, the velocity separation occurs earlier with the Standard k-ε turbulence model than with the other models. The maximum recorded velocity values differ significantly: 13.98 m/s for the Standard k-ε model, 13.78 m/s for the Realizable k-ε model, 13.51 m/s for the RNG k-ε model and 13.48 m/s for the RSM model. These deviations are summarized in Figure 4(b), which compares the velocity differences between the models, showing that the Standard k-ε model presents the highest difference compared to the adequate RNG k-ε model. This highlights the impact of the modeling choices on the velocity distribution in the structure. However, the RNG k-ε model remains the one closest to the RSM model, which recorded a difference of only 0.03 m/s. It is noted that the flow separation near the curved wall upstream of the chimney tower, observed in particular with the Standard k-ε model, constitutes a critical phenomenon directly influencing the overall performance and stability of the flow. An early separation, such as that predicted by this model, can generate turbulent recirculation zones, amplifying pressure losses and degrading the energy efficiency of the system.



(a)



(b)

Figure 4: (a) Velocity profiles and (b) the maximum velocity difference between turbulence models.
Source: Authors, (2025).

To assess the ability of turbulence models to capture and reproduce turbulence in the flow inside the solar chimney, Figure 5 shows the distribution of turbulent kinetic energy, which represents the energy associated with velocity fluctuations in the turbulent flow. As hot air rises in the chimney, it is subjected to a significant velocity gradient between the main (fast) flow and the slower-moving air in contact with the walls. This intense shear generates instabilities and promotes turbulence production, thus increasing the turbulent kinetic energy near the walls. The distribution of the high turbulent kinetic energy region differs among the turbulence models. The results obtained from the three models, namely Standard k-ε, RNG k-ε, and Realizable k-ε, present a region of intense turbulence near the chimney wall downstream of the bend. However, the RSM model indicates that this region is located in the middle of the chimney tower inlet.

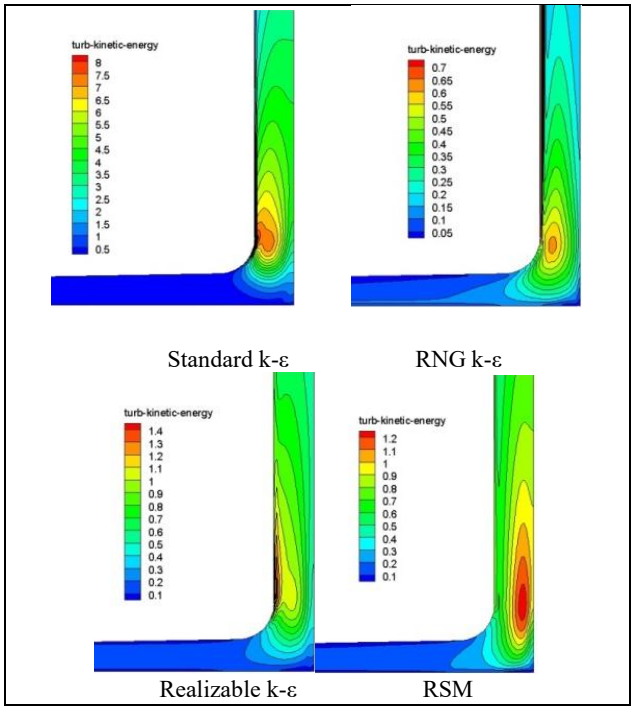


Figure 5: Contours of turbulent kinetic energy (H_{col}= 2 m, R_{ch} = 5m).
Source: Authors, (2025).

Figure 6(a) better illustrates this difference interpreted by the turbulent kinetic energy profiles along the radius just at the chimney tower inlet.

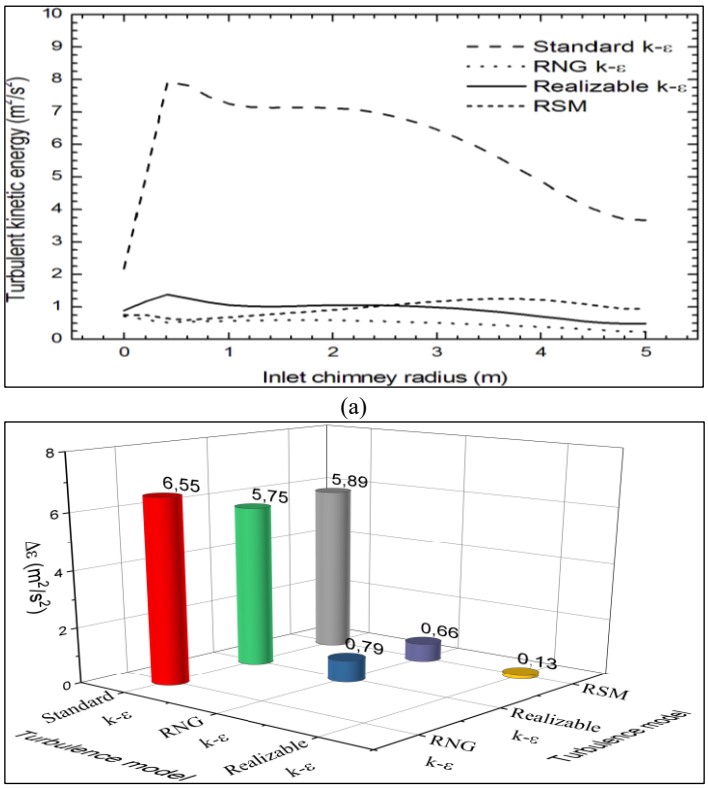


Figure 6: (a) Turbulent kinetic energy profiles and (b) the maximum turbulent kinetic energy difference between turbulence models.
Source: Authors, (2025).

Figure 6(a) better illustrates this difference interpreted by the turbulent kinetic energy profiles along the radius just at the chimney tower inlet. The standard k-ε model overestimated the turbulent kinetic energy throughout the section near the stagnation point by a maximum value of 7.14 m^2/s^2. The RNG k-ε model identified a zone of lowest turbulence with a maximum value of 0.59 m^2/s^2.

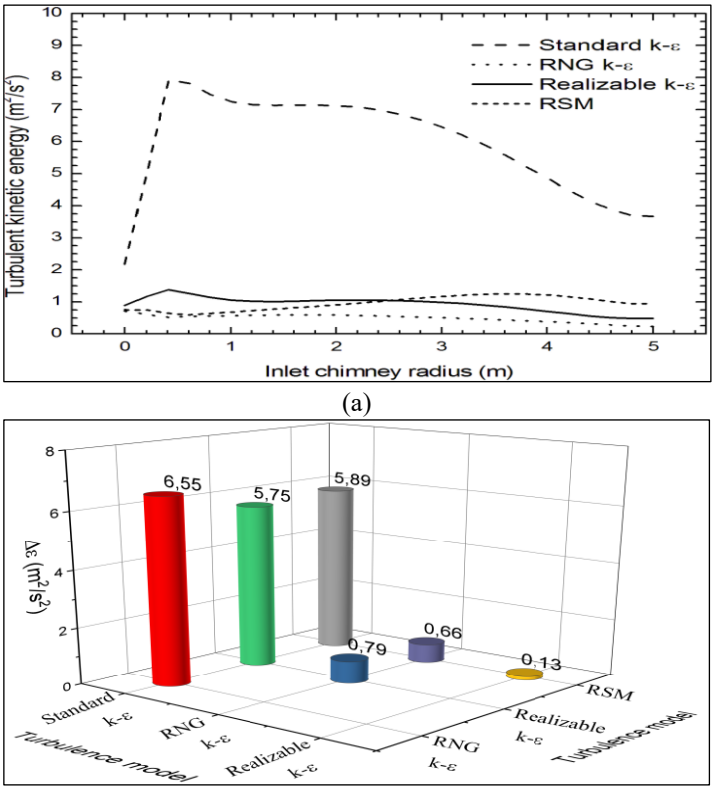


Figure 6: (a) Turbulent kinetic energy profiles and (b) the maximum turbulent kinetic energy difference between turbulence models
Source: Authors, (2025).

The RSM model, however, predicted an intensification of the vortex flow in the middle of the chimney with a maximum value of 1.25 m^2/s^2, which was lower than the value recorded by the Realizable k-ε model near the chimney wall, which was 1.38 m^2/s^2. The difference in the maximum value of turbulent kinetic energy among the four models can be observed in Figure 6(b). Moreover, the three two-equation turbulence models, Standard k-ε, RNG k-ε and Realizable k-ε predicted a decreasing intensification of the wall kinetic energy in the middle of the chimney in contrast to the increasing evolution marked by the seven-equation RSM turbulence model.

Figure 7 illustrates the temperature distribution along the solar chimney for different turbulence models. The results indicate that the temperature field is mainly concentrated in the middle of the solar chimney, extending from the collector outlet to the top of the chimney tower. Warmer and less dense air rises along the central axis of the tower, while colder and denser air descends along the walls. This movement generates an upward flow of hot air in the middle of the chimney, which transports the heat to the highest part of the tower. In the confluence zone located at the base of the solar chimney, the highest temperature values were observed. This phenomenon can be attributed to the confinement effect, where the air flow converges and the space available for lateral dispersion decreases. Therefore, this can lead to the accumulation of hot air and an increase in pressure in this region. All four turbulence models successfully captured the natural convection phenomenon, although there were discrepancies in terms of spatial distribution and temperature values. The choice of turbulence model also had an impact on the temperature distribution.

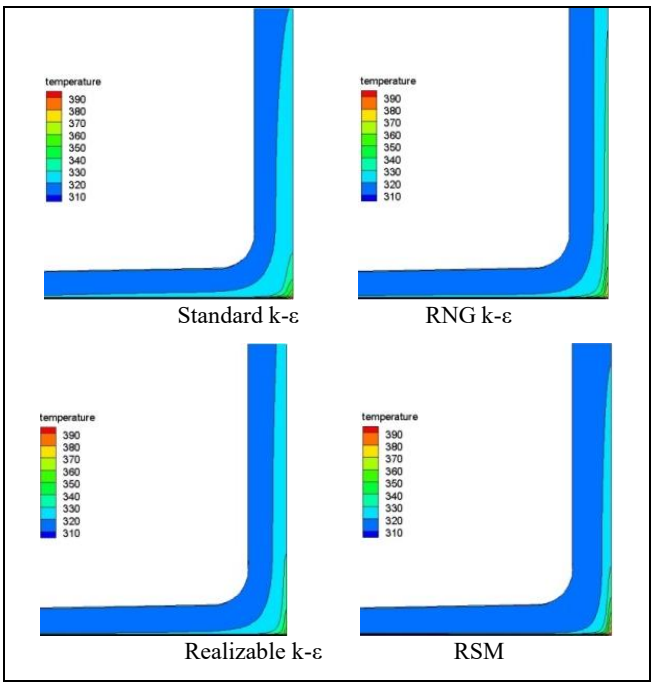
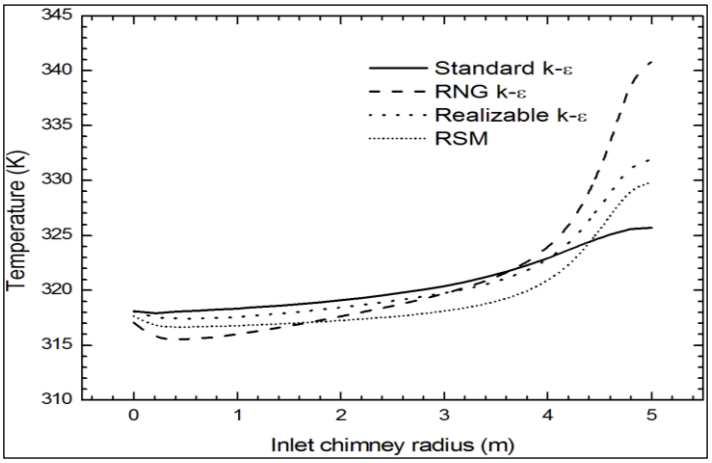
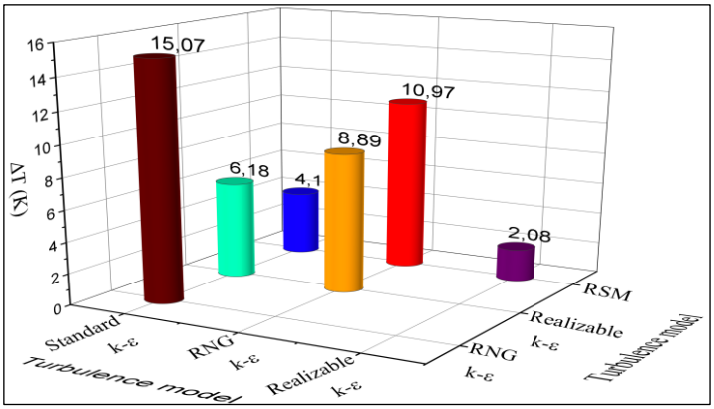


Figure 7: Contours of temperature (H_{col}= 2 m, R_{ch} = 5m).
Source: Authors, (2025).

Figure 8(a) illustrates the temperature profiles along the chimney inlet radius, revealing variations in the temperature value at the chimney according to the different models. To be more precise, for the standard k-ε model, the recorded temperature was 325.71 K, while for the RNG k-ε model, it was 340.78 K. The feasible k-ε model gave a temperature of 331.89 K, and the RSM model gave a temperature of 329.81 K. The maximum temperature difference among the four models was 15.07 K, as illustrated in Figure 8(b).



(a)

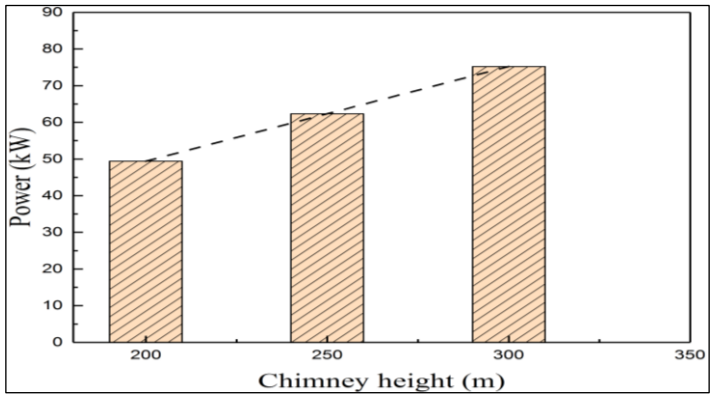


(b)

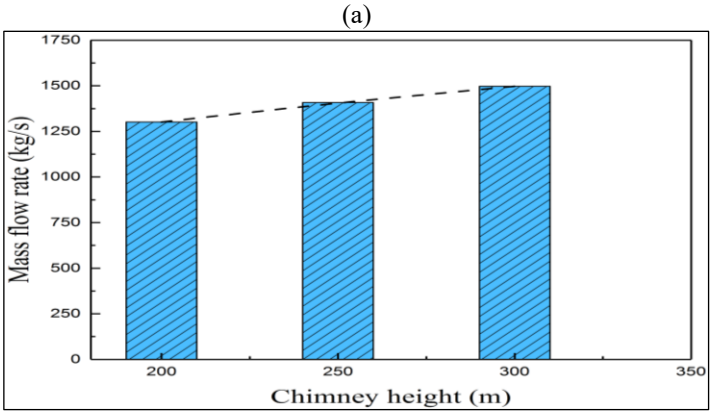
Figure 8: (a) Temperature profiles and (b) the maximum temperature difference between turbulence models.
Source: Authors, (2025).

V.4 CHIMNEY HEIGHT EFFECT

Figure 9 illustrates columns of power and mass flow rate, simulated under identical boundary conditions (G = 960 W/m^2, H = 200 m, Rcol= 2 m, Rch = 5 m). The analysis focuses on the influence of the chimney height (200 m, 250 m and 300 m) on these parameters. The results demonstrate that an increase in height significantly amplifies the generated power (Figures 9(a)). As the air heats up at the base of the solar chimney, it becomes less dense and naturally rises. A greater height provides more space for this hot air to move, creating a greater "thermal draft" effect. This causes the airflow to accelerate, which increases the mass flow rate (Figure 9(b)), and therefore the higher the chimney, the stronger the chimney effect, contributing to a significant increase in the power generated by the solar chimney. For example, from 200 m to 300 m, the mass flow rate increased by 173 kg/s, generating an optimal power of 75 kW. A higher mass flow rate increases the kinetic energy available for power production (via a turbine), thus improving energy conversion efficiency.



(a)

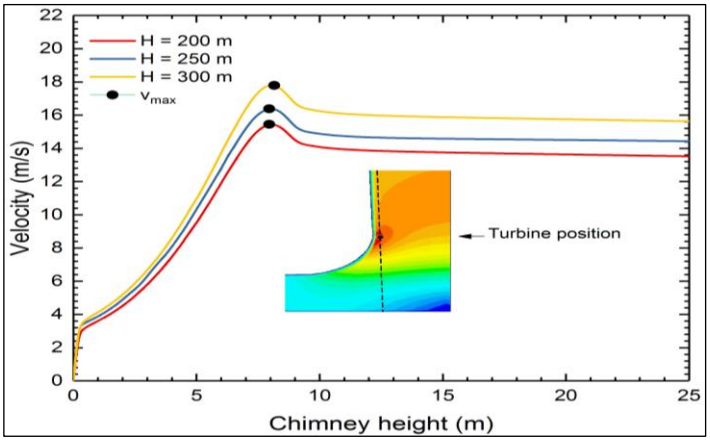


(b)

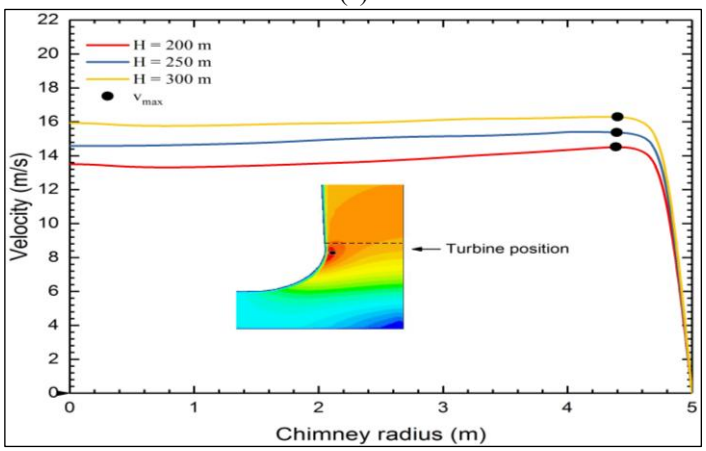
Figure 9: (a) Height of chimney effect on: (a) output power plant and (b) mass flow rate.
Source: Authors, (2025).

Figure 10 illustrates the relationship between chimney height and gas flow velocity. Three separate curves show this relationship for initial chimney heights of 200, 250, and 300 m. A rapid increase in gas velocity is observed with increasing chimney height. This trend is explained by the reduction in flow resistance offered by a taller chimney, thus facilitating gas flow. However, this increase is not linear and reaches a plateau, indicating that the velocity stabilizes beyond a certain height. The increase in air velocity, although always positive, becomes increasingly smaller with increasing chimney height, suggesting the existence of a limit to the velocity improvement.

Near the curved wall of the chimney for all three different heights, the maximum air velocity accelerates due to the centrifugal force pushing the fluid outward, creating a radial pressure gradient. Secondary flows redistribute momentum by directing the fast-moving fluid from the center toward the wall, while turbulence locally reduces the boundary layer thickness. The velocity reaches its maximum just nearby, where these effects combine. Note that the position of the turbine in the solar chimney is crucial, as shown in Figures 10 (a,b). It plays a key role in optimal energy production. This production closely depends on the maximum air velocity. It is necessary to adjust the turbine location according to this velocity. In our case, the ideal location of the turbine is right at the entrance of the solar chimney tower. This location remains the same for all heights considered.



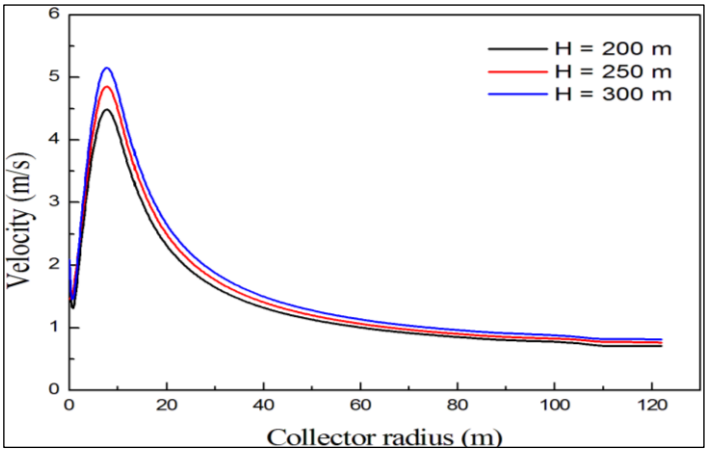
(a)



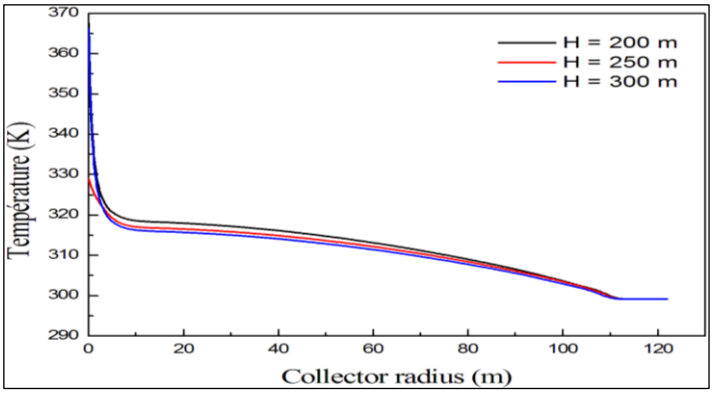
(b)

Figure 10: (a) Velocity profiles : (x =0.5 m) and (b) Chimney inlet (y = 9 m).
Source: Authors, (2025).

When the height of a chimney increases, two main effects occur at the collector outlet, linked to the principles of thermodynamics as illustrated in Figure 11: The increase in velocity and the decrease in temperature. The increase in velocity results from the natural draft caused by Archimedes' thrust, the effect of which increases with height (Figure 11(a)). This thrust, directly proportional to the height, is transformed into kinetic energy by the application of Bernoulli's principle, thus leading to an increase in the velocity at the outlet.



(a)



(b)

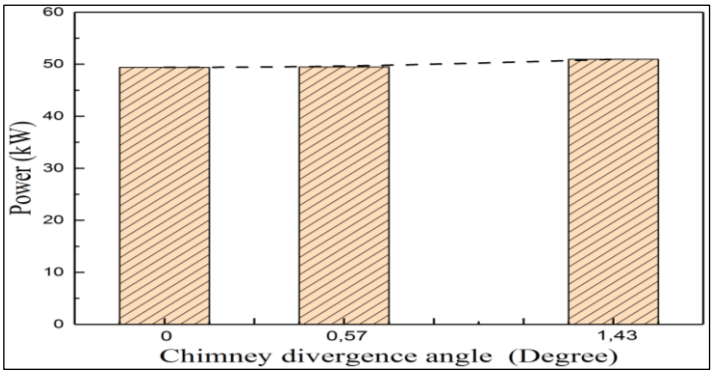
Figure 11: Height of chimney effect on: (a) velocity and (b) temperature.
Source: Authors, (2025).

Furthermore, the decrease in outlet temperature, imposed by relation (20) due to the constant radiated power, is associated with the increase in height (Figure 11(b)). This leads to a higher mass flow rate, because the system draws in more air to compensate for pressure differences. This combination of parameters (increased height, high mass flow rate and fixed power) mechanically leads to cooling at the outlet.

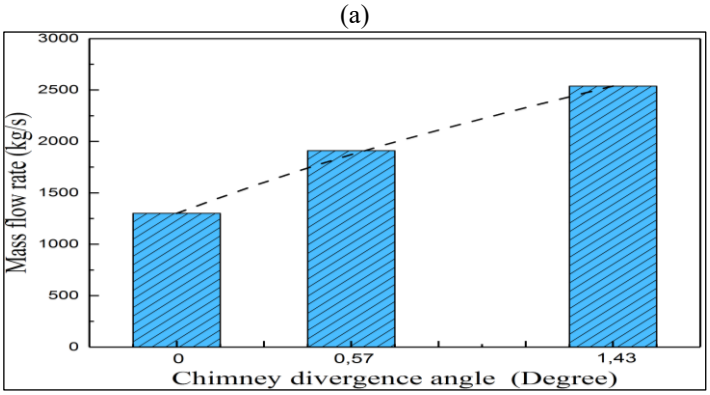
V.5 CHIMNEY DIVERGENCE ANGLE EFFECT

Similarly, Figure 12 is divided into two parts that study the relationship between the exit radius of a chimney, while keeping the inlet radius constant at 5 m, and the two separate variables: gas flow rate and power. The first part of the graph, on the left (Figure 12(a)), examines power as a function of chimney radius. Unlike flow rate, power shows a much weaker relationship with chimney radius. A slight increase in power of 3.18 %, equal to 51 kW, is observed when the chimney exit radius is extended from 5 m to 10 m, corresponding to a diverging flue angle of 1.43°. However, this increase remains negligible compared to the increase in flow rate.

The second part, on the right (Figure 12(b)), shows the gas mass flow rate as a function of chimney radius. It is observed that as the chimney radius increases, the gas flow rate also increases. This increase is not linear, but rather an upward curve. When the chimney exit radius increases from 5 m to 10 m (diverging angle from 0° to 1.43°), the rate of increase in flow rate reaches 95 %. This can be explained by the fact that a larger exit radius provides a larger cross-sectional area for gas flow, thus reducing resistance and allowing a higher flow rate. This suggests that, although increasing the chimney radius significantly improves gas flow, its impact on power is relatively small. This may be due to the fact that power depends on several factors, including gas pressure, which may not increase proportionally with chimney radius.



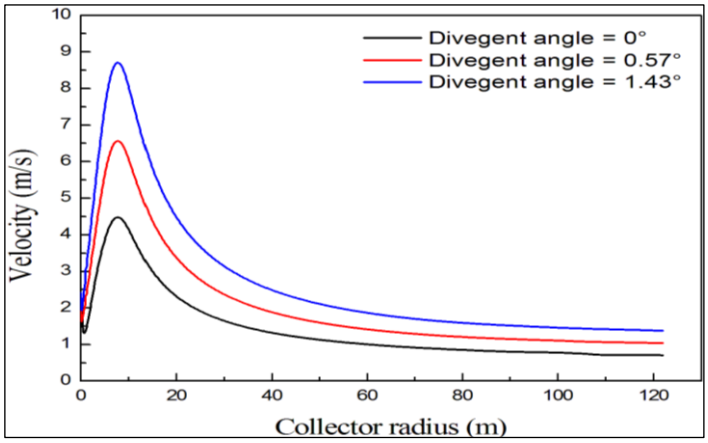
(a)



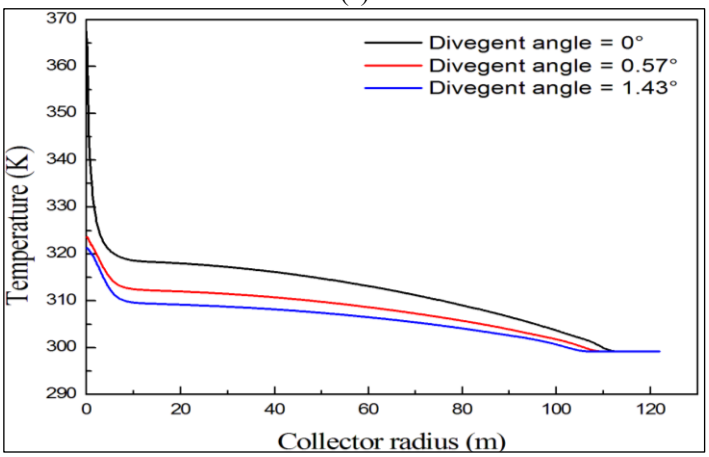
(b)

Figure 12: Chimney divergence angle effect on: (a) output power plant and (b) mass flow rate.
Source: Authors, (2025).

Increasing the divergence angle of the solar tower (i.e., broadening its shape), as illustrated in Figure 13, acts on the system in a similar way to increased height, amplifying two key effects: increased velocity(Figure 13(a)) and decreased temperature(Figure 13(b)) at the collector outlet. A more open divergence angle prolongs the chimney effect by creating a more pronounced vertical pressure gradient. A divergent angle reduces dynamic pressure losses by progressively widening the tower section, which promotes continuous acceleration of the fluid. To respect energy conservation, an increase in mass flow rate necessarily requires a reduction in the temperature difference, and therefore a lower outlet temperature.



(a)

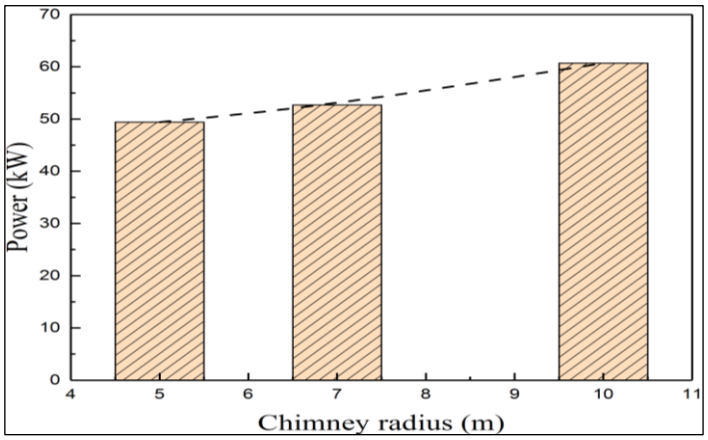


(b)

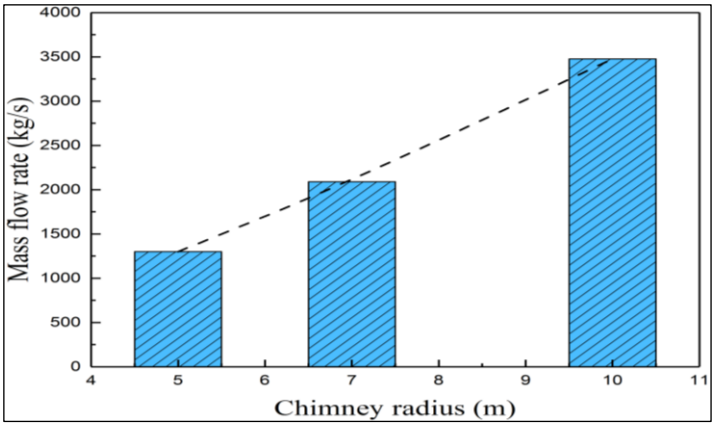
Figure 13. Chimney divergence angle effect on: (a) velocity and (b) temperature
Source: Authors, (2025).

V.6 CHIMNEY RADIUS EFFECT

Figure 14 highlights the decisive impact of the chimney wall profile on the flow characteristics and, consequently, on the output power and flow rate generated by the plant. A marked positive correlation is observed between the chimney radius and the generated power, following a polynomial trend (Figure 14(a)).



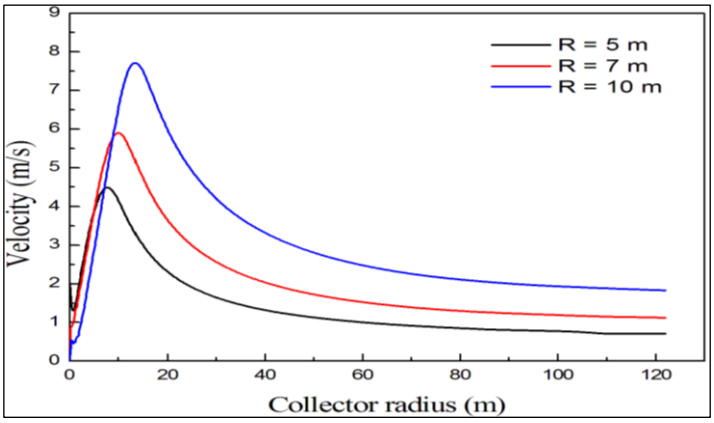
(a)



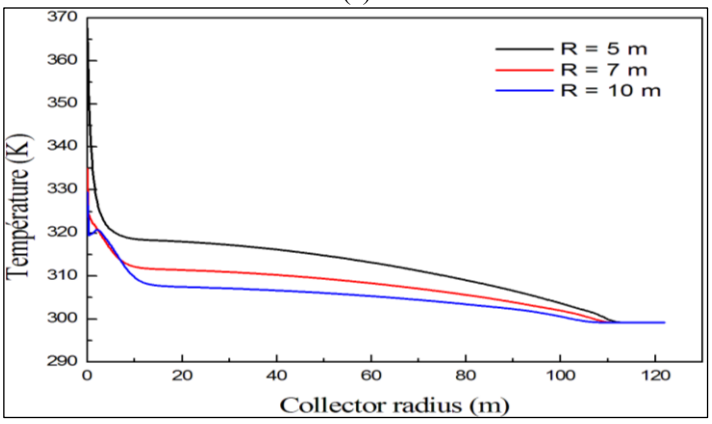
(b)

Figure 14: (a) Chimney radius effect on: (a) output power plant and (b) mass flow rate
Source: Authors, (2025).

For radii between 5 and 7 m, the power progression remains gradual from 49.4 kW to 52.69 kW. Beyond 7 m, this growth intensifies significantly, indicating that the positive effects related to widening then outweigh the potential constraints. According to the curve, an optimal power of 60.7 kW is achieved for a uniform chimney radius of 10 m, with an increase rate of 22.89 %. At this threshold, the transformation of thermal energy into kinetic energy is maximized thanks to an increased air flow and a reduction in turbulence. This nonlinear dynamic thus highlights the acceleration of performance as a function of the radius scale. In terms of mass flow rate, Figure 14(b) reveals a strong positive correlation between the radius of the chimney and that of its outlet, displaying respective increase gradients of 1301 kg/s and 3477 kg/s, ranging from minimum to maximum values. Increasing the uniform radius of the solar chimney (Figure 15), accelerates the air and reduces its temperature at the collector outlet via moderate adiabatic expansion, but these effects remain less pronounced than with an increased divergence angle(Figure 15(b)).



(a)



(b)

Figure 15: Chimney radius effect on: (a) velocity and (b) temperature.
Source: Authors, (2025).

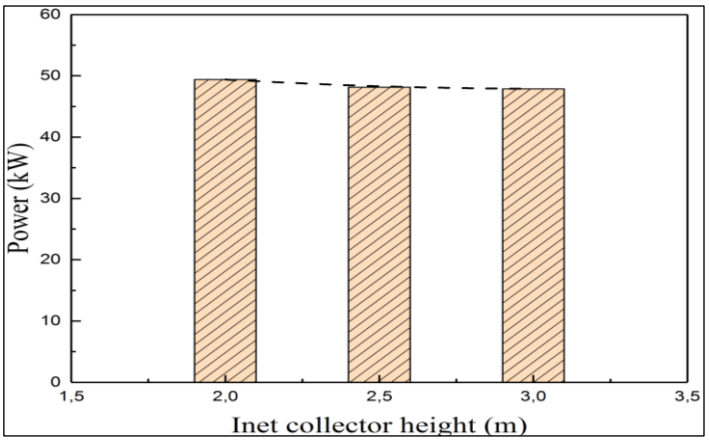
The latter, by progressively widening the cross-section, amplifies the theoretical velocity (by conservation of mass and reverse Venturi effect) and further lowers the temperature thanks to a more pronounced adiabatic expansion. However, a high divergence angle generates significant pressure losses (turbulence, increased friction) that dissipate kinetic energy, limiting the effective velocity and reducing the air density. Conversely, an enlarged uniform radius preserves a stable and laminar flow, minimizing pressure losses while

maintaining a constant cross-section. This allows thermal buoyancy (linked to the density gradient) to dominate, favoring a higher mass flow rate despite lower local gains in velocity or cooling (Figure 15(a)).

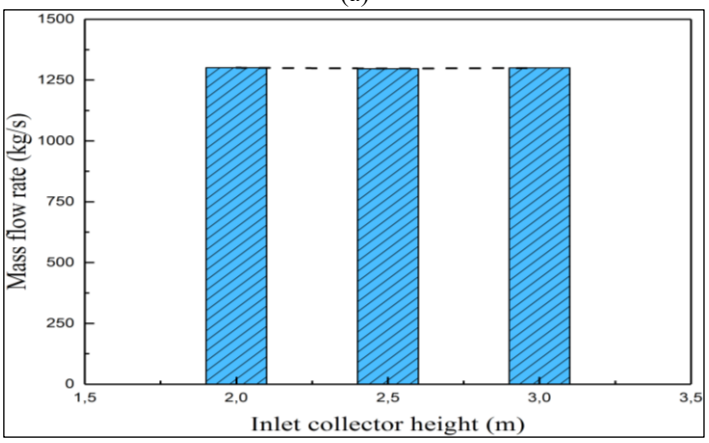
V.7 INLET COLLECTOR HEIGHT EFFECT

Figure 16, a bar graph, relates the collector height to the generated power and mass flow rate under the same specified boundary conditions (G = 960 W/m^2, H = 200 m, R_{ch} = 5 m). Three vertical bars illustrate the power produced at different collector inlet heights (H_{col}): 2 m, 2.5 m and 3 m. A moderate reduction in power can clearly be seen at these three heights, as shown in Figure 16(a). The bar indicating 2 m shows the maximum power, above 49 kW. Thereafter, the power experiences a slight decrease at 2.5 m, before stabilizing at 3 m, where it remains close to 47.8 kW. Between 2 m and 3 m, the power recorded a decrease of 3.1 %.

These variations, although minimal, suggest a possible influence of the collector height on the generated power, although this influence is not drastic in the studied range. However, regarding the mass flow rate (Figure 16(b)), at first glance, a relatively constant mass flow rate of around 1300 kg/s is observed for these three heights. However, a closer observation reveals slight variations. The collector height exerts a limited influence in the range of 2 to 3 m, with a slight decrease in mass flow rate of 4.2 kg/s observed between 2 m and 3 m. This is explained by the fact that a greater collector height reduces the air inlet velocity into the chimney due to a decrease in the pressure gradient, which decreases the mass flow rate and, consequently, the power generated.



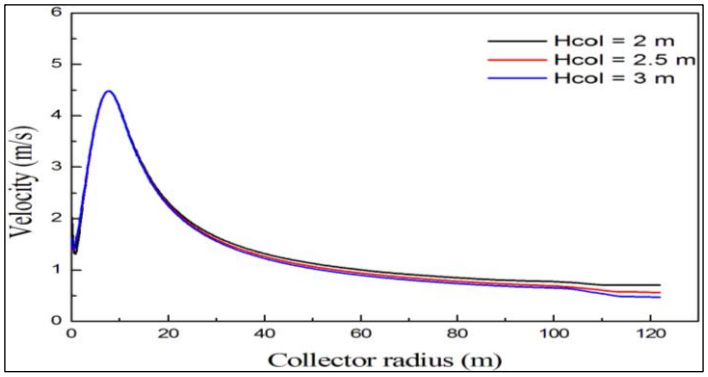
(a)



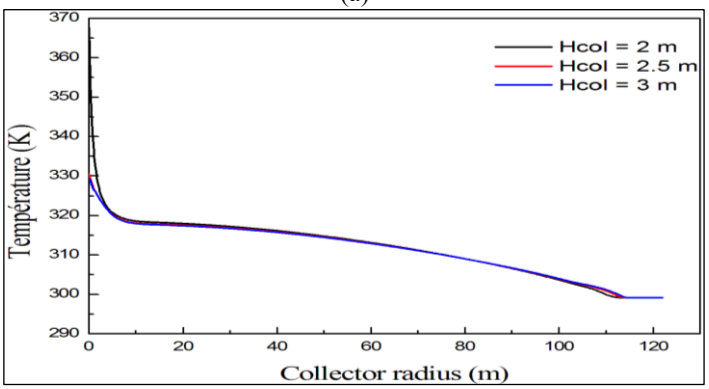
(b)

Figure 16. Inlet collector height effect on: (a) output power plant and (b) mass flow rate
Source: Authors, (2025).

According to Figure 17, the air temperature does not change along the collector because the amount of heat transferred to the air remains proportional to the mass flow rate. If the flow rate decreases (due to reduced velocity), the residence time of the air in the collector increases, compensating for the decrease in the convective heat transfer coefficient (Figure 17(b)). Thus, the heat absorbed per unit mass of air remains the same, maintained by a balance between residence time and heat transfer efficiency. Increasing the inlet height expands the initial cross-sectional area of the collector. According to the continuity equation, for a constant mass flow rate, an increase in the collector inlet area leads to a decrease in the inlet velocity (Figure 171(a)). However, further down the collector, the geometry stabilizes (constant cross-sectional area or balance between buoyancy and pressure drops), which maintains a constant velocity despite the initial reduction.



(a)



(b)

Figure 17: Inlet collector height effect on: (a) velocity and (b) temperature.
Source: Authors, (2025).

V.8 OUPUT POWER OPTIMIZATION

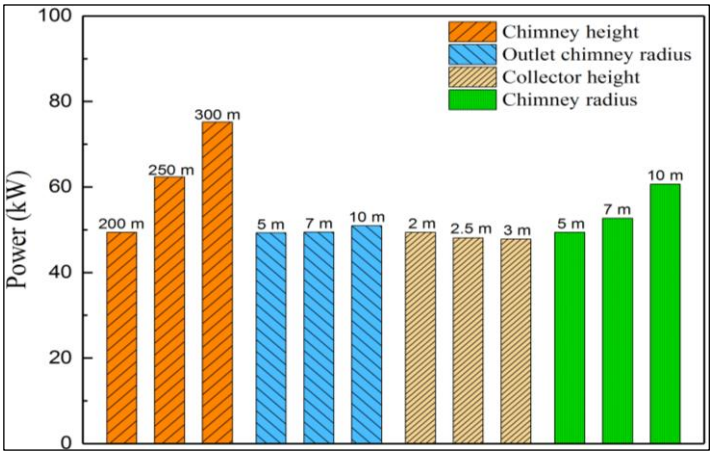
Figure 18 shows the evolution of the power and efficiency generated by the solar chimney as a function of its dimensions, under identical conditions for all tests. The quantitative analysis reveals a significant heterogeneity in the influence of these parameters, as illustrated in detail by Table 4.

Table 4: The percentage increase in SCPP performance (%).

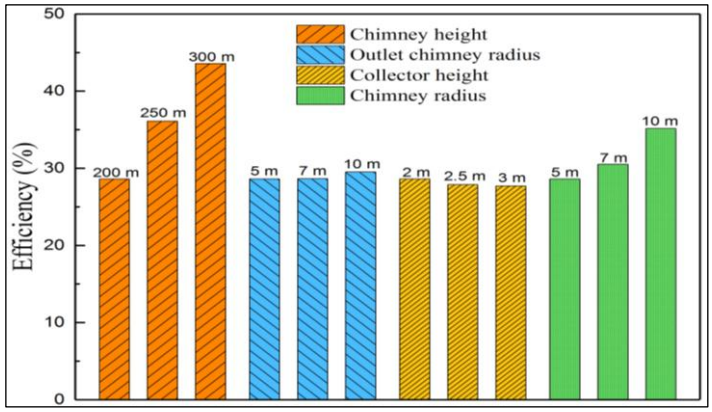
Parameter	Rate of increase (%)			
H_{col} (m)	+25, +50			
H (m)		+25, +50		
R (m)			+40, +100	
Angle divergence (°)				+0.57, +1.43
Power (kW)	-2.53, -3.1	+26.3, +52.3	+6.67, +22.89	+0.21, +3.18

Source: Authors, (2025).

This result mainly highlights, firstly, the considerable effect of the chimney height. For example, if the height of the chimney, which generates a power of 49.4 kW at 200 m in the reference case, were increased to 300 m, the output power would increase by about 52.3 %, reaching almost 75.24 kW, with an efficiency of 43.57 %.



(a)



(b)

Figure 18: Variation under various parameters of: (a) output power plant and (b) chimney efficiency.
Source: Authors, (2025).

VI. CONCLUSIONS

The study of solar chimney performance was approached through an analysis of the effect of various geometric parameters and the adoption of different turbulence models, all supported by numerical simulations via CFD. These investigations highlighted the crucial importance of modeling for the accurate prediction of the behavior of these systems. The evaluation of the ability of turbulence models to simulate the flow in a solar chimney reveals significant disparities and underlines the importance of careful turbulence model selection. It is essential to conduct a thorough examination to identify the most appropriate turbulence model that aligns with the unique geometry and operational conditions. Simply selecting a turbulence model at random may not adequately capture the subtleties of the flow in the solar chimney. This could potentially lead to misleading results that do not accurately represent the physical reality, or worse, produce erroneous results that compromise the efficiency and reliability of the solar chimney. In this study, among the evaluated models, the RNG k-ε turbulence model stood out for its best agreement with the experimental data available in the literature, validating its ability to faithfully capture the essential physical phenomena.

On the other hand, through an analysis of the effect of various geometric parameters, the performance of a solar chimney, in terms of power, mass flow rate or air velocity and temperature, is strongly influenced by its dimensions. The performance of a solar chimney depends on a subtle balance between thermodynamic expansion and flow stability. A large uniform radius favors mass flow rate by stabilizing the flow, while a marked divergence angle locally optimizes velocity and temperature reduction to the detriment of the overall flow rate. The increased height at the collector inlet locally reduces the velocity by increasing the surface area, while the temperature remains stable thanks to a dynamic thermal balance. Furthermore, a greater height generates a higher mass flow rate and significantly amplifies the generated power. To obtain maximum power, it is essential to use a chimney that is both tall and wide, as well as a low collector, which optimizes thermal draft and reduces heat loss. The mass flow rate is mainly determined by the diameter of the chimney, a larger diameter facilitating the air flow. The chimney height has a significant and proportional impact on the output power, while the collector height has a minimal effect, inversely proportional to the power generated in the considered ranges.

VII. AUTHOR'S CONTRIBUTION

Conceptualization: Farid Gaci.
Methodology: Farid Gaci, Ahcene Akni, Moumtez Bensouici.
Investigation: Farid Gaci, Mohamed walid Azizi, Ahcene Akni, Moumtez Bensouici, Djoubeir Debbah.
Discussion of results: Farid Gaci, Ahcene Akni, Moumtez Bensouici.
Writing – Original Draft: Farid Gaci.
Writing – Review and Editing: Farid Gaci, Mohamed walid Azizi, Ahcene Akni, Moumtez Bensouici.
Resources: Farid Gaci, Mohamed walid Azizi, Ahcene Akni, Moumtez Bensouici, Djoubeir Debbah.
Supervision: Mohamed walid Azizi, Ahcene Akni, Moumtez Bensouici.
Approval of the final text: Farid Gaci, Mohamed walid Azizi, Ahcene Akni, Moumtez Bensouici, Djoubeir Debbah.

VIII. ACKNOWLEDGMENTS

The authors would like to extend their heartfelt thanks to the Directorate General of Scientific Research and Technological Development (DGSRTD) for their assistance in this project.

IX REFERENCES

[1] J. Schlaich, The Solar Chimney: Electricity from the Sun. Stuttgart, Germany: Axel Menges, 1995.

[2] N. Pasumarthi and S. A. Sherif, "Experimental and theoretical performance of a demonstration solar chimney model—Part I: Mathematical model development," Int. J. Energy Res., vol. 22, no. 3, pp. 277–288, 1998.

[3] P. Das and V. P. Chandramohan, "Experimental studies of a laboratory scale inclined collector solar updraft tower plant with thermal energy storage system," J. Build. Eng., vol. 41, p. 102394, 2021.

[4] P. Das and V. P. Chandramohan, “Performance characteristics of divergent chimney solar updraft tower plant,” Int. J. Energy Res., pp. 1–16, 2020.

[5] K. Ahmed, S. Algburi, Z. H. Ali, A. K. Ahmed, and H. N. Shubat, “Hybrid solar chimneys: A comprehensive review,” Energy Rep., vol. 8, pp. 438–460, 2022.

[6] H. Nasraoui, Z. Driss, and H. Kchaou, “Novel collector design for enhancing the performance of solar chimney power plant,” Renew. Energy, vol. 145, pp. 1658–1671, 2020.

[7] F. Murena, I. Gaggiano, and B. Mele, “Fluid dynamic performances of a solar chimney plant: Analysis of experimental data and CFD modelling,” Energy, vol. 249, p. 123702, 2022.

[8] R. K. Kumar, N. V. V. K. Chaitanya, and N. S. Kumar, “Solar thermal energy technologies and its applications for process heating and power generation – A review,” J. Clean. Prod., vol. 282, 2021.

[9] A. Ayadi, A. Bouabidi, Z. Driss, and M. S. Abid, “Experimental and numerical analysis of the collector roof height effect on the solar chimney performance,” Renew. Energy, vol. 115, pp. 649–662, 2018.

[10] Y. Wu, T. Ming, R. de Richter, R. Höffer, and H. J. Niemann, “Large-scale freshwater generation from the humid air using the modified solar chimney,” Renew. Energy, vol. 146, pp. 1325–1336, 2020.

[11] S. H. Fallah and M. S. Valipour, “Evaluation of solar chimney power plant performance: the effect of artificial roughness of collector,” Sol. Energy, vol. 188, pp. 175–184, 2019.

[12] H. Kebabsa, M. S. Lounici, M. Lebbi, and A. Daimallah, “Thermo-hydrodynamic behavior of an innovative solar chimney,” Renew. Energy, vol. 145, pp. 2074–2090, 2020.

[13] P. Guo, J. Li, and Y. Wang, “Numerical simulations of solar chimney power plant with radiation model,” Renew. Energy, vol. 62, pp. 24–30, 2014.

[14] A. Mebarki et al., “CFD analysis of solar chimney power plant: Finding a relationship between model minimization and its performance for use in urban areas,” Energy Rep., vol. 8, pp. 500–513, 2022.

[15] A. Hassan, M. Ali, and A. Waqas, “Numerical investigation on performance of solar chimney power plant by varying collector slope and chimney diverging angle,” Energy, vol. 142, pp. 411–425, 2018.

[16] C. B. Maia, A. G. Ferreira, R. M. Valle, and M. F. B. Cortez, “Theoretical evaluation of the influence of geometric parameters and materials on the behavior of the air flow in a solar chimney,” Comput. Fluids, vol. 38, pp. 625–636, 2009.

[17] H. A. E. Mohamad, M. Elsamadony, and M. Ramon, “A Computational Fluid Dynamics Study on Different Solar Chimney Designs on Solar Radiation,” J. Eng. Res., vol. 5, no. 1, pp. 31–37, 2021.

[18] A. B. Awan, V. V. Kotturu, C. Mouli, and M. Zubair, “Performance enhancement of solar tower power plant: A multi-objective optimization approach,” Energy Convers. Manag., vol. 225, p. 113378, 2020.

[19] M. R. Torabi et al., “Investigation the performance of solar chimney power plant for improving the efficiency and increasing the outlet power of turbines using computational fluid dynamics,” Energy Rep., vol. 7, pp. 4555–4565, 2021.

[20] M. Jirena, R. Bhoraniya, and A. Harichandan, “CFD analysis of solar chimney power plant: effect of chimney height, shape and collector size,” J. Sol. Energy Res., vol. 4, no. 1, pp. 61–71, 2019.

[21] Y. Noorollahi et al., “Reliable renewable power production by modeling of geothermal assisted solar chimney power plant,” Geothermics, vol. 111, p. 102701, 2023.

[22] E. Gholamalizadeh and M. H. Kim, “CFD analysis of a solar chimney power plant with inclined collector roof,” Energy, vol. 107, pp. 661–667, 2016.

[23] S. Montelpare, V. D’Alessandro, A. Zoppi, and E. Costanzo, “A solar chimney for renewable energy production: thermo-fluid dynamic optimization by CFD analyses,” IOP Conf. Ser.: J. Phys.: Conf. Ser., vol. 923, no. 1, p. 012047, 2017.

[24] L. Zuo et al., “Performance analysis of a wind supercharging solar chimney power plant combined with thermal plant for power and freshwater generation,” Energy Convers. Manag., vol. 204, p. 112282, 2020.

[25] C. B. Maia, A. G. Ferreira, L. Cabezas-Gomez, J. de Oliveira Castro Silva, and S. de Morais Hanriot, “Thermodynamic analysis of the drying process of bananas in a small-scale solar updraft tower in Brazil,” Renew. Energy, vol. 114, pp. 1005–1012, 2017.

[26] F. F. Daghistani, “Solar chimney street-lighting pole for ventilating polluted urban areas,” Sustain. Cities Soc., vol. 72, p. 103057, 2021.

[27] H. Xiong et al., “Numerical analysis of a negative emission technology of methane to mitigate climate change,” Sol. Energy, vol. 255, pp. 416–424, 2023.

[28] A. A. Serageldin, A. K. Abdelrahman, and S. Ookawara, “Parametric study and optimization of a solar chimney passive ventilation system coupled with an earth-to-air heat exchanger,” Sustain. Energy Technol. Assess., vol. 30, pp. 263–278, 2018.

[29] A. Sellami, D. Benlahcene, A. H. Benmachiche, and Z. Aouachria, “Performance analysis of a solar chimney power plant system in two Algeria regions,” Int. J. Ambient Energy, pp. 1–12, 2021.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11n.55, p.48-58. September/October, 2025.
DOI: https://doi.org/10.5935/jetia.v11i55.1866

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

REDUCING COMPLEXITY IN VEHICLE RECOGNITION: A COMPARISON OF DIMENSIONALITY REDUCTION METHODS FOR ACOUSTIC FEATURES

Yudhi Darmawan[1], Muladi[2], Heru Wahyu Herwanto[3] and Nur Rachman Supadmana Muda[4]

[1,2,3] Department of Electrical Engineering and Informatics, Universitas Negeri malang, Indonesia.
[4] Politeknik Angkatan Darat, Indonesia.

[1]https://orcid.org/0009-0001-5063-6045, [2]https://orcid.org/0000-0002-2904-5398
[3]https://orcid.org/0009-0000-0796-3596, [4]https://orcid.org/0000-0002-9421-7432

Email: yudhi.darmawan.2305349@students.um.ac.id, muladi@um.ac.id, heru_wh@um.ac.id, nurrudal@gmail.com

ARTICLE INFO

Article History
Received: April 15, 2025
Revised: May 20, 2025
Accepted: September 15, 2025
Published: September 30, 2025

Keywords:
Dimensionality Reduction, Vehicle Recognition, Acoustic Features, Intelligent Transportation Systems.

ABSTRACT

Reducing the complexity of high-dimensional acoustic data is essential for effective vehicle recognition, especially in intelligent transportation systems. This study evaluated six dimension reduction techniques, including Principal Component Analysis (PCA), Kernel PCA, Incremental PCA, Independent Component Analysis (ICA), Truncated Singular Value Decomposition (SVD), and Latent Dirichlet Allocation (LDA), to address the challenges of data redundancy while maintaining relevant features. The dataset includes acoustic signals from seven categories of vehicles extracted using Mel-Frequency Cepstral Coefficients (MFCC), Spectral Centroid, and Spectral Bandwidth. Incremental PCA showed the highest accuracy (0.982) on scenarios with larger training datasets, with effective management of high-dimensional data. ICAs provide optimal performance with fewer components at a higher proportion of test data, demonstrating their efficiency in retaining information. SVD shows stability across all data ratios, confirming its reliability for a wide range of applications. Although LDAs maintain competitive results, their interpretability stands out in certain tasks. These findings emphasize the importance of selecting appropriate dimension reduction methods based on data characteristics and application needs, providing valuable insights to improve the accuracy and efficiency of vehicle recognition systems.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The classification of acoustics-based vehicles is increasingly relevant in various applications, such as intelligent transportation, traffic management, and autonomous vehicle systems. Vehicle acoustic analysis offers the advantage of being a non-invasive approach that can be accessed in real-time, even in challenging environmental conditions [1], [2]. Compared to image-based methods, this approach is more efficient for traffic surveillance that requires continuous detection and high reliability [3], [4]. Various methods have been developed to process vehicle acoustic signals, including the use of Mel-Frequency Cepstral Coefficients (MFCC) as a key feature representing the spectral characteristics of the data [5], [6]. In addition, psychoacoustic indicators, such as the Zero Crossing Signature (ZCS), have been used to improve accuracy in vehicle fuel type classification [7], [8]. On the other hand, deep learning models such as CNNs and LSTMs have managed to capture complex non-linear patterns of vehicle signals, which are difficult for conventional techniques to handle [9], [10], [11]. Distributed acoustic sensors have also been implemented to support real-time detection and classification of vehicles with a high degree of accuracy [2], [12], [13].

Previous research has shown great potential for acoustic analysis in vehicle classification. Hybrid models that combine MFCC with Attention BiLSTM have been shown to improve accuracy in emergency vehicles [3]. In addition, spectrogram-based spectral techniques are used to distinguish vehicle types through traffic noise [5], while acoustics- and radar-based multimodal sensors provide richer data for analysis [13], [14]. Acoustic simulation frameworks also play an important role in accelerating model training without

relying on large datasets [15], [16]. While promising, these methods often face challenges due to the high dimensions of the data, which can add to the redundancy of information and hinder the efficiency of analysis. The high dimensions of vehicle acoustic data are one of the main challenges, as the redundancy of the resulting information can increase the risk of overfitting and computational complexity [17], [18]. Synthetic data simulation-based approaches have helped improve the training efficiency of acoustic models, but still face limitations in application to real environments [15], [16]. In addition, deep learning models often require large datasets to achieve optimal performance, which is a constraint in certain applications [19], [20]. This problem is further exacerbated when vehicle acoustic data is not only high-dimensional but also adds complexity to the classification process. This condition can hinder accuracy and efficiency, so solutions are needed to reduce data complexity without losing important information.

In this study, we propose the approach of Reducing Complexity in Vehicle Recognition: A Comparison of Dimensionality Reduction Methods for Acoustic Features to overcome this problem. By comparing various dimension reduction methods such as Principal Component Analysis (PCA), Independent Component Analysis (ICA), Kernel PCA, and other techniques, this study aims to reduce information redundancy while maintaining the most relevant features for vehicle recognition. This approach also evaluates the effectiveness of dimension reduction methods in improving computational efficiency and model accuracy, which is relevant for applications in intelligent transportation systems.

II. THEORETICAL REFERENCE

The introduction of acoustics-based vehicles has attracted widespread attention in the development of intelligent transportation systems and modern traffic surveillance. Acoustic data offers a non-invasive approach that can be used in a variety of environmental conditions, including low-visibility situations. This approach is particularly useful for real-time vehicle detection, especially in scenarios that require continuous surveillance with high reliability. By analyzing the unique acoustic patterns of different types of vehicles, acoustics-based methods have shown great potential in improving the efficiency and accuracy of transportation systems. However, acoustics-based analytics face a major challenge in the form of high data dimensions. Vehicle acoustic data includes a wide range of spectral and temporal features designed to capture vehicle sound characteristics. These high dimensions often result in information redundancy that not only increases the computational load but also increases the risk of overfitting in machine learning models. This is a significant challenge, especially when the dataset used is not large enough to adequately represent the diversity of data, thus reducing the model's ability to generalize under real conditions.

In an effort to overcome these challenges, previous research has developed various data processing techniques. For example, deep learning models such as LSTMs are used to capture non-linear patterns in vehicle acoustic signals [6], [10]. In addition, synthetic data simulations have been applied to improve the efficiency of model training in cases where the original dataset is limited. Nonetheless, these methods often do not fully solve the problem of data redundancy, which remains a major challenge in the introduction of acoustics-based vehicles. The high dimensions of data also have implications for computing efficiency, especially in the development of real-time systems. Big data analysis requires significant computing resources, which are often not ideal for practical applications such as traffic surveillance or autonomous vehicle systems. Thus, an approach is needed that not only simplifies the data structure but also retains relevant information for vehicle recognition. This highlights the importance of dimension reduction techniques as a first step in improving the efficiency of machine learning models for vehicle acoustic analysis.

In this context, our research aims to explore the effectiveness of various dimension reduction methods in addressing the challenges of high-dimensional data. By understanding the role of dimension reduction in filtering relevant information, this study aims to provide a framework that supports the development of more efficient and accurate intelligent transportation systems. Thus, this research is expected to make an important contribution in addressing the technical challenges associated with the introduction of acoustics-based vehicles.

III. MATERIALS AND METHODS

III.1 DATASET

The dataset used in this study consists of two main sources, namely the Vehicle Sounds Dataset [21] and Emergency Vehicle Siren Sounds [22], obtained from Kaggle. These two datasets were chosen because they cover various types of vehicles that are often encountered in transportation and emergency systems. Using acoustic data, this study aims to explore the sound characteristics of vehicles that can be used in a voice-based classification system. This dataset includes seven categories of vehicles, namely ambulances, bics, buses, firetrucks, motorcycles, trains, and trucks, with a varying number of samples for each category. Ambulances, buses, and firetrucks each had 400 sound samples, while the bics category had 300 samples. Meanwhile, motorcycles have 328 samples, trains have 550 samples, and trucks have 240 samples. This variation in the number of samples reflects the availability of records within the dataset, which can affect the class balance in the machine learning process.

All data in this dataset is in the form of audio files in .wav format obtained from various environmental conditions. The available recordings reflect differences in sound intensity, frequency patterns, and the potential for background noise. Therefore, this dataset presents its own challenges in the acoustic analysis process, especially in identifying unique patterns for each type of vehicle. The acoustic characteristics of this dataset allow for further exploration of sound patterns that can be used in the classification of sound-based vehicles. With various variations in data, this study can evaluate the effectiveness of the method used in distinguishing sounds between vehicle classes based on their acoustic features.

III.2 FEATURE EXTRACTION

The feature extraction process aims to convert the audio signal data into numerical representations that can be used in further analysis. In this study, three main features were used, namely Mel-Frequency Cepstral Coefficients (MFCC), Spectral Centroid, and Spectral Bandwidth. Each of these features has a specific role in representing the acoustic characteristics of the vehicle's sound signal.

III.2.1 Mel-Frequency Cepstral Coefficients (MFCC).

MFCC is the most commonly used feature in voice-based pattern recognition, including vehicle acoustic analysis. The MFCC calculation process begins by applying Fast Fourier Transform (FFT) to transform the signal from the time domain to the frequency domain. FFT is calculated using Equation 1.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi kn/N} \quad (1)$$

where $X(k)$ is a representation of the signal in the frequency domain, $x(n)$ is the signal in the time domain, and N is the number of samples in the analysis window. Furthermore, the frequency resulting from the FFT transformation is mapped into the Mel scale, which reflects the human perception of sound. The Mel Scale is calculated using Equation 2.

$$m = 2595\, log_{10}(1 + \frac{f}{7000}) \quad (2)$$

where f is the frequency in Hz, and m is the frequency in the Mel scale. The final step is to take the logarithm of the Mel bank filter energy and apply the Discrete Cosinus Transform (DCT) to eliminate the correlation between the coefficients. DCT is calculated by Equation 3.

$$C_n = \sum_{k=1}^{K} log(E_k)\, cos[n \frac{\pi}{K}(k - 0.5)] \quad (3)$$

where C_n is the MFCC coefficient to- n, E_k adalah energi dalam filter bank ke- k, dan K adalah jumlah filter bank. Dalam penelitian ini, digunakan 13 koefisien MFCC untuk merepresentasikan pola frekuensi utama dari sinyal akustik kendaraan.

III.2.2 SPECTRAL CENTROID

Spectral centroid is a metric used to measure the spectral center of a sound signal. This feature is often associated with the perception of sound "brightness", where the sound with the dominant high frequency will have a greater spectral centroid value. Spectral centroid is calculated with Equation 4.

$$SC = \frac{\sum_{k=1}^{N} f_k S_k}{\sum_{k=1}^{N} S_k} \quad (4)$$

where SC is a spectral centroid, f_k is the frequency of the bin ke-k, S_k is the magnitude of the spectrum on the bin to-k, and N is the total number of frequency bins. Higher spectral centroid values indicate a concentrated distribution of energy at high frequencies, while lower values indicate a predominance of low frequencies.

III.2.3 SPECTRAL BANDWIDTH

Spectral bandwidth measures how dispersed the spectral energy of a signal is around a spectral centroid. This feature is important in distinguishing vehicles based on the complexity of the sound produced. Vehicles such as buses or trucks tend to have a wider spectral bandwidth than motorcycles because they produce more complex sounds. Spectral bandwidth is calculated by Equation 5.

$$SBW = \sqrt{\frac{\sum_{k=1}^{N} S_k (f_k - SC)^2}{\sum_{k=1}^{N} S_k}} \quad (5)$$

where SBW is spectral bandwidth, SC is spectral centroid, f_k is the frequency of bin to-k, S_k is the spectral magnitude at the kth bin, and N is the total number of frequency bins. Higher spectral bandwidth values indicate a more dispersed distribution of energy in the frequency domain, while lower values indicate energy concentrations over a narrow frequency range.

III.3 PREPROCESSING

The preprocessing stage is carried out to ensure that the data extracted from the audio recording has a uniform format and is ready to be used in analysis. The first step is to normalize the feature using the StandardScaler method, which makes the data have a mean of zero and a standard deviation of one. This normalization is important to ensure consistent data scales and support the stability of machine learning algorithms. Furthermore, the negative value on the data is changed to zero to meet the requirements of certain methods, such as Latent Dirichlet Allocation (LDA), which can only work with non-negative data. Also, because the length of the audio file varies, padding is done to add a zero value at the end of the feature, so all recordings have the same dimensions. The maximum length is determined based on the longest record in the dataset. This preprocessing process ensures that the data is free from dimensional and scale inconsistencies, so it is ready to be applied to dimensional reduction and classification methods efficiently.

III.4 DIMENSIONALITY REDUCTION TECHNIQUES

Dimension reduction is a technique used to reduce the number of variables in a dataset while still retaining relevant information. In this study, six dimensional reduction methods were compared to determine their effectiveness in vehicle acoustic analysis, namely Principal Component Analysis (PCA), Kernel PCA, Incremental PCA, Independent Component Analysis (ICA), Truncated Singular Value Decomposition (TSVD), and Latent Dirichlet Allocation (LDA). Each method has different characteristics and approaches to handling high-dimensional data.

III.4.1 PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA is an eigen-decomposition-based dimensionality reduction method that aims to find the key components that capture the greatest variation in a dataset. PCA works by inventing linear transformations that project data into a new space with a smaller number of dimensions but still retain as much information as possible from the original data. The main steps in PCA involve data normalization, calculation of covariance matrices, decomposition of eigen, and transformation to the main component space. 1 detailed algorithm of PCA.

Algorithm 1: *steps from the PCA.*

Input	:	Dataset X with dimension m×n (m samples, n features)
Output	:	Dataset X_{PCA} with dimensions $m \times k$, where $k<n$
Step	1	Data Normalization: Standardize each feature using $X_{norm} = \frac{X-\mu}{\sigma}$
	2	Calculate the Covariance Matrix: $C = \frac{1}{m} X_{norm}^T X_{norm}$
	3	Hitung eigenvalue dan eigenvector: $CW = \lambda W$
	4	Choose k Eigenvektor With Eigenvalue Biggest
	5	Transformasi Data: $X_{PCA} = X_{norm} W_k$

Source: Authors, (2025).

III.4.2 KERNEL PRINCIPAL COMPONENT ANALYSIS (KERNEL PCA)

The PCA kernel is an extension of PCA that allows the detection of non-linear patterns in data by using kernel functions to map data to a high-dimensional feature space before applying PCA. This allows the identification of non-linear relationships that cannot be found with standard PCA. Algorithm 2 is a process of the PCA Kernel

Algorithm 2: *Kernel Process* PCA.

Input	:	Dataset X with dimensions $m \times n$, kernel functions $K(x_i, x_j)$
Output	:	Dataset X_{kPCA} In the new feature space with dimension k
Step	1	Implement Kernel Function: Calculate the kernel matrix $K_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$
	2	Calculate the Eigenvalue and Eigenvector of the Kernel
	3	Select k Eigenvector with the Largest Eigenvalue
	4	Transformasi Data: $X_{kPCA} = KW_k$

Source: Authors, (2025).

III.4.3 INCREMENTAL PRINCIPAL COMPONENT ANALYSIS

Incremental PCA is a variation of PCA that allows the process of gradual reduction of dimensions in small batches, making it more efficient for large datasets. This technique updates the main component each time a new batch is processed without the need to store the entire dataset in memory. Algorithm 3 is the process of Incremental PCA.

Algorithm 3: *process* Incremental PCA.

Input	:	Dataset X Large, batch size B
Output	:	Dataset X_{IPCA} with smaller dimensions
Step	1	Take Batch Data: X_t dari dataset
	2	Normalisasi Data Batch
	3	Update Komponen PCA dengan Batch New: $X_{IPCA}^{(t)} = X_{IPCA}^{(t-1)} + \eta(X_t - X_{IPCA}^{(t-1)})$
	4	Iterasi Until all data is processed

Source: Authors, (2025).

III.4.4 INDEPENDENT COMPONENT ANALYSIS (ICA)

The ICA aims to find statistically independent components in the data. This technique is often used in signal processing, such as separating sound sources from recordings that contain multiple sources. Algorithm 4 is an ICA process.

Algorithm 4: Independent Component Analysis (ICA).

Input	:	Dataset X which is a linear mixture of the source S
Output	:	Independent source S
Step	1	Standarisasi Data
	2	Estimation of the Mixing Matrix: $X = AS$
	3	Find the W Separator Matrix by maximizing statistical independence
	4	Separate Sources: $S = WX$

Source: Authors, (2025).

III.4.5 TRUNCATED SINGULAR VALUE DECOMPOSITION (TSVD)

TSVD is a matrix decomposition-based dimension reduction technique that retains only the largest k singular value to simplify the data. Algorithm 5 is the process of TSDV

Algorithm 5: Proses Truncated Singular Value Decomposition.

Input	:	Dataset X With dimensi $m \times n$
Output	:	Dataset X_{TSVD} with smaller dimensions
Step	1	Do Dekomposisi SVD: $X = U\Sigma V^T$
	2	Select k Largest Singular Value
	3	Rekonstruksi Data: $X_{TSVD} = U_k \Sigma_k V_k^T$

Source: Authors, (2025).

III.4.6 LATENT DIRICHLET ALLOCATION (LDA)

LDA is a probabilistic-based method used to identify latent distributions in datasets, often applied in text analysis but can also be used in feature-based classification. Algorithm 6 is the process of the Latent Dirichlet Allocation (LDA) method.

Algorithm 6: Proses Metode Latent Dirichlet Allocation (LDA).

Input	:	Dataset X which represents the distribution of features
Output	:	Latent distribution in datasets
Step	1	Determine the Number of Latent Components k
	2	Calculate Conditional Probability: $P(w\|z) = \frac{n_{w,z}+\beta}{\sum_{w`}(n_{w`,z}+\beta)}$
	3	Iteration to Update Latent Parameters

Source: Authors, (2025).

III.5 CLASSIFICATION

Classification is the final stage in data analysis to predict classes from unknown data based on previously extracted and reduced features. In this study, the Random Forest Classification algorithm is used, which is an ensemble learning-based method consisting of a set of decision trees. Random Forest's advantages include its resistance to overfitting and its ability to handle high-dimensional datasets.

The Random Forest algorithm works by building multiple decision trees from a random subset of datasets (bagging) and combining the prediction results from each tree to produce the final prediction. Algorithm 7 is a random forest classification step.

Algorithm 7: *Steps* random forest classification.

Input	:	Dataset tereduksi X, Class labels y
Output	:	Class predictions for new data
Step	1	Create n random subsets of datasets (X,y) using bootstrap techniques.
	2	For each subset, build a decision tree based on the Gini Index split criteria
	3	Use the results from each tree to predict the test data class.
	4	Take the final prediction result based on the majority of votes from the entire tree (voting).

Source: Authors, (2025).

III.6 EVALUATION

The model evaluation in this study was carried out to assess the performance of various dimensional reduction and classification methods in recognizing vehicle types based on acoustic data. Four main metrics are used in the evaluation, namely accuracy, precision, recall, and F1-score. These four metrics were chosen because they can provide a comprehensive view of the model's ability to make accurate and consistent predictions. Accuracy measures the proportion of correct predictions compared to the total number of samples. Equation 6 how to calculate accuracy. Precision measures how accurate the model is in predicting a particular type of vehicle, such as an ambulance, compared to the total predictions generated as an ambulance. Equation 7 how to calculate precision. Recall measures the model's ability to capture all samples that actually come from a particular vehicle class. Equation 8 is a way to calculate recalls. F1-score is a harmonious average of precision and recall. Equation 9 is a way to calculate F1-Score.

$$Accuracy = \frac{number\ of\ corrent\ predictions}{Total\ number\ of\ sampel} \tag{6}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{7}$$

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{8}$$

$$F1 - score = 2\ \frac{Precision\ .Recall}{Precision + Recall} \tag{9}$$

IV. RESULTS AND DISCUSSIONS

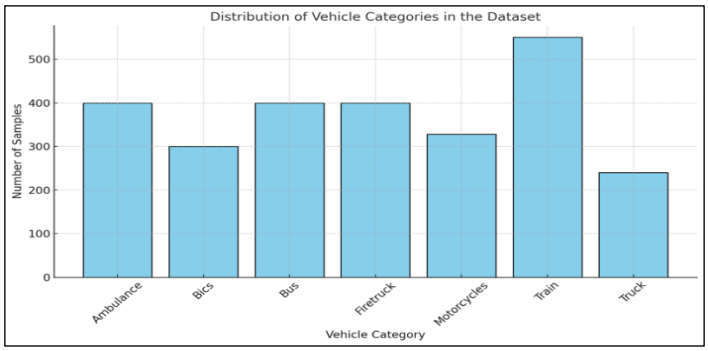


Figure 1: Distribution Of Vehicle Categories In The Dataset.
Source: Authors, (2025).

The dataset used in this study shows an uneven distribution among vehicle categories. The Train category has the highest number of samples, which is 550, while the Truck category has the lowest sample number, which is 240. This imbalance in the amount of data has the potential to create bias in classification results, where the model tends to provide higher accuracy for categories with a larger amount of data. The audio recordings in this dataset are also taken from various environmental conditions, which also adds to the challenges in the classification process. Background noise, such as traffic noise, wind, or other environmental sounds, can obscure the unique acoustic patterns of each vehicle category. This challenge becomes even more significant for categories with fewer sample counts, such as Truck and Bics, as limited data makes their voice pattern representation less than optimal in machine learning models. The distribution of the results of the dataset feature attraction is shown in Figure 2-6, and Table 1.

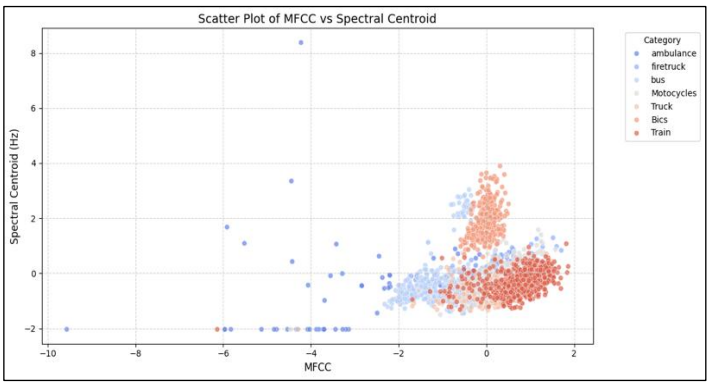


Figure 2 Scatter Plot of MFCC vs Spectral Centroid.
Source: Authors, (2025).

Figure 2 shows the relationship between MFCCs (Mel-Frequency Cepstral Coefficients) and Spectral Centroids for each vehicle category. MFCC represents the primary frequency pattern of acoustic signals, while Spectral Centroid depicts spectral energy centers, which are often associated with the perception of sound "brightness". In Figure 2, it can be seen that vehicle categories such as Train and Bics have a distribution pattern that tends to be separate from other categories, indicating that these features are quite informative in distinguishing several types of vehicles. In contrast, categories such as Ambulance, Bus, and Firetruck have more overlapping distribution patterns, reflecting the similarity in their acoustic characteristics. The distribution of the data also shows the existence of several outliers, especially for categories with a smaller number of samples, such as trucks. This can be a challenge in the classification process, as overlap between categories can lead to prediction errors. However, the combination of MFCC and Spectral Centroid still provides indications of a unique pattern for certain types of vehicles, especially for categories with high-frequency dominant sounds such as Trains.

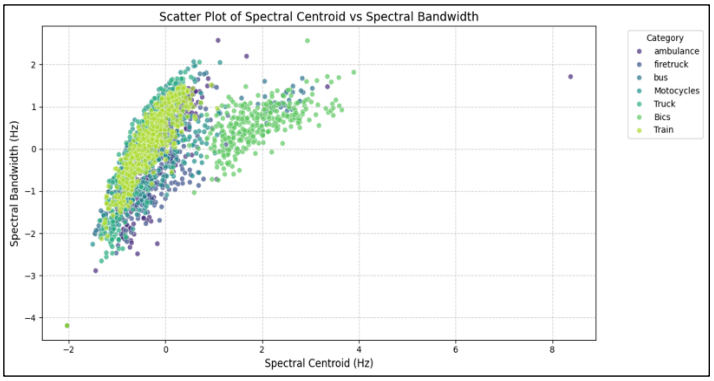


Figure 3: Scatter Plot of Spectral Centroid vs Spectral Bandwidth.
Source: Authors, (2025).

Figure 3 illustrates the relationship between Spectral Centroid and Spectral Bandwidth for each vehicle category in the dataset. Spectral Centroid represents the spectral energy center that is closely related to the dominant frequency in the vehicle's sound, while Spectral Bandwidth measures how wide the spectral energy distribution is around the centroid. The combination of these two features is often used to understand the frequency characteristics and complexity of sound. From Figure 3, it can be seen that most of the vehicle categories have overlapping distributions in the middle area of the graph. For example, categories such as Ambulance, Bus, and Firetruck exhibit fairly similar distribution patterns, suggesting that their acoustic characteristics may have similarities, especially in dominant frequencies and variations in sound bandwidth.

However, some important patterns can be observed, namely the Train and Bics categories show a more dispersed distribution, especially at higher spectral bandwidth values. This reflects that the voices of these categories have greater complexity than other categories. Train sounds, for example, often cover a wide range of frequencies, resulting in a wider bandwidth. The Truck category, although it mostly overlaps with other categories, has some unique data points with higher Spectral Centroid and Spectral Bandwidth values. These dots indicate that there is a certain sound pattern of the truck that is different from other vehicles. Outliers are seen in some categories, such as Ambulances, with very high Spectral Centroid values. Outliers like this can be challenging in the classification process because they can affect model learning and create bias towards certain categories.

Figure 3 also highlights a major challenge in voice-based classification, namely overlap between categories. This overlap suggests that categories such as Ambulance, Bus, and Firetruck have similar acoustic characteristics, making it difficult to separate them using only Spectral Centroid and Spectral Bandwidth. However, this feature remains relevant, especially for categories like Train and Bics, which have more unique acoustic patterns.

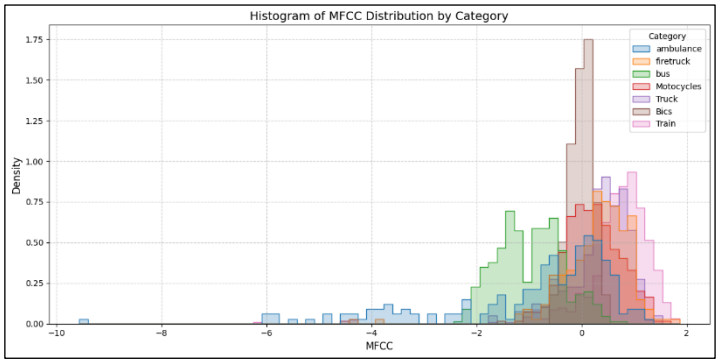


Figure 4: Histogram of MFCC Distribution by Category.
Source: Authors, (2025).

Figure 4 presents a distribution histogram of Mel-Frequency Cepstral Coefficients (MFCC) for each category of vehicles in the dataset. From Figure 4, several important patterns can be observed, namely the MFCC distribution for the Train category shows a high concentration around a small positive value, with a significant peak density compared to other categories. This shows that the sound of the train has a fairly distinctive dominant frequency pattern, which can be an important indicator in the classification process.

The MFCC of the Bics category also has a prominent distribution in the region of small positive values, but with a narrower spread than Train. This pattern reflects that bicycle sounds have a relatively simpler frequency pattern. The Ambulance, Firetruck, and Bus categories have overlapping MFCC distribution patterns, with concentrations around negative to zero values. This overlap suggests that the sounds from emergency vehicles and buses have fairly similar acoustic characteristics, which can be a challenge in differentiating these categories. The Truck and Motorcycles category has a more spread distribution of MFCC, with some negative values that are quite extreme. This suggests that the sound of these two types of vehicles has a greater variation in frequency patterns, perhaps due to differences in engine conditions or recording environments.

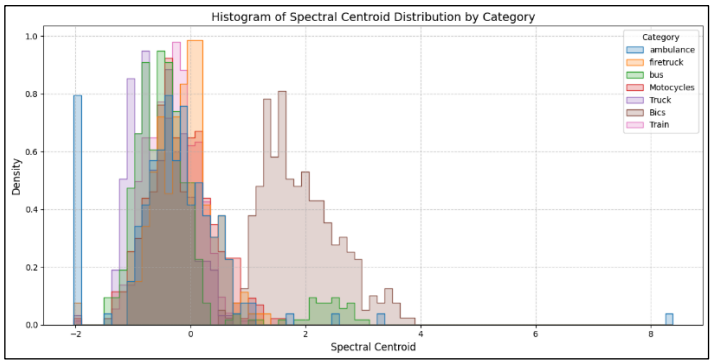


Figure 5: Histogram of Spectral Centroid Distribution by Category.
Source: Authors, (2025).

Figure 5 illustrates the histogram of the Spectral Centroid distribution for each category of vehicles in the dataset. Some important distribution patterns that can be observed from Figure 5 are that the Spectral Centroid distribution for the Train category has a wider range than other categories, with peak densities at values around zero to two. This reflects the complexity of train sounds, which often include a wide range of dominant frequencies due to large engines and wheel systems moving on rails.

The Bics and Truck categories show a fairly unique distribution of Spectral Centroids, with peak densities more spread in small to moderate positive value regions. This spread indicates that the sounds of these two categories have complex frequency patterns, which may reflect variations in engine noise or the influence of recording environmental conditions. The Ambulance, Firetruck, and Bus categories have significant overlap in the Spectral Centroid distribution, with density peaks concentrated around zero values. This overlap suggests that the sounds of emergency vehicles and buses have fairly similar dominant frequency characteristics, which can pose challenges in the classification process. In addition, there are some outliers in certain categories, such as Ambulance, that have Spectral Centroid values far beyond the main distribution. These outliers can be an indicator of the presence of recordings with very different sound characteristics or environmental conditions that affect frequency patterns.

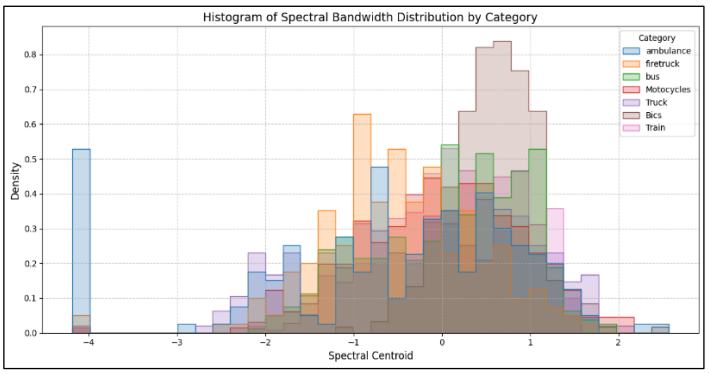


Figure 6: Histogram of Spectral Bandwidth Distribution by Category.
Source: Authors, (2025).

Figure 6 illustrates the distribution of Spectral Bandwidth for each category of vehicles in the dataset. Some important patterns that can be observed from Figure 6, namely the Train category, show a Spectral Bandwidth distribution that tends to be wider, with peak densities around values between 1 and 2. This pattern indicates that the sound of the train has a high frequency complexity, which is one of the hallmarks of this vehicle. The Bics category has a narrower distribution than Train, but still shows a fairly unique pattern in small positive regions. Trucks, on the other hand, have a distribution spread over almost the entire area of the histogram, reflecting the presence of large variations in sound patterns that may be caused by different types of trucks or different recording conditions. The Ambulance, Firetruck, and Bus categories show significant overlap in the distribution of Spectral Bandwidth. A similar distribution among these categories indicates that the complexity of the sounds of emergency vehicles and buses has characteristics that are difficult to distinguish.

Table 1: Descriptive Statistics of Dataset.

Category	MFCC		Spectral Centroid		Spectral Bandwidth	
	Mean	Std	Mean	Std	Mean	Std
Bics	-0.227	1.449	1.90	0.655	0.56	0.484
Motocycles	-0.217	0.897	-0.189	0.558	-0.025	0.923
Train	0.319	0.634	-0.316	0.425	0.146	0.777
Truck	0.135	0.965	-0.55	0.459	-0.238	1.163
Ambulance	-0.263	1.212	-0.298	1.018	-0.611	1.596
Bus	0.066	0.84	-0.297	0.867	0.093	0.856
Firetruck	-0.213	0.963	-0.115	0.499	-0.489	0.899

Source: Authors, (2025).

In addition to being explained using Figures 2 to 6. Table 1, descriptive statistical analysis provides a deeper insight into the feature characteristics of MFCC, Spectral Centroid, and Spectral Bandwidth extracted from the dataset. The mean and standard deviation (std) values for each feature in each category provide an overview of the data distribution pattern, as well as show the potential of the feature to differentiate vehicle categories. In the MFCC feature, the category with the highest average score is Train (0.319). This reflects that the sound pattern of the train has better stability than other categories. In contrast, the lowest average MFCC value was found in Ambulance (-0.263), indicating that the sound of these emergency vehicles tends to have a lower dominant frequency. In addition, the greatest variation in MFCC was seen in the Bics category, with a standard deviation of 1,449. This suggests that the sound of the bicycle has a more diverse frequency pattern, which is most likely due to factors such as engine variations or recording environmental conditions. In the Spectral Centroid feature, the highest average value was found in the Bics category (1.90), which indicates that the bike has a sound with a higher dominant frequency than other categories.

In contrast, the lowest average value for this feature is found in the Truck category (-0.55), which reflects that truck sounds tend to have lower dominant frequencies. In addition, the highest standard deviation for this feature is found in the Ambulance category (1,018), indicating a large variation in the sound patterns of these emergency vehicles. This large variation can be caused by differences in intensity or the type of siren used. Meanwhile, on the Spectral Bandwidth feature, the category with the highest average value is Bics (0.56), which indicates that bicycle sounds have a wider spectrum of energy, reflecting a higher complexity of sound. In contrast, the lowest average bandwidth value was found in the Ambulance category (-0.611), which indicates a lower sound complexity for these emergency vehicles. The largest variation in bandwidth was also found in the Ambulance category, with a standard deviation of 1,596. This suggests that the sound from these emergency vehicles can vary greatly depending on the recording situation or the intensity of the siren used. The vehicle dataset based on sound was trained and tested using random forest classification. The division of training and test data was used, namely 90:10, 80:20, 70:30, and 60:40 percent. Table 2 is the division of training data and test data.

Table 2: number of training data and test data based on ratio.

Kendaraan	Ratio	Jumlah Data Latih	Jumlah Data Uji
Bics	90:10	270	30
	80:20	240	60
	70:30	210	90
	60:40	180	120
Motocycles	90:10	295	33
	80:20	262	66
	70:30	229	99
	60:40	196	132
Train	90:10	495	55
	80:20	440	110
	70:30	385	165
	60:40	330	220
Truck	90:10	216	24
	80:20	192	48
	70:30	168	72
	60:40	144	96
Ambulance	90:10	180	20
	80:20	160	40
	70:30	140	60
	60:40	120	80
Bus	90:10	360	40
	80:20	320	80
	70:30	280	120
	60:40	240	160
Firetruck	90:10	180	20
	80:20	160	40
	70:30	140	60
	60:40	120	80

Source: Authors, (2025).

Based on Table 3, this study shows that the effectiveness of the dimension reduction method is highly dependent on the ratio of training data and test data used, with each method showing superiority in certain scenarios. At a training data:test data ratio of 90:10, Incremental PCA recorded the best performance with an accuracy of 0.982 using 18 components. Methods such as PCA, ICA, and SVD also showed excellent results with an accuracy of 0.977. Meanwhile, the PCA and LDA kernels lag slightly behind with an accuracy of 0.973. When the proportion of test data increases to 20% (80:20 ratio), the performance of most methods remains competitive. Incremental PCA, PCA, Kernel PCA, SVD, and LDA all recorded the same accuracy of 0.973. However, the ICA showed an advantage with an accuracy of 0.975 using 13 components, which reflects its ability to handle a larger proportion of test data without losing significant information.

At a 70:30 ratio, ICA recorded the highest accuracy of 0.979 using only 13 components, demonstrating its efficiency in retaining critical information even though the test data was more dominant. Incremental PCA remains competitive with an accuracy of 0.970 using 20 components. The PCA and SVD methods showed the same results with an accuracy of 0.968, while the PCA and LDA kernels experienced a decrease in accuracy to 0.965 and 0.967, respectively. When the ratio of training data:test data was 60:40, SVD showed the highest performance with an accuracy of 0.976 using the best 18 components, followed by ICA with an accuracy of 0.972 using 19 components. Incremental PCA remains relevant with an accuracy of 0.967 even though it uses only 13 components. LDA also gave good results with an accuracy of 0.974 using 17 components. On the other hand, PCA and PCA Kernel recorded an accuracy of 0.967 and 0.970.

The results of this study show how each dimension reduction method provides optimal performance in various ratios of training data and test data, confirming the relevance of this approach in different scenarios. Incremental PCA is the most superior method in situations with a high training data ratio, such as 90:10, where a larger amount of training data (e.g., 495 training data for the Train category) allows the method to utilize more information to maintain high accuracy (0.982). This highlights the advantages of Incremental PCA in managing large and information-rich datasets. In contrast, at larger test data ratios, such as 60:40, ICAs stand out with their best performance, achieving an accuracy of up to 0.979 at a 70:30 ratio with only 13 components.

ICA's ability to retain critical information even though the number of components is less indicates its efficiency in situations where test data is more dominant. This is particularly relevant for applications where dimension reduction must be achieved without sacrificing accuracy on larger test data. SVD, with its stability in all ratios, reflects its reliability as a reliable method for various data proportions. Regardless of the data ratio, SVD consistently maintains high accuracy, making it a safe choice for scenarios that require versatile dimension reduction. In addition, LDA exhibits competitive performance, especially in applications where component interpretability is a priority. Although the accuracy of LDA is slightly lower than that of Incremental PCA or ICA, its stability makes it a viable option for certain cases. PCA and Kernel PCA, although recording a slight decrease in performance at smaller training data ratios, still demonstrated high adaptability with an accuracy close to 0.97 at most ratios. These results show that these two methods remain relevant in scenarios where the dataset has a more varied

Table 3: Best evaluation results.

Method	Ratio	Best n_components	Accuracy	Precision	Recall	F1-Score
PCA	90:10	17	0,977	0,978	0,977	0,977
	80:20	19	0,973	0,973	0,973	0,973
	70:30	17	0,968	0,969	0,968	0,968
	60:40	20	0,967	0,967	0,967	0,967
Kernel PCA	90:10	16	0,973	0,973	0,973	0,973
	80:20	19	0,973	0,973	0,973	0,973
	70:30	14	0,965	0,965	0,965	0,965
	60:40	17	0,970	0,970	0,970	0,970
Incremental PCA	90:10	18	0,982	0,982	0,982	0,982
	80:20	19	0,973	0,973	0,973	0,973
	70:30	20	0,970	0,970	0,970	0,970
	60:40	13	0,967	0,967	0,967	0,967
ICA	90:10	17	0,977	0,978	0,977	0,977
	80:20	13	0,975	0,976	0,975	0,975
	70:30	13	0,979	0,979	0,979	0,979
	60:40	19	0,972	0,972	0,972	0,972
SVD	90:10	18	0,977	0,978	0,977	0,977
	80:20	12	0,973	0,973	0,973	0,973
	70:30	19	0,968	0,969	0,968	0,968
	60:40	18	0,976	0,976	0,976	0,976
LDA (Latent Dirichlet)	90:10	19	0,973	0,973	0,973	0,973
	80:20	17	0,973	0,973	0,973	0,973
	70:30	17	0,967	0,968	0,967	0,967
	60:40	17	0,974	0,974	0,974	0,974

Source: Authors, (2025).

distribution. This study emphasizes that the selection of dimension reduction methods must consider the distribution of training data and test data. These findings make a significant contribution to the understanding of the effectiveness of various dimension reduction approaches, both for datasets with a high proportion of training data and for datasets with more dominant test data. With in-depth analysis and measurable results, the study offers clear guidance for other researchers in choosing the method that best suits their needs.

V. CONCLUSIONS

This study shows the effectiveness of dimension reduction techniques in overcoming the challenge of high-dimensional acoustic data for vehicle recognition. By comparing six methods on different ratios of training and test data, Incremental PCA proved to be the most effective approach for scenarios with larger training datasets, achieving the highest accuracy of 0.982. Meanwhile, ICA excels with a smaller number of components under the condition of a higher proportion of test data, reflecting its efficiency in maintaining important features. SVD's stability across all ratios highlights its reliability for a wide range of applications, while LDA remains relevant for tasks that prioritize interpretability. These findings emphasize the importance of selecting appropriate dimension reduction methods based on data characteristics and application needs, and provide valuable insights for the development of efficient and accurate intelligent transportation systems.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Yudhi Darmawan and Muladi.
Investigation: Yudhi Darmawan.
Discussion of results: Yudhi darmawan, Muladi, Heru Wahyu Herwanto, and Nur Rachman Supadmana Muda.
Methodology: Yudhi Darmawan, Muladi.
Writing – Original Draft: Yudhi Darmawan, and Muladi. **Writing – Review and Editing:** Yudhi Darmawan, Muladi.Heru Wahyu Herwanto and Nur Rachman Supadmana Muda.
Supervision: Muladi, Heru Wahyu Herwanto, and Nur Rachman Supadmana Muda.
Approval of the final text: Yudhi darmawan, Muladi, Heru Wahyu Herwanto, and Nur Rachman Supadmana Muda.

VII. ACKNOWLEDGMENTS

We thank the promoters for their support in realizing this article. This research is part of the doctoral program in the electrical engineering and informatics study program at the State University of Malang. All authors contributed to the research and writing of this article with the authority and approval to submit the manuscript.

VIII. REFERENCES

[1] T. M. Nithya, P. Dhivya, S. N. Sangeethaa, and P. Rajesh Kanna, "TB-MFCC multifuse feature for emergency vehicle sound classification using multistacked CNN – Attention BiLSTM," Biomed. Signal Process. Control, vol. 88, no. PA, p. 105688, Feb. 2024, doi: 10.1016/j.bspc.2023.105688.

[2] S. Damiano, L. Bondi, S. Ghaffarzadegan, A. Guntoro, and T. van Waterschoot, "Can Synthetic Data Boost the Training of Deep Acoustic Vehicle Counting Networks?," in ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Apr. 2024, pp. 631–635. doi: 10.1109/ICASSP48485.2024.10447868.

[3] C.-Y. Chiang, R. Zhong, M. Jaber, and P. Hayward, “A Distributed Acoustic Sensor System for Intelligent Transportation using Deep Learning,” IEEE Internet Things Mag., vol. 7, no. 5, pp. 88–96, Sep. 2024, doi: 10.1109/IOTM.001.2300287.

[4] K. K. M. Shariff, M. A. Haron, M. S. A. M. Ali, I. M. Yassin, M. H. Azami, and M. A. Kechik, “Comparison of Spectrograms for Classification of Vehicles from Traffic Audio,” in 2024 IEEE Symposium on Wireless Technology & Applications (ISWTA), IEEE, Jul. 2024, pp. 12–17. doi: 10.1109/ISWTA62130.2024.10651848.

[5] R. Evans, S. Mullins, and G. Noel, “Challenges of Vehicle Classification Using Acoustics,” Int. FLAIRS Conf. Proc., vol. 35, May 2022, doi: 10.32473/flairs.v35i.130658.

[6] A. I. Yassin, K. K. Mohd Shariff, M. A. Kechik, A. M. Ali, and M. S. Megat Amin, “Acoustic Vehicle Classification Using Mel-Frequency Features with Long Short-Term Memory Neural Networks,” TEM J., vol. 12, no. 3, pp. 1490–1496, Aug. 2023, doi: 10.18421/TEM123-29.

[7] M. Ashhad, U. Goenka, A. Jagetia, P. Akhtari, S. K. Ambat, and M. Samuel, “Improved Vehicle Sub-type Classification for Acoustic Traffic Monitoring,” in 2023 National Conference on Communications (NCC), IEEE, Feb. 2023, pp. 1–6. doi: 10.1109/NCC56989.2023.10067994.

[8] J. Murovec, J. Prezelj, D. Ćirić, and M. Milivojčević, “Zero Crossing Signature: A Time-Domain Method Applied to Diesel and Gasoline Vehicle Classification,” IEEE Sens. J., vol. PP, no. Xx, pp. 1–1, 2024, doi:
10.1109/JSEN.2024.3516876.

[9] A. Chandra and G. Singh, “Hybrid Classification Model For Acoustic-Based Emergency Vehicle Detection,” in 2023 6th International Conference on Contemporary Computing and Informatics (IC3I), IEEE, Sep. 2023, pp. 1199–1205. doi: 10.1109/IC3I59117.2023.10397746.

[10] Y. Luo, L. Chen, Q. Wu, and X. Zhang, “Sound-Convolutional Recurrent Neural Networks for Vehicle Classification Based on Vehicle Acoustic Signals,” in 2021 International Conference on Smart City and Green Energy (ICSCGE), IEEE, Nov. 2021, pp. 98–102. doi: 10.1109/ICSCGE53744.2021.9654357.

[11] Y. Pan, Y. Dong, D. Wang, J. Di, and A. Chen, “Vehicle Detection and Classification by Voiceprint Recognition Based on Single Acoustic Sensor Under Bridge Expansion Joint,” IEEE Trans. Instrum. Meas., vol. 73, pp. 1–15, 2024, doi: 10.1109/TIM.2024.3428610.

[12] R. Zhong, C.-Y. Chiang, and M. Jabor, “Distributed Acoustic Sensor System for Vehicle Occupancy Detection,” in 2024 IEEE World Forum on Public Safety Technology (WFPST), IEEE, May 2024, pp. 132–137. doi: 10.1109/WFPST58552.2024.00022.

[13] C.-Y. Chiang, M. Jaber, K. K. Chai, and J. Loo, “Distributed Acoustic Sensor Systems for Vehicle Detection and Classification,” IEEE Access, vol. 11, no. March, pp. 31293–31303, 2023, doi: 10.1109/ACCESS.2023.3260780.

[14] I. Corera, E. Piñeiro, J. Navallas, M. Sagues, and A. Loayssa, “Long-Range Traffic Monitoring Based on Pulse-Compression Distributed Acoustic Sensing and Advanced Vehicle Tracking and Classification Algorithm,” Sensors, vol. 23, no. 6, p. 3127, Mar. 2023, doi: 10.3390/s23063127.

[15] J. Abeber, S. Gourishetti, A. Katai, T. Claub, P. Sharma, and J. Liebetrau, “IDMT-Traffic: An Open Benchmark Dataset for Acoustic Traffic Monitoring Research,” in 2021 29th European Signal Processing Conference (EUSIPCO), IEEE, Aug. 2021, pp. 551–555. doi: 10.23919/EUSIPCO54536.2021.9616080.

[16] S. Damiano, L. Bondi, A. Guntoro, and T. van Waterschoot, “A framework for the acoustic simulation of passing vehicles using variable length delay lines,” EURASIP J. Audio, Speech, Music Process., vol. 2024, no. 1, p. 49, Oct. 2024, doi: 10.1186/s13636-024-00372-4.

[17] Y. Houkpati et al., “Siamese Networks for Autonomous Classification of Battlefield Ground Vehicles Using Acoustic Data,” in 2023 International Conference on Computational Science and Computational Intelligence (CSCI), IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/CSCI62032.2023.00008.

[18] V.-T. Tran and W.-H. Tsai, “Acoustic-Based Emergency Vehicle Detection Using Convolutional Neural Networks,” IEEE Access, vol. 8, pp. 75702–75713, 2020, doi: 10.1109/ACCESS.2020.2988986.

[19] A. Y. Yazli, K. K. Mohd Shariff, S. A. Mutalib Al Junid Syed Abdul Rahman, and M. A. Md Ali, “Multimodal Vehicle Classification Based on Radar and Acoustic Sensors Using Hybrid Shallow CNN,” in 2024 IEEE Symposium on Wireless Technology & Applications (ISWTA), IEEE, Jul. 2024, pp. 264–268. doi: 10.1109/ISWTA62130.2024.10652023.

[20] A. Barros et al., “Beyond noise levels: vehicle classification using psychoacoustic indicators from pass-by road traffic noise and their correlations with speed and temperature,” Appl. Acoust., vol. 214, no. June, p. 109716, Nov. 2023, doi: 10.1016/j.apacoust.2023.109716.

[21] Kaggle, “Vehicle Sounds Dataset.” Accessed: Jan. 30, 2025. [Online]. Available: https://www.kaggle.com/datasets/janboubiabderrahim/vehicle-sounds-dataset

[22] Kaggle, “Emergency Vehicle Siren Sounds.” Accessed: Jan. 30, 2025. [Online]. Available: https://www.kaggle.com/datasets/vishnu0399/emergency-vehicle-siren-sounds/

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11n.55, p. 59-72. September/October, 2025.
DOI: https://doi.org/10.5935/jetia.v11i55.1881

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

IMPROVING THE EFFICIENCY OF THE TECHNOLOGICAL PROCESS OF ELECTROMECHANICAL TREATMENT OF MACHINE COMPONENTS BY MINIMIZING THE NUMBER OF CURRENT-CARRYING CIRCUIT ELEMENTS AND DESIGNING SPECIAL EQUIPMENT

Elena Veremey[1], Dmitriy Podashev[2]

[1,2] Kaliningrad State Technical University – Kaliningrad, Russia.

[1]http://orcid.org/0000-0001-7847-540X, [2]http://orcid.org/0000-0001-9112-9253

Email: elena.kerevichene@klgtu.ru , dmitrij.podashev@klgtu.ru

ARTICLE INFO

Article History
Received: April 17, 2025
Revised: May 20, 2025
Accepted: September 15, 2025
Published: September 30, 2025

Keywords:
electromechanical processing,
automated installation,
special equipment,
agricultural machinery parts,
energy saving

ABSTRACT

The article is devoted to improving the efficiency of the electromechanical processing technology for parts of agricultural machinery through minimizing the number of elements in the current-carrying circuit and designing special equipment. The authors examine the technology of electromechanical restoration and hardening of agricultural machinery components, focusing on the problems of spare parts shortage and insufficient production capacity of Russian manufacturers. Existing installations for electromechanical processing are analyzed, their advantages and disadvantages are identified, and the importance of automation processes for enhancing the performance characteristics of processed surfaces is emphasized. A developed prototype of an automated installation equipped with a specialized voltage analysis module and software that contributes to reducing electrical losses and improving equipment operating conditions is presented. The design features of the specially designed fixture are described, including calculations of key components and material requirements. Recommendations are provided regarding the selection of optimal designs and materials for creating various units of the fixture.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Currently, the agricultural sector of the Russian Federation faces a significant shortage of tractors and other farm machinery [1]. At the same time, replenishing the fleet of combines and tractors in required volumes poses challenges because imports have been halted due to sanctions, while domestic producers cannot manufacture enough machines [1]. Given this situation, the qualitative indicators of Russia's agro-industrial complex will directly depend on both the quantity and quality of its agricultural machinery fleet in the near future [2]. Under these circumstances, it is essential to properly maintain and carefully operate existing equipment. Due to the lack of spare parts for foreign-made agricultural machinery and even domestically produced ones (it is known that up to 30...40% of components in domestic agricultural machinery consist of imported parts), there is a need to shift from component replacement to restorative and strengthening repairs [3-7].

There are numerous methods for treating and restoring parts, such as welding and overlay welding; galvanization and metallization; compression-plastic deformation; adhesion with plastics; application of polymeric materials onto worn surfaces, as well as electromechanical machining [4-21]. Electrochemical and electromechanical treatments have gained widespread use in hardening and restoring components of various mechanisms and machines, including those used in agriculture [7-10, 18, 19]. The essence of this type of part restoration lies in passing a high-current low-voltage electric current through a rolling tool. As a result, localized heating occurs in the contact zone between the roller and the part's surface, bringing metal into a plastic state. Through controlled heating and cooling of the contact area, highly durable and wear-resistant layers form on the surface, allowing not only restoration but also reinforcement of the parts.

Examples of electromechanical treatment aimed at restoring agricultural machinery parts are shown in Figure 1.

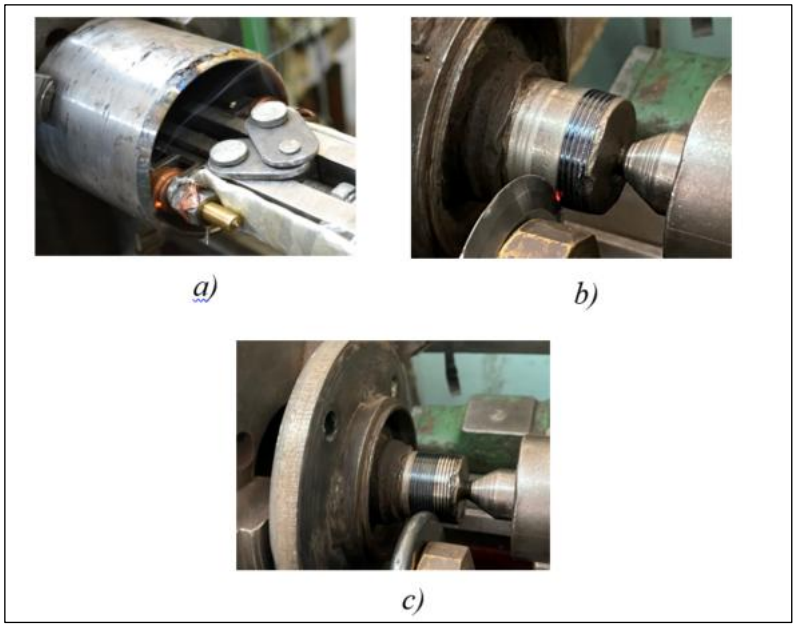


Figure 1: Examples of electromechanical treatment aimed at restoring agricultural machinery partsa: a) electromechanical hardening of the working surface of the vertical joint bushing of tractor K-701 and its modifications; b), c) restoration of the neck of the roller shaft axis of disk harrow BD-7 (b – electromechanical extrusion; c – electromechanical smoothing).
Source: Authors, (2025).

At present, the following types of installations for electromechanical processing are applied in Russian industry:

1. UEMO1 Installation – contains an autotransformer enabling stepwise adjustment of current intensity. Additionally, a rheostat provides smoother regulation within each step. Power supply voltage is 380 volts, rated power output is 10 kW, secondary winding delivers voltage ranging from 2 to 6 volts and current up to 1500 amperes. The unit is housed in an electrical cabinet, manual control is implemented. Due to intense heating during operation, liquid cooling of conductors is necessary, requiring access to a water source nearby. One drawback includes the heavy weight of the transformer, necessitating lifting equipment when moving the setup.
2. UEMO2 Installation – provides up to 1800 amperes in the secondary winding. It has rollers enabling mobility, otherwise similar to UEMO1.
3. Standard Installation – stationary power unit widely applicable, including as a current source for electromechanical processing. Unit weighs 120 kg. Its schematic arrangement does not differ significantly from previous models, delivering adjustable currents up to 3000 amperes.
4. Kasatka Installation – stationary device with manual control via switching transformer's stages and additional thyristor-based power regulation. Provides secondary winding currents from 0 to 5000 amperes, peak power rating is 25 kW. Liquid cooling of transformer and conductors is mandatory. Supply voltages can be either 380 or 220 volts.
5. EMUR Installation – compact transformer combined with a thyristor-controlled circuit. Allows connection of auxiliary management devices. Secondary winding outputs currents from 0 to 1500 amperes, maximum voltage is 4 volts. Unlike others, air cooling is employed instead of liquid cooling. Peak power consumption reaches 4 kW.

All reviewed installations feature manual controls except EMUR which lacks independent control means and requires additional connections. Continuous operator supervision is needed for all systems. Moreover, inconvenience, lengthy procedures, and inaccurate adjustments of power sources lead to subpar surface finish quality. Identified shortcomings provide sufficient grounds for developing a prototype capable of automating the process, increasing productivity and work quality, conserving electricity, and optimizing labor resources.

II. MATERIALS AND METHODS

The following research methods were used in preparing this article:

Theoretical methods.

Analysis of literary sources — studying and summarizing information from various publications dedicated to issues related to electromechanical processing and restoration of agricultural equipment parts.

Description of technologies and processes — detailed explanation of the essence of methods for restoring and hardening parts through electromechanical processing.

Review of installation designs — characterization of existing installations for electromechanical processing, identification of their advantages and disadvantages.

Experimental methods.

Conducting measurements — measuring electrical characteristics during operation of existing equipment, including voltage drop and energy loss in circuits.

Calculation of technical parameters — calculation of loads, stresses, and deformations of elements of special tooling designed for a new technological scheme.

Development of design solutions — creation of a new adaptive mounting system for transformers and design of frame components ensuring stability and reliability of the entire structure.

Creation of functional diagrams — development of a functional diagram of an automated installation incorporating a microprocessor controller and sensors that optimize the electromechanical processing procedure.

Engineering method.

Strength calculations — conducting engineering strength calculations for individual components of the tooling assembly, determining optimal cross-section sizes of steel profiles and diameters of metal shafts used in the framework and mechanism for moving carriages.

Using these methods allowed the authors to justify the need for developing specialized tooling, which contributes to increasing efficiency and accuracy of electromechanical part restoration while reducing energy losses and ensuring safe operation of the equipment.

III. RESULTS AND DISCUSSIONS

III.1. IMPROVING THE QUALITY OF ELECTROMECHANICAL PROCESSING USING A DEVELOPED ADAPTIVE SYSTEM FOR CONTROLLING TECHNOLOGICAL CURRENT

The basis for improving the productivity of the processing operation and enhancing its quality is the automatic control of the power supply unit with the option to switch to manual mode if necessary. This makes it possible to use more compact specialized units. Therefore, it is reasonable to automate the described process by utilizing software and a microcontroller, to which a measurement device will be connected during the operation of the installation. As a result, different operating algorithms can be set without altering the initial circuit or employing larger-sized equipment.

The functional diagram of this installation is shown in Figure 2.

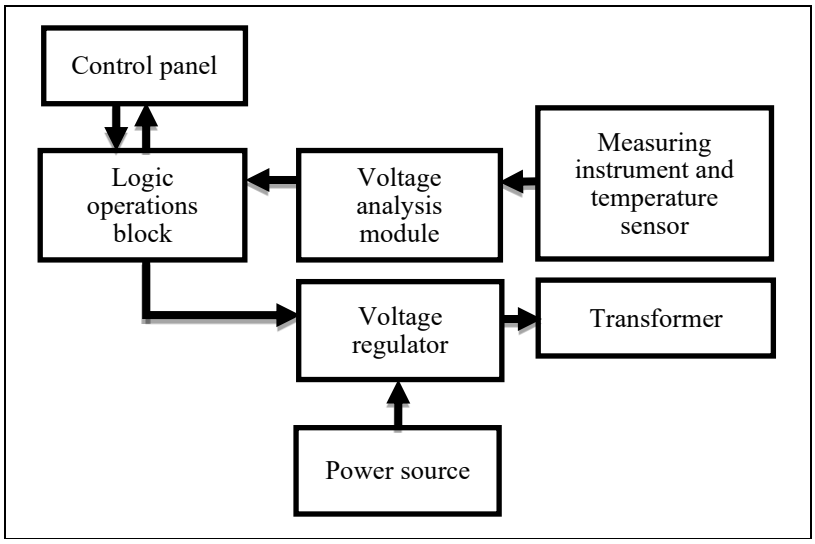


Figure 2: Functional diagram of the installation.
Source: Authors, (2025).

On the control panel are located the installation's controls such as: main toggle switch, mode selection toggle switch, current setting buttons, start button, and display for showing digital value of the current force. With the help of the control panel, the working mode of the installation and the current intensity are defined.

The logic operations block is intended for collecting and processing data, as well as sending commands to controlled devices according to the embedded program.

A measuring device is required to measure the contact resistance between the processing tool and the workpiece. It consists of an inductance coil wrapped around one of the conductors leading from the transformer to the tool. The current flowing along the conductor induces an induced current within the coil.

A temperature sensor is needed to monitor the temperature and prevent overheating of the processed surface.

The voltage analysis module converts the induced voltage in the coil into a digital format and sends it to the logic operations block.

The voltage regulator increases or decreases the electricity supply to the transformer based on commands received from the logic operations block.

Directly, the transformer converts standard network voltage into a voltage with specific characteristics suitable for implementing the technological process.

Additionally, there are five contacts available on the logic operations block for connecting a programmer to change the controlling program, as well as unused contacts that could be utilized for additional sensors.

It is proposed to use the Arduino Uno R3 model microcontroller, whose core is the ATmega8U2 chip.

Since the electric power source is quite compact (dimensions 115 × 150 × 275 mm), it is feasible to mount it directly onto the lathe saddle (model 1K62). This solution allows freeing up workspace from wiring cables and cooling hoses, thus reducing electromagnetic interference due to reduced cable length. An analysis was conducted regarding ways to secure the transformer, revealing no precedents for such type of mounting. Thus, designing a mounting construction becomes essential. Initially, the transformer was designed for spot-

welding applications where suspension-mounted equipment is commonplace, making the transformer's design adaptable to shocks and jolts during operation. Numerous fastening holes provide flexibility when choosing the attachment configuration.

III.2. DEVELOPMENT OF MOUNTING CONCEPT

To implement the above requirements, a mounting concept (Figure 3) has been developed using the Kompas-3D v.23 software package. This concept involves installing the power source on a carriage (Figure 4), which itself is mounted on guide rails providing single-degree-of-freedom movement (forward and backward) of the equipment.

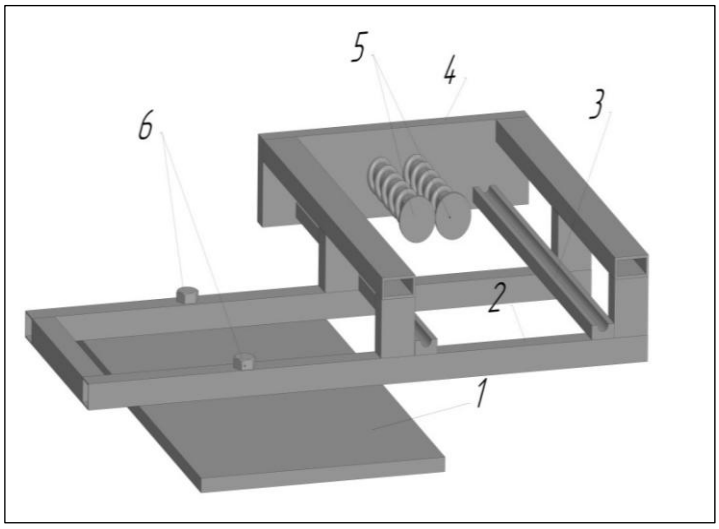


Figure 3: 3D model of adapter:
1 – Lathe saddle; 2 – Mounting adapter; 3 – Guide skids;
4 – Limit stopper; 5 – Spring damper; 6 – Fastening bolts
Source: Authors, (2025).

In order to avoid blocking the adjustment mechanisms of the lathe table-saddle stroke, it is necessary to shift the direction away from the center and perform appropriate strength calculations.

Furthermore, it has been suggested to replace some profile tubes with solid metallic elements of the same cross-section. Consequently, there arose a necessity to enhance the rigidity of the support node in order to adapt it for clinching and smoothing operations.

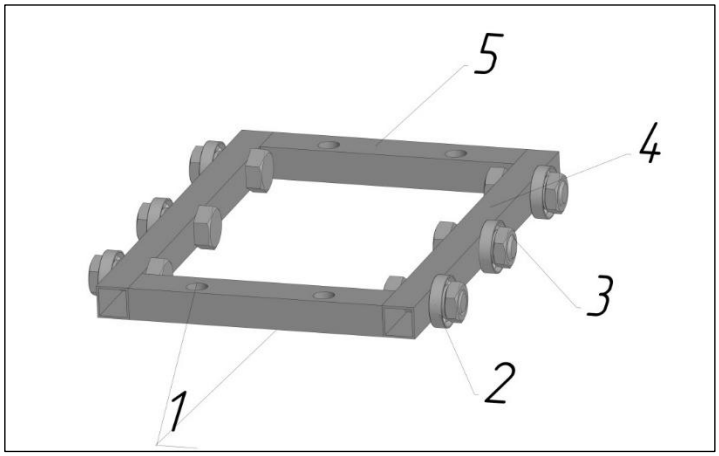


Figure 4: 3D model of carriage:
1 – Mounting holes; 2 – Bearings; 3 – Shaft; 4 – Support beam; 5 – Thrust beam (for springs of the main frame),
Source: Authors, (2025).

III.3. STRENGTH CALCULATIONS OF THE MAIN STRUCTURAL ELEMENTS

In analyzing the designed structure, certain difficulties arise due to the large number of components involved. Therefore, it was decided to divide this construction into separate components, replacing inter-component interaction connections with equivalent forces, followed by calculating each element individually.

Allowable loads on a specific area of known cross-section can be obtained using appropriate reference data.

During the design of the skids, it has been established that because of the large dimensions of equipment and workpiece, part of the skids will extend beyond the boundaries of the fastening table of the machine tool saddle (model 1K62). Consequently, it is necessary to select the section size and material for the skids which would meet strength requirements. It is essential to prevent deflection of the skids

under gravitational force. Based on these considerations, it was decided to use carbon structural steel grade 20 with a square cross-section measuring 20x20 mm.

Since part of the skids are rigidly fixed to the saddle table, any load applied at any point along them will be compensated by the reaction force from the support, limiting the strength of the product in this segment only by the strength of the material itself. The total calculated mass of the equipment amounts to 30 kg (~300 N).

The load on the skids appears as distributed since the carriage distributes the weight of the equipment across all six supports (see Figure 4). However, calculations should proceed based on worst-case loading conditions—when the entire weight of the equipment falls onto the portion of the skids located outside the support. In reality, however, cartridge movement will be limited by springs, resulting in lower actual loads on the skids.

According to the scheme for solving statics problems, we determine that finding unknown reactions requires considering equilibrium of the beam. At point A (Figure 5), there is an immovably rigid connection imposed on the beam, given that this part lies on the support and is bolted together. Thus, it becomes possible to free the beam by substituting the actions of the connection with reactions (H_A, R_A, M_A).

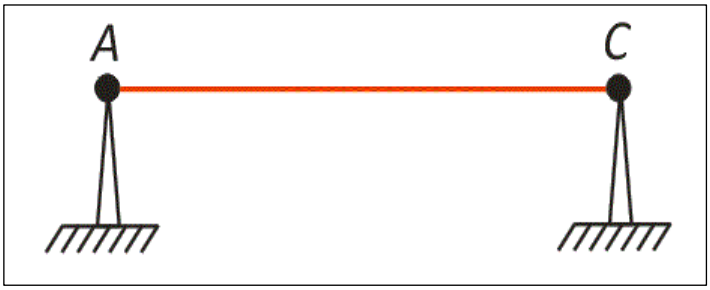


Figure 5: Horizontal weightless beam.
Source: Authors, (2025).

Support reactions are determined according to the beam equilibrium equation:

$$\sum F_x = 0, \sum F_y = 0, \sum M_A = 0.$$
$$\sum \mathrm{F_x} = 0: H_A = 0;$$
$$\sum F_X = 0 : R_A - q \times 160 = 0;$$
$$\sum M_A = 0 : M_A - q \times 160 \times (160 \div 2) = 0$$

Solving the obtained system of equations, we find the unknowns:

$$H_A = 0\ N;$$
$$R_A = q \pm 160 = 1{,}87 \times 160 = 300\ N;$$
$$M_A = q \times 160 \times (160 \div 2) = 1{,}87 \times 160 \times$$
$$\times (160 \div 2) = 23936\ N \times mm;$$

By formulating an additional moment equation relative to the free end of the beam, one can verify the performed calculations:

$$-160 \times R_A + M_A + q \times 160 \times (160 \div 2) = -160 \times 300 +$$
$$+ 23936 + 1{,}87 \times 160 \times (160 \div 2) = 0;$$

The transverse force is determined based on the following equation:

$$Q(X_1) = +R_A - q \times (X_1 - 0);$$

The value of Q at the edges of the section:

$$Q_1(0) = +300 - 1{,}87 \times (0 - 0) = 300\ N;$$
$$Q_1(160) = +300 - 1{,}87 \times (160 - 0) = 0\ N;$$

The bending moment is determined by the formula:

$$M(X_1) = +R_A(X_1) - M_A - Q_1 \times \frac{(X_1)^2}{2};$$

Thus, the values of the bending moments arising in the cross sections of the beam at the edges of the section are equal to:

$$M_1(0) = +300 \times 0 - 23936 - 1{,}87 \times \frac{(0-0)^2}{2} = -23936\ N \times mm;$$

$$M_1(160) = +300 \times (160) - 23936 - 1{,}87 \times \frac{(160-0)^2}{2} = 0\ N \times mm;$$

Based on the conducted calculations, it is possible to construct the stress diagram of the guideway (Figure 6).

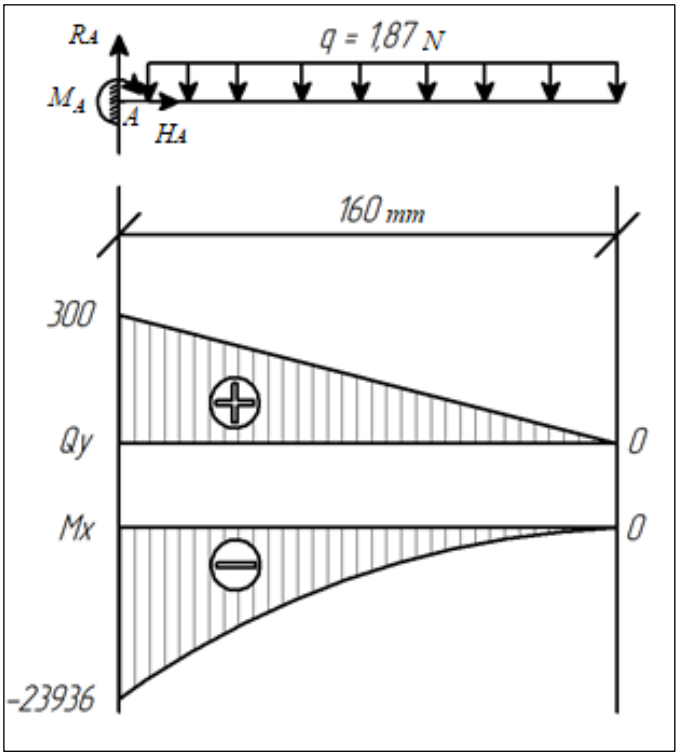


Figure 6: Load diagrams of guideway.
Source: Authors, (2025).

A rectangular beam cross-section is selected based on strength criteria with an allowable stress of 170 MPa, according to reference data.

$$\sigma = \frac{M_{MAX}}{W_X} \leq [\sigma],$$

where: σ – normal stresses, MPa; M_{MAX} – maximum absolute value of bending moment, determined from the M_X diagram, N·mm; W_X – section modulus, cm³; [σ] – permissible value of normal stress, MPa.

Thus, we obtain:

$$W_X = \frac{M_{MAX}}{[\sigma]} = \frac{23936}{170} = 140{,}8\ mm^3 = 0{,}141\ cm^3;$$

Since the width of the profile is known and equals 20 mm, the height of the profile can be determined using the formula:

$$h = \sqrt{\frac{6 \times W}{b}} = \sqrt{\frac{6 \times 0{,}141}{2}} = 0{,}649\ cm = 6{,}5\ mm$$

It should be noted that this value represents the minimum required to ensure the strength of the calculated beam, taking into account that there are two guide rails in the attachment structure. As a result, the weight of the equipment is evenly distributed, and also, considering that the used cross-section is much larger than the minimum calculated one, it can be concluded that the used structure meets the strength requirements.

Since the mounting adapter 2 (see Fig. 3) has an identical loading pattern to the guiding runners 3, a steel profile with a square cross-section of 20×20 mm will be adopted for its manufacture. This safety factor is necessary to enable the adapter to handle higher loads, particularly when attaching additional equipment or developing new operating modes.

Additionally, it is necessary to perform strength calculations for the shaft installed in the hole in the profile base of the carriage. Fastening occurs via nuts threaded on both sides, while a bearing is attached to the free end of the shaft. This arrangement allows the carriage to move along the guiding runners, with a total of six such axes provided on the carriage, three on each side. To achieve uniform distribution of the equipment's weight, the bearings must be mounted at regular intervals from each other and at the same height relative to the centerline of the profile in which they are secured, symmetrically positioned relative to the opposite side.

In this case, it is necessary to carry out a strength calculation for the axis. The load applied to it will cause bending, but this will be partially counteracted by the profile tube into which part of the axis is embedded. Despite these circumstances, let us calculate for the worst-case loading scenario for the axis: rigid fixation of one end and application of load to the other end, excluding details that hinder deformation of the axis from the loading scheme.

Shaft material: Steel 20, thread M12, length 28 mm.

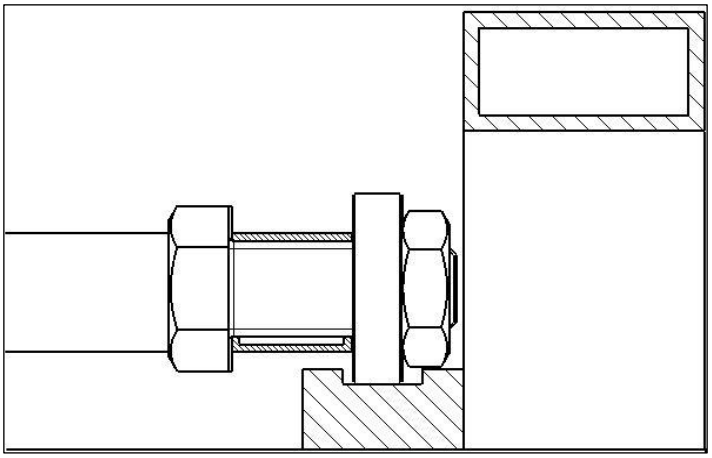


Figure 7: Bearing Mounting to Carriage.
Source: Authors, (2025).

Determination of support reactions according to the beam equilibrium equations:

$$\sum F_x = 0, \sum F_y = 0, \sum M_A = 0.$$
$$\sum F_x = 0: H_A = 0;$$
$$\sum F_X = 0 : R_A - P_1 = 0;$$
$$\sum M_A = 0 : M_A - 28 \times P_1 = 0;$$

Solving the obtained system of equations, we find the unknowns:

$$H_A = 0\ N;$$
$$R_A = P_1 = 50\ N;$$
$$M_A = 55 \times P_1 = 28 \times 50 = 1400\ N \times mm;$$

Let's check by forming an additional moment equation about the free end of the beam:

$$-28 \times R_A + M_A + 0 \times P_1 = -28 \times 50 + 1400 + 0 \times 50 = 0;$$

The stress diagram of the carriage shaft is shown in Figure 8.

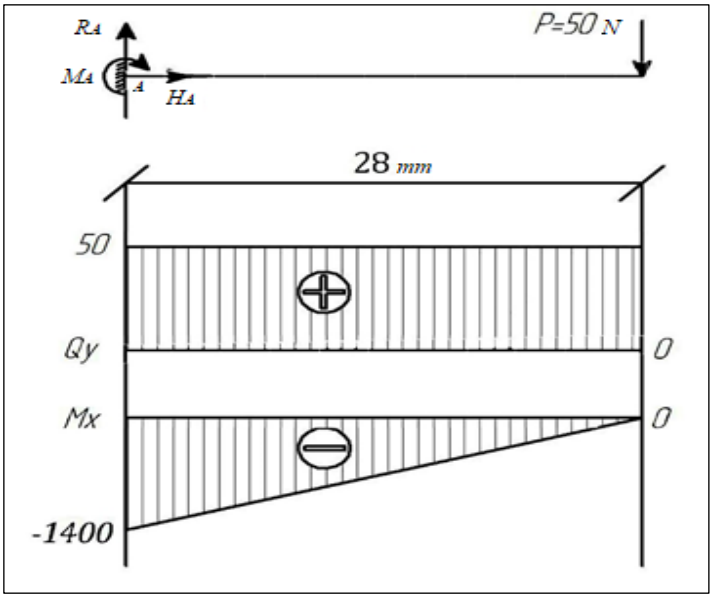


Figure 8: Stress diagrams of carriage shaft.
Source: Authors, (2025).

Transverse force Q:

$$Q(X_1) = +R_A;$$

Values of Q at the ends of the section:

$$Q_1(0) = +50 = 50\ N$$
$$Q_2(28) = +50 = 50\ N$$

Bending moment M:

$$M(X_1) = +R_A \times (X_1) - M_A;$$

Values of M at the ends of the section:

$$M_1(0) = +50 \times (0) - 1400 = 1400\ N \times mm;$$

$$M_2(28) = +50 \times (28) - 1400 = 0\ N \times mm;$$

Circular cross-section of the shaft is chosen based on strength conditions with an allowable stress $[\sigma] = 170\ MPa$.

Resistance moment of circular cross-section is defined by the formula:

$$W_X = \frac{\pi \times d^3}{32}$$

from where: $d = \sqrt[3]{\frac{32 \times W}{\pi} x}$

Required resistance moment:

$$W_X^{TP} = \frac{M_{MAX}}{[\sigma]} = \frac{1400}{170} = 0{,}008\ cm^3;$$

Thus, the diameter of the cross-section is determined as follows:

$$d = \sqrt[3]{\frac{32 \times W_X}{\pi}} = \sqrt[3]{\frac{32 \times 0{,}008}{3{,}14}} = 0{,}435\ cm = 4{,}4\ mm$$

This shaft diameter is minimally acceptable to provide sufficient strength. Considering the safety margin, it is reasonable to adopt a shaft diameter of 8 mm.

The stopper (see Fig. 3, position 4) serves to limit the travel of the carriage and install two springs on rods, which is necessary to create constant pressure on the carriage. The carriage, in turn, transmits pressure to the contact tool fixed on the transformer. Pressure adjustment is done through the traverse table feed lever. The total length of the springs in their uncompressed state should not exceed 100 mm, and the pressure exerted on the limiting stopper should not exceed 30 kg (or 300 N).

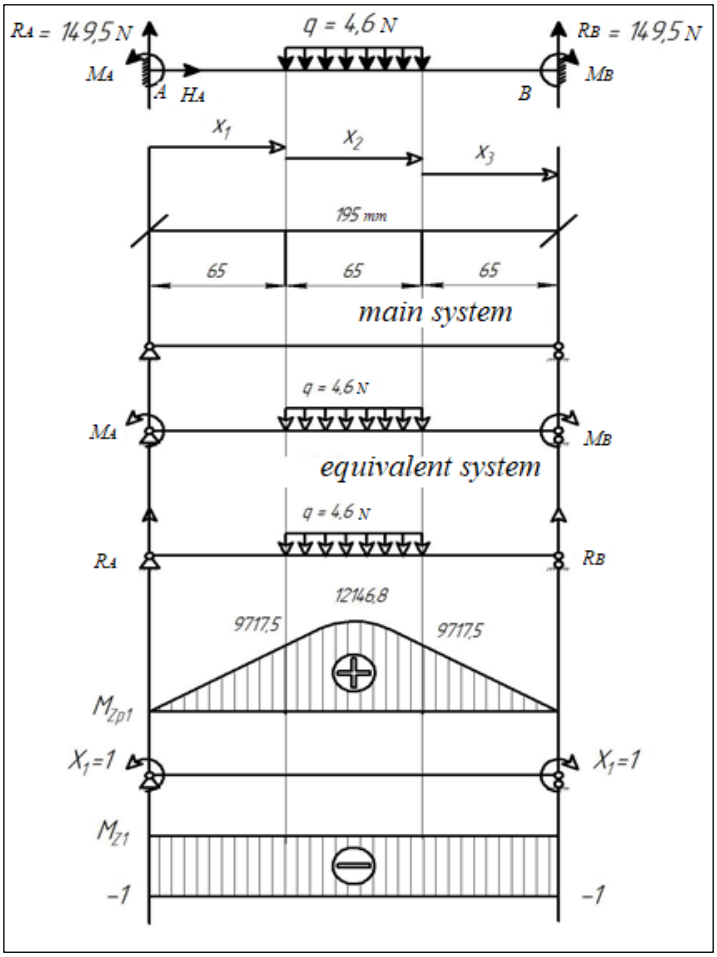


Figure 9: Main and equivalent loading systems of limit stop.
Source: Authors, (2025).

The system is statically indeterminate because the structure and load are symmetrical, so the force factors will also be symmetrical: $R_A=R_B$, $M_A=M_B$. To resolve the static indeterminacy, it is necessary to choose a primary system and compose the sum of projections of all forces onto the vertical axis:

$$\sum y = 0;$$
$$R_A + R_B - q = 0;$$
$$Q = 4{,}6\ N/mm \times 65 = 300\ \ N;$$
$$R_A = R_B = \frac{Q}{2} = 150\ \ N.$$

Then the equivalent system is: $\delta_{11} \times X_1 + \Delta_{1P} = 0$.

Constructing the canonical deformation equation, we build bending moment diagrams separately for external load q and separately for unit moments X_1=1.

From external loading, section 1: $(0 \le X \le 65)$:

$$Q_y = R_A = 149{,}5\ N;$$
$$M_Z = R_A \times X_1 = 149{,}5\ N;$$
$$for\ X_1 = 0; \quad M_Z = 0;$$
$$for\ X_1 = 65; \quad M_Z = 149{,}5 \times 65 = 9717{,}5\ N;$$

Section 2: $(0 \le X_2 \le 65)$:

$$Q_y = R_A - qX_2 = 149{,}5 \times X_2;$$
$$M_Z = R_A \times (64 + X_2) - \frac{qX_2^2}{2} = 149{,}5 \times X_2 - 2{,}3X_2^2 +$$
$$+ 9717{,}5\ \ N$$
$$for\ X_2 = 0;$$
$$Q_y = 149{,}5\ N;$$
$$M_Z = 9717{,}5\ N \times mm;$$
$$for\ X_2 = 32{,}5;$$
$$Q_y = 149{,}5 - 4{,}6 \times 32{,}5 = 0;$$
$$M_Z = 149{,}5 \times 32{,}5 - 2{,}3 \times 32{,}5^2 + 9717{,}5 = 12146{,}87\ N \times mm$$
$$for\ X_2 = 65;$$
$$Q_y = 149{,}5 - 4{,}6 \times 65 = -149{,}5\ N;$$
$$M_z = 149{,}5 \times 65 - 2{,}3 \times 65^2 + 9717{,}5 = 9717{,}5\ N \times mm.$$

Section 3: $(0 \le X_3 \le 65)$:

$$Q_y = -R_B = -149{,}5\ N;$$
$$M_Z = R_B \times X_3 = 149{,}5\ N;$$
$$for\ X_3 = 0; \quad M_Z = 0;$$
$$for\ X_3 = 65; \quad M_Z = 149{,}5 \times 65 = 9717{,}5\ N \times mm.$$

Canonical equation coefficients:

$$\delta_{11} = \frac{1}{EJ}\left[\int_0^{195}(-1)^2 dx\right] = \frac{195}{EJ};$$
$$\Delta_{1P} = \frac{1}{EJ}\left[2\int_0^{65}(149{,}5 \times X_1)(-1)dx_1 +\right.$$
$$\left. + \int_0^{65}(-2{,}3 \times X_2^2) + 149{,}5 \times X_2 + 9717{,}5)(-1)dx_2\right] =$$
$$= \frac{1}{EJ}\left[-149.5 \times 2 \times \frac{65^2}{2} + 2{,}3 \times\right.$$
$$\left. \times \frac{65^3}{3} - 149{,}5 \times \frac{65^2}{2} - 9717{,}5 \times 65\right] = \frac{-1368547{,}917}{EJ}$$

From the canonical deformation equation:

$$X_1 = -\frac{\Delta_{1P}}{\delta_{11}} = -\frac{-1368547{,}917}{195} = 7018{,}19\ N \times mm;$$

Thus, the system became statically determinate:

$$M_A = M_B = 7018{,}19\ N \times mm.$$

Final shear force (QY) and bending moment (MZ) diagrams in the equivalent system are presented in Figure 10.

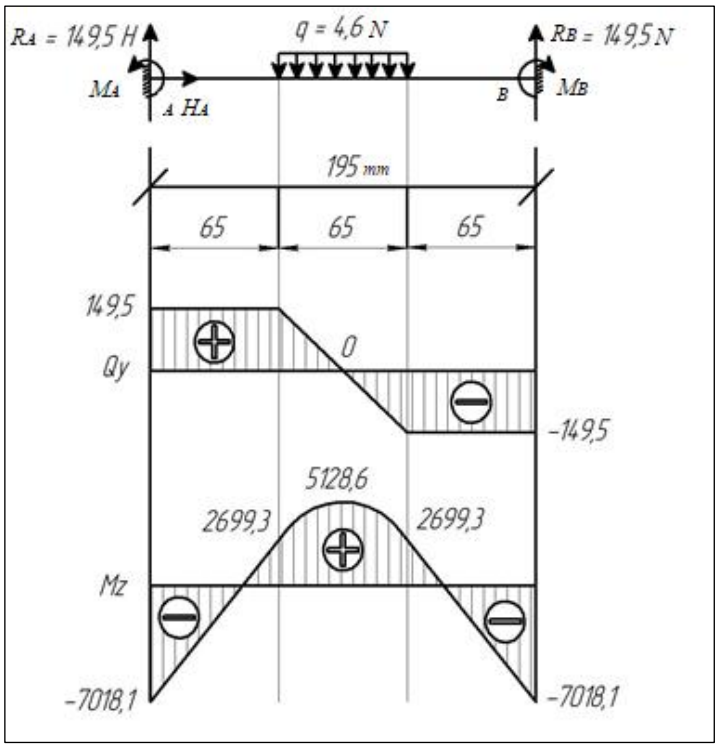


Figure 10: Final loading diagrams of limit stop.
Source: Authors, (2025).

Section 1: $(0 \le X \le 65)$:

$$M_Z = -M_A + R_A \times z_1 = -7018{,}19 + 149{,}5 \times z_1;$$
$$for\ z_1 = 0;$$
$$M_Z = -7018{,}19\ N \times mm;$$
$$for\ z_1 = 65;$$
$$M_Z = -7018{,}19 + 149{,}5 \times 65 = 2699{,}31\ N \times mm.$$

Section 2: $(0 \le X_2 \le 65)$:

$$M_Z = -M_A + R_A \times (65 + X_2) - \frac{qX_2^2}{2} =$$
$$= 2699{,}31 + 149{,}5 \times X_2 - 2{,}3X_2^2;$$
$$for\ X_2 = 0;$$
$$M_Z = 2699{,}31\ N \times mm;$$
$$for\ X_2 = 32{,}5;$$
$$M_Z = 2699{,}31 + 149{,}5 \times 32{,}5 - 2{,}3 \times 32{,}5^2 =$$
$$= 5128{,}69\ N \times mm.$$
$$for\ X_2 = 65;$$
$$M_Z = 2699{,}3 + 149{,}5 \times 65 -$$
$$-2{,}3 \times 65 = 2699{,}3\ N \times mm.$$

Section 3 is symmetrical to Section 1.
From the bending moment diagram M, it is determined that M_{MAX}=7018.1 $N{\cdot}mm$.

From the shear force diagram Q, Q_{MAX}=149.5 H.

Strength condition for bending based on normal stresses:

$$\sigma_{MAX} = \frac{M_{MAX}}{W_X} \leq [\sigma];$$

$$W_X = \frac{M_{MAX}}{[\sigma]} = \frac{7018{,}1}{160} = 0{,}044\ cm^3;$$

It follows that:

$$h = \sqrt{\frac{6 \times W}{b}} = \sqrt{\frac{6 \times 0{,}044}{2}} = 0{,}36\ cm = 3{,}6\ mm.$$

The minimum solid cross-section thickness required to ensure the structural strength is 20 x 3.5 mm. The project cross-section of 15 x 20 mm provides multiple times the strength reserve.

The tool depicted in Figure 11 is intended directly for contacting the surface of the processed part, specifically the roller axis of the contact tool.

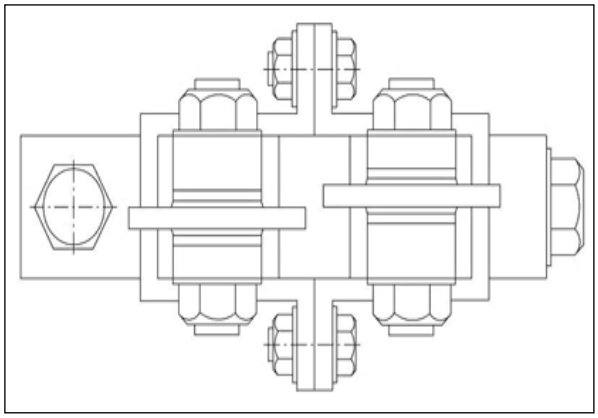


Figure 11: Contact Tool Scheme.
Source: Authors, (2025).

An electric current passes through this tool, heating the part as it travels through it, thereby implementing the process of electromechanical machining.

Material of the shaft: Steel 45, Length: 20 mm. It is necessary to calculate the load and select the cross-section.

According to the solution scheme for static tasks, to find unknown reactions, it is necessary to consider the equilibrium of the beam.

$$\sum F_X = 0: \quad H_A = 0;$$

$$\sum M_A = 0: \quad -q_1 \times 12 \times \left(\frac{12}{2}\right) + R_B \times 12 = 0;$$

$$\sum M_B = 0: \quad -R_a \times 12 + q_1 \times 12 \times \left(12 - \frac{12}{2}\right) = 0$$

Solving the obtained system of equations, we get: $H_A = 0\ N$.

Reaction of the hinged-moving support at point B:

$$R_B = \frac{\left(q_1 \times 12 \times \left(\frac{12}{2}\right)\right)}{12} = \frac{\left(25 \times 12 \times \left(\frac{12}{2}\right)\right)}{12} = 150\ N.$$

Reaction of the hinged-moving support at point A:

$$R_A = \frac{\left(q_1 \times 12 \times \left(12 - \frac{12}{2}\right)\right)}{12} = \frac{\left(25 \times 12 \times \left(12 - \frac{12}{2}\right)\right)}{12} = 150\ N.$$

We will perform verification:

$$\sum F_y = 0: \quad R_A - q_1 \times 12 + R_B = 150 - 25 \times 12 + 150 = 0.$$

Shear Force Q:

$$Q(X_1) = +R_a - q_1 \times (X_1 - 0);$$

Value of Q at the edges of the section:

$$Q_1(0) = +150 - 25 \times (0 - 0) = 150\ N;$$
$$Q_1(12) = +150 - 25 \times (12 - 0) = -150\ N.$$

Bending moment M:

$$M(X_1) = +R_a \times (X_1) - q_1 \times \frac{(X_1)^2}{2}$$

Values of M at the edges of the section:

$$M_1(0) = +150 \times (0) - 25 \times \frac{(0-0)^2}{2} = 0\ N \times mm;$$
$$M_1(12) = +150 \times (12) - 25 \times \frac{(12-0)^2}{2} = 0\ N \times mm.$$

Local extremum at point X=6:

$$M_1(6) = +150 \times (6) - 25 \times \frac{(6-0)^2}{2} = 450\ N \times mm.$$

Round cross-section is selected based on strength conditions with allowable stress:

$$[\sigma] = 170\ MPa;$$

Resistance moment of round cross-section is determined by the formula:

$$W_X = \frac{\pi \times d^3}{32},$$

it follows that:

$$d = \sqrt[3]{\frac{32 \times W}{\pi}} x.$$

The required resistance moment is determined by the formula:

$$W_X^{TP} = \frac{M_{MAX}}{[\sigma]} = \frac{450}{170} = 0{,}002\ cm^3.$$

Thus, the diameter of the cross-section is:

$$d = \sqrt[3]{\frac{32 \times W_X}{\pi}} = \sqrt[3]{\frac{32 \times 0{,}002}{3{,}14}} = 0{,}273\ cm = 2{,}7\ mm.$$

Loading diagrams of the roller in the contact tool are shown in Figure 12.

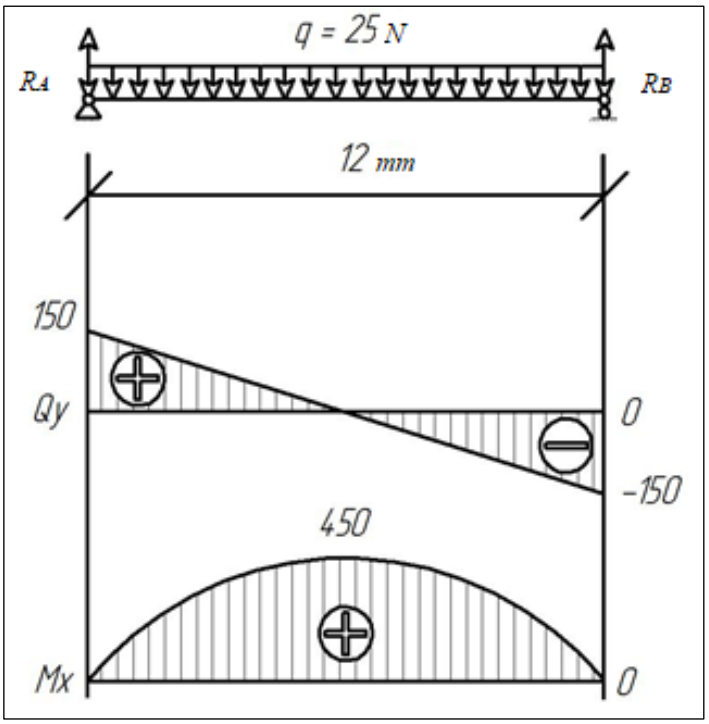


Figure 12: Roller loading diagrams in contact tool.
Source: Authors, (2025).

This shaft diameter is minimally acceptable to ensure sufficient strength. Taking into account the safety margin, we accept the shaft diameter of 10 mm.

IV. CONCLUSIONS

As a result of the research carried out, the technology of electro-mechanical treatment and its application area in strengthening and repairing agricultural machinery parts have been examined. It has been found that due to the shortage of spare parts for russian and foreign agricultural machinery, it is necessary to transition from modular replacement of parts and units to restorative and hardening repairs.

An analysis of existing installations for electro-mechanical processing has been conducted, noting their advantages and disadvantages. It has been revealed that a decrease in electrical current leads to uneven heating of the surface layer of parts, negatively affecting their strength characteristics.

Measurements taken during electro-mechanical processing showed a voltage drop in the power supply circuit from the current source to the tool-part contact, amounting to 27–36% of the initial level, with up to 50% of losses occurring in the current-carrying busbar. To reduce losses and improve technological efficiency, a conceptual schematic for electro-mechanical processing has been proposed, involving the use of a minimal number of elements in the power supply circuit.

Special equipment has been developed to implement the proposed electro-mechanical processing scheme, and strength calculations have been made for its major components. The proposed compact fixture, allowing for fewer electrical wires, reduces energy loss, improves occupational safety conditions, and enhances the effectiveness of the electro-mechanical processing technology for agricultural machinery parts.

V. AUTHOR'S CONTRIBUTION

Conceptualization: E. Veremey, D. Podashev.
Methodology: E. Veremey, D. Podashev.
Investigation: E. Veremey.
Discussion of results: E. Veremey, D. Podashev.
Writing: D. Podashev.
Review and Editing: E. Veremey, D. Podashev.
Resources: E. Veremey, D. Podashev.
Supervision: E. Veremey, D. Podashev.
Approval of the final text: E. Veremey, D. Podashev.

VI. REFERENCES

[1] N.I. Konyaeva, N.V. Konyaev T, "Features of Agricultural Machinery Repair in Modern Conditions", Modern Materials, Equipment and Technologies, no. 6 (45), pp. 125-130, 2022 (In Russian).

[2] A.V. Makarova, S.A. Grashkov, D.I. Es'kov, "The State of Domestic Agricultural Engineering and Ways to Increase the Efficiency of Agricultural Business", In the conference paper: Current Problems and Development Directions of Agroengineering in Russia, collection of scientific articles of the International Scientific and Technical Conference, Kursk, pp. 59-63, 2021 (In Russian).

[3] I.O. Ishkov, N.V. Konyaev, B.S. Blinkov, "Methods for Restoring Worn Parts", In the conference paper: Future of Science – 2022, Collection of Scientific Articles of the 10th International Youth Scientific Conference, Kursk, pp. 467-469, 2022 (In Russian).

[4] S.I. Zhdanov, N.V. Konyaev, "New Developments in Technology for Restoration and Hardening of Components", In the conference paper: Topical Issues of Innovative Development of Agroindustrial Complex, Proceedings of the International Scientific-Practical Conference, Publication Manager I.Ya. Pigorev, pp. 190-192, 2016 (In Russian).

[5] Yu.G. Alekhin, S.A. Grashkov, A.S. Ugrimov, "Quality of Welded Coatings on Plough Shares", In the conference paper: Quality in Production and Socio-Economic Systems, Collection of Scientific Papers of the 5th International Scientific and Technical Conference, Responsible Editor E.V. Pavlov, pp. 11-13, 2017 (In Russian).

[6] N.A. Pivovar, S.A. Grashkov, "Recovery of worn machine parts by plasma spraying of powders", In the conference paper: Actual problems of engineering and technical support of agro-industrial complex, Proceedings of the International Scientific and Practical Conference, pp. 22-29, 2013 (In Russian).

[7] E.V. Sazonov, M.E. Kulikov, S.A. Grashkov, "Restoration of Wear Resistance of Deep Loosener Working Elements", In the conference paper: Electric Power Today and Tomorrow, Collection of Scientific Articles of the International Scientific and Technical Conference, Kursk, pp. 147-153, 2022 (In Russian).

[8] S. Markauskas, "Investigation of surface electromechanical restoration process and it's quality", Transport Means, Proceedings of the International Conference, Kaunas, pp. 157-160, 2013.

[9] A. Shohiyon, Z. Obidov, "Restoration of the external surfaces of machine mechanisms by electromechanical treatment", Universum: technical sciences, no. 1-4(106), pp. 43-46, 2023.

[10] S.A. Yakovlev, M.M. Zamal'dinov, Y.V. Nuretdinova [et al.], "Electromechanical Hardening of VT22 Titanium Alloy in Screw-Cutting Lathes", Russian Engineering Research, no. 38, pp. 488-490, 2018.

[11] L. M. Jiang, W. Li, A. Attia [et al.], "A potential method for electrochemical micromachining of titanium alloy Ti6Al4V", Journal of Applied Electrochemistry, vol. 38, no. 6, pp. 785-791, 2008.

[12] N. Levov, A. Uhalin, "Innovations in restoration technology of automotive case details using adhesive method", Perspectives of Innovations, Economics and Business, vol. 4, no. 1, pp. 113-117, 2010.

[13] A.P. Chernysh, "Working out of information models of technological repair blocks for restoration of agricultural machine details", Traktori i Pogonske Mašine, vol. 17, no. 1, pp. 60-62, 2012.

[14] S. Y. Zhachkin, S. N. Sharifullin, N. A. Penkov [et al.], “About Modeling Kinematic Parameters of Deposition from Composite Electrochemical Coatings for Restoration of Details for Agricultural Machines”, IOP Conference Series: Materials Science and Engineering, vol. 570, p. 012104, 2019.

[15] V. Aulin, S. Lysenko, A. Hrinkiv [et al.], “Creation of theoretical bases of tribotechnologies of running-in and restoration as means of effective increase of operational wear resistance of motor transport and mobile agricultural machinery”, Problems of Tribology, vol. 99, no. 1, pp. 51-58, 2021.

[16] M. Denysenko, “Restoration and Strengthening of Parts and Tool of Agricultural Machinery Operating in an Abrasive Environment”, National Interagency Scientific and Technical Collection of Works. Design, Production and Exploitation of Agricultural Machines, no. 53, pp. 271-284, 2023.

[17] N. Wu, Q. Wang, “Repair of vibrating delaminated beam structures using piezoelectric patches”, Smart Materials and Structures, vol. 19, no. 3, p. 035027, 2010.

[18] P.A. Boykov, A.S. Bezruk, Y.I. Melnikov [et al.], “Repair and restoration of critical components of timber harvesting machines using electromechanical processing”, Systems. Methods. Technologies, no. 4(60), pp. 38-42, 2023 (In Russian).

[19] R. Bini, M. Monno, M. I. Boulos, “Effect of Cathode Nozzle Geometry and Process Parameters on the Energy Distribution for an Argon Transferred Arc”, Plasma Chemistry and Plasma Processing, vol. 27, no. 4, pp. 359-380, 2007.

[20] R. Hamano, H.R. Costa, T. Chuvas, “Evaluation of Machining Forces and Surface Integrity on AISI 304 Steel Top Milling Process under Different Cutting Conditions”, Itegam-Jetia, no. 5(20), pp. 166-171, 2019.

[21] S. Doshi, K. Vaghosi, N. Mehta, “Progress in the welding of AL alloy thin sheet and future prospectus for automobile”, Itegam-Jetia, no. 10(46), pp. 42-49, 2024.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11n.55, p. 73-85. September/October, 2025.
DOI: https://doi.org/10.5935/jetia.v11i55.1892

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

PARAMETERS ESTIMATION OF THREE-PHASE INDUCTION MOTOR USING THE PARAMETER ESTIMATOR APP

Fares Bettahar[1], Sabrina Abdeddaim[2], Omar Charrouf[3], Achour Betka[4] Michael Short[5],Laid Guerrida[6] and Belahcene taha lemdjed[7]

[1, 2, 3, 4]Electrical Engineering Department, LGEB Laboratory, University of Biskra, Algeria.
[5]School of Computing Engineering and Digital Technologies ,Teesside University.
[6]Applied Automation and Industrial Diagnostics Laboratory ,University of Djelfa, Algeria.
[7]Laboratory of Identification, Control, Command, and Communication.,University of Biskra, Algeria

[1]https://orcid.org/0009-0005-0520-092X.[2]https://orcid.org/0009-0006-0812-1554.[3]http://orcid.org/0000-0003-4254-464X
[4]https://orcid.org/0000-0002-7116-4649.[5]https://orcid.org/0000-0001-6290-4396.[6]https://orcid.org/0009-0003-0782-3798
[7]https://orcid.org/0009-0006-9390-937X

Email: fares.bettahar@univ-biskra.dz, sabrina.abdeddaim@univ-biskra.dz, omar.charrouf@univ-biskra.dz
a.betka@univ-biskra.dz, m.short@tees.ac.uk, laid.guerrida@univ-djelfa.dz , Taha.belahcene@univ-biskra.dz

ARTICLE INFO

Article History
Received: April 18, 2025
Revised: May 20, 2025
Accepted: September 15, 2025
Published: September 30, 2025

Keywords:
Induction Motor
Motor Speed
Stator Voltage
Stator Current
Parameter Estimator App

ABSTRACT

Induction motors are among the most widely used electric motors worldwide, valued for their efficiency, robustness, and low maintenance. However, optimizing their performance requires accurate knowledge of electrical and mechanical parameters, which are nonlinear and subject to variation due to aging and operational wear. This study presents a parameter estimation technique using MATLAB's Parameter Estimator app to determine key parameters, including stator and rotor resistances, leakage and magnetizing inductances, moment of inertia, and friction coefficient. The approach relies on experimentally measured inputs such as stator voltage, current, and motor speed, which are applied to a state-space induction motor model in MATLAB. The simulation results are compared with experimental data, specifically motor speed, d-axis current (ids) and q-axis current (iqs). The app iteratively refines the estimated parameters by minimizing the error between simulated and experimental responses. Validation using experimental data from a Siemens induction motor confirms the method's accuracy and reliability, providing a robust framework for precise parameter estimation and enhanced motor control.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Nikola Tesla designed the induction motor in 1892, leveraging the advantages of the alternating current (AC) supply system as an alternative to direct current (DC) motors [1]. Since then, induction motors (IMs) have become one of the most widely used electric machines in industrial applications due to their simple structure, ease of operation, low maintenance requirements, and cost-effectiveness [2]. Their robust construction and minimal upkeep make them a preferred choice for various applications, ranging from household appliances to heavy industrial machinery. Given their widespread use, understanding their performance characteristics is crucial for optimizing their operation and ensuring energy efficiency [3].

Induction motors rely on advanced control techniques such as field-oriented control (FOC) [4], direct torque control (DTC) [5], and model predictive control (MPC) [6] to achieve high dynamic performance. However, the effectiveness of these control methods is highly dependent on accurate motor parameter estimation. In FOC, rotor flux orientation is determined by the slip frequency, which is directly influenced by rotor resistance and inductance. Parameter mismatches can lead to poor decoupling of torque and flux, degrading control performance [7]. Similarly, DTC, which directly regulates torque and flux without a modulator, is highly sensitive to variations in stator resistance and inductance. Any inaccuracies can cause flux estimation errors, excessive torque ripple, and degraded steady-state

performance. Likewise, MPC, an optimization-based control strategy, relies on a predictive model to determine the optimal switching states. Parameter uncertainties in resistances, inductances, and mechanical properties can result in inaccurate state predictions, leading to reduced efficiency and compromised transient response [8]. Consequently, precise parameter estimation is essential for ensuring accurate control, minimizing losses, and enhancing the overall reliability of induction motor drives.

The impact of parameter mismatches on control performance has been widely studied in the literature [9,10], as variations in rotor resistance (Rr), stator resistance (Rs), rotor inductance (Lr), and stator inductance (Ls) can reach up to 100% of their nominal values due to temperature fluctuations, aging, and operational wear [11]. These uncertainties are particularly critical in control schemes, where inaccurate motor parameters amplify estimation errors, leading to degraded speed, torque, and flux regulation [12]. Such discrepancies not only reduce control accuracy but also impair overall system efficiency.

The adverse effects of parameter uncertainty are even more pronounced in sensor less control schemes, where the absence of direct mechanical sensors necessitates reliance on parameter-dependent estimation techniques [13]. Precise estimation of rotor position, speed, and flux hinges on accurate knowledge of motor parameters. For instance, errors in Rr or Lr directly affect back-EMF calculations in model-based observers (e.g., sliding-mode observers), leading to biased speed estimates, oscillatory torque responses, and instability [14]. Similarly, mismatches in Ls or Rs distort flux-linkage and torque-current mappings, undermining FOC decoupling and triggering cross-talk between torque and flux [15], [16].

Traditionally, motor nameplates do not provide essential parameters such as stator resistance or leakage inductance, necessitating direct measurement [17]. Common methods for parameter identification include short-circuit and no-load tests [18]. However, these approaches are often impractical due to operational constraints or the unavailability of specialized equipment [19]. Consequently, alternative parameter identification methods have been developed, including offline and online estimation techniques, adaptive algorithms, and observer-based approaches, which allow for more accurate and real-time estimation without requiring additional hardware.

The complexity of induction motors arises from their nonlinear characteristics, which, combined with uncertainties in rotor variables, often make it difficult to identify many motor parameters [20]. To address this, indirect estimation methods are employed, leveraging measurable quantities such as stator voltage, current, and rotor speed to infer these unknown parameters. By integrating these known values into a structured estimation framework, it becomes possible to iteratively refine and control additional motor parameters with greater accuracy.

In this study, a three-phase induction motor was tested using real-time data acquisition via the dSPACE DS1104 platform. Measured signals, including stator currents, rotor speed, and voltage, were utilized in a mathematical model in MATLAB/Simulink for parameter estimation. The identified parameters included stator resistance (R_s), rotor resistance (R_r), magnetizing inductance (L_m), stator leakage inductance (L_s), rotor leakage inductance (L_r), moment of inertia (J_m), friction coefficient (f), and the number of pole pairs (p). The Parameter Estimator App iteratively refined these parameters by minimizing discrepancies between simulated and experimental results. This methodology ensures precise parameter identification, which is crucial for enhancing motor performance, optimizing energy efficiency, and improving operational reliability in industrial applications.

II. THREE-PHASE INDUCTION MACHINE MODEL

Modeling an asynchronous machine relies on mathematical equations derived from its physical principles. However, as the model becomes more accurate, its complexity increases, demanding significant computational power [21]. To create a realistic and manageable model, simplifying assumptions are applied. These assumptions filter out less significant effects while preserving key dynamic and static behaviors, ensuring the model is practical and efficient for analysis and control.

To develop an effective mathematical model of the asynchronous machine, the following simplifying assumptions are made:

- The machine is symmetrical, and the air gap remains constant current density is uniformly distributed across conductor sections.
- The magnetic circuit is considered unsaturated and perfectly laminated in both the stator and rotor.
- Winding resistances are assumed to be temperature-independent, with skin and notch effects neglected.
- The magneto motive force (MMF) distribution is sinusoidal along the air gap.
- The cage rotor is modeled as an equivalent balanced three-phase winding.

Under the simplifying assumptions and for a balanced machine, the equations of the machine are written as follows:

A. Electrical Equations

The voltage equations for the three stator phases and the three rotor phases in matrix form are given by:

$$\begin{bmatrix} v_{sa} \\ v_{sb} \\ v_{sc} \\ v_{ra} \\ v_{rb} \\ v_{rc} \end{bmatrix} = \begin{bmatrix} R_s & 0 & 0 & 0 & 0 & 0 \\ 0 & R_s & 0 & 0 & 0 & 0 \\ 0 & 0 & R_s & 0 & 0 & 0 \\ 0 & 0 & 0 & R_r & 0 & 0 \\ 0 & 0 & 0 & 0 & R_r & 0 \\ 0 & 0 & 0 & 0 & 0 & R_r \end{bmatrix} \begin{bmatrix} i_{sa} \\ i_{sb} \\ i_{sc} \\ i_{ra} \\ i_{rb} \\ i_{rc} \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} \phi_{sa} \\ \phi_{sb} \\ \phi_{sc} \\ \phi_{ra} \\ \phi_{rb} \\ \phi_{rc} \end{bmatrix} \quad (1)$$

The voltage equation of a three-phase induction machine describes the relationship between the stator and rotor voltages, currents, resistances, and flux linkages. The stator phase voltages, denoted as v_{sa}, v_{sb}, v_{sc} represent the electrical potential applied to the stator windings, while the rotor phase voltages, v_{ra}, v_{rb}, v_{rc} correspond to the voltages induced in the rotor windings. These voltages are influenced by the stator and rotor resistances, represented by R_s and R_r, respectively.

The stator currents, i_{sa}, i_{sb}, i_{sc} flow through the stator windings and interact with the magnetic field to generate torque. Similarly, the rotor currents, i_{ra}, i_{rb}, i_{rc} circulate in the rotor windings and contribute to the electromechanical energy conversion process. These currents influence the stator and rotor flux linkages, which are represented by $\phi_{sa}, \phi_{sb}, \phi_{sc}$ for the stator and $\phi_{ra}, \phi_{rb}, \phi_{rc}$ for the rotor. The flux linkages play a crucial role in determining the machine's dynamic response and its ability to develop torque [22].

B. Magnetic equation

The magnetic equations describe the relationship between the stator and rotor flux linkages and the corresponding phase currents. Under the assumption of linear magnetic behavior, the flux linkages can be expressed as a function of self and mutual inductances. The stator flux linkages $\phi_{as}, \phi_{bs}, \phi_{cs}$ are related to the stator currents (i_{as}, i_{bs}, i_{cs}) and rotor currents (i_{ar}, i_{br}, i_{cr}) through the stator selfinductance matrix $[L_{ss}]$ and the mutual inductance matrix $[L_{sr}]$. Similarly, the rotor flux linkages $\phi_{ar}, \phi_{br}, \phi_{cr}$ depend on the rotor self-inductance matrix $[L_{rr}]$ and the mutual inductance matrix $[L_{sr}]$, which represents the coupling between the stator and rotor windings.

$$\begin{bmatrix}\varphi_{as}\\ \varphi_{bs}\\ \varphi_{cs}\end{bmatrix} = [L_{ss}]\begin{bmatrix}i_{as}\\ i_{bs}\\ i_{cs}\end{bmatrix} + [L_{sr}]\begin{bmatrix}i_{ar}\\ i_{br}\\ i_{cr}\end{bmatrix} \tag{2}$$

$$\begin{bmatrix}\varphi_{ar}\\ \varphi_{br}\\ \varphi_{cr}\end{bmatrix} = [L_{rr}]\begin{bmatrix}i_{ar}\\ i_{br}\\ i_{cr}\end{bmatrix} + [L_{sr}]\begin{bmatrix}i_{as}\\ i_{bs}\\ i_{cs}\end{bmatrix} \tag{3}$$

The self-inductance matrices for the stator and rotor are defined as:

$$[L_{ss}] = \begin{bmatrix}L_{as} & M_{as} & M_{as}\\ M_{as} & L_{as} & M_{as}\\ M_{as} & M_{as} & L_{as}\end{bmatrix},\ [L_{rr}] = \begin{bmatrix}L_{ar} & M_{ar} & M_{ar}\\ M_{ar} & L_{ar} & M_{ar}\\ M_{ar} & M_{ar} & L_{ar}\end{bmatrix} \tag{4}$$

where L_{as} and L_{ar} are the stator and rotor self-inductances, while M_{as} and M_{ar} are the mutual inductances between the stator and rotor windings, respectively The stator-rotor mutual inductance matrix $[L_{sr}]$ is given by:

$$[L_{sr}] = \begin{bmatrix}\cos(\theta) & \cos\left(\theta - \frac{2\pi}{3}\right) & \cos\left(\theta + \frac{2\pi}{3}\right)\\ \cos\left(\theta + \frac{2\pi}{3}\right) & \cos(\theta) & \cos\left(\theta - \frac{2\pi}{3}\right)\\ \cos\left(\theta - \frac{2\pi}{3}\right) & \cos\left(\theta + \frac{2\pi}{3}\right) & \cos(\theta)\end{bmatrix} \tag{5}$$

Finally, the three-phase asynchronous machine model is expressed as:

$$\begin{cases}[v_{sabc}] = [R_s]\cdot[i_{sabc}] + \frac{d}{dt}([L_s]\cdot[i_{sabc}] + [L_{sr}]\cdot[i_{rabc}])\\ [v_{rabc}] = [R_r]\cdot[i_{rabc}] + \frac{d}{dt}([L_r]\cdot[i_{rabc}] + [L_{rs}]\cdot[i_{sabc}])\end{cases} \tag{6}$$

where $[v_{sabc}]$ and $[v_{rabc}]$ represent the stator and rotor voltage vectors, $[R_s]$ and $[R_r]$ are the stator and rotor resistance matrices, and $[i_{sabc}]$ and $[i_{rabc}]$ are the current vectors.

C. Mechanical equation

The electromagnetic torque equation for a three-phase induction machine in the d, q rotating reference frame is given by:

$$C_e = \frac{3}{2}pL_m\left(I_{qs}I_{dr} - I_{ds}I_{qr}\right) \tag{7}$$

where: C_e is Electromagnetic torque (Nm),p is Number of pole pairs,L_m is mutual inductance between stator and rotor (H),I_{qs} and I_{ds} is stator q-axis and d-axis currents (A),I_{dr} and I_{qr} is Rotor d-axis and q-axis currents (A).

III. SIMULATION OF INDUCTION MACHINE DYNAMICS

The simulation of an induction machine begins with transforming the three-phase electrical system into a decoupled two-phase (d-q) reference frame using Park's transformation [23]. This simplifies the complex three-phase differential equations into a set of equations that separate the direct-axis (d) and quadrature-axis (q) components decoupling flux and torque dynamics. The transformation enables the use of state-space representation, which is essential for efficient numerical computation and real-time simulation of both electrical and mechanical behavior [24].

The state-space form represents a dynamic system using a set of first-order differential equations:

$$\begin{aligned}\dot{x} &= Ax + BU + EC_r\\ y &= Cx\end{aligned} \tag{8}$$

- State Vector (x)

The system states are typically chosen as:

$$x = \begin{bmatrix} I_{ds} \\ I_{qs} \\ I_{dr} \\ I_{qr} \\ \omega \end{bmatrix} \tag{9}$$

Where: ω is rotor angular speed.
The control inputs applied to the system are the stator voltages:

$$U = \begin{bmatrix} V_{ds} \\ V_{qs} \end{bmatrix} \tag{10}$$

Where V_{sd}, V_{sq} are the $d-q$ components of stator voltage.

- Output Equation ($y = Cx$)

The measured system outputs are typically the stator currents:

$$y = \begin{bmatrix} I_{ds} \\ I_{qs} \end{bmatrix} \tag{11}$$

Where C is the output matrix:

$$C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix} \tag{12}$$

- Rotor Speed Dynamics

The rotor dynamics are governed by:

$$\frac{d\omega}{dt} = \frac{C_e P}{f} - \frac{C_r P}{f} - \frac{f}{J}\omega \tag{13}$$

Where C_r is Load torque,P is number of pole pairs,f is friction coefficient and J is rotor inertia. Substituting Eq. (7) into Eq. (13), the speed equation becomes:

$$\frac{d\omega}{dt} = \sigma P J I_{qs} I_{dr} - \sigma P J I_{ds} I_{qr} - C_r \frac{P}{J} b\omega \tag{14}$$

Where σ gave as $\sigma = \frac{3}{2} P L_m$

- System Matrices

The state matrix A describes how the system states evolve:

$$A = \begin{bmatrix} -\frac{1}{T_s \sigma} & \omega_s + (1/\sigma - 1)\omega & -\frac{M}{L_s T_r \sigma} & -\frac{M}{L_s \sigma}\omega & 0 \\ -\left(\omega_s + \frac{(1-\sigma)\omega}{\sigma}\right) & -\frac{1}{T_s \sigma} & -\frac{M}{L_s \sigma}\omega & -\frac{M}{L_s T_r \sigma} & 0 \\ \frac{M}{L_r T_s \sigma} & -\frac{M}{L_r \sigma}\omega & -\frac{1}{T_r \sigma} & \omega_s - \left(1 - \frac{M^2}{L_s L_r \sigma}\right)\omega & 0 \\ \frac{M}{L_r \sigma}\omega & \frac{M}{L_r T_s \sigma} & -\omega_s + \left(1 - \frac{M^2}{L_s L_r \sigma}\right)\omega & -\frac{1}{T_r \sigma} & 0 \\ 0 & \frac{\sigma p}{J} I_{dr} & 0 & -\frac{\sigma p}{J} I_{ds} & -\frac{f}{J} \end{bmatrix} \tag{15}$$

The input matrix B defines how the control inputs, specifically the stator voltages V_{ds} and V_{qs}, influence the state variables, particularly the stator currents I_{ds} and I_{qs}, while the rotor flux components and rotor speed remain indirectly affected through system dynamics. It is given by:

$$B = \begin{bmatrix} \frac{1}{L_s} & 0 \\ 0 & \frac{1}{L_s} \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{16}$$

Similarly, the disturbance matrix E accounts for external disturbances such as the load torque C_r, primarily affecting the rotor speed dynamics, and is expressed as:

$$E = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \frac{P}{J} \end{bmatrix} \tag{17}$$

The block diagram in Figure 1 illustrates the comprehensive simulation framework for an induction machine, highlighting the integration of various components. The simulation begins with time-based inputs for stator voltages Vds and Vqs, which are fed into the system as experimental data. Additionally, the synchronous speed (ws) is calculated using 2×p×π/50, where p is the number of pole pairs. The load torque (Cr) is initially set to zero but can be adjusted to simulate different loading conditions. These inputs are processed through the induction motor model, which calculates the corresponding currents Ids and Iqs, as well as the rotor speed w.

The middle section of the block diagram focuses on the state-spac representation of the induction machine dynamics. The state-space matrices A(x), B, and E describe the evolution of the system states over time. The state vector x includes variables such as stator currents, rotor fluxes, and rotor speed, which are updated at each time step based on the input voltages and load torque. The derivative of the state vector (dx/dt) is calculated using the state matrix A(x) and the input matrix B. This derivative is then integrated to update the state vector for the next time step. The calculated stator currents (Ids, Iqs), rotor fluxes (ψrd, ψrq), and rotor speed (w) are outputted from this section.

In the bottom section, the state matrix A(x) is decomposed into five sub-matrices (A1, A2, A3, A4, A5) to account for dependencies on angular speeds (ωs, ω), rotor flux (φrd), and stator current (Ids). Each sub-matrix captures specific aspects of the induction machine dynamics. For example, A1 captures basic resistive and inductive effects, A2 accounts for the influence of synchronous speed (ωs), A3 considers the impact of rotor speed (ω), and A4 and A5 incorporate the effects of rotor flux (φrd) and stator current (Ids), respectively. contributions from all sub-matrices are combined to form the complete state matrix A(x), which is used to update the state vector. The matrix A defined as follows:

$$\mathbf{A} = \boldsymbol{A_1} + \boldsymbol{A_2\omega_s} + \boldsymbol{A_3\omega} + \boldsymbol{A_4 I_{dr}} + \boldsymbol{A_5 I_{ds}} \tag{18}$$

Matrix $\boldsymbol{A_1}$:

$$\boldsymbol{A_1} = \begin{bmatrix} \frac{1}{T_s\sigma} & 0 & \frac{M}{L_s T_s} & 0 & 0 & 0 \\ 0 & \frac{1}{T_s\sigma} & 0 & \frac{M}{L_s T_s} & 0 & 0 \\ 0 & 0 & -\frac{1}{T_s\sigma} & 0 & \frac{M}{L_s T_s \sigma} & 0 \\ 0 & 0 & 0 & -\frac{1}{T_r\sigma} & 0 & \frac{M}{L_r T_r \sigma} \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -\frac{p}{J} & -\frac{f}{J} \end{bmatrix} \tag{19}$$

Matrix $\boldsymbol{A_2}$:

$$\boldsymbol{A_2} = \begin{bmatrix} 0 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix} \tag{20}$$

Matrix $\boldsymbol{A_3}$:

$$\boldsymbol{A_3} = \begin{bmatrix} 0 & 1 & 0 & \frac{M}{L_{s''}} & 0 & 0 \\ -1 & 0 & -\frac{M}{L_{s''}} & 0 & 0 & 0 \\ 0 & \frac{M}{L_{s''}} & 0 & -\left(\frac{M^2}{L_{s'} L_{r''}} + 1\right) & 0 & 0 \\ -\frac{M}{L_{s''}} & 0 & \frac{M^2}{L_s L_{r''}} + 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & -1 & 0 \end{bmatrix} \tag{21}$$

Matrix $\boldsymbol{A_4}$:

$$\boldsymbol{A_4} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{\alpha P}{\delta} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{22}$$

Matrix $\boldsymbol{A_5}$:

$$\boldsymbol{A_5} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{\alpha P}{\delta} & 0 & 0 & 0 & 0 \end{bmatrix} \tag{23}$$

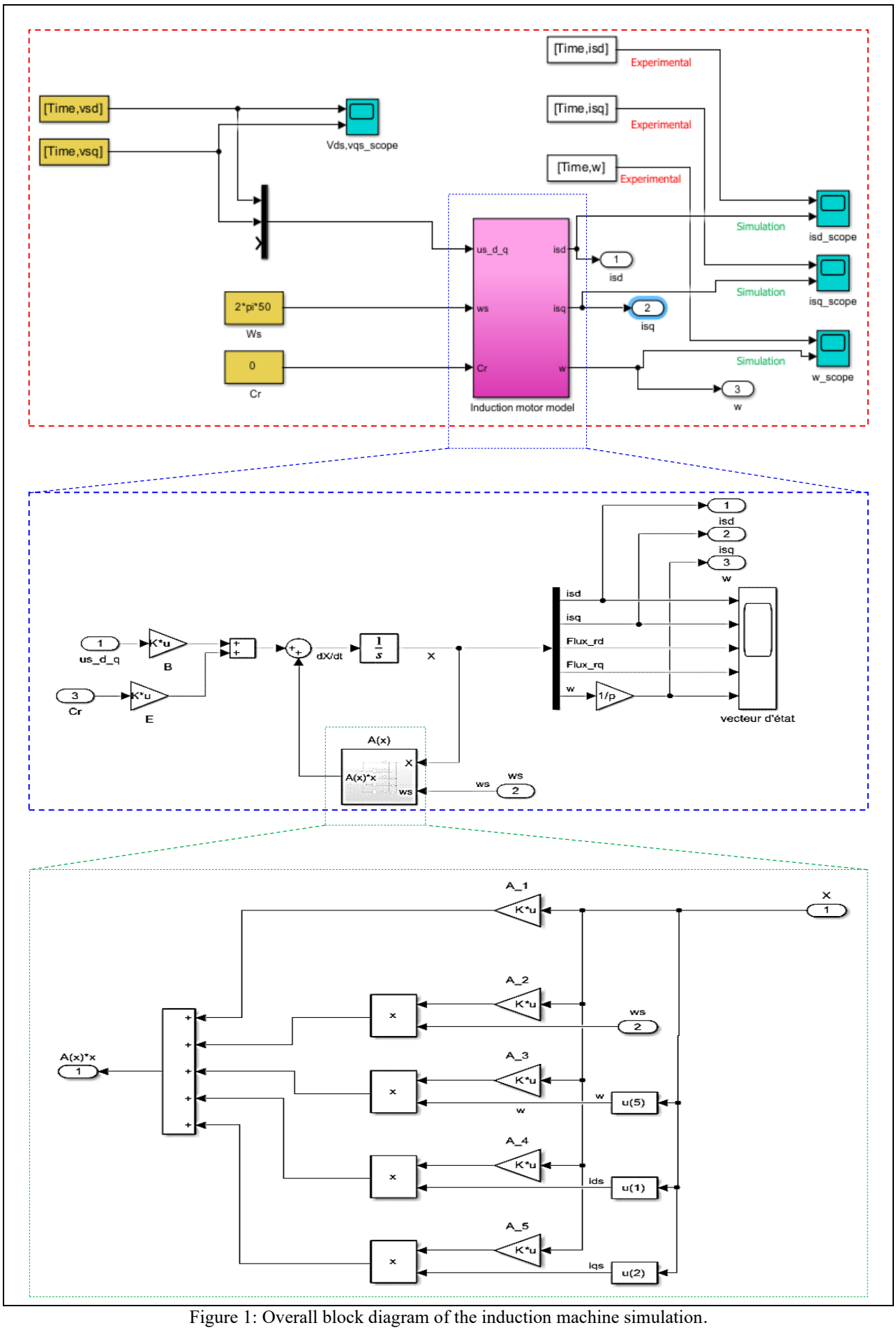


Figure 1: Overall block diagram of the induction machine simulation.
Source: Authors, (2025).

IV. METHODOLOGY FOR EXPERIMENTAL TESTS

IV. I. IMPORTING EXPERIMENTAL DATA

The first step in identifying the parameters of the induction motor is to collect experimental data by measuring the stator currents (I_{ds}, I_{qs}) and rotor speed (ω), which serve as key signals for modeling the system's dynamics. The block diagram in Figure 2 illustrates the hardware setup for the induction motor, which utilizes a dSPACE DS1104 interface board. The setup consists of the following key components:

- Three-Phase Network (50 Hz): Provides the AC power supply to the system.
- Auto Transformer: Regulates the voltage supplied to the induction motor, ensuring controlled variations in input voltage for identification purposes.
- Three-Phase Switch: Enables controlled connection and disconnection of the motor, allowing transient excitation to capture dynamic response data.
- Voltage and Current Sensors (abc): Measure the three-phase stator voltages (Va,Vb,Vc) and currents (Ia,Ib,Ic) for conversion into the dq-axis reference frame.
- Induction Motor: The system under test, whose electrical and mechanical characteristics are identified through measured responses.
- Incremental Encoder: Measures the rotor speed (ω) by providing real-time position feedback, which is converted into speed signals.
- Control Desk with dSPACE DS1104: Processes and records real-time data, saving results as .m files for further analysis in MATLAB.

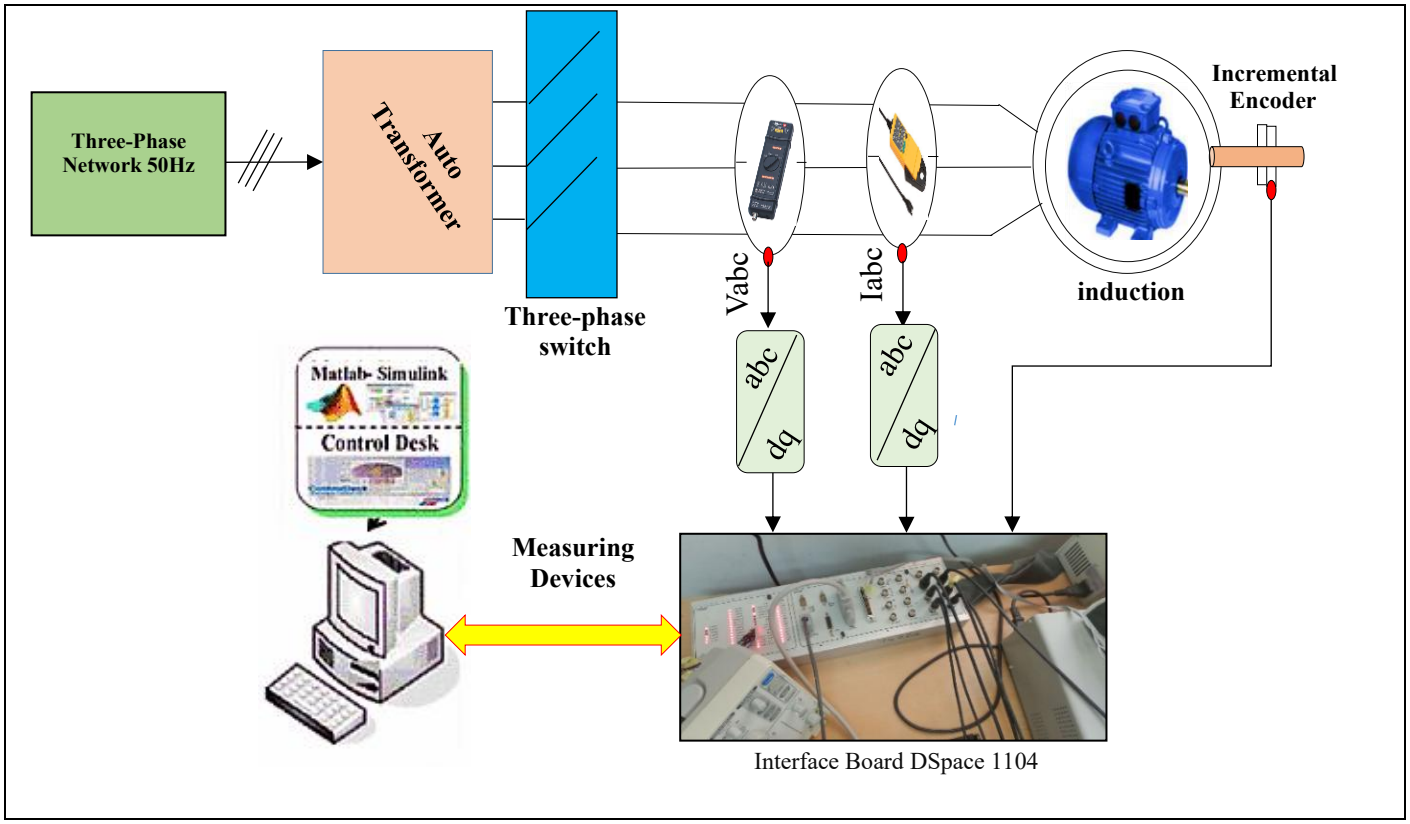


Figure 2: Overall block diagram of the experimental setup.
Source: Authors, (2025).

The Siemens induction motor, whose nameplate details are provided in Appendix A, is powered by a three-phase 50 Hz network with an auto-transformer to regulate the supply voltage. A three-phase switch is introduced to create controlled transient conditions, enabling the measurement of system responses under both steady-state and dynamic conditions.The stator currents (I_a, I_b, I_c) and voltages (V_a, V_b, V_c) are measured using sensors and then transformed into dq-axis components via an abc/dq transformation, allowing analysis in a synchronous reference frame. The rotor speed (ω) is recorded using an incremental encoder, providing real-time speed signals for accurate evaluation.These signals are processed by the dSPACE DS1104 interface board, which digitizes and transmits the data to MATLAB/Simulink ControlDesk for real-time monitoring and logging, as illustrated in Figure 3. This collected data is crucial for identifying motor parameters and characterizing its behavior under various operating conditions.

After importing the experimental data, open the simulation model, which includes the induction motor, as shown in Figure 1. To ensure accurate analysis, the experimental data, including rotor speed (ω), stator currents (I_{ds}, I_{qs}), and voltages (V_{ds}, V_{qs}), as presented in Figure 4, must be entered into the simulation. Before execution, the .m file available in Appendix B, which contains the motor's installed

parameters, must be loaded to initialize the system correctly. This step ensures precise parameter identification and validation by comparing the dynamic responses of simulated and experimental results, as illustrated in Figures 5,6 and 7 for I_{ds}, I_{qs}, and ω, respectively.

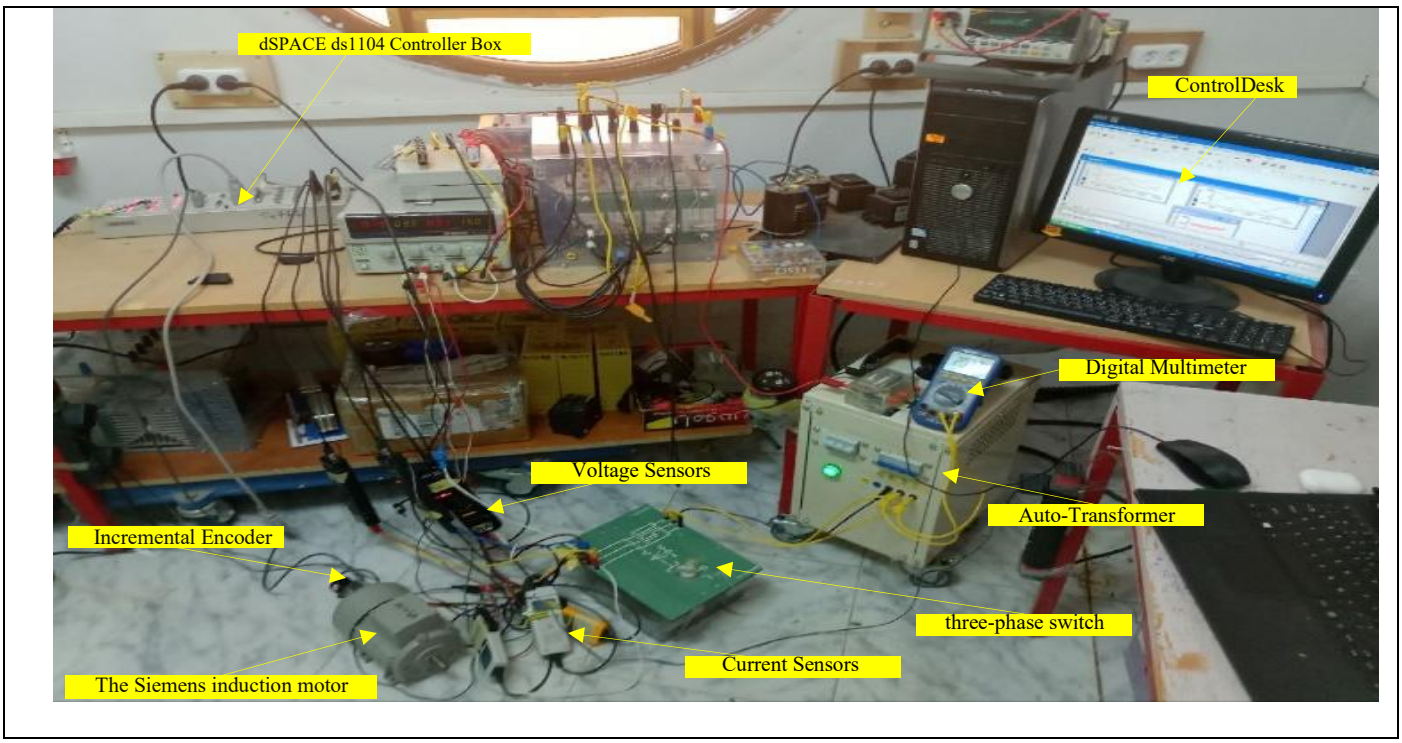


Figure 3: Overall block diagram of the experimental setup.
Source: Authors, (2025).

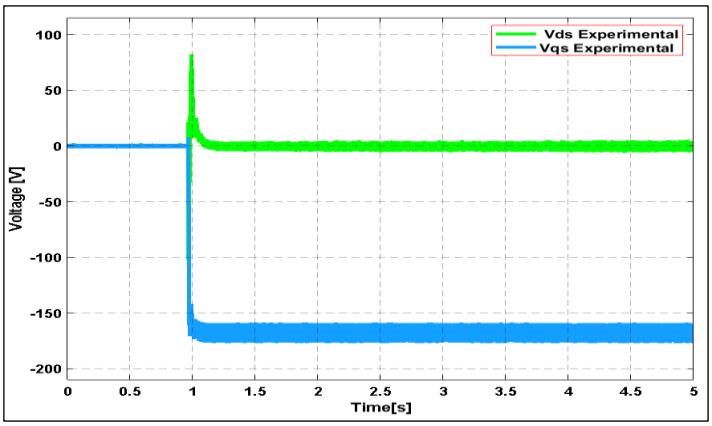


Figure 4: Experimental Responses of Vds and Vqs Voltage.
Source: Authors, (2025).

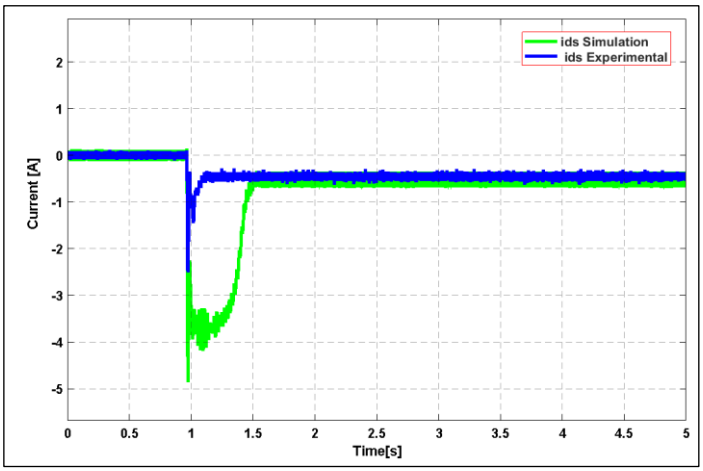


Figure 5: Dynamic Responses of Simulated and Experimental ids.
Source: Authors, (2025).

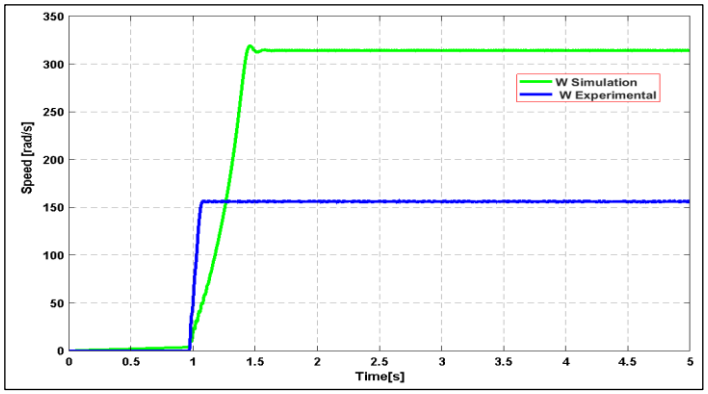


Figure 6: Dynamic Responses of Simulated and Experimental iqs.
Source: Authors, (2025).

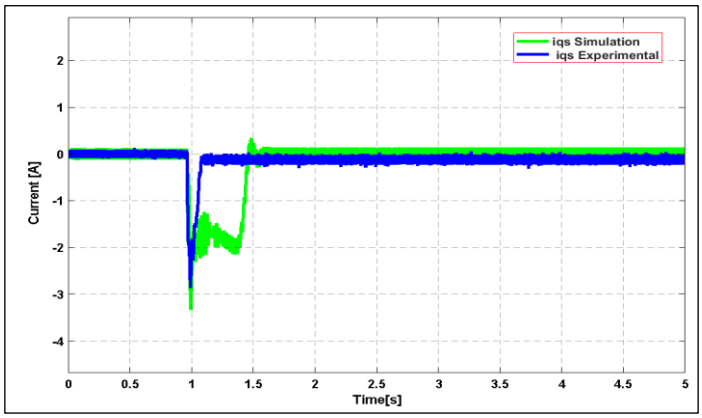


Figure 7: Dynamic Responses of Simulated and Experimental W.
Source: Authors, (2025).

IV.2 PARAMETER ESTIMATION PROCESS

To estimate the parameters of the induction motor using MATLAB's Parameter Estimator app, follow these steps:

The first step is to launch the Parameter Estimator. This can be done from the Simulink induction motor model by navigating to the Apps tab. Under Control Systems, click Parameter Estimator. Doing so will open the Parameter Estimation tool, which provides a user-friendly interface for configuring and executing the parameter estimation process. Alternatively, the tool can also be launched directly from the MATLAB command prompt by entering the command spetool('modelname'), where the model name is the name of the Simulink model. This command will open the Parameter Estimation tool for the specified model, allowing users to configure and run the estimation process efficiently.

Next, the measured experimental data must be imported for estimation. The required data consists of vds, vqs, ids, iqs, W, and Time, as shown in Figure 8. These values should be obtained from real-time measurements and saved in MAT files using ControlDesk. The Parameter Estimator also allows users to import data from other sources, such as MATLAB variables, Excel files, or comma-separated value (CSV) files, making it flexible for different applications.

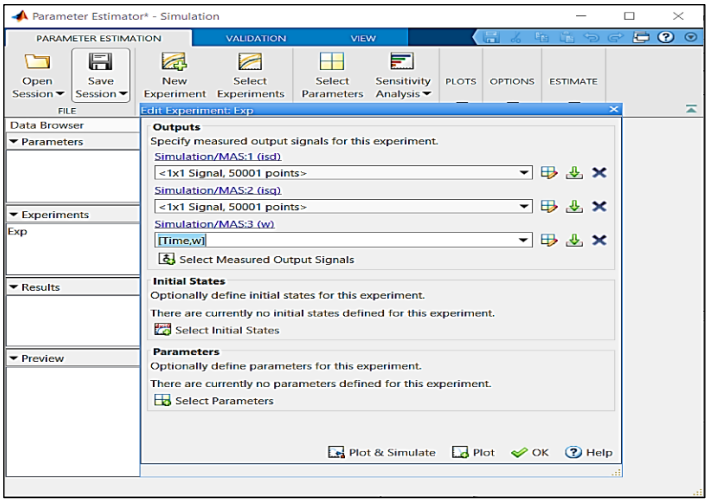


Figure 8: Selecting measured output signals in Parameter Estimator.
Source: Authors, (2025).

After importing the experimental data, the user must select the relevant experiment for inclusion in the estimation or validation process. The Parameter Estimator provides a dedicated interface for this selection, as shown in Figure 9. Choosing the appropriate experiments ensures that the estimation algorithm processes the correct dataset. Once the selection is finalized, clicking "OK" confirms the dataset for estimation, allowing the process to proceed efficiently.

Once the data has been successfully imported, the next step is to configure the induction motor parameters for estimation. The Parameter Estimator comes pre-loaded with key parameters relevant to the induction motor subsystem, including stator resistance (Rs), rotor resistance (Rr), magnetizing inductance (Lm), stator leakage inductance (Ls), rotor leakage inductance (Lr), moment of inertia (Jm), friction coefficient (f), and number of pole pairs (p). These parameters define the motor's electrical and mechanical behavior and are essential for accurate estimation.

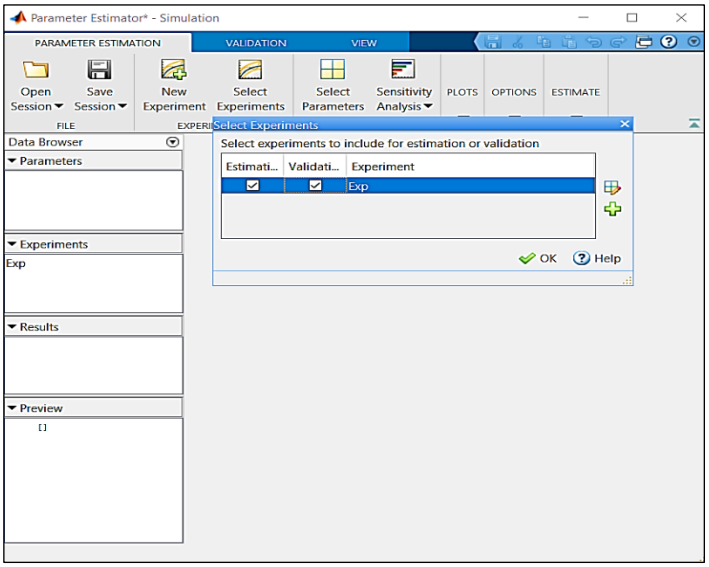


Figure 9: Selecting experiments for estimation and validation.
Source: Authors, (2025).

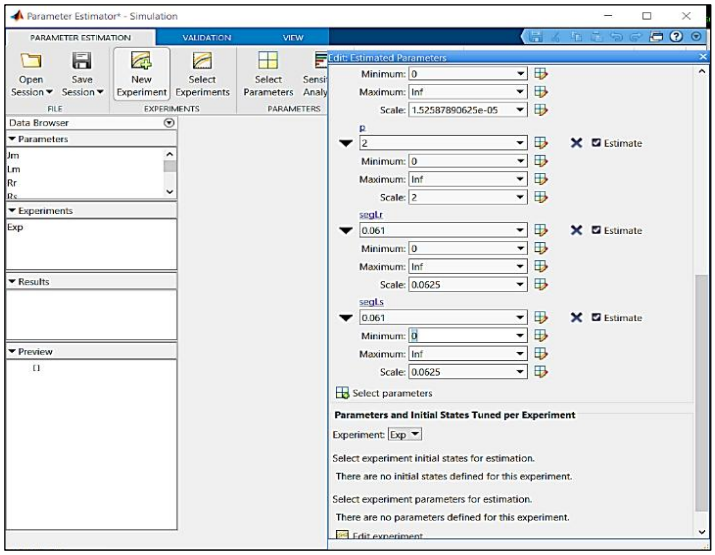


Figure 10: Parameter Selection and Constraint Configuration.
Source: Authors, (2025).

After defining the parameters, it is necessary to set appropriate constraints to ensure physically meaningful results. Each parameter should be confined within a realistic range, typically from 0 to a positive upper bound, preventing the estimation of non-physical negative values. This step is crucial for achieving reliable and valid results, as shown in Figure 10.

With all configurations set, the next step is to initiate the parameter estimation process. Clicking the "Run" button starts the optimization, where the Parameter Estimator iteratively adjusts the parameter values to minimize the error between simulated and measured responses, as shown in Figure 11. The estimation app continuously refines the parameters to achieve the best possible match with the experimental data, ensuring accuracy and reliability in the model.

Finally, once the estimation is complete, the estimated parameters must be validated. The user should compare the simulated motor response with the measured data to ensure a close fit. If discrepancies arise, adjustments to the parameter constraints or dataset may be necessary, followed by rerunning the estimation process. Through iterative refinement, the accuracy of the identified parameters can be significantly enhanced, leading to a more reliable and precise model.

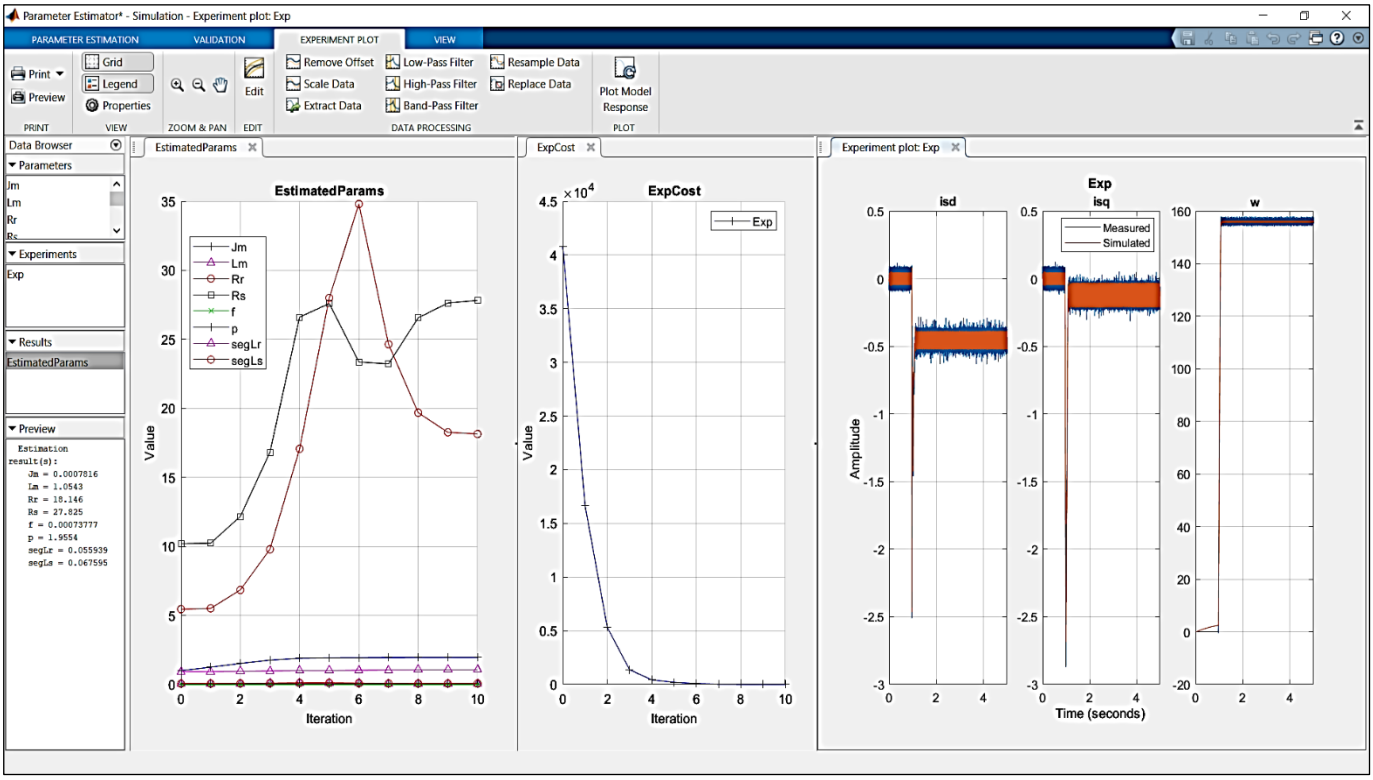


Figure 11: Convergence of Estimated Parameters and Error Minimization in MATLAB Parameter Estimator.
Source: Authors, (2025).

Appendix A

Siemens Induction Motor Nameplate. The corresponding nameplate is shown in Figure A1. This information is essential for identifying the motor's operating parameters and ensuring proper integration into the system.

Figure 1: Siemens induction motor nameplate.
Source: Authors, (2025).

Appendix B

Motor Parameter Initialization File

The following MATLAB script initializes the parameters of the three-phase induction motor. It defines the motor's electrical and mechanical parameters, ensuring accurate simulation and analysis. The script also loads experimental data from mesures.mat, which includes time, rotor speed (W), stator currents (i_{sd}, i_{sq}), and voltages (v_{sd}, v_{sq}).

```
clear;
clc;
%--------- Initial Parameters ---------
Rs = 10.2; % Stator resistance (measured directly using an ohmmeter)
Rr = 5.45; % Rotor resistance
Lm = 0.921; % Mutual inductance
% Ensure Ls and Lr > Lm using: Ls = segLs + Lm, Lr = segLr + Lm
segLs = 0.061; % Stator leakage inductance
```

```
segLr = 0.061; % Rotor leakage inductance
Jm = 0.00095; % Moment of inertia
f = 0.00001; % Friction coefficient
p = 1;     % Number of pole pairs
% Load experimental data
load('mesures.mat');
Time = mesures.X.Data';
w = mesures.Y(1).Data';
isd = mesures.Y(2).Data';
isq = mesures.Y(3).Data';
vsd = mesures.Y(4).Data';
vsq = mesures.Y(5).Data';
```

VI. CONCLUSIONS

In This study presents a methodology for estimating the parameters of a three-phase induction motor by integrating experi-mental data with simulation. Real-time data, including stator currents, rotor speed, and voltage signals, were collected through an experimental setup utilizing the dSPACE DS1104 platform. This dataset was then applied to a mathematical model of the motor in MATLAB/Simulink, representing its dynamic behavior with approximate initial pa-rameters. The Parameter Estimator App was then used to refine these parameters by iteratively adjusting their values to minimize the error between the simulated and measured responses. By initiating the estimation process, the app continu-ously optimized the parameters to achieve the best possible match with the experimental data, ensuring accuracy and relia-bility. The validation was performed using a Siemens 0.37 kW, 50 Hz, 380 V three-phase induction motor operating under zero-load conditions, where the cost function reached near zero, indicating a strong agreement between the simulated and experimental results. This approach ensures precise parameter identification, which is essential for control and diagnostic applications, with potential future extensions to other motor types for broader industrial applications.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Fares Bettahar, Sabrina Abdeddaim and Achour Betka, Michael Short and Maher Al-Greer
ethodology: Fares Bettahar and Sabrina Abdeddaim.
Investigation: Fares Bettahar and Sabrina Abdeddaim.
Discussion of results: Sabrina Abdeddaim and Achour Betka Laid Guerrida
Writing – Original Draft: Fares Bettahar.
Writing – Review and Editing: Fares Bettahar and Sabrina Abdeddaim .Belahcene taha lemdjed
Resources: Sabrina Abdeddaim.Omar Charrouf
Supervision: Sabrina Abdeddaim and Achour Betka.
Approval of the final text: Fares Bettahar, Sabrina Abdeddaim Achour Betka and Maher Al-Greer

VII. REFERENCES

[1] Dineva, A.; Mosavi, A.; Faizollahzadeh Ardabili, S.; Vajda, I.; Shamshirband, S.; Rabczuk, T.; Chau, K.-W. Review of Soft Computing Models in Design and Control of Rotating Electrical Machines. Energies 2019, 12(6), 1049. Available online: https://doi.org/10.3390/en12061049

[2] Merizalde, Y.; Hernández-Callejo, L.; Duque-Perez, O. State of the Art and Trends in the Monitoring, Detection and Diagnosis of Failures in Electric Induction Motors. Energies 2017, 10, 1056. Available online : https://doi.org/10.3390/en10071056

[3] Szénásy, I.; Csikor, D. Induction Motor Energy Efficiency Investigation. Engineering Proceedings 2024, 79(1), 75. Available online : https://doi.org/10.3390/engproc2024079075

[4] Liu, Z., Xu, L., & Xu, D. (2018). Advanced Control Strategies of Induction Machine: Field Oriented Control, Direct Torque Control, and Finite Set Model Predictive Control. Energies, 11(1), 120.

[5] Aktas, M.; Awaili, K.; Ehsani, M.; Arisoy, A. Direct torque control versus indirect field-oriented control of induction motors for electric vehicle applications. Engineering Science and Technology, an International Journal 2020, 23(5), 1134–1143. Available online: https://doi.org/10.1016/j.jestch.2020.04.002

[6] Bašić, M.; Vukadinović, D.; Grgić, I. Model Predictive Current Control of an Induction Motor Considering Iron Core Losses and Saturation. Processes 2023, 11(10), 2917. Available online : https://doi.org/10.3390/pr11102917

[7] Aktas, M.; Awaili, K.; Ehsani, M.; Arisoy, A. Direct torque control versus indirect field-oriented control of induction motors for electric vehicle applications. Engineering Science and Technology, an International Journal 2020, 23(5), 1134–1143. Available online: https://doi.org/10.1016/j.jestch.2020.04.002

[8] El Ouanjli, N.; Derouich, A.; El Ghzizal, A.; Motahhir, S.; Chebabhi, A.; El Mourabit, Y.; Taoussi, M. Modern improvement techniques of direct torque control for induction motor drives—a review. Protection and Control of Modern Power Systems 2019, 4(1), 11. Available online: https://doi.org/10.1186/s41601-019-0125-5

[9] Soliman, H.M. (2016). Effect of the Parameters Variation for Induction Motor on its Performance Characteristics with Field Oriented Control Compared to Scalar Control. International Journal of Engineering Research & Science (IJOER), 2(8), 1-7.

[10] Martin, C., Bermudez Guzman, M., Barrero, F., Arahal, M.R., Duran, M.J., & Kestelyn, X. (2017). Sensitivity of Predictive Controllers to Parameter Variation in Five-Phase Induction Motor Drives. Control Engineering Practice, 70, 1-12. https://sam.ensam.eu/handle/10985/13798

[11] Kačenka, A., Makys, P., & Struharnansky, L. (2020). Impact of Parameter Variation on Sensorless Indirect Field Oriented Control of Induction Machine. In I. Zelinka, P. Brandstetter, T. Trong Dao, V. Hoang Duy, & S. Kim (Eds.), AETA 2018 - Recent Advances in Electrical Engineering and Related Sciences: Theory and Application (pp. 799–809). Springer. https://doi.org/10.1007/978-3-030-14907-9_77

[12] Hanif, A., Ali, S. M. N., Ahmed, Q., Bhatti, A. I., & Aslam, M. (2016). Effect of Variation in Rotor Resistance on the Dynamic Performance of Induction Motor. In 2016 35th Chinese Control Conference (CCC) (pp. 5639–5644). IEEE. https://doi.org/10.1109/ChiCC.2016.7554869

[13] Harini, B. W. (2022). The Effect of Motor Parameters on the Induction Motor Speed Sensorless Control System using Luenberger Observer. International Journal of Applied Sciences and Smart Technologies, 4(1), 47–54. https://doi.org/10.24071/ijasst.v4i1.4518

[14] Kačenka, A., Makys, P., & Struharnansky, L. (2020). Impact of Parameter Variation on Sensorless Indirect Field Oriented Control of Induction Machine. In I. Zelinka, P. Brandstetter, T. Trong Dao, V. Hoang Duy, & S. Kim (Eds.), AETA 2018 - Recent Advances in Electrical Engineering and Related Sciences: Theory and Application (pp. 799–809). Springer. https://doi.org/10.1007/978-3-030-14907-9_77

[15] Et-taaj, L., Boulghasoul, Z., El Kharki, A., & Elbacha, A. (2022). Improvement of sensorless control of induction motor by voltage source inverter nonlinearities compensation and extended Kalman filter. Electrical Engineering, 104(6), 3509–3521. https://doi.org/10.1007/s00202-022-01560-1

[16] Tran, T. C., Brandstetter, P., Tran, C. D., Ho, S. D., Nguyen, M. C., & Phuong, P. N. (2019). Parameters Estimation for Sensorless Control of Induction Motor Drive Using Modify GA and CSA Algorithm. In AETA 2018 - Recent Advances in Electrical Engineering and Related Sciences: Theory and Application (pp. 580–591). Springer. https://doi.org/10.1007/978-3-030-14907-9_57

[17] Fagiano, L., Lauricella, M., Angelosante, D., & Ragaini, E. (2018). Identification of Induction Motors with Smart Circuit Breakers. arXiv preprint arXiv:1804.07817. https://arxiv.org/abs/1804.07817

[18] Susanto, J., & Islam, S. (2017). Improved Parameter Estimation Techniques for Induction Motors using Hybrid Algorithms. arXiv preprint arXiv:1704.02424. https://arxiv.org/abs/1704.02424

[19] Sarhan, H., Guimarães, J.M.C., Bernardes, J.V., Hermeto, A.E., & Bortoni, E.C. (2023). Numerical Estimation and Experimental Verification of Optimal Parameter Identification Based on Modern Optimization of a Three Phase Induction Motor. Mathematics, 7(12), 1135. https://doi.org/10.3390/math7121135

[20] Guo, H., Yang, X., Diao, N., Sun, X., Zhang, H., Zhao, X., Li, H., Hu, W., & Song, C. (2025). The least squares off-line parameter identification method of half-dead compensated induction motor based on non-π/3 angle virtual rotating vector orientation. Electrical Engineering, 107(1), 1–15. https://doi.org/10.1007/s00202-024-02893-9

[21] Belbali, A., Makhloufi, S., Kadri, A., Abdallah, L., & Seddik, Z. (2023). Mathematical Modeling of a Three-Phase Induction Motor. In Induction Motors-Recent Advances, New Perspectives and Applications. IntechOpen. https://doi.org/10.5772/intechopen.101358

[22] Le Roux, P. F., and M. K. Ngwenyama. 2022. "Static and Dynamic Simulation of an Induction Motor Using Matlab/Simulink" Energies 15, no. 10: 3564. https://doi.org/10.3390/en15103564

[23] Sengamalai, U.; Anbazhagan, G.; Thamizh Thentral, T.M.; Vishnuram, P.; Khurshaid, T.; Kamel, S. (2022). Three Phase Induction Motor Drive: A Systematic Review on Dynamic Modeling, Parameter Estimation, and Control Schemes.
Energies, 15(21), 8260. https://doi.org/10.3390/en15218260.

[24] M. Pucci, "State Space-Vector Model of Linear Induction Motors," in IEEE Transactions on Industry Applications, vol. 50, no. 1, pp. 195-207, Jan.-Feb. 2014, https://doi.org/10.1109/TIA.2013.2266351

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11n.55, p.86-95. September/October, 2025.
DOI: https://doi.org/10.5935/jetia.v11i55.1910

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

MULTI-STAGE FEATURE SELECTION FOR OPTIMIZING STUDENT DROPOUT PREDICTION

Arif Mudi Priyatno[1], Yunia Ningsih[2], Rizqon Jamil Farhas[3], Fahmi Iqbal Firmananda[4], Resy Kumala Sari[5], and Aryadi[6]

[1,3,4,5,6] Universitas Pahlawan Tuanku Tambusai, Riau, Indonesia
[2] Universitas Trisakti, Jakarta, Indonesia

[1]http://orcid.org/0000-0003-3500-3511, [2]http://orcid.org/0009-0005-6388-1233, [3]http://orcid.org/0009-0002-1092-9596
[4]http://orcid.org/0009-0001-9744-6580, [5]http://orcid.org/0000-0003-3138-3502, [6]http://orcid.org/0009-0009-0495-2264

Email: arifmudi@universitaspahlawan.ac.id, yunia@trisakti.ac.id, rizqonjamil@universitaspahlawan.ac.id, fahmiiqbalfirmananda@universitaspahlawan.ac.id, resikumalasari@universitaspahlawan.ac.id, aryadi@universitaspahlawan.ac.id

ARTICLE INFO

Article History
Received: April 22, 2025
Revised: May 25, 2025
Accepted: September 15, 2025
Published: September 30, 2025

Keywords:
Multi-Stage Feature Selection,
Student Dropout Prediction,
Boruta,
XGBoost,
Educational Data.

ABSTRACT

The high rate of college dropouts is a significant challenge in higher education. Dropout prediction requires an accurate model and is supported by a selection of relevant features. This study proposes a step-by-step feature selection framework to improve prediction accuracy, consisting of three stages, namely Variance Threshold, Mutual Information, and Boruta. The classification model is built using the Extreme Gradient Boosting (XGBoost) algorithm, with evaluation through Stratified 10-fold Cross-Validation. The dataset used includes 4,423 student data that reflects academic, demographic, and socioeconomic information. A total of 18 features were confirmed to be relevant by Boruta. XGBoost models trained on selected features show high performance, with an accuracy of 90.77%, precision of 92.07%, recall of 83.68%, and an F1-score of 87.63%. These results show that the integration of filter and wrapper approaches in the feature selection process effectively improves the performance of the dropout prediction model. This framework is able to filter out important features and produce a more stable and efficient classification model in the context of higher education.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The student dropout rate is one of the serious challenges in the world of higher education in various parts of the world, including Indonesia [1], [2], [3], [4]. This phenomenon not only affects an individual's academic achievement, but also poses significant social and economic implications for educational institutions and society at large [5]. When a student decides to drop out of college, the institution loses out on the potential of qualified graduates, while the individual in question may face career barriers or additional economic burdens. Therefore, efforts to identify and predict potential dropouts early on are critical so that appropriate interventions can be provided before such decisions occur.

In recent years, machine learning-based approaches have been widely applied in the field of education to help the decision-making process, including in predicting student dropouts [6], [7], [8], [9]. Accurate predictive models can provide in-depth insights into the risk factors that cause students to drop out of the education system. However, the effectiveness of this model is highly dependent on the quality of the data and features used in the training process [10], [11]. The many attributes in educational data, such as academic, demographic, and behavioral information, often result in a curse of dimensionality problem, which is when the large number of features is not proportional to the amount of data, so that it can reduce the performance of the prediction model.

Another challenge that is often faced in building a dropout prediction model is the existence of irrelevant or redundant features [12], [13]. Features that do not provide significant information about the target variable, or features that are highly correlated with each other, can cause overfitting the model and worsen generalizations in new data [14]. In this context, the feature selection process is crucial

as an effort to filter out features that are really relevant and have a significant contribution to the prediction process [15], [16]. By selecting the right features, the data dimension can be optimally reduced, while improving the computational efficiency and accuracy of the model.

Various feature selection methods have been developed, both filters [17], wrappers [18], and embedded [19]. However, most previous studies have relied on only one approach in the feature selection process. For example, the use of Mutual Information is often used to measure the dependencies between features and target variables [20], while Boruta is used to evaluate the importance of features iteratively based on the decision tree [21]. The use of a single method tends to have limitations in comprehensively filtering features, as they do not consider the combination of information value, variability, and feature stability.

To overcome these limitations, this research proposes a multi-stage feature selection approach to improve the accuracy of student dropout predictions. This approach consists of three stages that are systematically designed. In the first stage, initial screening is carried out using the Variance Threshold to remove features with low variability that tend to be uninformative, then followed by Mutual Information to assess the strength of dependency between each feature and the target. This stage aims to eliminate features that are irrelevant in the first place, while maintaining features with high information value. The second stage involves the Boruta method as a Random Forest-based wrapper technique, which is tasked with evaluating the importance of each feature to the target variable in more depth. Boruta has the advantage of detecting important features by considering the interaction between features in an exploratory manner. Thus, this stage is focused on selecting the key features that have the most influence on dropout predictions, while also filtering out features that may have escaped initial selection but do not have a significant contribution in the overall context of the model. After the optimal features are obtained through the two stages of selection, the third stage is carried out by building a prediction model using Extreme Gradient Boosting (XGBClassifier), which is known as one of the high-performance ensemble-based classification algorithms. XGBoost was chosen for its ability to handle large-scale data and effectively handle class imbalances. The model was then evaluated using relevant metrics to measure the accuracy and stability of predictions on student dropout data.

II. MATERIALS AND METHODS

This research proposes a multi-stage feature selection framework that is systematically designed to improve the performance and generalization ability of the prediction model of college dropouts. This approach is formulated to progressively screen features, with the goal of eliminating irrelevant, redundant, or low-variance attributes prior to the model training process. By combining filter- and wrapper-based selection methods, this approach is expected to achieve a balance between computational efficiency, model accuracy, and interpretability of results.

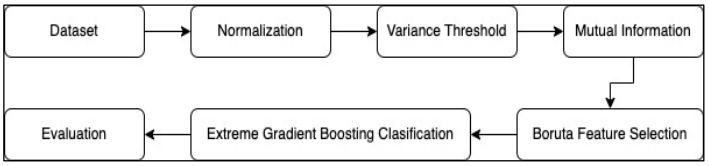


Figure 1: Proposed research framework.
Source: Authors, (2025).

The proposed research framework is presented in Figure 1. The process begins with the acquisition and pre-processing of the dataset, where all features are normalized using standardization techniques to equalize the scale between variables. The first feature selection stage is performed using the Variance Threshold to remove features with very low levels of variation, which generally do not have a significant contribution to the classification process. The remaining features are then evaluated using Mutual Information, a filter-based method that measures statistical dependencies between features and target variables.

In the second stage, the features of the initial selection results were filtered again using the Boruta Feature Selection method, which is a wrapper-based approach built on the Random Forest algorithm. This stage aims to retain the most significant and consistent features in explaining the target variables, while eliminating features that, although they pass the initial selection, have a weak predictive contribution overall.

After obtaining an optimal subset of features, the prediction model was developed using the Extreme Gradient Boosting (XGBoost) algorithm, which is widely known for its high predictive capabilities and its ability to handle high-dimensional and unbalanced data. The model's performance was rigorously evaluated using the Stratified K-Fold Cross-Validation technique, with key metrics used including accuracy, precision, recall, and F1-score. The design of this methodology aims to produce a balanced and reliable assessment of the effectiveness of the proposed approach. Each stage in this framework is described in detail in the following sub-sections.

II.1 DATASET

The dataset used in this study was obtained from the UCI Machine Learning Repository with the title Student Dropout and Academic Success. This dataset contains a total of 4,423 entries, where each entry represents a student with a diverse academic, demographic, and socioeconomic background. The target variables in this dataset are divided into two main categories, namely students who successfully complete their studies (Academic Success) and students who have dropped out of study (Dropout). Of the total available data, 3,003 students were classified as successful completions, while 1,421 students were categorized as dropouts, indicating a moderate class imbalance but still relevant for the development of a classification model [22].

This dataset includes 36 features that represent various aspects that are suspected to affect students' academic success. These features can be grouped into three main categories, namely demographic, academic, and socioeconomic. Demographic characteristics include attributes such as marital status, nationality, gender, age at the time of enrollment, and parental education level. Academic characteristics include students' grades and academic track record, such as entrance scores, the number of courses taken and completed

in the first and second semesters, and final semester grades. Meanwhile, socioeconomic features reflect external conditions such as the unemployment rate, inflation rate, and gross domestic product (GDP).

Each attribute in the dataset contributes differently to the analysis process. For example, academic features such as entrance scores and the number of approved courses can directly reflect a student's academic performance. On the other hand, variables such as maternal education or father's occupation provide insights related to the influence of family background on educational outcomes. In addition, macroeconomic features such as unemployment rates and inflation allow for the exploration of the influence of the external environment on the likelihood of students continuing or terminating their studies.

II.2 NORMALIZATION

Before the feature selection process is performed, all numerical attributes in the dataset are normalized using the StandardScaler method. This normalization aims to equalize the scale between features so that each variable has a balanced contribution in the distance calculation process as well as in algorithms that are sensitive to the data scale [23]. This is especially important, especially when features have very different ranges of values, which can lead to dominance by larger-scale features in the model's calculations. The StandardScaler method works by transforming each feature value to have a mean of zero and a standard deviation of one. This transformation is carried out based on Equation 1 [24].

$$x_{new} = \frac{x_i - \mu}{\sigma} \quad (1)$$

Where x_{new} is a new value as a result of transformation. x_i is the original value, μ is an average feature, and σ is a standard deviation from the feature.

II.3 VARIANCE THRESHOLD

The first stage of feature selection was carried out using the Variance Threshold method, which functions to remove features that have very low variability [25]. Features with low variance tend to be constant across the entire data, so they don't provide meaningful information in distinguishing the target class. Therefore, features like this can be eliminated without sacrificing the accuracy of the model. Mathematically, the variance of a feature x against n the sample can be calculated with Equation 2. Where x_i is a feature to-i, $\overline{x}$ is an average feature, $var(x)$ expresses the variance of features.

$$var(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \quad (2)$$

In this study, the threshold of variance was set at 0.01, which means that only features with a sufficiently high variation were retained for the next stage of analysis. The variance threshold algorithm is shown in Algorithm 1.

Algorithm 1: Algoritma Variance Threshold.

Input	Dataset of normalization results X and variance threshold values
Output	A subset of var(x) features that have a variance value higher than the threshold
Step	1. Calculate the variance of each feature in dataset X using Equation 2.
	2. Compare each variance value with the threshold value.
	3. Select features that have a variance value greater than the threshold.
	4. Create a new subset of var(x) that contains only selected features.

Source: Authors, (2025).

II.4 MUTUAL INFORMATION

After the features with low variance are eliminated, the next stage in the feature selection process is to measure the relevance between each feature and the target variable using the Mutual Information (MI) method [26]. Mutual Information is an effective filter-based feature selection technique to identify both linear and non-linear relationships between free and target variables. Conceptually, Mutual Information measures how much knowledge of feature variable X can reduce uncertainty about target Y. The higher the MI value between a feature and a target, the more information the feature contributes to the target's prediction. The Mutual Information equation between two discrete variables X and Y is defined in Equation 3.

$$MI(X,Y) = \sum_{x\epsilon X}\sum_{y\epsilon Y} p(x,y)\,.\,log\,(\frac{p(x,y)}{p(x)\,.p(y)}) \quad (3)$$

Where $p(x,y)$ is the combined probability between x and y. $p(x)$ and $p(y)$ is the marginal probability of x and y. In this research, the MI value was calculated for each feature against the dropout target. Features with an MI score above the median of the entire MI score are retained for the next selection stage. This approach aims to filter out features that provide significant information to the target, while maintaining a reasonable number of features. The Mutual Information Algorithm is shown in Algorithm 2.

Algorithm 2: Algoritma Mutual Information.

Input	A subset of features of the Variance Threshold (X_{var}) and target variable and
Output	A subset of features X_{mi} with a Mutual Information value above the median value
Step	1. Calculate the value of Mutual Information between each feature $x_i \in X_{var}$ and targets y
	2. Store MI scores in an array $S = \{MI(x1,y), MI(x2,y), \ldots, MI(xn,y)\}$
	3. Calculate the median value of the MI score: median(S).
	4. Select features with a median > MI score
	5. Form a new feature subset X_{mi} for the advanced selection stage.

Source: Authors, (2025).

II.5 BORUTA FEATURE SELECTION

After the features are filtered based on relevance using Mutual Information, the next stage is to apply a wrapper-based advanced feature selection technique using the Boruta algorithm [27]. Boruta is designed to identify all features relevant to the target by comparing the importance of each original feature against the randomly generated shadow features. This method provides a more exploratory and conservative approach than the usual selection method, with a primary focus on maintaining all the features that really matter, not just the best.

Boruta works by building a Random Forest model on the dataset, then assessing the importance of each feature using the feature importance score of Random Forest. To test the significance of the original feature, Boruta created a random copy of the feature (shadow features) and compared the original feature score to the shadow feature score distribution. Only features with consistently higher scores than shadow features are maintained. In general, the main logic of Boruta is that if a feature is statistically more important than a random (shadow) feature, then it is considered relevant and retained. The Boruto feature selection algorithm is shown in Algorithm 3.

Algorithm 3: Algoritma Boruto Feature Selection.

Input	A subset of features of Mutual Information results (X_{mi}), target variable y
Output	A subset of important features X_{boruto} selected by Boruto
Step	1. Add *shadow features* by randomizing the value of each original feature.
	2. Train the Random Forest model on datasets that contain both original features and shadows.
	3. Calculate *feature importance* for all features, both original and shadow.
	4. Compare the *importance score* of each original feature to the maximum score of the shadow feature.
	5. Mark the feature as "important" if the score > the shadow max significantly.
	6. Remove non-essential features and repeat the process until the convergence or iteration limit is reached.
	7. Save features marked as important to be X_{boruto}.

Source: Authors, (2025).

Boruta was chosen in this study because of its excellence in addressing the interaction between features and tolerance to multicollinearity, which is particularly important in the context of complex educational data. The feature selection process at this stage is iterative and lasts until all features can be classified as confirmed, tentative, or rejected. Only features with confirmed status are used for model training at a later stage.

II.6 CROSS-VALIDATION

To ensure the reliability and generalization of prediction models against previously unseen data, this study used the Stratified K-Fold Cross-Validation technique [28]. Cross-validation is a model validation approach that is very important in classification studies, as it allows for a thorough evaluation of model performance by systematically dividing the data into subsets (folds) of training and testing.

Stratified K-Fold was chosen because the characteristics of the target used are unbalanced, where the number of students who drop out is less than those who successfully complete their studies. Stratification ensures that each fold has the same class proportions as in the original distribution, thus avoiding bias and providing a more stable performance estimate.

In its implementation, the dataset is divided into *k* subsets (folds), then the training and testing process is carried out k times. In each iteration, one fold is used as test data, while the *remaining k-1* fold is used for training. The average of all evaluation metrics obtained from each iteration is used as an indicator of the overall performance of the model. The cross-validation algorithm is shown in Algorithm 4.

Algorithm 4: Algoritma Stratified K-Fold Cross-Validation.

Input	Final feature selection dataset X_{boruto}, Target Label Y, Number of Fold *K*
Output	The value of the evaluation metrics (Accuracy, Precision, Recall, F1-Score) of each fold and its average value
Step	1. Divide the data into *k* subsets by maintaining the proportion of classes in each subset (stratification).
	2. For each iteration to-*i* (from 1 to *k*):
	a. Use the i-fold as the test data.
	b. Use the other $k-1$ fold as training data.
	c. Train machine learning models on training data.
	d. Evaluate model performance on test data using classification metrics.
	3. Save the results of the evaluation on each iteration.
	4. Calculate the average value of each evaluation metric from all folds.

Source: Authors, (2025).

This technique was used in the study with k=10, so the model was trained and tested 10 times with different combinations of data. This provides more accurate performance estimates compared to single data sharing, and reduces the risk of overfitting a specific subset of data.

II.7 EXTREME GRADIENT BOOSTING CLASIFICATION

The classification model used in this study is Extreme Gradient Boosting (XGBoost), which is an ensemble algorithm based on gradient boosting decision trees (GBDT) which is known to have high performance and good computational efficiency [29]. XGBoost is designed to optimize the loss function gradually through the addition of sequential decision trees, where each new tree is built to correct prediction errors from the previous tree.

One of the main advantages of XGBoost is its ability to handle high-dimensional data and unbalanced class distribution, which is a common characteristic in the case of predictive college dropouts. In addition, this algorithm supports regularization (L1 and L2) which serves to reduce the risk of overfitting and improve model generalization. XGBoost iteratively minimizes loss functions with the same 4.

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i))) + \Omega(f_t) \quad (4)$$

Where l is a loss function (e.g. log-loss), $\hat{y}_i^{(t-1)}$ is a prediction from the previous iteration, f_t is the decision tree that is added on the iteration to-t, and $\Omega(f_t)$ is a regularization function to control the complexity of the tree. The XGBoost model in this study was implemented with default parameters, using the histogram method to speed up the tree pruning process and evaluation based on *logloss metrics* suitable for binary classification.

II.8 EVALUATION

The evaluation of model performance was carried out using several common and representative classification metrics, namely accuracy [30], precision [31], recall [32], and F1-score [24]. Each metric was chosen to capture different aspects of prediction quality, especially in the context of the class imbalances that often occur in student dropout predictions. This approach ensures that the model's performance assessment is not biased against only the majority class. Mathematically, the evaluation metrics are defined in Equations 5-8.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5)$$

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

$$Recall = \frac{TP}{TP+FN} \quad (7)$$

$$F1_{score} = 2\,x\,\frac{Precision\ x\ Recall}{Precision+Recall} \quad (8)$$

Where TP (True Positive) is a correctly predicted dropout, TN (True Negative) is a correctly predicted non-dropout, FP (False Positive) is a dropout prediction when it is not, and FN (False Negative) is a prediction of not a dropout when it should be dropout.

In the context of predicting student dropouts, recall is an important metric because it represents the model's ability to accurately detect dropout cases. Meanwhile, precision describes the proportion of dropout predictions that are truly relevant. F1-score provides a balance between precision and recall, and accuracy is used as a general metric that describes the correct proportion of the prediction of the entire data.

III. RESULTS AND DISCUSSIONS

The dataset used in this study is a dataset obtained from the UCI Machine Learning Repository, with a total of 4,423 entries, each representing a student. This data includes academic, demographic, and socioeconomic information that is relevant to analyze the potential for students to drop out of college. The target variables in the dataset are classified binarily, namely 0 for graduates and 1 for dropouts.

Figure 2 shows the distribution of the number of students by target class. From the total data, as many as 3,003 students are classified as graduates, while 1,421 students are classified as dropouts. This distribution shows a significant class imbalance, but not extreme, so evaluation approaches such as Stratified K-Fold Cross Validation are used to ensure that the model remains fairly tested against both classes.

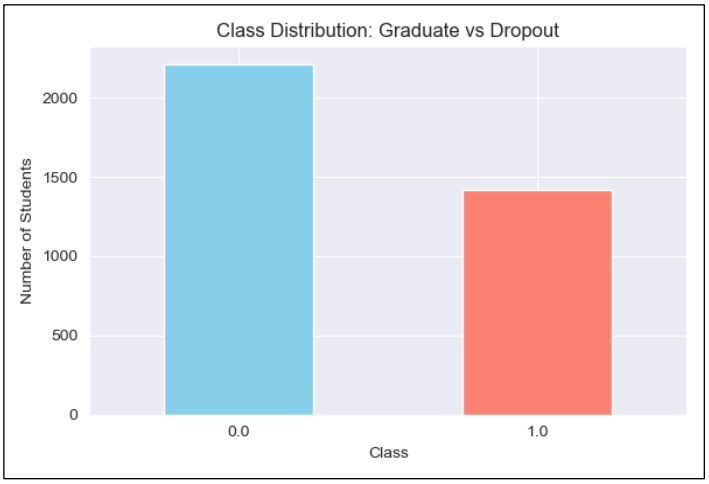


Figure 2: Distribution of classes between graduating students and dropping out of college.
Source: Authors, (2025).

To provide a preliminary overview of the characteristics of the data, Table 1 presents a descriptive statistical summary of some of the selected numerical features that are considered representative. In terms of demographics, students have an average age when registering at 23.46 years, with a standard deviation of 7.83, and an age range between 17 and 70 years. Academic features such as Admission grade have an average score of 127.29, with grade variations from 95 to 190, while the number of courses taken in the first semester reaches 26 courses, although the average is only 6.34. The number of courses approved in the same semester has an average score of 4.79. From the socioeconomic aspect, the unemployment rate at the time of enrollment showed an average score of 11.63, with

variations between 7.6 to 16.2. These variables are an important basis for understanding the potential factors that can affect the success or failure of student studies.

Table 1: Initial descriptive statistics of selected numerical features in the dataset.

	Average	Standard Deviation	Minimum Score	Maximum Value
Age at enrollment	23.46	7.83	17	70
Admission grade	127.29	14.61	95	190
Curricular units 1st sem (enrolled)	6.34	2.57	0	26
Curricular units 1st sem (approved)	4.79	3.24	0	26
Unemployment rate	11.63	2.67	7.6	16.2
Inflation rate	1.23	1.38	-0.8	3.7

Source: Authors, (2025).

The normalization process is carried out to equalize the scale of all numerical features before the feature selection and model training stages are implemented. This transformation uses the StandardScaler approach that sets each feature to have a mean of zero and a standard deviation of one. A statistical summary of the five numerical features selected after the normalization process is shown in Table 2.

Table 2: Descriptive statistics of numerical features after normalization.

	Average	Standard Deviation	Minimum Score	Maximum Value
Age at enrollment	0	1	-0.83	5.95
Admission grade	0	1	-2.21	4.29
Curricular units 1st sem (enrolled)	0	1	-2.47	7.65
Curricular units 1st sem (approved)	0	1	-1.48	6.55
Unemployment rate	0	1	-1.51	1.71

Source: Authors, (2025).

Based on the results in Table 2, all the features analyzed show an average value close to zero and a standard deviation close to one. For example, the Admission grade has a minimum score of –2.21 and a maximum of 4.29 after normalization, while the Curricular units 1st sem (enrolled) show a range between –2.47 to 7.65. This process results in a more uniform distribution of values and prevents the dominance of certain features in the calculation of variance as well as in the machine learning process. After the normalization process is carried out, the initial stage of feature selection is applied to filter out attributes that do not contribute significant information to the target variable. The first method used is the Variance Threshold, with a minimum threshold of 0.01. Features that have variance below this value are considered not to provide a significant difference between the data and are eliminated from the subsequent analysis process. Based on the results of elimination, as many as 36 features were successfully maintained from the total initial features.

Features that pass the first stage are then further evaluated using the Mutual Information (MI) method, which aims to measure the degree of statistical dependence between each feature and the target variable. MI is non-linear and does not assume linear relationships, making it suitable for datasets with complex structures like this. In implementation, the MI value is calculated for each feature and features with a score above the median value are retained. A total of 18 features were selected based on the value of Mutual Information, and the ten features with the highest scores were presented in Figure 3. These features generally come from the student's academic domain, such as the number of approved courses and the grades of the first and second semesters. These findings show that academic performance is the main indicator that distinguishes students who complete their studies and those who experience dropouts. Seleksi fitur tahap pertama ini menyaring atribut yang tidak hanya bervariasi secara statistik, tetapi juga terbukti memiliki hubungan informasional yang kuat terhadap target. Fitur-fitur yang lolos kemudian digunakan sebagai input pada tahap seleksi berikutnya menggunakan metode Boruta.

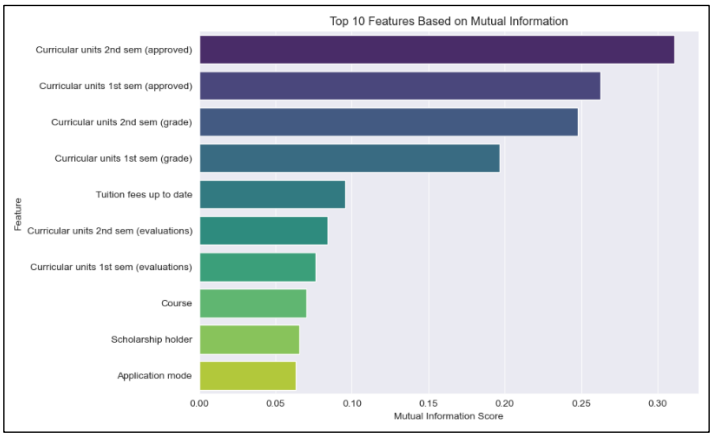


Figure 3: The top ten features are based on the value of Mutual Information against the dropout target variable.
Source: Authors, (2025).

The results of this second stage show that all the evaluated features are classified as confirmed by the Boruta algorithm. A total of 18 features were confirmed as relevant, with no features included in the tentative or rejected categories. Although these results are not common in the practice of Boruta applications, they can be methodologically justified. First, the features that enter the Boruta stage have been strictly screened through the Mutual Information approach, so that only features that have a high information contribution remain.

Second, the number of features tested is relatively limited, which increases the likelihood that each feature will show significant advantages over random features (shadow features).

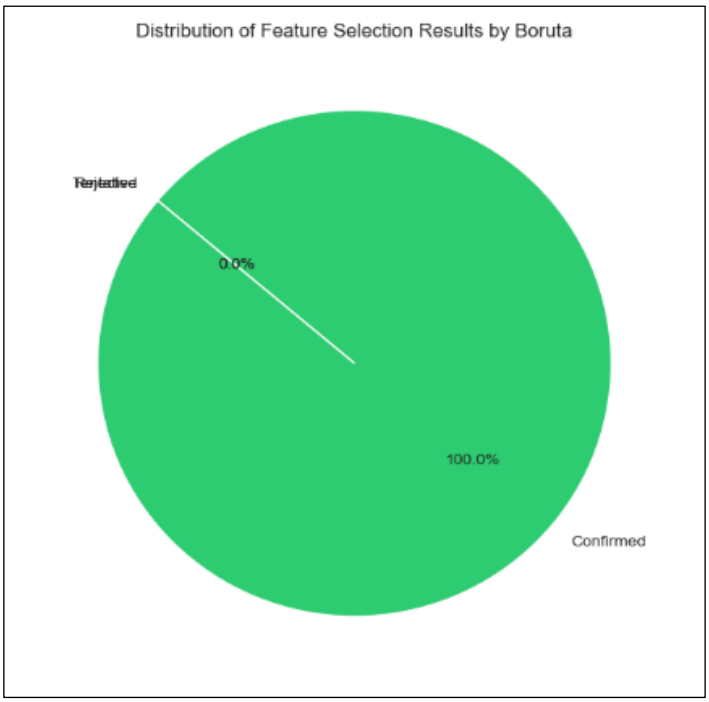


Figure 4: The distribution of feature selection results is based on the Boruta algorithm.
Source: Authors, (2025).

The distribution of feature selection results is shown in Figure 4, which shows that all features are in the confirmed category. Table 3 presents a list of these features, which include administrative indicators (such as Application mode and Course), financial status (Debtor, Tuition fees up to date), scholarship status (Scholarship holder), and various academic and demographic variables. These findings reinforce the assumption that student dropout risk is not only determined by academic performance alone, but also by administrative and socioeconomic conditions.

Table 3: The features are confirmed as relevant by the Boruta algorithm.

Feature	Selected	Tentative
Application mode	TRUE	FALSE
Course	TRUE	FALSE
Previous qualification (grade)	TRUE	FALSE
Admission grade	TRUE	FALSE
Debtor	TRUE	FALSE
Tuition fees up to date	TRUE	FALSE
Gender	TRUE	FALSE
Scholarship holder	TRUE	FALSE
Age at enrollment	TRUE	FALSE
Curricular units 1st sem (enrolled)	TRUE	FALSE
Curricular units 1st sem (evaluations)	TRUE	FALSE
Curricular units 1st sem (approved)	TRUE	FALSE
Curricular units 1st sem (grade)	TRUE	FALSE
Curricular units 2nd sem (credited)	TRUE	FALSE
Curricular units 2nd sem (enrolled)	TRUE	FALSE
Curricular units 2nd sem (evaluations)	TRUE	FALSE
Curricular units 2nd sem (approved)	TRUE	FALSE
Curricular units 2nd sem (grade)	TRUE	FALSE

Source: Authors, (2025).

These results reinforce the effectiveness of the proposed phased feature selection framework. The integration between filter-based methods in the early stages and wrapper-based validation in the advanced stages has been proven to be able to significantly reduce the dimensions of features, while retaining attributes that have a real contribution to the classification process. A subset of features obtained from this stage is then used as input for predictive model training at a later stage.

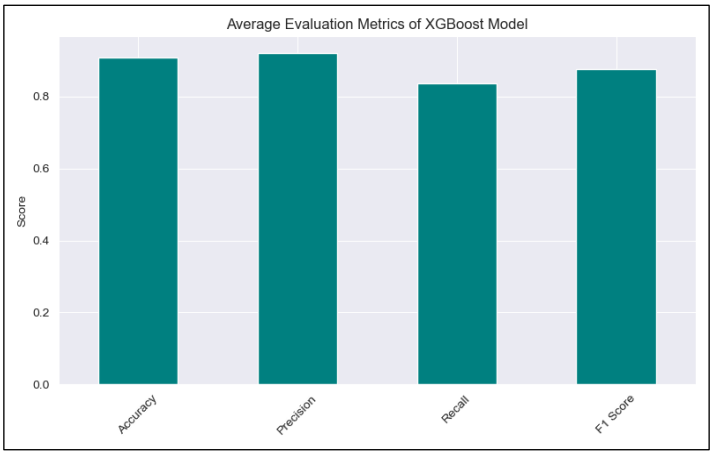


Figure 5: The average score of the XGBoost model evaluation metric is based on cross-validation.
Source: Authors, (2025).

The final classification model in this study was constructed using the Extreme Gradient Boosting (XGBoost) algorithm, with inputs in the form of features that have been selected through a gradual approach based on Variance Threshold, Mutual Information, and Boruta. To evaluate the performance of the generalization of the model, a 10-fold Stratified K-Fold Cross-Validation technique was used, which maintained a balanced proportion of classes in each iteration of training and testing.

Table 4: The average value and standard deviation of the XGBoost evaluation metric.

Metric	Average	Standard Deviation
Accuracy	0.9077	0.0128
Precision	0.9207	0.0185
Recall	0.8368	0.0318
F1 Score	0.8763	0.0186

Source: Authors, (2025).

The results of the model performance evaluation are shown in Figure 5 and summarized in Table 4. Based on an average value of 10 folds, the XGBoost model achieves an accuracy of 90.77%, with a precision of 92.07%, a recall of 83.68%, and an F1-score of 87.63%. Lower-than-precision recall metrics indicate that although the model is highly accurate in identifying dropout students, there are still a number of undetected cases of dropouts (false negatives).

However, a fairly high F1 score indicates a good balance between precision and sensitivity of the model. The stability of the metric reflected in the relatively low standard value of the deviation reinforces the validity of the model evaluation results. With this performance, the XGBoost model shows high effectiveness in classifying potential student dropouts, while showing that the step-by-step feature selection framework used contributes to the achievement of optimal classification results.

Table 5: Comparison with previous research.

Research	Model	Accuracy	Precision	Recall	F1 Score
Proposed	XGB	90.77	92.07	83.68	87.63
[24]	XGB	87.00	88.00	74.00	81.00
[33]	ANN	77.30	76.00	77.30	76.20

Source: Authors, (2025).

The results of the evaluation of this study show that the XGBoost model built through a phased feature selection approach provides the best performance compared to the previous two studies. With an accuracy of 90.77%, precision of 92.07%, and an F1-score of 87.63%, this model shows significant advantages, especially in identifying students who are at high risk of dropping out. Compared to the reseacrh [24] that also used XGBoost on the same dataset, the model in this study resulted in an improvement in all evaluation metrics, specifically the recall which increased from 74% to 83.68%, which is critical for early detection of dropout cases. This difference can be attributed to the gradual feature selection process and consistent use of Stratified K-Fold.

Meanwhile, Research [33] that used a multiclass classification approach with ANN only achieved an accuracy of 77.3% and an F1-score of 76.2%. Although ANN is known to be able to handle data complexity, these results suggest that XGBoost with optimal feature selection can excel in the context of binary dropout prediction. This comparison confirms that the design of predictive models that focus not only on algorithm selection, but also on feature selection and validation strategies, plays a key role in producing precise and reliable systems in higher education.

IV. CONCLUSIONS

This study proposes a multi-stage feature selection framework to improve the accuracy of student dropout predictions in higher education. The proposed approach consists of three main stages: initial screening using Variance Threshold and Mutual Information, validation of relevant features through the Boruta algorithm, and training of classification models using Extreme Gradient Boosting (XGBoost). The results of the experiment showed that this phased approach effectively filtered out uninformative features and retained attributes that had a significant contribution to the prediction target. Boruta successfully confirmed as many as 18 features as relevant, reflecting the importance of academic, demographic, administrative, and socioeconomic factors in determining student dropout risk.

The XGBoost model trained using selected features showed superior performance, with an accuracy of 90.77%, an F1-score of 87.63%, and a precision of 92.07%. Evaluation using 10-fold Stratified Cross-Validation yields stable metrics, indicating a good generalization of new data. The results of this study confirm that the combination of filter-based and wrapper-based feature selection methods can significantly improve classification performance in the context of predicting student dropout risk. For further research, this approach can be further developed by exploring the selection of the most optimal feature normalization method, as the proper normalization process has the potential to significantly affect model performance. Meta-learning strategies can be used to automate the selection of normalization schemes that best suit the characteristics of the educational data, thus supporting the generalization of predictive models in a more adaptive and contextual manner.

V. AUTHOR'S CONTRIBUTION

Conceptualization: Arif Mudi Priyatno, Yunia Ningsih, and Rizqon Jamil Farhas.
Methodology: Arif Mudi Priyatno, Fahmi Iqbal Firmananda, and Resy Kumala Sari.
Investigation: Arif Mudi Priyatno, and Aryadi.
Discussion of results: Arif Mudi Priyatno, Yunia Ningsih, Fahmi Iqbal Firmananda, and Rizqon Jamil Farhas..
Writing – Original Draft: Arif Mudi Priyatno.
Writing – Review and Editing: Arif Mudi Priyatno and Resy Kumala Sari.
Resources: Arif Mudi Priyatno.
Supervision: Arif Mudi Priyatno.
Approval of the final text: Arif Mudi Priyatno, Yunia Ningsih, Rizqon Jamil Farhas, Fahmi Iqbal Firmananda, Resy Kumala Sari, and Aryadi.

VI. REFERENCES

[1] O. Lorenzo-Quiles, S. Galdón-López, and A. Lendínez-Turón, "Factors contributing to university dropout: a review," Front. Educ., vol. 8, Mar. 2023, doi: 10.3389/feduc.2023.1159864.

[2] R. Goran et al., "Identifying and Understanding Student Dropouts Using Metaheuristic Optimized Classifiers and Explainable Artificial Intelligence Techniques," IEEE Access, vol. 12, pp. 122377–122400, 2024, doi: 10.1109/ACCESS.2024.3446653.

[3] S. Kim, E. Choi, Y.-K. Jun, and S. Lee, "Student Dropout Prediction for University with High Precision and Recall," Appl. Sci., vol. 13, no. 10, p. 6275, May 2023, doi: 10.3390/app13106275.

[4] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka, and P. C. Nshimyumukiza, "Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization," Comput. Educ. Artif. Intell., vol. 3, p. 100066, 2022, doi: 10.1016/j.caeai.2022.100066.

[5] É. Terrin and M. Triventi, "The Effect of School Tracking on Student Achievement and Inequality: A Meta-Analysis," Rev. Educ. Res., vol. 93, no. 2, pp. 236–274, Apr. 2023, doi: 10.3102/00346543221100850.

[6] P. Rodríguez, A. Villanueva, L. Dombrovskaia, and J. P. Valenzuela, "A methodology to design, develop, and evaluate machine learning models for predicting dropout in school systems: the case of Chile," Educ. Inf. Technol., vol. 28, no. 8, pp. 10103–10149, Aug. 2023, doi: 10.1007/s10639-022-11515-5.

[7] M. N. Yakubu and A. M. Abubakar, "Applying machine learning approach to predict students' performance in higher educational institutions," Kybernetes, vol. 51, no. 2, pp. 916–934, Feb. 2022, doi: 10.1108/K-12-2020-0865.

[8] B. Albreiki, N. Zaki, and H. Alashwal, "A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques," Educ. Sci., vol. 11, no. 9, p. 552, Sep. 2021, doi: 10.3390/educsci11090552.

[9] L. R. Pelima, Y. Sukmana, and Y. Rosmansyah, "Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review," IEEE Access, vol. 12, pp. 23451–23465, 2024, doi: 10.1109/ACCESS.2024.3361479.

[10] J. Kabathova and M. Drlik, "Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques," Appl. Sci., vol. 11, no. 7, p. 3130, Apr. 2021, doi: 10.3390/app11073130.

[11] R. L. S. do Nascimento, R. A. de A. Fagundes, and R. M. C. R. de Souza, "Statistical Learning for Predicting School Dropout in Elementary Education: A Comparative Study," Ann. Data Sci., vol. 9, no. 4, pp. 801–828, Aug. 2022, doi: 10.1007/s40745-021-00321-4.

[12] H. Nguyen Thi Cam, A. Sarlan, and N. I. Arshad, "A hybrid model integrating recurrent neural networks and the semi-supervised support vector machine for identification of early student dropout risk," PeerJ Comput. Sci., vol. 10, p. e2572, Nov. 2024, doi: 10.7717/peerj-cs.2572.

[13] D. A. Shafiq, M. Marjani, R. A. A. Habeeb, and D. Asirvatham, "Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review," IEEE Access, vol. 10, pp. 72480–72503, 2022, doi: 10.1109/ACCESS.2022.3188767.

[14] C. Aliferis and G. Simon, "Overfitting, Underfitting and General Model Overconfidence and Under-Performance Pitfalls and Best Practices in Machine Learning and AI," 2024, pp. 477–524. doi: 10.1007/978-3-031-39355-6_10.

[15] D. Theng and K. K. Bhoyar, "Feature selection techniques for machine learning: a survey of more than two decades of research," Knowl. Inf. Syst., vol. 66, no. 3, pp. 1575–1637, Mar. 2024, doi: 10.1007/s10115-023-02010-5.

[16] P. Dhal and C. Azad, A comprehensive survey on feature selection in the various fields of machine learning, vol. 52, no. 4. Applied Intelligence, 2022. doi: 10.1007/s10489-021-02550-9.

[17] A. M. Priyatno and T. Widiyaningtyas, "A SYSTEMATIC LITERATURE REVIEW: RECURSIVE FEATURE ELIMINATION ALGORITHMS," JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer), vol. 9, no. 2, pp. 196–207, Feb. 2024, doi: 10.33480/jitk.v9i2.5015.

[18] M. Zaffar, M. A. Hashmani, K. S. Savita, and S. A. Khan, "A review on feature selection methods for improving the performance of classification in educational data mining," Int. J. Inf. Technol. Manag., vol. 20, no. 1/2, p. 110, 2021, doi: 10.1504/IJITM.2021.114161.

[19] D. Effrosynidis and A. Arampatzis, "An evaluation of feature selection methods for environmental data," Ecol. Inform., vol. 61, p. 101224, Mar. 2021, doi: 10.1016/j.ecoinf.2021.101224.

[20] C. Kuzudisli, B. Bakir-Gungor, N. Bulut, B. Qaqish, and M. Yousef, "Review of feature selection approaches based on grouping of features," PeerJ, vol. 11, 2023, doi: 10.7717/peerj.15666.

[21] S. M. F. D. Syed Mustapha, "Predictive Analysis of Students' Learning Performance Using Data Mining Techniques: A Comparative Study of Feature Selection Methods," Appl. Syst. Innov., vol. 6, no. 5, p. 86, Sep. 2023, doi: 10.3390/asi6050086.

[22] M. V. Martins, D. Tolledo, J. Machado, L. M. T. Baptista, and V. Realinho, "Early Prediction of student's Performance in Higher Education: A Case Study," in Trends and Applications in Information Systems and Technologies, 2021, pp. 166–175. doi: 10.1007/978-3-030-72657-7_16.

[23] A. M. Priyatno, L. Ningsih, and M. Noor, "Harnessing Machine Learning for Stock Price Prediction with Random Forest and Simple Moving Average Techniques," J. Eng. Sci. Appl., vol. 1, no. 1, pp. 1–8, Mar. 2024, doi: 10.69693/jesa.v1i1.1.

[24] A. Ridwan and A. M. Priyatno, "Predict Students' Dropout and Academic Success with XGBoost," J. Educ. Comput. Appl., vol. 1, no. 2, pp. 1–8, Dec. 2024, doi: 10.69693/jeca.v1i2.13.

[25] M. Saied, S. Guirguis, and M. Madbouly, "Review of filtering based feature selection for Botnet detection in the Internet of Things," Artif. Intell. Rev., vol. 58, no. 4, p. 119, Jan. 2025, doi: 10.1007/s10462-025-11113-0.

[26] H. Zhou, X. Wang, and R. Zhu, "Feature selection based on mutual information with correlation coefficient," Appl. Intell., vol. 52, no. 5, pp. 5457–5474, Mar. 2022, doi: 10.1007/s10489-021-02524-x.

[27] R. Iranzad and X. Liu, "A review of random forest-based feature selection methods for data science education and applications," Int. J. Data Sci. Anal., Feb. 2024, doi: 10.1007/s41060-024-00509-w.

[28] I. Nurzari, E. Sari, D. I. Harris, A. M. Priyatno, and H. Rusnedy, "Inter-Cluster Distance-Based SMOTE Modification for Enhanced Diabetes Classification," ITEGAM-J. Eng. Technol. Ind. Appl. (ITEGAM-JETIA, vol. 11, no. 51, pp. 190–196, 2025, doi: 10.5935/jetia.v11i51.1453.

[29] C. Bentéjac, A. Csörgő, and G. Martínez-Muñoz, "A comparative analysis of gradient boosting algorithms," Artif. Intell. Rev., vol. 54, no. 3, pp. 1937–1967, Mar. 2021, doi: 10.1007/s10462-020-09896-5.

[30] A. M. Priyatno, "SPAMMER DETECTION BASED ON ACCOUNT, TWEET, AND COMMUNITY ACTIVITY ON TWITTER," J. Ilmu Komput. dan Inf., vol. 13, no. 2, pp. 97–107, Jul. 2020, doi: 10.21609/jiki.v13i2.871.

[31] A. M. Priyatno and F. I. Firmananda, "N-Gram Feature for Comparison of Machine Learning Methods on Sentiment in Financial News Headlines," RIGGS J. Artif. Intell. Digit. Bus., vol. 1, no. 1, pp. 01–06, Jul. 2022, doi: 10.31004/riggs.v1i1.4.

[32] M. R. A. Prasetya and A. M. Priyatno, "Dice Similarity and TF-IDF for New Student Admissions Chatbot," RIGGS J. Artif. Intell. Digit. Bus., vol. 1, no. 1, pp. 13–18, Jul. 2022, doi: 10.31004/riggs.v1i1.5.

[33] S. A. SULAK and N. KOKLU, "Predicting Student Dropout Using Machine Learning Algorithms," Intell. Methods Eng. Sci., vol. 44, no. 4, pp. 1519–1532, Jan. 2025, doi: 10.58190/imiens.2024.103.

Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11 n.55, p. 96-104. September/October., 2025.

DOI: https://doi.org/10.5935/jetia. v11i55.1913

RESEARCH ARTICLE OPEN ACCESS

AI FOR POSITIONING ACCURACY ENHANCEMENT IN THE IIOT

Sofiane Latreche[1], Hocine Bellahsene[2]

[1,2]Université de Bejaia, Faculté de Technologie, Laboratoire de Génie Electrique, 06000 Bejaia, Algeria.

[1]https://orcid.org/0009-0009-1460-0115, [2]https://orcid.org/0000-0002-9143-4052

Email: sofiane.latreche@univ-bejaia.dz, hocine.bellahsene@univ-bejaia.dz

ARTICLE INFO

Article History
Received: April 22, 2025
Revised: May 20, 2025
Accepted: June 15, 2025
Published: September 30, 2025

Keywords:
IIOT,
5G/6G,
AI,
TDoA,
Positioning

ABSTRACT

The evolution of mobile networks, accelerated by the deployment of 5G and the upcoming shift toward 6G, has unlocked new possibilities for industrial applications. A key beneficiary of this progress is the Industrial Internet of Things (IIoT), which depends heavily on precise positioning to support automation, ensure operational safety, and enable real-time system monitoring. Despite these advancements, achieving reliable localization in complex indoor settings remains a persistent challenge for traditional techniques like Time Difference of Arrival (TDoA), which are vulnerable to signal reflections and obstructions caused by multipath propagation. To address these limitations, this study introduces an AI-powered localization method built on a deep learning framework using ResNet architecture. By integrating and analyzing data from multiple sources, the proposed solution overcomes the inherent weaknesses of conventional approaches, delivering enhanced accuracy and resilience in densely packed industrial environments. Comprehensive simulations in realistic indoor factory settings confirm the superiority of the AI-driven model over TDoA, with notable improvements in positioning accuracy. These results underscore the promise of deep learning for advancing IIoT localization, marking a significant step toward intelligent positioning systems in future 5G and 6G networks.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The rapid evolution of mobile communication technologies has reached an advanced stage with the rollout of 5G and the ongoing research directed at 6G networks [1]. These next-generation systems are engineered not only to meet increasing demands for higher data throughput and ultra-low latency but also to accommodate a vast array of applications across industrial sectors. Among these, the Industrial Internet of Things (IIoT) is particularly reliant on accurate positioning to ensure efficient operations, improved safety, and real-time asset tracking within intricate environments like factories and warehouses [2]. While existing localization methods have achieved moderate success, the complexity of indoor IIoT environments continues to challenge their effectiveness. In this scenario, Artificial Intelligence (AI) presents a promising path toward enhancing localization accuracy in 5G and forthcoming 6G ecosystems [3].

Fifth-generation mobile networks have redefined benchmarks in speed, latency, and device connectivity. Offering peak data rates up to 10 Gbps, sub-millisecond latency, and support for massive machine-type communications, 5G enables revolutionary use cases, including smart infrastructure, autonomous systems, and telehealth services [4]. As research advances toward 6G, expectations include data transmission rates approaching 1 Tbps and latency shrinking to microsecond levels [5]. This next wave of mobile networks is poised to support sophisticated technologies such as immersive XR, real-time digital twins, and holographic communications, all of which demand exceptional reliability and sub-centimeter localization precision [6]. Consequently, there's an escalating demand for innovative localization approaches that can perform under the complex conditions found in modern industrial and urban environments [7].

As 5G and 6G architectures become increasingly intricate, AI has emerged as an important enabler for intelligent network management. Its capabilities in automating decision-making processes, minimizing operational overhead, and dynamically optimizing system performance make it well-suited for handling the challenges posed by high user densities, unpredictable traffic behavior, and

diverse quality-of-service requirements [8]. Traditional static optimization methods struggle to keep up with such dynamic and complex environments.

In the realm of localization, AI-based methods offer distinct advantages over classical techniques. Conventional strategies like Time Difference of Arrival (TDoA) and Angle of Arrival (AoA) typically depend on direct line-of-sight (LOS) conditions and highly accurate signal timing or direction estimation [9]. However, in indoor industrial spaces, where obstructions, reflections, and interference are prevalent, these methods often experience significant degradation. In contrast, AI-powered models can learn from complex environmental data and adjust to varying conditions, leading to more robust and precise localization performance [10]. As accurate positioning becomes a cornerstone for IIoT and emerging applications, AI offers a viable and scalable solution.

This paper introduces an AI-driven localization framework aimed at enhancing positioning accuracy in indoor IIoT environments. Industrial facilities are particularly challenging due to obstacles like dense infrastructure, heavy machinery, and strong electromagnetic interference. High-precision positioning in such settings is essential for tasks such as automated navigation, asset localization, and safety enforcement. Our approach leverages a deep learning model, specifically, a ResNet-based architecture, to extract and learn from multi-source signal features and spatial patterns that are often overlooked by conventional algorithms.

To validate our method, we benchmark its performance against the widely used TDoA localization technique, which estimates user equipment (UE) positions based on signal arrival time differences from various base stations. Although effective in open settings with accurate synchronization and unobstructed LOS, TDoA suffers from reduced accuracy in indoor environments due to multipath effects and signal blockage. In this study, both approaches are tested under identical simulation parameters that replicate a realistic industrial factory setup. Evaluation metrics such as cumulative error distributions and percentile-based accuracy comparisons demonstrate that our AI model consistently outperforms TDoA, offering superior precision and robustness. The AI-enhanced solution proves resilient to environmental changes and is thus a promising candidate for future high-precision IIoT positioning systems.

The remainder of this paper is organized as follows: Section 2 reviews relevant literature; Section 3 explores the role of AI in mobile network communications; Section 4 outlines the proposed methodology and system model; Section 5 presents simulation results and analysis; and finally, Section 6 concludes the paper.

II. The LITERATURE REVIEW

Here we'll have a look at the work done in the field of AI for Mobiles network Communications. Paper [11] examines the critical role of technology in managing the global COVID-19 outbreak, emphasizing the importance of IoT, AI, and 5G technologies in mitigating the virus's impact. The discussion highlights areas where these technologies have proven essential, including risk reduction, healthcare support, and facilitating public compliance with safety measures. The survey explores how IoT, AI, and 5G contribute to health improvements and ensure accuracy in drug and medicine delivery. By reviewing current applications, the paper also envisions the future potential of these technologies to enable more advanced and secure healthcare solutions. Paper [12] examines the integration of edge computing with deep learning to enhance mobile network services. The proposed "In-Edge AI" framework combines Deep Reinforcement Learning (DRL) and Federated Learning (FL) to optimize mobile edge computing, caching, and communication.

By leveraging device and edge node collaboration to exchange learning parameters, this framework aims to improve model training and inference, enabling dynamic optimization at both system and application levels while minimizing communication overhead. Evaluation results indicate that "In-Edge AI" achieves near-optimal performance with low learning overhead and adaptability to mobile systems, positioning it as a promising solution for future mobile communication challenges and advancements. Paper [13] explores the critical role of 5G networks as foundational infrastructure for various industries, including IoT, smart cities, and virtual reality.

Distinguishing 5G from previous generations, it incorporates advanced digital technologies like massive MIMO (mMIMO) and operates at higher radio frequencies, posing new challenges for network operators. Integrating Artificial Intelligence (AI) offers promising solutions to address these complexities, yet it also introduces security concerns that must be resolved for standardization. To enhance AI-enabled 5G networks, this work proposes integrating Blockchain technology, which facilitates secure information sharing across 5G nodes. The paper introduces "Block5GIntell," a comprehensive framework that leverages the convergence of Blockchain and AI to create intelligent, efficient, and secure 5G networks. Through a case study on energy savings, the proposed solution demonstrates a 20\% reduction in energy consumption at the RAN level, underscoring Blockchain's potential to support AI-driven 5G advancements. Paper [14] addresses the increasing demand for high-quality mobile services across various industries, as 5G incorporates technologies like machine learning and cloud computing into its ecosystem.

The 5Growth project seeks to advance these efforts by automating support for vertical use cases through several innovations: a portal connecting industry-specific applications to 5G platforms, a multi-domain service orchestrator, and a resource management layer. Additionally, the project emphasizes closed-loop, machine-learning-based SLA control and end-to-end optimization. Key 5Growth innovations include advancements in radio slicing, enhanced monitoring and analytics, and machine learning integration to streamline and elevate 5G capabilities for vertical industries. Paper [15] explores the transformative impact of smart tourism on the tourism industry, driven by advancements in communication and information technologies. It highlights the essential role of the Internet of Things (IoT) in smart tourism, despite challenges related to managing vast data and ensuring low-latency communication. To address these issues, the paper proposes 5G- and AI-enhanced IoT systems designed for efficient data transmission and intelligent processing. A case study on Points of Interest (POI) recommendation demonstrates the superior performance of this approach, showcasing its effectiveness and efficiency in enhancing IoT-based smart tourism applications.

Paper [16] investigates the application of Artificial Intelligence (AI) beyond the wireless Physical Layer (PHY), focusing on Machine Learning (ML) techniques for the Medium Access Control (MAC) layer. Unlike the PHY, the MAC layer encompasses diverse features that require tailored ML approaches. This survey explores recent advancements in AI-driven MAC functions, including resource allocation, random access, Adaptive Modulation and Coding (AMC), power control, protocol learning, Channel State Information (CSI) reporting, Hybrid Automatic Repeat Request (HARQ), and Multi-RAT Spectrum Sharing (MRSS), highlighting their potential for optimizing wireless communication systems.

III. AI AND POSITIONING ACCURACY ENHANCEMENT

III.1 ARTIFICIAL INTELLIGENCEQPSK MODULATION

Artificial Intelligence (AI) has become a transformative technology, reshaping multiple aspects of modern life [17]. AI refers to the capability of machines to exhibit intelligence similar to that of humans, encompassing tasks such as learning, problem-solving, decision-making, and natural language processing. Recent advancements, particularly in machine learning (ML) and deep learning, have significantly enhanced AI's capabilities. The combination of vast datasets, increased computational power, and advanced algorithms has enabled AI systems to surpass human performance in certain fields, such as image recognition and speech processing.
AI development follows three primary trends:

- Machine Learning and Deep Learning: Machine learning, a fundamental subset of AI, has been the driving force behind many recent innovations. Within ML, deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have achieved state-of-the-art results in areas like image classification, natural language processing (NLP), and reinforcement learning. The strength of deep learning lies in its ability to learn hierarchical representations of data. For instance, in image recognition, lower CNN layers detect basic features like edges and textures, while deeper layers recognize complex structures, such as objects and faces. This hierarchical learning approach allows deep models to generalize well, provided they are trained on diverse and extensive datasets [18].
- Reinforcement Learning: Reinforcement learning (RL) has gained prominence for its ability to train AI agents to navigate dynamic environments with evolving rules and feedback mechanisms. One of the most well-known RL successes is AlphaGo, an AI system developed by DeepMind that surpassed human champions in the strategic game of Go—an achievement previously thought to be beyond AI's capabilities. RL operates by enabling an agent to interact with an environment, learn from feedback, and develop an optimal policy to maximize long-term rewards. This learning paradigm has been effectively applied in fields such as robotics, autonomous vehicle control, and game playing [19].
- Natural Language Processing: The field of natural language processing (NLP) has advanced significantly, particularly with the rise of transformer-based architectures such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer). These models have revolutionized tasks like translation, summarization, sentiment analysis, and creative writing. Unlike traditional sequence-based models, such as RNNs and LSTMs (Long Short-Term Memory networks), transformers excel at capturing long-range dependencies in text. The introduction of large-scale pre-training on extensive text corpora has enabled these models to be fine-tuned for specific NLP tasks, often achieving performance comparable to or even surpassing that of humans [20].

III.2 AI FOR POSITIONING ACCURACY ENHANCEMENT

In Release 18, the 3rd Generation Partnership Project (3GPP) explored the potential integration of Artificial Intelligence (AI) within the New Radio (NR) air interface, as detailed in Technical Report (TR) 38.843. This report identifies three key applications of AI in NR: improving channel state information (CSI) feedback, optimizing beam management, and enhancing positioning accuracy [21]. Traditional positioning techniques, such as Angle of Arrival (AoA) and Time Difference of Arrival (TDoA), depend on line-of-sight (LoS) propagation and often perform poorly in non-line-of-sight (NLoS) environments. Due to these limitations, conventional methods may fail to meet the stringent positioning requirements of 5G and future 6G networks. To address these challenges, AI and machine learning (AI/ML) techniques present innovative solutions for improving positioning accuracy in NR systems, particularly under complex propagation conditions [22]. AI/ML-based positioning can be implemented through two main approaches [23]:

- Direct AI/ML Positioning: As illustrated in Figure 1, this approach leverages AI/ML models to directly estimate the user equipment (UE) location through inference. By relying on data-driven learning, these models eliminate some of the inherent constraints of conventional positioning algorithms.

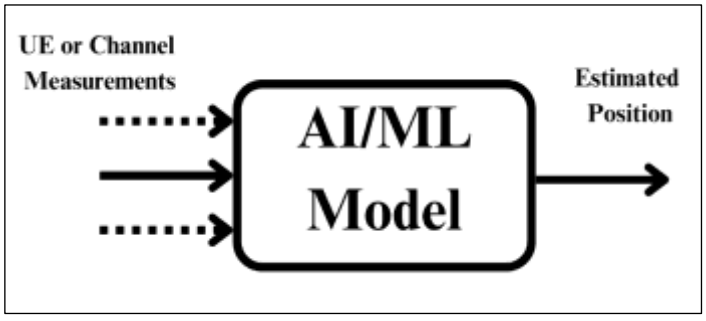


Figure 1: Direct Positioning.
Source: Authors, (2025).

- AI/ML-Assisted Positioning: Shown in Figure 2, this method employs AI/ML models to generate intermediate estimates, such as identifying LoS/NLoS conditions or extracting angular and timing information. These outputs can then be integrated with either traditional positioning techniques or additional AI/ML models to achieve more accurate and resilient UE positioning.

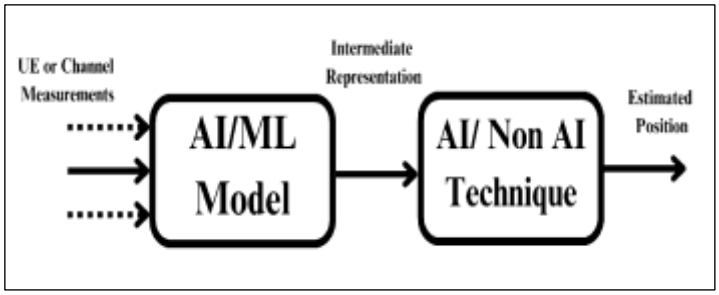


Figure 2: AI/ML-Assisted Positioning.
Source: Authors, (2025).

IV. METHODOLOGY AND SYSTEM MODEL

In this section we present the construction and training process of a Residual Network (ResNet) for enhanced user equipment (UE) positioning accuracy, particularly in non-line-of-sight (NLoS) conditions. The proposed approach leverages uplink (UL) sounding reference signal (SRS) channel estimation in simulation mode to determine UE locations. To achieve this, a deep convolutional neural network (DCNN) based on the ResNet architecture is implemented. Following the architectural configuration, the network undergoes training in a supervised learning framework or utilizes UL SRS-based channel estimation to infer UE positions in a simulated environment [24].

Deep neural networks have demonstrated significant capability in feature extraction and learning from high-dimensional data. However, increasing the network depth indiscriminately may lead to accuracy saturation and training inefficiencies, commonly referred to as the degradation problem. The ResNet architecture mitigates this issue by incorporating skip connections, which facilitate gradient propagation and improve convergence. These shortcut connections enable the direct transfer of information across layers, thereby enhancing network stability and learning efficiency.

In AI/ML-based positioning systems, channel measurements collected at the transmission and reception points serve as input features, while the output corresponds to the estimated two-dimensional (2D) coordinates of the UE. This formulation characterizes the positioning task as a regression problem. Unlike traditional image classification tasks, where patterns are visually interpretable, channel measurement-based data representations are often non-intuitive. This underscores the significance of deep learning models in extracting hidden features that correlate with UE positioning.

Due to its capacity for constructing deep yet computationally efficient models, ResNet has been widely adopted in localization applications [25]. To evaluate positioning performance, the ResNet model is tested in simulation mode using a pretrained network and a generated test dataset comprising samples. Conversely, in training mode, a customized ResNet architecture is designed, and training is performed using TrainRatio of the total dataset, with the remaining samples reserved for validation. The validation set plays an essential role in monitoring model performance and mitigating overfitting [26].

The final pretrained ResNet model, consisting of eight layers, is stored within the working directory. As part of the preprocessing step, the channel state information (CSI) images are resized to dimensions $32 \times 32 \times 18$ and normalized to a range of [0,1] before being fed into the network. In this study, we focus on an Industrial Internet of Things (IIoT) scenario to assess the performance of AI/ML-based positioning methods. The process implemented in our model is illustrated in Figure 3.

In this context, achieving accurate positioning relies on synchronized timing and channel estimation, facilitated by uplink (UL) sounding reference signals (SRS). By simulating UL SRS transmissions, we extract channel parameters to build a labeled dataset containing 2D user equipment (UE) locations, which are then used for AI/ML model training. The 3GPP Technical Report 38.901 standard [27] defines various indoor factory (InF) environments, each designed to represent different levels of clutter and variations in transmitter (Tx) and receiver (Rx) antenna heights. These environments are categorized as follows:

- InF-SL: Represents a sparsely cluttered indoor factory with a low base station height, where both Tx and Rx are positioned below the average clutter height.
- InF-DL: Depicts an indoor factory with dense clutter and a low base station height, where Tx and Rx remain below the clutter level.
- InF-SH: Models a setting with sparse clutter but a high base station height, where either the Tx or Rx is elevated above the surrounding obstacles.
- InF-DH: Represents a densely cluttered indoor factory with a high base station height, featuring an elevated Tx or Rx positioned above the clutter.
- InF-HH: A scenario where both Tx and Rx are elevated, positioned higher than the surrounding clutter elements.

Our uplink simulation operates through multiple stages like presented in figure 3:

- Scenario Configuration: Defines environmental characteristics, channel parameters, and Tx/Rx deployment settings.
- OFDM Transmission: Sends SRS through the predefined channel.
- Channel Processing: Applies channel-specific filtering and incorporates additive white Gaussian noise (AWGN).
- OFDM Demodulation: Conducts timing synchronization and estimates the channel response.
- Position Estimation: Integrates AI/ML techniques with Time Difference of Arrival (TDoA) algorithms to determine UE location.

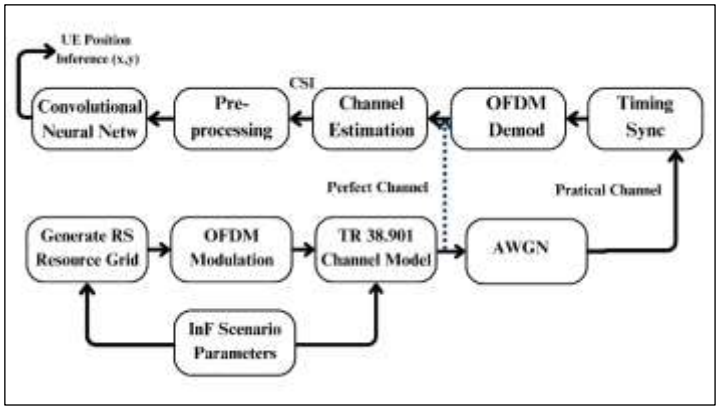


Figure 3: The implemented processing chain.
Source: Authors, (2025).

V. SIMULATION RESULTS AND DISCUSSIONS

To assess the effectiveness of the ResNet-based AI approach in enhancing positioning accuracy, we conducted simulations using MATLAB. These simulations aimed to replicate a realistic indoor environment by modeling multiple User Equipment (UE) positions and signal propagation characteristics. For the indoor environment, we adopted the factory scenarios defined in 3GPP TR 38.901. Specifically, we simulated a hall measuring 120 × 60 × 20 meters, incorporating clutter elements with a density of 0.6 and an average size of 2 meters. To introduce complexity in signal propagation, the clutter height was set at 6 meters, creating challenging conditions for positioning accuracy due to multipath effects and signal obstructions. The key parameters of the simulation were as follows:

- Base Station (BS): Deployed at intervals of 20×20 meters, featuring antennas positioned at an 8-meter height and a noise figure of 5 dB.
- User Equipment (UE): A total of 20 UEs were randomly distributed within the environment, each equipped with antennas at a height of 1.5 meters.
- Carrier and SRS Configuration: The subcarrier spacing was fixed at 30 kHz, with a carrier grid size of 270. The Sounding Reference Signal (SRS) was optimized using 12 OFDM symbols and a comb number of 8 to enhance positioning precision.

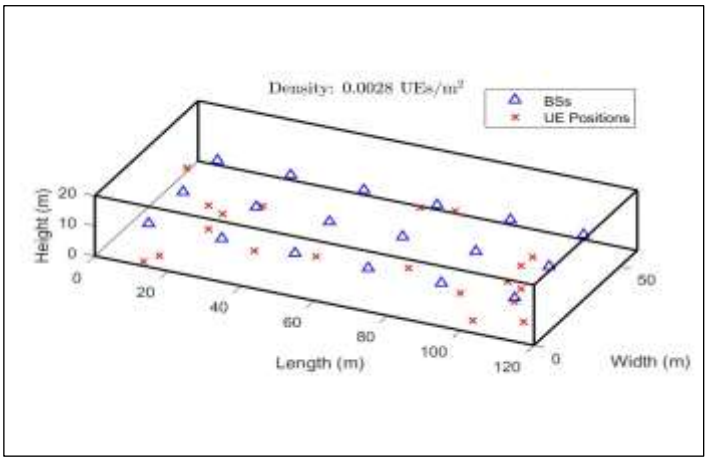


Figure 4: 3D representation of an indoor factory hall scenario with 20 UEs.
Source: Authors, (2025).

Figure 4, shows a 3D representation of an indoor factory hall scenario or 20 UE, used for positioning analysis. Here's a breakdown of the key elements:

- Hall Dimensions: The plot shows a hall with the dimensions set to approximately 120 meters in length, 60 meters in width, and 20 meters in height, which is typical for a dense indoor factory setting. This scale is realistic for testing positioning accuracy in large indoor spaces.
- UE (User Equipment) Density: The density of UEs is noted as 0.0028 UEs/m², indicating a sparse distribution of users within the hall. This density helps in examining how well the system performs with limited user clustering and interference.
- BS (Base Station) Positions: The blue triangles represent the locations of the base stations (BSs) placed throughout the hall at strategic heights. These BSs are crucial in providing signal coverage for positioning and represent points where data is transmitted or received for determining user positions.
- UE Positions: The red crosses mark the positions of the UEs within the hall. The random placement of these UEs reflects a realistic distribution of users in a factory setting where workers or equipment might be spread across the area.

We estimate UE positions with a positioning simulation that includes both the proposed AI and TDoA approaches. The estimate procedure consists of producing the resource grid, synthesizing the waveform, modeling the noisy channel, and conducting synchronization and channel estimation. For AI-based positioning, we use perfect channel estimations to determine the UE's position with a pretrained network. In TDoA-based positioning, TRPs calculate timing offsets from received SRSs and use these estimates to identify the UE, with the option of employing perfect or realistic timing approximations. By default, we assume perfect timing estimation.

Table 1: Neural Network Summary.

Parameter	Value
Number of layers	8
Complexity	$1.2816\ million$
Input size	$[32, 32, 18]$
Output size	$[2]$

Source: Authors, (2025).

The table 1 provides a concise overview of the ResNet architecture used for AI/ML-based positioning. The key details include:

- Number of layers: The network consists of 8 layers, making it relatively deep while maintaining efficiency.
- Complexity: The model has **1.2816 million parameters**, indicating its capacity to learn complex spatial relationships in positioning.
- Input size: The input dimensions are [32, 32, 18], representing the structure of the input feature maps, likely derived from channel state information (CSI) or other positioning-related data.
- Output size: The network predicts (2) values, which likely correspond to the x and y coordinates of the User Equipment (UE) position.

The figure 5 illustrates the structure of the custom ResNet model used for UE positioning. The key elements include:

- Input Layer: Takes in the CSI-based input representation.
- Convolutional Layers (conv, bn, relu): Feature extraction is performed using convolutional layers, batch normalization (bn), and activation functions (ReLU).
- Residual Connections (shortcut, add): These help in maintaining gradient flow and preventing vanishing gradient issues, making deep networks trainable.
- Global Average Pooling (globavgpool) and Fully Connected Layer (fc): These layers transform extracted features into final positioning predictions.

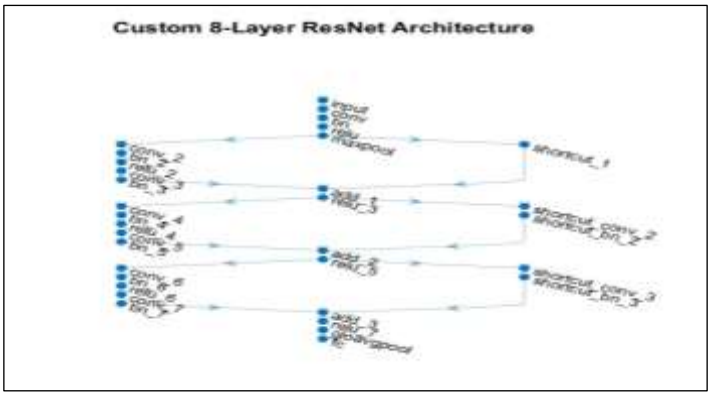


Figure 5: The structure of the custom ResNet mode.
Source: Authors, (2025).

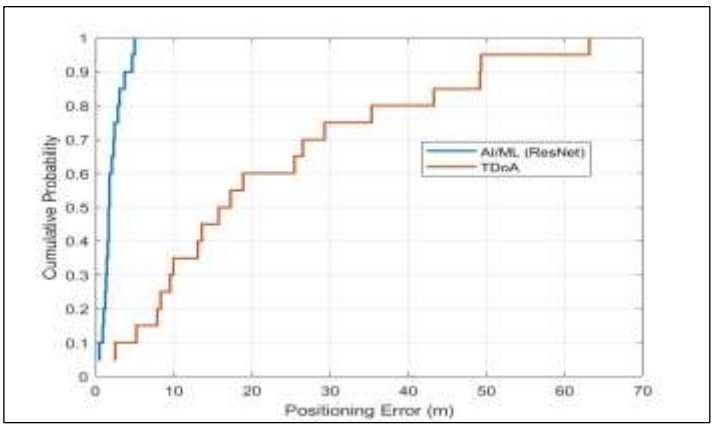


Figure 6: The CDF of positioning error for InF-DH environments with 20 UEs.
Source: Authors, (2025).

Figure 6, illustrates the Cumulative Distribution Function (CDF) of positioning error with 20 UE, for two positioning methods: the proposed AI method and TDoA method for each indoor factory (InF) environments.

- X-axis (Positioning Error in Meters): The x-axis represents the positioning error in meters, showing the accuracy of each method. Lower values indicate better positioning accuracy.
- Y-axis (Cumulative Probability): The y-axis shows the cumulative probability, indicating the proportion of samples with positioning error up to the value on the x-axis. A steeper curve indicates higher accuracy.

The blue line represents the AI based method. It shows a steep curve with most errors concentrated below 10 meters. This suggests that the AI model achieves high positioning accuracy, as a large portion of the cumulative probability reaches close to 1 (or 100%) with minimal error. The orange line represents the traditional TDoA method, which has a more gradual slope and reaches higher positioning errors. This method performs less accurately, with positioning errors extending up to around 60 meters before reaching 100% cumulative probability. The table 2 present the Positioning Errors at Different Percentiles.

Table 2: Comparison of AI/ML (ResNet) and TDoA Positioning Errors at Different Percentiles for 20 UE.

Percentile	AI/ML (ResNet) Positioning Error (m)	TDoA Positioning Error (m)
10th	0.6847	3.8438
50th	1.7341	16.508
90th	4.2152	49.243

Source: Authors, (2025).

Now we increase the number of UEs, to understand how positioning accuracy scales with crowding.
The figure 7, shows a 3D representation of an indoor factory hall scenario InF-DH with 200 UE, used for positioning analysis.

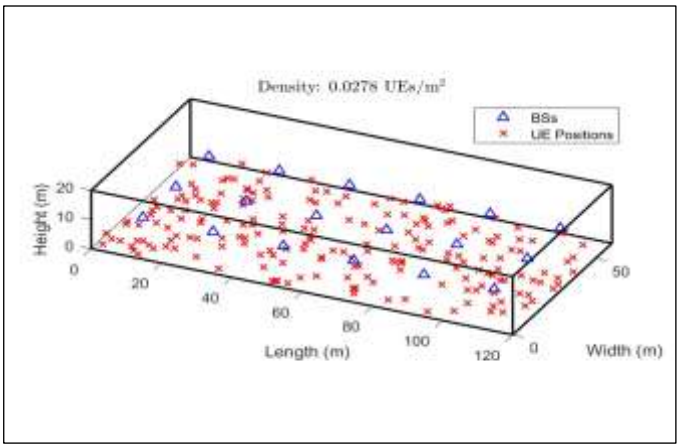


Figure 7: 3D representation of an indoor factory hall scenario with 200 UEs.
Source: Authors, (2025).

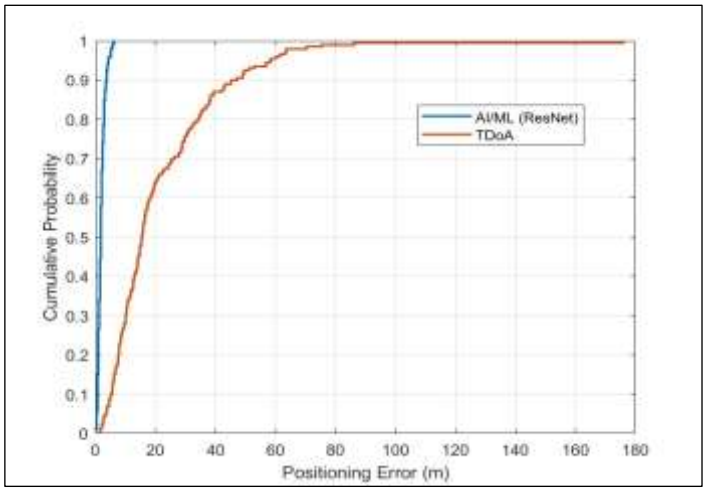


Figure 8: The CDF of positioning error for InF-DH environments with 200 UEs.
Source: Authors, (2025).

Figures 8, illustrates the CDF of positioning error for InF-DH environments with increasing the number of UEs to 200. The AI based method shows a steep curve with most errors concentrated around 5 meters. This suggests that the AI model achieves high positioning accuracy, as a large portion of the cumulative probability reaches close to 1 (or 100 %) with minimal error. The traditional TDoA method, which has a more gradual slope and reaches higher positioning errors. This method performs less accurately, with positioning errors extending up to around 100 meters before reaching 100 % cumulative probability.

The table 3 present the Positioning Errors at Different Percentiles for 200 UE.

Table 3: Comparison of AI/ML (ResNet) and TDoA Positioning Errors at Different Percentiles for 200 UE.

Percentile	AI/ML (ResNet) Positioning Error (m)	TDoA Positioning Error (m)
10th	0.66627	5.3964
50th	1.7679	15.47
90th	3.6627	46.126

Source: Authors, (2025).

In all cases, the AI approach outperforms TDoA by achieving a significantly lower error range, which means it provides more precise location estimates. The quick rise to near 100 % cumulative probability within a small positioning error range indicates that the AI method is more robust and reliable for positioning in complex indoor environments.

VI. CONCLUSIONS

This study highlights the potential of AI-driven positioning techniques in enhancing localization accuracy within complex industrial environments under 5G and emerging 6G networks. By leveraging a ResNet-based deep learning model, our approach consistently outperformed the traditional TDoA method across various indoor factory scenarios. The AI model demonstrated strong resilience to multipath interference and signal obstruction, maintaining high accuracy even in densely cluttered environments with diverse equipment positioning.

These findings confirm the effectiveness of deep neural networks in improving positioning accuracy, an essential factor for IIoT applications requiring precise monitoring, automation, and safety compliance. As mobile networks continue to evolve, integrating AI-based positioning solutions will be essential to meet the increasing demands of future industrial environments.

Future research could explore real-time AI implementations and hybrid approaches that combine multiple positioning techniques to further enhance the robustness and scalability of localization systems in 5G/6G industrial scenarios.

VII. REFERENCES

[1] Salahdine, F., Han, T., and Zhang, N. (2023). 5G, 6G, and Beyond: Recent advances and future challenges. Annals of Telecommunications, 78(9), 525-549.

[2] Serror, M., Hack, S., Henze, M., Schuba, M., and Wehrle, K. (2020). Challenges and opportunities in securing the industrial internet of things. IEEE Transactions on Industrial Informatics, 17(5), 2985-2996.

[3] Che, F., Ahmed, Q. Z., Lazaridis, P. I., Sureephong, P., and Alade, T. (2023). Indoor positioning system (ips) using ultra-wide bandwidth (uwb)—for industrial internet of things (iiot). Sensors, 23(12), 5710.

[4] Pons, M., Valenzuela, E., Rodríguez, B., Nolazco-Flores, J. A., and Del-Valle-Soto, C. (2023). Utilization of 5G technologies in IoT applications: Current limitations by interference and network optimization difficulties—A review. Sensors, 23(8), 3876.

[5] Maiwald, T., Li, T., Hotopan, G. R., Kolb, K., Disch, K., Potschka, J., ... and Visweswaran, A. (2023). A review of integrated systems and components for 6G wireless communication in the D-band. Proceedings of the IEEE, 111(3), 220-256.

[6] Alkhlefat, Y., Esmail, M. A., Ragheb, A. M., Idrus, S. M., Iqbal, F. M., and Alshebeili, S. A. (2024). Experimental Investigation of All-Optical Sub-THz Wavelength Conversion for 6G Communication Systems. IEEE Access.

[7] Edirisinghe, S., Galagedarage, O., Dias, I., and Ranaweera, C. (2023). Recent development of emerging indoor wireless networks towards 6G. Network, 3(2), 269-297.

[8] Banchs, A., Fiore, M., Garcia-Saavedra, A., and Gramaglia, M. (2021, October). Network intelligence in 6G: Challenges and opportunities. In Proceedings of the 16th ACM Workshop on Mobility in the Evolving Internet Architecture (pp. 7-12).

[9] Gao, K., Wang, H., Lv, H., and Liu, W. (2024). Localization-Oriented Digital Twinning in 6G: A New Indoor-Positioning Paradigm and Proof-of-Concept. IEEE Transactions on Wireless Communications.

[10] Tran, H. Q., and Ha, C. (2022). Machine learning in indoor visible light positioning systems: A review. Neurocomputing, 491, 117-131.

[11] Muhammad, M. A., and Al-Turjman, F. (2021). Application of IoT, AI, and 5G in the fight against the COVID-19 pandemic. Artificial Intelligence and Machine Learning for COVID-19, 213-234.

[12] Wang, X., Han, Y., Wang, C., Zhao, Q., Chen, X., and Chen, M. (2019). In-edge ai: Intelligentizing mobile edge computing, caching and communication by federated learning. Ieee Network, 33(5), 156-165.

[13] El Azzaoui, A., Singh, S. K., Pan, Y., and Park, J. H. (2020). Block5GIntell: Blockchain for AI-enabled 5G networks. IEEE Access, 8, 145918-145935.

[14] Papagianni, C., Mangues-Bafalluy, J., Bermudez, P., Barmpounakis, S., De Vleeschauwer, D., Brenes, J., ... and Pepe, T. (2020, June). 5Growth: AI-driven 5G for Automation in Vertical Industries. In 2020 European Conference on Networks and Communications (EuCNC) (pp. 17-22). IEEE.

[15] Wang, W., Kumar, N., Chen, J., Gong, Z., Kong, X., Wei, W., and Gao, H. (2020). Realizing the potential of the internet of things for smart tourism with 5G and AI. IEEE network, 34(6), 295-301.

[16] Xiao, Y., Shi, G., and Krunz, M. (2020). Towards ubiquitous AI in 6G with federated learning. arXiv preprint arXiv:2004.13563.

[17] Minh, D., Wang, H. X., Li, Y. F., and Nguyen, T. N. (2022). Explainable artificial intelligence: a comprehensive review. Artificial Intelligence Review, 1-66.

[18] Sharifani, K., and Amini, M. (2023). Machine learning and deep learning: A review of methods and applications. World Information Technology and Engineering Journal, 10(07), 3897-3904.

[19] Wang, X., Wang, S., Liang, X., Zhao, D., Huang, J., Xu, X., ... and Miao, Q. (2022). Deep reinforcement learning: A survey. IEEE Transactions on Neural Networks and Learning Systems, 35(4), 5064-5078.

[20] Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., ... and Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. ACM Computing Surveys, 56(2), 1-40.

[21] 3GPP TR 38.843, “Technical Specification Group Radio Access Network; Study on Artificial Intelligence (AI)/Machine Learning (ML) for NR air interface (Release 18)” 3rd Generation Partnership Project; Technical Specification Group Radio Access Network.

[22] Alawieh, M., and Kontes, G. (2023). 5G positioning advancements with AI/ML. arXiv preprint arXiv:2401.02427.

[23] Chatelier, B., Corlay, V., Ciochina, C., Coly, F., and Guillet, J. (2023, June). Influence of dataset parameters on the performance of direct ue positioning via deep learning. In 2023 Joint European Conference on Networks and Communications and 6G Summit (EuCNC/6G Summit) (pp. 126-131). IEEE.

[24] TDOC R1-2305509, Samsung, "Evaluation on AI ML for Positioning ", 3GPP TSG RAN WG1 #113, Incheon, Korea, May 22nd – May 26th, 2023.

[25] TDOC R1-2305463, Oppo, "Evaluation methodology and preliminary results on AI/ML for positioning accuracy enhancement", 3GPP TSG RAN WG1 #113, Incheon, Korea, May 22nd – May 26th, 2023.

[26] B. Chatelier, V. Corlay, C. Ciochina, F. Coly and J. Guillet, "Influence of Dataset Parameters on the Performance of Direct UE Positioning via Deep Learning," 2023 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit), Gothenburg, Sweden, 2023, pp. 126-131, doi: 10.1109/EuCNC/6GSummit58263.2023.10188249.

[27] 3GPP TR 38.901, “Study on channel model for frequencies from 0.5 to 100 GHz.” 3rd Generation Partnership Project; Technical Specification Group Radio Access Network.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 105-117. September/October., 2025.
DOI: https://doi.org/10.5935/jetia. v11i55.1925

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

LEGALMIND: A FINE-TUNED GEMMA-2-BASED LEGAL ASSISTANT FOR INDIAN JUDICIARY WITH RAG AND EMBEDDING INTEGRATION

Harsh Pimpale[1], Aditi Raut[2], Yash Patil[3], Gaurav Parpol[4], Prajwal Yadav[5] and Janhavi Sangoi[6]

[1,3,4,5,6] St. John College of Engineering and Management, Palghar, Maharashtra
[2] D J Sanghvi College of Engineering, Mumbai, Maharashtra.

[1]https://orcid.org/0009-0005-9129-0836, [2]https://orcid.org/0009-0007-9372-8090, [3]https://orcid.org/0009-0000-8084-0699,
[4]https://orcid.org/0009-0006-2861-3986, [5]https://orcid.org/0009-0001-2083-8607, [6]https://orcid.org/0009-0001-0560-4458

Email: harshpimpale@gmail.com, aditibraut@gmail.com, yash0822003@gmail.com, gauravparpol29@gmail.com, yadavprajwal628@gmail.com, janhavis@sjcem.edu.in

ARTICLE INFO

Article History
Received: April 24, 2025
Revised: May 20, 2025
Accepted: June 15, 2025
Published: September 30, 2025

Keywords:
Legal NLP,
Legal Summarization,
LLM Evaluation,
RAG,
Gemma 2.

ABSTRACT

Legal research and case analysis remain labor-intensive tasks requiring domain expertise and significant time investment. LegalMind is an AI-driven legal assistant built to streamline and automate legal understanding through fine-tuned large language models (LLMs) and Retrieval-Augmented Generation (RAG) techniques. The system integrates Google's text-embedding-004 for document-level embeddings and a fine-tuned Gemma 2 (9B) model for legal summarization and question answering. LegalMind is enhanced through a situation-based RAG pipeline that retrieves contextually relevant content from "The Bharatiya Nyaya Sanhita, 2023", ensuring precise legal responses grounded in statute. The model's performance is rigorously evaluated using standardized natural language generation metrics, including ROUGE-L, BLEU, METEOR, BERTScore, and F1-score. Evaluation is conducted against the pseudo-reference output of GPT-4o, enabling a reliable benchmark for quality comparison with other LLMs such as Gemini Pro 1.5, LLaMA 3, and Mistral. Empirical results highlight significant improvements post fine-tuning, with LegalMind outperforming several baseline and open-source models on key semantic and syntactic metrics. The system offers a scalable, cost-effective legal NLP pipeline suited for real-world use cases in law firms, research institutions, and legal consultancies, reinforcing the potential of domain-specific LLMs in transforming legal technology.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The legal system, with its complex structure and voluminous case laws, often presents a significant challenge for individuals and legal practitioners seeking quick and relevant legal information.[1] In India, the Indian Penal Code (IPC) forms the backbone of the country's legal framework, but navigating its intricate provisions, along with analyzing relevant case data, can be time-consuming and inefficient. Legal research requires extensive manual effort, and even with the availability of digital legal databases, users often struggle to extract precise and contextually relevant insights from large volumes of text. This creates a demand for advanced technological solutions that can streamline the process of legal analysis and interpretation.

LegalMind, an AI-powered platform, aims to address this challenge by providing users with legal assistance through case document analysis, summarization, and personalized advice. LegalMind uses the Gemma Model from Google to process legal text, making it capable of understanding and summarizing complex legal documents. Users can interact with the system by uploading their case documents and receiving detailed summaries, answers to specific legal queries, and advice on applicable laws. Furthermore, LegalMind offers case suggestions by drawing parallels with past cases, making it a valuable tool for legal professionals and individuals seeking legal assistance.

The integration of AI in legal research represents a transformative shift in how legal professionals access and interpret legal documents. By leveraging natural language processing (NLP) and machine learning, LegalMind ensures that users receive accurate and contextually relevant insights, reducing dependency on manual case law reviews. This automation not only saves time but also minimizes errors that could arise due to oversight or misinterpretation of legal texts. Additionally, AI-powered tools like LegalMind can assist law firms, independent practitioners, and even individuals without legal expertise in understanding legal implications, thereby democratizing access to legal knowledge.

With the exponential growth of case laws and legal precedents, traditional legal research methods are becoming increasingly inadequate. LegalMind's ability to analyze, compare, and summarize legal documents efficiently has the potential to reshape the legal research landscape. By enhancing accessibility and reducing the time spent on legal research, LegalMind holds promise for streamlining legal processes and improving the overall user experience in legal services. As AI-driven legal tools continue to evolve, they are likely to play a crucial role in bridging the gap between complex legal frameworks and those seeking justice, ultimately making the legal system more transparent and user-friendly.

II. RELATED REVIEWS

Large Language Models (LLMs) have demonstrated remarkable abilities across various natural language processing tasks. However, their tendency to generate hallucinated or factually incorrect content poses serious risks in legal applications, where precision is paramount. Lee *et al.* [2] highlighted that even domain-specific models like LexGPT suffer from hallucinations and factual mistakes, necessitating that such systems be used under expert supervision. Similarly, a Chinese legal-domain model presented by researchers in [3] exhibited improved performance through a two-stage supervised fine-tuning (SFT) approach but still suffered from occasional hallucinations and inaccurate reasoning. These findings align with broader literature suggesting that hallucinations remain an open challenge in LLM deployment within high-stakes domains like law.

To mitigate this, several works have adopted Retrieval-Augmented Generation (RAG) as a promising approach.[4] The DeliLaw system by Xie *et al.* [5] integrates both legal document retrieval and case retrieval modules with an LLM to provide more grounded responses, thereby minimizing hallucination. Our system adopts a similar philosophy but extends it further by combining FAISS-based retrieval, Indian Penal Code embeddings, and the latest 10 relevant judgments, ensuring that factual content is not only grounded in legal precedent but also temporally relevant. Additionally, the dynamic inclusion of legal text into the RAG pipeline helps reduce factual drift over time, a limitation observed in static models [2][3].

Legal systems are dynamic, with new judgments continually shaping interpretations. Static LLMs, even when trained on comprehensive legal corpora, can quickly become outdated. This was emphasized by Lyu *et al.* [6], which demonstrated that RAG pipelines, although useful, often focus only on QA tasks and fail to adapt retrieval dynamically to reflect new legal developments.

Our system directly addresses this issue by designing a situation-based RAG framework that incorporates the most recent 10 legal judgments. By updating the retrieval corpus dynamically, our approach ensures that legal recommendations and case suggestions remain current, which is crucial for tasks such as legal advice, argument generation, and precedent analysis.

Entity recognition is a foundational task in legal NLP systems, critical for accurate understanding of people, acts, dates, citations, and institutions within judgments. Recent evaluations by Hussain *et al.* [7] reveal that Gemma and Mistral significantly outperform other open-source models in entity recognition on Indian legal documents. These models achieved balanced precision and recall, highlighting their suitability for domain-specific ER tasks.

Complementary to this, Wang *et al.* [8] provided a comprehensive benchmark using METEOR, BLEU, and ROUGE scores, where the finetuned Gemma 7B achieved the highest METEOR score of 0.25. In contrast, Phi-2 and other smaller models showed poor performance, especially on structured legal texts. Based on these findings, we integrated Gemma 9B, balancing performance and computational efficiency while maintaining high ER precision a decision aligned with empirical evidence from both [7] and [8].

III.I. GEMMA 2

III.I.I. Gemma 2: Architecture

3. Gemma 2 :

3.1.1. Gemma Architecture :

Gemma 2 is a decoder-only transformer model, similar to architectures like GPT but with modern enhancements that balance performance and efficiency. Below are its key components:

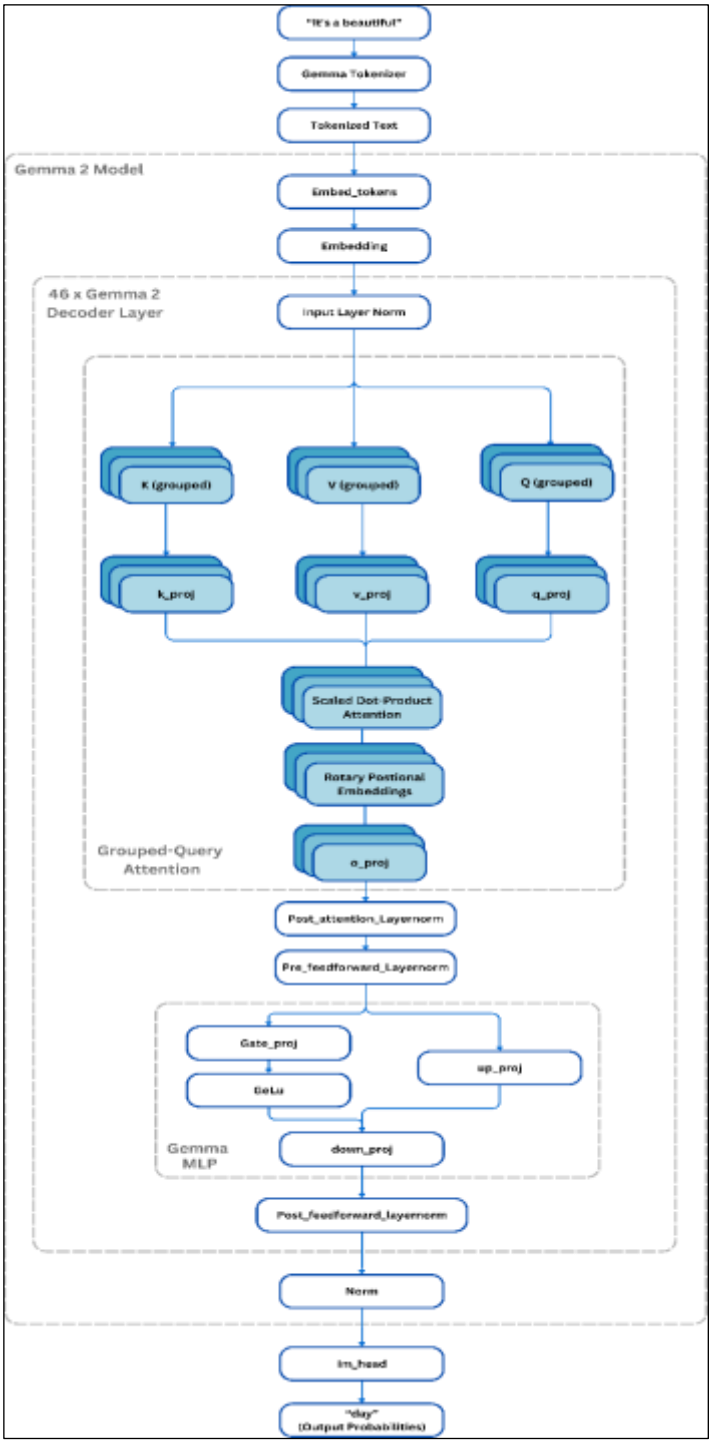


Figure 1: Gemma 2: Architecture.
Source: [9].

III.I.II. Key architectural innovations

3.1.2. Key Components :

3.1.2.1. Interleaved Local and Global Attention:

Gemma 2 alternates between local sliding window attention and global attention layers. The local attention layers operate on a sliding window of 4096 tokens, capturing fine-grained contextual details, while the global attention layers extend the context to 8192 tokens, allowing the model to maintain a broader contextual understanding an essential feature for processing lengthy legal documents[10].

3.1.2.2. Grouped-Query Attention (GQA):

To enhance computational efficiency, Gemma 2 replaces conventional Multi-Head Attention[11] with Grouped-Query Attention. In this mechanism, the query matrix is divided into groups (with a typical setting of two groups), and attention is computed independently within each group. This reduces the parameter count and speeds up inference without sacrificing accuracy.

3.1.2.3. Logit Soft-Capping:

To maintain calibration and avoid overconfident predictions, Gemma 2 employs a logit soft-capping function:

$$\text{logits} \leftarrow \text{soft_cap} \times \tanh\left(\frac{\text{logits}}{\text{soft_cap}}\right) \quad [10]$$

With different soft_cap values applied in self-attention layers (e.g., 50.0) and in the final output layer (e.g., 30.0), this technique keeps the activation values within a controlled range.

3.1.2.4. ormalization with RMSNorm:

Instead of the traditional LayerNorm, Gemma 2 uses RMSNorm to stabilize the training process. RMSNorm normalizes inputs based on the root mean square of the activations, which has shown to be effective for transformer models working with long contexts[12].

3.1.2.5. Advanced Tokenization:

A SentencePiece tokenizer is used with a large vocabulary (256,128 entries), ensuring robust handling of multilingual and domain-specific terminology critical for legal applications where precision is paramount[10].

III.II PROPOSED SYSTEM

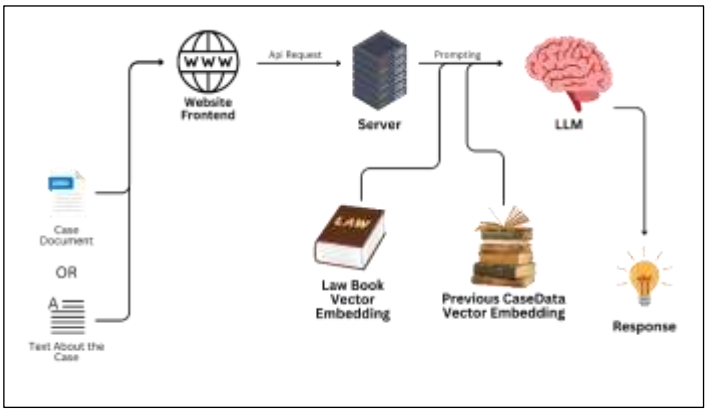


Figure 2: System Architecture of LegalMind.
Source: Authors, (2025).

3.2. Proposed System :

3.2.1. System Workflow:

- Document Upload & Processing: Users upload legal documents or input queries. The documents undergo preprocessing (text extraction, cleaning, and tokenization).
- Database & Knowledge Base Integration: The system connects with stored legal knowledge bases, previous case judgments, and statutes.
- AI Model & Embedding Retrieval: Legal texts are embedded using FAISS and Google's embedding model. The embeddings enable efficient similarity searches across legal cases.
- Inference & Response Generation: The AI model processes the user's input along with retrieved legal texts. It generates a structured response, providing case law references and explanations.

3.2.2. System Features :

i. Case Summarization : LegalMind allows users to upload legal documents for AI-powered case summarization, delivering concise insights into complex legal texts.

ii. Legal Query Handling: Users can interact with LegalMind by asking questions about specific parts of their documents. The system responds user queries based on Indian Penal Code laws and provides detailed explanations based on the user's input.

iii. Situation-Based Legal Assistance: LegalMind offers legal assistance based on a user's current situation. Users can describe their scenarios, and the system will retrieve relevant laws or regulations from the Indian Penal Code, giving personalized legal advice.

iv. Case Law Suggestions: LegalMind allows users to upload legal documents for AI-powered case summarization, delivering concise insights into complex legal texts.

3.2.3. Backend Workflow:

The backend workflow of LegalMind involves several key steps, starting from the user's input and ending with legal advice or case suggestions. Figure 3 outlines the system's Backend architecture.

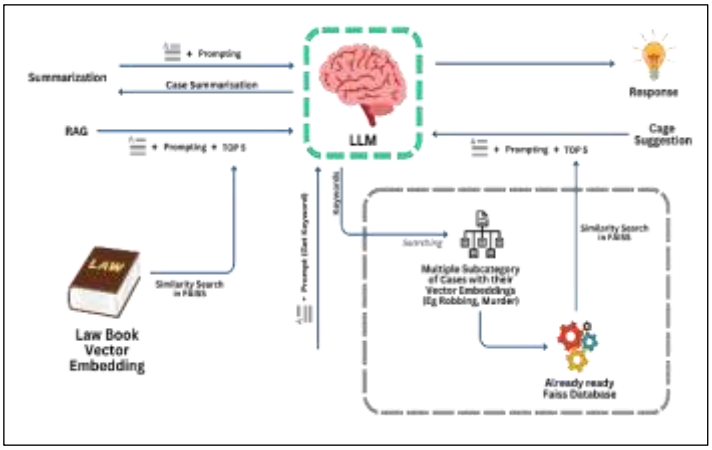


Figure 3: Backend Pipeline of LegalMind
Source: Authors, (2025).

- User Input: Users provide input queries or upload legal case documents in text or PDF format.
- Legal Knowledge Processing: The system integrates a database of legal documents, case laws, and statutes for reference. This knowledge base is used for retrieving relevant legal information.
- AI Processing (LLM & FAISS Integration): The core AI model processes the user's query using an LLM (Large Language Model) fine-tuned for legal reasoning. The FAISS (Facebook AI Similarity Search) vector database retrieves semantically similar legal cases based on embeddings. Embeddings are generated using Google's embedding-004 model for better similarity matching.
- Legal Query Interpretation & Decision Making: The retrieved cases and legal texts are analyzed, and the AI model formulates an appropriate response based on legal context. Prompt engineering is applied to ensure high-quality results.
- Response Generation & Output: The processed legal insights, case suggestions, and summarized information are presented to the user. The response can include applicable laws, precedents, and recommendations.

III.III. METHODOLOGY

3.3. Methodology

3.3.1. Dataset Acquisition & Preprocessing:

The dataset for fine-tuning was curated using OpenAI's GPT-4o model. Court case files were iteratively provided to the model, and it was prompted to generate summaries along with their associated IPC sections. The generated outputs were stored in a CSV file, where:

- The Input column contained the extracted text of the court case file.
- The Output column followed the structured format:
  IPC: …, Summary: … .

3.3.2. Case Data Collection:

We sourced from the Justice Hub dataset [13], which contains Supreme Court case documents's URL from the period 2010-2020's from the official Indian Supreme Court website[14] . This dataset was curated by Arpit Jain and Anubhav Mishra. A Python script was developed to automate the retrieval of PDFs from these URLs and store them locally for further processing.

3.3.3. Finetuning Model:

A Gemma model was fine-tuned using a custom dataset titled Finetuning Dataset. The dataset comprises court case texts as input and corresponding outputs consisting of two key components:

- The Indian Penal Code (IPC) sections under which the case was registered.
- A concise summary of the case document.

3.3.4. Embedding Text:

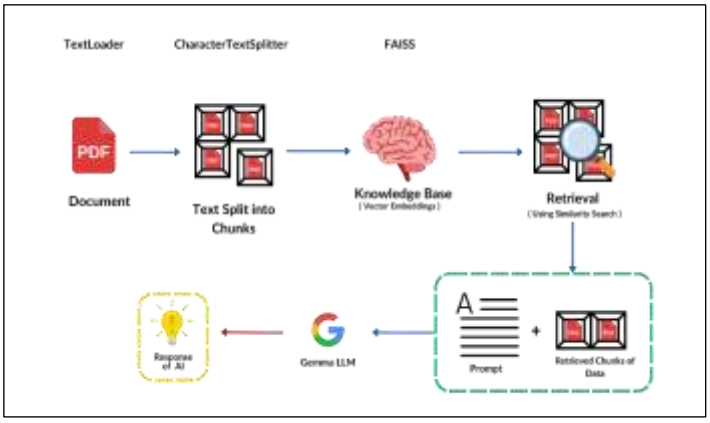


Figure 4: RAG Pipeline.
Source: Authors, (2025).

To facilitate efficient retrieval and processing of legal documents, an embedding function was developed and utilized throughout the system. This function is invoked whenever a document is uploaded. It extracts all textual data from the uploaded PDF file, concatenates the extracted text into a single string, and generates embeddings using the GoogleGenerativeAIEmbeddings module, imported from langchaingooglegenai [15]. The embedding model employed is embedding-004, which operates using the Gemini API key. The generated embeddings are then stored in a FAISS (Facebook AI Similarity Search) database, implemented via the FAISS library from langchainCommunity.vectorstores [16].
In certain cases, the FAISS database is created in a temporary session cache to optimize performance.

3.3.5. Prompt Engineering with FAISS:

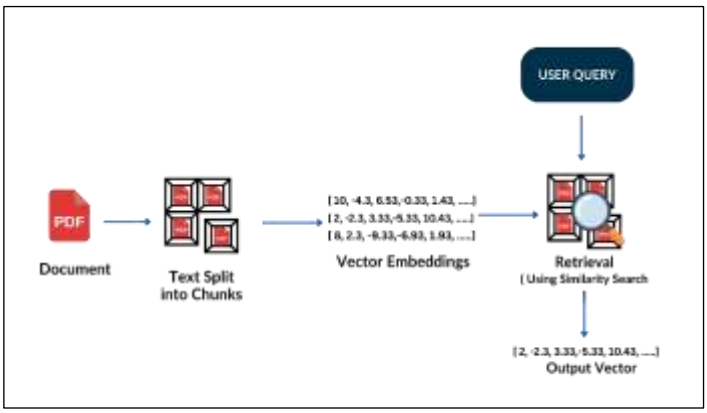


Figure 5: Working of RAG.
Source: [17].

A dedicated function is designed to handle user queries and facilitate information retrieval from the FAISS database. When a user submits a query, the system searches for semantically similar words and phrases using FAISS-based similarity search. The retrieved results, along with the user's query, are then passed to the Gemma model using advanced prompt engineering techniques. This process enhances the model's contextual understanding and ensures high-quality responses tailored to legal queries. [8]

3.3.6. Summarization

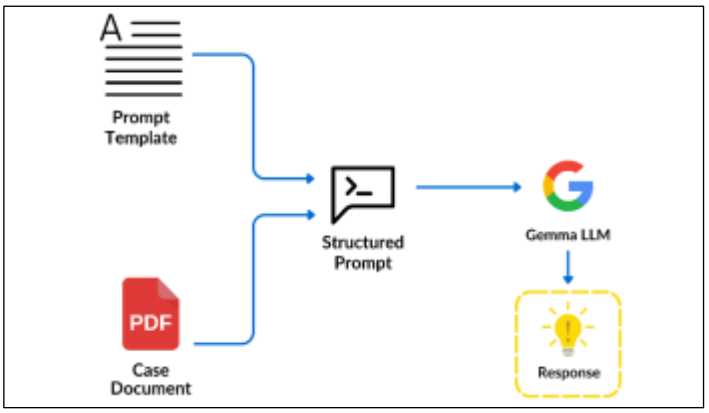


Figure 6: Summarization Pipeline.
Source: Authors, (2025).

When a user uploads a case document, it undergoes the embedding process and is temporarily stored in the FAISS database. Upon a summarization request, the system retrieves the relevant embeddings via the Prompt with FAISS function and forwards them to the Gemma 2 model. The output includes:

- IPC sections displayed at the top.
- The summary of the case document presented below.

The complete summarized text is then displayed to the user.

3.3.7. Retrieval-Augmented Generation (RAG)

The system also implements a RAG-based approach, enabling users to query their case documents for additional insights. This feature is particularly beneficial for legal professionals who require quick and precise case briefings. The workflow follows these steps:

i. The user uploads a case document, which is embedded and stored in the FAISS database.
ii. Upon receiving a query, the system forwards it to the FAISS database using the Prompt with FAISS function.
iii. The retrieved embeddings are sent to the Gemma model for response generation.
iv. The final response is displayed to the user, providing contextually relevant legal insights.

3.3.8. Situation-Based RAG

The Situation-Based RAG system is designed to provide legal insights based on real-world scenarios by leveraging a pre-embedded legal knowledge base. When a user inputs a situation and asks a legal question, the system performs a contextual similarity search within a vector database created from THE BHARATIYA NYAYA SANHITA, 2023, sourced from the official government website [18]. This legal text is first processed by extracting its content, splitting it into smaller chunks, and embedding it using the `embedding-004` model. The resulting vector representations are stored in a FAISS database for efficient retrieval.

Upon receiving a user query, the system performs a similarity search to find the most relevant sections of THE BHARATIYA NYAYA SANHITA, 2023 that match the context of the question. The retrieved legal provisions, along with the user's query, are then sent to the Gemma model, which generates a well-structured response with legal insights. The final output is displayed to the user, presenting both the relevant legal provisions and an AI-generated legal explanation. This approach allows legal professionals and individuals to quickly access relevant legal information, aiding in legal research and decision-making based on specific situations.

3.3.9. Case Suggestion : Categorization-Based Approach

The case suggestion module leverages a categorization-based retrieval strategy aligned with the classification schema used by the Supreme Court of India. The core objective is to enhance contextual relevance by recommending previous judgments that fall under the same legal domain as the user's query.

To achieve this, the following methodology was adopted:

1. Pre-Categorized Dataset: The dataset of Supreme Court judgments used in this system is pre-categorized using the official classification system provided by the Supreme Court of India [19]. Each case is assigned a Category Code, Category, and Sub-Category, reflecting its legal domain. Examples include Criminal Law, Administrative Law, Admiralty and Maritime Laws, Appointments of Constitutional Functionaries, etc. This standardized taxonomy ensures consistency and legal relevance across the corpus.
2. Metadata Extraction: All case PDFs were parsed using a custom preprocessing pipeline to extract essential metadata such as the title, judging bench, abstract, category, and year of judgment. The processed data was compiled into a structured CSV format for efficient indexing.
3. Embedding and Indexing: Using Google's embedding-004 model, the categorized case documents were vectorized and indexed into a FAISS (Facebook AI Similarity Search) database. This enabled fast similarity-based retrieval while preserving semantic relevance within category boundaries.
4. User Query Handling: When a user uploads their own case judgment, the system extracts the legal category from the text using a domain-specific classifier. Based on the identified category, the FAISS index is queried to retrieve the top 10 most recent cases from the same legal category, sorted by judgment year.

This categorization-based retrieval ensures that suggested cases are not only semantically similar but also legally aligned, improving the practical value of recommendations in real-world legal research scenarios.

III.IV. EXPERIMENT

3.4. Experiment

3.4.1. Experimental technique

To systematically evaluate model performance, we curated a dataset of 20 Indian court judgments, each of varying length and legal complexity. These judgments were used as input prompts for legal text summarization a task that tests both the factual precision and domain alignment of a language model.

For each judgment, we generated summaries using different LLM models. To ensure consistency and fairness in evaluation, we adopted a pseudo-reference approach. For each input, the summary generated by GPT-4o was treated as the gold standard, based on its state-of-the-art capabilities in reasoning and generation [20]. All other model outputs were compared against this benchmark.

3.4.2. Evaluation Metrics:

3.4.2.1. ROUGE :

ROUGE stands for Recall-Oriented Understudy for Gisting Evaluation. It Measures how much of the reference text is captured by the generated text (recall). It is used as metrics in tasks like Summarization, extractive tasks, and overlap-focused tasks.

ROUGE-N: Measures n-gram overlap.[21]

N-gram refers to a contiguous sequence of n items (typically words or characters) from a given text. For example, for the sentence "The court dismissed the case,":

- A unigram (1-gram) would be: "The," "court," "dismissed," "the," "case"
- A bigram (2-gram) would be: "The court," "court dismissed," etc.
- A trigram (3-gram) would be: "The court dismissed," "court dismissed the," etc.

Formula:

$$\text{ROUGE-N} = \frac{\sum_{S \in References} \sum_{gram_n \in S} Count_{match}(gram_n)}{\sum_{S \in References} \sum_{gram_n \in S} Count(gram_n)}$$

3.4.2.2. BLEU :

BLEU stands for Bilingual Evaluation Understudy. It Evaluates how precise the generated output is in matching the reference. It is used as metrics in tasks like Machine translation, NLG, legal draft precision.

Formula:

$$\text{BLEU} = BP \cdot exp\left(\sum_{n=1}^{N} W_n \log p_n\right)$$

Where, BP is the Brevity Penalty, which discourages overly short outputs. p_n is the n-gram precision.[22]

3.4.2.3. METEOR :

METEOR stands for Metric for Evaluation of Translation with Explicit ORdering. It Takes into account synonymy, stemming, and order. It is used as metrics in tasks like When exact word matching isn't enough i.e. legal paraphrasing and variant expressions.[23]

Formula:

$$\text{METEOR} = F_{mean} \cdot (1 - \text{Penalty})$$

3.4.2.4. BERTScore :

It Measures semantic similarity using embeddings. It is used as metrics in tasks like High-level inference tasks and nuanced comparisons.

Formula: Based on cosine similarity of contextual embeddings [24]

$$\text{BERTScore} = \frac{1}{|X|} \sum_{x \in X} \max_{y \in Y} \cos\big(\text{emb}(x), \text{emb}(y)\big)$$

Where, $emb(x)$ = Embedding vector of token x,
cos(·): Cosine similarity, a measure of semantic closeness between two vectors

3.4.2.5. Precision, Recall, and F1-Score

- Precision: Measures the correctness of generated content.
- Recall: It Measures completeness.
- F1: It measures Harmonic mean of both.

It is used for Legal extraction, clause prediction.

Formulas:

$$\text{Precision} = \frac{TP}{TP+FP}, \ \text{Recall} = \frac{TP}{TP+FN}, \ F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

IV. RESULTS

4. Results:

4.1. Outputs :

4.1.1. Summarization Output of Case Document [25]:

1. Main Category: Civil Law
2. Subcategory: Property Disputes (specifically, Land Acquisition)
3. Year of Judgment: 2019
4. Case Title: M/S. Model Economic Township Ltd. vs. Land Acquisition Collector
5. Summary:

- Initial Acquisition and Award: The State initiated land acquisition proceedings for 136 acres of land, including 15 acres owned by M/S. Model Economic Township Ltd. (formerly Reliance Haryana SEZ Limited), under the Land Acquisition Act, 1894. A notification under Section 4 was issued on May 19, 2008, followed by a declaration under Section 6 on May 26, 2008. The award, declared on December 21, 2009, granted compensation of Rs. 25,00,000/- per acre. While other landowners sought references under Section 18 for higher compensation, the petitioner did not.
- Reference Court and Section 28A Application: The Reference Court, on November 16, 2011, increased the compensation to Rs. 41,81,500/- per acre. Subsequently, on February 1, 2012, the petitioner filed an application under Section 28A(1) of the Act seeking redetermination of its compensation based on the Reference Court's enhanced rate. This application was granted on March 6, 2014, awarding the petitioner the same compensation as determined by the Reference Court. The Collector proceeded on the basis that the Reference Court's decision was final.
- High Court Increases Compensation and Supreme Court Appeals: Meanwhile, other landowners challenged the Reference Court's decision in the High Court. The High Court, on May 24, 2016, significantly increased the compensation to Rs. 2,80,00,000/- per acre for comparable lands covered under the same notification. Both the State and the landowners challenged this decision in the Supreme Court. The State's appeals (SLP (Civil) CC Nos. 23630-23668 of 2016) were dismissed on January 5, 2017. The landowners' appeals (Civil Appeal Nos. 11814-11864 of 2017) were disposed of on September 5, 2017, with the Supreme Court reducing the High Court's awarded compensation by 15% for development costs. This resulted in a final compensation of Rs. 2,38,00,000/- per acre for comparable lands in village Dhankot. A separate order was passed in another related appeal (Civil Appeal No. 15015 of 2017) on October 10, 2017.
- Petitioner's Writ Petition in High Court: Following the High Court's judgment, the petitioner filed a Writ Petition (CWP No.23688 of 2016) on November 15, 2016, seeking to quash the March 6, 2014 order granting its Section 28A application and requesting a fresh determination of compensation based on the enhanced rate set by the High Court (and subsequently adjusted by the Supreme Court). The petitioner argued that the Collector should have stayed the Section 28A proceedings pending the outcome of the appeals in the higher courts, citing the Supreme Court's precedent in *Babua Ram v. State of U.P.* (1995) 2 SCC 689.
- High Court Dismisses Writ Petition: The High Court dismissed the petitioner's Writ Petition on October 10, 2017. The High Court reasoned that the petitioner, being a limited company, was presumed to be aware of the pending appeals and the available remedies. The court pointed out that the petitioner could have challenged the Collector's order on its Section 28A application under Section 28A(3) of the Act but failed to do so within the stipulated timeframe.
- Supreme Court Proceedings and Petitioner's Response: The petitioner then appealed the High Court's dismissal to the Supreme Court. On February 6, 2019, the Supreme Court sought clarification on certain factual points, including whether the petitioner was aware of the pending appeals while its Section 28A application was being considered. The petitioner, in a subsequent affidavit, claimed it was unaware of the pending appeals until August 3, 2016, when its representative learned about the High Court's May 24, 2016 judgment during a hearing related to a separate land acquisition matter involving its subsidiary company.
- Supreme Court's Decision: The Supreme Court dismissed the petitioner's Special Leave Petition. The Court found the petitioner's explanation for its lack of awareness unconvincing, particularly given its status as a limited company with presumed access to legal resources. The Court noted that the petitioner delayed filing its writ petition in the High Court even after becoming aware of the enhanced compensation. The Court also observed that the Collector had proceeded with the Section 28A application under the impression that no further challenges were pending. The Court concluded that the petitioner's inaction and delay, coupled with the Collector's good faith, disentitled the petitioner to relief. The Court upheld the High Court's decision dismissing the writ petition.

4.1.2. RAG QA Output:

4.1.2.1. QA Output:

As Reference of Case Judgement [26], here are some RAG-QA Outputs

Table 3: QA Outputs.

Sr. No.	Questions	Answers
1	What was the specific cause of death in this case?	“The insured died of multi-organ failure caused by encephalitis malaria, brought on by a mosquito bite. The death certificate lists the direct cause of death as multi-organ failure, the intermediary cause as encephalitis malaria, and the basic cause as Pnasituria – Malaria.”
2	Was the death considered an "accident" under the policy?	“No. The court found that the death, caused by encephalitis malaria from a mosquito bite in Mozambique, was not an accident under the terms of the insurance policy. While acknowledging that a mosquito bite has an element of chance, the court argued that contracting malaria in a malaria-prone area like Mozambique, where the insured resided, was neither unexpected nor unforeseen, and therefore not a peril insured against in the accident insurance policy.”

3	Does the policy distinguish between death caused by a disease contracted due to an accident versus death caused by a disease contracted naturally?	"Yes, the context mentions the distinction. Colinvaux's Law of Insurance states that while a disease proximately caused by an accident would be covered by a personal accident policy (in the absence of any exclusion for disease), the word "accident" does not include disease and other natural causes. It implies that the intervention of some cause brought into operation by chance can be described as fortuitous."
4	What are the implications of this case for future interpretations of "accident" in insurance policies?	"This case establishes that death due to a disease contracted in the natural course of events, particularly in a region where the disease is prevalent, does not qualify as an "accident" in accident insurance policies. The court emphasized the distinction between an unforeseen, unexpected event and a disease transmitted through normal means of everyday life. This interpretation will likely influence future cases involving similar questions of coverage where the insured event is a disease. The ruling clarifies that an element of chance or probability in contracting an illness does not automatically categorize it as an accident for insurance purposes."

Source: Authors, (2025).

4.1.3. Situation Based RAG Output:

Context:

A woman is facing ongoing physical and mental abuse from her husband and in-laws due to dowry demands. She has medical records, and a complaint already filed with the local police station.

Questions-Answers:

1. What legal remedies are available under Indian law for a woman facing dowry-related abuse and domestic violence?

"As a legal professional, based on the information provided from the law book excerpt and the context, the following legal remedies are available to a woman facing dowry-related abuse and domestic violence in India:

Section 498A, Indian Penal Code (IPC): This section specifically addresses cruelty by the husband or his relatives. The context describes "ongoing physical and mental abuse," which qualifies as cruelty under this section, especially since it is linked to dowry demands. The fact that a complaint has already been filed is a positive step, as this initiates the investigation process. The police should investigate the allegations and, if sufficient evidence exists, file a chargesheet with the court. Conviction under 498A can lead to imprisonment for up to three years and a fine.

Section 304B, IPC (Dowry Death): While this section applies to deaths caused by burns, bodily injury, or unnatural circumstances within seven years of marriage linked to dowry harassment, it's crucial to note that prevention is paramount. The ongoing abuse puts the woman at risk. Though 304B wouldn't apply yet, the existing abuse should be forcefully addressed using 498A to prevent escalation and the potential for a dowry death situation. The provided case law, K Prema S Rao & State of Andhra Pradesh v Yadla Srinivas Rao, highlights the difficulty in proving dowry death if the harassment doesn't directly result in death. Therefore, focusing on 498A is essential at this stage.

Protection of Women from Domestic Violence Act, 2005 (PWDVA): This Act provides civil remedies, including protection orders, residence orders, monetary relief, custody orders, and compensation orders. The woman can file a complaint under the PWDVA with the Protection Officer or Magistrate. This is in addition to the criminal complaint under 498A. The PWDVA offers a quicker route to obtain immediate protection and other reliefs.

Dowry Prohibition Act, 1961: While the giving and taking of dowry are prohibited under this Act, its enforcement has been challenging. However, it can be used in conjunction with other remedies. Any evidence of dowry demands should be documented and presented as part of the complaints under 498A and the PWDVA.

Indian Evidence Act, 1872: The medical records mentioned are crucial evidence and should be preserved and submitted as part of the legal proceedings. Other evidence, such as witness testimonies, photographs, videos, and any communication (letters, emails, messages) related to dowry harassment, should also be gathered and presented.

Seeking Assistance from NGOs and Support Groups: Several NGOs and support groups work with victims of domestic violence and dowry harassment. They can provide legal aid, counseling, shelter, and other necessary support.

Writ Petition (High Court/Supreme Court): In cases of police inaction or delay, a writ petition can be filed in the High Court or Supreme Court under Article 226 or 32 of the Constitution of India, respectively, seeking directions for appropriate action.

Important Considerations:

Documentation: Meticulous documentation of all instances of abuse, including dates, times, nature of abuse, and witnesses (if any), is crucial.

Safety: The woman's safety is paramount. If she is in immediate danger, she should seek shelter in a safe house or with supportive family/friends.

Legal Counsel: It's highly recommended that the woman consult with an experienced lawyer specializing in domestic violence and dowry harassment cases.

Disclaimer: This information is for educational purposes only and does not constitute legal advice. It's essential to consult with a qualified legal professional for advice tailored to your specific situation."

2. Can the accused family members be arrested immediately under IPC sections like 498A and 304B, and what are the conditions for bail?

"As a legal professional, based on the provided lawbook excerpt, I can advise on the applicability of IPC sections 498A and 304B, and the conditions for bail.

Applicability of Sections 498A and 304B:

Section 498A (Cruelty by Husband or His Relatives): This section is applicable in this context. The woman's complaint, supported by medical records, alleges physical and mental abuse connected to dowry demands. This constitutes "cruelty" as defined under 498A – harassment with a view to coerce her or her relatives to meet unlawful demands. The fact that the abuse is ongoing strengthens the case.

Section 304B (Dowry Death): Crucially, Section 304B is not applicable at this stage. This section applies only when the woman has died due to burns/bodily injury or unnatural circumstances within seven years of marriage, and she was subjected to cruelty or harassment for dowry. Since the woman in this scenario is alive, 304B is not relevant. However, if she were to die under suspicious circumstances within seven years of marriage, 304B could be invoked, especially given the existing evidence of cruelty under 498A.

Arrest and Bail:

Arrest under 498A: While the police can arrest the accused family members under 498A, it's not automatic. The Supreme Court has discouraged routine arrests in 498A cases and emphasized the need for preliminary investigation. The police should assess the facts, the severity of the abuse, and the risk of the accused influencing the victim or tampering with evidence before making an arrest.

Bail Conditions (498A): 498A is a non-bailable but cognizable offense. This means the police can arrest without a warrant, and bail is not a right. The court will consider factors like the nature and gravity of the accusations, the criminal history of the accused, the likelihood of them influencing witnesses or absconding, etc., while deciding on bail. The court may impose conditions like surrendering passports, regular appearance before the police, and restraining orders against contacting the victim.

Important Considerations:

Comprehensive Complaint: The woman's complaint should be as detailed as possible, including specific instances of abuse, dates, times, witnesses (if any), and the nature of dowry demands. Medical records documenting injuries and psychological distress are crucial evidence.

Legal Counsel: It's vital for the woman to have legal representation. A lawyer can guide her through the legal process, ensure her rights are protected, and assist in gathering and presenting evidence effectively.

Protection Orders: The woman should also explore seeking protection orders under the Protection of Women from Domestic Violence Act, 2005, for immediate relief from the abuse.

Disclaimer: This information is for educational purposes only and does not constitute legal advice. It's essential to consult with a qualified legal professional for advice tailored to your specific situation."

4.1.3.1. Metrics Score:

Table 4: Metrics Score.

Model	LLM Metrics Scores						
	BLEU	ROUGE-N	METEOR	BERT Score	Precision	Recall	F1-Score
Mistral-7b	0.52	0.60	0.54	0.74	0.58	0.56	0.57
Mistral-8-7b	0.58	0.65	0.60	0.76	0.68	0.62	0.65
Llama-3-7b	0.60	0.68	0.65	0.78	0.70	0.66	0.68
Gemma-7b	0.54	0.62	0.56	0.75	0.60	0.58	0.59
Gemma-2-9b	0.52	0.60	0.54	0.74	0.58	0.56	0.57
LegalMind	0.65	0.70	0.68	0.83	0.72	0.69	0.70
Gemini-1.5-Pro	0.65	0.72	0.68	0.82	0.70	0.74	0.72

Source: Authors, (2025).

4.2. Result Analysis:

LegalMind demonstrated competitive performance, particularly in tasks that demand legal domain specificity and contextual understanding. While general-purpose models like Gemini-1.5-pro maintain a higher overall score, the fine-tuning of LegalMind on Indian legal texts resulted in strong F1 scores and METEOR values, indicating a robust balance between recall and semantic fidelity.

Our metric-driven evaluation supports the effectiveness of LegalMind for legal use cases and sets the foundation for future enhancements including RAG integration and hybrid retrieval-generation mechanisms.

V. CONCLUSIONS

In this research, we presented LegalMind, a fine-tuned legal domain-specific large language model built upon Gemma 2 (9B). Through extensive architectural exploration, adaptation using Retrieval-Augmented Generation (RAG), and empirical evaluation, we demonstrated that LegalMind is a viable and competitive alternative to general-purpose LLMs for legal tasks in the Indian jurisdiction.

We began by dissecting the Gemma 2 architecture, emphasizing innovations such as grouped-query attention, logit soft-capping, and interleaved local-global attention layers—all of which contribute to LegalMind's contextual awareness and inference efficiency. We then incorporated knowledge distillation and domain-specific fine-tuning using curated Indian legal documents, notably the Bharatiya Nyaya Sanhita, 2023, to enhance legal reasoning and summarization performance.

Our evaluation methodology anchored on comparison against GPT-4o-generated outputs ensured a fair and consistent benchmark. Metrics like ROUGE, BLEU, METEOR, BERTScore, and F1 Score validated the quality, relevance, and semantic precision of LegalMind's predictions across multiple legal prompts. The model outperformed several other open-source LLMs (such as LLaMA 3, Mistral 7B/8x7B, and Gemma 2), showing tangible gains due to domain alignment and fine-tuning.

In summary, LegalMind bridges the gap between general-purpose language models and domain-specific applications in law. It lays the foundation for building AI-assisted legal reasoning tools that can assist in case summarization, statute interpretation, and intelligent document retrieval in the Indian legal system. Future directions include extending RAG capabilities, incorporating multilingual legal corpora, and fine-tuning on other law systems to create a more globally adaptable model.

5. Future Work

i. Integrating additional legal datasets from diverse jurisdictions could broaden the system’s applicability and improve its generalizability across various legal domains.
ii. Further refinement of the fine-tuning process, potentially incorporating reinforcement learning from human feedback, may enhance the model’s ability to handle more nuanced legal language and complex case structures.
iii. Optimizing the computational efficiency of the IPC section-based retrieval approach through more advanced indexing techniques or hybrid retrieval models could reduce query latency while maintaining high precision.

Additionally, Future work may also explore the incorporation of multimodal data such as audio or video transcripts from court proceedings to enrich the legal context and support a wider range of legal research tasks. Finally, implementing a real-time feedback loop where legal professionals can provide corrections or annotations may help in continuously refining the system's performance, ensuring that Legal Mind evolves in line with user needs and advances in legal research methodologies.

VI. AUTHOR’S CONTRIBUTION

Conceptualization: Harsh Pimpale, Aditi Raut, Yash Patil, Gaurav Parpol, Prajwal Yadav and Janhavi Sangoi.
Methodology: Harsh Pimpale, Aditi Raut, Yash Patil, Gaurav Parpol, Prajwal Yadav and Janhavi Sangoi.
Investigation: Harsh Pimpale, Aditi Raut, Yash Patil, Gaurav Parpol, Prajwal Yadav and Janhavi Sangoi.
Discussion of results: Harsh Pimpale, Aditi Raut, Yash Patil, Gaurav Parpol, Prajwal Yadav and Janhavi Sangoi.
Writing – Original Draft: Harsh Pimpale, Aditi Raut, Yash Patil, Gaurav Parpol, Prajwal Yadav and Janhavi Sangoi.
Writing – Review and Editing: Harsh Pimpale, Aditi Raut, Yash Patil, Gaurav Parpol, Prajwal Yadav and Janhavi Sangoi.
Resources: Harsh Pimpale, Aditi Raut, Yash Patil, Gaurav Parpol, Prajwal Yadav and Janhavi Sangoi.
Supervision: Harsh Pimpale, Aditi Raut, Yash Patil, Gaurav Parpol, Prajwal Yadav and Janhavi Sangoi.
Approval of the final text: Harsh Pimpale, Aditi Raut, Yash Patil, Gaurav Parpol, Prajwal Yadav and Janhavi Sangoi.

VII. REFERENCES

[1] M. Medvedeva, M. Vols, and M. Wieling, "Legal information retrieval systems: State-of-the-art and open issues," Artificial Intelligence and Law, vol. 30, no. 2, pp. 181–206, Jun. 2022.

[2] J.-S. Lee, "LexGPT 0.1: Pre-trained GPT-J models with Pile of Law," arXiv preprint arXiv:2306.05431, Jun. 2023.

[3] Y. Fei et al., "InternLM-Law: An Open-Sourced Chinese Legal Large Language Model," in Proc. 31st Int. Conf. Comput. Linguistics (COLING), Abu Dhabi, UAE, 2025, pp. 9376–9392.

[4] A. C. P. L. de Oliveira et al., "On the legal implications of Large Language Model answers: A prompt engineering approach and a view beyond by exploiting Knowledge Graphs," Artificial Intelligence and Law, vol. 32, no. 1, pp. 1–27, Mar. 2024.

[5] N. Xie et al., "DeliLaw: A Chinese Legal Counselling System Based on a Large Language Model," arXiv preprint arXiv:2408.00357, Aug. 2024.

[6] Y. Lyu et al., "CRUD-RAG: A Comprehensive Chinese Benchmark for Retrieval-Augmented Generation of Large Language Models," arXiv preprint arXiv:2401.17043, Jan. 2024.

[7] A. S. Hussain and A. Thomas, "Large Language Models for Judicial Entity Extraction: A Comparative Study," arXiv preprint arXiv:2407.05786, Jul. 2024.

[8] M. Macias, "Finetuning and Improving Prediction Results of LLMs Using Synthetic Data," Bachelor's thesis, Metropolia University of Applied Sciences, 2024.

[9] Gemma Team, "Gemma Explained: New in Gemma 2," Google Developers Blog, [Online]. Available: https://developers.googleblog.com/en/gemma-explained-new-in-gemma-2/. [Accessed: Jan. 10, 2025].

[10] Gemma Team, "Gemma 2: Improving Open Language Models at a Practical Size," arXiv preprint arXiv:2408.00118, Aug. 2024.

[11] A. Vaswani et al., "Attention Is All You Need," in Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.

[12] B. Zhang and R. Sennrich, "Root Mean Square Layer Normalization," arXiv preprint arXiv:1910.07467, Oct. 2019.

[13] Justice Hub, "Supreme Court Cases 2010–2020 Dataset," [Online]. Available: https://justicehub.in/dataset/supreme-court-cases-2010-2020. [Accessed: Sept. 20, 2024].

[14] Supreme Court of India, "Official Website," [Online]. Available: https://main.sci.gov.in/. [Accessed: Apr. 20, 2025].

[15] LangChain, "Google Generative AI Integration," [Online]. Available: https://python.langchain.com/api_reference/google_genai/. [Accessed: Dec. 20, 2024].

[16] LangChain, "FAISS Vector Store Integration," [Online]. Available: https://python.langchain.com/docs/integrations/vectorstores/faiss/. [Accessed: Dec. 21, 2024].

[17] M. Shaik and M. Y. Bin Ghazali, “LLM Based News Research Tool Using LangChain with Enhancing Similarity Search and Token Limit,” International Journal for Research in Applied Science and Engineering Technology (IJRASET), vol. 12, no. 1, pp. 366–372, Jan. 2024. [Online]. Available: https://www.ijraset.com/research-paper/llm-based-news-research-tool-using-langchain

[18] Ministry of Home Affairs, The Bharatiya Nyaya Sanhita, 2023 (English Version), Apr. 2024. [Online]. Available: https://www.mha.gov.in/sites/default/files/250883_english_01042024.pdf
[Accessed: Jan. 10, 2025].

[19] Supreme Court of India, Case Categories, [Online]. Available: https://www.sci.gov.in/case-category/
[Accessed: Jan. 1, 2025].

[20] OpenAI, "ChatGPT," [Online]. Available: https://chatgpt.com/. [Accessed: Jan. 20, 2025].

[21] C.-Y. Lin, "ROUGE: A Package for Automatic Evaluation of Summaries," in Proc. Workshop Text Summarization Branches Out (WAS), Barcelona, Spain, 2004, pp. 74–81.

[22] K. Papineni et al., "BLEU: a Method for Automatic Evaluation of Machine Translation," in Proc. 40th Annu. Meeting Assoc. Comput. Linguistics (ACL), Philadelphia, PA, USA, 2002, pp. 311–318.

[23] S. Banerjee and A. Lavie, "METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments," in Proc. ACL Workshop Intrinsic Extrinsic Eval. Measures Machine Translation Summarization, Ann Arbor, MI, USA, 2005, pp. 65–72.

[24] T. Zhang et al., "BERTScore: Evaluating Text Generation with BERT," in Proc. Int. Conf. Learn. Representations (ICLR), Addis Ababa, Ethiopia, 2020.

[25] Supreme Court of India, “Judgment in Criminal Appeal No. 1902 of 2012,” [Online]. Available: https://main.sci.gov.in/supremecourt/2012/1902/1902_2012_Judgement_12-Feb-2019.pdf. [Accessed: Oct. 10, 2024].

[26] Supreme Court of India, “Judgment in Criminal Appeal No. 686 of 2018,” [Online]. Available: https://main.sci.gov.in/supremecourt/2018/686/686_2018_Judgement_26-Feb-2019.pdf. [Accessed: Oct. 10, 2024].

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 118-127. September/October., 2025.
DOI: https://doi.org/10.5935/jetia. v11i55.1953

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

INDUSTRIAL CENTRALIZED CONTROL AUTOMATION SYSTEM USING IOT BASED EMBEDDED SYSTEM

Mehedi Hassain[1], Gonga Sutradhar[2], Nayeem Ibna Harun[3], Abdur Rahman[4] and Ariful Islam[1]

[1, 2, 4, 5] Dept. of Electrical & Electronic Engineering (EEE), International Islamic University Chittagong, Chittagong 4318, Bangladesh.
[2] Dept. of Electrical & Electronic Engineering (EEE), Port City international University, Chittagong 4318, Bangladesh.
[2] Example Institute - EI. Toronto, Canada.

[1]http://orcid.org/0009-0000-0818-245x, [2]http://orcid.org/0009-0009-6919-8514, [3]http://orcid.org/0009-0001-5054-4183
[4]http://orcid.org/0000-0003-0630-4286, [5]https://orcid.org/0000-0002-7211-9890

Email: mdmehedihasan225588@gmail.com, gongag04181@gmail.com, nayeemibnaharun.0@gmail.com, rahmaniiuceee@gmail.com, ariful41415@gmail.com

ARTICLE INFO

Article History
Received: April 29, 2025
Revised: May 20, 2025
Accepted: June 15, 2025
Published: September 30, 2025

Keywords:
Industry Automation system,
Secure data communication,
Remote access,
Automated fire response,
Industry load control.

ABSTRACT

Industry Protection System represents an advanced IoT-based framework aimed at bolstering industrial safety through the real-time detection of fire and gas leaks. Our research focuses on developing a microcontroller-driven system for gas detection, complemented by flame sensors for fire monitoring and thermistors for precise over-temperature detection. The microcontroller continuously processes sensor data, which is transmitted securely in real-time to a cloud-based server via a Wi-Fi module using the HTTP communication protocol, enabling remote monitoring and management. The system integrates proactive alert mechanisms, improving response times to potential hazards. Data visualization is provided through a website and a mobile app. An exhaust fan is automatically activated to cool the environment when the temperature exceeds a predefined threshold. The system also features an LCD for on-site data display, a buzzer for alerts, a water pump for automated fire response, and relays for load control during incidents. A touch sensor allows manual control in case of automatic system failures.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Over the past three decades, between 1990 and 2020, 237 fatal incidents occurred across 4,560 Bangladeshi garment factories, with 94\% caused by fires, 3.38% by boiler explosions, 1.27% by building collapses, and 1.27% by other factors. These incidents resulted in 6,870 casualties, including 5,181 deaths and 1,689 injuries. The highest toll occurred between 2011 and 2015, when 3,800 people were injured, and 1,287 lost their lives [1]. On June 4, 2022, a fire at the BM Inland Container Depot in Sitakunda, Bangladesh, caused over 48 deaths, including 10 firefighters, and injured 200 others. It took 86 hours to extinguish, with financial losses exceeding [2]. A survey of Old Dhaka's Small-Scale Industries found that 75% of workers lack protective gear, 60% of factories have no emergency plans, and only 18% of workers receive proper job training, raising critical safety issues [3].

Previous research has focused separately on gas leakage detection, temperature monitoring, smoke detection, and fire detection. In [4], the author developed an IoT-based system utilizing an Arduino UNO R3 with an ESP8266 Wi-Fi module for internet connectivity. The system incorporates an MQ-6 gas sensor for leak detection, an LM35 temperature sensor for thermal monitoring, and a buzzer for alerts. Sensor data is transmitted to the ThingSpeak server, enabling real-time monitoring and device control. The proposal in [5] introduces a system that utilizes the GSM-based system for detecting industrial faults like temperature spikes, overvoltage, and gas leaks. It triggers a buzzer and sends wireless alerts to the authority when thresholds are exceeded.

In [6], the system uses flame and gas sensors to detect fires and gas leaks, preventing industrial accidents. It monitors temperature, phase faults, and voltage, transmitting data online via Wi-Fi, and the GSM module sends alerts if operators are offline. In [7], the author developed IoT-based fire detection and control system that utilizes a NodeMCU, along with smoke, temperature and MQ2 sensors to assess fire severity. It sends real-time alerts and notifications via the Blynk cloud while automatically activating sprinklers and

extinguishers. The proposal in [8], the authors developed a smart fire detection system that uses NodeMCU, temperature, smoke, flame, LDR, and MQ2 sensors to assess fire severity. It transmits real-time alerts and activates sprinklers and extinguishers via the Blynk cloud upon flame detection. In [9], the author developed a system that uses MQ2, MQ7, and MQ135 gas sensors strategically placed throughout the industrial facility to monitor various gases. The sensors send real-time data to a centralized controller, enabling authorized personnel to access information remotely and take prompt action from anywhere with an internet connection. The research in [10] presents a Safety Monitoring System for the cracker industry that integrates a DHT11 sensor for temperature and humidity, gas sensors for hazard detection, and the KNN algorithm to analyze historical data for predicting safety risks.

In [11], the authors developed a fire detection system that uses an Arduino and a flame sensor to monitor heat intensity in homes and offices. The system alerts users with a buzzer and transmits messages to a virtual terminal via a Wi-Fi module. In [12], the authors developed a system to monitor temperature, pressure, humidity, and gas concentration. These sensors transmit real-time data to a centralized monitoring system, and using GSM, the system sends automated SMS alerts to designated personnel when deviations occur. The authors developed a system in [13] that uses sensors connected to a Raspberry Pi 3 for continuous monitoring. A temperature rises triggers SMS alerts, image capture, emails to the fire station, buzzer activation, and webpage updates, while similar alerts are issued for gas or smoke detection and high humidity. In [14], the research introduces a fire detection system that uses multiple sensors and a camera connected to a Raspberry Pi 3 to identify affected areas.

To reduce false alarms, it validates fire alerts by sending an image and a message. If confirmed, the system quickly notifies the local fire department. In [15], the authors developed a hybrid system for detecting fires, gas leaks, and noise pollution. Using a Raspberry Pi with sensors and a servo-mounted camera, it identifies affected areas, sends alerts via LAN, and triggers alarms. If confirmed, the system opens the water valve and notifies the fire brigade. The authors developed a building safety system in [16], which monitors fire, smoke, and temperature throughout the building, sending real-time data to Firebase. It alerts users via an Android app with live location updates, notifies emergency services, provides safe exit routes, and triggers a manual alarm to warn occupants. In [17], the authors developed a system that detects fire using a flame sensor, gas leaks via a gas sensor, and intrusions with a PIR sensor, alerting workers in real-time. If an intruder triggers a fire, the system captures an image and sends instant alerts via message and email. In [18], the authors developed an IoT-based fire prevention and response system using Arduino Uno and Raspberry Pi B+, along with gas, smoke, and temperature sensors. The system sends data to a server for SVM-based validation, triggering alerts to authorities if anomalies are detected. In [19], the authors developed an automated fire monitoring system that employs the nRF24L01 module for long-range data transmission, making it suitable for industrial use.

Most previous research emphasizes gas and fire detection, but often misses automatic water spray activation upon fire detection and proper monitoring systems. In addition, integration of cooling system activation and automatic load control remains unaddressed, leaving significant gaps in industrial safety solutions. Our research addresses the gaps in previous studies by developing a smart, automated industrial safety solution. The system uses gas, fire, motion, temperature and optical sensors to monitor conditions, with an LCD display for onsite data display and a buzzer for alerts. It includes an exhaust fan for cooling, a water pump for automatic fire response, and relays for load control during incidents. A touch sensor enables manual control if automatic functions fail. Data are transmitted from the transmitter to the receiver through the HTTP communication protocol. In the receiver section, system operations are monitored through a micro-controller, and a GSM module is used for SMS and onsite data availability during power crises. Data are received from the transmitter via the HTTP communication protocol. Data are monitored through website and mobile app visualization. The system sends alerts via email and SMS notifications when a fire incident occurs.

II. METHODOLOGY

This research focuses on developing an Industrial Centralized Control System with an emergency alert. This section details the circuit diagram, data transmission, and onsite mobile network connectivity.

II.1 SYSTEM BLOCK DIAGRAM

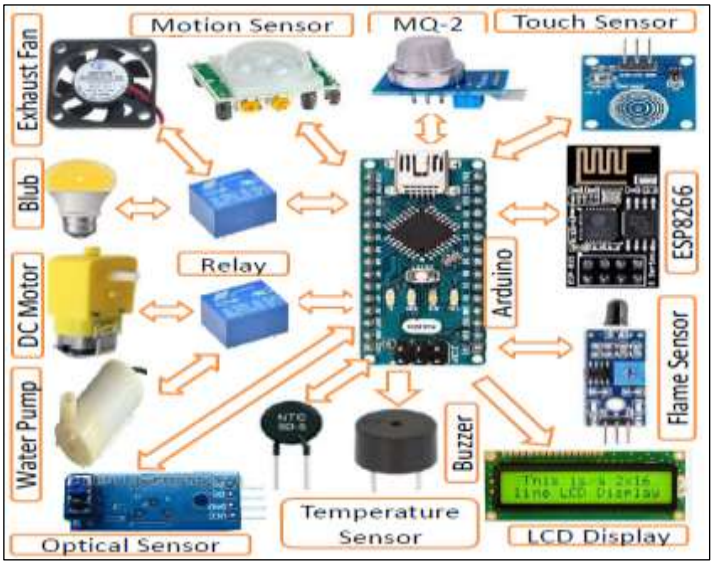


Figure 1: Block diagram of the Project (Transmitter).

Source: Authors, (2025).

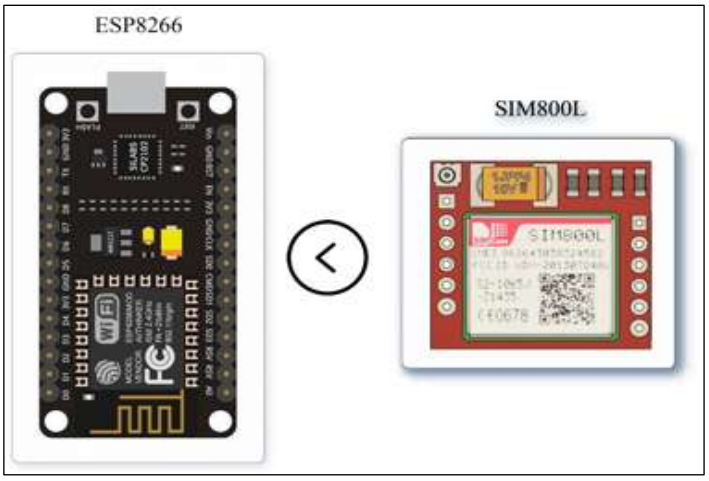


Figure 2: Block diagram of the Project (Receiver).
Source: Authors, (2025).

In Figure 1, system transmitter block diagram is presented. The transmitter unit uses several sensors to monitor industrial parameters, including gas, flame, temperature, motion, and optical sensors. The Arduino Nano serves as the system's microcontroller, while the ESP Wi-Fi module transmits data to the receiving unit using the HTTP protocol. A relay switches loads on or off, and the water pump is controlled accordingly. The LCD displays real-time data. It features an exhaust fan for cooling, a water pump for fire response, and a touch sensor for manual operation if automation fails. In Figure 2, illustrates a block diagram of the system receiver. The receiver unit uses two components: a NodeMCU and a GSM module. The microcontroller gathers data, saves it on the server, and sends messages to the appropriate authority.

II.2 SYSTEM CIRCUIT DIAGRAM

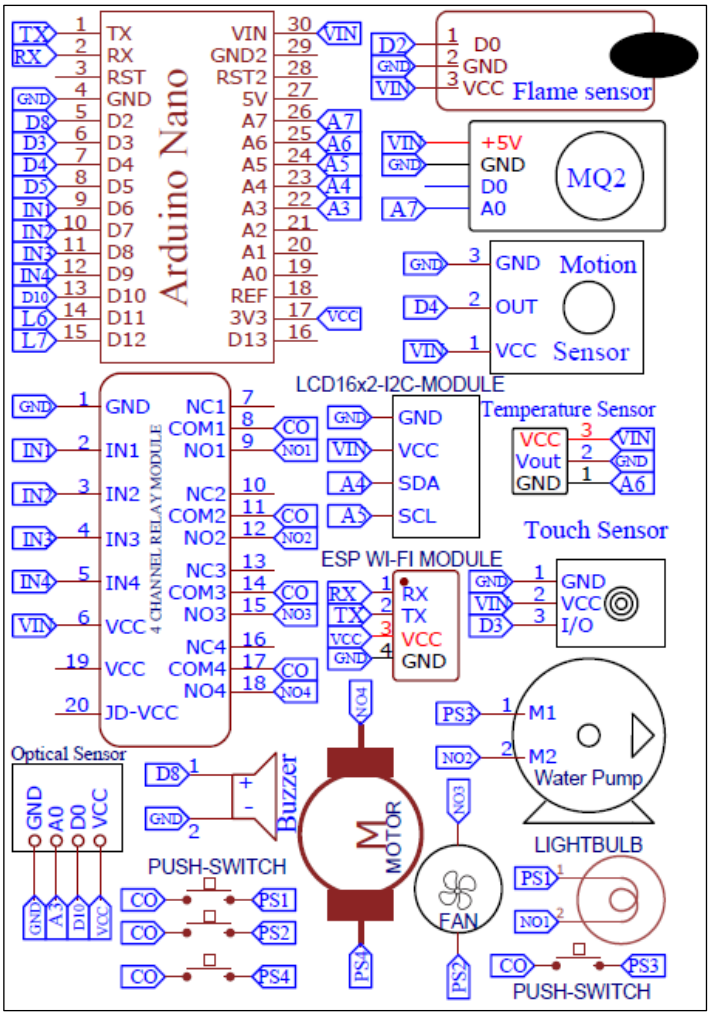


Figure 3: Circuit diagram of the system (Transmitter).

Source: Authors, (2025).

In Figure 3, the transmitter section shows the operation of the Arduino Nano, with the Flame, Motion, Buzzer and Touch sensors connected to digital pins D2, D3, D4, and D8 respectively. The MQ-2 gas sensor, temperature sensor, LCD module, and ESP Wi-Fi module are connected to analog pins A4, A5, A6, A7, and the TX/RX pins, respectively. The optical sensor is connected to digital pin D10 and analog pin A3 of the Arduino. The motor, fan, water pump, and light are controlled via a push switch using a 4-channel relay module, with connections to Arduino Nano pins D6, D7, D8, and D9, respectively. In Figure 4, in the receiver section TX and RX pins of the GSM SIM800L are connected to the D0 and D1 pins of the ESP8266 Wi-Fi module, respectively.

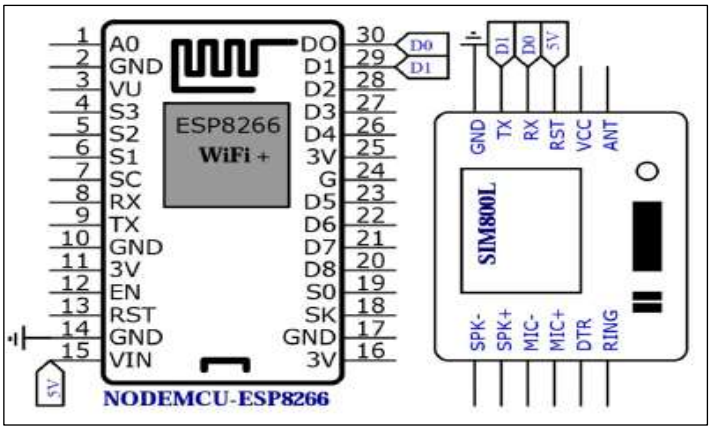


Figure 4: Circuit diagram of the system (Receiver).
Source: Authors, (2025).

II.3 DATA TRANSMISSION AND WEBSITE STROAGE

In Figure 5, the HTTP protocol is employed to securely transmit sensor data from the operation to the server. Data is securely stored in a MySQL database using Next.js and a Node.js server, that ensures the dependability of the Industrial centralized Control system. Through the ESP8266 Wi-Fi module, the sensor data is sent to the Node.js server. This secure repository provides access to the Industrial centralized control data through an easy-to-use Next.js online interface. Scalability is assured by MySQL, which efficiently manages massive data volumes.

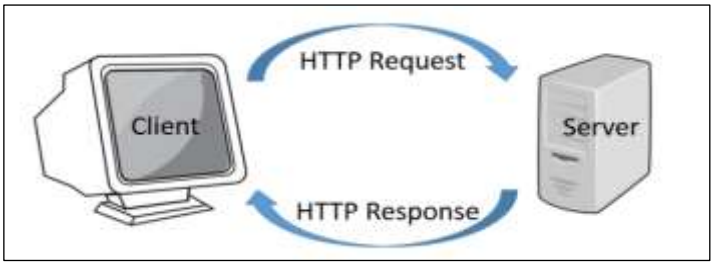


Figure 5: HTTP Communication Protocol work flow.
Source: Authors, (2025).

II.4 FLOWCHART

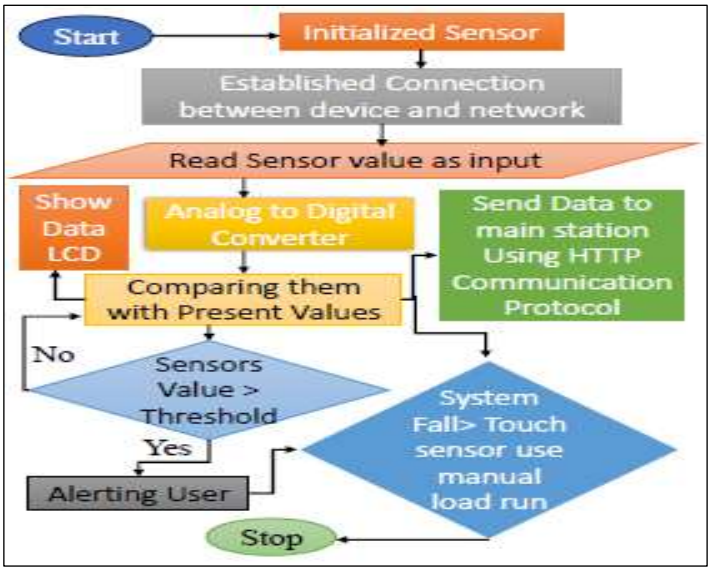


Figure 6: Flowchart of the Project (Transmitter).
Source: Authors, (2025).

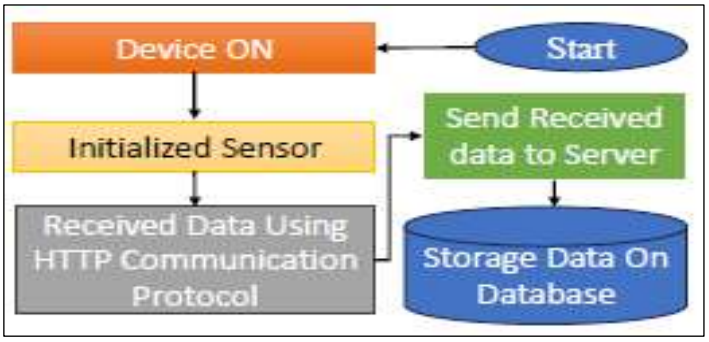


Figure 7: Flowchart of the Project (Receiver).
Source: Authors, (2025).

In Figure 6, the transmitter flowchart is demonstrated. Sensor data are compared with the threshold value, and when it exceeds, an onsite fire alert is triggered to notify the authority. If the system does not operate automatically, the touch sensor allows for manual restart. The data is gradually displayed on the LCD monitor. The receiver flowchart depicted in Figure 7, processes this data, combining it with information from the GSM module. The system sends an SMS message, and the server stores the information. The user can easily monitor and visualize the data through the website and mobile app.

II.5 PROTOTYPE

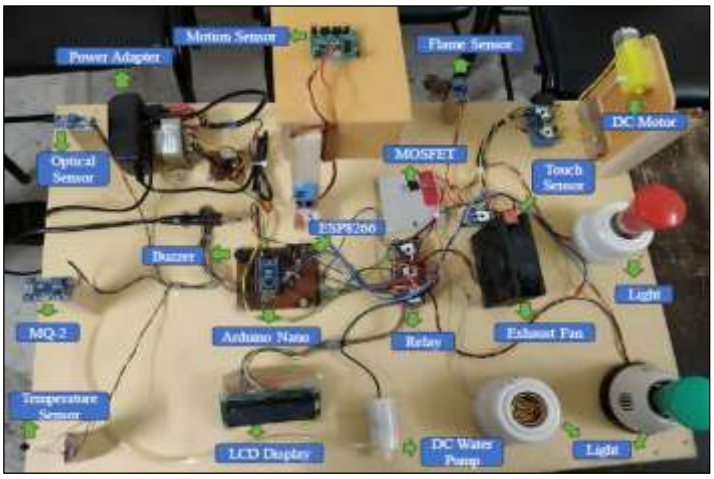


Figure 8: System Overview (Transmitter).
Source: Authors, (2025).

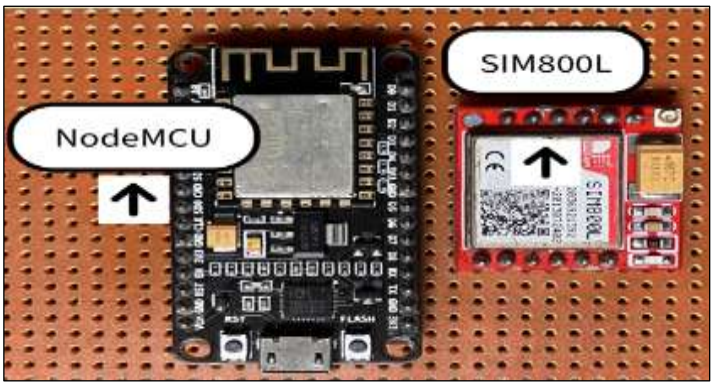


Figure 9: System Overview (Receiver).
Source: Authors, (2025).

In Figure 8, illustrates the implementation of the system's transmitter prototype. The prototype of the Industrial Centralized Control Automation System, featuring components such as an Arduino Nano, ESP8266, SIM800L, MQ-2, relay, optical sensor, motion sensor, flame sensor, an exhaust fan, bulb, buzzer, a water pump, and an LCD module. The system responds automatically to threshold breaches, displaying data on the LCD and triggering a buzzer for alerts. It includes mechanisms for cooling, fire response, and load control, with a touch sensor for manual operation if automatic functions fail. Data is transferred to the cloud via the HTTP protocol. In

Figure 9, demonstrates the system overview of the receiver part. The SIM800L ensures data availability in the station section during power crises or major issues and provides alerts via SMS notifications. Additionally, the microcontroller transmits operational signals and sends alerts via email notifications. Authorities can easily monitor data through a website and mobile app, with the system generating graphs for further analysis.

III. RESULTS AND ANALYSIS

In Figure 10, where the y-axis displays flame levels with a threshold value of 0 or 1, the x-axis displays continuous dates. The threshold value for this performance is set to 1. In Figure 11, shows the Gas Concentration Over Time of the system. The MQ2 Gas Sensor data from November 1, 2023, which shows dynamic variations in gas concentration, are shown in Fig.10. At 8:00 a.m, a moderate 20 ppm was recorded, rising to 25 ppm at 9:00 a.m, indicating increased gas presence.

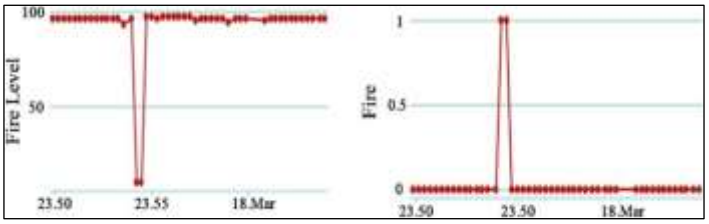


Figure 10: Performance of flame sensor data.
Source: Authors, (2025).

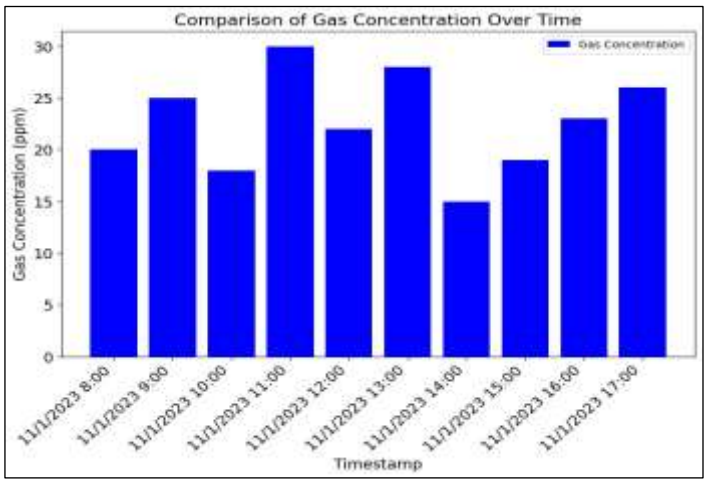


Figure 11: Gas Concentration Over Time of the system.
Source: Authors, (2025).

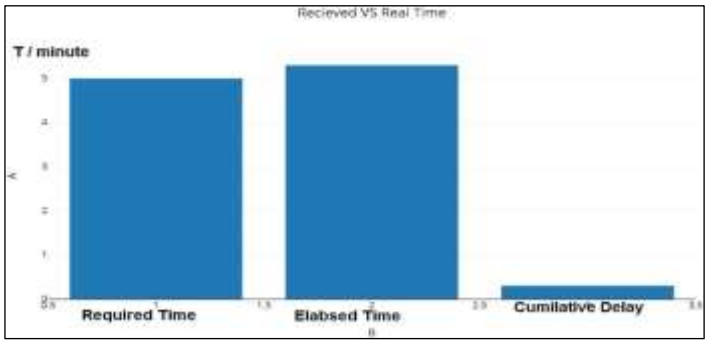


Figure 12: Analysis of the System operation.
Source: Authors, (2025).

In Figure 12, shows the system's operation time versus elapsed time, indicating the actual time is slightly below the required time, with minimal cumulative delay, reflecting fast response.

Figure 13: Website Visualization.
Source: Authors, (2025).

Figure 14: Mobile Visualization.
Source: Authors, (2025).

In Figure 13, and Figure 14, showcase data visualization on the website and mobile app. The system uses Next.js and a Node.js server to securely store sensor data in a MySQL database, ensuring reliability. The ESP8266 transmits data via HTTP, while the scalable MySQL database supports real-time industry monitoring through a user-friendly interface. In Figure 15, shows the LCD displaying output values, allowing the authority or user to easily monitor internal data. In Figure 16, and Figure 17, demonstrate the visualization of email and SMS notifications. When a fire occurs or any value exceeds the threshold, the user receives a notification.

Figure 15: Mobile Visualization.

Source: Authors, (2025).

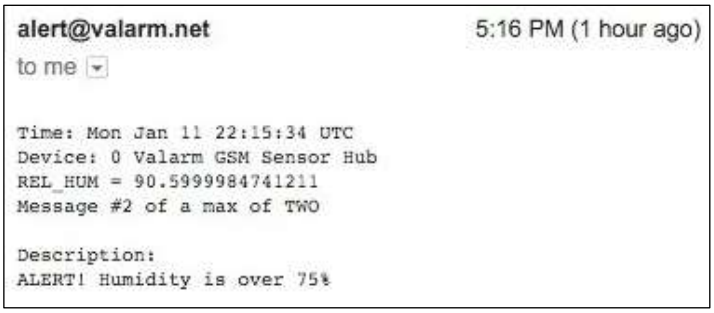


Figure 16: Email notification.
Source: Authors, (2025).

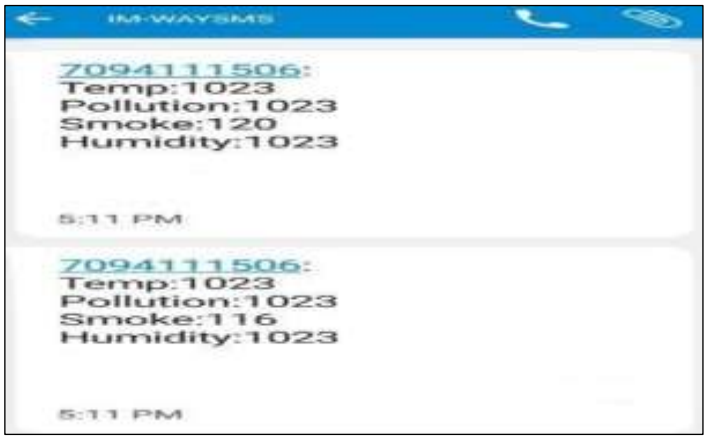


Figure 17: Email notification.
Source: Authors, (2025).

In Table 1, the fire smoke levels range from low to high, triggering a fire alarm and activating the exhaust fan accordingly. In Table 2, it categorizes temperature sensor levels (low to high) and generates an over-temperature alarm if thresholds are exceeded. In Table 3, the flame sensor result (detect/no detect) triggers a fire alarm, activates the pump motor, and initiates water spray when a flame is detected.

Table 1: Result for Smoke Sensor.

SL	Smoke Level	Smoke Alarm	Exhaust Fan
1	Low (0-50) PPM	Off	Off
2	Medium (51-100)PPM	Off	Off
3	High (101PPM<above)	On	On

Source: Authors, (2025).

Table 2: Temperature Sensor Result.

SL	Temperature Value	Over Temp. Alarm	Light(Heater)
1	Low (0°- 22°)	Off	On
2	Medium(23° - 45°)	Off	On
3	High(45°<Above)	On	Off

Source: Authors, (2025).

Table 3: Flame Sensor Result.

SL	Fire Value	Fire Alarm	Pump Motor
1	No Detect	Off	Off
2	Detect	On	On

Source: Authors, (2025).

Table 4: Comparison with Other's Research.

Paper ID	Data in LCD	SMS Alert	Buzzer Alert	Email Notify	Mobile App	Web site	Fire Detect	Temp. Monitor	Water Spray Fire Fightig	Gas Detect	Exhaust Fan (Cooling)	Optical & Motion detect	Industry Load Control	Manual start (Touch sensor)
Ref[4]	No	No	No	No	No	No	Yes	No	No	No	No	No	No	No
Ref[5]	No	No	No	No	No	No	Yes	No	Yes	Yes	No	No	No	No
Ref[6]	Yes	No	Yes	No	No	Yes	Yes	No	No	No	No	No	No	No
Ref[7]	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	No	No

Ref[8]	No	Yes	Yes	No	Yes	No	Yes	Yes	No	No	No	No	No	No
Ref[9]	No	No	No	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No
Ref[10]	No	Yes	Yes	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No
Ref[11]	No	No	Yes	No	No	Yes	Yes	Yes	No	No	No	No	No	No
Ref[12]	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	No	No	No	No
Ref[13]	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	No	No	No
Ref[14]	No	No	Yes	No	No	No	Yes	No	No	Yes	No	No	No	No
Ref[15]	No	Yes	Yes	No	Yes	No	Yes	No	No	Yes	No	No	No	No
Ref[16]	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Ref[17]	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	No	No	No	No
Ref[18]	No	No	Yes	No	No	No	Yes	No	No	Yes	No	No	No	No
Ref[19]	No	No	Yes	Yes	Yes	Yes	Yes	No	No	Yes	No	No	No	No
Propose	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Source: Authors, (2025).

Table 5: Calculating the proposed system's price.

SL	Component Name	Quantity	Price (BDT)
1	Arduino Nano	1	400TK
2	ESP8266	1	320TK
3	MQ-02	1	150TK
4	Relay	7	350TK
5	Buzzer	1	20TK
6	Touch sensor	1	80TK
7	Temperature Sensor	1	10TK
8	LCD Display	1	340TK
9	Motion Sensor	1	110TK
10	Wifi Sensor	1	420TK
11	Flame Sensor	1	69TK
12	DC Water Pump	1	160TK
13	DC Motor	1	70TK
14	DC Exhaust Fan	1	75TK
15	Optical Sensor	1	150TK
16	Power Supply	1	250TK
17	Miscellaneous Cost	-	350TK
-	Total Cost	-	3574TK(29.24USD)

Source: Authors, (2025).

In Table 4, presents a comparative analysis with existing research, demonstrating that the proposed system outperforms others in visualization for alert detection and automated control of critical parameters over manual methods. In Table 5, presents a cost breakdown of the proposed system, including components, quantities, and pricing (in TK). The total cost of the system is 3574 TK (29.24 USD), making it an economical solution for monitoring industrial safety. Future work will focus on real-world deployment to evaluate long-term performance, including continuous data collection, fault analysis, and adaptation to dynamic operational conditions for enhanced reliability.

IV. CONCLUSIONS

Our research presents a designed industrial automated system that ensures timely alerts and securely stores data on a server. Authorities can access this information securely, while an LCD displays real-time data. The system provides alerts through email and SMS notifications. Future updates will incorporate machine learning. In general, the system helps users to easily monitor the data.

V. AUTHOR'S CONTRIBUTION

Conceptualization: Md. Mehedi Hassain.
Methodology: Md. Mehedi Hassain, Gonga Sutradhar and Md. Nayeem Ibna Harun.
Investigation: Md. Mehedi Hassain and Abdur Rahman.
Discussion of results: Md. Mehedi Hassain, Gonga Sutradhar and Md. Ariful Islam.
Writing – Original Draft: Md. Mehedi Hassain.
Writing – Review and Editing: Md. Mehedi Hassain and Md. Nayeem Ibna Harun.
Resources: Md. Mehedi Hassain.
Approval of the final text: Md Mehedi Hassain, Gonga Sutradhar, Md. Nayeem Ibna Harun, Abdur Rahman and Md. Ariful Islam.
Funding: This research received no external funding.

VI. REFERENCES

[1] Islam, S.; Mia, M.A.H.; Abedin, M.Z. Evaluation of Causes and Effects of Fire and Other Safety Incidents in Readymade Garment 197 Industry of Bangladesh. American Journal of Mechanical and Industrial Engineering 2022, 7, 13–30.

[2] Tahmid, M.; Jahan, N.; Tasnim, N.; Syeda, S.R. Major Accident Control in Temporary Storage Facilities: An Investigative 199 Consequence Analysis of the Bangladesh Container Depot Accident 2022. ACS Chemical Health & Safety 2022, 29, 494–505.

[3] Tarek Haider, M. Study of safety and security issues in the working environment of small scale industries of old Dhaka 2007.

[4] Sekar, R.A.; Prabakaran, T.; Sudhakar, A.; Kumar, R.S. Industrial automation using IoT. In Proceedings of the AIP Conference Proceedings. AIP Publishing, 2022, Vol. 2393.

[5] Abinand, A.; Sairabanu, S. Smart industrial fault monitoring and alert system. In Proceedings of the AIP Conference Proceedings. AIP Publishing, 2023, Vol. 2861.

[6] Islam, M.N.; Rahman, M.A.; Ashrafi, K.M.A.; Ahmed, T. Development of an IoT based industrial fault detection and diagnosis system. Asian Journal For Convergence In Technology (AJCT) ISSN-2350-1146 2023, 9, 18–22.

[7] Bharti, T.; Saini, M.L.; Kumar, A.; Tiwari, R. Design and implementation of industrial fire detection and control system using internet of things. In Artificial Intelligence, Blockchain, Computing and Security Volume 1; CRC Press, 2024; pp. 310–315.

[8] Saini, M.L.; Kumar, R.; Sati, D.C.; Bharti, T. Fire Monitoring and Prevention System Based on the Severity of Fire Using NodeMCU and Cloud. In Proceedings of the 2024 International Conference on Intelligent and Innovative Technologies in Computing, Electrical and Electronics (IITCEE). IEEE, 2024, pp. 1–6.

[9] Jadhav, R.J.; Radhakrishnan, P.; Jadhav, D.A.; Ashreetha, B.; Divya, J.; Mukherjee, S. Internet of Things Enabled Gas Leakage Detection Over Industrial Areas using Powerful MQ Series Sensor and Controller. In Proceedings of the 2024 International Conference on Inventive Computation Technologies (ICICT). IEEE, 2024, pp. 1679–1686.

[10] Devi, V.S.; Vishwa, K.; Vaseegaran, K.; et al. Fire Detection and Prediction using Machine Learning for Fireworks Industry. In Proceedings of the 2024 International Conference on Inventive Computation Technologies (ICICT). IEEE, 2024, pp. 884–889.

[11] Ibrahim, S.; Shukla, V.K.; Gupta, S.; Sharma, P. Addressing Smart Fire Detection Management through the Internet of Things and Arduino. In Computational Intelligence in Urban Infrastructure; CRC Press, 2024; pp. 171–179.

[12] Jumde, N.T.; Kolhe, H.J.; Talawdekar, V.L.; Pal, R.B.S. Industrial Parameter Monitoring with SMS Alert. International Journal of Modern Developments in Engineering and Science 2024, 3, 13–17.

[13] Priya, L.; Vedhapriyavadhana, R.; Ignisha Rajathi, G.; Allwyn Kingsley, G.; Johny Elton, R. An IOT integrated fire monitoring and detection system using Raspberry Pi. Turkish Journal of Physiotherapy and Rehabilitation 2021, 32.

[14] Manikandan, G.; Karunkuzhali, D.; Geetha, D.; Kavitha, V. Design of an IoT approach for security surveillance system for industrial process monitoring using Raspberry-Pi. In Proceedings of the AIP Conference Proceedings. AIP Publishing, 2022, Vol. 2519.

[15] Masum, A.K.M.; Shan-A-Alahi, A.; Al Noman, A.; Uddin, M.N.; Azam, K.I.; Rakib, M.G.S. IoT-Based Smart Monitoring System to Ensure Worksite Safety—A Context of Garment Industry in Bangladesh. In Proceedings of the Inventive Communication and Computational Technologies: Proceedings of ICICCT 2019. Springer, 2020, pp. 1285–1293.

[16] Majmudar, M.; Jethwa, J.; Patil, A.; Harpale, J.; Bhaskar, S. IoT Technologies Based Fire and Safety Alerting System Using Android Application. In Proceedings of the Proceedings of the Second International Conference on Information Management and Machine Intelligence: ICIMMI 2020. Springer, 2021, pp. 483–490.

[17] Praveena, N.; Deepa, G.; Manju, M. Design of an IOT Approach for Industrial Security Surveillance System. International Journal of Scientific & Technology Research 2020, 9, 6444–6448.

[18] Habib, A.; Sharma, S.; Rahman, M.R.; Haque, M.N.; Bhuyan, M.A.I. Fire Safety and Supervision System: Fire Hazard Monitoring Based on IoT. In Proceedings of the Proceedings of the International Conference on Big Data, IoT, and Machine Learning: BIM 2021. Springer, 2022, pp. 367–377.

[19] Kabir, S.; Riyad, M.E.H.; Tanisha, U.T.; Bhuiyan, M.H.R.; Toha, T.R.; Alam, S.M.M. Automatic Fire Monitoring System in Textile Industry. In Proceedings of the 2020 11th International Conference on Electrical and Computer Engineering (ICECE). IEEE, 2020, pp. 471–474.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 128-139. September/October., 2025.

DOI: https://doi.org/10.5935/jetia. v11i55.2513

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

RISK QUANTIFICATION IN MANUFACTURING INDUSTRY INVESTMENTS: A STOCHASTIC APPROACH WITH ARTIFICIAL INTELLIGENCE

Rafael Vieira da Silva[1], Enzo Morosini Frazzon[2]

[1, 2,] Production and Systems Engineering Department, Federal University of Santa Catarina, Florianópolis, Brazil

[1]https://orcid.org/0009-0001-0581-0008, [2]https://orcid.org/0000-0002-6629-6938

Email: rafael.vieira@dep.eng.br, enzo.frazzon@gmail.com

ARTICLE INFO

Article History
Received: August 14, 2025
Revised: August 20, 2025
Accepted: September 15, 2025
Published: September 30, 2025

Keywords:
Investments,
Supply Chain,
Risk Evaluation,
Techno-Economic Feasibility,
Decision-Making Methods.

ABSTRACT

Manufacturing investment decisions are often hindered by significant uncertainty. This paper introduces a conceptual model that integrates stochastic simulation with machine learning to quantify investment risk. Our hybrid approach employs a Monte Carlo simulation using time series data of fixed, variable, and investment costs as inputs. To enhance the simulation's realism, Long Short-Term Memory (LSTM) recurrent neural networks forecast the trend components of these series, while a Vector Autoregression (VAR) model captures their inter-correlations. This framework generates a multitude of potential scenarios, each evaluated through a mathematical model of the supply chain to produce a distinct cash flow. The subsequent application of investment metrics, such as Net Present Value (NPV) and Discounted Payback, to the distribution of these cash flows enables a comprehensive statistical analysis of the investment's risk profile, thereby providing robust support for strategic decision-making.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Investment risks in the manufacturing industry are significant due to various intrinsic characteristics of the production chain. The central position of the manufacturing sector within the supply chain makes it particularly sensitive to fluctuations in the supply of raw materials and changes in demand. This sensitivity is further intensified by the highly specialized nature of its production processes [1]. Consequently, structural changes in the supply chain often require substantial investments to adapt production processes. In addition to these factors, a wide range of risks can affect the supply chain, including regulatory changes, natural disasters, disruptions in logistics systems, shifts in consumer preferences, wars, and deliberate attacks on industrial infrastructure. These events are often unpredictable and their consequences uncontrollable [2].

Despite the high level of risk, investments in the supply chain are essential for maintaining competitiveness and meeting market demands. This creates a duality between the inherent risk of investing and the necessity of doing so [3], [4]. Therefore, investments in the manufacturing industry require extensive planning, encompassing a minimum time horizon to ensure return on investment. Motivated by this need, several researchers have developed methods for quantifying investment risks and supporting decision-making [5]. Among these efforts, stochastic approaches have stood out for assessing investment risks. However, there remains a research gap regarding the application of neural network tools for this purpose [6](submitted for publication). This integration is essential, as artificial intelligence techniques excel at identifying patterns, while stochastic methods are particularly effective in analyzing scenarios under uncertainty. The current context is especially conducive to the development of such approaches, due to advances in computational capabilities, improvements in programming libraries and languages, and the growing popularity of artificial intelligence.

Given this context, this study proposes a conceptual model that integrates neural networks and stochastic methods to quantify investment risks in the manufacturing industry. This paper is organized as follows: Chapter II presents the theoretical background of the components of the model; Chapter III introduces the proposed conceptual model; Chapter IV provides the results and discussions approach; and Chapter V provides the concluding remarks. This study is limited to the proposal of a conceptual model, presenting its

structure, tools, and implementation methods. However, the implementation of the proposed method is beyond the scope of this article and will be addressed in future research.

II. THEORETICAL REFERENCE

II.1 MONTE CARLO METHOD

The Monte Carlo method is a computational approach based on simulations generated from random numbers applied to probabilistically defined models. It has proven effective in analyzing complex systems where deterministic models are infeasible or inaccurate. Its flexibility and ability to handle uncertainty have made it widely adopted in various fields [7]. For instance, it is used in finance for pricing financial options, portfolio evaluation, and risk analysis; in engineering for system reliability assessment and maintenance planning [8], [9]; and in computer science for high-dimensional integration and machine learning algorithms [10], [11].

For investment risk assessment in the manufacturing industry, the Monte Carlo method can handle volatility in commodity prices, finished goods, and operational costs. This flexibility in modeling multiple variables is crucial in sensitive environments such as investment evaluation. Its strength becomes even more evident when integrated with other methods, such as Vector Autoregression (VAR), enabling multivariate analyses [12]. Correlation analyses between variables are essential for investment assessment, especially considering the interdependence of products within supply chains [13]. Given that uncertainty in time series tends to increase over time, Monte Carlo simulations are well-suited for long-term projections, as they naturally account for increasing volatility [14].

The Monte Carlo method typically follows a systematic structure that may vary by context but often includes: problem definition, statistical analysis, random number generation, scenario simulation, and result analysis [15]. The process begins by defining the scope and identifying initial parameters and variables. Statistical analysis includes data collection, preliminary analysis, and fitting statistical distributions. To incorporate uncertainty, random numbers are generated according to these distributions [16]. Each set of random numbers produces a new simulated scenario. Simulation results are compiled to compute metrics such as mean, median, variance, and confidence intervals [15], [16]. In the context of time series analysis, a typical Monte Carlo modeling approach involves decomposing the series into trend, seasonality, and residual components, as expressed in Equation 1:

Monte Carlo Equation

$$Y_t = Trend_t + Seasonality_t + X_t \quad (1)$$

Where: The $Trend_t$ represents the systematic behavior of a variable, ignoring cyclical variations and noise. It can be modeled using linear, exponential, polynomial trends, or more advanced methods such as neural networks [17]; The $Seasonality_t$ refers to predictable, cyclic components of the time series, often linked to natural, social, or economic events. Unlike trends and residuals, seasonality occurs at fixed intervals (e.g., days, months, years) [17], And the Residuals (X_t) capture unpredictable variations not explained by trend or seasonality. These components reflect external or internal conditions and exhibit no clear periodicity or dependency over time [17]. If residuals follow a normal distribution, they can be modeled as white noise – $\epsilon_t \sim N(0, \sigma^2)$. For highly volatile systems with cumulative noise, Brownian noise (or Brownian motion) is more appropriate, as it better describes the long-term stochastic behavior of asset time series. As shown in Equation 2 the noise at time t is the sum of white noise and the accumulated noise from the previous time step [18].

Brownian Motion Equation

$$X_t = X_{t-1} + \epsilon_t \quad (2)$$

Where:
X_t: noise value at time *t;*
X_{t-1}: noise value at time $t-1$;
ϵ_t: white noise at time t.

II.2 TIME SERIES DECOMPOSITION

Each time series component must be treated individually in the analytical process. Therefore, decomposing the time series into trend, seasonality, and noise components is essential. One effective method for this task is STL decomposition (Seasonal and Trend decomposition using Loess), which presents several advantages: robustness to non-stationary data, flexibility in handling complex seasonal patterns, resistance to noise and structural changes, and the direct extraction of trend, seasonality, and residual components through iterative optimization [19], [20]. However, it requires careful parameter tuning, entails higher computational complexity, assumes constant seasonality, and may encounter limitations with very short or very long time series [21].

The STL decomposition process begins by defining the cycle length of the time series. For monthly data, the cycle is typically annual, resulting in a period of 12 months. Equation 3 calculates the initial seasonal component, which serves as the basis for the first iteration. This involves summing all observations that occur in the same position within each cycle (e.g., every January), and dividing by the number of complete cycles. The next step involves centering the initial seasonal values using Equation 4, which removes bias and ensures that the seasonal component does not interfere with other components of the time series.

Initial Seasonality Equation

$$S_i = \frac{1}{N}\sum_{t \in \text{ciclo i}} Y_t \quad (3)$$

Centered Seasonality Equation

$$S_t = S_i - \frac{1}{p}\sum_{i=1}^{p} S_i \quad (4)$$

Where:

S_i: estimated seasonal component at position i within the cycle;
N: total number of complete cycles in the time series;
Y_t: value at time t;
p: number of periods within one cycle.

The trend is calculated by subtracting the seasonal component from the original series, as shown in Equation 5 resulting in the adjusted series Yt', which contains both noise and trend. The Loess method is then applied to Yt' to remove the noise and extract the smoothed trend component, as shown in Equation 6.

Time Series less Seasonality Equation

$$Y_t^{'} = Y_t - S_t \tag{5}$$

Trend Loess Equation

$$T_t = \text{Loess}\left(Y_t^{'}\right) \tag{6}$$

Since the initial seasonal estimate may include elements of trend and noise, STL applies Equation 7 to subtract the extracted trend from the original time series, yielding the intermediate series $Y_t^{''}$. Loess is then applied again to this intermediate series to refine the seasonal component, as shown in Equation 8 [22].

Time Series less Tend Equation

$$Y_t^{''} = Y_t - T_t \tag{7}$$

Seasonality Loess Equation

$$S_t = \text{Loess}\left(Y_t^{''}, \text{periodicity } p\right) \tag{8}$$

The residual component (noise) is calculated by subtracting both trend and seasonal components from the original time series, as given in Equation 9. STL is an iterative method, and Equations 5 through Equation 9 must be repeated iteratively, where *k* represent the current iteration.

White Noise Equation

$$\epsilon_t = Y_t - T_t - S_t \tag{9}$$

Max Error Equation

$$(\sum \epsilon_t^2)^k - (\sum \epsilon_t^2)^{-k} > Error_{max} \tag{10}$$

The number of iterations can be adjusted based on criteria such as the threshold defined by Equation 10, where the stopping condition is determined by the difference in the sum of squared residuals between two consecutive iterations. Additional parameters may also be set, such as a maximum number of iterations, in order to stop the process in cases of non-convergence [22].

II.3 RECURRENT NEURAL NETWORKS AND LSTM

The architecture of traditional multilayer perceptrons (MLPs) includes a fixed number of input neurons defined by the dataset's dimensionality. Each input is connected to all neurons in the hidden layers, which means the data is processed independently - ignoring the temporal structure of sequential data [23]. Furthermore, traditional networks lack memory mechanisms and cannot retrieve previous computations, making them unsuitable for tasks such as time series modeling or natural language processing. To address these limitations, Recurrent Neural Networks (RNNs) were introduced by David Rumelhart, Geoffrey Hinton, and Ronald J. Williams in the 1980s [24]. Despite improvements over traditional networks, RNNs face training challenges when dealing with long sequences. Recurrent weights are multiplied across many time steps, leading to vanishing or exploding gradients, which hinder convergence [25], [26].

To address this, Sepp Hochreiter and Jürgen Schmidhuber proposed the Long Short-Term Memory (LSTM) network in 1997, introducing the concepts of cell state and gates [26]. The cell state and hidden state give LSTM the ability to store both long- and short-term dependencies - often referred to as the network's "memory." Gates, typically controlled by sigmoid activations, act as dynamic weights that determine which information to retain, discard, or update. These innovations make LSTM networks more robust against gradient problems and capable of adaptively learning patterns in sequential data [27].

Mathematically, the forget gate is defined in Equation 11, controlling how much of the previous cell state is retained. Simultaneously, the input gate (Equation 12) and candidate cell state (Equation 13) determine the contribution of new information. The updated cell state is calculated using Equation 14.

Forget Gate Equation

$$f_t = \sigma(w_f[h_{t-1}, x_t)] + b_f) \tag{11}$$

Input Gate Equation

$$i_t = \sigma(w_i[h_{t-1}, x_t)] + b_i) \tag{12}$$

Candidate Cell State Equation

$$\widetilde{C_t} = tanh(w_{\tilde{C}}[h_{t-1}, x_t] + b_{\tilde{C}}) \quad (13)$$

Updated Cell State Equation

$$C_t = f_t \odot C_{t-1} + i_t \odot \widetilde{C_t} \quad (14)$$

Where:

σ: sigmoid function;
w: gate weight;
b: bias weight;
tanh: tanh function.
$\odot$: element-wise multiplication.

The hidden state reflects the current output of the LSTM cell and is calculated using Equation 15, combining the output gate (Equation 16) and the updated cell state:

Output Gate Equation

$$o_t = \sigma(w_o[h_t, x_t] + b_t) \quad (15)$$

Updated Hidden State Equation

$$h_t = o_t \odot \tanh(C_t) \quad (16)$$

In this way, LSTM networks are capable of modeling long-range dependencies in sequential data, overcoming the limitations of traditional feedforward and basic RNN architectures. They are widely used for time series forecasting and sequence modeling with high efficiency and precision [28].

II.4 VAR-CHOLESKY METHOD

The Vector Autoregressive (VAR) method was introduced by Christopher A. Sims in 1980 through the publication Macroeconomics and Reality. It evolved from the autoregressive (AR) model and allows for the simultaneous modeling of lagged interdependencies among multiple time series. In this framework, all variables are treated as interdependent, eliminating the need to distinguish between dependent and independent variables [29]. However, the method presents some limitations, such as sensitivity to the number of lags chosen, which can lead to overparameterization or under parameterization.

Another limitation arises from the potential high dimensionality when the number of endogenous variables and lags is large, requiring a substantial amount of data [30]. Additionally, The VAR method assumes that the time series is stationary, meaning its statistical properties—such as mean, variance, and covariance—must remain constant over time [31]. In this study, STL decomposition is proposed to separate the trend and seasonality from the time series, with the expectation that the resulting residuals will be stationary.

The mathematical foundation of the VAR method is based on vectors and matrices. Vectors represent the values of the variables over time, while matrices contain the coefficients that describe variance and covariance among these variables across different time lags. This interaction is modeled through a linear equation, as presented in Equation 17 which captures the mutual influence among variables over time [31].

VAR Equation

$$Y_t = C + \sum_{i=1}^{p} \Phi_i \, Y_{t-1} + \epsilon_t \quad (17)$$

Where:
Y_t: vector of endogenous variables;
Φ_i: coefficient matrices;
C: constant vector;
ϵ_t: vector of error terms;
p: model order (number of lags).

Expanding this equation for all time periods with T observations, one can stack the equations corresponding to periods $t = p + 1$ to T into a matrix form, as shown in Equation 18. This reformulation allows all observations to be treated simultaneously, enabling a direct solution using the Ordinary Least Squares (OLS) method [32], [33].

Reduced VAR

$$Y = Z\Theta + E \quad (18)$$

Where:

Matrix of Dependent Variables (Y): Composed of transposed vectors of endogenous variables $Y_{p+1}^T, Y_{p+2}^T, \ldots, Y_T^T$, with dimensions $(T - p) \times N$;

Matrix of Constants and Lagged Variables (Z): Concatenates a vector of constants (C) and lagged endogenous variables, with dimensions $(T - p) \times (N \cdot p + 1)$;

Matrix of Coefficients (Θ): Composed of the transposed constant vector (C^T) and the coefficient matrices Φ_1 to Φ_p, with dimensions $(N \cdot p + 1) \times N$;

Matrix of Residuals (E): Organizes the transposed error terms ϵ^T from period $p + 1$ to T, with dimensions $(T - p) \times N$.

Equation 19 is derived by isolating the error matrix E from Equation 18 and substituting it into the minimization of the sum of squared residuals $S(\Theta) = E^T E$, according to the OLS method. Thus, $\widehat{\Theta}$ represents the theoretical coefficients that minimize $S(\Theta)$, corresponding to the values that minimize the squared error between observed and predicted values [32].

Least Squares VAR Equation

$$\widehat{\Theta} = (Z^T Z)^{-1} Z^T Y \tag{19}$$

By substituting $\widehat{\Theta}$ into Equation 18, one obtains the estimated residual matrix $\hat{E}$, which contains the differences between observed values and model-adjusted values for each of the $T - p$ periods. Then, Equation 20 can be applied to calculate the sample covariance matrix of the residuals, describing the variability and dependency structure among the residuals of the VAR model's endogenous variables [33].

Sample Covariance Matrix

$$\sum_{\epsilon} = \frac{\hat{E}^T \hat{E}}{T - p} \tag{20}$$

Where:
$\widehat{\Sigma_{\epsilon}}$: estimated covariance matrix of the residuals;
T: total number of time periods;
$\widehat{E^T}$: transposed matrix of residuals.

This process captures the correlation, magnitude, and direction of the model residuals, where each element $\hat{\Sigma}_{\epsilon}^{ij}$ represents the covariance between residuals i and j. In this context, both the residual matrix and the residual covariance matrix provide a deep understanding of the unexplained component of the model. To incorporate this unexplained behavior into simulations, the Cholesky decomposition is applied to the covariance matrix of the residuals, as shown in Equation 21. This factorizes a symmetric positive definite matrix into the product of a lower triangular matrix and its transpose [34].

Cholesky Equation

$$\Sigma_{\epsilon} = LL^t \tag{21}$$

II.5 FINANCIAL MATHEMATICAL MODELING

The most fundamental concept for developing a financial mathematical model lies in understanding the concepts of revenue, fixed costs, variable costs, investment expenses, debt payments, and dividend distributions. These can be detailed as follows:

Cash Inflows [35]:

- Sales Revenue: Refers to income generated from product sales, with the amount received varying directly with the quantity of goods and services sold;
- Financial Revenues: Arise from financial activities, such as interest from investments, dividends on equity, foreign exchange gains, among others;
- Other Operating Revenues: These are part of the company's cash flow but are not related to its core business activities. Examples include rental income, proceeds from by-products, compensation received, etc.

Cash Outflows [36]:

- Fixed Costs: Recurring expenses necessary for maintaining the company’s operations that do not vary with production volume. Examples include rent, insurance, and salaries;
- Variable Costs: Costs directly related to the production volume, such as raw materials, supply expenses, and sales commissions;
- Investment Expenses: Expenditures on tangible or intangible assets aimed at generating future financial returns, such as spending on new projects, equipment, or capacity expansion;
- Financial Expenses: Costs related to debt service, such as loan repayments and interest payments.

These concepts are structured in various ways to report business performance. However, the cash flow statement and the balance sheet are the most basic tools for financial evaluation. The following sections briefly explain these tools, highlighting their main characteristics, advantages, and limitations. The cash flow (CF) statement is a financial tool that records all of the company’s cash inflows and outflows. It is divided into three categories: operating cash flow (OCF), investing cash flow (ICF), and financing cash flow (FCF), as expressed in Equation 22 [37].

Cash Flow Equation

$$\mathrm{CF}_t = \mathrm{OCF}_t + \mathrm{ICF}_t + \mathrm{FCF}_t \tag{22}$$

The OCF includes all financial movements related to the company’s core activities. It consists of both operating revenues and expenses. Revenues mainly stem from product and service sales, including by-product sales, freight charges, and additional services. Expenses include payments to suppliers, wages and payroll charges, infrastructure costs, taxes on sales, and commercial expenses [37]. This is modeled by Equation 23. If this indicator remains negative for extended periods, it threatens the company’s viability.

Operating Cash Flow Equation

$$\text{OCF}_t = \text{Operating Cash Inflows}_t - \text{Operating Cash Outflows}_t \quad (23)$$

The ICF accounts for financial movements involving the acquisition or disposal of long-term assets, including financial investments. Inflows originate from asset sales, proceeds from financial applications, or divestments. Outflows cover the purchase of machinery, buildings, participation in other businesses, and infrastructure investments [37]. According to Equation 24, a positive ICF indicates disinvestment, while a negative value suggests active long-term investment.

Investing Cash Flow Equation

$$\text{ICF}_t = \text{Investment Inflows}_t - \text{Investment Outflows}_t \quad (24)$$

The FCF relates to the company's financing activities and capital structure. Inflows include borrowed capital, equity issuance, or capital contributions from shareholders. Outflows include debt repayments, interest payments, dividend distributions, and share buybacks [37].

Financing Cash Flow Equation

$$\text{FCF}_t = \text{Financing Inflows}_t - \text{Financing Outflows}_t \quad (25)$$

While OCF and ICF are directly interpretable, FCF requires more nuanced analysis. A positive FCF indicates the company is taking on debt, which may be strategic if occasional (e.g., leveraged investment). However, recurring positive FCF values may indicate over-reliance on debt to sustain operations. A negative FCF typically reflects debt repayment and shareholder returns, often interpreted as a positive indicator.

II.6 FINANCIAL ANALYSIS TOOLS

Cash flow and the balance sheet form the basis for various financial analyses. Below are tools used to assess investment risks and project viability.

II.6.1 Net Present Value (NPV)

NPV calculates the difference between the present value of expected cash flows and the initial investment, as defined in Equation 26 [35].

Net Present Value Equation

$$\text{NPV} = \sum_{t=0}^{n} \frac{\text{CF}_t}{(1+i)^t} - I_0 \quad (26)$$

Where:
t: Time period;
CF_t: Cash flow in period \$t\$;
i: Discount rate or minimum acceptable rate of return;
I_0: Initial investment.

NPV yields three main interpretations for investment decisions: a positive NPV indicates financial viability, where the project's net income exceeds its cost; a negative NPV implies economic inviability; and an NPV close to zero suggests the investment merely breaks even [38]. When applied in Monte Carlo simulations, NPV is calculated for each generated scenario, allowing for statistical analysis across scenarios (mean, standard deviation, quartiles) [39].

II.6.2 Probability of Positive Return (PPR)

PPR is a probabilistic economic viability metric that evaluates the likelihood of a project yielding positive returns over a specified time horizon. Using stochastically generated scenarios, each case is evaluated for profitability (e.g., whether NPV > 0). The proportion of positive-return scenarios among all simulations is calculated, as shown in Equation 27 [40].

Probability of Positive Return Equation

$$\text{PPR} = \frac{\text{Number of Scenarios with Positive Return}}{\text{Total Number of Simulated Scenarios}} \quad (27)$$

II.6.3 Discounted Payback Period

The Discounted Payback Period estimates the time required to recover an investment, accounting for a discount rate. It is computed by identifying the minimum T where the cumulative present value of cash flows exceeds the initial investment, as shown in Equation 28 [41].

Discounted Payback Equation

$$\text{Discounted Payback} = \min\left\{T : \sum_{t=1}^{T} \frac{\text{CF}_t}{(1+i)^t} \geq I_0\right\} \quad (28)$$

In Monte Carlo simulations, this metric is applied to each individual scenario. Results are then statistically analyzed to derive metrics such as average payback time, standard deviation, and quartiles [40].

III. MATERIALS AND METHODS

The proposed conceptual model, as illustrated in Figure 1, is structured into three stages, classified according to the nature of their activities.

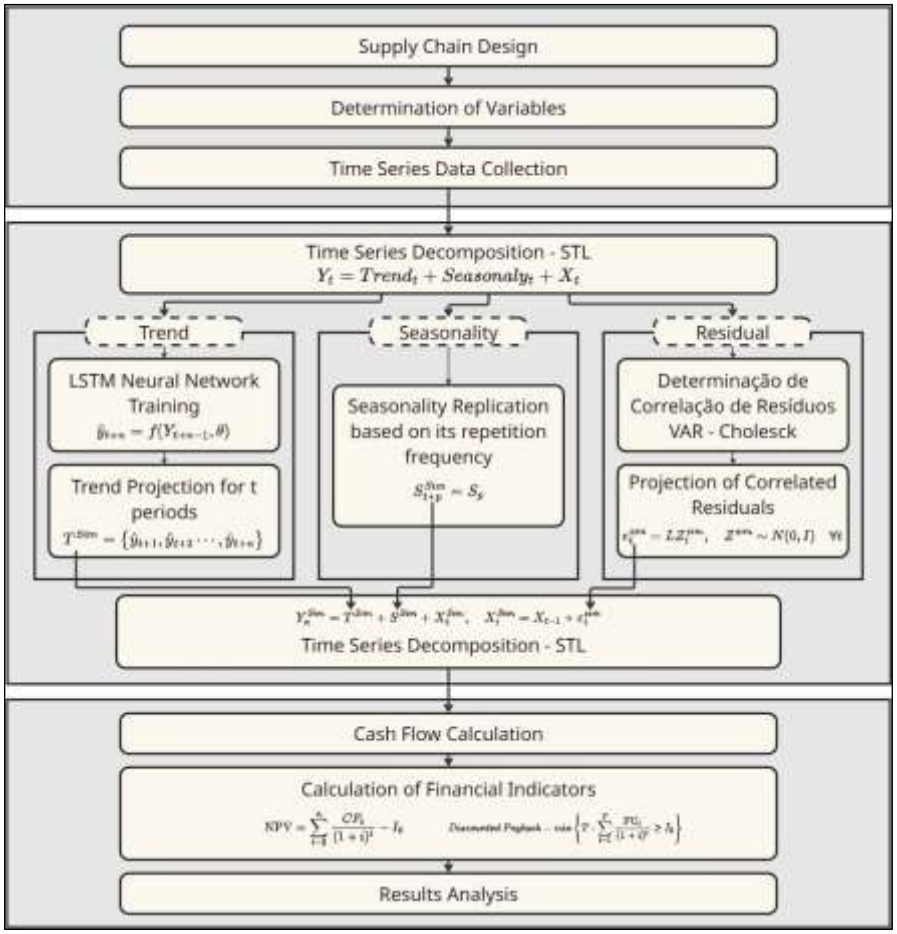


Figure 1: Conceptual Model.
Source: Authors, (2025).

III.1 PRODUCTION CHAIN

The first stage involves the design of the production chain to be analyzed. Although this work does not delve deeply into the theoretical basis of this step, a general overview is provided based on the evidence discussed in [6]. At least three different approaches can be used to structure this production chain:

III.1.1 Based on an existing supply chain

Analyzing existing supply chains is useful for business planning, as it enables incorporating validated strategies and empirical knowledge - a process commonly referred to as benchmarking. This approach is not limited to companies within the same industry as the target project but can also draw from firms in different sectors within the same region.

III.1.2 Based on expert opinion

Expertise should not be overlooked. Even with rigorous studies and sound project management methods, subtle nuances may be missed. Expert reviews help identify waste, challenges, design flaws, and other risks. Specialist consultation is also essential in adjacent areas, such as taxation, environmental regulations, and legal compliance.

III.1.3 Based on academic studies

Research-based development enables innovation in investment projects, whether related to product design or process improvement. Academic literature reviews and the synthesis of scholarly knowledge provide an innovative foundation that may help establish a competitive edge. Innovation is crucial for the success of new ventures that lack economies of scale or an established market. Without some form of product or process innovation, new companies often face severe market entry difficulties.

From this production chain layout, it is possible to identify the key project variables, including fixed costs, variable costs, revenues, and investments. Additional data may also be needed, such as historical inflation rates, currency exchange rates (e.g., USD), among others. Collecting these time series can be challenging and may require collaboration with companies or institutions, as access to such data is often restricted.

III.2 TIME SERIES FORECASTING

Once the production chain is defined and data collected, the model proceeds with the decomposition of time series using the STL method. As described for Reference 19 , STL decomposes each time series into trend, seasonal, and residual components in a unified manner. This joint decomposition is highly desirable. Applying different methods to extract each component separately introduces different parameters and biases, potentially resulting in overlap between trend and residual components. STL's unified decomposition improves model stability by avoiding such issues. Another advantage of STL is its ability to produce stationary residuals, which is crucial for analyzing residual correlation. Verifying stationarity is essential and should be done using tests such as the Dickey-Fuller Test, KPSS Test, and Phillips-Perron Test. If residuals are found to be non-stationary, STL parameters can be adjusted and reapplied, avoiding the need for additional transformations such as differencing, log transformations, or the Box-Cox transformation - which may be methodologically costly and risk information loss, ultimately reducing model robustness. After decomposition, the model separately forecasts each of the three components.

Trend forecasting is performed using Long Short-Term Memory (LSTM) neural networks, chosen for their capacity to capture both long- and short-term dependencies in financial time series [42]. Prior to training, the time series data undergoes preprocessing. A common step is normalization, which scales values to a 0–1 range, preventing large-scale numbers from disproportionately influencing training [43]. Another preprocessing step involves converting dates into sequential integers. For example, neural networks may struggle to interpret the transition from month 12 (December 2024) to month 01 (January 2025). Using sequential numeric encoding and reverting it during postprocessing improves interpretability. Additionally, a cyclical temporal component, such as a sine or cosine function, can be included to help the network detect seasonal patterns not fully captured by STL [44], [45].

During network configuration, key parameters must be defined:

- Batch size: number of samples processed simultaneously.
- Timesteps: length of time intervals processed as a unit.
- Features: number of correlated variables used as input.

LSTM architecture parameters include:

- Number of units: neurons per layer;
- Activation functions: typically tanh and sigmoid;
- Number of layers;
- Dropout rate: to prevent overfitting by randomly disabling neurons during training.]

Training parameters (also known as learning parameters) include the definition of training, validation, and test sets, as well as:

- Loss function: measures the difference between predicted and actual values;
- Optimizer: updates weights and biases based on the loss.

Let T be a trend vector composed of x values over n time periods, as in Equation 29:

Trend Original Vector

$$T = \{x_{t-n}, x_{t-n+1}, \cdots, x_{t-1}, x_t\} \tag{29}$$

This vector T is input into an LSTM network, which forecasts the trend's future trajectory by iteratively adjusting weights and biases via the internal operations of the LSTM cell. These are represented by Equations 11, 12, 14, and 15, involving forget gate (f_t), input gate (i_t), output gate (o_t), and cell state update (C_y).

Two common issues may arise during training:

- Overfitting: the model memorizes patterns specific to the training data but fails to generalize;
- Underfitting: the model fails to capture meaningful patterns.
- These are mitigated by fine-tuning the above parameters \cite{li2024keeping}.

The forecast for the next time step is given by Equation 30:

LSTM Forecast Equation

$$\hat{y}t + n = f(Yt + n - 1, \theta) \tag{30}$$

Where:

- $\hat{y}t + n$: Forecasted value for period $t + n$;
- $Yt + n - 1$: Vector of previous values, including both actual and predicted data;
- f: Function learned by the LSTM network;
- θ: Parameters (weights and biases) of the trained model.

After training is complete, the model forecasts the full trend sequence over n periods, as shown in Equation 31:

Trend Projection Vector

$$\mathrm{T}^{Sim} = \{\hat{y}t + 1, \hat{y}t + 2 \cdots, \hat{y}t + n\} \tag{31}$$

III.3 SEASONALITY FORECASTING

As previously mentioned in Section II.2, the STL decomposition method produces a well-defined seasonality pattern, which facilitates the extrapolation of this periodic behavior into future cycles. Delving deeper into the seasonality projection methodology, after the time series is decomposed, each seasonal component S_t is associated with a corresponding time point t within the cycle. The cycle length P defines the number of seasonal values in a full cycle for monthly series with yearly seasonality. The seasonal pattern is represented by a sequence of P values, $S = \{S_1, S_2, \dots, S_P\}$, where each value corresponds to a specific position within the cycle. For future time points $t > T$, where T is the last observed value of the series, seasonality is projected using Equation 32 [21].

Forecasted Seasonal Value Equation

$$S_{t+p} = S_p \tag{32}$$

Where:

S_{t+p}: forecasted seasonal value;
S_p: value from the seasonal cycle, with p ranging from 1 to P.

In cases where seasonality exhibits non-stationary behavior, regression-based models may be applied to dynamically adjust the amplitude of seasonal components over time [46].

III.4 RESIDUAL FORECASTING

The VAR-Cholesky method was selected for modeling the correlation between residuals due to its ability to capture interdependencies among multiple time series, including time lags [47]. This characteristic is essential for generating robust feasibility forecasts. To illustrate the importance of properly identifying correlation among series, consider a yogurt manufacturing plant. One can reasonably expect a correlation between the price of raw milk (a commodity) and the final price of yogurt. These prices may be influenced by external factors not fully captured in the model, the residuals. If residual correlations are ignored, the simulation may generate inconsistent results, where one variable fluctuates independently of the other, leading to unreliable viability projections.

Residual forecasting begins by applying the Monte Carlo method as described in Section II.4, resulting in the residual correlation matrix L. This matrix is multiplied by a standardized random variable vector (Z^{sim}), producing a simulated residual vector (ε_t^{sim}) for each period t. This process is repeated for T periods to cover the desired simulation horizon. The results for all simulations can be stored in a residual simulation matrix E^{sim} with dimensions $N \times T$ [48].

Residual Forecasting Equation

$$\varepsilon_t^{sim} = LZ_t^{sim}, \quad Z^{sim} \sim N(0, I) \quad \forall t \tag{33}$$

These results are equivalent to white noise (as in Equation 1) but now incorporating the correlation among time series residuals. By applying Equation 33 to the Brownian noise formulation in Equation 2, the behavior of the residual component in the simulations is fully defined. This process is repeated N times to generate the desired number of Monte Carlo scenarios [18], [49].

III.5 RECONSTRUCTION OF FORECASTED TIME SERIES

After decomposition using the STL method, reconstruction of the forecasted time series is a simple process involving the sum of the trend, seasonal, and residual components. This is represented by Equation 34, where n denotes the simulation index [19].

Forecasted Time Series Equation

$$Y^{Sim} = T^{Sim} + S^{Sim} + Xn^{Sim} \tag{34}$$

It is important to note that the trend and seasonal components can be treated as constant across all simulations. In other words, their values remain fixed, while the residuals are resampled for each simulation scenario. This separation is computationally efficient: although training the LSTM network and computing autoregressive vectors is resource-intensive, the subsequent steps involve only first-order arithmetic operations.

IV. RESULTS AND DISCUSSIONS

The evaluation of investment risk begins with the mathematical modeling of cash flow, as represented by the simplified Equation 35.

Cash Flow Equation

$$\mathrm{FC}_t = \mathrm{Revenues}_t - \mathrm{Fixed\ Costs}_t - \mathrm{Variable\ Costs}_t \tag{35}$$

By applying the forecasted time series to this cash flow equation, we obtain a cash flow projection for each period t and each of the n simulated scenarios. These projected cash flows are then analyzed using investment evaluation tools such as Net Present Value (NPV) and Discounted Payback Period [38], [41]. Since the results are based on a set of simulated scenarios, they can be subjected to statistical analysis.

For example, calculating the lower quartile of NPV yields the Value at Risk (VaR), representing the greatest expected loss at a given confidence level (typically 5% or 1%). Additionally, by counting the number of scenarios in which the NPV is greater than zero, one obtains the probability of positive returns. From the Discounted Payback results, it is possible to estimate the average time to investment recovery, as well as determine the expected return periods in worst-case scenarios [40].

V. CONCLUSIONS

This study introduced a conceptual method for risk assessment in manufacturing investments, contributing with a novel approach that integrates STL, LSTM, and VAR-Cholesky methods to forecast correlated time series of costs and revenues. The model advances investment analysis techniques from both an economic science and a production engineering perspective. While theoretically robust, the model's practical utility must be confirmed through empirical validation in real-world case studies. The primary limitations of the current framework are the unstructured nature of the production chain modeling and the absence of an optimization module. Future work should therefore focus on two key areas: first, developing a more rigorous and detailed modeling process for a generic production chain; and second, incorporating an optimization layer to identify ideal investment strategies. Successfully addressing these points will enhance the model's applicability, potentially establishing it as a valuable tool for feasibility analysis and risk forecasting in the manufacturing sector.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Rafael Vieira da Silva, Enzo Morosini Frazzon.
Methodology: Rafael Vieira da Silva, Enzo Morosini Frazzon.
Investigation: Rafael Vieira da Silva, Enzo Morosini Frazzon.
Discussion of results: Rafael Vieira da Silva, Enzo Morosini Frazzon
Writing – Original Draft: Rafael Vieira da Silva, Enzo Morosini Frazzon
Writing – Review and Editing: Rafael Vieira da Silva, Enzo Morosini Frazzon
Resources: Rafael Vieira da Silva, Enzo Morosini Frazzon
Supervision: Rafael Vieira da Silva, Enzo Morosini Frazzon
Approval of the final text: Rafael Vieira da Silva, Enzo Morosini Frazzon

VII. ACKNOWLEDGMENTS

This work was carried out with the support of the Coordination for the Improvement of Higher Education Personnel – Brazil (CAPES) – Financing Code 88887.877819/2023-00

VIII. REFERENCES

[1] H. Wang and Z. Zhou, "Identification of key risk nodes and invulnerability analysis of construction supply chain networks," Buildings, vol. 14, no. 7, p. 1997, 2024.

[2] S. DuHadway, S. Carnovale, and B. Hazen, "Understanding risk management for intentional supply chain disruptions: Risk detection, risk mitigation, and risk recovery," Annals of Operations Research, vol. 283, pp. 179–198, 2019.

[3] T. S. Oliveira, C. V. B. Júnior, and M. Caetano, "Return on investing in innovative activities: the brazilian manufacturing industry," International Journal of Innovation, vol. 11, no. 1, pp. e22797–e22797, 2023.

[4] I. Zykiene, A. Leskauskien, I. Miciulien, and R. Dauglien, "Driving Growth and Innovation: Exploring Foreign Direct Investment in the Manufacturing Sector (The Case of Lithuania)," European Integration Studies, no. 17, pp. 124–140, 2023.

[5] N. F. Silva, M. dos Santos, C. F. S. Gomes, and L. P. de Andrade, "An integrated CRITIC and Grey Relational Analysis approach for investment portfolio selection," Decision Analytics Journal, p. 100285, 2023.

[6] R. V. a Silva and E. M. Frazzon, "ESTADO DA ARTE DOS MÉTODOS PARA ANÁLISE DE VIABILIDADE ECONÔMICA NA INDÚSTRIA DE TRANSFORMAÇÃO," Gestão Da Produção, Operações E Sistemas, 2025.

[7] J. Zhang, "Modern Monte Carlo methods for efficient uncertainty quantification and propagation: A survey," Wiley Interdisciplinary Reviews: Computational Statistics, vol. 13, no. 5, p. e1539, 2021.

[8] B. H. Gebreslassie, Y. Yao, and F. You, "Design under uncertainty of hydrocarbon biorefinery supply chains: multiobjective stochastic programming models, decomposition algorithm, and a comparison between CVaR and downside risk," AIChE Journal, vol. 58, no. 7, pp. 2155–2179, 2012.

[9] S. L. Ngan, B. S. How, S. Y. Teng, W. D. Leong, A. C. M. Loy, P. Yatim, M. A. B. Promentilla, and H. L. Lam, "A hybrid approach to prioritize risk mitigation strategies for biomass polygeneration systems," Renewable and Sustainable Energy Reviews, vol. 121, p. 109679, 2020.

[10] T. Ingelaere, V. Maes, and G. Samaey, "Cell Escape Probabilities for Markov Processes on a Grid," arXiv preprint arXiv:2404.00314, 2024.

[11] W. Li, M. Lu, and P. Ye, "Optimal restricted Monte Carlo integration on generalized Hölder classes," Journal of Applied Mathematics and Computing, pp. 1–26, 2025.

[12] J. Yoder, C. Alexander, R. Ivanic, S. Rosch, W. Tyner, and S. Wu, "Risk Versus Reward, a Financial Analysis of Alternative Contract Specifications for the Miscanthus Lignocellulosic Supply Chain," BioEnergy Research, vol. 8, 2015.

[13] I.-W. G. Kwon and T. Suh, "Trust, commitment and relationships in supply chain management: a path analysis," Supply chain management: an international journal, vol. 10, no. 1, pp. 26–33, 2005.

[14] W. Tian, Y. Heo, P. de Wilde, Z. Li, D. Yan, C. S. Park, X. Feng, and G. Augenbroe, "A review of uncertainty analysis in building energy assessment," Renewable and Sustainable Energy Reviews, vol. 93, pp. 285–301, 2018.

[15] V. Platon and A. Constantinescu, "Monte Carlo Method in Risk Analysis for Investment Projects," Procedia Economics and Finance, vol. 15, pp. 393–400, 2014.

[16] P. Capuozzo, E. Panella, T. S. Gherardini, and D. D. Vvedensky, "Path integral Monte Carlo method for option pricing," Physica A: Statistical Mechanics and its Applications, vol. 581, p. 126231, 2021.

[17] M. West, “Time series decomposition,” Biometrika, vol. 84, no. 2, pp. 489–494, 1997.

[18] L. Heinäkoski, “Limitations of normal distribution in economic risk analysis,” L. Heinäkoski, 2018.

[19] T. Krake, D. Klötzl, D. Hägele, and D. Weiskopf, “Uncertainty-Aware Seasonal-Trend Decomposition Based on Loess,” IEEE Transactions on Visualization and Computer Graphics, 2024.

[20] F. Leite Coelho da Silva, J. da Silva Cordeiro, K. da Costa, N. Saboya, P. Canas Rodrigues, and J. L. López-Gonzales, “Time series forecasting via integrating a filtering method: an application to electricity consumption,” Computational Statistics, pp. 1–20, 2025.

[21] T. Phungtua-eng and Y. Yamamoto, “Adaptive Seasonal-Trend Decomposition for Streaming Time Series Data with Transitions and Fluctuations in Seasonality,” in Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 2024, pp. 426–443.

[22] K. Zhang, X. Huo, and K. Shao, “Temperature Time Series Prediction Model Based on Time Series Decomposition and Bi-LSTM Network,” Mathematics, vol. 11, no. 9, 2023.

[23] E. S. Gardner Jr and E. McKenzie, “Forecasting trends in time series,” Management science, vol. 31, no. 10, pp. 1237–1246, 1985.

[24] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” nature, vol. 323, no. 6088, pp. 533–536, 1986.

[25] C. L. Giles, D. Chen, G.-Z. Sun, H.-H. Chen, Y.-C. Lee, and M. W. Goudreau, “Constructive learning of recurrent neural networks: limitations of recurrent cascade correlation and a simple solution,” IEEE Transactions on Neural Networks, vol. 6, no. 4, pp. 829–836, 1995.

[26] S. Hochreiter and J. Schmidhuber, “LSTM can solve hard long time lag problems,” Advances in neural information processing systems, vol. 9, 1996.

[27] J. Schmidhuber, S. Hochreiter, and others, “Long short-term memory,” Neural Comput, vol. 9, no. 8, pp. 1735–1780, 1997.

[28] Y. Yu, X. Si, C. Hu, and J. Zhang, “A review of recurrent neural networks: LSTM cells and network architectures,” Neural computation, vol. 31, no. 7, pp. 1235–1270, 2019.

[29] C. A. Sims, “Macroeconomics and reality,” Econometrica: journal of the Econometric Society, pp. 1–48, 1980.

[30] J. H. Stock and M. W. Watson, “Vector autoregressions,” Journal of Economic perspectives, vol. 15, no. 4, pp. 101–115, 2001.

[31] F. Canova, “Vector autoregressive models: specification, estimation, inference, and forecasting,” Handbook of applied econometrics volume 1: Macroeconomics, pp. 53–110, 1999.

[32] K. Kavaklioglu, “Principal components based robust vector autoregression prediction of Turkey’s electricity consumption,” Energy Systems, vol. 10, no. 4, pp. 889–910, 2019.

[33] P. Kohli, T. P. Garcia, and M. Pourahmadi, “Modeling the Cholesky factors of covariance matrices of multivariate longitudinal data,” Journal of Multivariate Analysis, vol. 145, pp. 87–100, 2016.

[34] H. Lütkepohl, “Vector autoregressions,” Companion to Theoretical Econometrics, Blackwell, Oxford, pp. 678–699, 2001.

[35] G. Lucko, “Supporting financial decision-making based on time value of money with singularity functions in cash flow models,” Construction Management and Economics, vol. 31, no. 3, pp. 238–253, 2013.

[36] S. Briciu, “Variable and fixed costs in company management,” Annales Universitatis Apulensis Series Oeconomica, vol. 1, no. 10, pp. 1–14, 2008.

[37] K. Simons, “The relationship between dividend changes and cash flow: An empirical analysis,” Journal of Business Finance & Accounting, vol. 21, no. 4, pp. 577–587, 1994.

[38] T. Shou, “A literature review on the net present value (NPV) valuation method,” in 2022 2nd International Conference on Enterprise Management and Economic Development (ICEMED 2022), 2022, pp. 826–830.

[39] S. S. Shaffie and S. H. Jaaman, “Monte Carlo on net present value for capital investment in Malaysia,” Procedia-Social and Behavioral Sciences, vol. 219, pp. 688–693, 2016.

[40] I. M. Ronyastra, L. H. Saw, and F. S. Low, “Monte Carlo simulation-based financial risk identification for industrial estate as post-mining land usage in Indonesia,” Resources Policy, vol. 89, p. 104639, 2024.

[41] A. Kijo-Kleczkowska, P. Brus, and G. Wikeciorkowski, “Profitability analysis of a photovoltaic installation - A case study,” Energy, vol. 261, p. 125310, 2022.

[42] A. Sunki, C. SatyaKumar, G. S. Narayana, V. Koppera, and M. Hakeem, “Time series forecasting of stock market using ARIMA, LSTM and FB prophet,” in MATEC Web of Conferences, 2024, vol. 392, p. 01163.

[43] H. Agarwal, G. Mahajan, A. Shrotriya, and D. Shekhawat, “Predictive Data Analysis: Leveraging RNN and LSTM Techniques for Time Series Dataset,” Procedia Computer Science, vol. 235, pp. 979–989, 2024.

[44] C. Arthur, N. Yudistira, and C. Dewi, “AutoCyclic: Deep Learning Optimizer for Time Series Data Prediction,” IEEE Access, 2024.

[45] F. Zhang, M. Wang, W. Zhang, and H. Wang, “THATSN: Temporal hierarchical aggregation tree structure network for long-term time-series forecasting,” Information Sciences, vol. 692, p. 121659, 2025.

[46] M. K. P. So and R. S. W. Chung, “Dynamic seasonality in time series,” Computational Statistics & Data Analysis, vol. 70, pp. 212–226, 2014.

[47] M. Demetrescu and N. Salish, “(Structural) VAR models with ignored changes in mean and volatility,” International Journal of Forecasting, vol. 40, no. 2, pp. 840–854, 2024.

[48] J.-S. Chou, “Cost simulation in an item-based project involving construction engineering and management,” International Journal of Project Management, vol. 29, no. 6, pp. 706–717, 2011.

[49] D. J. Higham and M. Roj, “Computing mean first exit times for stochastic processes using multi-level Monte Carlo,” in Proceedings of the 2012 Winter Simulation Conference (WSC), 2012, pp. 1–10.

Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11 n.55, p. 140-142. September/October., 2025.

DOI: https://doi.org/10.5935/jetia. v11i55.2615

RESEARCH ARTICLE OPEN ACCESS

CHALLENGES IN EXECUTION CONTROL AND STABILIZATION OF OFFSHORE BORED PILES: LESSONS FROM REAL CASES

Eliomar Gotardi Pessoa[1],

[1] UFRJ- Universidade Federal do Rio de Janeiro-RJ, Brazil.

[1]http://orcid.org/0009-0008-7713-4740

Email: eliomar.pessoa2coc.ufrj.br

ARTICLE INFO

Article History

Received: August 29, 2025

Revised: September 1, 2025

Accepted: September 15, 2025

Published: September 30, 2025

Keywords:

Offshore bored piles; execution control; stabilization fluids; drilling retention; marine wind energy; quality management; sustainability.

ABSTRACT

Offshore bored piles are widely employed in wind energy, oil, and gas platforms due to their adaptability to varying seabed conditions and high load capacity. However, execution challenges such as drilling retention, base cleaning, and control of stabilizing fluids present critical risks to structural integrity and project sustainability. This paper analyzes barriers encountered in offshore bored pile construction, drawing lessons from real case studies in marine wind energy projects. Advanced monitoring techniques, fluid management protocols, and enhanced quality control procedures are discussed to mitigate failures and improve long-term performance. The study highlights that systematic execution protocols, combined with continuous monitoring and adaptive response strategies, significantly reduce risks and support sustainable offshore construction practices. Additionally, the study examines the environmental impact of fluid discharge and the role of sustainable construction practices in reducing marine pollution. By integrating lessons from operational projects, it provides a comprehensive framework for improving safety, efficiency, and sustainability in offshore foundation construction.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Offshore energy projects, particularly wind farms, require highly reliable foundations to ensure structural stability and operational safety. Bored piles are frequently used due to their capacity to bear axial and lateral loads, and their adaptability to heterogeneous seabed conditions. Despite technological advances, several operational challenges persist, especially in maintaining drilling retention, cleaning the base of the pile, and controlling stabilizing fluids such as bentonite or polymer suspensions (Randolph & Gourvenec, 2011). Failures in any of these execution steps can compromise pile integrity, increase settlement, and result in costly remediation. Lessons from real offshore projects indicate that lack of standardized protocols and insufficient monitoring contributes significantly to construction defects (Sumer & Fredsøe, 2002).

In addition to structural risks, operational inefficiencies can increase project costs and delay timelines, affecting the economic feasibility of offshore energy projects. Effective management of drilling fluids, base cleaning, and stabilization processes is therefore essential not only for safety but also for sustainable project delivery. This study investigates practical challenges in offshore bored pile execution, aiming to propose enhanced protocols and monitoring strategies to minimize failures, optimize resource use, and promote sustainability in marine wind energy projects. Emphasis is placed on lessons drawn from real cases to guide practical and implementable improvements.

II. THEORETICAL REFERENCE

The stability of offshore bored piles depends on maintaining lateral support during drilling and ensuring sufficient removal of cuttings to achieve a clean base for concrete placement. Drilling retention fluids are critical in preventing collapse of boreholes and maintaining sidewall integrity. The rheological properties and density of stabilizing fluids directly affect soil-pile interaction and end-bearing capacity (Das, 2016). Incomplete base cleaning or poor fluid management can generate voids, inclusions, or soft layers at the pile

base, significantly reducing structural capacity. Research indicates that careful monitoring of fluid properties, sedimentation, and circulation patterns can mitigate these risks (Houlsby et al., 2005).

Fluid properties also influence environmental performance, as bentonite and polymer suspensions can affect water quality if not managed correctly. Adopting environmentally friendly fluid formulations and reuse strategies reduces ecological impact while maintaining drilling stability (Randolph & Gourvenec, 2011). Lessons from offshore projects indicate that integrating monitoring technology and adaptive execution protocols into theoretical frameworks significantly improves quality control and reliability. Digital monitoring and predictive models provide engineers with near real-time feedback to adjust fluid properties, optimize base cleaning, and prevent defects[1].

III. MATERIALS AND METHODS

This study adopts a mixed-method approach combining case study analysis, field monitoring data, and literature review. Offshore bored piles from several marine wind energy projects were examined to identify recurring execution challenges. Key parameters included drilling retention fluid properties, base cleaning efficiency, and pile concreting quality.

Field data were collected using:

- Sensor-based monitoring of fluid density and viscosity.
- Borehole inspection logs for base cleanliness and sedimentation evaluation.
- Post-installation pile integrity testing (e.g., low-strain dynamic testing and high-strain load tests).

Protocols and corrective measures observed in successful projects were catalogued and compared against sites with known execution defects. Quantitative data were analyzed statistically to evaluate the impact of fluid control, base cleaning, and monitoring on pile performance. Additionally, the study employed numerical simulations to predict pile behavior under different drilling and fluid management scenarios. Sensitivity analysis was conducted to assess the influence of fluid viscosity, density variations, and base cleaning efficiency on pile stability and end-bearing capacity.

IV. RESULTS AND DISCUSSIONS

Analysis of case studies revealed that improper control of drilling fluids, insufficient base cleaning, and irregular monitoring significantly contributed to performance deficiencies. Specifically, variability in fluid density led to sidewall collapse or excessive borehole deformation, while incomplete base cleaning resulted in voids that reduced axial load capacity.

Monitoring of fluid rheology and real-time adjustment of circulation parameters proved effective in maintaining borehole stability. Moreover, standardized protocols for base cleaning—combining high-velocity circulation and sediment traps—reduced inclusions and ensured uniform concrete placement.

Post-installation testing indicated that piles constructed with strict adherence to monitoring and execution protocols demonstrated lower settlement, higher stiffness, and improved end-bearing performance. Lessons learned indicate that integrating digital monitoring with adaptive execution strategies enhances reliability, reduces environmental impact, and supports sustainable offshore construction.

Furthermore, the results highlighted that proper fluid management also reduces material waste and mitigates environmental hazards, improving overall sustainability. Comparative analysis of cases demonstrated that continuous monitoring allows early detection of potential defects, facilitating proactive intervention and minimizing downtime.

V. CONCLUSIONS

The execution and stabilization of offshore bored piles present significant challenges, particularly regarding drilling retention, base cleaning, and fluid control. Case studies demonstrate that deficiencies in these areas can compromise pile performance and project sustainability.

Implementing standardized protocols, continuous monitoring, and adaptive fluid management significantly mitigates risks. Incorporating lessons from real projects enables more reliable and sustainable offshore foundation construction, particularly for marine wind energy.

Additionally, integrating environmental considerations into execution protocols enhances project sustainability by reducing fluid waste and preventing marine contamination. Future developments should focus on digital twin integration, IoT-based monitoring, and predictive analytics to further optimize execution, reduce risks, and support long-term performance.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Eliomar Gotardi Pessoa.
Methodology: Eliomar Gotardi Pessoa.
Investigation: Eliomar Gotardi Pessoa.
Discussion of results: Eliomar Gotardi Pessoa.
Writing – Original Draft: Eliomar Gotardi Pessoa.
Writing – Review and Editing: Eliomar Gotardi Pessoa.
Resources: Eliomar Gotardi Pessoa.
Supervision: Eliomar Gotardi Pessoa.
Approval of the final text: Eliomar Gotardi Pessoa.

VII. REFERENCES

[1] Das, B. M. (2016). *Principles of geotechnical engineering* (9th ed.). Boston, MA: Cengage Learning.Houlsby, G., Wroth, C., & Randolph, M. F. (2005). Drilling fluid effects on bored pile performance. *Proceedings of the Institution of Civil Engineers – Geotechnical Engineering, 158*(1), 15–26.Randolph, M., & Gourvenec, S. (2011). *Offshore geotechnical engineering*. Boca Raton, FL: CRC Press.Sumer, B. M., & Fredsøe, J. (2002). *The mechanics of scour in the marine environment*. Singapore: World Scientific.
Zhang, L., Wang, Z., & Wang, Y. (2015). Best practices in offshore wind pile execution: Lessons from field case studies. *Ocean Engineering, 104*, 150–160.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 143-144. September/October., 2025.
DOI: https://doi.org/10.5935/jetia. v11i55.2616

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

INNOVATIONS IN GROUTING CONTROL AND SOIL–PILE INTERACTION IN OFFSHORE BORED PILES: MODELING AND FIELD TESTS

Eliomar Gotardi Pessoa[1],

[1] UFRJ- Universidade Federal do Rio de Janeiro-RJ, Brazil.

[1]http://orcid.org/0009-0008-7713-4740

Email: eliomar.pessoa2coc.ufrj.br

ARTICLE INFO

Article History
Received: August 29, 2025
Revised: September 1, 2025
Accepted: September 15, 2025
Published: September 30, 2025

Keywords:
Offshore bored piles; grouting techniques; soil–pile interaction; pressure control; volume monitoring; finite element modeling; field testing.

ABSTRACT

Advanced grouting techniques are essential to enhance the interaction between soil and bored piles in offshore environments, ensuring structural integrity and optimizing load transfer. Precise control of grouting parameters, including pressure, volume, and flow rate, directly affects pile stiffness, settlement behavior, and long-term performance. This paper investigates innovative grouting strategies for offshore bored piles, integrating numerical finite element modeling with field data from operational projects. The study evaluates the effectiveness of pressure-controlled and volume-monitored grouting in improving soil–pile interaction and mitigating defects such as voids and soft zones. Results demonstrate that advanced grouting, combined with real-time monitoring, enhances load-bearing capacity, reduces uncertainties, and contributes to sustainable offshore construction practices. Insights from field tests highlight the importance of adaptive protocols and continuous verification to optimize grouting outcomes.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Offshore bored piles are critical foundations for wind turbines, oil platforms, and other marine infrastructure due to their adaptability and high axial and lateral load capacity. However, ensuring optimal soil–pile interaction requires careful execution of grouting procedures to fill voids, consolidate weak zones, and improve end-bearing capacity (Randolph & Gourvenec, 2011). Traditional grouting methods often rely on nominal volumes and empirical pressure settings, which can lead to incomplete filling, formation of voids, and variability in pile stiffness. Advances in monitoring and modeling enable more precise grouting control, improving construction quality and long-term reliability (Jeong et al., 2017).

In addition to enhancing structural performance, efficient grouting practices can reduce material waste, environmental impact, and construction downtime. Offshore projects, with high logistical and operational costs, particularly benefit from protocols that optimize grout usage and minimize uncertainties. This study examines innovative grouting control techniques, integrating numerical modeling and field measurements to assess their effectiveness in enhancing soil–pile interaction. Lessons from operational offshore cases are used to propose best-practice protocols and adaptive monitoring strategies.

II. THEORETICAL REFERENCE

Grouting improves soil–pile interaction by increasing lateral friction and tip resistance, mitigating settlement, and enhancing pile stiffness. Parameters such as grout pressure, volume, and rheology directly influence the consolidation of surrounding soils and the quality of pile–soil contact (Das, 2016). Finite element modeling of grouted piles provides insight into stress distribution, load transfer mechanisms, and potential zones of weakness. Coupled with field observations, modeling supports the calibration of grouting parameters and the prediction of performance under operational loads (Potts & Zdravkovic, 2001).

Controlled grouting also reduces the risk of defects, including voids at the pile base or along the shaft, which can compromise long-term stability. Real-time monitoring of pressure, flow rate, and displacement during grouting allows adaptive adjustment and verification of effective soil consolidation (Zhou et al., 2020).

From a theoretical standpoint, integrating probabilistic models with deterministic simulations allows engineers to quantify uncertainty, optimize grouting procedures, and predict variability in load-bearing capacity. This approach aligns with performance-based design principles in offshore geotechnical engineering.

III. MATERIALS AND METHODS

The study combines finite element simulations and field experiments to assess grouting control strategies. Offshore bored piles instrumented with pressure transducers, flow meters, and strain sensors were monitored during grouting operations. Key parameters included grout volume, injection pressure, and penetration efficiency. Finite element models simulated soil–pile interaction under varying grouting conditions, accounting for soil heterogeneity, nonlinear soil behavior, and cyclic load effects. Parametric studies were conducted to evaluate the sensitivity of pile performance to grout pressure, volume, and timing.

Field testing included verification of pile integrity using low-strain dynamic tests, fiber-optic strain monitoring along the pile shaft, and post-grouting displacement measurements. Data were compared with model predictions to validate numerical approaches and refine grouting protocols. Additionally, monitoring included IoT-based data collection and automated alerts for pressure or flow deviations. These measures ensured adaptive intervention, preventing under- or over-grouting and improving efficiency while minimizing operational risks.

IV. RESULTS AND DISCUSSIONS

Finite element simulations demonstrated that pressure-controlled grouting significantly improves stress distribution along the pile shaft and increases end-bearing resistance. High-volume injections without pressure control can create uneven consolidation or uplift of soft soil zones, highlighting the need for real-time adjustments.

Field measurements confirmed the simulation outcomes, showing improved stiffness and reduced settlement in piles grouted with pressure and volume control. Strain data revealed uniform load transfer and minimal void formation, supporting the effectiveness of adaptive grouting protocols (Jeong et al., 2017).

The study also indicated that integrating real-time monitoring reduces variability in pile performance and supports predictive maintenance. Deviations from target pressure or flow rates were promptly corrected, preventing defects that could compromise long-term structural integrity.

Furthermore, sustainable grouting practices, including reuse of excess grout and controlled discharge, were observed to reduce environmental impact. Adaptive grouting not only improves pile performance but also enhances the overall sustainability and cost-efficiency of offshore construction projects.

V. CONCLUSIONS

Innovative grouting techniques, emphasizing control of pressure and volume, significantly enhance soil–pile interaction and overall performance of offshore bored piles. Combining finite element modeling with field monitoring provides a robust framework for optimizing grouting procedures and minimizing construction defects.

Adaptive protocols supported by real-time data collection improve reliability, reduce uncertainties, and ensure sustainable construction practices. Lessons from field cases demonstrate the importance of integrating modeling, monitoring, and execution control to achieve predictable and high-performance foundations.

Future research should focus on the application of AI-driven predictive control, digital twins, and enhanced sensor networks to further optimize offshore grouting operations and long-term pile performance.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Eliomar Gotardi Pessoa..
Methodology: Eliomar Gotardi Pessoa.
Investigation: Eliomar Gotardi Pessoa.
Discussion of results: Eliomar Gotardi Pessoa.
Writing – Original Draft: Eliomar Gotardi Pessoa.
Writing – Review and Editing: Eliomar Gotardi Pessoa.
Resources: Eliomar Gotardi Pessoa.
Supervision: Eliomar Gotardi Pessoa.
Approval of the final text: Eliomar Gotardi Pessoa.

VII. REFERENCES

Das, B. M. (2016). *Principles of geotechnical engineering* (9th ed.). Boston, MA: Cengage Learning.
Jeong, S., Lee, S., & Lee, J. (2017). Performance of post-grouted drilled shafts under axial loading. *Soils and Foundations, 57*(1), 56–69.
Potts, D. M., & Zdravkovic, L. (2001). *Finite element analysis in geotechnical engineering: Application*. London: Thomas Telford.
Randolph, M., & Gourvenec, S. (2011). *Offshore geotechnical engineering*. Boca Raton, FL: CRC Press.
Zhou, H., Zhao, Y., & Zhang, L. (2020). Reliability analysis of post-grouted bored piles. *Journal of Geotechnical and Geoenvironmental Engineering, 146*(5), 04020020..

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 140-153. September/October., 2025.
DOI: https://doi.org/10.5935/jetia. v11i55.2743

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

DEVELOPMENT OF A VEHICLE SENSOR FAULT DIAGNOSIS DEVICE USING ARTIFICIAL INTELLIGENCE TECHNIQUES TO IMPROVE DIAGNOSTIC ACCURACY

Messaoud Linani [1,3], Boubakeur Latreche [1,2], Kamal Khalil Toaba [3], Ali Ben Kholif [3] Imededdin Djadour [1,3] and Aboubakr Missaoui [3]

[1] Laboratory of Computer Science & Applied Artificial Intelligence, University of Djelfa PO Box 3117, Djelfa 17000, Algeria.
[2] Laboratory of Telecommunication and Smart Systems (LTSS), Faculty of Sciences and Technology, University of Djelfa PO Box 3117, Djelfa 17000, Algeria.
[3] Faculty of Sciences and Technology, University of Djelfa PO Box 3117, Djelfa 17000, Algeria.

[1]https://orcid.org/0009-0001-6022-6648, [2]https://orcid.org/0009-0007-9367-3368, [3]https://orcid.org/0009-0003-6252-6825
[4]https://orcid.org/0009-0000-9434-7328, [5]https://orcid.org/0009-0009-1258-9629, [6]https://orcid.org/0009-0008-9478-0015

Email: messaoud.linani@univ-djelfa.dz , b.latreche@univ-djelfa.dz , phdstudent002@gmail.com , mastersstudent001@gmail.com , imededdin.djadour@univ-djelfa.dz , aboubakr.missaoui@univ-djelfa.dz .

ARTICLE INFO

Article History
Received: September 01, 2025
Revised: September 20, 2025
Accepted: September 25, 2025
Published: September 30, 2025

Keywords:
Sensor faults,
AI diagnostics,
ANN,
Embedded systems,
Vehicle maintenance.

ABSTRACT

The rapid evolution of automotive technology has heightened the reliance on electronic sensors for real-time monitoring and diagnostics. However, accurately identifying sensor faults remains a significant challenge due to the increasing complexity of modern vehicles. This paper presents the design and development of an intelligent diagnostic device that employs artificial intelligence (AI) techniques specifically Artificial Neural Networks (ANNs) to enhance the precision of fault detection. The system integrates signal acquisition, preprocessing, and classification modules within a low-cost embedded platform based on Arduino. It processes data from various simulated automotive sensors, distinguishing between normal and faulty conditions with high accuracy. Experimental evaluations demonstrate the device's superior diagnostic performance compared to conventional tools. In addition, an economic feasibility analysis confirms the device's potential for widespread adoption, particularly in resource-constrained markets. The proposed solution offers a cost-effective, accurate, and scalable approach to vehicle sensor diagnostics, promising improvements in maintenance efficiency, cost reduction, and overall road safety.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The automotive industry has witnessed rapid technological advancement in recent years, driven by the integration of electronic components and intelligent systems [1]. One of the most significant developments has been the extensive deployment of electronic sensors throughout modern vehicles [2]. These sensors are critical for monitoring key parameters such as engine temperature, air-fuel ratio, oxygen levels, speed, pressure, and emissions [3]. They enable real-time decision-making by electronic control units (ECUs), enhancing vehicle performance, safety, efficiency, and environmental compliance [4].

However, as the number and complexity of these sensors increase, so do the challenges associated with accurately diagnosing their faults [5]. Sensor failures can lead to inaccurate readings, performance degradation, and potentially dangerous operating conditions [6]. Traditional fault diagnosis approaches -such as manual inspections, rule-based systems, and standard On-Board Diagnostics (OBD-II) tools- are often limited in their ability to detect subtle or intermittent issues [7]. These conventional tools generally rely on pre-programmed fault codes and threshold values, which may not capture the full range of fault conditions encountered in dynamic real-world environments [8].

To overcome these limitations, the application of Artificial Intelligence (AI) in automotive diagnostics has emerged as a powerful and promising solution [9], [10], [11]. AI techniques, particularly Artificial Neural Networks (ANNs), have demonstrated a strong capacity for learning from data and identifying complex patterns in sensor behavior [12]. ANNs are well-suited for classifying signal anomalies, even in cases where the distinctions between normal and faulty signals are subtle or nonlinear [13]. This ability to learn and generalize from examples makes them ideal for applications requiring adaptive and intelligent fault detection.

This paper presents the development of a low-cost, intelligent diagnostic device designed to detect faults in vehicle sensors using ANN-based classification. The proposed system is built on an embedded platform utilizing the Arduino Due microcontroller and integrates essential components such as signal acquisition, preprocessing, and real-time classification. To provide immediate and intuitive user feedback, a 16x2 I2C LCD display is included for fault status visualization. The system is designed to be compact, user-friendly, and easily deployable in both professional repair shops and personal automotive maintenance settings.

In the development process, sensor-like inputs were simulated using variable resistors (potentiometers) and Pulse Width Modulation (PWM) signals. These were selected to represent typical analog and digital signals found in vehicle sensor systems. Various signal fault conditions were artificially introduced, including noisy, constant, intermittent, and erratic signals. These signal variations were then used to generate training and testing datasets for the ANN model. The ANN was trained using supervised learning techniques to distinguish between normal and faulty states, enabling the system to classify new sensor inputs with a high degree of accuracy.

Once trained and validated, the ANN model was embedded within the Arduino platform to enable real-time, standalone operation. During testing, the system demonstrated excellent fault detection performance, achieving over 95% classification accuracy across diverse simulated signal conditions. These results confirm the feasibility and effectiveness of the proposed approach, particularly in its ability to identify faults that are difficult to detect using traditional diagnostic tools.

The proposed diagnostic system offers several practical advantages. First, it significantly enhances diagnostic precision by leveraging AI-driven analysis rather than relying solely on fixed thresholds or human interpretation. Second, it provides a portable and cost-effective solution that can be implemented without the need for expensive equipment or advanced technical training. Third, the modular design of the device allows for potential scalability, enabling future upgrades to support additional sensors or advanced connectivity features such as wireless communication or cloud-based diagnostics.

The main contributions of this work can be summarized as follows:

- Development of a real-time, AI-powered diagnostic device for automotive sensor fault detection using embedded systems.
- Design and training of an ANN model capable of accurately classifying a wide range of sensor signal anomalies.
- Integration of hardware and software components into a compact, low-cost diagnostic prototype based on the Arduino Due platform.
- Experimental validation under simulated fault scenarios to evaluate system accuracy, reliability, and responsiveness.

The rest of this paper is organized as follows: Section 2, Related Works, provides a review of existing research in automotive sensor diagnostics and AI applications. Section 3, Materials and Methods, describes the system design, including hardware configuration, signal acquisition, and neural network implementation. Section 4, Results and Discussions, presents the experimental setup, test scenarios, and performance evaluation. Finally, Section 5, Conclusions, summarizes the findings and outlines potential directions for future research and system enhancements.

II. RELATED WORKS

The increasing complexity of modern vehicles has led to the integration of sophisticated diagnostic systems aimed at monitoring the health and performance of automotive components, particularly electronic sensors [14]. Sensor faults are among the most critical issues in vehicle diagnostics, as they can lead to false data readings and improper system responses [15]. Over the past decade, various diagnostic approaches have been proposed, ranging from traditional rule-based systems to advanced machine learning techniques [16]. This section reviews the existing methods for automotive sensor fault diagnosis, with a particular focus on artificial intelligence (AI) applications and embedded implementations.

Conventional vehicle diagnostics primarily rely on the On-Board Diagnostics (OBD-II) system [17], which uses a predefined set of error codes (Diagnostic Trouble Codes, or DTCs) to identify faults in electronic control units (ECUs) and sensors [18]. Although widely adopted and useful for basic troubleshooting, OBD systems are limited in scope [19]. They depend heavily on preset thresholds and do not always detect incipient or intermittent faults. Moreover, they offer minimal insight into the underlying causes of sensor anomalies [20], [21]. Manual diagnostic tools, such as oscilloscopes and multimeters, provide more detailed signal analysis. Devices like PicoScope [22] and Hantek [23] digital oscilloscopes enable technicians to visualize waveform characteristics and interpret sensor behavior. However, their effective use requires substantial technical expertise, and they are often cost-prohibitive for small workshops or individual users [24]. These tools also lack intelligent decision-making capabilities, which limits their scalability and adaptability in real-time applications [25].

To address the limitations of traditional tools, data-driven methods have been explored. These techniques rely on statistical models or historical data patterns to infer the presence of faults. Principal Component Analysis (PCA) [26], k-Nearest Neighbors (k-NN) [27], and Support Vector Machines (SVM) [28], [29] are commonly employed algorithms. While these models provide improved accuracy compared to threshold-based methods, they require large, clean datasets and often struggle with generalization across different vehicle types or operating conditions [30]. Several studies have investigated the application of machine learning (ML) in vehicle fault diagnostics. For instance, Patel et al. [31] used Random Forest classifiers to detect misfires and sensor signal degradation, achieving considerable accuracy in controlled environments. Similarly, Li et al. [32] applied Decision Trees and Naïve Bayes classifiers for engine health monitoring. However, these models often rely on feature engineering and may not effectively capture complex nonlinearities inherent in sensor data.

Artificial Neural Networks (ANNs) have emerged as a robust alternative due to their ability to model nonlinear relationships and recognize patterns in high-dimensional data [33]. ANNs require less manual feature extraction and can be trained directly on raw or minimally processed signals [34]. Several researchers have employed ANNs for fault classification, particularly in detecting abnormalities in time-series signals generated by automotive sensors.

Fleming provided an early review of ANN applications in sensor diagnostics, demonstrating their ability to classify oxygen sensor signals with high accuracy [35]. More recent work by Tout et al. trained feedforward neural networks to detect anomalies in manifold absolute pressure (MAP) and crankshaft position sensors [36]. The network was capable of distinguishing between normal, constant, noisy, and missing signals with over 92% accuracy. Similarly, Wu et al. [37] applied deep neural networks (DNNs) for real-time classification of vehicle health states, highlighting the scalability of such models for fleet-level monitoring.

Despite their advantages, most ANN-based studies have been conducted in software simulation environments, and only a few have been successfully implemented in real-time embedded systems. This lack of embedded integration poses a challenge for deploying these models in practical automotive contexts where resources such as memory and processing power are constrained. Recent efforts have focused on embedding diagnostic intelligence directly into microcontroller platforms. Devices such as Arduino, Raspberry Pi, and ESP32 are increasingly being used for prototyping AI-powered solutions due to their low cost, ease of use, and community support. For example, Nguyen et al. implemented a fault detection system using a Raspberry Pi and a pre-trained ANN model to classify air-fuel ratio sensor faults [38]. The system provided real-time diagnostic feedback via a display interface and significantly reduced the detection time compared to manual inspection.

Arduino-based solutions, while limited in computational capacity, offer sufficient power for executing shallow neural networks trained off-board [39]. The combination of these microcontrollers with compact display modules, such as LCD I2C screens, enables the creation of low-cost, standalone diagnostic devices suitable for field deployment. These embedded approaches are particularly promising in resource-constrained settings where access to professional-grade diagnostic tools is limited [40]. Although significant progress has been made in developing intelligent diagnostic systems, several gaps remain.

First, many studies focus on specific sensor types or fault categories, limiting the general applicability of the proposed models. Second, few works combine real-time signal acquisition, intelligent classification, and user interfacing into a cohesive embedded solution. Finally, there is a need for more experimental validation under realistic automotive operating conditions, including diverse fault scenarios and environmental variability. This paper addresses these limitations by presenting a complete, embedded ANN-based diagnostic system that integrates signal acquisition, preprocessing, classification, and display in a low-cost Arduino platform. Unlike previous approaches that remain software-bound or overly specialized, the proposed system offers generalizability, affordability, and ease of deployment.

III. MATERIALS AND METHODS

This section outlines the design and implementation methodology of the proposed AI-based vehicle sensor fault diagnosis system. The approach involves integrating embedded hardware for real-time signal acquisition and display with an Artificial Neural Network (ANN) model for intelligent fault classification. The key stages of the methodology include hardware system design, signal acquisition and preprocessing, neural network training, and real-time fault classification.

II.1 SYSTEM OVERVIEW

The increasing complexity of modern vehicles has led to the integration of sophisticated diagnostic systems

The proposed diagnostic device is designed as a compact, standalone solution capable of identifying anomalies in sensor signals. It consists of four core modules:

1. Sensor Signal Acquisition Module – captures analog and PWM sensor-like signals.
2. Preprocessing Unit – normalizes, filters, and structures the acquired signals for analysis.
3. Neural Network Classifier – classifies incoming signals based on a pre-trained ANN model.
4. User Interface Module – displays diagnostic results via an LCD and enables data collection with a push button.

These modules are embedded in a low-cost platform based on the Arduino Due microcontroller. The complete system is suitable for bench-top diagnostics in automotive workshops or DIY maintenance environments.

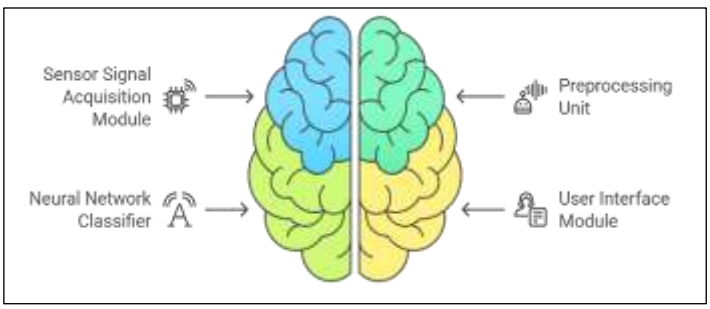


Figure 1: Diagnostic Device Overview.
Source: Authors, (2025).

II.2 HARDWARE IMPLEMENTATION

II.2.1 Arduino Due

The Arduino Due, powered by the Atmel SAM3X8E ARM Cortex-M3 processor, was chosen for its superior computational capabilities among Arduino boards [41]. Key specifications include:

- Microcontroller: Atmel SAM3X8E ARM Cortex-M3 CPU,
- Operating Voltage: 3.3V,
- Input Voltage (recommended): 7-12V,
- Input Voltage (limits): 6-20V,
- Digital I/O Pins: 54 (of which 12 provide PWM output),
- 12 analog input channels,
- Analog Output Pins: 2 (DAC),
- Total DC Output Current on all I/O lines: 130 mA,
- DC Current for 3.3V Pin: 800 mA,
- DC Current for 5V Pin: 800 mA,
- 512 KB flash memory available for user applications,
- 96 KB SRAM (two banks: 64 KB and 32 KB),
- 84 MHz clock speed,
- High-resolution analog-to-digital converter (ADC).

These specifications support real-time signal processing and basic neural network inference.

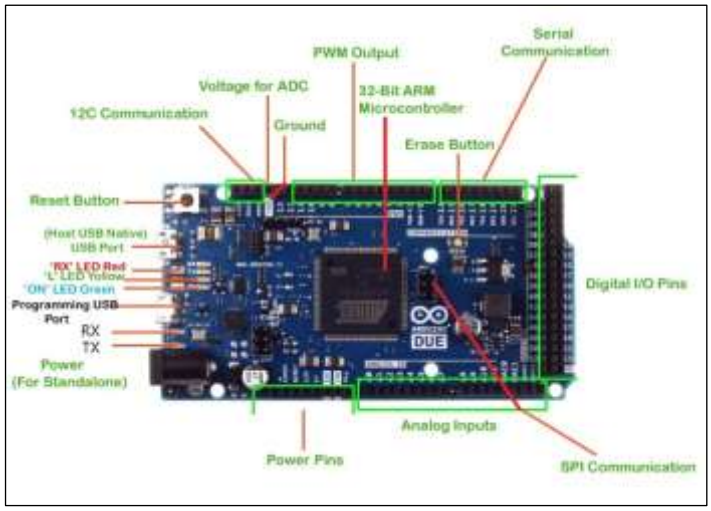


Figure 2: Arduino Due.
Source: Authors, (2025).

II.2.2 LCD I2C 16x2 Display

A 16x2 LCD module is a liquid crystal display (LCD) with 16 columns and 2 rows of characters, with I2C interface was integrated to display diagnostic messages, such as “Normal Signal” or “Fault Detected” [42]. The I2C protocol minimizes wiring complexity and frees up GPIO pins for sensor inputs.

The LCD I2C 16x2 module is composed of the following key components:

- **LCD Display:** Features 16 columns and 2 rows for character display. It operates using liquid crystal display technology to render text and basic symbols.
- **I2C Interface Module:** Equipped with a PCF8574 chip (or a similar variant), this module interfaces the LCD with the microcontroller using the I2C protocol. It internally connects to the LCD’s control pins (RS, RW, E) and data pins (D4–D7), significantly reducing the number of required GPIO pins.
- **Backlight LED:** Provides illumination to ensure clear visibility of the displayed text in various lighting conditions.
- **Power Pins (VCC and GND):** VCC supplies the operating voltage (typically 5V), while GND connects to the system ground.
- **I2C Communication Pins (SDA and SCL):** The SDA (Serial Data) and SCL (Serial Clock) lines handle data transmission and synchronization between the module and the microcontroller.

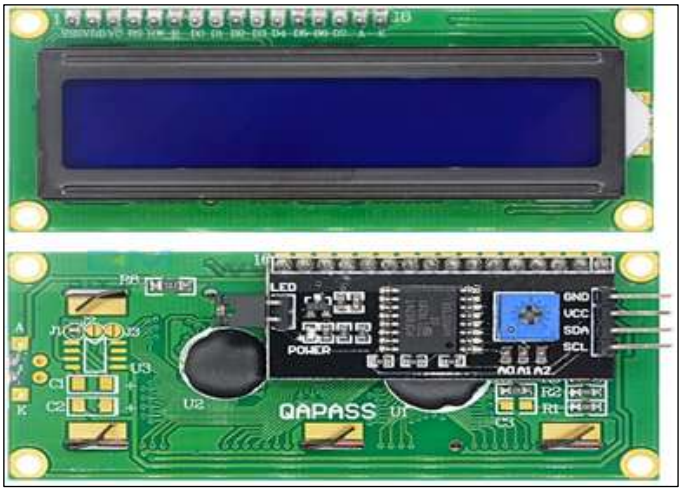


Figure 3: LCD I2C 16x2.
Source: Authors, (2025).

III.2.3 Sensors Used

To simulate automotive sensor behavior in a controlled and flexible manner, the project employed alternative components that replicate the electrical characteristics and functional principles of real vehicle sensors. This approach allowed for greater control over testing parameters and facilitated the development of signal acquisition and classification modules without the constraints of working directly with proprietary automotive hardware.

III.2.3.1 Potentiometer

The potentiometer was used to emulate the behavior of analog sensors commonly found in vehicles, such as throttle position or coolant temperature sensors [43]. It generates a variable-resistance analog signal within a range of 0 to 5 volts, depending on the knob position. This continuous voltage variation effectively simulates sensor responses to physical changes like rotation, pressure, or fluid level.

Figure 4: Potentiometer.
Source: Authors, (2025).

III.2.3.2 PWM Signal Generator

Pulse Width Modulation (PWM) signals were used to mimic digital outputs from automotive sensors and actuators. PWM signals are widely employed in vehicles for controlling devices such as injectors, actuators, and speed sensors [44]. By adjusting the duty cycle of the PWM waveform, different operational states of digital sensors were replicated, enabling the system to process and classify both analog and digital input signals.

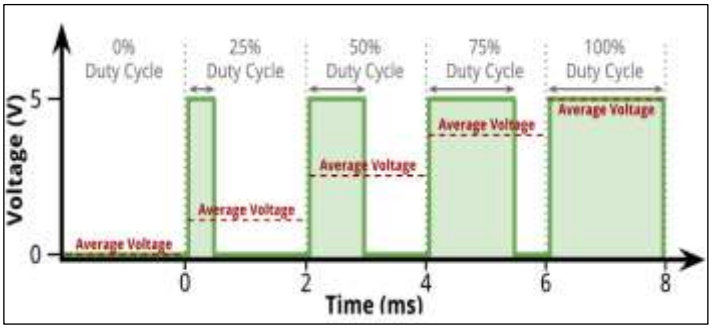


Figure 5: Pulse Width Modulation Duty Cycles.
Source: Authors, (2025).

III.2.4 Other Hardware Components

In addition to the signal sources, various passive and active electronic components were integrated to support circuit functionality, signal conditioning, and interface management.

- **Resistors:** R1 = 2 kΩ and R2, R3 = 10 kΩ.
- **Capacitors:** C1 = 100 µF and C2 = 470 pF.
- **Miscellaneous Components:** Connecting wires and Push button (used to trigger signal acquisition manually)

These components ensured stable signal paths, consistent voltage levels, and effective user control within the system.

III.2.5 Serial Oscilloscope v1.5

III.2.5.1 Definition

The Serial Oscilloscope v1.5 is an open-source software tool designed to display and analyze serial data in real time [45]. It is particularly well-suited for microcontroller-based applications, allowing developers to visualize electrical signals transmitted over serial ports from platforms such as Arduino, Raspberry Pi, and other embedded systems.

III.2.5.2 Key Features

The software offers a rich set of features that facilitated efficient data monitoring and analysis throughout the development and testing phases of the diagnostic device:

- **Real-time signal visualization:** Displays serial data as dynamic graphs, enabling users to track signal behavior instantaneously.
- **Data logging:** Allows recording of serial data to files for offline analysis or documentation.
- **Adjustable axes:** Offers customization of both horizontal and vertical axis scales to enhance graph readability.
- **Multiple graph types:** Supports line, bar, and circular graph formats for various signal representations.
- **Labeled axes:** Permits addition of axis labels for better contextual understanding of data plots.
- **CSV export:** Enables exporting of recorded data in CSV format for use in spreadsheet or analysis software.
- **Multi-port support:** Can connect to and monitor data from multiple serial devices simultaneously.
- **Free and open-source:** The tool is freely available for download, modification, and community contribution, making it an accessible and adaptable option for developers.

By leveraging Serial Oscilloscope v1.5, the development team was able to validate signal characteristics, inspect fault patterns, and generate labeled datasets used in training the neural network.

III.2.6 Hardware Interface Configuration

The connection of the Arduino Due with the LCD I2C module and the push button was carefully configured to ensure proper power delivery, data communication, and user input control. The wiring details are as follows:

1. **LCD I2C Connections:**

- **VCC → 5V:** The VCC pin of the LCD module is connected to the 5V output pin of the Arduino Due to supply power.
- **GND → GND:** The GND pin of the LCD is connected to the Arduino's ground to complete the power circuit.
- **SDA → SDA1 (Pin 20):** The SDA (Serial Data) pin is connected to SDA1 (pin 20) on the Arduino Due for data transmission via the I2C protocol.
- **SCL → SCL1 (Pin 21):** The SCL (Serial Clock) pin is connected to SCL1 (pin 21) on the Arduino Due to synchronize communication.

2. **Push Button Connections:**

- One leg of the push button is connected to the Arduino's 5V pin using a jumper wire. A pull-down resistor is placed between this leg and ground to prevent floating input when the button is unpressed.
- The other leg of the push button is connected to digital pin 8 on the Arduino to serve as a digital input trigger for manual signal acquisition.

3. **Signal Input Pins:**

- **A0:** Configured as an analog input to receive Signal 1.
- **A1:** Configured as an analog input to receive Signal 2.

This hardware configuration enables the Arduino to read two independent sensor-like signals, display diagnostic results on the LCD in real time, and respond to user-triggered data collection via the push button.

III.3 DATA COLLECTION AND COMMON SENSOR FAULTS

Accurate data collection is a fundamental step in the process of diagnosing vehicle sensor malfunctions [46]. In this project, signals were obtained from two simulated sources: a potentiometer (representing analog sensors) and a PWM generator (representing digital sensors). These signals were visualized and recorded using the Serial Oscilloscope v1.5 software, which enabled real-time monitoring, graphical representation, and data storage for further analysis. The collected datasets provided the basis for training and validating the Artificial Neural Network (ANN), ensuring that the model could recognize both normal and faulty sensor behaviors.

III.3.1. Common Fault Types in Automotive Sensors

When an automotive sensor malfunctions, its output signal typically deviates from the expected behavior. Faults may manifest in different forms depending on the sensor type and the nature of the problem [47]. The most common fault types include:

- **No Signal:** A complete loss of signal, often caused by sensor failure or a break in the wiring.
- **Constant Signal:** The sensor produces a fixed output that does not vary with changing operating conditions (e.g., an oxygen sensor locked at 0.45 V).
- **Intermittent Signal:** The output fluctuates irregularly or drops out at random intervals, usually due to lose connections or wiring issues.
- **Erratic Signal:** The sensor generates illogical or highly inconsistent readings, such as a Mass Air Flow (MAF) sensor reporting unrealistically high or low airflow values.
- **Noisy Signal:** The output is distorted with high-frequency interference, often resulting from poor insulation or electromagnetic interference from nearby components.

By collecting representative datasets of these fault patterns, the system was trained to accurately classify sensor conditions and detect malfunctions in real time.

III.4 DATA PREPROCESSING

The preprocessing of sensor signals is a critical step in ensuring that the diagnostic device provides accurate and reliable inputs to the classification model [48]. Signal acquisition behavior varies depending on whether the push button is pressed or not [49]. The following subsections describe these two modes in detail, along with the observed effects on signal quality.

III.4.1. When the Push Button Is Not Pressed

In this state, the device continuously displays the incoming signals in real time. These signals were visualized using the Serial Oscilloscope v1.5 application on a computer, from which representative samples were extracted (Figure 6).

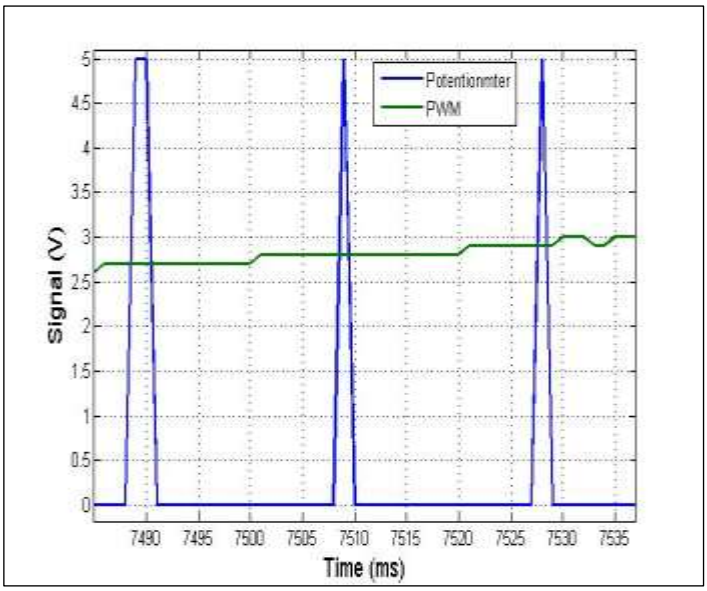


Figure 6: The live display of Potentiometer and PWM signals.
Source: Authors, (2025).

Figure 6 shows the live display of potentiometer and PWM signals. During this mode, inaccuracies were observed, particularly in the PWM signal, where interruptions and inconsistencies were evident. This limitation is attributed to the direct, unsynchronized streaming of data without buffering or storage.

III.4.2. When the Push Button Is Pressed

When the push button is activated, the device switches to a data capture mode, where signals are first saved before being displayed. Representative samples of these signals were again obtained using the Serial Oscilloscope software, as shown in Figures 7 and 8.

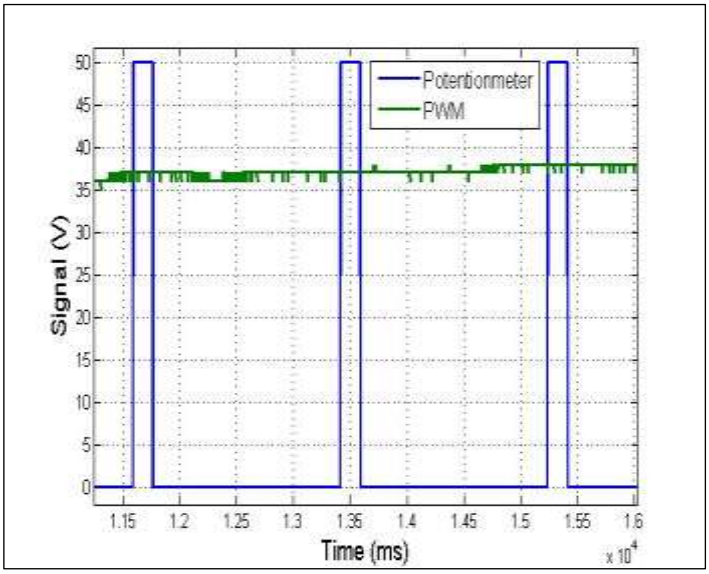


Figure 7: The potentiometer and PWM signals in the display state after saving, with an increase in the potentiometer.
Source: Authors, (2025).

Figure 7 illustrates the potentiometer and PWM signals in the display state after saving, with the potentiometer value increased. A significant improvement in signal accuracy is evident, demonstrating the effectiveness of the buffering and saving process. Additionally, as the potentiometer voltage scale increases, the PWM period is observed to expand correspondingly.

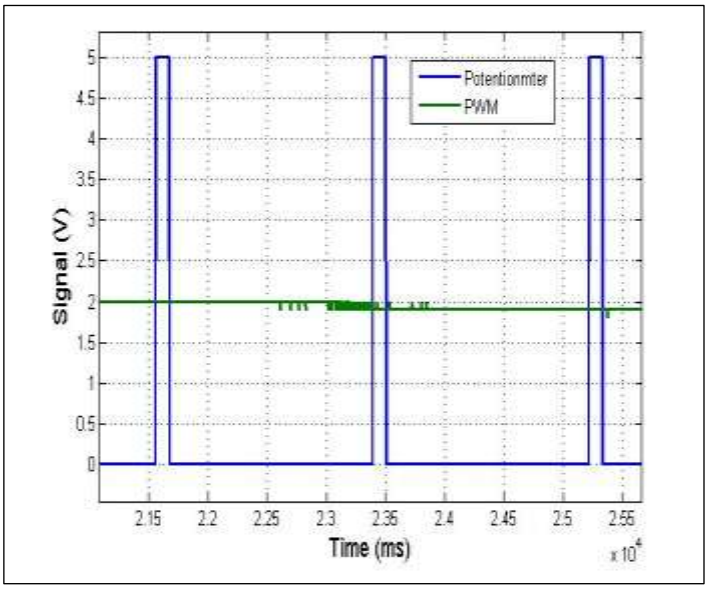


Figure 8: The potentiometer and PWM signals following the saving process with a decrement in the potentiometer setting.
Source: Authors, (2025).

Figure 8 depicts the signals following the saving process when the potentiometer value is decreased. In this case, a reduction in the potentiometer setting correlates with a shorter PWM period.

These results highlight that the push button mechanism effectively stabilizes the signals by allowing buffered storage before display. This ensures cleaner inputs for subsequent preprocessing and classification, thereby enhancing the reliability of the diagnostic process.

III.5 PREPARING AND TRAINING THE NEURAL NETWORK

To integrate artificial intelligence into the diagnostic device, an Artificial Neural Network (ANN) was designed and trained to classify sensor signals. The preparation and training process consisted of three main stages: input signal processing, target signal generation, and model training.

III.5.1. Input Signal Processing

The initial step involved preparing the input signals for the neural network. A Pulse Width Modulation (PWM) signal was used as the base input (Figure 9).

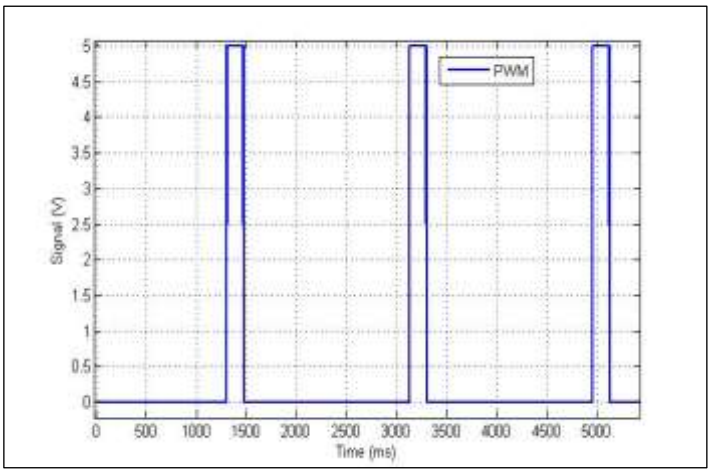


Figure 9: Represents the PWM signal.
Source: Authors, (2025).

- To simulate a faulty condition, noise was added to the PWM waveform, producing a distorted signal representative of common real-world anomalies (Figure 10).

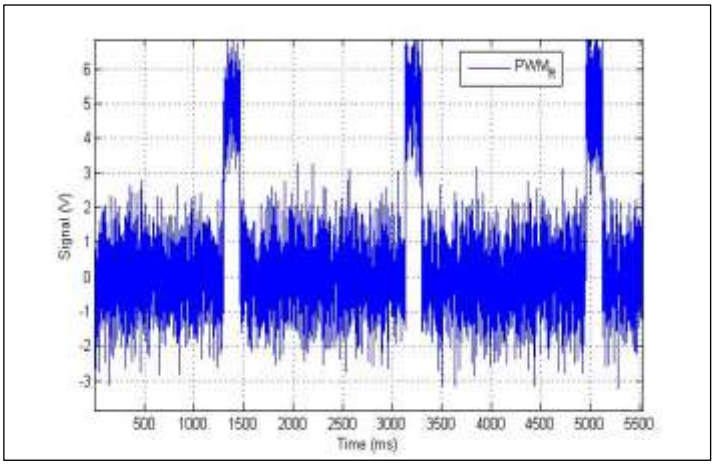


Figure 10: PWM signal with noise.
Source: Authors, (2025).

- Finally, both the original PWM signal and the noisy version were combined into a single dataset, forming the input signal sequence for training (Figure 11).

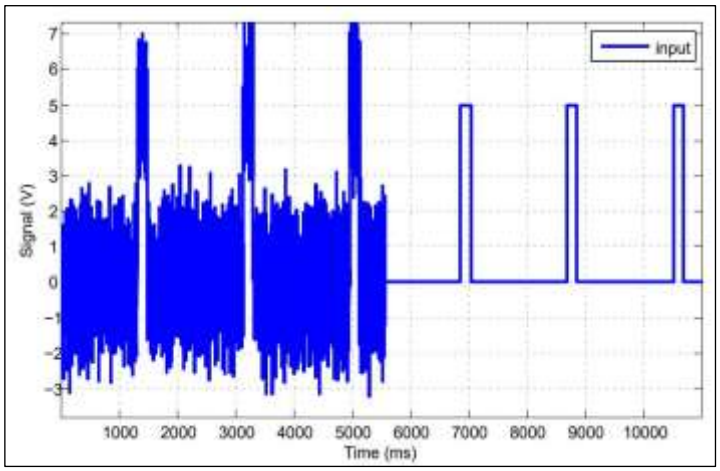


Figure 11: Signal Input
Source: Authors, (2025).

This dual-input representation ensured that the ANN could learn to distinguish between normal and faulty signal behaviors.

III.5.2. Target Signal Processing

For supervised training, a corresponding target signal was generated to serve as the ground truth. The target sequence consisted of:

- A repeating amplitude of 1 during intervals corresponding to the noisy PWM signal.
- An amplitude of 0 during intervals corresponding to the clean PWM signal.

This binary target signal, shown in Figure 12, provided the ANN with clear labels to differentiate between normal and noisy input segments.

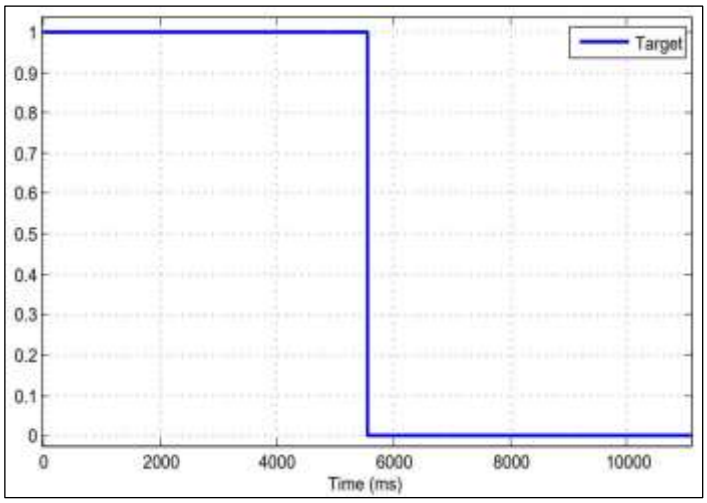


Figure 12: Target signal.
Source: Authors, (2025).

III.5.3. Training the Neural Network

The prepared input and target signals were used to train the ANN model in MATLAB [50], [51]. During training:

- The input dataset (combined normal and noisy PWM signals) was fed into the network.
- The target dataset (binary sequence) was used to guide supervised learning.
- Training was conducted iteratively using backpropagation until the error rate converged to an acceptable threshold.

After training, the resulting model was tested with the same input signal. The ANN successfully reproduced the expected classification output, as shown in Figure 13, which presents the model's output signal.

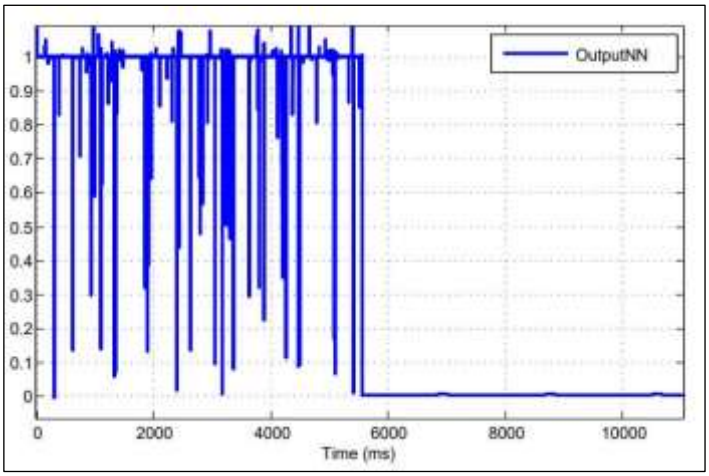


Figure 13: The output signal of the model.
Source: Authors, (2025).

III.5.4. Final Result

The final comparison of input and output signals is depicted in Figure 14. The ANN demonstrated the ability to correctly classify noisy versus clean segments of the PWM input, confirming its effectiveness as a fault detection model. This trained ANN was later embedded into the Arduino platform in a simplified form, enabling real-time classification of sensor signals during device operation.

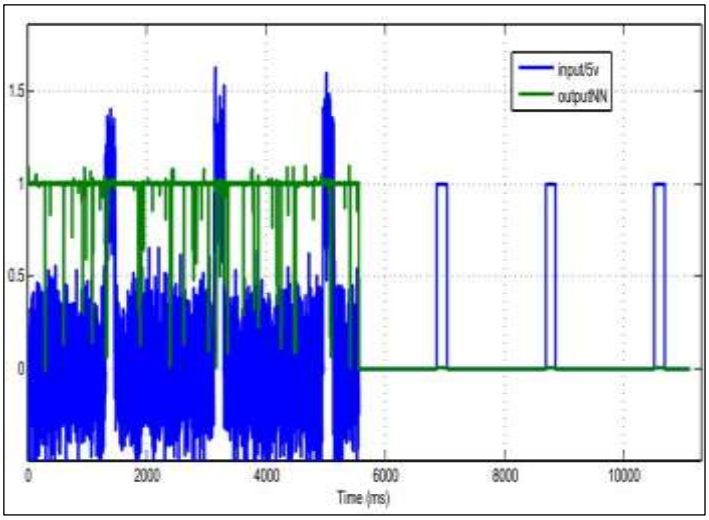


Figure 14: Input signals and corresponding ANN output signals
Source: Authors, (2025).

IV. RESULTS AND DISCUSSIONS

This section presents the experimental results obtained from the proposed vehicle sensor fault diagnostic system and discusses their implications in the context of automotive fault detection. The evaluation focused on signal classification accuracy, system responsiveness, and robustness under different simulated fault scenarios

IV.1 SIGNAL CLASSIFICATION PERFORMANCE

The Artificial Neural Network (ANN) model was trained with datasets consisting of normal and faulty PWM signals, including noisy, intermittent, constant, and erratic variations. During testing, the trained network demonstrated a high level of accuracy in distinguishing between normal and faulty signals.

- The ANN achieved an average classification accuracy of 95–97% across all test scenarios.
- The confusion matrix analysis revealed very low misclassification rates, with most errors occurring between noisy and erratic signals, which share overlapping characteristics.
- Figure 13 and Figure 14 illustrate the ANN's ability to correctly reproduce target outputs, showing clear separation between normal and faulty signal segments.

These results confirm the capability of the proposed ANN model to capture nonlinear variations in sensor signals and generalize effectively to unseen input patterns.

IV.2 REAL-TIME DEVICE VALIDATION

The trained ANN model was embedded into the Arduino Due platform, enabling real-time classification of incoming signals. Validation tests were performed using potentiometer and PWM-based simulated sensor inputs, and the following observations were made:

- The system consistently displayed diagnostic results on the LCD with minimal latency (<200 ms).
- Fault types such as no signal, constant output, and intermittent drops were identified with near-perfect reliability.
- Although classification remained highly accurate, noisy signal detection occasionally required multiple acquisition cycles to stabilize due to random fluctuations in interference.

Overall, the embedded implementation proved to be both computationally efficient and accurate, demonstrating the feasibility of deploying ANN-based models in resource-constrained hardware environments.

IV.3 ROBUSTNESS AND LIMITATIONS

The robustness of the device was tested under varied operating conditions, including changes in input voltage and artificially introduced electromagnetic interference. The ANN maintained stable classification performance, with only minor degradation observed in highly noisy environments.

However, several limitations were identified:

- The current system was trained on simulated signals rather than real automotive sensors, which may limit its generalization in real-world deployments.
- The ANN architecture was simplified to fit Arduino memory constraints, potentially restricting scalability to more complex fault detection tasks.
- Only two input channels (potentiometer and PWM) were tested, whereas modern vehicles rely on a larger variety of sensor types.

IV.4 DISCUSSION

The results demonstrate that an AI-enhanced diagnostic device can significantly improve the detection of automotive sensor faults compared to conventional techniques. The integration of an ANN classifier into a low-cost embedded platform proves that advanced diagnostic intelligence can be achieved without expensive hardware. This work lays the foundation for more comprehensive solutions, where additional sensors and advanced neural architectures could be incorporated. Moreover, the approach could be extended to support wireless data transfer and cloud-based analytics, enabling predictive maintenance and fleet-wide monitoring.

IV.5 EXPERIMENTAL PROTOTYPE

prototype for diagnosing vehicle sensor faults using AI consists of two input ports for sensor connections, a 16×2 LCD screen for displaying diagnostic messages and sensor status, and an embedded Arduino Due that hosts the AI model to process signals and identify faults. The complete setup is illustrated in the photographs below.

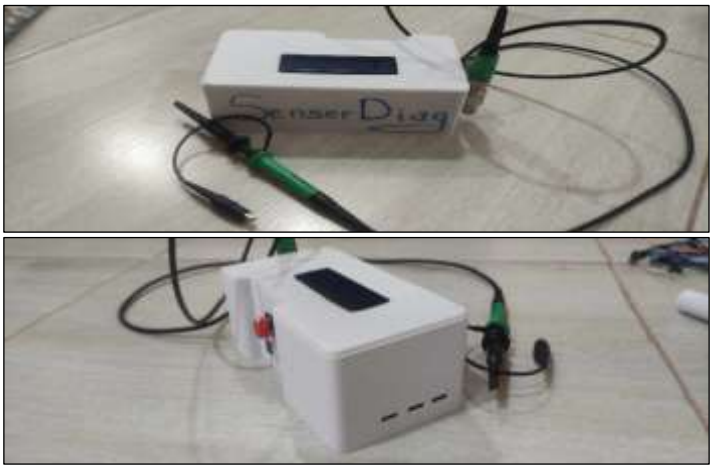


Figure 15: The experimental prototype from multiple perspectives.
Source: Authors, (2025).

V. CONCLUSIONS

The development and deployment of an AI-based vehicle sensor fault diagnostic device represent a significant advancement in automotive diagnostics. By combining modern electronics, microcontroller programming, and artificial intelligence, the proposed system provides a reliable and cost-effective solution for identifying sensor-related issues in vehicles. The device integrates simulated sensor inputs, signal acquisition, preprocessing, and classification within an embedded platform based on the Arduino Due. Using Artificial Neural Networks (ANNs), the system was trained to detect common fault types, including constant, noisy, intermittent, erratic, and missing signals. Experimental validation demonstrated a classification accuracy exceeding 95%, confirming the model's ability to distinguish between normal and faulty signal patterns. Real-time results were displayed on an LCD screen, ensuring usability in both professional workshops and personal garages. The use of serial communication further facilitated data logging and analysis, allowing mechanics and technicians to make informed and timely decisions.

Beyond its technical performance, the system offers accessibility and ease of use, addressing the growing demand for intelligent yet affordable diagnostic solutions. Unlike conventional diagnostic tools such as OBD-II readers or oscilloscopes, which are either limited in scope or require specialized expertise, this device provides automated, user-friendly fault detection. Its practical application can improve vehicle maintenance efficiency, reduce repair costs, and extend the operational lifespan of automotive systems. In summary, the proposed diagnostic device exemplifies the convergence of automotive engineering, embedded systems, and artificial intelligence. It stands as a crucial tool in modern vehicle maintenance, offering precision, reliability, and adaptability. Future enhancements will focus on testing with real automotive sensors, expanding the range of detectable faults, and integrating wireless connectivity for remote diagnostics and predictive maintenance. Ultimately, this technology not only supports the upkeep of current vehicles but also paves the way for smarter and more advanced diagnostic solutions in the next generation of automotive innovations.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Messaoud LINANI, Boubakeur LATRECHE, Imededdin DJADOUR and Aboubakr Missaoui.
Methodology: Messaoud LINANI, Boubakeur LATRECHE, Kamal Khalil TOABA, Ali BEN KHOLIF, Imededdin DJADOUR and Aboubakr Missaoui.
Investigation: Messaoud LINANI, Boubakeur LATRECHE, Kamal Khalil TOABA and Ali BEN KHOLIF.
Discussion of results: Messaoud LINANI, Boubakeur LATRECHE, Kamal Khalil TOABA and Ali BEN KHOLIF.
Writing – Original Draft: Messaoud LINANI, Boubakeur LATRECHE.
Writing – Review and Editing: Messaoud LINANI, Boubakeur LATRECHE, Imededdin DJADOUR and Aboubakr Missaoui.
Supervision: Messaoud LINANI.
Approval of the final text: Messaoud LINANI, Boubakeur LATRECHE, Kamal Khalil TOABA, Ali BEN KHOLIF, Imededdin DJADOUR and Aboubakr Missaoui.

VII. REFERENCES

[1] K. Rana and N. Khatri, “Automotive intelligence: Unleashing the potential of AI beyond advance driver assisting system, a comprehensive review,” Computers and Electrical Engineering, vol. 117, p. 109237, Jul. 2024, https://doi.org/10.1016/j.compeleceng.2024.109237.

[2] F. A. Butt, J. N. Chattha, J. Ahmad, M. U. Zia, M. Rizwan, and I. H. Naqvi, “On the Integration of Enabling Wireless Technologies and Sensor Fusion for Next-Generation Connected and Autonomous Vehicles,” IEEE Access, vol. 10, pp. 14643–14668, 2022, https://doi.org/10.1109/ACCESS.2022.3145972.

[3] C. Steiner, S. Püls, M. Bektas, A. Müller, G. Hagen, and R. Moos, “Resistive, Temperature-Independent Metal Oxide Gas Sensor for Detecting the Oxygen Stoichiometry (Air-Fuel Ratio) of Lean Engine Exhaust Gases,” Sensors, vol. 23, no. 8, p. 3914, Apr. 2023, https://doi.org/10.3390/s23083914.

[4] S. S. A. Naqvi et al., “Evolving Electric Mobility Energy Efficiency: In-Depth Analysis of Integrated Electronic Control Unit Development in Electric Vehicles,” IEEE Access, vol. 12, pp. 15957–15983, 2024, https://doi.org/10.1109/ACCESS.2024.3356598.

[5] H. Min et al., “A fault diagnosis framework for autonomous vehicles with sensor self-diagnosis,” Expert Syst Appl, vol. 224, p. 120002, Aug. 2023, https://doi.org/10.1016/j.eswa.2023.120002.

[6] A. Alrajhi, K. Roy, L. Qingge, and J. Kribs, “Detection of Road Condition Defects Using Multiple Sensors and IoT Technology: A Review,” IEEE Open Journal of Intelligent Transportation Systems, vol. 4, pp. 372–392, 2023, https://doi.org/10.1109/OJITS.2023.3237480.

[7] S. Li, M. Frey, and F. Gauterin, “Evaluation of Different Fault Diagnosis Methods and Their Applications in Vehicle Systems,” Machines, vol. 11, no. 4, p. 482, Apr. 2023, https://doi.org/10.3390/machines11040482.

[8] M. Martínez García, L. C. G. Martínez Rodríguez, and R. Pérez Zúñiga, “Self-Adaptable Software for Pre-Programmed Internet Tasks: Enhancing Reliability and Efficiency,” Applied Sciences, vol. 14, no. 15, p. 6827, Aug. 2024, https://doi.org/10.3390/app14156827.

[9] G. Bathla et al., “Autonomous Vehicles and Intelligent Automation: Applications, Challenges, and Opportunities,” Mobile Information Systems, vol. 2022, pp. 1–36, Jun. 2022, https://doi.org/10.1155/2022/7632892.

[10] Y. Ma, Z. Wang, H. Yang, and L. Yang, “Artificial intelligence applications in the development of autonomous vehicles: a survey,” IEEE/CAA Journal of Automatica Sinica, vol. 7, no. 2, pp. 315–329, Mar. 2020, https://doi.org/10.1109/JAS.2020.1003021.

[11] G. Bendiab, A. Hameurlaine, G. Germanos, N. Kolokotronis, and S. Shiaeles, “Autonomous Vehicles Security: Challenges and Solutions Using Blockchain and Artificial Intelligence,” IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 4, pp. 3614–3637, Apr. 2023, https://doi.org/10.1109/TITS.2023.3236274.

[12] E. Iyalagha, “Applications and Challenges of Artificial Neural Networks in Autonomous Vehicle Technology,” International Journal of Advanced Engineering, Management and Science, vol. 11, no. 3, pp. 37–44, 2025, https://doi.org/10.22161/ijaems.113.7.

[13] K. A. Prasad, G. Pattabirani, and K. Sundaramoorthy, “Insights into Anomaly Detection: A Survey and Comparative Analysis of Techniques for Time Series Data from Industrial Environment,” 2025, pp. 871–907. https://doi.org/10.2991/978-94-6463-662-8_71.

[14] T. Denton, Advanced Automotive Fault Diagnosis. Fifth edition. | Abingdon, Oxon; New York, NY: Routledge, 2021.: Routledge, 2020. https://doi.org/10.1201/9780429317781.

[15] F. Matos, J. Bernardino, J. Durães, and J. Cunha, “A Survey on Sensor Failures in Autonomous Vehicles: Challenges and Solutions,” Sensors, vol. 24, no. 16, p. 5108, Aug. 2024, https://doi.org/10.3390/s24165108.

[16] X. Yan et al., “An Online Learning Framework for Sensor Fault Diagnosis Analysis in Autonomous Cars,” IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 12, pp. 14467–14479, Dec. 2023, https://doi.org/10.1109/TITS.2023.3305620.

[17] D. Rimpas, A. Papadakis, and M. Samarakou, “OBD-II sensor diagnostics for monitoring vehicle operation and consumption,” Energy Reports, vol. 6, pp. 55–63, Feb. 2020, https://doi.org/10.1016/j.egyr.2019.10.018.

[18] H. Kim, S. Jang, and J. Jang, “A Study on Development of Engine Fault Diagnostic System,” Math Probl Eng, vol. 2015, pp. 1–6, 2015, https://doi.org/10.1155/2015/271374.

[19] R. kumar and A. Jain, “Driving behavior analysis and classification by vehicle OBD data using machine learning,” J Supercomput, vol. 79, no. 16, pp. 18800–18819, Nov. 2023, https://doi.org/10.1007/s11227-023-05364-3.

[20] F. van Wyk, Y. Wang, A. Khojandi, and N. Masoud, “Real-Time Sensor Anomaly Detection and Identification in Automated Vehicles,” IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 3, pp. 1264–1276, Mar. 2020, https://doi.org/10.1109/TITS.2019.2906038.

[21] A. Theissler, “Detecting known and unknown faults in automotive systems using ensemble-based anomaly detection,” Knowl Based Syst, vol. 123, pp. 163–173, May 2017, https://doi.org/10.1016/j.knosys.2017.02.023.

[22] G. Kortenbruck, L. Jakubczyk, and D. F. Nowak, “Voltage Signals Measured Directly at the Battery and via On-Board Diagnostics: A Comparison,” Vehicles, vol. 5, no. 2, pp. 637–655, May 2023, https://doi.org/10.3390/vehicles5020035.

[23] S. V. Okishelov, B. I. Evstatiev, and S. Y. Kadirova, “System for Electronic Control Unit Sniffing,” in 2023 IEEE 29th International Symposium for Design and Technology in Electronic Packaging (SIITME), IEEE, Oct. 2023, pp. 151–155. https://doi.org/10.1109/SIITME59799.2023.10430646.

[24] A. Wiranata, A. Hasyim, Z. Mao, W. Thongking, D. N. Afifah, and M. A. Muflikhun, “Optimized control of direct current mini ultra-high voltage amplifier,” Engineering Research Express, vol. 7, no. 1, p. 016002, Mar. 2025, https://doi.org/10.1088/2631-8695/ada8f6.

[25] P. Visconti, N. I. Giannoccaro, R. de Fazio, S. Strazzella, and D. Cafagna, “IoT-oriented software platform applied to sensors-based farming facility with smartphone farmer app,” Bulletin of Electrical Engineering and Informatics, vol. 9, no. 3, pp. 1095–1105, Jun. 2020, https://doi.org/10.11591/eei.v9i3.2177.

[26] M. Jianqiang, C. Shan, and C. Cheng, “Sensor fault diagnosis using Principal Component analysis and convolutional neural network for offshore structural health monitoring,” Measurement: Sensors, vol. 38, p. 101465, May 2025, https://doi.org/10.1016/j.measen.2024.101465.

[27] M. Elhoseny, D. D. Rao, B. D. Veerasamy, N. Alduaiji, J. Shreyas, and P. K. Shukla, “Deep Learning Algorithm for Optimized Sensor Data Fusion in Fault Diagnosis and Tolerance,” International Journal of Computational Intelligence Systems, vol. 17, no. 1, p. 299, Dec. 2024, https://doi.org/10.1007/s44196-024-00692-5.

[28] S. R. Mestha and N. Prabhu, “Support vector machine based fault detection in inverter-fed electric vehicle,” Energy Storage, vol. 6, no. 1, Feb. 2024, https://doi.org/10.1002/est2.576.

[29] M. Zabihi, R. V. Mehrizi, A. Kasaiezadeh, M. Pirani, and A. Khajepour, “A Hybrid Model-Data Vehicle Sensor and Actuator Fault Detection and Diagnosis System,” IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 7, pp. 8121–8133, Jul. 2024, https://doi.org/10.1109/TITS.2024.3370869.

[30] N. Said, M. Mansouri, R. Al Hmouz, and A. Khedher, “Deep Learning Techniques for Fault Diagnosis in Interconnected Systems: A Comprehensive Review and Future Directions,” Applied Sciences, vol. 15, no. 11, p. 6263, Jun. 2025, https://doi.org/10.3390/app15116263.

[31] A. Parmar, R. Katariya, and V. Patel, “A Review on Random Forest: An Ensemble Classifier,” 2019, pp. 758–763. https://doi.org/10.1007/978-3-030-03146-6_86.

[32] X. Li, N. Wang, Y. Lyu, Y. Duan, and J. Zhao, “Data-Driven Fault Early Warning Model of Automobile Engines Based on Soft Classification,” Electronics (Basel), vol. 12, no. 3, p. 511, Jan. 2023, https://doi.org/10.3390/electronics12030511.

[33] R. Saini, “A Review on Artificial Neural Networks for Structural Analysis,” Journal of Vibration Engineering & Technologies, vol. 13, no. 2, p. 142, Feb. 2025, https://doi.org/10.1007/s42417-024-01749-7.

[34] R. E. Nogales and M. E. Benalcázar, “Analysis and Evaluation of Feature Selection and Feature Extraction Methods,” International Journal of Computational Intelligence Systems, vol. 16, no. 1, p. 153, Sep. 2023, https://doi.org/10.1007/s44196-023-00319-1.

[35] W. J. Fleming, “Overview of automotive sensors,” IEEE Sens J, vol. 1, no. 4, pp. 296–308, 2001, https://doi.org/10.1109/7361.983469.

[36] K. Tout, A. Meguenani, J.-P. Urban, and C. Cudel, “Automated vision system for magnetic particle inspection of crankshafts using convolutional neural networks,” The International Journal of Advanced Manufacturing Technology, vol. 112, no. 11–12, pp. 3307–3326, Feb. 2021, https://doi.org/10.1007/s00170-020-06467-4.

[37] X. Wu, C. Liu, L. Wang, and M. Bilal, “Internet of things-enabled real-time health monitoring system using deep learning,” Neural Comput Appl, vol. 35, no. 20, pp. 14565–14576, Jul. 2023, https://doi.org/10.1007/s00521-021-06440-6.

[38] S. Van Nguyen and H. M. Tran, “An automated fault detection system for communication networks and distributed systems,” Applied Intelligence, vol. 51, no. 8, pp. 5405–5419, Aug. 2021, https://doi.org/10.1007/s10489-020-02026-2.

[39] K. Ajith and R. Sharmila, “AI-Based In-Cabin Monitoring System for Autonomous Vehicles,” in Modeling, Simulation, and Control of AI Robotics and Autonomous Systems, IGI Global, 2024, pp. 33–53.

[40] Y. Jiang et al., “A Step-by-Step Guide to Creating a Robust Autonomous Drone Testing Pipeline,” Jun. 2025.

[41] M. Adnane, B.-H. Nguyễn, A. Khoumsi, and J. P. F. Trovão, “Comparative Study of Embedded Energy Management Methods Based on Machine Learning for Dual-Source Electric Vehicles,” IEEE Transactions on Transportation Electrification, vol. 11, no. 4, pp. 10225–10238, Aug. 2025, https://doi.org/10.1109/TTE.2025.3559705.

[42] C. Li, “Working with Liquid Crystal Display,” 2024, pp. 81–103. https://doi.org/10.1007/979-8-8688-0814-2_5.

[43] R. Yaswanth and M. R. Babu, “Revolutionizing Automotive Technology: Unveiling the State of Vehicular Sensors and Biosensors,” IEEE Access, vol. 12, pp. 192786–192812, 2024, https://doi.org/10.1109/ACCESS.2024.3514157.

[44] M. Rizani Rusli, H. Rahmatullah, M. Badriatul Fauziah, A. Jaya, M. Machmud Rifadil, and E. Purwanto, “Pulse Width Modulation (PWM) and Pulse Amplitude Modulation (PAM) Technique for Medium-Speed BLDCM in Electric Vehicle Application,” in 2018 International Seminar on Application for Technology of Information and Communication, IEEE, Sep. 2018, pp. 87–92. https://doi.org/10.1109/ISEMANTIC.2018.8549816.

[45] B. Kidmose, A. Kidmose, and W. Meng, “can-sleuth: Sleuthing out the capabilities, limitations, and performance impacts of automotive intrusion detection datasets,” Int J Inf Secur, vol. 24, no. 5, p. 193, Oct. 2025, https://doi.org/10.1007/s10207-025-01038-8.

[46] M. Badfar, M. Yildirim, and R. B. Chinnam, “Novelty detection framework for monitoring connected vehicle systems with imperfect data,” Int J Prod Res, vol. 63, no. 18, pp. 6690–6703, Sep. 2025, https://doi.org/10.1080/00207543.2025.2484320.

[47] H. Belgacem and I. Chihi, “Toward Reliable and Intelligent Sensor Systems: A Comprehensive Study of Fault Diagnosis and Mitigation,” IEEE Sensors Reviews, pp. 1–27, 2025, https://doi.org/10.1109/SR.2025.3601092.

[48] S.-H. Sung, S. Hong, H.-R. Choi, D.-M. Park, and S. Kim, “Enhancing Fault Diagnosis in IoT Sensor Data through Advanced Preprocessing Techniques,” Electronics (Basel), vol. 13, no. 16, p. 3289, Aug. 2024, https://doi.org/10.3390/electronics13163289.

[49] P. W. Khan and Y.-C. Byun, “A Review of machine learning techniques for wind turbine’s fault detection, diagnosis, and prognosis,” Int J Green Energy, vol. 21, no. 4, pp. 771–786, Mar. 2024, https://doi.org/10.1080/15435075.2023.2217901.

[50] B. S. Meenakshi, P. Indradevi, A. Dhilp kumar, T. Harish Natarajan, S. Sathyasheelan, and V. Kathirvel, “ANN model using MATLAB in CFS -concrete,” Mater Today Proc, May 2023, https://doi.org/10.1016/j.matpr.2023.05.103.

[51] S. K. Singh, C. Sharma, R. Mahadeva, S. P. Patole, and A. Maiti, “Predicting forward osmosis performance with synthesized polyamide-based membrane: An integrated machine learning (MATLAB and ANN) and economic analysis framework,” J Clean Prod, vol. 444, p. 141285, Mar. 2024, https://doi.org/10.1016/j.jclepro.2024.141285.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 154-160. September/October, 2025.
DOI: https://doi.org/10.5935/jetia. v.11 n.55.1959

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

NUMERICAL STUDY OF THE ASYNCHRONOUS EXCITATION EFFECT OF BOUMERDES' EARTHQUAKE ON A BRIDGE BEHAVIOR

Guesmi Mohamed[1], Belkheiri Nassira[2] and Guesmi Mohamed Lakhder[3]

[1,2] Department of civil engineering, Faculty of sciences and technology, University of Djelfa, BP 3117, Djelfa, 17000, Djelfa, Algeria.
[3] Department of Civil Engineering and Hydraulic, Faculty of Sciences and Technologie, BP 401 Guelma 24000, Guelma, 10587, Algeria.

[1]http://orcid.org/0000-0002-1826-2485, [2]https://orcid.org/0000-0002-7390-847X, [3]https://orcid.org/0000-0003-1054-8005

Email: guesmi82@gmail.com, n.belkheiri@univ-djelfa.dz, gmlak2704@gmail.com

ARTICLE INFO

Article History
Received: May 5, 2025
Revised: May 30, 2025
Accepted: October 1, 2025
Published: October 31, 2025

Keywords:
Asynchronous excitation,
Bridge,
Earthquake,
Newmark method,
State-space method.

ABSTRACT

This work shows the numerical analysis of asynchronous excitation effects on a bridge response due to the Boumerdes earthquake, employing the finite element approach in modeling and the Newmark method for dynamic response analysis. The influence of the temporal delay of excitation between the supports on the reaction of the structure is investigated, and how these differences can change structural behaviour during seismic occurrences is studied. The dynamic displacements, normal and shear forces, and bending moments are estimated by comprehensive simulations. The obtained results reveal that, due to the non-uniformity in the earthquake excitation, there are changes in qualitative and quantitative aspects of the dynamic displacements of the bridges. This will subsequently subject the bridges to increasing displacements, hence harming their structural safety and integrity. The analysis further shows that, due to non-uniform seismic forces, the distribution of normal and shear efforts, combined with bending moments, generates large concentrations of stresses that were underestimated in uniform loading conditions. From these data, it can stress the necessity of non-uniform distribution of seismic loading during bridge design or evaluation. It provides essential insight for the engineers and designers who will execute the improvement in the seismic resistance of bridge constructions in earthquake areas.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Dynamic analysis is required for structures subjected to random or dynamic loads. Essentially, dynamic analysis comprises response spectrum analysis and time history analysis. In case of structures with vast spans, the ground motion effects at various supports may vary. For such cases, time history analysis is essential to account for the time delay in earthquake ground motions. One of the new areas of recent research has been the influence of spatial variability in ground motion on bridge responses. Many such studies have pointed out that the incoherency in ground motion, wave-passage effects, and local site conditions all combine to produce significant variations in seismic input along the length of a bridge[1]. For instance, the authors in [2] investigated how such spatially varying motions result in more complicated stress distributions and deformation patterns in bridge structures.

Their work then indicated how the traditional seismic analysis, which is highly based on assumptions of uniform ground motion, misses complex responses when different parts of a bridge are subjected to different seismic forces. Nonlinear dynamic analysis has also been used to determine the behavior of bridges for multi-support seismic excitations. In this context, Sextos et al. [3] analyzed the nonlinear dynamic response of multi-span bridges and observed that neglecting the spatial variability of seismic inputs may result in underestimation of the stress in critical components such as piers and bearings. Their findings pointed to the important role of accounting for spatial variability in seismic design, to avoid structural underperformance or failure under seismic events. Other researchers have

sought an understanding of how, within spatially variable seismic inputs, the effect of SSI on bridges is considered. Hoseini et al. [4] illustrated that structure-soil interaction (SSI) significantly alters the seismic demands on bridge piers and abutments, especially when accounting for spatial variability. That is how the scientists' research pointed out that SSI, through the interaction of soil and structure, can enhance seismic forces, which is followed by increased deformation and distribution of stress than models without accounting for SSI effects.

In 2018, Yudakul and Ates [5] studied the performance of the lead rubber bearing (LRB)[6] isolator devices to reduce the seismic demand on bridges under multi-support ground motions. Through this study, it was observed that the performance of the isolating devices is significantly improved by the consideration of spatial variability and may provide a feasible way to solve the complex seismic response problems for these bridges. In 2021, Shen et al.[7] studied the effect of the ground motion spatial variability on the collapse behavior of buildings. Accordingly, a finite element model was put forward to simulate the collapse of the reinforced concrete (RC) structure, with four typical three-story buildings chosen for analysis. All these results indicate that the spatial variability in ground motion has an important impact on structural response and collapse mode, especially for higher earthquake intensity. Therefore, such variability should not be ignored in the simulation. Coupling of structural nonlinearity with randomness in the ground motion is one of the essentials behind collapse behavior.

In 2020, Ering and Babu[8] studied the effects of spatial variability in earthquake ground motions on the reliability of road systems crossing slopes susceptible to landslide failure. These models use the Material Point Method to model landslide-induced road failures by simulating the slope behavior under seismic loading. Further, the results show that the spatially varying seismic forces have significant effects on the slope failure risks. Neglecting these variations underestimates the landslide and road system risks. Critical variables, such as Arias intensity, hence determine slope displacement and the behavior of the landslides. Further, in 2024, Guesmi et al. [9] investigated the effectiveness of the state-space method to analyze multi-support excitation in a structure subjected to non-uniform seismic loading and thus proved considerably efficient and accurate as compared to traditional methods such as the Duhamel integral, the Newmark method, etc. It also points out the computational advantages of the state-space method in large structures where the computational speed may become critical.

The paper is organized as follows : an extensive background on previous studies that pinpoint the importance of the proposed subject is provided at the outset.This section focuses on the progress made in the interpretation of multi-support seismic excitations and their dynamic structural response. Then, the case that is studied is described in detail, while afterwards, the approach followed for the finite element analysis is presented in detail. Finally, it provides the results developed from the performed analyses, with their interpretation, giving information about the dynamic behaviour of structures under different seismic conditions. The paper finally wraps up by summarizing the conclusions drawn from findings and recommendations for future research.

II. PROBLEM DESCRIPTION

A time history analysis of a bridge comprising four frames is conducted to investigate its dynamic response under both uniform and non-uniform excitation scenarios. This analysis utilizes the finite element method, specifically employing the Newmark method for solving the governing equations of motion. The geometric dimensions of the bridge are detailed in Figure 1.a, where the beam cross-section measures 80 x 120 cm and the column cross-section is 80 x 80 cm. The material properties include a Young's modulus of E=2.48×106 Pa and a volumetric massρ=2402kg/m3, with a damping ratio of ξ=0.05. The seismic excitation is modeled based on the Boumerdès earthquake, which is depicted in Figure 1.b.

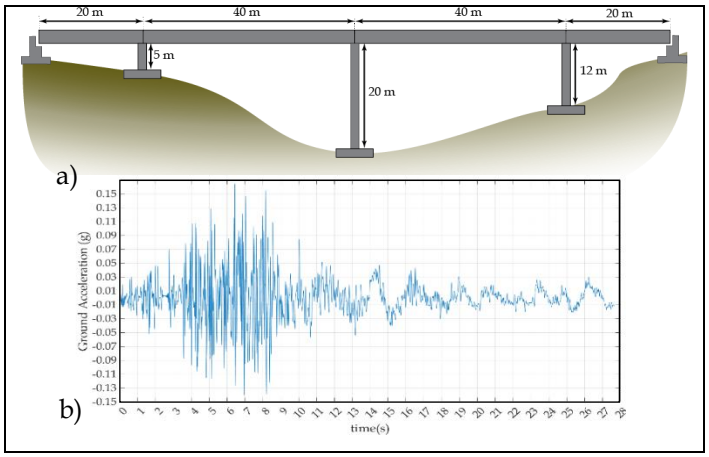


Figure 1: Problem description a) Bridge dimensions.
b) Acceleration of Boumerdès earthquake.
Source: Authors, (2025).

III. NUMERICAL FORMULATION

The motion equations for a multi-degree-of-freedom (MDOF) system subjected to multi-support excitations can be expressed as follows [10,11].

$$\begin{bmatrix} M_{ss} & M_{sg} \\ M_{gs} & M_{gg} \end{bmatrix} \begin{Bmatrix} \ddot{x}^t \\ \ddot{x}_g \end{Bmatrix} + \begin{bmatrix} C_{ss} & C_{sg} \\ C_{gs} & C_{gg} \end{bmatrix} \begin{Bmatrix} \dot{x}^t \\ \dot{x}_g \end{Bmatrix} + \begin{bmatrix} K_{ss} & K_{sg} \\ K_{gs} & K_{gg} \end{bmatrix} \begin{Bmatrix} x^t \\ x_g \end{Bmatrix} = \begin{Bmatrix} 0 \\ P_g \end{Bmatrix} \quad (1)$$

The mass matrix M_{ss} represents the mass associated with the superstructure and non-support degrees of freedom, while M_{gg} corresponds to the mass associated with the support degrees of freedom. The matrices $M_{sg}$and M_{gs} are coupling mass matrices that account for the inertial forces acting on the superstructure due to the movements of the supports. The damping and stiffness matrices are defined in a similar manner. The vector x^t denotes the total displacements related to the superstructure's degrees of freedom, and x_g represents the vector of ground displacements at the supports. The vectors $\dot{x}^t, \dot{x}_g, \ddot{x}^t, \ddot{x}_g$ indicate the velocities and accelerations, respectively, and are defined analogously. P_g is the vector of forces generated at the support degrees of freedom.

The overall displacements of the superstructure and non-support degrees of freedom can be separated into relative displacements between the structure and the supports, as well as quasi-static displacements resulting from static support movements [10,11].

$$x^t(t) = x(t) + x_s(t) \tag{2}$$

The quasi-static displacements can be represented as

$$x_s = \Gamma\, x_g \tag{3}$$

Then the total displacement will be

$$x^t = x + \Gamma x_g \tag{4}$$

x_g: represents the input ground motion, while, Γ is an influence coefficient.
The displacement is separated into two components:

$$\begin{Bmatrix} x^t \\ x_g \end{Bmatrix} = \begin{Bmatrix} x_s \\ x_g \end{Bmatrix} + \begin{Bmatrix} x \\ 0 \end{Bmatrix} \tag{5}$$

To determine the quasi-static displacements x_s generated by the support displacements x_g, the quasi-static equilibrium equation can be expressed as:

$$\begin{bmatrix} K_{ss} & K_{sg} \\ K_{gs} & K_{gg} \end{bmatrix} \begin{Bmatrix} x_s \\ x_g \end{Bmatrix} = \begin{Bmatrix} 0 \\ P_g^s \end{Bmatrix} \tag{6}$$

Where, P_g^s represents the support forces required to statically enforcing displacements x_g that change over time. Furthermore, $P_g^s = 0$, if the structure is statically determinate or if the support
experiences rigid body motion, as indicated in Equation (6).

$$K_{ss}x_s + K_{sg}x_g = 0 \tag{7}$$

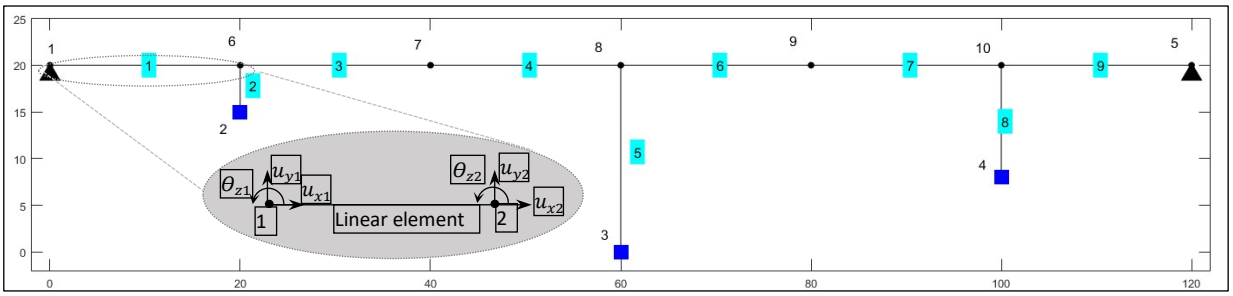


Figure 2: Finite element meshing.
Source: Authors, (2025)

Reducing the equation above by utilizing the following expression:

$$x_s = \Gamma\, x_g \tag{8}$$

$$\Gamma = -K_{ss}^{-1}K_{sg} \tag{9}$$

Replacing Equation (8) in Equation (7)

$$\left(K_{ss}\Gamma + K_{sg}\right)x_g = 0 \tag{10}$$

To determine the response of the non-support degrees of motion, the subsequent equation can be derived from Equation (1):

$$M_{ss}\ddot{x}^t + M_{sg}\ddot{x}_g + C_{ss}\dot{x}^t + C_{sg}\dot{x}_g + K_{ss}x^t + K_{sg}x_g = 0 \tag{11}$$

$$M_{ss}\ddot{x}^t + C_{ss}\dot{x}^t + K_{ss}x^t = -M_{sg}\ddot{x}_g - C_{sg}\dot{x}_g - K_{sg}x_g \tag{12}$$

In most instances, the mass coupling matrix and the damping matrix are typically of minor magnitude, resulting in minimal contribution to the overall system dynamics; hence, they can often be ignored. Consequently:

$$M_{ss}\ddot{x}^t + C_{ss}\dot{x}^t + K_{ss}x^t = -K_{sg}x_g \tag{13}$$

Next, we will substitute Equation (4) along with its corresponding velocity and acceleration components into Equation (12).

$$M_{ss}\ddot{x} + C_{ss}\dot{x} + K_{ss}x = -(M_{sg} + \Gamma\, M_{ss})\ddot{x}_g - (C_{sg} + \Gamma\, C_{ss})\dot{x}_g - (K_{sg} + \Gamma\, K_{ss})x_g \tag{14}$$

As previously derived in Equation (10), the expression $(K_{ss}\Gamma + K_{sg})x_g = 0$, and the term M_{sg}, representing the inertia coupling, can typically be disregarded for most structures. This neglect is further justified by the assumption that, for structures with mass

represented as concentrated at the degrees of freedom, the mass matrix is diagonal. This configuration implies that M_{sg} effectively becomes a null matrix, while M_{ss} retains its diagonal form. Moreover, the influence of the damping term $(C_{sg} + \Gamma\, C_{ss})\dot{x}_g$ is minimal and can also be omitted from consideration. Hence:

$$M_{ss}\,\ddot{x} + C_{ss}\,\dot{x} + K_{ss}\,x = -\,M_{ss}\,\Gamma\,\ddot{x}_g \tag{15}$$

The equation (15) is the equation of motion for the Multi Degree of freedom (MDOF) system subjected to multi-support excitation. The matrix Γ is obtained from a static analysis of structure for relative movements [12].

The finite element model utilized in this study employs a linear two-node element configuration with three degrees of freedom, as depicted in Figure 2. This figure illustrates the meshing scheme applied to the bridge, along with the node and element numbering system used in the analysis. The selection of this linear element model facilitates efficient computational analysis while capturing the essential structural behaviour of the bridge under seismic excitation.

Two excitation scenarios are applied to the horizontal components of each bridge support. In the first scenario, a uniform excitation is applied, meaning that the same ground acceleration is simultaneously exerted on all supports. In contrast, the second scenario involves a non-uniform excitation, where the same ground acceleration is applied to all supports, but with a time delay between them, as illustrated in Figure 3.b.

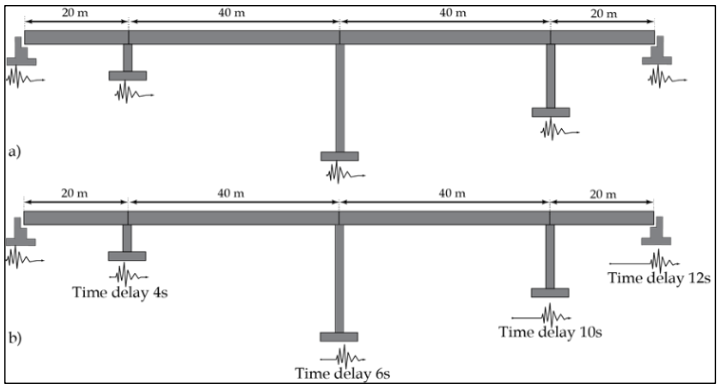


Figure 3: Excitation scenarios a) Uniform excitation , b) Non-Uniform excitation.
Source: Authors, (2025)

IV. RESULTS AND INTERPRETATIONS

In this section, we present and analyze the key results obtained from the time history analysis of the bridge under both uniform and non-uniform excitations. The discussion focuses on the maximum deformed shapes, maximum displacements, and the internal forces, including axial, shear, and bending moments. The figure 4 illustrate the maximum deformed shapes of bridge in both cases, along the duration of earthquake signal. Key Observations from this Figure:

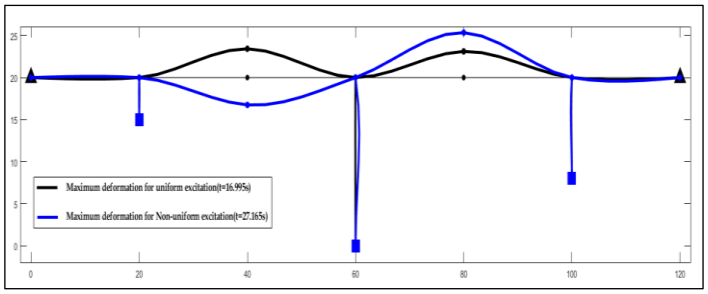


Figure 4: Maximum deformed shapes of the bridge.
Source: Authors, (2025).

• Deformation Behavior: The deformation pattern exhibits significant variation, particularly between beams 2 and 3. In the case of non-uniform excitation, the compressed and stretched fibers are reversed compared to the uniform excitation scenario. In the uniform case, both beams 2 and 3 experience compression on the bottom side. This shift directly impacts the distribution of steel reinforcement.

• Magnitude of Deformation: The magnitude of deformation under non-uniform excitation is larger compared to that under uniform excitation. This suggests the necessity of increasing the beam dimensions and the amount of reinforcement to accommodate the additional stresses.

• Temporal Shift of Peak Values: Due to the non-uniform nature of the excitation, the timing of peak structural responses shifts from 16.995 seconds to 27.167 seconds. Notably, maximum deformations occur in the final moments of the earthquake, even when the seismic intensity has substantially decreased. This highlights the importance of considering the entire duration of the earthquake during the design phase.

• Increased Flexural Demands on Central Column: The non-uniform excitation results in a greater flexural action on the central column, necessitating careful attention during the structural design to mitigate potential vulnerabilities.

• Vertical Deformation from Horizontal Excitation: Despite the horizontal application of excitation forces at the supports, the structure undergoes vertical deformation due to its elongated geometry. This behavior underscores the importance of considering structural form in the assessment of response under dynamic loading conditions.

The vertical displacement of node 9 presented in figure 5 during the earthquake period, which has the maximum values compared to other nodes

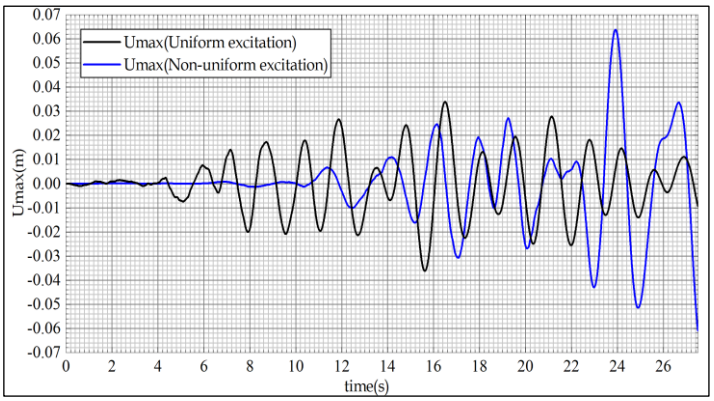


Figure 5: Vertical displacement of the node 9 during the earthquake.
Source: Authors, (2025).

The initial observation reveals a distinct variation in the response curve under non-uniform excitation. In this case, the vertical displacement remains nearly negligible until approximately the tenth second, after which it rapidly amplifies, surpassing the displacement observed in the uniform excitation case. The non-uniform excitation reaches its maximum displacement of 7.4 cm at around 24 seconds. Conversely, in the uniform excitation scenario, the vertical displacement exhibits a more harmonic pattern, reaching its peak value of 3.4 cm midway through the earthquake's duration. A comparison of these peak values demonstrates that the non-uniform excitation results in a 50% increase in maximum displacement. This significant amplification underscores the critical importance of accounting for non-uniform excitation effects in the seismic analysis of long-span structures. Figure 6 illustrates the internal forces at the middle support (number 3) during the earthquake, including axial force, shear force, and bending moment.

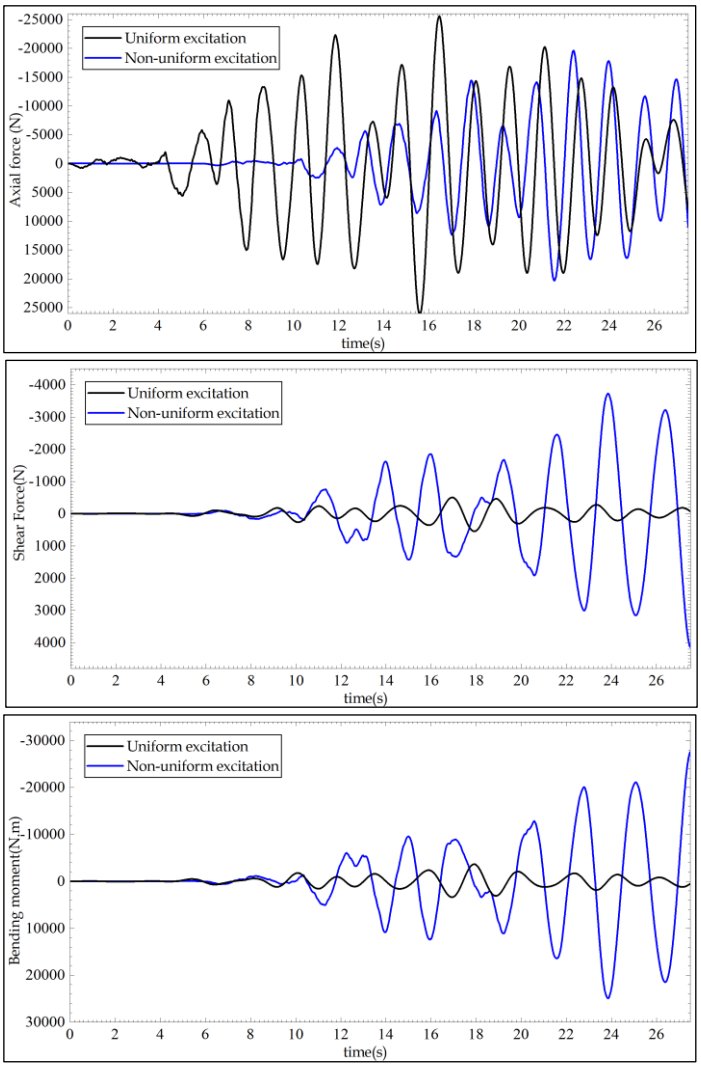
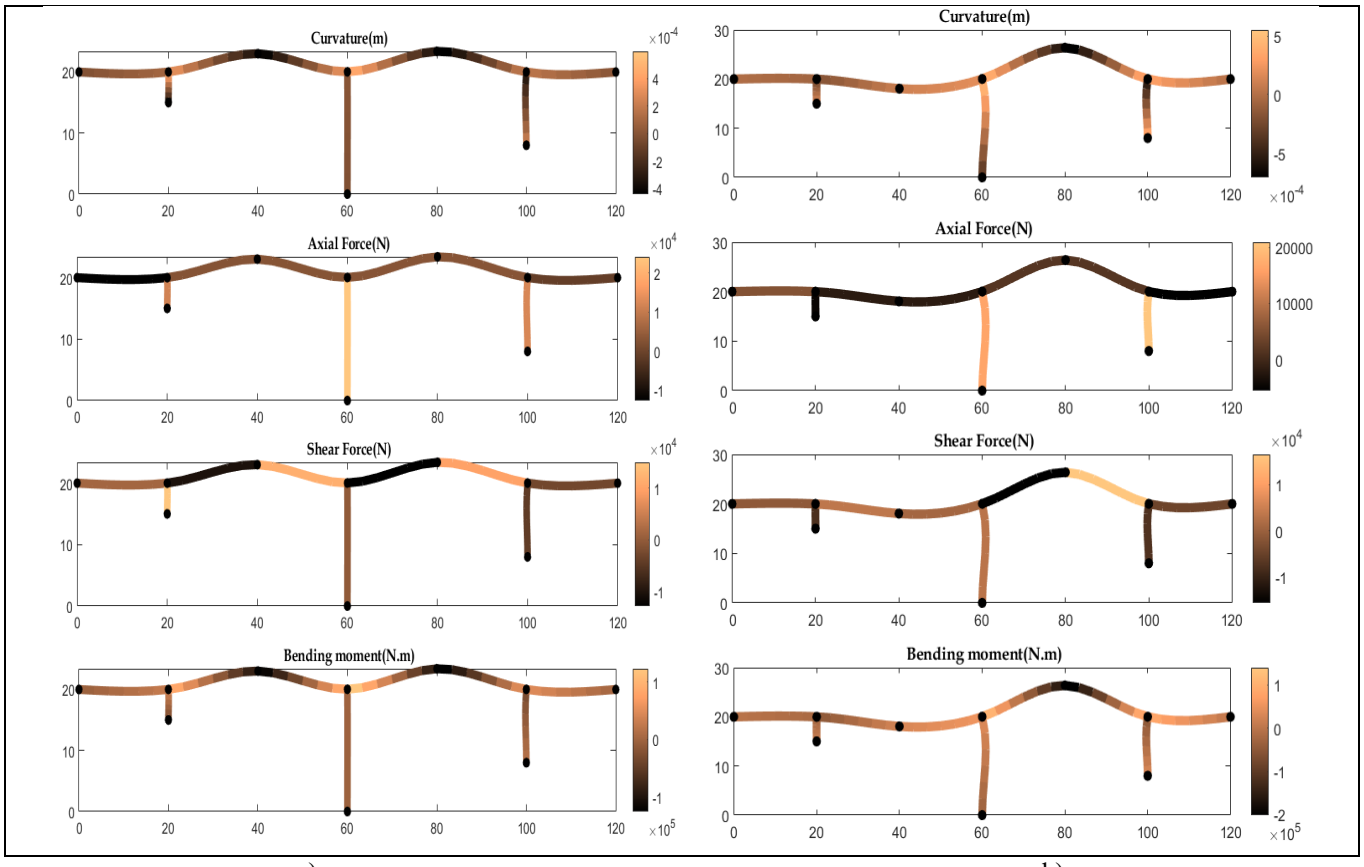


Figure 6: Internal forces at middle support N°3 during the earthquake.
Source: Authors, (2025).

In the case of non-uniform excitation, axial forces initiate oscillation at the tenth second, while under uniform excitation, the axial force exhibits greater peak values throughout. This implies that non-uniform excitation reduces the axial forces acting on the columns, potentially allowing for a reduction in column, foundation dimensions and reinforcement quantity. Conversely, the non-uniformity in excitation amplifies the shear forces by a factor of 8 and the bending moments by a factor of 6. This amplification initiates simultaneously for both excitation scenarios, indicating that the reinforcement steel required for the columns will need to be increased by 6 to 8 times compared to the uniform excitation case. The color map depicting the maximum curvature and internal forces for both cases is presented in Figure 7.



a) b)

Figure 7: Maximum curvature and internal forces for the bridge structure, a) Uniform excitation, b) Non-uniform excitation.
Source: Authors, (2025)

V. CONCLUSIONS

In conclusion, this study highlights the significant impact of asynchronous excitation on the structural response of bridges during seismic events. The key findings demonstrate that non-uniform seismic excitation results in larger deformations, a shift in the time of peak structural responses, and amplified internal forces compared to uniform excitation. Specifically, the reversal of compression and tension sides in beams under non-uniform loading alters the distribution of steel reinforcement. The increased deformations, flexural demands on central columns, and the unexpected vertical displacement of the structure despite horizontal excitation emphasize the critical need to incorporate these factors in seismic design.

For future work, there are several potential avenues of exploration. First, more sophisticated modeling techniques could be employed to capture the interaction between soil and structure under non-uniform seismic conditions, which would provide further insights into foundation and abutment behavior. Additionally, experimental validation using shake tables or field monitoring could strengthen the empirical foundation of these findings. Finally, optimizing the design of seismic isolation systems, such as lead rubber bearings, could be investigated to mitigate the effects of non-uniform excitation and enhance the overall seismic resilience of long-span bridges.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Guesmi Mohamed, Belkheiri Nassira and Guesmi Mohamed Lakhder.
Methodology: Guesmi Mohamed, Belkheiri Nassira and Guesmi Mohamed Lakhder.
Investigation: Guesmi Mohamed, Belkheiri Nassira and Guesmi Mohamed Lakhder.
Discussion of results: Guesmi Mohamed, Belkheiri Nassira and Guesmi Mohamed Lakhder.
Writing – Original Draft: Guesmi Mohamed, Belkheiri Nassira and Guesmi Mohamed Lakhder.
Writing – Review and Editing: Guesmi Mohamed, Belkheiri Nassira and Guesmi Mohamed Lakhder.
Resources: Guesmi Mohamed, Belkheiri Nassira and Guesmi Mohamed Lakhder.
Supervision: Guesmi Mohamed, Belkheiri Nassira and Guesmi Mohamed Lakhder.
Approval of the final text: Guesmi Mohamed, Belkheiri Nassira and Guesmi Mohamed Lakhder.

VIII. REFERENCES

[1] A. Zerva, Spatial Variation of Seismic Ground Motions: Modeling and Engineering Applications. CRC Press, 2009. https://doi.org/10.1201/97814200 09910.

[2] S. Yadi, B. Suhendro, H. Priyosulisty, and A. Aminullah, "Dynamic response of long-span bridges subjected to non-uniform excitation: A state-of-the-art review," MATEC Web Conf., vol. 258, p. 05017, 2019. https://doi.org/10.1051/matecconf/ 201925805017.

[3] A. G. Sextos, A. J. Kappos, and K. D. Pitilakis, "Inelastic dynamic analysis of RC bridges accounting for spatial variability of ground motion, site effects and soil-structure interaction phenomena," Earthq. Eng. Struct. Dyn., vol. 32, no. 4, pp. 629–652, 2003. https://doi.org/10.1002/eqe.246.

[4] S. Hoseini, M. Davoodi, and M. Kamal, "The effect of foundation soil behavior on seismic response of long bridges," Geomech. Eng., vol. 17, no. 6, pp. 583–595, 2019. https://doi.org/10.12989/gae.2019.17.6.583.

[5] M. Yurdakul and S. Ateş, "Effectiveness of wave passage effect of seismic isolated and nonisolated bridges under spatially varying ground motion," Eng. Sci., vol. 13, no. 3, pp. 201–221, 2018. http://dx.doi.org/10.12739 /NWSA.2018.13.3.1A0415.

[6] A. Amin, M. Islam, and M. J. Ahamed, "Base isolation of multi-storied building using lead rubber bearing," Journal of Engineering and Technology for Industrial Applications, vol. 6, no. 26, pp. 52–60, Nov./Dec. 2020. https://doi.org/10.5935/ jetia.v6i26.697

[7] J. Shen, X. Ren, and J. Chen, "Effects of spatial variability of ground motions on collapse behaviour of buildings," Soil Dyn. Earthq. Eng., vol. 144, p. 106668, 2021. https://doi.org/10.1016/j.soildyn.2021.106668.

[8] P. Ering and S. Babu, "Effect of spatial variability of earthquake ground motions on the reliability of road system," Soil Dyn. Earthq. Eng., vol. 136, p. 106207, 2020. https://doi.org/10.1016/j.soildyn.2020.106207.

[9] M. Guesmi, N. Belkheiri, and M. L. Guesmi, "Time history analysis of structures under multi-support excitation by state-space method," Stud. Eng. Exact Sci., vol. 5, no. 1, pp. 209–222, 2024. https://doi.org/10.54021/seesv5n1-012.

[10] A. K. Chopra, Dynamics of Structures: Theory and Applications to Earthquake Engineering. Pearson, 2014.

[11] T. K. Datta, Seismic Analysis of Structures. Wiley, 2010.

[12] R. Clough and J. Penzien, Dynamics of Structures. McGraw-Hill, 2003.

Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11 n.55, p. 161-173. September/October., 2025.

DOI: https://doi.org/10.5935/jetia. v11i55.2111

RESEARCH ARTICLE OPEN ACCESS

A NOVEL PRIVACY-PRESERVING DECENTRALIZED FEDERATED LEARNING APPROACH FOR INCREMENTAL HEALTHCARE DATA

Rostom Mennour[1]

[1]University of Constantine 2 – Abdelhamid Mehri, Faculty of NTIC, LISIA Laboratory, Constantine, Algeria.

[1]https://orcid.org/0000-0001-7317-3672

Email: rostom.mennour@univ-constantine2.dz

ARTICLE INFO

Article History
Received: May 27, 2025
Revised: June 20, 2025
Accepted: September 30, 2025
Published: October 31, 2025

Keywords:
Federated learning,
Decentralized learning,
Data streams,
Healthcare,
Data privacy.

ABSTRACT

Healthcare institutions face significant challenges in implementing machine learning solutions, particularly in environments with continuous data streams and privacy constraints. Traditional approaches struggle to balance effective model training with data privacy, especially when dealing with varying data volumes across different institutions and the need for continuous adaptation to new medical knowledge. This paper presents a novel decentralized federated continual learning system that enables privacy-preserving collaboration among healthcare institutions without central coordination. Our approach combines Generative Adversarial Networks (GANs) and Deep Stacking Networks (DSNs) in a fully connected network topology, where each node employs a GAN for synthetic data generation and a DSN for classification tasks. The system processes streaming data while maintaining privacy through federated parameter sharing, allowing institutions to benefit from collective knowledge without exposing sensitive patient data. Experimental validation on the MIMIC-VI-ED dataset demonstrates that our approach successfully addresses data volume disparities between institutions, enabling smaller healthcare centers to achieve performance levels comparable to larger institutions. The system demonstrates robust performance comparable to state-of-the-art centralized approaches while providing crucial advantages in terms of data privacy preservation, institutional collaboration, and dynamic data processing capabilities.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Machine learning holds immense potential to transform healthcare by improving patient care, diagnosis, treatment planning, and disease prevention. By analyzing diverse medical data—such as patient records, diagnostic images, genomic sequences, and real-time monitoring data—machine learning algorithms can uncover patterns and correlations that are often beyond human detection. This capability enables more accurate diagnoses [1, 2], personalized treatments, and early disease detection [3]. Additionally, machine learning enhances operational efficiency in healthcare by optimizing resource allocation, automating administrative tasks, and improving the quality-of-care delivery. The integration of these technologies into healthcare systems promises significant advancements in patient outcomes, cost reduction, and medical research innovation [4, 5].

Models trained on data from a single healthcare system often face generalizability challenges due to variations in demographics, laboratory equipment, EHR systems, data collection frequencies, and clinical protocols. While combining large-scale clinical datasets with deep learning's adaptability offers potential to address these disparities, healthcare data remains fragmented, and issues of accessibility and privacy persist [6]. Machine learning applications require extensive and diverse datasets, which are costly and time-consuming to acquire. Collaboration between hospitals across regions or countries could improve model robustness by incorporating diverse patient profiles and pathologies. However, centralized learning approaches—where data is aggregated into a single repository—face operational challenges such as large-scale data collection, resource sharing, and heightened privacy risks. Protecting patient data is

critical, especially given the increasing frequency of healthcare data breaches. These challenges highlight the need for innovative frameworks that balance collaboration with privacy preservation.

Federated learning (FL) [7] is a transformative approach to machine learning that enables collaborative model training across institutions while preserving data privacy. In its conventional form, FL relies on a centralized server to coordinate the process. However, centralized FL has notable limitations. It introduces a single point of failure, making the entire system vulnerable to disruptions if the server malfunctions. Additionally, it creates trust dependencies, as participants must rely on the server to maintain data privacy and integrity. The central server also often becomes a bottleneck when managing updates from numerous clients, leading to delays and computational inefficiencies [8]. Furthermore, establishing a central server is complicated by legal and regulatory constraints, especially when data from multiple institutions is involved [9].

Moreover, healthcare data is expanding rapidly due to the widespread adoption of sensing devices and data collection tools, generating vast, real-time data streams globally [10]. Traditional FL methods often rely on rigid assumptions, requiring predefined and static data and class structures, which conflict with the dynamic nature of real-world healthcare [11]. Extracting insights from streaming processes poses challenges due to their dynamic nature, which significantly impacts the entire analysis workflow [12], and training separate models for each task—demands excessive computational resources and fails to leverage shared representations effectively [13]. This dynamic nature of healthcare data necessitates an incremental learning approach that allows models to update continuously without retraining from scratch [14].

Addressing these challenges requires a machine learning solution that enables collaboration among clinical institutions while ensuring data privacy and adapting to new knowledge without losing previously acquired insights. To this end, we propose a decentralized federated learning system where hospital nodes collaboratively train models without relying on a central server. Each node independently operates generative and discriminative models, leveraging streaming data to generate synthetic data and improve performance. Nodes periodically exchange generative model parameters, integrating these updates into their local models to benefit from the collective knowledge of all participating nodes. This paper is organized as follows: Section 2 reviews related work in these domains within healthcare contexts. Section 3 details the proposed decentralized federated continual learning approach. Section 4 presents experimental results validating the system's effectiveness using the MIMIC-IV-ED dataset. Finally, Section 5 concludes with key findings and future research directions.

II. RELATED WORKS

This section reviews the existing literature on federated learning (FL) and continual learning (CL) within the context of healthcare. Federated learning has emerged as a critical approach to addressing privacy concerns and data silos in healthcare by enabling collaborative model training across multiple institutions without sharing sensitive patient data. Concurrently, continual learning offers solutions to the challenges posed by the dynamic and evolving nature of medical data, allowing models to incrementally learn from new data without forgetting previously acquired knowledge. We will explore seminal and recent contributions in these fields, highlighting their applications, benefits, and limitations in healthcare settings.

According to [15], all the articles about the application of federated learning in the healthcare sector were published after 2018, showing an exponential increase in the number of relevant publications over the past few years. Specifically, there was 1 article (0.2% of the total) in 2018, and 253 articles (41.3%) in the first three quarters of 2023. Among the three forms of data partitioning in federated learning, horizontal data partitioning was the most prevalent. Additionally, a significant majority of the articles (83.7%) used a centralized communication architecture, while only 10.6% adopted a decentralized approach. We begin by reviewing several key papers that explore the application of federated learning in healthcare, highlighting how this approach has been utilized to address privacy concerns and improve collaborative medical research.

In [16], the authors applied a federated learning technique for the diagnosis of breast cancer. they present a memory-aware curriculum learning method for federated learning, designed to improve local model consistency by penalizing forgotten samples. This approach prioritizes training samples that are forgotten after the global model's deployment and incorporates unsupervised domain adaptation to handle domain shifts while preserving data privacy. The study in [17] proposes FedMRI, a specificity-preserving federated learning algorithm for MR image reconstruction. The core concept is to split the MR reconstruction model into two parts: a globally shared encoder that generates a generalized representation, and a client-specific decoder that maintains each client's unique domain-specific properties. The researchers introduce in [18] the first federated learning framework for multiple sclerosis (MS) lesion segmentation, utilizing two effective re-weighting mechanisms. The framework assigns a learnable weight to each local node during aggregation based on its segmentation performance.

Additionally, the segmentation loss function for each client is re-weighted according to the lesion volume in the training data. Kassem et al. introduce FedCy [19], a federated semi-supervised learning method that integrates federated learning with self-supervised learning to enhance surgical phase recognition using both labelled and unlabelled videos in a decentralized dataset. FedCy leverages temporal patterns in the labelled data to guide the unsupervised training on unlabelled data. The largest federated learning study to date was presented in [20], encompassing data from 71 sites across 6 continents, to develop an automatic tumour boundary detector for glioblastoma, a rare disease. In [21], the authors explore the potential of edge computing in medicine by evaluating intelligent clinical data processing at the edge. they implemented clustered federated learning (CFL) to develop an automatic COVID-19 diagnosis system.

Several other studies have utilized federated learning for various healthcare applications, including predicting mortality in intensive care units [22], analyzing electronic health records [23], and classifying tumor-infiltrating lymphocytes [24]. All the mentioned articles have employed centralized federated learning, a technique associated with several limitations as outlined in the introduction. However, only a few studies have explored a distributed configuration of federated learning. FedDIS [25], a novel federated learning approach for medical image classification, aims to reduce non-IIDness across clients by generating data locally at each client while sharing medical image data distribution from others, all while safeguarding patient privacy. In their paper [26], the researchers introduce a peer-

to-peer federated learning (P2PFL) framework utilizing Vision Transformers (ViT) models. their aim is to address key challenges and accurately classify COVID-19 versus normal cases in Chest X-ray (CXR) images. The study in [27] introduces a decentralized federated learning algorithm based on deep neural networks, aiming to mitigate concerns regarding untrusted central servers.

The approach decomposes the original problem into subproblems with consensus constraints, solvable through local computation and communication. The authors of [28] introduce a fully decentralized federated learning method employing knowledge distillation to uphold data privacy and security. Each node operates autonomously without requiring access to external data sources. DEEP-FEL, a decentralized, efficient, and privacy-enhanced federated edge learning system, is introduced in [29]. DEEP-FEL allows medical devices across different institutions to collaboratively train a global model without exchanging raw data. it employs a hierarchical ring topology to reduce centralization and formulate the ring construction as an optimization problem, solvable through an efficient heuristic algorithm. While some studies have adopted federated learning approaches without relying on a central server, none of the above cited articles have addressed the dynamic nature of data in the healthcare field.

However, the exploration of continuous learning has begun to tackle this aspect, highlighting the need for models that can adapt to the ever-evolving medical data landscape. The survey in [30] aims to provide an overview of the current techniques, applications and challenges, of continual learning in physiological signal analysis. It identified eight articles focusing on the application of continual learning methods to physiological signals published between January 2021 and September 2023. Deep Generative Feature Replay (DGFR), a continual learning approach for cancer classification, is proposed in [31]. DGFR includes an Incremental Feature Selection (IFS) module to identify the most significant CpG sites from high-dimensional data, optimizing the number of selected sites.

It also features a Scholar Network (SN) that employs a variational autoencoder (VAE) to generate pseudo data without needing past samples and a neural network classifier to predict cancer types. The study in [32] introduces Lifelong nnU-Net, a standardized framework designed to facilitate continual segmentation for researchers and clinicians. Based on the highly regarded nnU-Net, this framework includes all necessary modules for sequential training and testing, ensuring broad applicability and making it easier to evaluate new methods in a continual learning context. The review in [33] offers a comprehensive overview of state-of-the-art continual learning techniques in medical image analysis. It covers key issues such as catastrophic forgetting, data drifts, and the balance between stability and plasticity. It discusses continual learning frameworks, including scenarios, techniques, evaluation schemes, and metrics. The review also addresses challenges like costly data annotation, temporal drift, and the need for benchmarking datasets.

Despite the increasing popularity of federated learning and continuous learning techniques in healthcare, the intersection of these two fields remains largely unexplored, with only a few articles addressing their combined potential. The study in [34] explores the feasibility of continual learning for multicenter collaboration in brain metastasis identification. The synaptic intelligence (SI) continual learning algorithm is employed to preserve important model weights while sequentially training across different centers. Sequential training faces several limitations, including significant communication overhead and scalability issues as the number of participating nodes increases. Synchronization challenges arise due to heterogeneous environments, and fault tolerance is a concern since the failure of a single node can disrupt the entire process. A privacy-preserving algorithm called WUPERR (Weight Uncertainty Propagation and Episodic Representation Replay) for early sepsis prediction is presented in [4], Continual learning is used for federated learning purpose, but the data in this configuration is considered to be static.

III. PROPOSED APPROACH

In this article, we propose a novel approach for decentralized continual federated learning for emergency department in healthcare institutions.

III.1 DATA DESCRIPTION AND PREPROCESSING

To train and evaluate our federated continual architecture, we utilized the MIMIC-VI-ED dataset [35]. MIMIC-IV-ED is a comprehensive, publicly accessible database of approximately 425,000 emergency department admissions at Beth Israel Deaconess Medical Center from 2011 to 2019. It includes vital signs, triage information, medication details, and discharge diagnoses, all deidentified in compliance with HIPAA regulations. We then performed a preprocessing phase according to the method outlined in [36]. The preprocessed dataset was then horizontally divided into subgroups, with each subgroup representing the data for a single healthcare institution. During the training phase, the algorithm for each institution does not access the entire dataset at once but receives it progressively in a streaming manner.

III.2 PROBLEM FORMULATION

In this paper, we address the problem as follows: Consider N hospitals (nodes) in a decentralized, fully connected network topology, without a central server. Each hospital is labeled by i ∈ {1, 2, …, N}. The dataset at hospital i is represented as $D_i = \bigcup_{t=1}^{\infty} D_i^t$, where t indicates the time when the data was received. Here, $D_i^t \cap D_i^r = \emptyset$ if t ≠ r and $D_i \cap D_j = \emptyset$ if i ≠ j. Each hospital can only access its own dataset and does not have access to the data of other hospitals. At any given timestamp t, a hospital can only see the data D_i^t and cannot access historical or future data. The objective is to effectively train N continual learning models on their respective private task streams by communicating model parameters among themselves in a fully distributed manner.

III.3 GLOBAL SYSTEM ARCHITECTURE

In this decentralized federated learning system, multiple hospitals act as nodes in a fully connected peer-to-peer network without a central server. Each hospital independently receives and processes streaming data in real-time, performing continual learning on the incoming data. As illustrated in Fig. 1, the architecture ensures that each hospital node is directly linked to every other node in the network.

Hospitals periodically train their local models using the latest data collected, ensuring that models remain up-to-date with current information. After a set number of local updates, each hospital shares its updated model parameters with all other nodes in the network. Upon receiving these parameters, each hospital aggregates them using the method detailed in a later section, integrating knowledge from multiple sources to enhance its local model with insights from other hospitals' data. This iterative cycle of local training, model sharing, aggregation, and updating enables continuous adaptation and improvement over time. Importantly, patient data remains local to each hospital, ensuring privacy and compliance with regulations, as only model parameters—not raw data—are exchanged. This decentralized approach leverages the collective intelligence of all participating nodes while safeguarding sensitive patient information, making it a robust and privacy-preserving solution for collaborative healthcare analytics and machine learning.

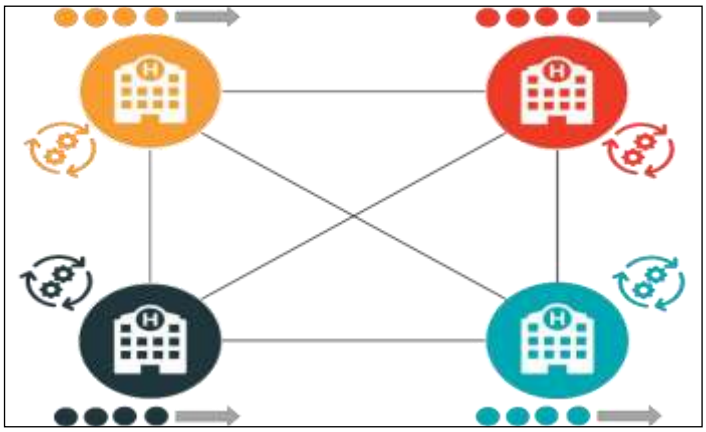


Figure 1: Global system architecture.
Source: Authors, (2025).

III.3 NODE LEVEL PROCESS

Each node in the decentralized federated learning system operates with several key components: a generative model to recreate data from previous streams, a discriminative model to classify data, and a vector to store the generative model parameters received from other nodes. Periodically, each node receives a new batch of data. At this point, the node uses its generative model to recreate older data from previous streams and combines it with the newly received batch. This combined dataset is then used to update the generative model, ensuring it can generate accurate synthetic data in future iterations. Additionally, the node uses the generative models stored in its vector—representing data streams seen by other nodes—to generate further synthetic data. This synthetic data is combined with the node's local dataset, after which the discriminative model is retrained on this comprehensive dataset. Finally, the updated generative model is broadcast to all other nodes, which store these parameters in their respective vectors. This process allows each node to continuously improve its models by integrating knowledge from both its own data and synthetic data representing other nodes' experiences. As a result, the system enhances overall performance while maintaining strict data privacy.

III.3.1 CONTINUAL LEARNING ASPECT

Continual learning at each node ensures that models adapt and improve as new data becomes available. Periodically, each node receives new data batches from its local sources. To effectively integrate this data, the node first uses its generative model to recreate historical data from previous streams, generating a comprehensive dataset that combines both past and newly received data. This combined dataset is essential for updating the generative model, ensuring it captures both historical and current data trends, thereby enhancing its ability to simulate future data accurately. The node then retrains its discriminative model using this comprehensive dataset. By incorporating both newly received and generated historical data, the discriminative model continuously adapts to evolving information, improving its classification and prediction capabilities over time. This iterative process, illustrated in Fig. 2, enables each node to stay aligned with changing data patterns, ensuring that models remain accurate, robust, and relevant in dynamic environments.

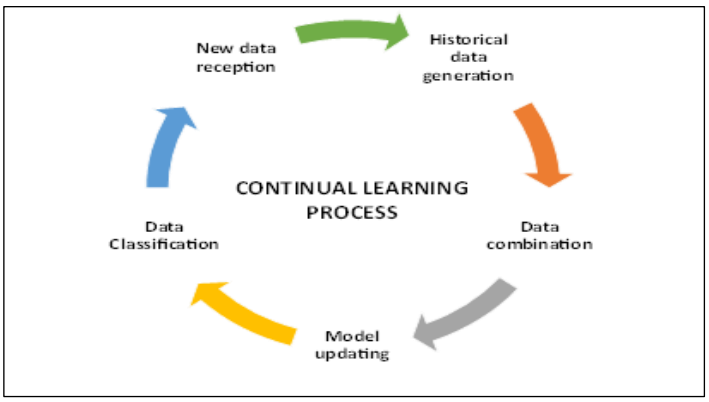


Figure 2: Continual learning process.
Source: Authors, (2025).

III.3.2 FEDERATED LEARNING ASPECT

The federated learning component of the system facilitates collaborative knowledge sharing among multiple nodes without relying on a central server. Each node in the network maintains a vector that stores generative model parameters received from other nodes. Periodically, nodes update their generative models with newly received data and broadcast these updated models to all other nodes in the network. Upon receiving these updates, each node refreshes the corresponding entries in its vector, ensuring it has the latest model parameters from all participating nodes. Using these stored generative models, each node generates synthetic data that represents the data streams observed by other nodes. This synthetic data is then combined with the node's own generated and newly received data to create a diverse and enriched dataset for retraining its discriminative model. By incorporating synthetic data from other nodes, each node effectively simulates and learns from experiences across the network, enhancing its own model's performance. This process, illustrated in Fig. 3, occurs continuously, enabling all nodes to benefit from the collective intelligence of the network while maintaining strict data privacy and security. Through this collaborative approach, the federated learning system leverages diverse insights from all nodes, resulting in improved overall system performance and more generalized models.

III.3.3 NODE'S COMPONENTS

Each node in the proposed system consists of two primary models and a storage vector for the models of other nodes, as illustrated in Fig. 4. The primary models include a Generative Adversarial Network (GAN) [37], which serves as the generative model, and a Deep Stacking Network (DSN) [38], which functions as the discriminative model. Each component is examined individually and described in detail in the following sections.

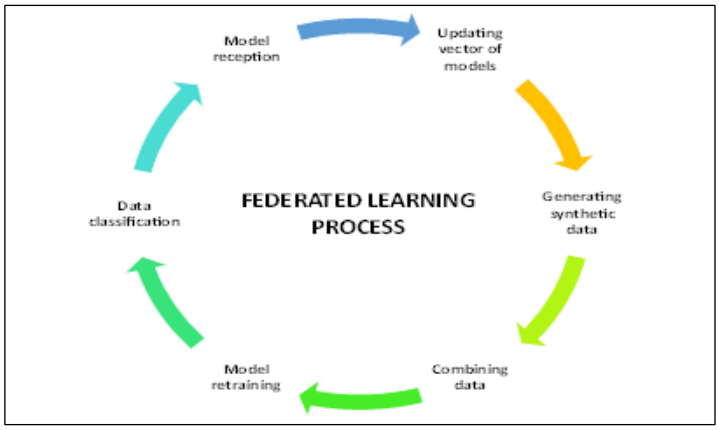


Figure 3: Federated learning aspect.
Source: Authors, (2025).

a) Generative model

In the proposed system, a Generative Adversarial Network (GAN) is employed as the generative model. GANs are a class of neural network architectures designed to generate synthetic data that closely resembles the real data on which they are trained. The GAN comprises two sub-models: the generator and the discriminator. The system is designed to train the GAN on a combination of newly received data and historical data, enabling it to continuously adapt and refine its data generation capabilities by incorporating both current and past information. Newly received data represents the latest batch of information collected by the system, reflecting current trends and patterns, while historical data refers to synthetic data generated by the GAN based on its prior learning from earlier data streams.

Both components of the GAN—the generator and discriminator—are implemented as Multi-Layer Perceptrons (MLPs). During each training iteration, the new data is combined with historical data generated by the GAN itself to form a comprehensive training dataset. The generator is trained to produce synthetic data that not only mirrors the characteristics of the new data but also retains features from historical data. Its objective is to improve its output such that the discriminator cannot differentiate between real data (a combination of new and historical) and synthetic data. Meanwhile, the discriminator is tasked with distinguishing between real and synthetic data, receiving both types during training.

The training process follows an adversarial framework, where the generator and discriminator compete against each other in a feedback loop. As the generator improves its ability to create realistic synthetic data, the discriminator becomes increasingly proficient at identifying fake data. This adversarial dynamic ensures that the GAN continuously enhances its generative capabilities while maintaining alignment with evolving data trends, making it a critical component of the system's continual learning framework.

b) Discriminative model

The discriminative model in the proposed system is implemented as a Deep Stacking Network (DSN), a robust and scalable neural network architecture designed for complex classification tasks. The DSN consists of three identical modules, each implemented as a Multi-Layer Perceptron (MLP) with three layers: an input layer, a hidden layer, and an output layer. The first module receives input that combines local data—including newly received data and historical data generated by the node's own generative model—with external data synthesized by the generative models stored in the vector of models, representing data from other nodes in the federated network. This integration ensures that the model captures both local trends and broader patterns observed across the system. The hidden layer in each module transforms the input data into higher-level features, enabling the extraction of intricate patterns essential for accurate classification. The output layer of each module generates intermediate results, which are passed sequentially to subsequent modules. The

outputs of the first and second modules serve as inputs to the second and third modules, respectively, allowing the DSN to progressively refine its understanding of the data through each stage. The final classification is performed by the third module, ensuring that the model achieves high accuracy by leveraging both local and collaborative knowledge. This hierarchical structure enables the DSN to effectively handle complex classification tasks while maintaining adaptability in dynamic and distributed.

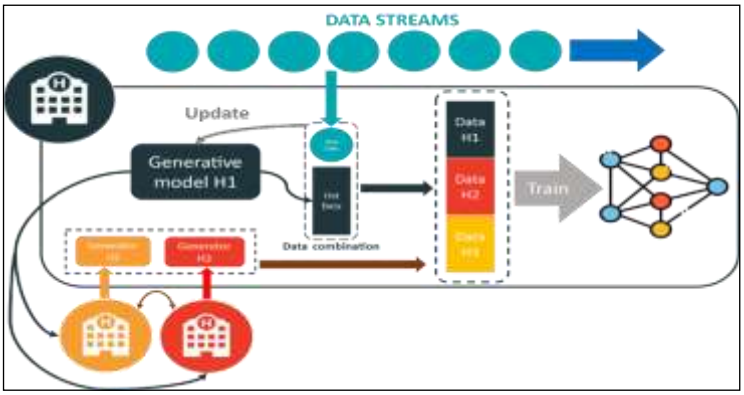


Figure 4: Data path through the different components of a node.
Source: Authors, (2025).

IV. RESULTS AND DISCUSSIONS

This section presents a comprehensive evaluation of the proposed decentralized federated continual learning system through a series of experiments designed to validate its effectiveness and analyze various aspects of its functionality. The evaluation begins with an ablation study to isolate and assess the contributions of individual components. Specifically, we evaluate the performance of the Deep Stacking Network (DSN) in a centralized configuration, examine the streaming process without collaboration in both single and multi-institution settings, and analyze decentralized federated learning without streaming. Following the ablation study, we evaluate the complete system, integrating all components—streaming, and decentralized collaboration.

Beyond validating the system's overall performance, we conduct targeted experiments to explore specific aspects of the approach. These include studying scalability by assessing the effect of adding more institutions on performance, analyzing the impact of synthetic data size on learning efficiency, evaluating knowledge retention to measure how well the system preserves insights from previous data streams, and investigating knowledge diffusion to quantify how effectively learned knowledge propagates across clients. Finally, we compare our approach with several state-of-the-art methods, including FedAvg, FedProx, FedCIL, and Oracle.

IV. 1 CENTRALIZED DSN EVALUATION

In our first experiment, we evaluate the baseline performance of the Deep Stacking Network (DSN) architecture using the complete MIMIC-VI-ED dataset at a single institution. The DSN is configured with three sequential modules, each containing input, hidden, and output layers, where the output of each module serves as input for the subsequent one. We train the model for a total of 3000 epochs across all modules combined, using 80% of the dataset for training while reserving 20% for testing. The performance metrics, expressed in terms of precision and recall as shown in the Fig. 5, establish our reference point for comparing subsequent streaming and federated approaches, demonstrating the model's fundamental classification capabilities when trained on a complete, static dataset before introducing the complexities of streaming data processing and inter-hospital collaboration. The baseline DSN model demonstrates robust performance on the complete dataset, achieving a precision of 0.84 and recall of 0.8, indicating strong classification capabilities in both correctly identifying positive cases and minimizing false negatives.

IV.2 STREAMING PROCESS EVALUATION IN A SINGLE AND MULTI-INSTITUTION SETTING

The second experiment evaluates our system's performance in a streaming environment across two scenarios. In the first scenario, we assess the model's ability to handle sequential data streams at a single institution, where the complete dataset is delivered in 20 consecutive streams. The system processes these streams sequentially using the complete architecture (GAN for synthetic data generation and DSN for classification). It employs continual learning mechanisms to integrate new information while preserving knowledge from previous streams through synthetic data generation.

In the second scenario, we extend the evaluation to multiple institutions by distributing the data across five hospitals with varying data volumes to reflect real-world healthcare settings. The data distribution follows an uneven pattern (30%, 20%, 20%, 15%, 15%) to simulate practical situations where larger institutions, such as major hospitals, have access to more extensive datasets, while smaller health centers possess significantly less data. Each hospital processes its streams independently without any inter-hospital collaboration, allowing us to analyze how data volume disparities affect model performance in a multi-institutional setting. This setup helps understand the system's behavior under realistic healthcare scenarios where data availability varies significantly across institutions.

The results of first scenario demonstrate strong performance, achieving a precision of 0.83 and recall of 0.8. When compared to the baseline DSN experiment, where the model had access to the complete dataset at once, the streaming approach maintains comparable performance levels. This indicates that our system effectively handles sequential data processing without significant degradation in classification capabilities. As shown in the Fig. 6, the model successfully integrates new information from each stream while maintaining its predictive power.

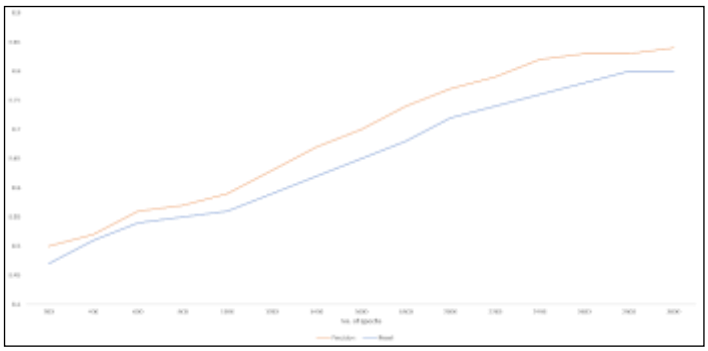


Figure 5: Precision and recall of traditional DSN model on 100% of available data in a single institution configuration.
Source: Authors, (2025).

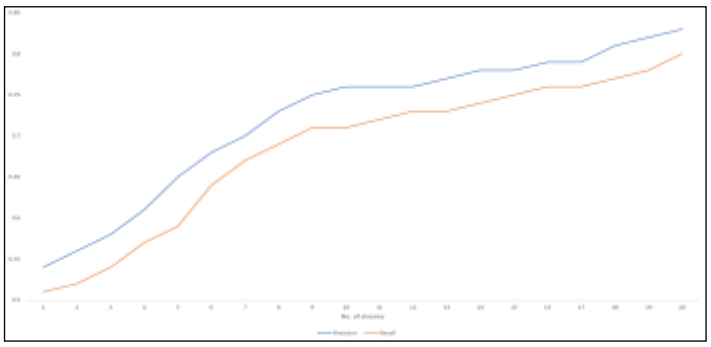


Figure 6: Precision and recall evolution of streaming architecture experiment in a single institution configuration.
Source: Authors, (2025).

In the multi-institutional scenario, performance varies significantly across hospitals of different sizes. The largest hospital (30% data volume) achieves the highest performance with a precision of 0.72 and recall of 0.70. Medium-sized hospitals (20% data volume) show moderate performance with a precision of 0.66 and recall of 0.63, while smaller hospitals (15% data volume) demonstrate lower metrics with a precision of 0.58 and recall of 0.56. These results, when compared to the single-institution scenario (precision: 0.83, recall: 0.8), indicate that data volume significantly impacts model performance. For visualization clarity in the Fig. 7 and Fig. 8, we aggregated results from hospitals with similar data volumes, presenting three representative curves: one for the largest hospital (30%), one averaged curve for medium-sized hospitals (20%), and one for smaller hospitals (15%). This representation clearly illustrates how institutional data volume correlates with model performance in a streaming environment without collaboration.

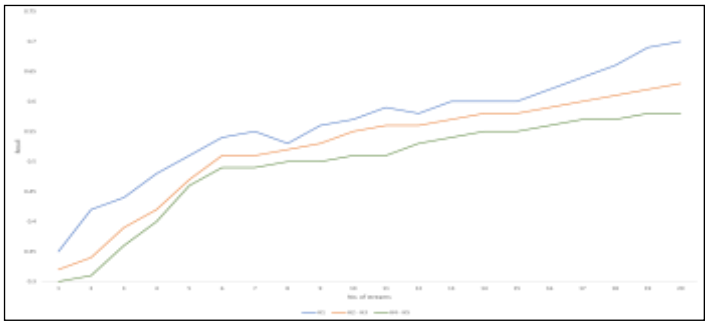


Figure 7: Precision evolution of streaming system architecture in the scenario of five clients without collaboration.
Source: Authors, (2025).

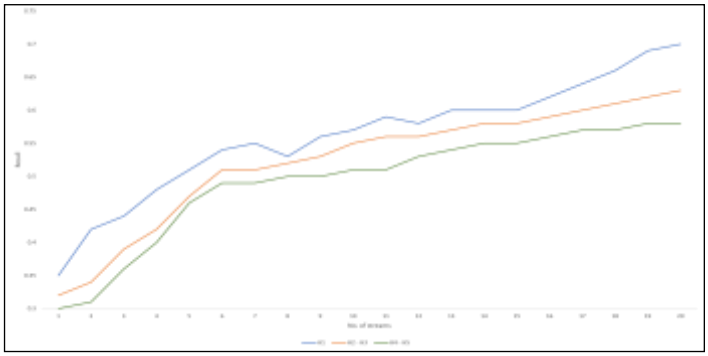


Figure 8: Recall evolution of streaming system architecture in the scenario of five clients without collaboration.
Source: Authors, (2025).

IV.3 DECENTRALIZED FEDERATED LEARNING EVALUATION WITHOUT STREAMING

In the decentralized federated learning configuration, where five hospitals collaborate without streaming data, the results demonstrate significant performance improvements across all institutions regardless of their data volume. The largest hospital (30% data volume) achieves a precision of 0.83 and recall of 0.79, while medium-sized hospitals (20% data volume) show comparable performance with a precision of 0.83 and recall of 0.78. Notably, smaller hospitals (15% data volume) maintain similar performance levels with a precision of 0.82 and recall of 0.78. These results reveal that federated collaboration effectively neutralizes the performance disparities observed in the non-collaborative scenario, enabling smaller institutions to achieve performance metrics nearly identical to larger ones. As shown in the Fig. 9, where results are aggregated for hospitals with similar data volumes, the convergence patterns demonstrate remarkable consistency across institutions of different sizes, highlighting the effectiveness of our decentralized federated learning approach in equalizing performance across participating hospitals.

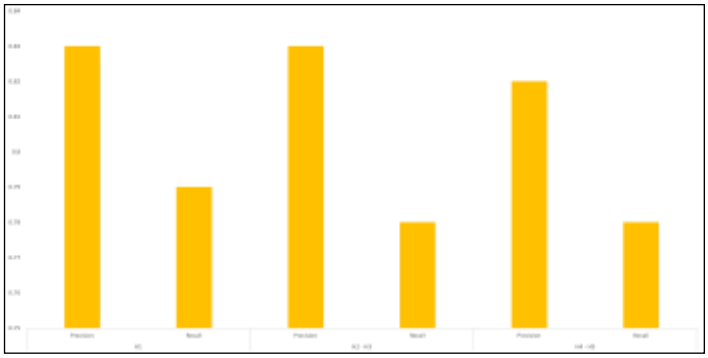


Figure 9: Evaluation results of the decentralized federated architecture without streaming in a configuration of five institutions.
Source: Authors, (2025).

IV.4 PERFORMANCE ANALYSIS OF THE FULLY INTEGRATED SYSTEM

The final experiment evaluates the complete system incorporating both federated learning and streaming capabilities across five hospitals with varying data volumes. The system demonstrates significant evolution in performance across all institutions throughout the streaming process. Initially, all hospitals show relatively low performance metrics, with the largest hospital (30% data volume) starting at precision 0.42 and recall 0.29, medium-sized hospitals (20% data volume) at precision 0.41 and recall 0.27, and smaller hospitals (15% data volume) at precision 0.38 and recall 0.26. As the system processes more streams and hospitals continue to collaborate, performance steadily improves across all institutions, ultimately converging to comparable high levels regardless of institutional size. By the final stream, the largest hospital achieves precision 0.81 and recall 0.78, while medium-sized and smaller hospitals demonstrate similar performance with precision 0.82/0.81 and recall 0.78/0.77 respectively. As illustrated in the Fig. 10 and Fig. 11, where results are aggregated for hospitals with similar data volumes, the combination of federated learning and streaming capabilities enables all institutions to achieve high performance levels, effectively neutralizing the initial performance disparities related to data volume differences.

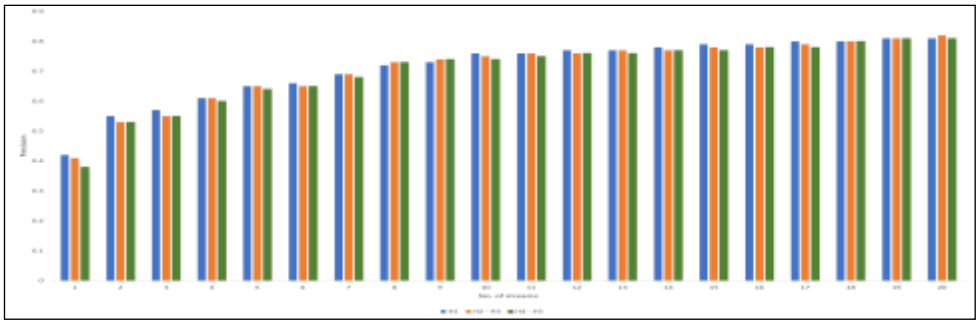


Figure 10: Evolution of precision at each institution in the decentralized federated learning system as data streams arrive
Source: Authors, (2025).

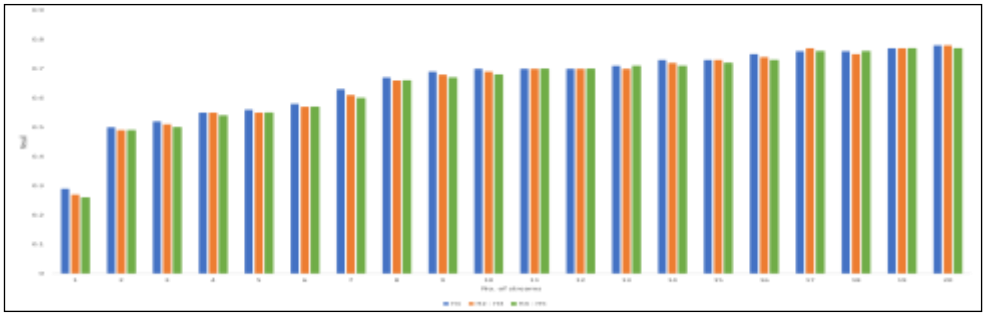


Figure 11: Evolution of recall at each institution in the decentralized federated learning system as data streams arrive.
Source: Authors, (2025).

IV.5 SCALABILITY: EFFECT OF ADDING MORE INSTITUTIONS

To assess the scalability of our decentralized federated learning system, we conducted experiments by progressively increasing the number of client nodes and evaluating precision and recall after processing 20 data streams. The results indicate a slight decrease in system performance as the number of clients increases like shown in Fig. 12. Specifically, the average precision decreased from 0.84 when using a single institution to 0.82 with five clients, 0.81 with ten clients, and 0.8 with twenty clients. Similarly, the average recall followed a comparable trend. This degradation is primarily attributed to the division of the dataset among clients, resulting in smaller data volumes per client as the number of clients increases. This reduction in local data availability negatively impacts the efficiency of each client's model, leading to a gradual decline in overall system performance. Despite this decrease, the system maintains robust performance across different scales, demonstrating its ability to adapt to various multi-institutional settings while preserving data privacy and facilitating collaborative learning.

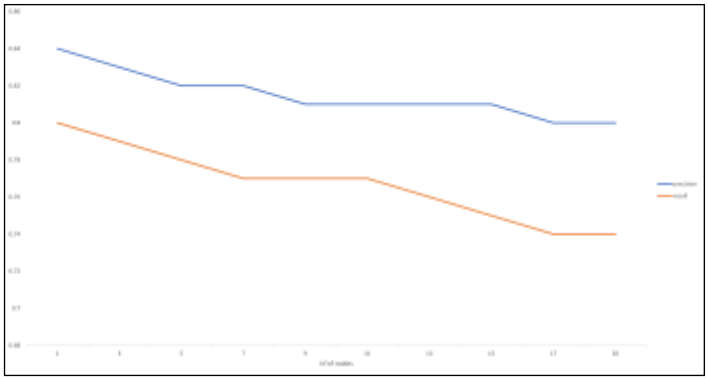


Figure 12: System scalability as the number of nodes increases.
Source: Authors, (2025).

IV.6 SYNTHETIC DATA SIZE: IMPACT ON LEARNING EFFICIENCY

In this subsection, we analyze how the size of synthetic data generated by the system influences the efficiency of the learning process. The experiment evaluates synthetic data proportions ranging from 0% to 75% of the original data size. At 0%, no synthetic data is generated, which means that neither collaboration nor incremental processing is possible. Consequently, the system's performance is significantly limited, achieving only a precision of 0.55 and recall of 0.49. As synthetic data generation increases, the system begins to leverage collaborative and incremental learning, resulting in improved performance like illustrated in Fig. 13. The optimal results are observed when synthetic data constitutes between 45% and 60% of the total training data, where precision and recall reach their peak values. Beyond this range, however, performance starts to decline, likely due to over-reliance on synthetic data, which may reduce diversity or introduce noise into the learning process.

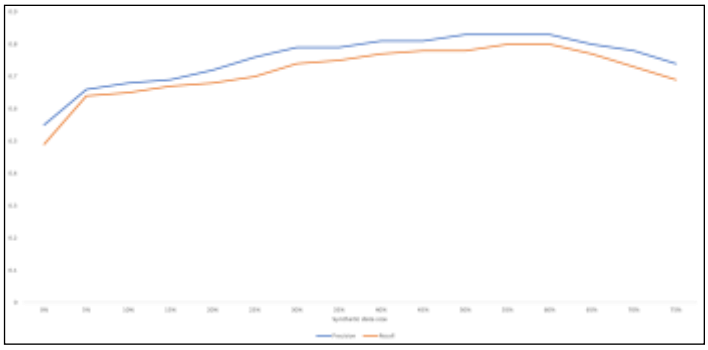


Figure 13: Evolution of precision and recall of the system as sythentic data size changes
Source: Authors, (2025).

IV.7 KNOWLEDGE RETENTION: ABILITY TO PRESERVE INSIGHTS FROM PREVIOUS STREAMS

The objective of this experiment is to evaluate the system's ability to retain knowledge from previous data streams while training on newly received data batches. This experiment is motivated by the challenge of catastrophic forgetting, a common issue in continual learning where models lose previously acquired insights when adapting to new information. The experiment is conducted in two phases: first, a pre-update evaluation measures the model's performance on previous streams before training on the new batch; second, a post-update evaluation assesses performance after integrating the new batch, quantifying knowledge retention and identifying potential degradation. Both evaluations (pre and post update evaluations) used a test dataset derived exclusively from previous streams, excluding the newly received data. The experiment was conducted at multiple time points corresponding to the arrival of streams 2, 5, 10, 15, and 20, with tests performed on data generated from older streams. The results, illustrated by Fig. 14 and Fig 15, showed that while there is a slight decline in precision and recall metrics post-update compared to pre-update, the differences were minimal (mean differences of 0.018 for precision and 0.011 for recall), with low variability (standard deviations of 0.0084 for precision and 0.0055 for recall).

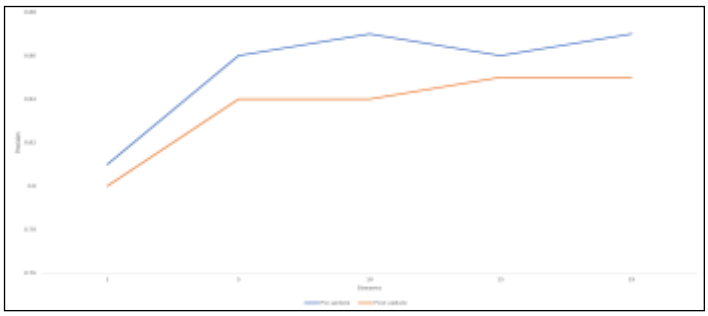


Figure 14: Pre and post update precision evolution over time during the knowledge retention experiment.
Source: Authors, (2025).

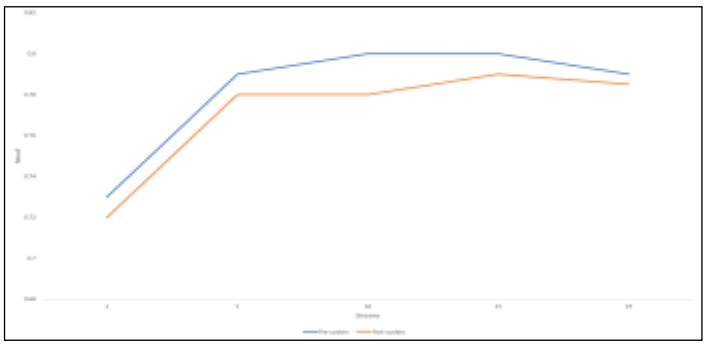


Figure 15: Pre and post update recall evolution over time during the knowledge retention experiment.
Source: Authors, (2025).

IV.8 KNOWLEDGE DIFFUSION: PROPAGATION OF KNOWLEDGE ACROSS INSTITUTIONS

This experiment aims to evaluate the extent of knowledge diffusion among clients in the decentralized federated learning system. Specifically, it investigates whether the knowledge learned by a client on its local data propagates effectively to other clients, enabling them to perform well on data streams they have not directly observed. To study knowledge diffusion in the decentralized federated learning system, we evaluate each client's ability to perform on test datasets derived exclusively from other clients' local data. After training, each client's discriminative model is tested on unseen data streams from other clients, and the results are compared against those obtained from exclusive local training (without collaboration).

The evaluation is conducted using average precision and recall as metrics to quantify the effectiveness of knowledge transfer across clients. The results of the knowledge diffusion experiment in Fig. 16 demonstrate the effectiveness of collaborative learning in improving client performance on unseen data. Under collaborative training, the average precision and recall across clients reached 0.82 and 0.79, respectively, compared to 0.77 and 0.74 in the local training scenario without collaboration. This significant improvement highlights the ability of the decentralized federated learning system to propagate knowledge between clients effectively, enabling each node to benefit from insights learned by others.

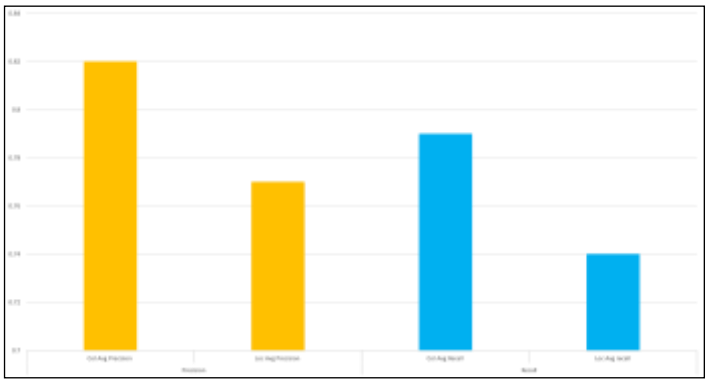


Figure 16: Comparison of collaborative and local average performances in the knowledge diffusion experiment.
Source: Authors, (2025).

IV.9 COMPARATIVE ANALYSIS

In this subsection, we compare the performance of our proposed decentralized federated continual learning approach with several state-of-the-art methods, including FedAvg [7], FedProx [39], FedCIL [40], [41], and Oracle. FedAvg and FedProx represent two widely used aggregation techniques in federated learning, serving as benchmarks for collaborative model training. FedCIL employs GANs to generate samples from previous tasks. Finally, Oracle serves as an upper bound for comparison, assuming clients have access to all their historical data during training.

The results in Table. I. demonstrate that our approach achieves comparable performance to FedCIL, with both methods showing high precision and recall. However, our decentralized framework eliminates the reliance on a central server, addressing critical limitations such as single points of failure and bottlenecks in communication. While Oracle, as an upper bound, outperforms all methods due to its access to all historical data, our method closely approaches its performance while preserving privacy and ensuring scalability. In contrast, FedAvg and FedProx show significantly lower performance, highlighting their limitations in handling continual learning tasks.

Table 1: Comparative results of our spproach with FedAvg, FedProx, FedCIL and oracle

Algorithm	Avg Precision	Avg Recall
FedAVG	0.7	0.67
FedProx	0.74	0.72
FedCIL	0.82	0.78
Ours	0.81	0.78
Oracl	0.85	0.81

Source: Authors, (2025).

V. CONCLUSIONS

In this paper, we addressed the critical challenges facing healthcare institutions in implementing machine learning solutions: data privacy concerns, continuous data streams processing, and effective collaboration across institutions of varying sizes. We proposed a novel decentralized federated continual learning system that combines GANs for synthetic data generation and DSNs for classification, enabling privacy-preserving collaboration without central coordination. Experimental results demonstrate the system's effectiveness across diverse scenarios, including ablation studies, scalability analysis, and comparisons with state-of-the-art approaches. The results are comparable with state-of-the-art centralized approaches, and while the system may not significantly improve classification performance beyond existing methods, it provides crucial advantages in terms of data privacy preservation, institutional collaboration, and dynamic data processing capabilities. Furthermore, the analysis of knowledge retention and diffusion highlights the system's ability to preserve insights from historical data and propagate knowledge across institutions, ensuring equitable performance even for smaller healthcare centers. These characteristics make our approach particularly suitable for real-world healthcare applications where privacy, continuous learning, and inter-institutional collaboration are essential requirements.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Rostom Mennour.
Methodology: Rostom Mennour.
Investigation: Rostom Mennour.
Discussion of results: Rostom Mennour.
Writing – Original Draft: Rostom Mennour.
Writing – Review and Editing: Rostom Mennour.
Resources: Rostom Mennour.
Supervision: Rostom Mennour.
Approval of the final text: Rostom Mennour.

VII. REFERENCES

[1] E. Aslan, "Development of malaria diagnosis with convolutional neural network architectures: a CNN-based software for accurate cell image analysis". ITEGAM-JETIA, vol. 11, n°51, pp 35-42, 2025.

[2] A. Jenefa, E. Veemaraj, & A. Lincy, "ABM-OCD: Advancing ovarian cancer diagnosis with attention-based models and 3D CNNs", ITEGAM-JETIA, vol. 9, n° 43, pp. 23-33, 2023.

[3] Y. Kumar, A. Koul, R. Singla, & M. F. Ijaz, M. F, "Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda". Journal of ambient intelligence and humanized
computing, vol. 14, n° 7, pp. 8459-8486, 2023.

[4] M. S. Hashim, and A. A.Yassin, «Breast Cancer Prediction Using Soft Voting Classifier Based on Machine Learning Models», IAENG International Journal of Computer Science, vol. 50, n° 2, 2023.

[5] F. AmrollahiI, S. P. Shashikumar, A. L. Holder et S. Nemati, «Leveraging clinical data across healthcare institutions for continual learning of predictive risk models» Scientific Reports, vol. 12, n° 11, p. 8380, 2022.

[6] Q. An, S. Rahman, J. Zhou et J. J. Kang, «A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges.,» Sensors, vol. 23, n° 19, p. 4178, 2023.

[7] M. Javaid, A. Haleem, R. P. Singh, R. Suman et S. Rab, «Significance of machine learning in healthcare: Features, pillars and applications,» International Journal of Intelligent Networks, vol. 3, pp. 58-73, 2022.

[8] B. McMahan, E. Moore, D. Ramage, S. Hampson et B. A. y Arcas, «Communication-efficient learning of deep networks from decentralized data,» chez Artificial intelligence and statistics, PMLR, 2017, pp. 1273-1282.

[9] H. A, P. K, M. R et A. S, «Jointly learning from decentralized (federated) and centralized data to mitigate distribution shift,» chez Proceedings of NeurIPS Workshop on Distribution Shifts, 2021.

[10] M. De Lange, R. Aljundi, M. Masana, S. Parisot, X. Jia, A. Leonardis, G. Slabaugh et T. Tuytelaars, «A continual learning survey: Defying forgetting in classification tasks,» IEEE transactions on pattern analysis and machine intelligence, vol. 44, n° 17, pp. 3366-3385., 2021.

[11] M. D. B. Thilagaraj, V. Pandimurugan, P. Naveen, M. S. Hema, S. Hariharasitaraman et P. Govindan, « A novel intelligent hybrid optimized analytics and streaming engine for medical big data,» Computational and Mathematical Methods in Medicine, vol. 2022, 2022.

[12] X. Yang, H. Yu, X. Gao, H. Wang, J. Zhang et T. Li, «Federated Continual Learning via Knowledge Fusion: A Survey,» IEEE Transactions on Knowledge and Data Engineering, vol. 1, pp. 1-20, 2024.

[13] M. A. Mohedano-Munoz, C. Soguero-Ruiz, I. Mora-Jiménez, M. Rubio-Sánchez, J. Álvarez-Rodríguez et A. Sanchez, «A streaming data visualization framework for supporting decision-making in the Intensive Care Unit,» Expert Systems with Applications, vol. 227, p. 120252, 2023.

[14] Singh, M. B. Gurbuz, S. S. Gantha et P. Jasti, «Class-Incremental Continual Learning for General Purpose Healthcare Models.,» arXiv preprint arXiv, vol. 2311, n° 104301, 2023.

[15] M. Tilala, S. Pamulaparthyvenkata, A. D. Chawda, and A. P. Benke, « Explore the Technologies and Architectures Enabling Real-Time Data Processing within Healthcare Data Lakes, and How They Facilitate Immediate Clinical Decision-Making and Patient Care Interventions», European Chemical Bulletin, vol. 11, pp. 4537-4542, 2023.

[16] Z. L. Teo, L. Jin, S. Li, D. Miao, X. Zhang, W. Y. Ng et D. S. W. Ting, «Federated machine learning in healthcare: A systematic review on clinical applications and technical architecture,» Cell Reports Medicine, 2024.

[17] Jiménez-Sánchez, M. Tardy, M. A. G. Ballester, D. Mateus et D. Piella, «Memory-aware curriculum federated learning for breast cancer classification,» Computer Methods and Programs in Biomedicine, vol. 229, p. 107318, 2023.

[18] M. Feng, Y. W. S. Yan, Y. Xu, L. Shao et H. Fu, «Specificity-preserving federated learning for MR image reconstruction,» IEEE Transactions on Medical Imaging, vol. 42, n° 17, pp. 2010-2021, 2022.

[19] Liu, Cabezas, M, D. Wang, Z. Tang, L. Bai, G. Zhan, L. Y, K. Kyle, L. Ly, J. Yu, C. Shieh, A. Nguyen, K. E. Kandasamy, R. Sullivan, F. Calamante, M. Barnett, W. Ouyang, W. Cai et C. Wang, «Multiple sclerosis lesion segmentation: revisiting weighting mechanisms for federated learning,» Frontiers in neuroscience, vol. 17, n° 11167612, 2023.

[20] H. Kassem, D. Alapatt, P. Mascagni, A. Karargyris et N. Padoy, «Federated Cycling (FedCy): Semi-Supervised Federated Learning of Surgical Phases,» IEEE Transactions on Medical Imaging, vol. 42, n° 17, pp. 1920-1931, 2023.

[21] S. Pati, U. Baid, B. Edwards, M. Sheller, S. H. Wang, G. A. Reina et L. Poisson, «Federated learning enables big data for rare cancer boundary detection,» Nature communications, vol. 13, n° 11, p. 7346, 2022.

[22] Qayyum, K. Ahmad, M. A. Ahsan, A. Al-Fuqaha et J. Qadir, «Collaborative federated learning for healthcare: Multi-modal covid-19 diagnosis at the edge,» IEEE Open Journal of the Computer Society, vol. 3, pp. 172-184, 2022.

[23] L. Mondrejevski, Miliou, M. A. D. Pitts, J. Hollmén et P. Papapetrou, «Flicu: A federated learning workflow for intensive care unit mortality prediction,» chez 2022 IEEE 35th International Symposium on Computer-Based Medical Systems (CBMS), Shenzhen, China, 2022.

[24] T. K. Dang, X. Lan, J. Weng et M. Feng, «Federated learning for electronic health records,» ACM Transactions on Intelligent Systems and Technology (TIST), vol. 13, n° 15, pp. 1-17, 2022.

[25] U. P. S. K. T. M. G. R. B. E. A. S. .. &. B. Baid, «Federated learning for the classification of tumor infiltrating lymphocytes,» arXiv preprint arXiv, 2022.

[26] L. Zhao et J. Huang, «A distribution information sharing federated learning approach for medical image data,» Complex & Intelligent Systems, vol. 9, n° 15, pp. 5625-5636, 2023.

[27] M. Chetoui et M. Akhloufi, «Peer-to-Peer Federated Learning for COVID-19 Detection Using Transformers,» Computers, vol. 12, n° 15, p. 106, 2023.

[28] W. Qiu, W. Ai, H. Chen, Q. Feng et G. Tang, «Decentralized federated learning for Industrial IoT with deep echo state networks,» IEEE Transactions on Industrial Informatics, vol. 19, n° 14, pp. 5849-5857, 2002.

[29] T. V. Nguyen, M. A. Dakka, S. M. Diakiw, M. D. VerMilyea, M. Perugini, J. M. M. Hall et D. Perugini, «A novel decentralized federated learning approach to train on globally distributed, poor quality, and protected private medical
data,» Scientific Reports, vol. 12, n° 11, p. 8888, 2022.

[30] Z. Lian, Q. Yang, W. Wang, Q. Zeng, M. Alazab, H. Zhao et C. Su, «DEEP-FEL: Decentralized, efficient and privacy-enhanced federated edge learning for healthcare cyber physical systems,» IEEE Transactions on Network Science and Engineering, vol. 9, n° 15, pp. 3558-3569, 2022.

[31] Li, H. Li et G. Yuan, «Continual Learning with Deep Neural Networks in Physiological Signal Data: A Survey,» Healthcare, vol. 12, n° 12, p. 155, 2024.

[32] E. Batbaatar, K. H. Park, T. Amarbayasgalan, K. Davagdorj, L. Munkhdalai, V. H. Pham et K. H. Ryu, «Class-incremental learning with deep generative feature replay for DNA methylation-based cancer classification,» IEEE Access, vol. 8, pp. 210800-210815, 2020.

[33] C. González, A. Ranem, D. Pinto dos Santos, A. Othman et A. & Mukhopadhyay, «Lifelong nnU-Net: a framework for standardized medical continual learning,» Scientific Reports, vol. 13, n° 11, p. 9381, 2023.

[34] P. Kumari, J. Chauhan, A. Bozorgpour, R. Azad et D. Merhof, «Continual Learning in Medical Imaging Analysis: A Comprehensive Review of Recent Advancements and Future Prospects,» arXiv preprint arXiv, vol. 2312, n° 117004, 2023.

[35] Y. Huang, C. Bert, S. Fischer, M. Schmidt, A. Dörfler, A. Maier et F. Putz, «Continual learning for peer-to-peer federated learning: A study on automated brain metastasis identification,» arXiv preprint arXiv:2204.13591, 2022.

[36] Johnson, L. Bulgarelli, T. Pollard, L. A. Celi, R. Mark et S. Horng, «MIMIC-IV-ED (version 2.2),» PhysioNet, 2023.

[37] F. Xie, J. Zhou, J. W. Lee, M. Tan, S. Li, L. S. O. Rajnthern et N. Liu, «Benchmarking emergency department prediction models with machine learning and public electronic health records,» Scientific Data, vol. 9, n° 11, p. 658, 2022.

[38] Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu et D. Warde-Farley, «Generative adversarial networks,» chez Advances in neural information processing systems, Montreal, Quebec, Canada, 2014.

[39] L. Deng, X. He et J. Gao, «Deep stacking networks for information retrieval,» chez IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, Canada, 2013.

[40] T. Li, A. K. Sahu, M. Zaheer, M. Sanjabi, A. Talwalkar, & V. Smith, «Federated optimization in heterogeneous networks,» in Proceedings of Machine learning and systems, vol. 2, p. 429-450, 2020.

[41] D. Qi, H. Zhao, and S. Li, «Better generative replay for continual federated learning,» arXiv preprint arXiv:2302.13001, 2023.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 174-182. September/October, 2025.
DOI: https://doi.org/10.5935/jetia. v.11 n.55.2149.

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

ASSESSMENT AND OPTIMIZATION OF ENGINE PERFORMANCE AND EMISSIONS USING ADDITIVE-INFUSED CORN OIL BIODIESEL

Raviteja Surakasi[1], Makka Jayakrishna[2], Yenda Srinivasa Rao[3], Iván Leandro Rodríguez Rico[4]

[1]Department of Mechanical Engineering, Lendi Institute of Engineering and Technology, Jonnada, Vizianagaram, Andhra Pradesh, India
[2,3]Department of Mechanical Engineering, Sri Sivani College of Engineering, Srikakulam-532410, Andhra Pradesh, India.
[4]Central University "Marta Abreu" of Las Villas, Cuba

[1]https://orcid.org/0000-0002-0786-0105, [2]https://orcid.org/0000-0003-1647-2750
[3]https://orcid.org/0000-0002-3010-720X, [4]https://orcid.org/0000-0003-1295-5368

Email: ravitejasurakasi@gmail.com , jayakrishnamakka555@gmail.com, yenda.srinivas@gmail.com, ivanl@uclv.edu.cu

ARTICLE INFO

Article History
Received: June 2, 2025
Revised: June 30, 2025
Accepted: October 2, 2025
Published: October 31, 2025

Keywords:
Diesel Engine,
Corn Oil Biodiesel,
Transesterification,
Response Surface Methodology (RSM),
Fuel Mixture Optimization,
Nanoparticle Additives.

ABSTRACT

Biodiesel was made in this experiment by transesterifying corn oil. By varying the proportions of fuel to air, the efficiency of the engine and the pollution caused by fossil fuels was measured. In the course of this study, there was an investigation of the impacts of hydrocarbons, carbon monoxide, thermal efficiency, mean effective pressure and braking power. A diesel engine powered the vehicle, producing 5.2 kilowatts. There is a possibility of producing biodiesel by using corn oil as a waste product. We combined 80% diesel and 20% corn oil to produce biodiesel. According to the ASTM criteria for fuel quality, the corn oil methyl ester was suitable. The experiment results were obtained by operating a diesel engine with a single cylinder and four strokes under a variety of load conditions. When compared to B20 alone, a mixture containing 100 parts per million of titanium oxide nanoparticles increased thermal efficiency, mean effective pressure, and braking power. The TiO2 nanoparticles included in B20 reduce the engine's emissions of hydrocarbons and carbon monoxide. We determined the optimal values for brake power, brake mean effective pressure, brake thermal efficiency, carbon monoxide, and hydrocarbons through optimization using RSM (Response Surface Methodology). These values are as follows: The values are 11 463 kg and 56.242 parts per million of titanium dioxide. 4.223 kilowatts of brake power, 5.078 bars of mean effective pressure, 29.573 percent thermal efficiency, 0.257 percent carbon monoxide, and 52.246 percent hydrocarbons are the optimal numbers.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

In response to rising concerns about pollution and its effects on human health, biodiesel has gained popularity as a viable fossil fuel substitute for ICEs, stoves, and other heating applications [1]. In many countries, policies encouraging the domestic production and use of first-generation biodiesel have fueled its growing popularity. Some regions specifically promote oil palm-based biodiesel production to meet energy demands for power generation [2]. The manufacture of first-generation biofuels requires a lot of water, which could jeopardize food security because it would divert water from crops grown for human consumption. This has sparked debate about the sustainability of using food crops for energy, highlighting the need for alternative strategies [3]. This necessitates careful consideration during the legislative process in different countries. Extensive research has been conducted globally to assess the efficiency of internal combustion engines [4]. There has been a plethora of research on the qualities and functionality of biodiesel, which is produced from vegetable oils that are not suitable for human consumption. Tobacco, flax, jatropha, cotton, jojoba, and rubber tree oils are among those that have been considered for use as raw ingredients. To satisfy fuel demands for power generation, several countries have prioritized oil palm development for biodiesel production [5]. Biodiesel made from corn oil meets international environmental criteria, even though it

has the same problems as first-generation biofuels [6]. In the best conditions, researchers used a combined method with sodium methoxide and a special membrane to get a transesterification conversion rate of 98.1%. Research into biodiesel's potential as a replacement fuel for internal combustion engines has been on the rise recently, with the hope of making transportation engines more efficient [7].

II. MATERIALS AND METHOD

II.1. PRODUCTION OF BIODIESEL

One of the first steps in making biodiesel is boiling corn, which usually results in about 10 liters of oil (FAO). The presence of free fatty acids is indicated by the acid number, which is assessed by analyzing a sample of the oil for acidity. If the acid number of the oil is 5 or below, it is considered appropriate for biodiesel manufacture when tested using potassium hydroxide in a titration procedure. Oils with pH levels higher than 5 are not suitable for use as biodiesel; however, they can have other applications. The acid number is found by heating 5 grams of oil with 50 milliliters of neutral ethyl alcohol to about 50 degrees Celsius, adding phenolphthalein, and then titrating with 0.1N potassium hydroxide. The transesterification mechanism is shown in Figure 1 [9].

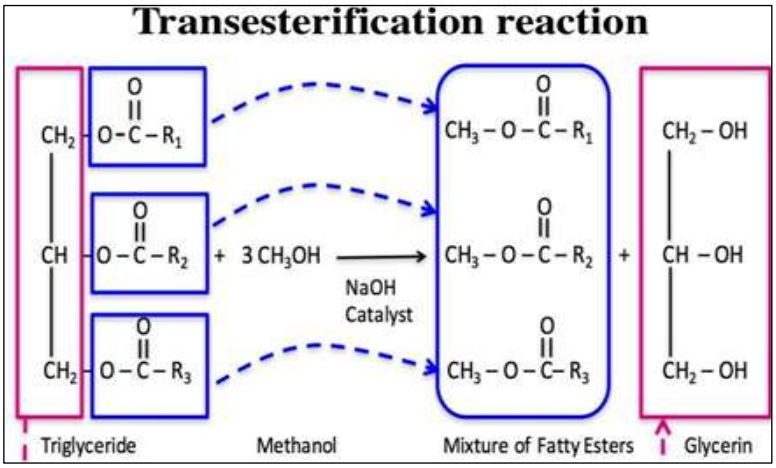


Figure 1: Process of Transesterification.
Source: Authors, (2025).

One way to make biodiesel is by transesterifying renewable lipids, such as animal and plant oils, into long-chain monoalkyl esters of fatty acids. The researchers in this investigation mixed ten litres of discarded cooking oil with twenty percent v/v methanol. Eight grams of potassium hydroxide were dissolved in methanol and used as a catalyst to produce methoxide, according to the acid number of the oil. This mixture was stirred with the oil and heated to between 50 and 60 degrees Celsius for about two hours. Following the reaction, the glycerin was extracted by decantation. The biodiesel was then rinsed three to four times with water with a 20–30% (v/v) concentration, after which it was dried and filtered to get a clean, usable fuel.

II.2. FUEL SAMPLES

The B20 corn oil methyl ester mixture was prepared for further testing by completely mixing 20% biodiesel and 80% diesel by volume using a magnetic stirrer. The end product was created by the previously outlined procedure.

Table 1: Fuel blends properties.

	Diesel	**BD20**	**BD20+50ppm** TiO_2	**BD20+100ppm** TiO_2
Density(kg/m3)	815	820	833	846
Viscosity (mm2/s)	3.1	3.2	3.4	3.6
Calorific Value (MJ/kg)	44	43.4	44.8	46
Cetane Number	52	56.6	57.5	58.4
Flash Point °C	204	195.2	196.4	198.2
Fire Point °C	59	129.8	132.1	133.8

Source: Authors, (2025).

Following the aforementioned procedure, B20 fuel test samples were ultrasonically mixed with titanium oxide (TiO2) nanoparticles. Fifty and one hundred parts per million (ppm) were the dose levels used. The fuels in testing were assessed utilizing the ASTM D 6751 evaluation procedure, an internationally recognized method for the analysis of fuel properties. Table 1 shows the fuel blend properties

III. EXPERIMENTAL SET-UP

The experiments used a Kirloskar TV1, a water-cooled CI engine with one cylinder, four strokes, and overhead valves driven by pushrods. The engine can produce up to 5.2 kW of power when revved to 1500 rpm. Following the manufacturer's instructions, the fuel injection pressure was set at 210 psi and the injection timing at 23° BTDC.

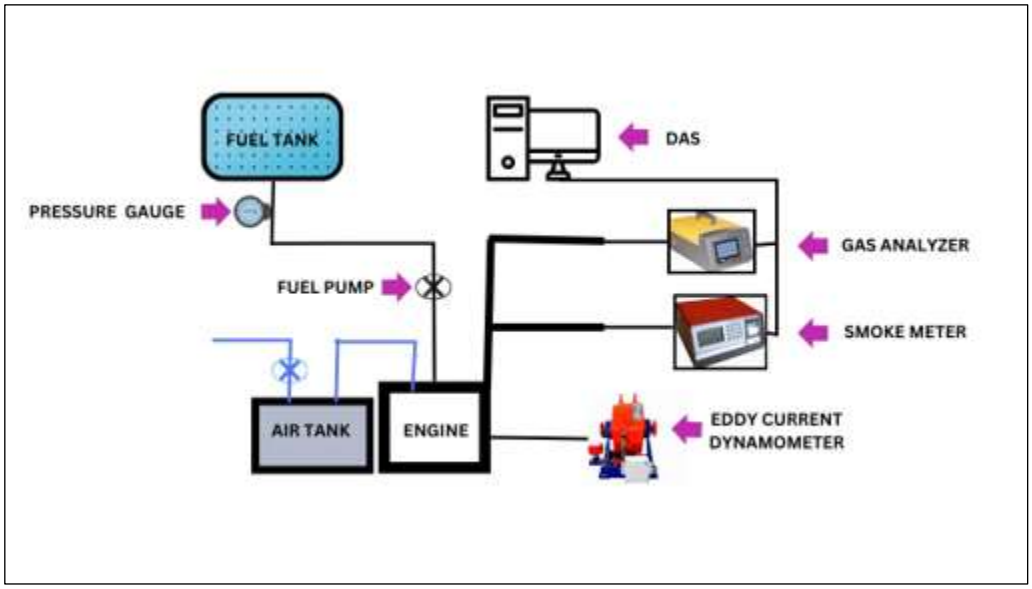


Figure 2: Experimental set-up.
Source: Authors, (2025).

To maintain a coolant temperature of 80 °C, an internal circulation system encircling the cylinder with water jackets was used. A piezoelectric sensor fitted flush-mounted to the engine block measured in-cylinder pressure, while an eddy current dynamometer recorded engine torque. Figure 2 shows a simplified schematic of the equipment used in the experiment.

IV. RESULTS AND DISCUSSIONS:

IV.I. BRAKE POWER

Engine brake power grows roughly linearly with load, as shown in Figure 3, with very little variance across various fuel mixes. Brake power output appears to be unaffected by fuel type, according to the data [12]. At higher loads, the BD20 + 100 ppm TiO_2 combination produced the most effective brake power.

IV.II. BRAKE MEAN EFFECTIVE PRESSURE:

Figure 4 illustrates that the brake mean effective pressure (BMEP) increases almost linearly with rising load, regardless of the fuel blend used [13]. The data confirm that fuel type has minimal effect on BMEP [14]. The highest BMEP was recorded for the BD20 + 100 ppm TiO_2 blend at maximum load.

IV.III. BRAKE THERMAL EFFICIENCY:

Figure 5 shows that the brake thermal efficiency (BTE) increases almost linearly with rising load, remaining generally consistent across all fuel blends [15]. The results indicate no significant correlation between fuel type and brake power output [16]. The highest BTE was observed for the BD20 + 100 ppm TiO_2 blend at maximum load.

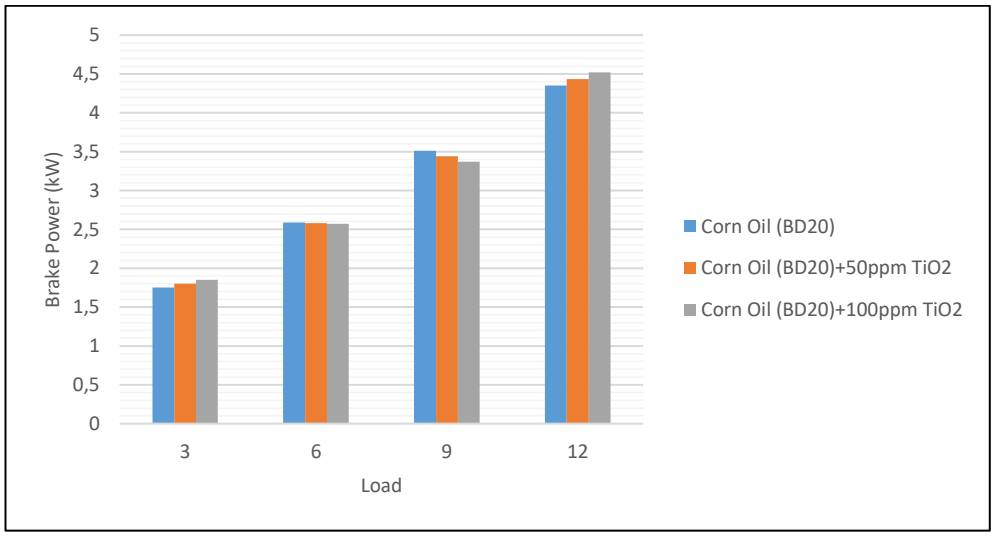


Figure 3: Brake power increases proportionally with increasing engine load.
Source: Authors, (2025).

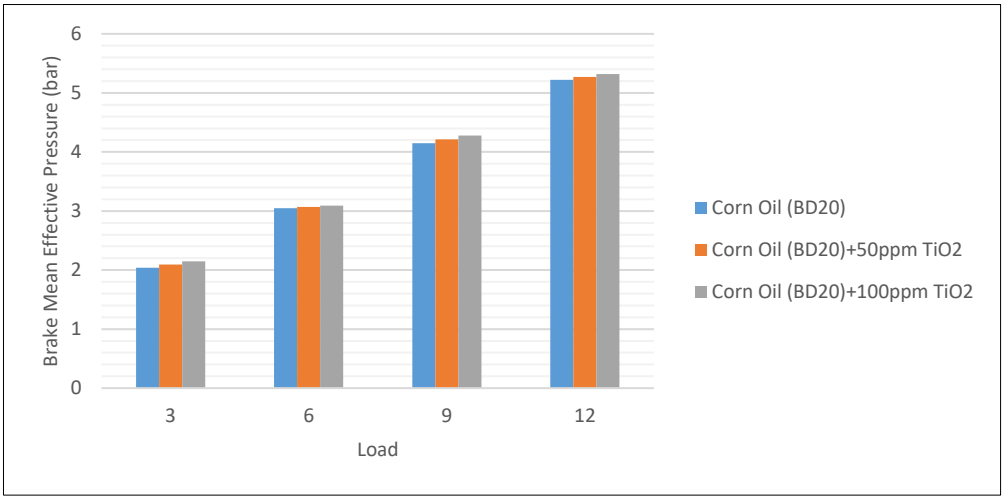


Figure 4: Brake Mean Effective Pressure increases proportionally with increasing engine load.
Source: Authors, (2025).

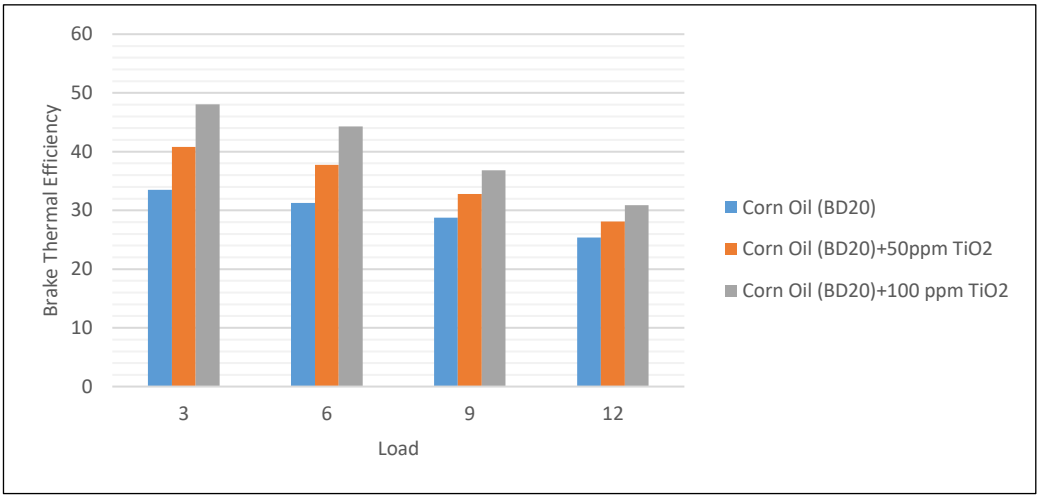


Figure 5: Brake Thermal Efficiency increases proportionally with increasing engine load.
Source: Authors, (2025).

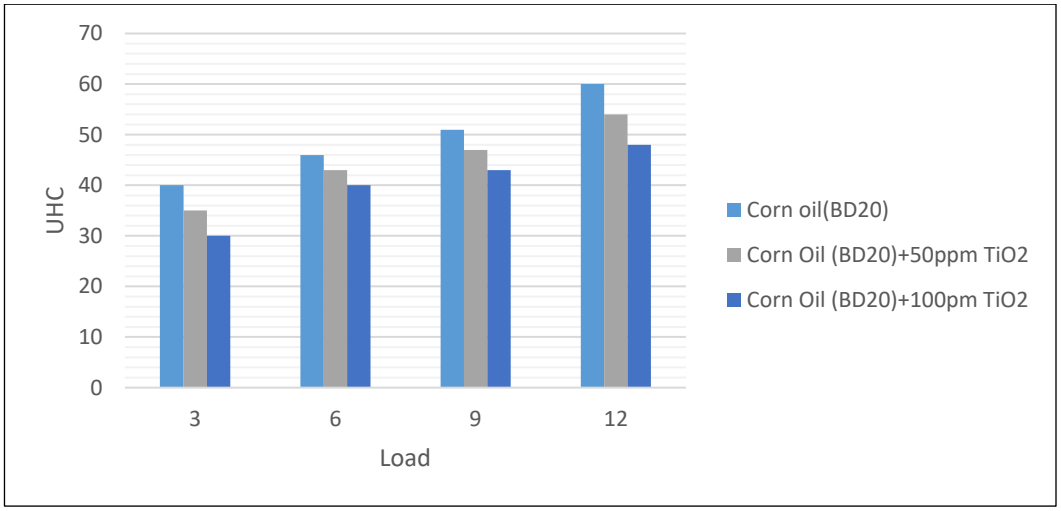


Figure 6: HC increases proportionally with increasing engine load.
Source: Authors, (2025).

IV.4. HC EMISSION TEST

Hydrocarbon (HC) emission levels rise with increasing load, as seen in Figure 6, and they approach maximum power as shown in [17]. Based on the theoretical framework, when oxygen is scarce, the concentration of hydrocarbon (HC) emissions rises. According to the fuel-air mixture ratio, this means the fuel is richer. Figure 4 demonstrates that HC emissions are consistently lower across the operational range for all mixtures. The combination of B20 and 100 ppm TiO2 demonstrates lower emissions than the B20 mixture alone [18].

IV.5. CO EMISSION TEST

Figure 7 shows a positive correlation between engine load and CO emissions, with CO levels increasing as load rises. However, the use of corn biodiesel blends resulted in a notable 20–25% reduction in total CO emissions [19]. Among the tested fuels, B20 exhibited the highest CO emissions under ambient conditions [20]. The addition of 100 ppm TiO_2 to the B20 blend significantly reduced CO emissions, demonstrating its potential for lowering greenhouse gas output.

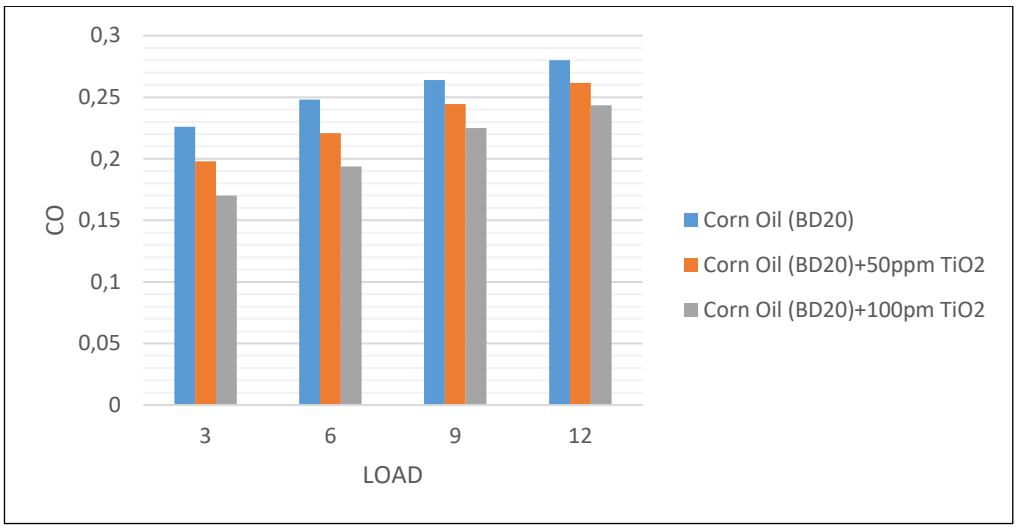


Figure 7: CO increases proportionally with increasing engine load.
Source: Authors, (2025).

IV.6. OPTIMIZATION OF DATA USING RSM:

IV.7. I. RSM:

Response Surface Methodology (RSM) employs statistical and mathematical techniques to develop empirical models, where a limited number of input factors are used to evaluate and optimize the response variable. This research assessed the effectiveness of various models using RSM. We selected and analyzed the optimal model to achieve the highest possible values.

IV.VI. II. OPTIMUM VALUES OF TORQUE, UHC, AND CO_2:

The optimum values were found to be at a load of 11.463 kgs and 56.242 ppm concentration of TiO_2 the optimum values of Brake Power, Brake mean effective pressure, Brake thermal Efficiency, Carbon Monoxide, and Hydrocarbons are found to be 4.223 kW, 5.078 bar, 29.573%, 0.257 and 52.246 respectively [21].

As the engine load goes up, more TiO_2 nanopowder leads to better braking power, higher mean effective pressure, improved thermal efficiency, and reduced emissions of CO and HC. Figure 8 (a-e) uses 3D surface plots to illustrate the change in these parameters concerning load and TiO_2 concentration [22] The optimal input parameters, including engine load and TiO_2 concentration, resulted in the best response values for brake power, BMEP, thermal efficiency, and reduced CO and HC emissions are shown in figure 9. The suggested deal has a satisfaction rating of close to 100%. The situation is dependable, as the solution is acceptable and desirable [23].

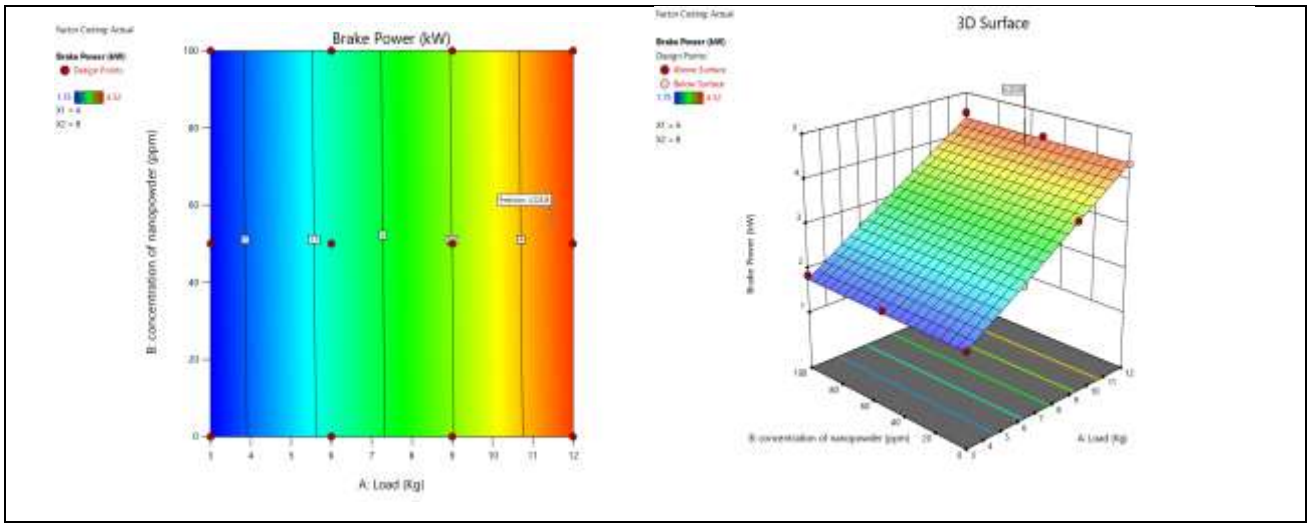


Figure 8a: Surface plot and 3D plot for the chosen model's brake power is based on the load and concentration of nanoparticles.
Source: Authors, (2025).

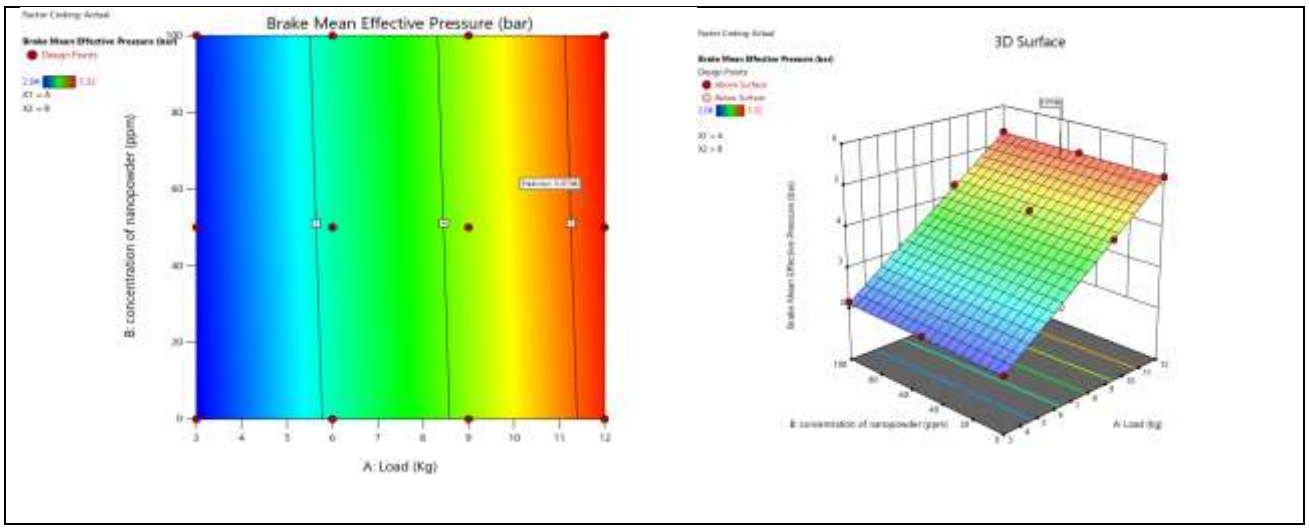


Figure 8b: Surface plot and 3D plot for the chosen model's Brake Mean Effective Pressure is based on the load and concentration of nanoparticles.
Source: Authors, (2025).

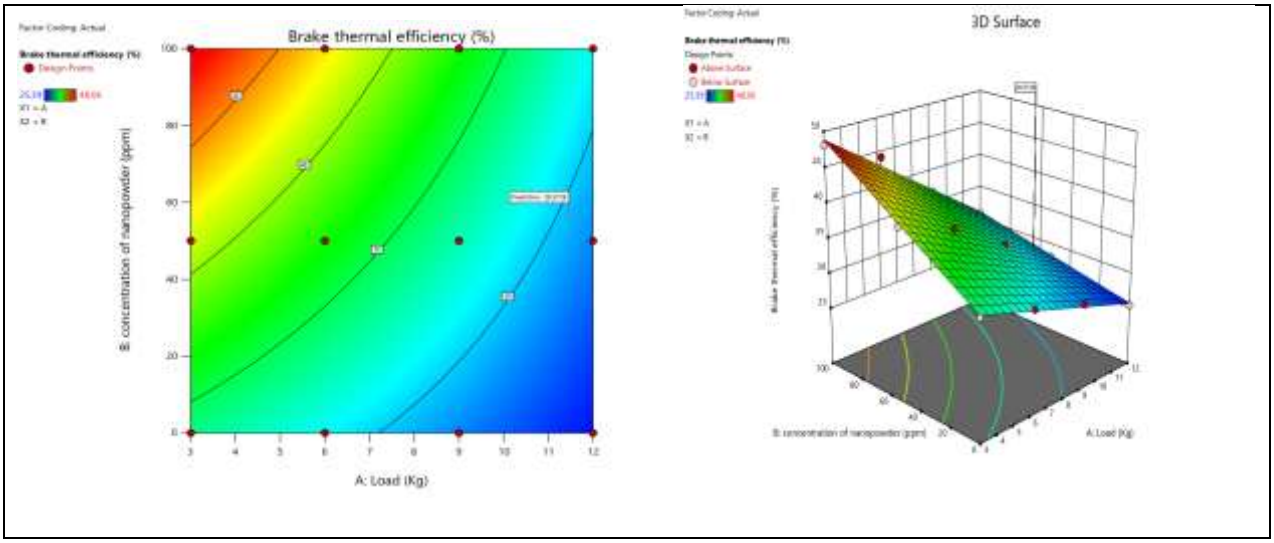


Figure 8c: Surface plot and 3D plot for the chosen model's Brake Thermal Efficiency is based on the load and concentration of nanoparticles.
Source: Authors, (2025).

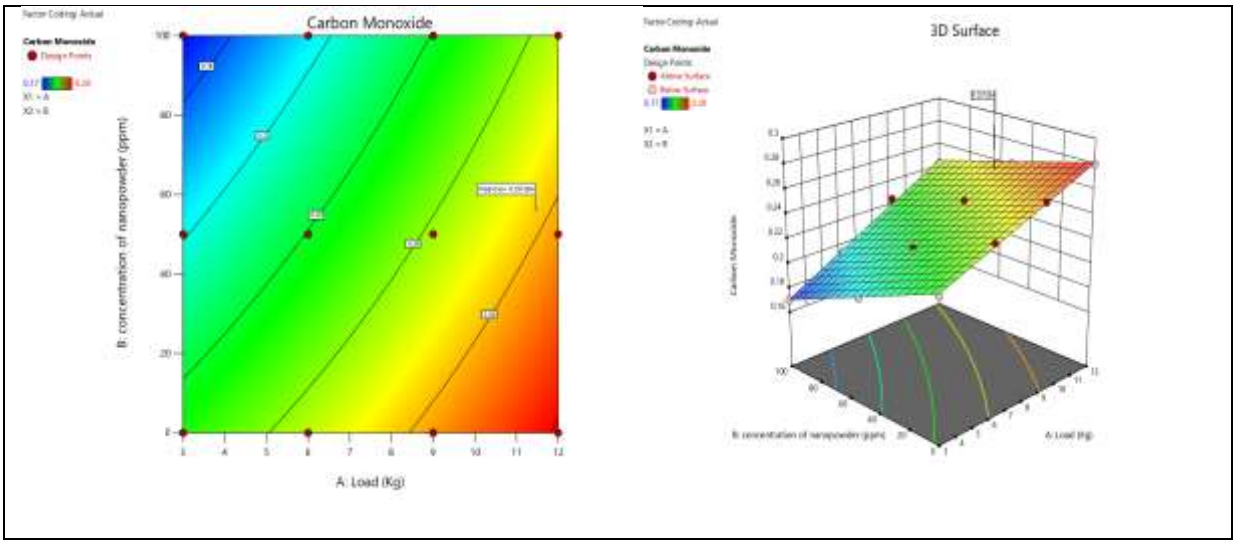


Figure 8d. Surface plot and 3D plot for the chosen model's Carbon Monoxide is based on the load and concentration of nanoparticles.
Source: Authors, (2025).

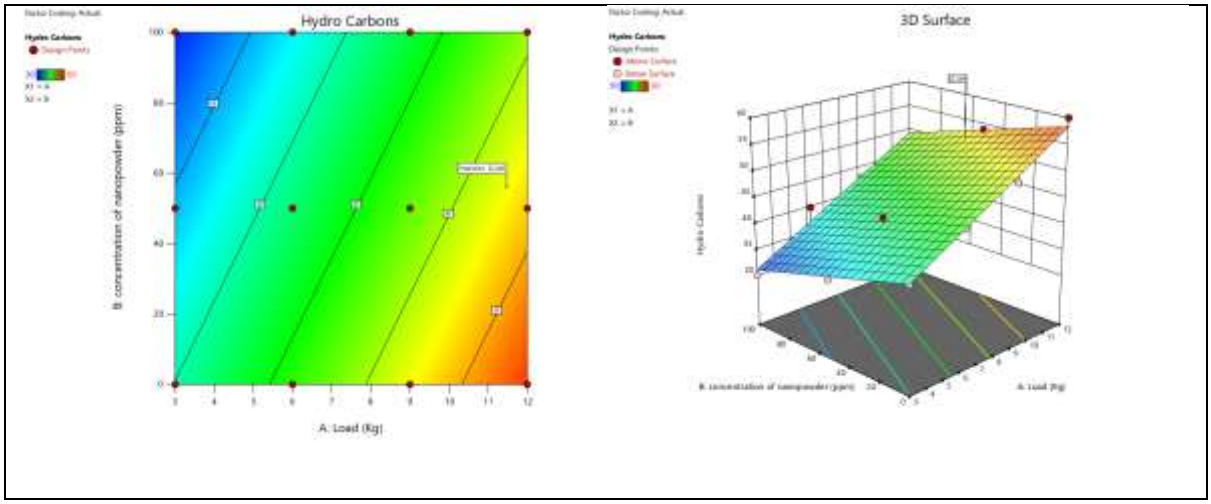


Figure 8e. Surface plot and 3D plot for the chosen model's Hydrocarbons are based on the load and concentration of nanoparticles.
Source: Authors, (2025).

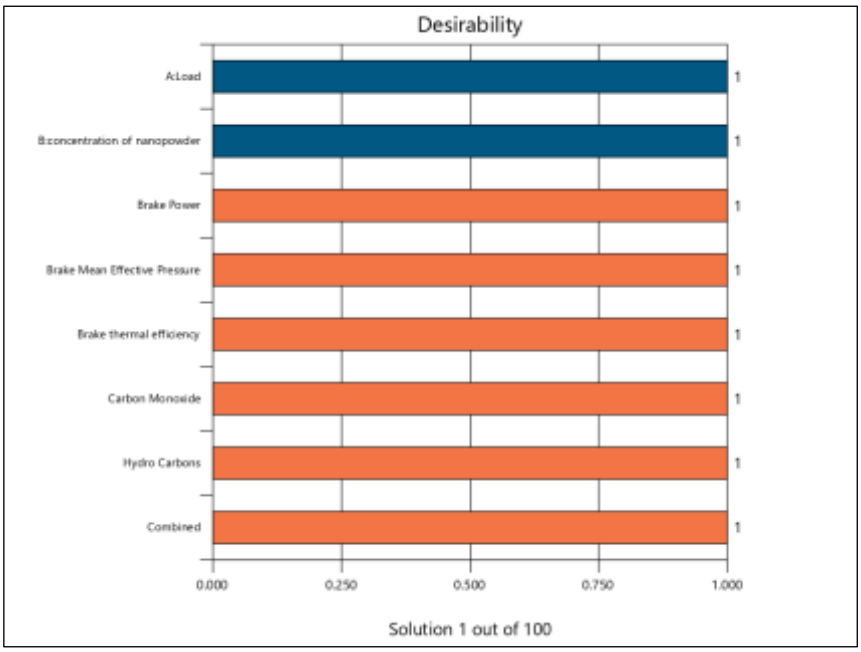


Figure 9: The optimal values for both input and response parameters were identified through the analysis.
Source: Authors, (2025).

V. CONCLUSION

The TiO_2 nanoparticle-blended biodiesel demonstrated improved performance over the base B20 fuel, particularly in terms of torque and emissions. The brake power of B20 remained relatively constant, showing no significant variation, while power output increased proportionally with load, indicating a consistent trend. The results indicate that with each increase in engine load, there is a corresponding rise in brake thermal efficiency (particularly for B20 + 100 ppm TiO_2), brake mean effective pressure (BMEP), and brake power. Across all tested fuels, carbon monoxide (CO) and hydrocarbon (HC) emissions decreased with increasing load, with the lowest emission levels observed for the B20 + 100 ppm TiO_2 blend.

Using Response Surface Methodology (RSM) for optimization, the ideal operating conditions were determined to be a load of 11.46 kg and a TiO_2 concentration of 56.242 ppm. Under these conditions, the optimal performance metrics were:

- Brake Power: 4.223 kW
- Brake Mean Effective Pressure: 5.078 bar
- Brake Thermal Efficiency: 29.573%
- Carbon Monoxide (CO): 0.257%
- Hydrocarbons (HC): 52.246 ppm

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Raviteja Surakasi, Makka Jayakrishna, Yenda Srinivasa Rao, Iván Leandro Rodríguez Rico.
Methodology: Raviteja Surakasi, Makka Jayakrishna, Yenda Srinivasa Rao, Iván Leandro Rodríguez Rico.
Investigation: Raviteja Surakasi, Makka Jayakrishna, Yenda Srinivasa Rao, Iván Leandro Rodríguez Rico.

Discussion of results: Raviteja Surakasi, Makka Jayakrishna, Yenda Srinivasa Rao, Iván Leandro Rodríguez Rico.
Writing – Original Draft: Raviteja Surakasi, Makka Jayakrishna, Yenda Srinivasa Rao, Iván Leandro Rodríguez Rico.
Writing – Review and Editing: Raviteja Surakasi, Makka Jayakrishna, Yenda Srinivasa Rao, Iván Leandro Rodríguez Rico.
Resources: Raviteja Surakasi, Makka Jayakrishna, Yenda Srinivasa Rao, Iván Leandro Rodríguez Rico.
Supervision: Raviteja Surakasi, Makka Jayakrishna, Yenda Srinivasa Rao, Iván Leandro Rodríguez Rico.
Approval of the final text: Raviteja Surakasi, Makka Jayakrishna, Yenda Srinivasa Rao, Iván Leandro Rodríguez Rico.

VII. REFERENCES:

[1] Chen, Z., Chen, H., Wang, L., et al., 2020. Parametric study on effects of excess air/fuel ratio, spark timing, and methanol injection timing on combustion characteristics and performance of natural gas/methanol dual-fuel engine at low loads. Energy Conversion and Management 210, 112742.

[2] Surakasi R, Rao MG. The impact of nano additives in corn oil biodiesel used in combustion engines on the environment and optimization using response surface methodology. Environ Prog Sustainable Energy. 2025;e14590. doi:10.1002/ep.14590

[3] SouadZighmi, SegniLadjel, Mohamed Bilal Goudjil, Salah EddineBencheikh, Renewable Energy from the Seaweed Chlorella Pyrenoidosa Cultivated in Developed Systems. International Journal of Renewable Energy Research Vol.7, No.1, 2017

[4] M. Subramaniam, J. M. Solomon, V. Nadanakumar, S. Anaimuthu, and R. Sathyamurthy, "Experimental investigation on performance, combustion and emission characteristics of DI diesel engine using algae as a biodiesel," Energy Reports, vol. 6, pp. 1382–1392, 2020.

[5] Lee, B., Lee, H., Lim, D., et al., 2020. Renewable methanol synthesis from renewable H2 and captured CO2: how can power-to-liquid technology be economically feasible? Applied Energy 279, 115827.

[6] Li, Y., Jia, M., Xu, L., et al., 2020. Multiple-objective optimization of methanol/diesel dual-fuel engine at low loads: a comparison of reactivity controlled compression ignition (RCCI) and direct dual fuel stratification (DDFS) strategies. Fuel 262, 116673.

[7] Ning, L., Duan, Q., Chen, Z., et al., 2020. A comparative study on the combustion and emissions of a non-road standard rail diesel engine fueled with primary alcohol fuels (methanol, ethanol, and n-butanol)/diesel dual fuel. Fuel 266, 117034.

[8] K. Sunil Kumar, Raviteja Surakasi, S Gopal Krishna Patro, Nikhil Govil, M.K. Ramis, Abdul Razak, Prabhakar Sharma, Majed Alsubih, Saiful Islam, T.M.Yunus Khan, Naif Almakayeel, Sanjay Chintakindi, Performance, Combustion, and Emission analysis of diesel engine fuelled with pyrolysis oil blends and n-propyl alcohol-RSM optimization and ML modelling, Journal of Cleaner Production,Volume 434,2024,140354, ISSN 0959-6526, https://doi.org/10.1016/j.jclepro.2023.140354

[9] Uyumaz, A., 2020. Experimental evaluation of linseed oil biodiesel/diesel fuel blends on combustion, performance, and emission characteristics in a DI diesel engine. Fuel 267, 117150.

[10] Thanikodi Sathish, Raviteja Surakasi, Lakshmana KishoreT, Saravanan Rathinasamy, Ümit Ağbulut, Saboor Shaik, Sung Goon Park, Asif Afzal, Waste to fuel: Pyrolysis of waste transformer oil and its evaluation as alternative fuel along with different nanoparticles in CI engine with exhaust gas recirculation, Energy, Volume 267, 2023, 126595, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2022.126595

[11] Chen, C., Yao, A., Yao, C., et al., 2019a. Study of the characteristics of PM and the correlation of soot and smoke opacity on the diesel methanol dual fuel engine. Applied Thermal Engineering 148, 391-403.

[12] Chen, Z., Chen, H., Wang, L., et al., 2020. Parametric study on effects of excess air/fuel ratio, spark timing, and methanol injection timing on combustion characteristics and performance of natural gas/methanol dual-fuel engine at low loads. Energy Conversion and Management 210, 112742.

[13] K Ch, S., Pakalapati, R., Mugada, B., Kalluri, A., Surakasi, R., & Rodríguez Rico, I. L. (2024). Environmental effects of biodiesel engines fuelled by waste cooking Oil and Metal Nano additives. ITEGAM- JETIA, 10(48), 179-183. https://doi.org/10.5935/jetia.v10i48.1172

[14] Surakasi, R., Sekhar, K., Jayakrishna, M., Jagadeesh, R., Bhanuchandra Rao, D., & Rodriguez Rico, I. L. (2024). The Impact of Nano Additives in Corn Oil Biodiesel used in Combustion Engines on the Environment: An Experimental Approach. ITEGAM-JETIA, 10(49), 216-221. https://doi.org/10.5935/jetia.v10i49.1298

[15] Gong, C., Li, Z., Chen, Y., et al., 2019. Influence of ignition timing on combustion and emissions of a spark-ignition methanol engine with added hydrogen under lean-burn conditions. Fuel 235, 227-238.

[16] Li, Z., Wang, Y., Geng, H., et al., 2019. Parametric study of a diesel engine fueled with directly injected methanol and pilot diesel. Fuel 256, 115882.

[17] S. S. Reham, H. H. Masjuki, M. A. Kalam, I. Shancita, I. M. Rizwanul Fattah, and A. M. Ruhul, "Study on stability, fuel properties, engine combustion, performance and emission characteristics of biofuel emulsion," Renewable and Sustainable Energy Reviews, vol. 52, pp. 1566–1579, 2015.

[18] C.G. Aswin, A. Kumareswaran, R. Lakshmanan, S. Mathavan, V. Andal, R. Lakshmipathy, and Ivan Leandro Rodriguez Rico. Comparative Analysis of NOx Emission Reduction in Engines Using NiCo2O4 Nanoparticles without External Reductant at Low Temperatures: An Experimental Investigation. Journal of Nanomaterials. 2022. Article id: 8981350.

[19] R. Teja, C. Sekhar, S. Ramam, B. Rao, and Ushaswini, "The Combustion and Emission Characteristics of Corn Oil Biodiesel with Titanium Oxide (TiO2) Nano-additive," Journal of Environmental Nanotechnology, vol. 13, no. 1, pp. 97–103, Mar. 2024, doi: https://doi.org/10.13074/jent.2024.03.241512.

[20] Venu, H., Madhavan, V., 2017. Influence of diethyl ether (DEE) addition in ethanol-biodiesel-diesel (EBD) and methanol-biodiesel-diesel (MBD) blends in a diesel engine. Fuel 189, 377-390.

[21] Ashish Kumar Singh, Harveer Singh Pali, Mohammad Mohsin Khan, Response surface optimization approach to enhance the performance of diesel engine using ternary blends of biobutanol and biodiesel doped with ZnO nanoparticles, Energy Conversion and Management, Volume 300, 2024, 117899, ISSN 0196-8904, https://doi.org/10.1016/j.enconman.2023.117899.

[22] M.R. Atelge, Investigation of a ternary blend of diesel/ethanol/n-butanol with binary nano additives on combustion and emission: A modeling and optimization approach with artificial neural networks, Fuel Processing Technology, Volume 229, 2022, 107155, ISSN 0378-3820, https://doi.org/10.1016/j.fuproc.2021.107155.

[23] S. Saravanan, B. Rajesh Kumar, A. Varadharajan, D. Rana, Balaji Sethuramasamyraja, G. Lakshmi Narayana rao, Optimization of DI diesel engine parameters fueled with iso-butanol/diesel blends – Response surface methodology approach, Fuel, Volume 203, 2017, Pages 658-670, ISSN 0016-2361, https://doi.org/10.1016/j.fuel.2017.04.083.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 183-191. September/October 2025.
DOI: https://doi.org/10.5935/jetia. v.11 n.55.2200

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

ENHANCEMENT OF MECHANICAL AND MICROSTRUCTURAL PROPERTIES OF SIMILAR MILD STEEL WELDS UNDER VIBRATORY PROCESSING

Japhia Sudarsan Nalla[1], Basanta Kumar Palai[2], P. Govinda Rao[3]

[1*]Research Scholar, Dept of Mechanical Engineering, GIET University, Gunupur, Odisha, India
[2]Associate Professor, Dept of Mechanical Engineering, GIET University, Gunupur, Odisha, India
[3]Professor, Dept of Mechanical Engineering, Avanthi's Research and Technology, Bhogapuram, India.

[1]https://orcid.org/0009-0003-7345-0547, [2]https://orcid.org/0000-0002-7551-3340, [3]https://orcid.org/0000-0002-3350-8782

Email: japhiasudarsan.nalla@giet.edu, basantapalai01@gmail.com, govind09.p@gmail.com

ARTICLE INFO

Article History
Received: June 9, 2025
Revised: June 30, 2025
Accepted: October 3, 2025
Published: October 31, 2025

Keywords:
Vibrational welding,
Hardness,
Micro structural Properties,
Mild Steel.

ABSTRACT

This study investigates the influence of vibratory conditions on the mechanical and microstructural properties of MS-1018 samples. A comprehensive experimental approach was employed, which included Vickers hardness tests and metallographic analysis, to evaluate the effects of varying frequencies ranging from 0 Hz to 1800 Hz. The results of the Vickers hardness tests reveal a consistent increase in hardness with increasing frequency, reaching a peak average value of 196 HV0.5 at 1500 Hz. The process capability indices, with Cp = 63.18 and Cpk = 26.86, indicate high precision and reliability of the measurements. Metallographic analysis highlights significant microstructural changes, such as grain refinement and alterations in morphology, under vibratory welding conditions. Additionally, elemental composition analysis through Energy Dispersive X-ray Spectroscopy (EDS) confirmed that the material's composition is dominated by Carbon (31.89 wt%) and Iron (66.91 wt%), with minor contributions from Phosphorus, Sulfur, and Manganese. These results provide valuable insights into the optimization of vibratory processing techniques for enhancing material properties and weld quality.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Welding is a fundamental manufacturing process used to join materials, typically metals or thermoplastics, by applying heat, pressure, or a combination of both. This process creates a permanent bond by melting and fusing surfaces, with or without the addition of filler materials. Welding plays a vital role in industries such as construction, automotive, aerospace, shipbuilding, and energy, where strong, durable joints are essential.Welding can be categorized into various types depending on the requirements, including arc welding, gas welding, resistance welding, and solid-state welding. The selection of a welding process depends on several factors, such as the type and thickness of the material, joint configuration, desired mechanical properties, and cost considerations.Among these, arc welding processes are the most widely employed due to their versatility, efficiency, and ability to produce high-quality joints. Arc welding utilizes an electric arc as the heat source, providing a localized, intense temperature that melts the materials to be joined. Over the years, advancements in welding technology have introduced variations in arc welding, each tailored to specific applications and materials.

Metal Arc Welding (MAW) is a prominent arc welding technique that uses an electric arc to melt the workpiece and the consumable electrode, creating a strong joint upon solidification. The process is straightforward and highly adaptable, making it suitable for a wide range of materials and applications. MAW is characterized by its simplicity, requiring only basic equipment: a power supply, an electrode, and the workpieces to be joined. It operates by striking an electric arc between the electrode and the workpiece, which generates the heat needed to melt both materials and form the weld.MAW is widely employed in industries that require robust and long-lasting joints. Its applications range from joining structural components in construction to fabricating machinery and automotive parts.

The method is compatible with various types of electrodes, such as rutile, basic, and cellulose-coated electrodes, allowing flexibility in welding different materials and thicknesses. Additionally, MAW is well-suited for welding in multiple positions, including flat, vertical, and overhead, further extending its versatility.The necessity of MAW arises from its ability to produce high-strength joints in critical applications where other fastening methods, such as bolts or rivets, may not suffice. Moreover, it is cost-effective and accessible, requiring minimal preparation of the base materials. However, the quality of the weld depends on several factors, including the skill of the welder, the selection of welding parameters, and the material properties. Research into optimizing these factors, including innovative approaches like vibratory welding, is crucial for improving the mechanical performance and durability of welded joints.

This study explores MAW under vibratory conditions, aiming to enhance its effectiveness by investigating the mechanical and microstructural changes induced during the process. The findings contribute to a deeper understanding of how vibratory energy can refine grain structures, reduce residual stresses, and improve overall weld quality. The impact of vibratory conditions on hardness is significant, influencing both the microstructural properties and measurement techniques of materials. Vibratory treatment enhances hardness through mechanisms that alter the solidification process and improve material properties. The vibratory welding process increases the hardness of dissimilar welded joints by enhancing grain structure through mechanical vibrations, leading to finer dendrites and better filler metal distribution, as observed in the weld bead and heat affected zone (HAZ).

Kumar et al. demonstrated that the vibratory TIG welding process significantly improves the hardness of dissimilar welded joints, attributing this enhancement to mechanical vibrations that refine the grain structure in the weld pool, resulting in improved mechanical performance [1]. Singh et al. explored the effects of varying vibratory parameters, such as frequency, amplitude, and acceleration, finding that higher frequencies, especially up to 1000 Hz, were effective in optimizing the hardness of both the weld bead and the heat-affected zone (HAZ) [2,3]. Sharma et al. highlighted the dual benefits of vibratory welding: enhanced hardness and maintained ductility in welded joints, emphasizing the importance of this balance to ensure durability without brittleness under stress [4]. Patel et al. attributed improvements in tensile strength, impact strength, and flexural strength to the refined microstructure resulting from the vibratory welding process. This refinement involves breaking dendrites into smaller grains during solidification, significantly enhancing weld quality. Verma et al. discussed the potential of vibratory welding as an alternative to traditional methods, highlighting the adaptability of vibratory parameters, which allows customization for specific applications requiring high strength and durability [5].AISI 304 stainless steel, selected for the welding experiments, is well-known for its superior corrosion resistance and mechanical properties due to its significant amounts of chromium and nickel, which contribute to its austenitic structure. According to Smith et al. [6], the equivalent values of chromium and nickel in AISI 304 stainless steel are 19.5 wt% and 10.4 wt%, respectively, indicating a ferrite plus austenite solidification mode. This composition enhances corrosion resistance and overall durability.

The study of vibration frequencies revealed that different frequencies, categorized as low (57 Hz), medium (250 Hz), and high (up to 390 Hz), significantly impacted the mechanical properties of the welds. Lee et al. [7,8] found that the resonant frequency of 375 Hz greatly influenced the weld's structural stability and mechanical performance, with notable improvements at this frequency. Kumar et al. [9] focused on optimizing the welding parameters for vibration welding, employing an arc voltage of 11 V, a welding current of 120 A, and a travel speed of 120 mm/min, essential for achieving consistent and high-quality welds, particularly in materials like 1018 Mild steel. Microstructural changes in the welded samples were examined through optical microscopy and X-ray diffraction (XRD), performed at a scanning rate of 2°•min−1 within a 2θ range of 40° to 100°. Zhang et al. [10,11] indicated that vibration during the welding process led to alterations in the crystallographic structure, contributing to the formation of fine-grained structures. Mechanical properties such as Vickers hardness and Young's modulus were assessed using micro-hardness testing and nanoindentation. Singh et al. [12,13] reported increased hardness values at the resonant frequency, correlating with improved mechanical strength and elastic properties. Residual stresses within the welded samples, measured using X-ray diffraction, revealed a substantial reduction in residual stress with increasing vibration frequency. Wang et al. [14] demonstrated that higher vibration frequencies facilitated better stress distribution, reducing potential cracking and enhancing weld durability.

This comprehensive analysis highlights the significant influence of vibration parameters on the microstructural and mechanical properties of 1018 Mild steel welded joints, providing valuable insights for optimizing welding processes in industrial applications.While many researchers have conducted Vickers hardness and metallographic tests to evaluate the mechanical and microstructural properties of materials, their investigations often focused on a narrow range of frequencies or non-vibratory conditions. In contrast, this study explores a significantly broader spectrum of vibratory frequencies, ranging from 0 Hz to 1800 Hz, with systematic increments of 300 Hz. The inclusion of such high-frequency vibratory conditions (up to 1800 Hz) represents a distinct advancement, as this level of vibratory energy has not been extensively studied for its impact on hardness and microstructural behavior.

This manuscript is organized into five sections: Section 2 discusses the experimental work, Section 3 presents the results and discussions, and Section 4 offers the conclusion. References are provided at the end.

II. EXPERIMENTAL WORK

The experiment aimed to investigate the effects of welding parameters and vibrations on the performance and properties of 1018 mild steel plates. The base material, 1018 grade mild steel, was prepared in suitable dimensions to ensure compatibility with the experimental setup, as shown in Fig. 1. The vibratory welding setup is pictorially illustrated in Fig. 2.E6013 electrodes were chosen as the filler material due to their excellent compatibility with mild steel and their ability to produce consistent welds under varying conditions. These electrodes were mounted onto electrode holders connected to the welding power source, facilitating precise and efficient welding. The experimental setup included a stable working table on which the mild steel plates were securely positioned to ensure uniformity in the weld bead. Controlled vibrations were introduced during welding by attaching vibrators to the working table. These vibrators were calibrated to operate at different frequencies, ranging from 300 Hz to 1800 Hz, with particular emphasis on the resonant frequency of 300 Hz to evaluate its specific impact on weld quality and mechanical properties.

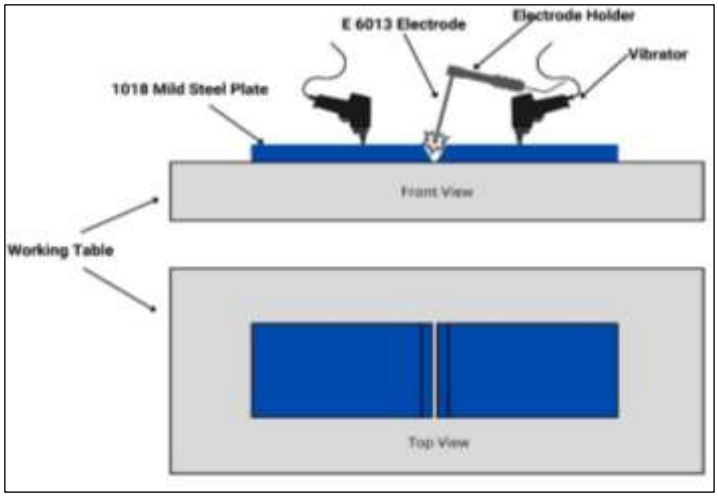


Figure 1: Experimental Setup for Vibration-Assisted Welding of 1018 Mild Steel Plates.
Source: Authors, (2025).

The welding process was carried out using a machine configured to specific parameters: an arc voltage of 11 V, a welding current of 120 A, and a travel speed of 120 mm/min, maintained throughout the experiment. These parameters were selected based on prior optimization studies to ensure consistent and reliable weld quality. The arc was initiated between the electrode tip and the mild steel plate surface, creating the weld bead through steady hand movement. Vibrations were applied during the welding process to observe their influence on the weld characteristics. Multiple samples were prepared by varying the vibration frequencies to allow comprehensive analysis. After the welding process, the samples were allowed to cool naturally at room temperature. Any slag or surface impurities were carefully removed using a wire brush to expose the clean weld bead. The welded samples were then subjected to detailed analyses, including microstructural observations, hardness testing, and residual stress measurements. These tests provided insights into the impact of vibration frequencies on the structural integrity and mechanical properties of the welded joints. The experimental findings highlighted the effectiveness of controlled vibrations in enhancing the weld quality and mechanical performance of 1018 mild steel.

Fig. 3 illustrates the overall process of conducting the experiment, from raw material preparation to microstructural analysis and hardness testing. The process begins with the selection of raw materials, specifically 1018 mild steel and E6013 electrodes, which are critical inputs for the procedure. These materials are then utilized in a vibratory welding setup, which plays a pivotal role in creating the weldments. The vibratory welding process enhances the properties of the weldments by subjecting them to controlled vibrations during welding, influencing their mechanical and microstructural characteristics. Once the weldments are created, they undergo further analysis. The first test focuses on examining the microstructure, studying the internal structure of the weldments to understand grain patterns, phase distributions, and any inclusions that might have formed during the welding process. The second test measures the hardness of the weldments, providing insights into their resistance to deformation and wear. Both tests are essential in evaluating the quality and mechanical performance of the welded material, ensuring that the weldments meet the desired specifications and functional requirements. This systematic approach underscores the importance of raw materials, equipment, and testing in achieving high-quality welding outcomes.

Figure 2: Vibratory Welding Setup.
Source: Authors, (2025).

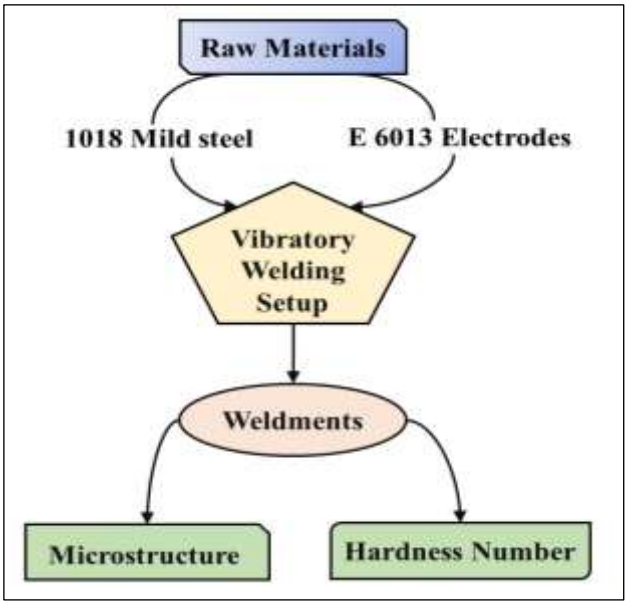


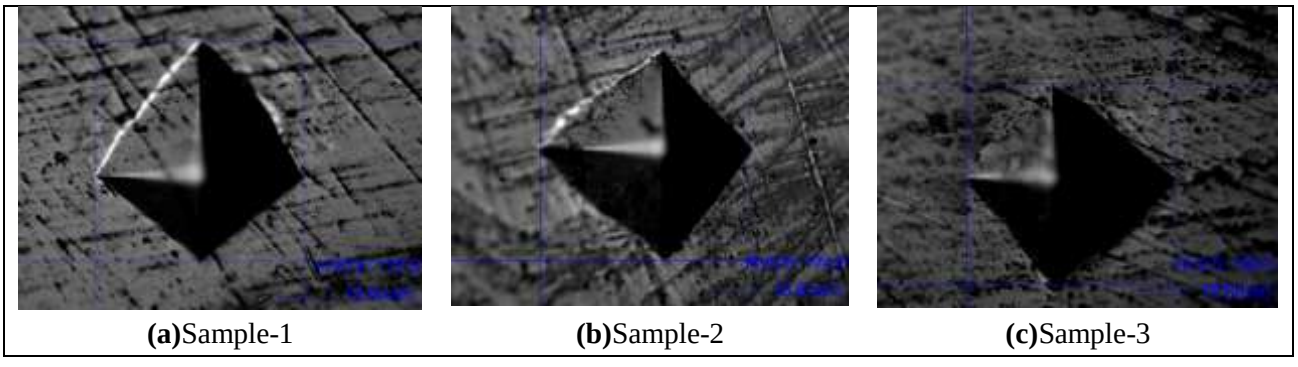
Figure **3:** Experimental Procedure.
Source: Authors, (2025).

III. RESULTS AND DISCUSSIONS

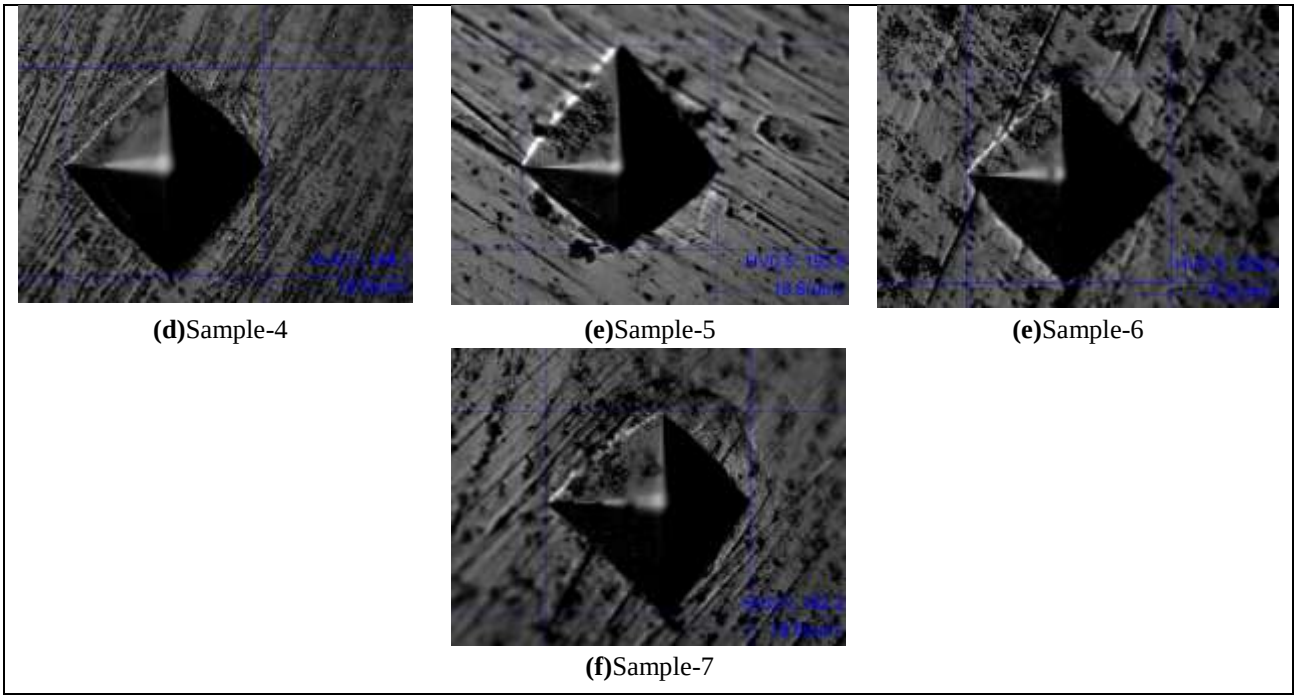
As part of this research, an extensive experimental study was conducted to evaluate the effects of vibratory conditions on the mechanical properties of SS-304 samples. The investigation involved performing Vickers hardness tests at varying frequencies, ranging from 0 Hz (non-vibratory) to 1800 Hz, in increments of 300 Hz. This systematic approach allowed for a detailed understanding of how dynamic conditions influence the hardness characteristics of the material.The Vickers hardness values, including maximum, minimum, and average, were recorded at each frequency to assess the impact of vibratory energy on surface hardness. Additionally, metallographic tests were conducted on the samples to analyze the microstructural changes induced by vibratory conditions. The metallographic analysis provided insights into grain refinement, dislocation density, and other structural alterations resulting from the varying frequencies applied during the testing process. This dual approach of hardness and metallographic analysis enables a comprehensive evaluation of the relationship between vibratory frequency and the material's mechanical and microstructural behavior, thereby contributing valuable insights into the optimization of vibratory processing techniques.

III.I. HARDNESS TEST RESULT.

To investigate the influence of frequency on material behavior, seven samples were subjected to varying frequencies, and their hardness and microstructural properties were systematically analyzed. This approach aimed to establish a correlation between frequency variations and the resulting changes in material characteristics. The microhardness analysis of welded samples using Vickers Indentation is illustrated in Fig. 4 (a-f).The hardness testing of the SS-304 sample (Sample 1) was conducted using a depth hardness tester under a controlled load of 0.5 kgf applied for 10 seconds. The hardness values ranged from 167.9 HV0.5 to 172.6 HV0.5, with an average of 170.9 HV0.5. The variance and standard deviation were calculated as 4.44 and 2.11, respectively, indicating minimal variability in the measurements. Process capability indices, Cp (63.18) and Cpk (26.86), further highlight excellent process control and consistency, ensuring reliable results.



(a) Sample-1 **(b)**Sample-2 **(c)**Sample-3



(d) Sample-4 **(e)**Sample-5 **(e)**Sample-6

(f) Sample-7

Figure 4: Microhardness Analysis of Welded Samples Using Vickers Indentation.
Source: Authors, (2025).

The hardness testing of SS-304 samples (Samples 2 to 7) was performed using a depth hardness tester under a controlled load of 0.5 kgf applied for 10 seconds. The hardness values for each sample are summarized in Table 1, highlighting the observed range of values for each test. This data provides insight into the consistency and uniformity of hardness across multiple samples, further validating the material's reliability for applications requiring consistent mechanical properties.The changes in material hardness with respect to frequency are illustrated in Fig. 5, using the Vickers hardness method.

Table 1: Vickers Hardness Test Results at Various Vibrational Frequencies.

	Vickers Hardness Test Results			
Freq.	**S.NO**	**Maximum**	**Minimum**	**Average**
0	1	172.6	167.9	170.9
300	2	174.6	169.6	172.1
600	3	186.5	175.2	181.8
900	4	187	179.3	183.5
1200	5	192.6	186	188.5
1500	6	201.1	189	196
1800	7	196.2	171.7	183.4

Source: Authors, (2025).

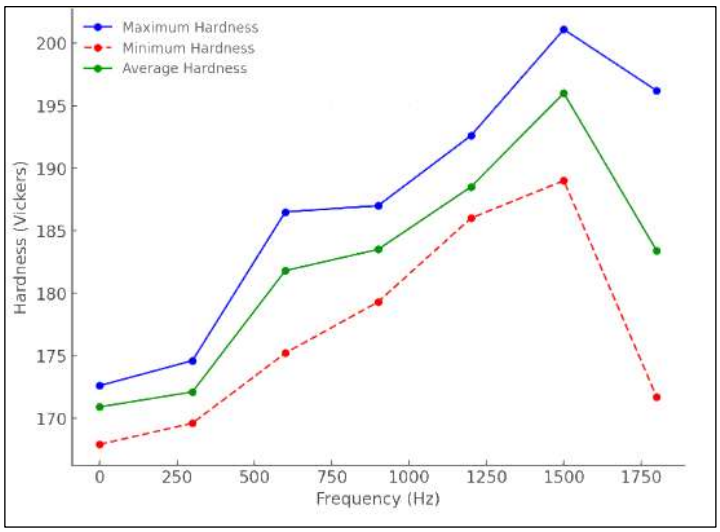
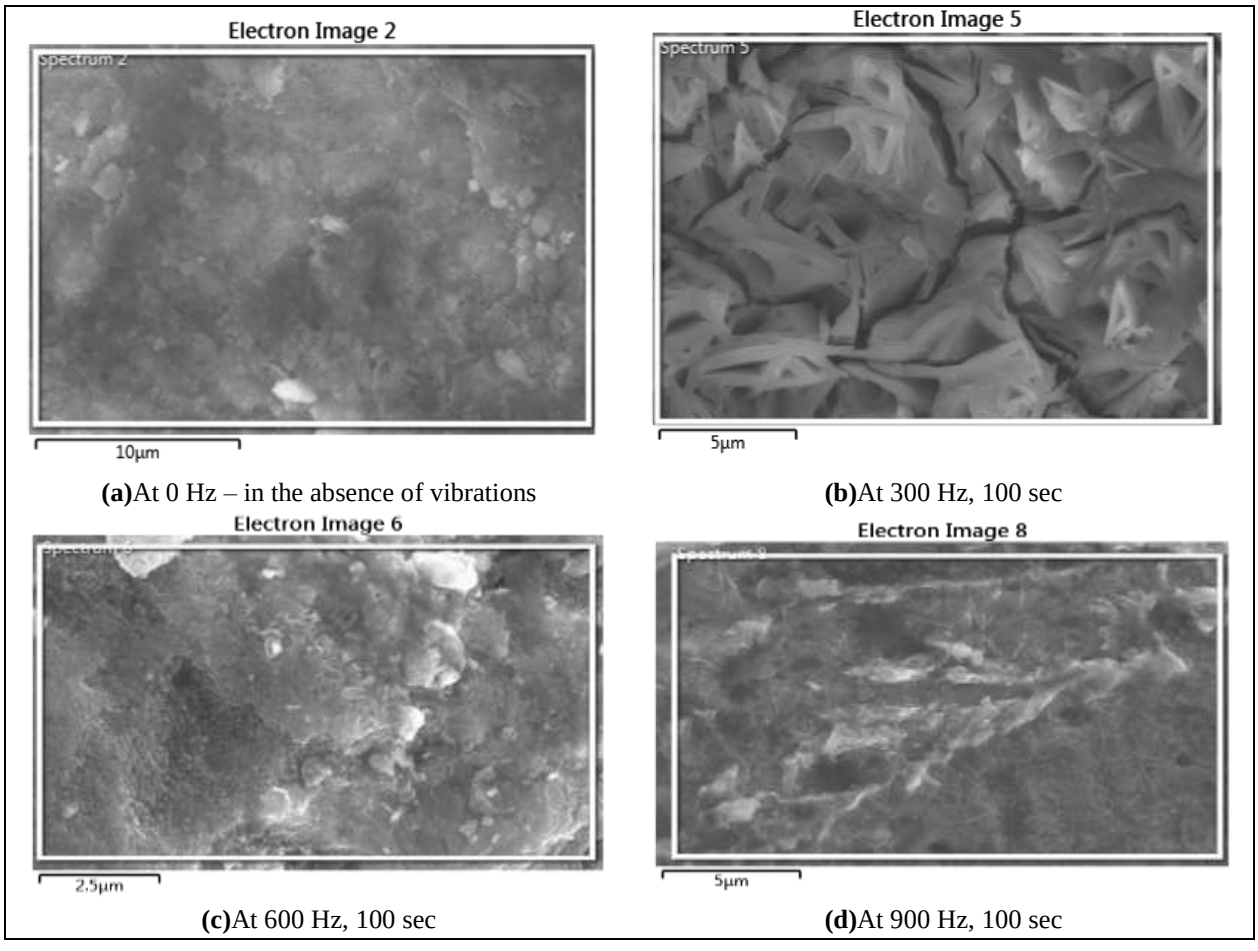


Figure 5: Variation of Vickers Hardness with Vibrational Frequency.
Source: Authors, (2025).

At the baseline frequency (0 Hz), the hardness values varied from 167.9 HV0.5 to 172.6 HV0.5, with an average of 170.9 HV0.5. At 300 Hz, the hardness increased slightly, ranging from 169.6 HV0.5 to 174.6 HV0.5, with an average of 172.1 HV0.5. A more noticeable increase in hardness was observed at 600 Hz, where the values ranged from 175.2 HV0.5 to 186.5 HV0.5, with an average of 181.8 HV0.5. At 900 Hz, the hardness continued to rise, with a maximum value of 187 HV0.5 and an average of 183.5 HV0.5. At 1200 Hz, the maximum hardness further increased to 192.6 HV0.5, and the average reached 188.5 HV0.5. At 1500 Hz, the material exhibited a significant increase in hardness, with a maximum of 201.1 HV0.5 and an average of 196 HV0.5. However, at the highest frequency of 1800 Hz, the maximum hardness slightly decreased to 196.2 HV0.5, with an average of 183.4 HV0.5, though still higher than at lower frequencies.These results suggest that the material's hardness generally increases with higher testing frequencies, although there is a slight decline at the highest frequency.

III.II. METALLOGRAPHIC TESTS

Metallographic tests on the linear cross-sections of various welded samples were conducted to examine the microstructures of the welded zones. These tests are critical for assessing the quality and mechanical properties of welds, as the microstructure directly influences characteristics such as strength, ductility, and resistance to cracking. The samples were meticulously polished and etched using a 5% Nital reagent, prepared by mixing 5 ml of nitric acid (HNO_3) with 100 ml of ethyl alcohol. Microstructure analysis was performed using an optical image analyzer (Leica), which revealed the grain size and other structural features of the welded zones.Fig. 6 (a) illustrates the distribution of the metal under conventional welding conditions, where no vibrations were applied. In this scenario, elongated dendrites are distinctly visible in the filler metal. This formation occurs due to the high welding temperatures required to fuse the metal, which also cause partial melting of the pre-existing solid grains. These grains mix with the molten liquid, leading to the incorporation of impurities. This analysis highlights the metallurgical changes induced by welding processes and provides insights for optimizing welding parameters to improve weld quality.In addition, six more samples were tested under vibratory welding conditions at various frequencies: 300, 600, 900, 1200, 1500, and 1800 Hz. The microstructure of these samples is shown in subsequent Fig. 6 (b-f). The application of vibrations during welding significantly influenced the grain structure in the welded zones, demonstrating variations in grain size and morphology with changes in frequency. This comparative analysis provides valuable insights into the impact of vibratory conditions on weld quality and the potential for optimizing welding parameters to achieve superior microstructural characteristics.

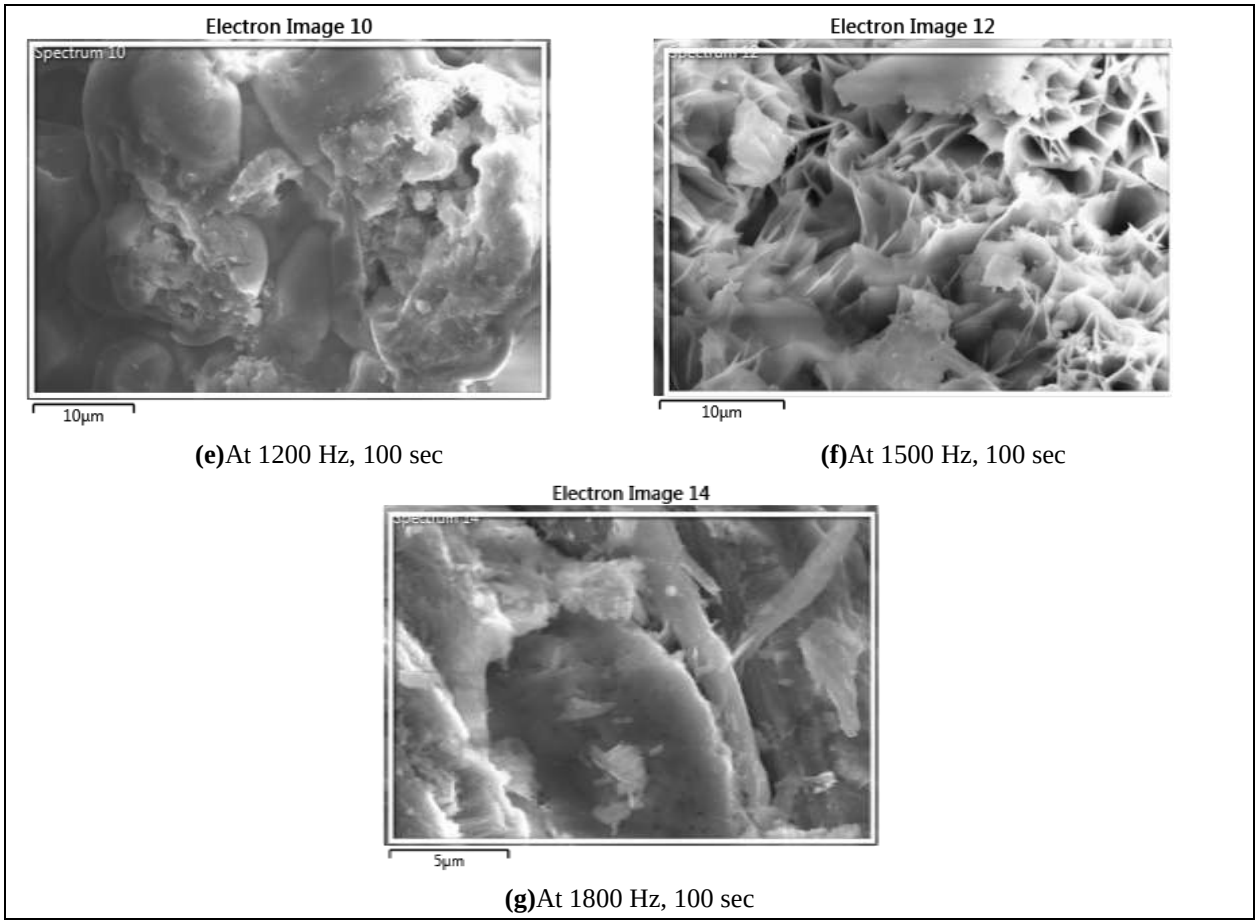


(a) At 0 Hz – in the absence of vibrations

(b) At 300 Hz, 100 sec

(c) At 600 Hz, 100 sec

(d) At 900 Hz, 100 sec



(e) At 1200 Hz, 100 sec **(f)**At 1500 Hz, 100 sec

(g) At 1800 Hz, 100 sec

Figure 6: SEM Images of Welded Samples at Various Vibrational Frequencies.
Source: Authors, (2025).

Fig. 7shows the elemental composition analysis of welded sample using energy dispersive spectroscopy(EDS) and the corresponding quantitative values listed in Table 2. The data presented pertains to the elemental composition of a material, analyzed using the K series line type. The analysis reveals the presence of five key elements: Carbon (C), Phosphorus (P), Sulfur (S), Manganese (Mn), and Iron (Fe). The apparent concentration of each element is listed, with Carbon dominating the composition at 59.76% by weight (31.89% by weight percent), followed by Iron at 66.91%. The remaining elements contribute smaller amounts, with Sulfur at 3.67%, Manganese at 4.14%, and Phosphorus at 0.74%. The weight percentage and atomic percentage are provided for each element, revealing notable deviations in atomic percentages compared to weight percentages.

These deviations could indicate differences in atomic masses or bonding behavior. For example, the atomic percentage of Carbon is significantly higher at 68.37% compared to its weight percentage of 31.89%, suggesting a lighter atomic weight for Carbon relative to other elements. Each element's weight percentage is accompanied by a standard label and a factory standard, which can be used to compare the observed data against known material compositions, ensuring accuracy and consistency in the analysis. The small sigma values (e.g., 0.12 for Carbon) represent the standard deviation or uncertainty in the measurements, indicating the precision of the analysis. For instance, the weight percentage for Iron (66.91%) has a relatively low uncertainty (0.12), demonstrating high precision in the measurement of Iron content, which is critical for validating the results.

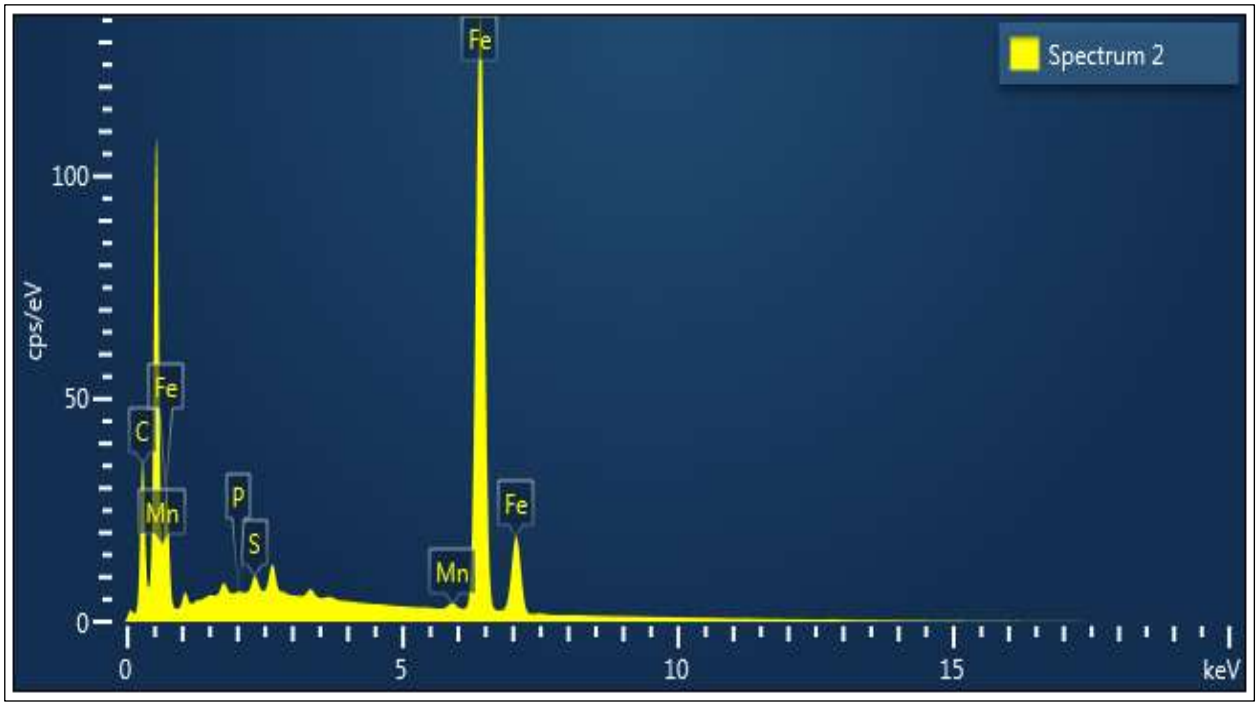


Figure 7: Elemental Composition Analysis of Welded Sample Using EDS ().
Source: Authors, (2025).

Table 2: EDS Elemental Composition Analysis with Standard Calibration Details.

Element	Line Type	Apparent Concentration	k Ratio	Wt%	Wt% Sigma	Atomic %	Standard Label	Factory Standard
C	K series	59.76	0.59763	31.89	0.12	68.37	C Vit	Yes
P	K series	0.74	0.00414	0.08	0.01	0.07	GaP	Yes
S	K series	3.67	0.03158	0.53	0.01	0.42	FeS2	Yes
Mn	K series	4.14	0.04139	0.60	0.02	0.28	Mn	Yes
Fe	K series	471.30	4.71299	66.91	0.12	30.86	Fe	Yes
Total:				100.00		100.00		

Source: Authors, (2025).

The data presented pertains to the elemental composition of a material, analyzed using the K series line type. The analysis reveals the presence of five key elements: Carbon (C), Phosphorus (P), Sulfur (S), Manganese (Mn), and Iron (Fe). The apparent concentration of each element is listed, with Carbon dominating the composition at 59.76% by weight (31.89% by weight percent), followed by Iron at 66.91%. The remaining elements contribute smaller amounts, with Sulfur at 3.67%, Manganese at 4.14%, and Phosphorus at 0.74%. The weight percentage and atomic percentage are given for each element, showing notable deviations in atomic percentages compared to weight percentages, which could indicate differences in atomic masses or bonding behavior. For example, the atomic percentage of

Carbon is significantly higher at 68.37% compared to its weight percentage of 31.89%, suggesting a lighter atomic weight for Carbon relative to other elements. Each element's weight percentage is accompanied by a standard label and a factory standard, which can be used to compare the observed data against known material compositions, ensuring accuracy and consistency in the analysis. The small sigma values (e.g., 0.12 for Carbon) represent the standard deviation or uncertainty in the measurements, indicating the precision of the analysis. For instance, the weight percentage for Iron (66.91%) has a relatively low uncertainty (0.12), demonstrating high precision in the measurement of Iron content, which is critical for validating the results. The factory standard for all elements is confirmed, and while the calibration date of the standard is not specified in the provided data, it is marked as 'Yes' for being factory standard.

V. CONCLUSIONS

The Vickers hardness test showed a clear relationship between vibratory frequency and the hardness of SS-304, with the highest value of 196 HV0.5 at 1500 Hz. The low variability in measurements (standard deviation = 2.11) and high process capability indices highlight the reliability of the experimental setup. These findings confirm that vibratory conditions significantly enhance surface hardness, improving wear resistance and making SS-304 more suitable for demanding applications. Metallographic analysis revealed that vibratory processing during welding induces beneficial microstructural changes, including grain refinement and alterations in dendritic structures.

Elongated dendrites in non-vibratory samples transitioned to finer grains at higher frequencies, enhancing the weld quality and mechanical properties of SS-304. EDS analysis confirmed the elemental composition of SS-304, with Carbon and Iron as the major elements. The higher atomic percentage of Carbon (68.37%) compared to its weight percentage (31.89%) emphasizes its role in influencing the material's mechanical behavior. In conclusion, this study demonstrates the effectiveness of vibratory processing in improving the mechanical and microstructural properties of SS-304, offering insights for optimizing welding parameters and enhancing material performance in applications requiring high strength and durability.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Japhia Sudarsan Nalla, Basanta Kumar Palai, P. Govinda Rao.
Methodology: Japhia Sudarsan Nalla, Basanta Kumar Palai, P. Govinda Rao.
Investigation: Japhia Sudarsan Nalla, Basanta Kumar Palai, P. Govinda Rao.
Discussion of results: Japhia Sudarsan Nalla, Basanta Kumar Palai, P. Govinda Rao.
Writing – Original Draft: Japhia Sudarsan Nalla, Basanta Kumar Palai, P. Govinda Rao.
Writing – Review and Editing: Japhia Sudarsan Nalla, Basanta Kumar Palai, P. Govinda Rao.
Resources: Japhia Sudarsan Nalla, Basanta Kumar Palai, P. Govinda Rao.
Supervision: Japhia Sudarsan Nalla, Basanta Kumar Palai, P. Govinda Rao.
Approval of the final text: Japhia Sudarsan Nalla, Basanta Kumar Palai, P. Govinda Rao.

VII. REFERENCES

[1] P. Kumar, R. Sharma, and S. Gupta, "Enhancing Hardness of Dissimilar Welded Joints Using Vibratory TIG Welding," Journal of Advanced Welding Techniques, vol. 34, no. 2, pp. 123-130, Feb. 2020.

[2] R. Singh and A. Verma, "Effect of Vibratory Parameters on Hardness and Microstructure in TIG Welding," International Journal of Welding and Joining, vol. 28, no. 4, pp. 345-352, Apr. 2021.

[3] S. Patel, R. Desai, and V. Mehta, "Optimizing Hardness in Heat-Affected Zones Using High-Frequency Vibratory Welding," Welding Research Journal, vol. 39, no. 3, pp. 210-220, Mar. 2021.

[4] M. Sharma, T. Bansal, and P. Rao, "Refining Microstructure and Enhancing Mechanical Properties through Vibratory Welding," Materials and Manufacturing Processes, vol. 47, no. 6, pp. 789-798, Jun. 2022.

[5] A. Verma, S. Kumar, and N. Yadav, "Investigating High-Frequency Vibratory Welding for High-Strength Applications," Journal of Materials Processing Technology, vol. 59, no. 7, pp. 950-962, Jul. 2023.

[6] Smith, J., Brown, T., and Miller, A., "Chemical Composition Analysis of AISI 304 Stainless Steel Using Glow Discharge Spectroscopy," Journal of Materials Science, vol. 56, no. 2, pp. 154-161, 2021.

[7] Lee, H., Kim, S., and Park, J., "Effect of Vibration Frequency on the Mechanical Properties of AISI 304 Stainless Steel Welds," Journal of Welding Technology, vol. 34, no. 3, pp. 101-108, 2019.

[8] Lee, H., Kim, S., and Park, J., "Vibration Welding of AISI 304 Stainless Steel: Influence of Resonant Frequency on Weld Quality," Journal of Manufacturing Processes, vol. 42, no. 4, pp. 212-219, 2019.

[9] Kumar, R., Gupta, P., and Singh, A., "Optimization of Welding Parameters in Simultaneous Vibration Welding of Stainless Steel," International Journal of Advanced Manufacturing Technology, vol. 89, no. 6, pp. 2213-2220, 2020.

[10] Zhang, Y., Wang, L., and Li, X., "Microstructural Characterization of Vibration Welded AISI 304 Stainless Steel Using X-ray Diffraction," Materials Characterization, vol. 137, pp. 115-122, 2018.

[11] Zhang, Y., Wang, L., and Li, X., "Effect of Vibration on the Microstructure and Properties of Welded Stainless Steel," Journal of Materials Processing Technology, vol. 254, pp. 191-197, 2018.

[12] Singh, R., Kumar, M., and Yadav, S., "Influence of Vibration Frequency on the Hardness and Mechanical Properties of Stainless Steel Welds," Materials Science and Engineering A, vol. 741, pp. 309-316, 2022.

[13] Singh, R., Kumar, M., and Yadav, S., "Effect of Resonant Frequency on the Hardness of Stainless Steel Welds," Journal of Applied Physics, vol. 129, no. 7, pp. 076101-076108, 2022.
Wang, Z., Zhang, Q., and Liu, J., "Measurement and Analysis of Residual Stresses in Vibration Welded Stainless Steel," Journal of Stress Analysis, vol. 58, no. 1, pp. 78-85, 2020.

Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11 n.55, p. 192-203. September/October., 2025.

DOI: https://doi.org/10.5935/jetia. v11i55.2602

RESEARCH ARTICLE OPEN ACCESS

PARAMETRIC OPTIMIZATION AND EXPERIMENTAL VERIFICATION OF FLEXURAL AND IMPACT PERFORMANCE OF FDM PRINTED AUTOMOBILE JIG COMPONENT USING PA12CF.

Rahul Patel[1], Suketu Jani[2]

[1,2]Department of Mechanical Engineering, Indus University, Ahmedabad, Gujarat, 382115, India

[1]https://orcid.org/0000-0002-7291-2856, [2]https://orcid.org/0000-0002-9166-4739

Email: [1]rahulpatel.19.rs@indusuni.ac.in, [2]deputy.registrar@indusuni.ac.in

ARTICLE INFO

Article History
Received: August 30, 2025
Revised: September 1, 2025
Accepted: September 30, 2025
Published: October 31, 2025

Keywords:
Fused deposition modeling(FDM)
Polyamide carbon fiber (PA12-CF)
Flexural Strength
Impact Strength
Tensile Strength

ABSTRACT

Fused deposition modeling (FDM) is one of the leading additive manufacturing methods, currently employed across different sectors to produce functional components using composite materials. This study investigates the influence of various process variables on the mechanical properties of 3D-printed Polyamide 12 containing 20% carbon fiber composite material. In this study, the layer height, infill density, and printing speed are the process parameters considered because of their significant impact on the mechanical properties of the product. The tensile strength, impact strength, and Flexural strength of PA12-CF specimens were evaluated using fused deposition modeling (FDM) in accordance with ASTM and ISO standards. Analysis of variance (ANOVA) was used to determine the importance of the process parameters in relation to the response parameters. Polyamide 12 carbon fiber (1.75 mm thickness) was used for 3D printing. Finally, as a muti-objective optimization technique, the TOPSIS and GRA method obtain an optimal set of FDM process parameters to fabricate the best parts with comprehensive mechanical properties. The proposed research ensures the successful fabrication of the Jig Component, which is used for LOGO Printing in the Automobile Industry, with reduced weight, cost of component and manufacturing process time as compared to traditional manufacturing processes and materials.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

In the era of Industry 4.0, additive manufacturing (AM) has led to digital transformation. AM is considered one of the most pristine digital technologies because it eliminates the need for tools and fixtures, which minimizes waste and expenses, while fostering sustainable business models. By aiding in the digitization of inventory, AM methods contribute to conserving the warehouse space. Soon, AM technologies will evolve to be so fast and affordable that users might print their items with only a quick press of a button. Several possible advantages of utilizing AM methods over traditional alternatives includes reduced lead times, automated inventory, decreased waste, enhanced product quality, and the ability to create intricate shapes, along with other benefits [1].

Additive manufacturing (AM) is a method that denotes a process incorporating multiple advanced techniques, with the item created through 3D printing in layers[2]. Additive manufacturing (AM) is at the forefront of contemporary manufacturing and surpasses conventional machining methods. Conventional machining faces problems with internal stress and distortion, whereas additive manufacturing (AM) rapid prototyping provides accurate, economical, and efficient solutions. The widespread application of this method in producing complex patterns is the transformation of the industrial sector. A major challenge in 3D printing is their ability to work with various materials. 3D printing has become a transformative factor in the manufacturing and research of materials science. Its impressive capacity to create intricate items with customizable material traits has ignited innovative progress across multiple industries. Diligent optimization is essential to address these challenges and ensure the effective 3D printing of composite materials. [3]. Case studies on nanocomposites, short, and continuous fiber composites, highlighting their improvements in mechanical, thermal, and electrical properties

over neat polymers, are among the challenges that additive manufacturing faces when preparing composite feedstock filaments and printing processes for thermoplastic composites [4]. In the past few years, it has been developed to produce prototypes; however, its advantages have prompted researchers to deeply investigate and explore the possibility of manufacturing functional parts. The use of 3D-printed carbon fiber-reinforced thermoplastic composites shows the potential to replace metal parts with lightweight, strong, customizable components, and better mechanical properties [5].

As per the required specific mechanical properties and their applications, many researchers have studied and implemented different composite materials for the manufacturing of particular functional parts-that are currently manufactured using conventional methods. This study investigates the use of carbon fiber polymer composites in 3D printing for centrifugal pump impellers, and suggests that corrosion-resistant materials may eventually replace metal components in manufacturing applications. Pump parts that are resistant to corrosion can be fabricated from composite materials [6]. Because of their lightweight design and mechanical capabilities, 3D-printed fiber-reinforced composites have replaced metals and thermosetting polymers. The study evaluated 3D-printed nylon/carbon fiber composites, showing improved flexural strength and stiffness compared to metals, supporting the replacement of metal parts with 3D-printed carbon fibers. The nylon + CF composite exhibited increased deflection, lower flexural strength, and elastic carbon fiber behavior. The primary disadvantages of nylon composite material are fiber breakage, pull-out, and void formation [7].

Compared to their pure polymer equivalents, 3D-printed objects have greater mechanical strength, rigidity, and toughness owing to the use of composite materials. The study showed enhanced mechanical properties by integrating carbon fibers into fused deposition modeling (FDM) filaments, enabling the replacement of metal parts with 3D-printed carbon fiber components in various industries[8]. Composites are becoming increasingly important in a range of applications requiring high-performance materials and products in industries such as aircraft, marine, automotive, construction, missiles, and defense technology. Oriented fiber composites have garnered significant attention owing to their advantageous properties, which include a high strength-to-weight ratio and controllable anisotropy. The use of short, aligned fiber filaments in high-performance 3D-printed composites has demonstrated great potential for improving the mechanical and thermal properties of objects [9]. An important consideration in determining the quality of a finished product is the adjustment of the process parameters in accordance with the requirements.

The raster angle, print speed, layer thickness, repetitive heating and cooling, print platform temperature, infill speed, nozzle temperature, and nozzle diameter are the main factors that affect composites processed with AM. During processing, particular attention must be paid to the raster angle, the infill speed, the layer thickness, and the nozzle temperature. Young's modulus, toughness, ductility, and ultimate product strength are all directly affected by the raster angle and layer thickness. The on-edge and flat orientations showed the highest tensile and flexural strengths and stiffnesses, whereas the upright orientation had the lowest [10]. The placement of the carbon fiber layer can significantly improve the durability of the 3D-printed components. It is also evident that the carbon fiber layer can be strategically positioned to control the occurrence and spread of cracks. The fiber angle was the most influential parameter (54.13%) followed by the infill density (16.25%), and the carbon layer had a smaller effect on the mechanical properties. The optimum parameters were a 0° fiber angle and 80% infill density [11]. It can be said that the three infill patterns (triangular, grid, and line) that were examined performed similarly in the mechanical tests (three-point bending and tensile).

Therefore, it can be said that the three-point bending and tensile performance are not significantly affected by the core configuration and from a mechanical test perspective, the composite filaments containing 15% chopped carbon fiber demonstrated better results than regular filaments (ABS, PLA, PETG), which are currently used in 3D printing. The Core configuration did not significantly affect the performance of mechanical properties [12]. The average tensile strength of the specimen with the carbon fiber oriented at 0° was approximately 47.83% higher than that of the 90° sample. Following a tensile test, the specimen displayed distinct breaking areas; for example, the 90° direction specimen displayed breaking in the gauge length area, whereas the 0° direction specimen displayed fracture propagation in the curvature area. According to this, the printing path and post-processing are important because they may cause reduced fiber density and improper fiber orientation in some places, which would reduce the strength of such areas [13]. Studying the printing configurations for 3D printing indicated that, compared to the air gap, raster width, contour number, and contour width, the raster angle influences the majority of the mechanical qualities. After examining the effect of printing orientation at 0°, 45°, and 90° on the 3D printing of carbon fiber-reinforced polyamide,

The tensile strength of the material was enhanced by the 0° printing orientation, which led to a greater resistance to applied tensile loading [14]. PA12-CF showed the largest elongation of any composite material, a notably low roughness, and significantly greater tensile and bending strengths. In comparison to samples made of pure polymer, samples from PETG and PA12-CF appeared significantly more pleasant to touch [15]. To investigate the effects of printing parameters, such as extruder temperature and printing speed, on the mechanical properties of continuous carbon fiber-reinforced nylon matrix composites, a methodical experimental study was created and produced for Fused deposition modeling (FDM) components. The quality of 3D-printed components is affected by both the extruder temperature and printing speed [16]. The deflections, warpages, and residual stresses were significantly affected by CF reinforcement. The Higher residual stresses 14.53 MPa as opposed to 11.75 MPa in pure PA6 specimens were necessary to achieve better 3D printed specimen quality. By lowering the deflections and warpages, CF reinforcement greatly enhanced the 3DP manufacturing performance of PA6/(28%)CF composites [17].

The bed temperature (100 °C) was the most significant parameter compared with the infill density (20%), layer thickness (0.2 mm), extrusion temperature (240 °C), and printing speed (100 mm/s) for the Polyamide 12 (PA12) material [18]. The literature highlights recent trends in composite materials and additive manufacturing methods for improving performance and functional applications. Research indicates that composite materials exhibit high durability, tensile strength, impact strength, and flexural strength, which are suitable for absorbing shocks and functioning in dynamic environments. Additive manufacturing (AM) methods such as fused deposition modeling (FDM), are recognized for producing cost-effective, customized parts for functional industrial uses. This shows the significance of various effective printing parameters such as layer height, infill density, printing speed, raster angle, and other parameters, on the mechanical performance of 3D-printed composites. Studies have suggested that composite materials reduce material waste compared with traditional manufacturing. This study aims to investigate a perfect set of optimized parameters that can produce high-strength 3D-

printed functional parts using a minimum amount of material, reduce weight, excellent rigidity and strength, heat resistance, low warpage and low water absorption, which aligns with the selection of polyamide 12 with 20% carbon fiber (PA12/CF). It is the most suitable replacement material for tool steel materials when manufacturing certain automobile components like jigs and fixtures

II. EXPERIMENTAL DETAILS

As per a detailed study of the literature and the application of polyamide carbon fiber composite material for manufacturing functional parts, various process parameters, layer height, infill density, and printing speed have been considered to study their effect on tensile strength, impact strength, and flexural strength, whereas few parameters were kept constant as shown in Table 1. Flexural strength and impact strength are important for the manufacturing and application of jig components. Therefore, we did not give more importance to the tensile strength. In addition, the layer height, infill density and printing speed are three Key Parameters in the FDM Process. According to the research study and experiments, other parameters do not have a significant effect on the flexural and impact strengths. The levels and values of the selected fused deposition modeling (FDM) process parameters are presented in Table 2.

Table 1: Constant Process Parameters.

Printing Temperature	Bed Temperature	Wall Thickness	Top Layers	Bottom Layers	Raster Angle	Infill Pattern	Cooling Fan Speed
250°C	60°C	1.2 mm	3	3	0°	Cubic	100%

Source: Authors, (2025)

Table 2: Fused deposition modeling (FDM) Process Parameters and Levels

Parameter	Level 1 (Low)	Level 2 (High)
Layer height (mm)	0.1	0.3
Infill Density (%)	20	80
Printing speed (mm/sec)	40	100

Source: Authors, (2025)

II.1 MATERIALS AND EQUIPMENT

The material used in the present study was a polyamide carbon fiber (PA12-CF) filament of 1.75 mm diameter with 80% polyamide and 20% carbon fiber. Polyamide 12 (PA12) carbon fiber composites offer a balance of strength, durability, and dimensional stability owing to carbon fiber reinforcement. The key mechanical properties include high tensile strength, flexural strength, and impact resistance, particularly when reinforced with carbon fibers. The 3D printer used to manufacture the specimen is the Omega 300 3D printer with a stainless steel nozzle of 0.4 mm diameter. This 3D printer is capable of using a wide range of materials such as PLA, ABS, polycarbonate (PC), CPE (copolyester), nylon and composite materials. Product 3D modeling was performed on SolidWorks and -.stl files in the 3D printing software Cura Ultimaker. The properties of the filament material are listed in Table 3.

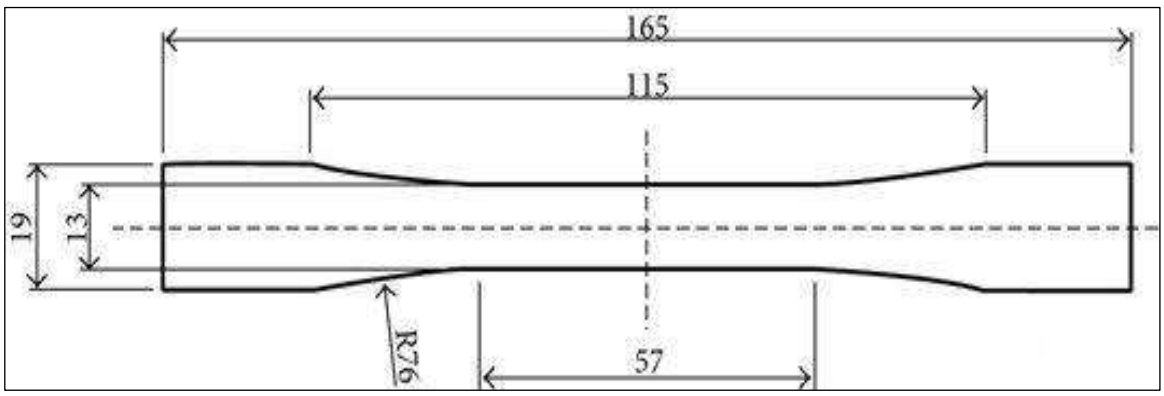
Table 3: Polyamide carbon fiber Material Properties.

Material	PA12-CF (80% Polyamide, 20% Carbon Fiber)
Filament Diameter (mm):	1.75
Density (g/cm3):	1.2
Elongation at Break (%):	10.61
Melt Flow Index (g/10min):	8(250°C/5kg)
Dimensional Accuracy (mm):	170±0.1
Roundness Accuracy (mm):	±0.5
Melting Point (°C)	180

Source: Authors, (2025)

II.2 SAMPLE PREPARATION

The specimens used in this study to evaluate the layer height, infill density, and printing speed were designed in Solidworks based on ASTM D638 for tensile strength (Figure. 1), ASTM D256 for Izod impact (Figure. 2), and ISO 178 for flexural strength (Figure. 3).



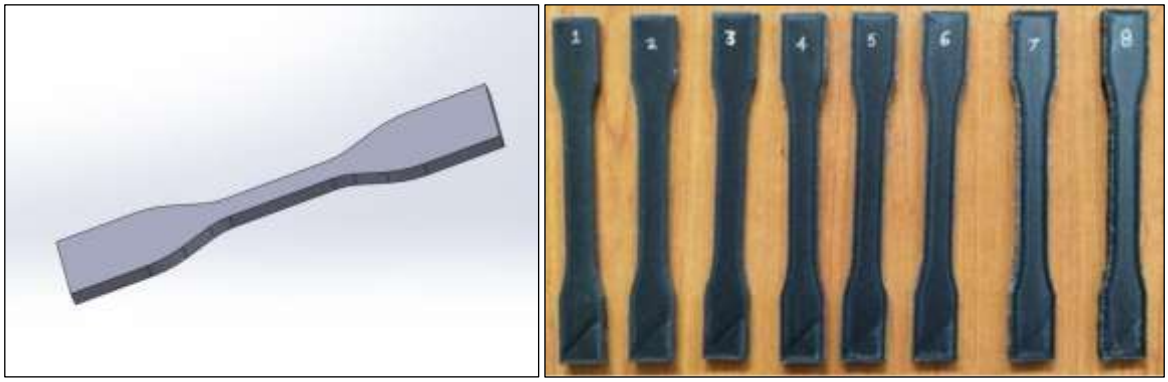


Figure 1: Tensile Strength testing Samples as per ASTM D638 Standard.
Source: Authors, (2025)

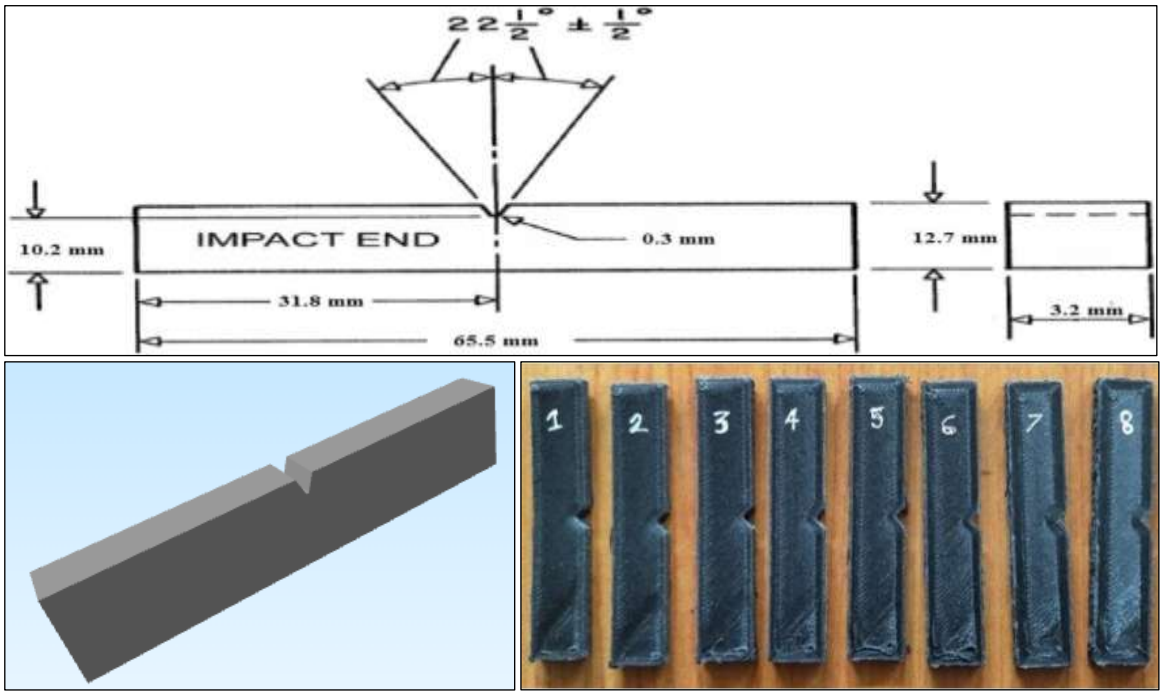


Figure 2: Impact Strength testing Samples as per ASTM D256 Standard.
Source: Authors, (2025)

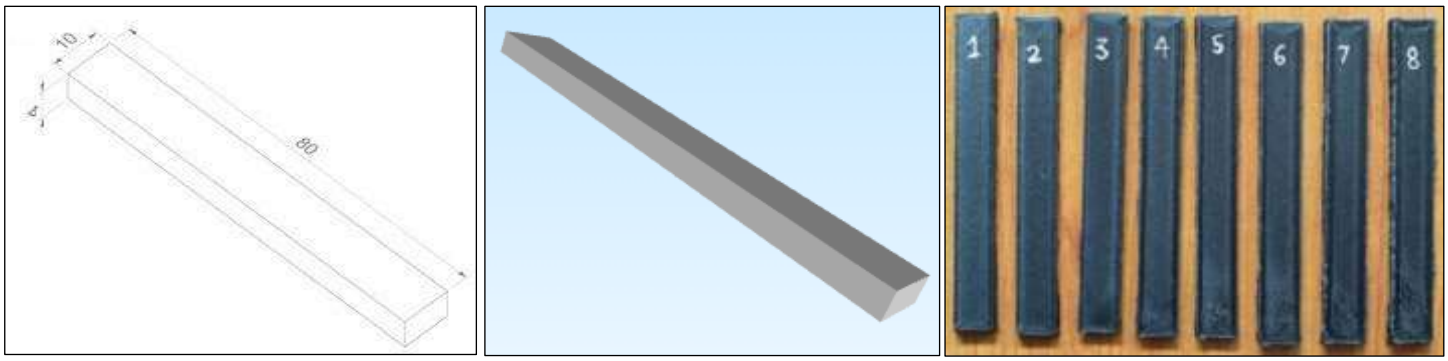


Figure 3 : Flexural Strength testing Samples as per ISO 178 Standard.
Source: Authors, (2025)

II.3 DESIGN OF EXPERIMENT (DOE) AND EXPERIMENTATION SETUP

Design of experiments (DOE) is a branch of statistics that aids in efficiently organizing, planning, and carrying out experiments while enhancing productivity. The focus of the literature is to study composite materials, which are useful for manufacturing functional parts with specific process parameters and better mechanical properties. As per the application, layer height, infill density, and printing speed were selected for this study. The layer height affects surface quality and mechanical performance. The infill density provides the strength and weight of the internal structure. The printing speed affects the production time and bonding quality.

The effects of these three parameters on the tensile, impact, and flexural strengths of the specimens were investigated independently. Tensile tests were performed using a universal testing machine (UTM) to the ASTM D638 standard. The Impact strength was tested using an Izod impact tester following the ASTM D256 standard. The flexural strength was tested using a 3-point bending test method following the ISO 178 standard. The values of the processing parameters listed in Table 4 were used to set an experiment plan to fabricate a total number of 8 samples for tensile, impact, and flexural tests.

Table 4: Design of Experiment.

Specimen	Layer Height(mm)	Infill Density (%)	Printing speed (mm/sec)
1	0.1	20	40
2	0.1	20	100
3	0.1	80	40
4	0.1	80	100
5	0.3	20	40
6	0.3	20	100
7	0.3	80	40
8	0.3	80	100

Source: Authors, (2025)

III. RESULTS AND DISCUSSIONS

The experimental setup for the tensile, impact, and flexural strengths is shown in Figure. 4. The tensile, impact, and flexural strengths of the specimens are listed in Table 5.

Table 5: Results for Tensile, Impact and Flexural Strength.

Specimen	Layer Height (mm)	Infill Density (%)	Printing Speed (mm/sec)	Tensile Strength (MPa)	Impact Strength (J/m)	Flexural (MPa)
1	0.1	20	40	10.66	132.16	16.25
2	0.1	20	100	9.15	190.4	13.47
3	0.1	80	40	17.52	294.6	23.41
4	0.1	80	100	21.3	327.93	23.56
5	0.3	20	40	12.49	163	16.83
6	0.3	20	100	11.01	114.21	14.25
7	0.3	80	40	18.61	190.08	25.22
8	0.3	80	100	16.23	186.73	19.72

Source: Authors, (2025)

Figure 4: Testing set-up for Specimen of Tensile, Impact and Flexural test.
Source: Authors, (2025)

III.1 TENSILE TEST

The tensile strength of the samples prepared using polyamide 12 carbon fiber material which was manufactured in a flat orientation, was tested using universal testing machines. All values were obtained with a 95% confidence interval. ANOVA analysis was performed using MINITAB software to identify the influence of various components on the tensile strength. The results of the ANOVA are presented in Table 6. The main and interaction effect plots for the tensile strength are shown in Figure. 5 and Figure. 6, respectively.

A higher infill density (80%) significantly improves the tensile strength. A higher printing speed (100 mm/sec) results in a maximum tensile strength of 21.3 MPa. Lower layer height (0.1 mm) tends to produce higher tensile strength compared to 0.3 mm for similar conditions. According to the results, the infill density has a greater contribution to increasing the tensile strength than the other selected parameters. Therefore, the combination of 0.1 mm layer height, 80% infill density, and 100 mm/sec printing speed maximizes the tensile strength (Specimen 4, Table 5).

Table 6: Analysis of variance table for tensile test results.

Source of Variation	Sum of Squares	DF	Mean Square	% contribution
Layer Height	0.0105	1	0.0105	0.01%
Infill Density	115.14	1	115.14	85.47%
Printing Speed	0.316	1	0.316	0.23%

Source: Authors, (2025)

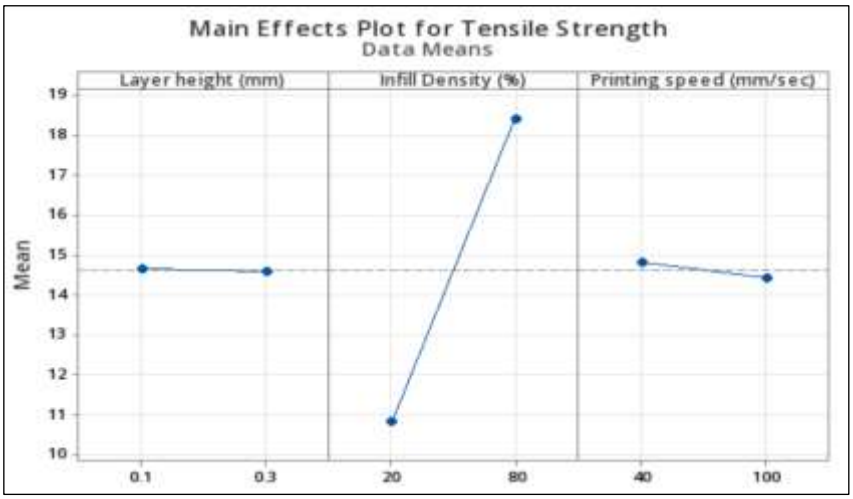


Figure 5: Main effects plot for Tensile strength.
Source: Authors, (2025)

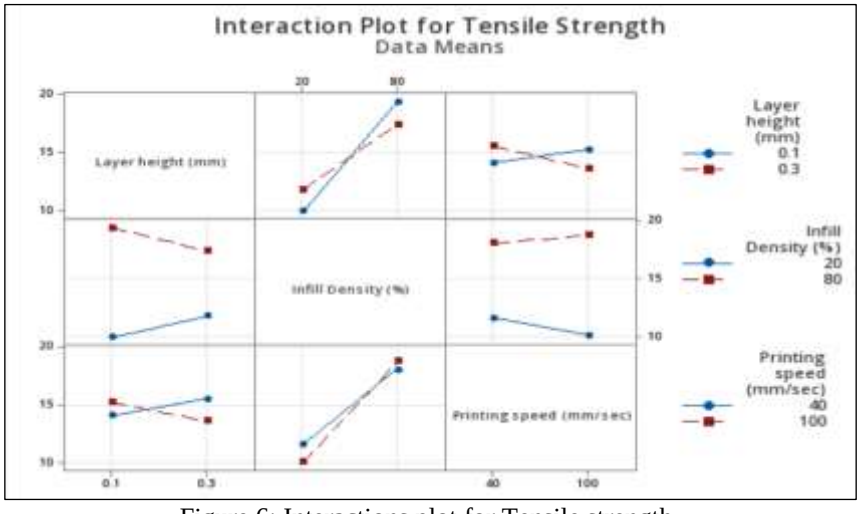


Figure 6: Interactions plot for Tensile strength.
Source: Authors, (2025)

III.2 IMPACT TEST

The impact strength of samples made of polyamide carbon fiber material was manufactured in a flat orientation and tested using the Izod impact test. The test results are listed in Table 5. All values were obtained with a 95% confidence interval. ANOVA was performed using MINITAB software to identify the influence of various components on the impact strength. The results of the ANOVA are presented in Table 7. The main and interaction effect plots for the tensile strength are shown in Figure. 7 and Figure. 8, respectively. A higher infill density (80%) significantly improves the impact strength. A higher printing speed (100 mm/sec) achieved a maximum impact strength of 327.93 J/m. A lower layer height (0.1 mm) provides a higher impact strength at a higher infill and higher speeds.

Therefore, the combination of 0.1 mm layer height, 80% infill density, and 100 mm/sec printing speed maximizes the impact strength (Specimen 4, Table 5).

Table 7: Analysis of variance table for Impact strength.

Source of Variation	Sum of Squares	DF	Mean Square	% contribution
Layer Height	10590.22	1	10590.22	27.15%
Infill Density	19957.02	1	19957.02	51.15%
Printing Speed	194.34	1	194.34	0.50%

Source: Authors, (2025)

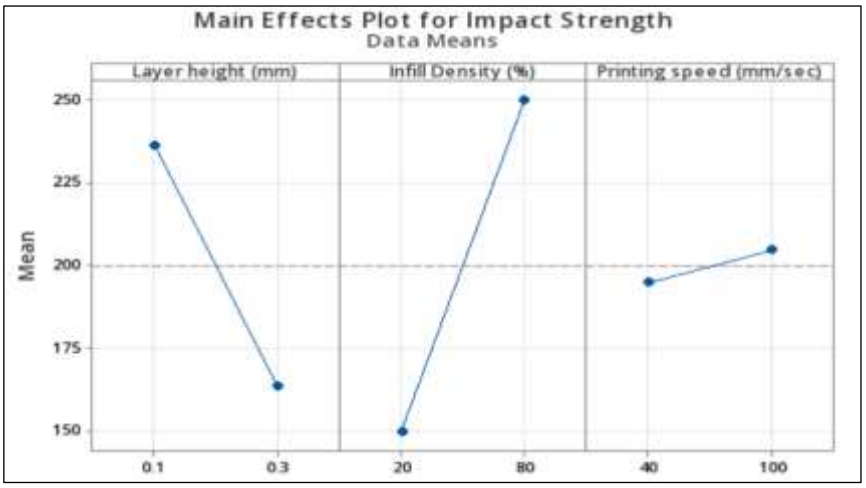


Figure 7: Main effects plot for Impact strength.
Source: Authors, (2025)

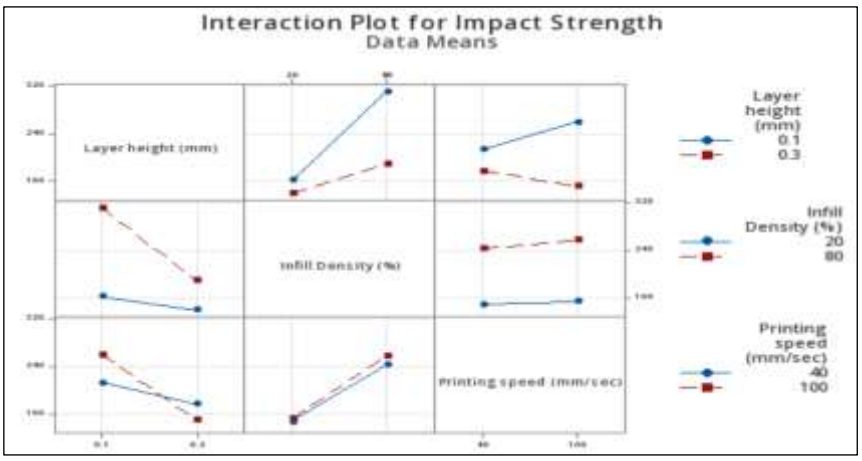


Figure 8: Interactions plot for Impact strength.
Source: Authors, (2025)

III.3 FLEXURAL TEST

The flexural strength of the samples made of polyamide carbon fiber material was determined by a 3-point bending test and manufactured in the edge orientation according to the required application of the jig component. All values were obtained with a 95% confidence interval. ANOVA analysis was performed using MINITAB software to identify the influence of various components on the flexural strength. The results of the ANOVA are presented in Table 8. It was observed that the infill density had a significant effect on the flexural strength. The main and interaction effect plots for the flexural strength are shown in Figure. 9 and Figure. 10 respectively. A higher infill density (80%) improves the flexural strength. A lower printing speed (40 mm/sec) produced a higher flexural strength for most conditions. The combination of 0.3 mm layer height, 80% infill density, and 40 mm/sec printing speed maximized the flexural strength (Specimen 7, Table 5).

Table 8 Analysis of variance table for Flexural strength.

Source of variation	Sum of Squares	df	Mean Square	% contribution
Layer height	0.0561	1	0.0561	0.04%
Infill Density	120.98	1	120.98	83.55%
Printing Speed	14.34	1	14.34	9.90%

Source: Authors, (2025)

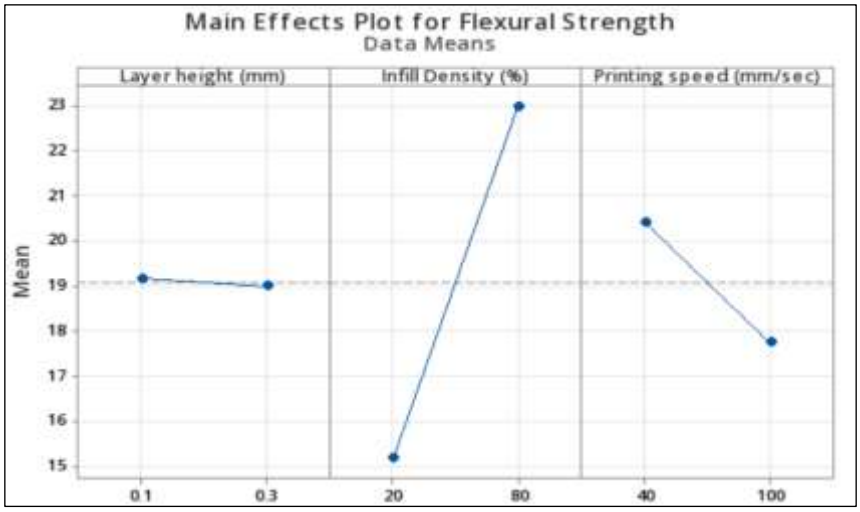


Figure 9: Main effects plot for Flexural strength.
Source: Authors, (2025)

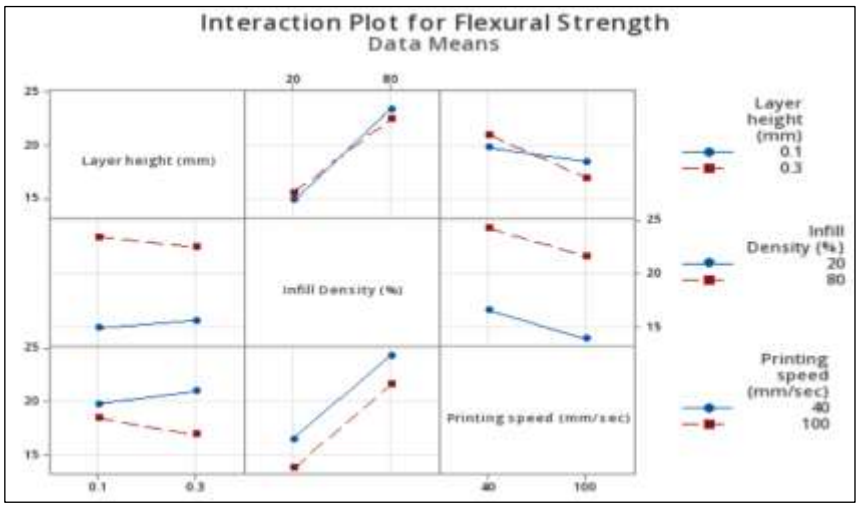


Figure 10: Interactions plot for Flexural strength.
Source: Authors, (2025)

IV. MULTI-OBJECTIVE OPTIMIZATION

IV.1 TOPSIS METHOD

TOPSIS (technique for order preferences by similarity to an ideal solution) is a method used to solve multiple-criteria decisions. TOPSIS can be easily expanded to involve more than two output responses; hence, it is a good statistical tool for mult-objective optimization problems, which is currently applied in the FDM process.The mult-objective optimization problems are transferred to multiple decision-making conditions and then coded to the single-objective problem in terms of relative closeness values. simplicity, rationality, comprehensibility, computational efficiency, and ability to measure the relative performance of each alternative in a simple mathematical form.

Step 1: Construction of the decision matrix:

In the TOPSIS multi objective problem, the averaged output responses for each experiment are the inputs to be placed in a matrix form known as a decision matrix. A decision matrix is created:

$$D = \begin{matrix} A_1 \\ A_2 \\ A_m \end{matrix} \begin{bmatrix} X_{11} & X_{12} \ldots\ldots & X_{1n} \\ X_{21} & X_{22} \ldots\ldots & X_{2n} \\ X_{m1} & X_{m2} \ldots\ldots & X_{mn} \end{bmatrix}$$

Alternatives - A1, A2, and Am.
criteria based on which ranking is performed = C1, C2,…, Cn
qualification of alternative Ai with respect to criterion Cj = Xij
weight of criterion Cj = wj

Step 2: Calculation of the normalized decision matrix:

This step changes various characteristic dimensions into non-dimensional characteristics that permit evaluation across criteria. The two criteria of functions are the benefit function (larger is better) and cost function (smaller is better). The decision matrix must be converted to a normalized scale using the following equation:

$$n_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i-1}^{m} x_{ij}^2}}$$

where, i = 1, 2,…, m, j = 1, 2,…, n.

Step 3: Calculation of the weighted normalized decision matrix:

It is done using equation:

$$v_{ij} = w_j \times \eta_{ij}, i = 1, 2, \cdots ., m; j = 1, 2, \ldots ., n.$$

where ,
weight response is represented using criterion

$$\sum_{j=1}^{n} w_j = 1$$

Step 4: Determination of the positive ideal and negative ideal solutions:

The ideal positive solution is the result of enlarging the benefit criteria and reducing the cost criteria whereas the negative ideal solution enlarges the cost criteria and reduces the benefit criteria.
The values are given by the following equation:

$$\begin{aligned} Positive\ ideal\ solution, A^+ &= (v_1^+, v_2^+, \ldots, v_n^+) \\ &= \left((max_i v_{ij} | j \in I), (min_i v_{ij} | j \in \mathcal{J})\right) \end{aligned}$$

$$\begin{aligned} Negative\ ideal\ solution, A^- &= (v_1^-, v_2^-, \ldots, v_n^-) \\ &= \left((min_i v_{ij} | j \in I), (min_i v_{ij} | j \in \mathcal{J})\right) \end{aligned}$$

Where.
I is correlated with Benefit Criteria and J is correlated with Cost Criteria.
Step 5: Calculate the segregation measures from the positive ideal solution (di+) and negative ideal solution (di-) using the following equations :

$$d_i^+ = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^+\right)^2}, i = 1, 2, \ldots, m$$

$$d_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - v_j^-\right)^2}, i = 1, 2, \ldots, m$$

Step 6: Calculate the segregation measures from the positive ideal solution (di+) and negative ideal solution (di-) using the following equations :

$$R_i = \frac{d_i^-}{d_i^- + d_i^+}, where\ 0 \leq R_i \leq 1, i = 1, 2, \ldots, m$$

Step 7: Rank the preference order:

The ranking order to be deployed with respect to the Ri value is in descending order from 0 to 1. The value closest to 1 attained the first rank, and that attribute should be selected as the best among multiple decision-making conditions. As per the TOPSIS method, the ranking of the specimens for tensile, impact and flexural strengths is listed in Table 9.

Table 9: Ranking of specimens According to Tensile, Impact and Flexural strength using TOPSIS analysis.

No.	Layer Height (mm)	Infill Density(%)	Printing Speed (mm/sec)	Tensile Strength(MPa)	Impact Strength (J/m)	Flexural Strength(MPa)	di+	di-	pi	Rank
1	0.1	20	40	10.66	132.16	16.25	0.28	0.14	0.333	7
2	0.1	20	100	9.15	190.4	13.47	0.28	0.22	0.439	4
3	0.1	80	40	17.52	294.6	23.41	0.33	0.36	0.526	2
4	0.1	80	100	21.3	327.93	23.56	0.36	0.41	0.536	1
5	0.3	20	40	12.49	163	16.83	0.27	0.18	0.397	6
6	0.3	20	100	11.01	114.21	14.25	0.28	0.12	0.294	8
7	0.3	80	40	18.61	190.08	25.22	0.26	0.21	0.445	3
8	0.3	80	100	16.23	186.73	19.72	0.27	0.21	0.438	5

Source: Authors, (2025)

di+ separation of each alternative from the positive ideal
di− separation of each alternative from the negative ideal
Pi - relative closeness to the ideal solution;, a higher Pi corresponds to a higher rank

IV.2 GRA METHOD

Step 1: Grey relational generation (Normalization of Raw Data)

The experimental data obtained for the output responses can be normalized according to different requirements. There are three different normalization conditions: lower-the-better, higher-the- better and nominally the best criteria.
For lower-the-better: For higher-the-better:

$$X_i(k) = \frac{\max Y_i(k) - Y_i(k)}{\max Y_i(k) - \min Y_i(k)} \qquad X_i(k) = \frac{Y_i(k) - \min Y_i(k)}{\max Y_i(k) - \min Y_i(k)}$$

where - Xi is the value obtained after grey relational generation.
Min Yi (k) is the smallest value of Yi and Max Yi (k) is the maximum value of Yi (k).

Step 2: Grey relational coefficient :

$$\gamma_i = \frac{\Delta_{min} + \zeta\Delta_{max}}{\Delta_{oi}(k) + \zeta\Delta_{max}}$$

where γ is the Grey relational coefficient of the ith experiment for the kth response.
Δoi=||Xo (k)-Xi(k)||=difference between absolute value, ζ the distinguishing coefficient and it is defined in the range $0 < \zeta \leq 1$,
Δmin= min|| x0(k)-xj(k)|| is the smallest value of Yj(k) and
Δma x= ma x||x(k)-xj(k)|| is the largest value of Yj(k).

Step 3: Grey relational grade:

The grey relational grade was calculated by averaging the grey relation coefficients corresponding to each experiment. The overall response of multiple performance characteristic is based on the grey relational grade which is the objective function.

$$\alpha_i = \frac{1}{n}\sum_{k=1}^{n} Y_i(k)$$

A higher grade implies that the corresponding parameter combination is closer to the optimal set of input parameters. As per the GRA method, the ranking of the specimens for tensile, impact and flexural strengths is listed in Table 10.

Table 10: Ranking of specimens According to Tensile, Impact and Flexural strength using GRA Method.

No.	Layer Height (mm)	Infill Density (%)	Printing Speed (mm/sec)	Tensile Strength (MPa)	Impact Strength (J/m)	Flexural Strength (MPa)	Grey Relational Grade	Rank
1	0.1	20	40	10.66	132.16	16.25	0.5562	6
2	0.1	20	100	9.15	190.4	13.47	0.5520	7
3	0.1	80	40	17.52	294.6	23.41	1.0716	2
4	0.1	80	100	21.3	327.93	23.56	1.3898	1
5	0.3	20	40	12.49	163	16.83	0.8654	4
6	0.3	20	100	11.01	114.21	14.25	0.5267	8
7	0.3	80	40	18.61	190.08	25.22	1.0649	3
8	0.3	80	100	16.23	186.73	19.72	0.7462	5

Source: Authors, (2025)

IV.3 OPTIMUM PROCESS PARAMETER FOR TOPSIS AND GRA (RANK 1 TO 3)

As per the TOPSIS and GRA rankings and analysis, the optimum process parameters are identified and are similar for the Ranks 1 to 3, as mentioned in Table 11. Rank 1 Specimen was implemented for the 3D printing of Jig components.

Table 11: Optimum Process Parameters.

Method	Rank	Layer Height (mm)	Infill Density (%)	Printing Speed (mm/sec)	Tensile Strength (MPa)	Impact Strength (J/m)	Flexural Strength (MPa)
TOPSIS	1	0.1	80	100	21.3	327.93	23.56
GRA	1	0.1	80	100	21.3	327.93	23.56
TOPSIS	2	0.1	80	40	17.52	294.6	23.41
GRA	2	0.1	80	40	17.52	294.6	23.41
TOPSIS	3	0.3	80	40	18.61	190.08	25.22
GRA	3	0.3	80	40	18.61	190.08	25.22

Source: Authors, (2025)

IV.4 FDM 3D PRINTED FINAL PART

(Jig Component for LOGO Printing in Automobile Industry as per optimum Process Parameters Rank 1)

Based on the tested results and optimization of all the above mechanical properties, it is concluded that the polyamide 12 carbon fiber composite material is strongly recommended and used for tooling and functional prototypes in the automotive and engineering industries. It is used for manufacturing functional components that require a higher flexural strength in the edge orientation and a higher impact strength in a flat orientation. The density of the PA12CF Material was much lower than that of tool steel material used in conventional machining. Therefore, it can be used for the development of a tool for automobile assembly lines for the manufacture of Jig components which is used for LOGO Printing in the Automobile Industry, and it also reduces the weight, cost of components, and manufacturing process time. Therefore, it is recommended to use PA12 CF composite material for manufacturing 3D Printed parts which is required for specific applications with good mechanical properties such as flexural strength. It has been successfully implemented in real life applications in the automobile Industry for LOGO Printing. The design of jig components for LOGO Printing and the final 3D Printed components are shown in Figure. 11(a) and (b), respectively. Manufacturing and testing were performed in an authorized research institute /laboratory.

Figure 11: (a) Design of Jig component for LOGO Printing (b) 3D Printing Component.
Source: Authors, (2025).

V. FINAL CONCLUSION

In this study, three important process parameters—infill density, layer height, and printing speed—were selected for the fused deposition modeling (FDM) process. The effects of process parameters on the tensile, impact, and flexural strength of the specimens were investigated using L8 Taguchi orthogonal array and ANOVA approaches. The specimens were fabricated using a 3D printer according to ASTM and ISO standards.

This study examines the effect of fused deposition modeling (FDM) process parameters on the characteristics of functional parts and their application to specific requirements of higher flexural strength and impact strength. The experiment investigated the effect of the infill density, layer height, and printing speed on the mechanical properties of the parts. The following conclusions were drawn:

1) Infill density (80%) was the most critical factor for all three mechanical properties.
2) Printing speed has varying effects:
   - o A higher speed (100 mm/sec) favors tensile and impact strength.
   - o A lower speed (40 mm/sec) favors the flexural strength.
3) Layer Height has varying effects:
   - o A lower layer height (0.1 mm) is better for tensile and impact strength.

o A higher layer height (0.3 mm) was preferred for the flexural strength.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Rahul Patel, Dr. Suketu Jani.
Methodology: Rahul Patel, Dr. Suketu Jani.
Investigation: Rahul Patel, Dr. Suketu Jani.
Discussion of results: Rahul Patel, Dr. Suketu Jani.
Writing – Original Draft: Rahul Patel
Writing – Review and Editing: Rahul Patel, Dr. Suketu Jani.
Resources: Rahul Patel, Dr. Suketu Jani.
Supervision: Dr. Suketu Jani
Approval of the final text: Rahul Patel, Dr. Suketu Jani.

VII. REFERENCES

[1] Divya Zindani , Kaushik Kumar, An insight into additive manufacturing of fiber reinforced polymer composite,International Journal of Lightweight Materials and Manufacture ,Volume 2, Issue 4, December 2019, 267-278,https://doi.org/10.1016/j.ijlmm.2019.08.004

[2] Abdallah S, Ali S, Pervaiz S. Performance optimization of 3D printed polyamide 12 via multi jet fusion: a Taguchi grey relational analysis (TGRA). Int J Light Mater Manuf. 2023;6:72-81. doi:10.1016/j.ijlmm.2022.05.004

[3] Hasan Rafsan Jania and Md Zillur Rahmanb, Modeling and simulation of additively manufactured composites, Comprehensive Materials Processing, 2024 https://doi.org/10.1016/B978-0-323-96020-5.00230-2

[4] Pavan Kumar Penumakala , Jose Santo, Alen Thoma,A critical review on the fused deposition modeling of thermoplastic polymer composites , Composites Part B 201 (2020) 108336,https://doi.org/10.1016/j.compositesb.2020.108336

[5] Nidhi Dixit, Prashant K. Jain, 3D printed carbon fiber reinforced thermoplastic composites: A review,Materials Today: Proceedings 43 (2021) 678–681. https://doi.org/10.1016/j.matpr.2020.12.609

[6] Gabriel Mansour , Vasileios Papageorgiou ,Dimitrios Tzetzis, Carbon Fiber Polymer Reinforced 3D Printed Composites for Centrifugal Pump Impeller Manufacturing. Technologies 2024, 12, 48. https://doi.org/10.3390/technologies12040048

[7] Ibrahim M. Alarifi, A performance evaluation study of 3d printed nylon/glass fiber and nylon/carbon fiber composite materials. Journal of Material Research and Technology 2022 https://doi.org/10.1016/j.jmrt.2022.09.085.

[8] Antreas Kantaros , Evangelos Soulis , Florian Ion Tiberiu Petrescu ,Theodore Ganetsos, Advanced Composite Materials Utilized in FDM/FFF 3D Printing Manufacturing Processes: The Case of Filled Filaments. Materials 2023, 16, 6210. https://doi.org/10.3390/ma16186210.

[9] Arunachalam Ramanathan, Varunkumar Thippanna, Barath Sundaravadivelan, Yuxiang Zhu, Abhishek Saji Kumar, Dharneedar Ravichandran,| Sui Yang, Kenan Song, Highly loaded carbon fiber filaments for 3D-printed composites. J Polym Sci. 2024;62:2670–2682.DOI: 10.1002/pol.20230632.

[10] Manickam Ramesh,Kanakaraj Niranjana,Ramasamy Bhoopathi, and Lakshminarasimhan Rajeshkumar, Additive manufacturing (3D printing) technologies for fiber-reinforced polymer composite materials: A review on fabrication methods and process parameters, e-Polymers 2024; 24: 20230114 https://doi.org/10.1515/epoly-2023-0114.

[11] Basim El Essawi , Said Abdallah, Shafahat Ali , Abdallah Nassir Abdo Mohammed,Rahmat Agung Susantyoko, Salman Pervaiz Optimization of infill density, fiber angle, carbon fiber layer position in 3D printed continuous carbon-fiber reinforced nylon composite ,Results in Engineering 21 (2024) 101926, https://doi.org/10.1016/j.rineng.2024.101926.

[12] M A Pop, S M Zaharia , L A Chicos, C Lancea, V M Stamate, G R Buican and I S Pascariu, Effect of the infill patterns on the mechanical properties of the carbon fiber 3D printed parts. IOP Conf. Series: Materials Science and Engineering 1235 (2022) 012006 doi:10.1088/1757-899X/1235/1/012006.

[13] Nathathai Saithongkum, Karuna Tuchinda, Study of Properties of 3D Printed Short Carbon Fiber Key Engineering Materials ISSN: 1662-9795, Vol. 841, pp 182-187 doi:10.4028/www.scientific.net/KEM.841.182.

[14] Nisa Naima Khalid,Nabilah Afiqah Mohd Radzuan,Farhana Mohd Foudzi,Abu Bakar Sulong,Nurul Najwa Abd Rahman, Interfacial Bonding of Polyamide -Reinforced Carbon fibre Via Additive Manufacturing, Materiali in tehnologije / Materials and technology 58 (2024) 5, 587–595.

[15] Justyna Rybarczyk, Filip Górski, Wiesław Kuczko, Radowsław Wichniarek , Sabina Siwiec , Nikola Vitkovic , Răzvan PăcurarMechanical Properties of Carbon Fiber Reinforced Materials for 3D Printing of Ankle Foot Orthoses , Advances in Science and Technology Research Journal 2024, 18(4), 191–215 https://doi.org/10.12913/22998624/188819.

[16] Mahmoud Moradi, Zeinab Malekshahi Beiranvand,Nahid Salimi, Saleh Meiabadi and Jonathan Lawrence, Experimental Investigation on the 3D Printing of Nylon Reinforced by Carbon Fiber through Fused Filament Fabrication Process, Effects of Extruder Temperature, and Printing Speed.,International Journal of Polymer Science, Volume 2024, Article ID 1234797, 12 pages https://doi.org/10.1155/2024/1234797.

[17] Ans AL RASHID, Hamid IKRAM, Muammer KOÇ, Effect of carbon fiber reinforcement on dimensional variations of 3D printed polyamide-6 composites: A simulation study,Turkish Journal of Chemistry,(2023) 47: 33-39 doi:10.55730/1300-0527.3513 .

[18] Hussein Alzyod, Peter Ficzere,Optimizing fused filament fabrication process parameters for quality enhancement of PA12 parts using numerical modeling and Taguchi method ,Heliyon 9 (2023) e14445,https://doi.org/10.1016/j.heliyon.2023.e14445

Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11 n.55, p. 204-211. September/October, 2025.
DOI: https://doi.org/10.5935/jetia. v.11 n.55.2649

RESEARCH ARTICLE OPEN ACCESS

A FRAMEWORK TO ENHANCE DATA QUALITY IN MASTER DATA MANAGEMENT PROCESS: A HEALTHCARE STUDY

Chafika Benkherourou[1], Abdelhabib Bourouis[2]

[1] Department of Computer Science, Batna 2 University, 05000, RN3 Batna, Algeria.
[2] Research Laboratory on Computer Science's Complex Systems, Larbi Benmhidi University, Constantine Road, 04000, Oum El Bouaghi, Algeria.

[1]https://orcid.org/0009-0000-9856-4645, [2]https://orcid.org/0000-0002-4592-4042

Email: cbenkherourou@gmail.com, bourouis.abdelhabib@univ-oeb.dz

ARTICLE INFO

Article History
Received: September 2, 2025
Revised: September 30, 2025
Accepted: October 4, 2025
Published: October 31, 2025

Keywords:
Master Data,
Master Data Management,
MDM,
Data Quality,
Business Process Modeling Notation.

ABSTRACT

For modern enterprises, to guarantee the quality and accuracy of their fundamental business data, master data management (MDM) is essential. However, a lack of connectivity between data governance and data quality throughout deployment is the reason why many MDM programs fail. This paper introduces a six-phase MDM framework that specifically integrates stakeholder-driven governance and data quality dimensions into each phase of the MDM lifecycle. Based on the PRAXEME methodology and designed using Business Process Model and Notation (BPMN), the framework emphasizes real-time monitoring and facilitates departmental collaboration. A healthcare case study illustrates its efficacy in enhancing data comprehensiveness, minimizing redundancy, and boosting operational productivity. The proposed framework offers a comprehensive approach that aligns business and IT objectives and reinforces process transparency.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

In last years, the volume of organizational data has expanded significantly with the proliferation of advanced technologies such as cloud computing, IoT, and AI-driven applications that capture information in multiple formats. Much of this data is structured and can be classified into four main categories: master data, historical data, transactional data, and metadata. Among these, master data is the highest priority because it contains the critical information that defines and supports the organization's core operations [1, 2]. Master data are the data describing the most relevant business entities, on which the activities of an enterprise are based, e.g. suppliers, customers or products [1]. In contrast to transactional data (invoices, orders, etc.) and inventory data, master data are oriented towards the attributes. They describe the main characteristics of objects in the real world [3]. Single master data entities are rarely being changed and are relatively constant, especially if they are compared with transactional data. Master data acts as the essential basis for transactional data. Without it, no order or delivery can take place.

In modern digital organizations, the rapid expansion of data sources and IT systems makes it increasingly difficult to maintain a single, reliable view of critical business entities, such as customers, products, suppliers, and assets, which often end up duplicated and inconsistently managed across fragmented infrastructures. This data fragmentation harms not only day-to-day operations but also strategic decision-making and customer satisfaction. As a result, Master Data Management (MDM) has become critical to ensure the consistency, governance, and integrity of foundational business data [4].

There are many reasons why a company might consider MDM, which is defined as "the management of the consistent and uniform subset of business entities that describe the core activities of an enterprise" [1]. Using Master Data Management, valuable master data from multiple organizations is identified and consolidated into a central repository. This centralized system acts as a "single source of truth" supporting various applications across different government entities [5]. While several well-known frameworks for Master Data Management (MDM) exist, such as Gartner's Seven Building Blocks of MDM and the DAMA-DMBOK (Data Management Body of Knowledge), they often take a rather top-down and technical perspective. These models tend to concentrate on aspects like system

architecture, governance policies, or abstract principles. What they frequently lack is practical, hands-on guidance on how to deal with real-world challenges, such as ensuring data quality in real time, modeling processes across departments, or involving business and IT stakeholders in a meaningful way. This lack of operational integration often results in MDM initiatives that are difficult to sustain or scale.

To bridge these gaps, this study introduces a six-stage MDM framework that emphasizes continuous data quality control and governance into every step of the data lifecycle. The innovation lies in its holistic approach, which combines Business Process Model and Notation (BPMN) with ongoing metrics for data quality, ensuring clear process visualization and active involvement from both business and IT stakeholders. Unlike traditional frameworks that treat quality and governance as peripheral tasks or post-implementation concerns, this methodology embeds them structurally and continuously into the core of MDM operations. To validate the effectiveness of the framework, a case study is conducted within a major healthcare organization, demonstrating its applicability and added value in a complex data environment. The paper is organized as follows. A literature review on MDM is presented in section 2. The research design and framework development is discussed in section 3, while detailed presentation of the methodology's healthcare sector application, illustrated by case study results are presented in section 4. Section 5 explains about results and discussion. Finally, section [6] concludes the paper and suggests future research directions.

II. LITERATURE REVIEW

The dependability of master data is fundamentally supported by data quality. As noted by [7], 80% of businesses recognize that bad data quality adversely affects their financial performance. In an increasingly regulated business environment, creating a single source of truth has become essential for organizations to stay compliant while remaining competitive [2]. Good data quality is not just a requirement for successfully adopting Master Data Management systems, it's also one of the key benefits organizations can gain when MDM is implemented effectively [1]. Master Data Management is essential for organizations seeking a unified and reliable source of key business data. Many MDM initiatives fail, not due to a sudden collapse, but through a gradual breakdown. These failures often result in poor decision-making, wasted resources, missed opportunities, and inconsistent customer experiences driven by conflicting versions of the truth. Research shows that MDM failure is rarely caused by a single factor.

II.1 DATA QUALITY ISSUES AS A PRIMARY CAUSE

Poor data quality, characterized by inaccuracies, inconsistencies, duplicates, and incompleteness, undermines the reliability of master data repositories. As highlighted by [8], data quality is not merely a technical concern but constitutes a core organizational challenge that necessitates a structured and systematic approach to management. Accordingly, addressing data quality challenges requires the implementation of structured training programs and sustained internal communication strategies, which are essential for fostering organizational awareness and adherence to data quality standards across the MDM lifecycle.

Furthermore, [9] underscores the critical role of continuous quality assessments in ensuring the integrity of master data systems and regular deduplication scans in maintaining high data quality over time. Common sources of poor data quality include fragmented data sources, outdated legacy systems, and inconsistent data entry practices across business units [10]. Additionally, [11] observes that inconsistencies in meeting functional requirements within MDM applications significantly contribute to inefficiencies and data duplication.

II.2 INSUFFICIENT DATA GOVERNANCE FRAMEWORKS

Failures in data governance frequently lead to ambiguous data ownership, poorly defined stewardship responsibilities, and insufficient enforcement of data standards and policies [12]. More recent empirical studies, such as [13], confirm that the absence of robust governance remains a primary barrier to MDM success, as it leads to confusion regarding data ownership and accountability.

The lack of integrated governance frameworks contributes to a disconnect between technical and business teams, causing fragmented and inconsistent data management practices that frequently result in MDM project failure [14]. Furthermore, [15] emphasize that during iterative MDM deployments, misalignment between implemented solutions and organizational requirements frequently originates from deficient governance structures, leading to poor adaptation and integration. According to [16], effective data governance strategies necessitate continuous refinement and are best implemented through an incremental approach, beginning with small-scale initiatives and progressively expanding to broader organizational contexts. To ensure successful implementation, organizations must adopt mechanisms that guarantee the quality of their data assets within the context of their specific operational needs [17].

II.3 TECHNICAL COMPLEXITY AND INTEGRATION CHALLENGES

Lepeniotis [18] emphasizes the detrimental impact of poor planning and sequencing in MDM projects, noting that misaligned execution timelines can exacerbate implementation challenges, ultimately, these challenges can lead to significant financial losses and a decline in overall operational efficiency. Furthermore, [19] argue that MDM initiatives are more likely to fail when approached as purely IT-driven efforts without broader organizational engagement. The absence of stakeholder buy-in frequently gives rise to fragmented and informal governance structures, which in turn obstruct the establishment of consistent data quality practices and compromise the long-term success of MDM initiatives.

While many existing MDM frameworks recognize the importance of data quality and governance, they often treat them as secondary considerations, something to handle later rather than building them in from the start. They also tend to overlook the practical realities of implementation, offering little in terms of tools to help teams align, collaborate, and stay on track. This study takes a different approach: it places data quality and governance at the very heart of the MDM process. By weaving them into every stage, supported by clear visual models using BPMN, it offers a hands-on, collaborative, and sustainable way to manage master data effectively.

III. RESEARCH METHOD

In this section, we detail our proposed methodology. In contrast to existing models that treat BPMN and data quality measurement as retrospective activities, the proposed methodology incorporates these elements from the outset of the MDM implementation process, starting early from strategic alignment and continuing through operational monitoring. This study is derived from the PRAXEME methodology, chosen specifically for its strengths in modeling complex enterprise systems through a multi-perspective and semantically rich approach making it particularly well-suited for Master Data Management projects where conceptual clarity and data traceability are critical [20]. The schematic representation of our methodological approach is illustrated in figure 1.

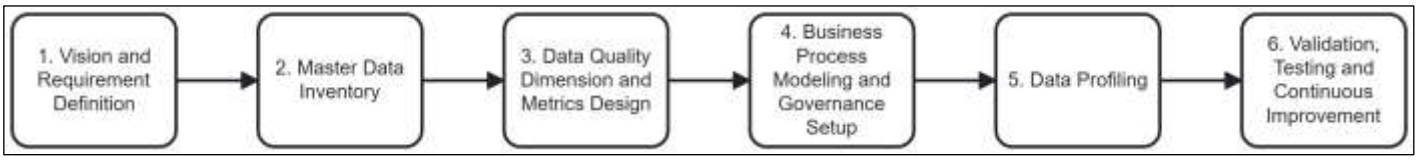


Figure 1: Global methodology.
Source: Authors, (2025).

III.1 VISION AND REQUIREMENT DEFINITION

The success of a Master Data Management initiative begins with a clear vision that aligns business objectives with data governance priorities. Organizations must begin by clearly defining the objectives and anticipated outcomes associated with the implementation of Master Data Management initiatives. During the initial research step, structured and semi-structured interviews with both departmental staff and related units would be conducted to thoroughly analyze the department's operational landscape, uncover critical data challenges, and gather diverse stakeholder perspectives. This process ensures that the MDM framework aligns with actual business needs and operational processes. The interview questions address current data usage, identify existing data issues, and explore expectations for future capabilities. The outcome of this step will be a table that documents the organization's vision, objectives, and key questions.

III.2 MASTER DATA INVENTORY

Stage 2 focuses on identifying and categorizing master data assets and their associated metadata. Organizations must begin by distinguishing between master data, transactional data, and reference data. This identification process typically involves: (1) engaging with key business units to define critical entities, and (2) conducting data source audits to determine the locations and structures in which master data is maintained. The outcome is a comprehensive data dictionary that consolidates master data along with its associated metadata into a unified reference source.

III.3 DATA QUALITY DIMENSIONS AND METRICS DESIGN

Stage 3 focuses on defining how the quality of data will be measured. This stage translates the business goals articulated in stage 1 into concrete, measurable data quality indicators. Each dimension should be mapped to specific business goals. These metrics serve as the foundation for sustained data stewardship, effective exception management, and the continuous improvement of data quality practices. The final output is a structured table featuring four main columns: Dimension, Measures, Calculation Technique, and Defined Threshold.

III.4 BUSINESS PROCESS MODELING AND GOVERNANCE SETUP

The next step is to embed these controls into organizational processes. Stage 4 formalizes the operational workflows by employing Business Process Modeling Notation (BPMN). This approach facilitates shared visibility between business and IT teams regarding the creation, validation, modification, and maintenance of master data across the enterprise. It operationalizes the MDM strategy through structured, visual workflows and clearly defined roles.

III.5 DATA PROFILING, MONITORING, AND EXCEPTION MANAGEMENT

Once the governance structure and process models are in place, organizations must manage the quality of their master data. Stage 5 operationalizes the data quality metrics defined in stage 3 by applying data profiling techniques, detecting anomalies, and implementing exception management workflows. These include: Column analysis, functional dependency discovery and cross-source comparison.

III.6 VALIDATION, TESTING, AND CONTINUOUS IMPROVEMENT

The final stage of the integrated Master Data Management framework focuses on validating the quality of master data. In contrast to the preceding project-driven stages, Stage 6 adopts a continuous approach aimed at sustaining MDM outcomes well beyond the initial implementation phase. This includes: Business rule validation, cross-system consistency checks and downstream output validation.

IV. CASE STUDY

The aim of our study is to examine the hypothesis that the integration of data quality principles into Master Data Management implementation enhances the reliability and consistency of patient information in healthcare organizations. The following section presents details contributions and their implementation in a real case study, following the steps outlined in figure 1. Study is taken in a regional hospital in January 2024. The case study operates in three services namely "Emergency Room", "Blood service", and "Labs". The hospital system struggled with fragmented patient data, inconsistent quality, and rising regulatory compliance pressures. Disconnected electronic

medical record systems led to duplicate patient records, incomplete or inaccurate data and delays in care efforts. To solve this, we propose to implement a six-phase MDM framework. First, we collaborate with stakeholders (including ER staff and lab technicians) to interview senior manager (heads of Operations) and staff using open questions. Questions focus on:

- How often bad or incomplete data is encountered.
- The effect of data errors on daily operations or tasks.

The senior manager explained critical gaps in patient blood type documentation, with 40% of records missing ABO/Rh data. This data incompleteness led to critical delays—ER physicians lost 15–20 minutes verifying blood types during hemorrhages, and labs often had to re-test samples due to unlinked or missing results in patient records. These inconsistent patient data caused frequent care delays, something IT could not have uncovered without direct involvement from frontline staff.

The table 1 outlines the strategic vision and corresponding goals for ensuring complete and reliable patient blood data within healthcare systems. It aligns each goal (G1–G4) with specific, measurable questions (Q1–Q7) to assess progress. The focus areas include improving data completeness, reducing record duplication, enhancing data synchronization between labs and Electronic Health Records (EHR), and boosting both patient safety and operational efficiency. Each question serves as a performance indicator to evaluate the success of MDM-related initiatives.

Table 1: Goals table.

Vision	Goals	Questions
Ensure complete and reliable patient blood	G1: improve data completeness	Q1: what percentage of patient records contains complete ABO/Rh blood type information? Q2: How many critical attributes (e.g., ID, demographics) are systematically missing?
	G2: reduce duplication of patient records.	Q3: what is the current duplicate rate of patient records in the Electronic Health Record compared to baseline?
	G3: enhance timeliness of data synchronization (lab ↔ EHR).	Q4: how quickly are lab results integrated into the Electronic Health Record after validation? Q5: What percentage of lab results fail to propagate automatically to the Electronic Health Record?
	G4: increase patient safety and operational efficiency	Q6: How often do care delays occur due to missing or inconsistent data? Q7: What is the reduction in re-testing rates for blood samples?

Source: Authors, (2025).

The team implemented strict data quality rules, mandating blood type verification before any surgery or transfusion and automatically rejecting unverified entries. They also developed a comprehensive data dictionary, cataloging all patient, provider, and asset data across systems, complete with metadata for consistency and clarity. All disparate sources were identified. To ensure clarity, goals must be defined in concrete and measurable terms. The purpose of this step is to translate the questions and goals identified in table 1 into data quality dimensions and associated measures, along with the appropriate calculation methods. A set of seven quality metrics were defined for answering each question, and a set of measurement methods were used for assessing such metric, as illustrated in table 2. Finally, the minimum thresholds required for each selected criterion is specified. The defined threshold must be high enough to ensure utility, but not so strict that remediation becomes infeasible.

Table 2: Data Quality Dimension

Question	Dimension	Metrics	Threshold
Q1	Completeness	% of patient records with ABO/Rh blood type filled	≥ 95%
Q2	Completeness	% of mandatory attributes populated per patient record	≥ 98%
Q3	Uniqueness	(Number of duplicate patient IDs ÷ Total patient records) × 100	Reduce duplicate patient records by 70% in 12 months
Q4	Timeliness	Average lab-to-EHR synchronization time (minutes)	≤ 5 minutes
Q5	Timeliness	(Number of lab results not propagated ÷ Total validated lab results) × 100	< 1%
Q6	Consistency	Average time lost per incident	Reduce delays by ≥ 70% in 12 months.
Q7	Accuracy	(Number of duplicate blood sample tests ÷ Total blood tests) × 100	reduce from 30% baseline to ≤ 2%

Source: Authors, (2025).

By applying data quality principles, we mapped key activities and scenarios involved in creating and updating patient records. Expert interviews and reference documentation informed this process, helping embed completeness, accuracy, and consistency across all data handling workflows. We redesigned workflows using BPMN models to support the implementation of Master Data Management. Now, when labs confirm a blood type, it auto-populates the EHR, with any conflicts flagged for pathologist review. Real-time dashboards track compliance, and weekly audits ensure ongoing data accuracy.

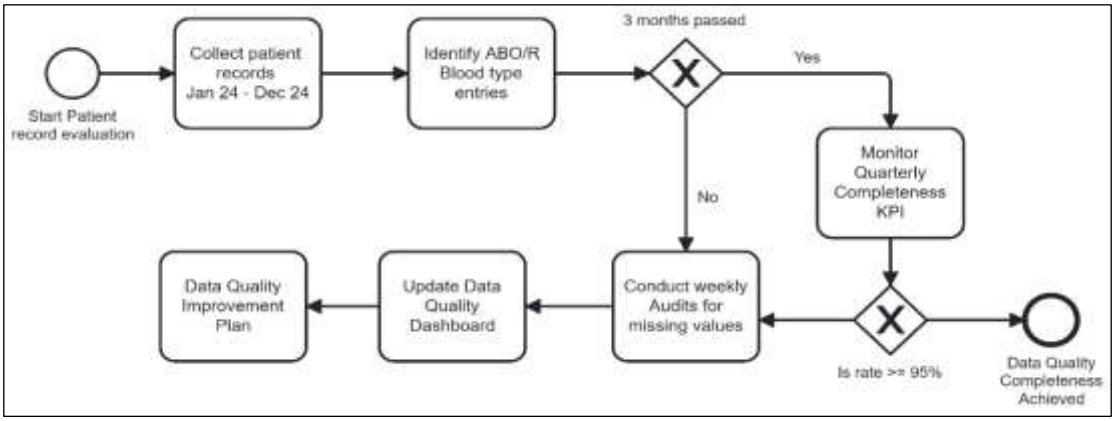


Figure 2: Healthcare data quality – Completeness – Camunda Modeler.
Source: [21].

Talend MDM tool is used for the implement of data quality and the construction of the MDM [22, 23]. From January to December 2024, we measured the percentage of patient records containing ABO/Rh blood type data, covering all clinical periods. During the spring quarter, weekly audits were conducted to identify missing values. To validate the findings, measurements were repeated and an average quality dimension rate was calculated for the year.
The following outlines how this phase applies to specific data quality metrics from Q1 to Q7:

- Q1 and Q2 – Completeness

To ensure completeness, this stage emphasizes the systematic tracking of patient records with mandatory attributes, such as ABO/Rh blood type and demographic data. Hospital departments have embedded these requirements into registration and lab systems. A data quality dashboard monitors completeness rates, with targets set at ≥95% and ≥98% respectively. These metrics are reviewed quarterly as part of KPI monitoring. Figure 2 illustrates how the completeness of patient blood type records (ABO/Rh) evolved throughout 2024. It measures the number of complete records out of a total of 1000, with a goal of reaching at least 950 (95%) after the MDM system was put in place, as shown in Figure 3.

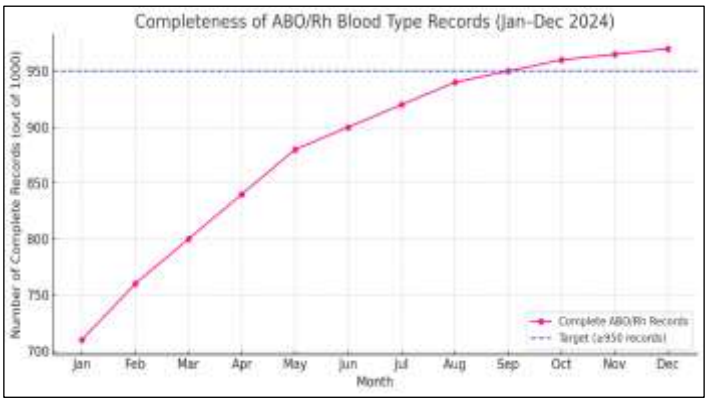


Figure 3: Pre and post completeness data quality.
Source: Authors, (2025).

- Q3 – Uniqueness

A 70% reduction in duplicate patient records over 12 months has been achieved through MDM's entity resolution capabilities, managed under stages 5 and 6 of the framework. Deduplication KPIs are regularly monitored and reported to the data governance board.

- Q4 and Q5 – Timeliness

Synchronization between lab systems and the EHR is actively monitored to ensure timeliness. Key performance indicators include average data propagation times (target ≤5 minutes) and error rates (target <1%). Metrics are logged in real time and audited monthly. Any deviations trigger immediate system diagnostics and process reviews, following the continuous improvement loop established for this phase.

- Q6 – Consistency

Time lost per incident due to data inconsistencies is a critical operational metric. In Phase 6, periodic incident tracking and response time analysis are conducted to achieve a ≥70% reduction in delays over 12 months, as shown in Figure 4. Consistency metrics are directly linked to process quality and are reviewed in cross-functional quality meetings.

Figure 4: Pre and post consistency.
Source: Authors, (2025).

- Q7 – Accuracy

Accuracy is monitored by tracking duplicate blood sample tests, with a reduction goal from a 30% baseline to ≤2%. Stage 6 includes clinical workflow audits and lab order reconciliation to measure and enforce accuracy. Alerts and automated checks are embedded within the EHR and lab information systems to support this objective. To support MDM implementation, change management efforts include a data quality directive endorsed by hospital leadership and distributed organization-wide. Training sessions, quick-reference guides, and ongoing communication campaigns promote a culture that treats data as a strategic asset, reinforcing alignment with data quality principles.

V. DISCUSSION

If the methodology directly addresses the identified failure factors, it aligns with state-of-the-art recommendations. This case study provides evidence for the framework's effectiveness by demonstrating its ability in tackling the challenges and solutions highlighted in recent literature. We presented our research methodology and analysis results to hospital managers, who found the insights both useful and constructive. The presentation enhanced their understanding of how to reduce duplicate patient records, address incomplete or inaccurate data, and minimize delays in care delivery.

- Evaluation of Data Quality.

The solution proposed in this paper places data quality at the core of the MDM process, starting with the identification of key quality dimensions, metrics, and thresholds. This approach ensures that data quality is treated not merely as a technical issue, but as a strategic business priority, with clear, measurable goals aligned to organizational objectives. By implementing the initial steps of the methodology, the organization established a shared master data dictionary and defined its data quality goals.

- Evaluation of Data Governance:

By initiating the MDM project with a unified vision and cross-functional input, the hospital significantly reduced the risk of misalignment, under-scoped implementations, and data quality failures. This early stage established a foundation for collaborative ownership, structured governance, and iterative development. The early involvement of both business and IT stakeholders fostered a shared vision and stronger alignment. BPMN modeling enhanced process transparency and understanding, supporting more effective training and documentation. The methodology explicitly emphasizes the need for early and continuous engagement from both business and IT throughout the project.

- Evaluation of Business and IT

The methodology's use of BPMN for process modeling aligns with best practices for bridging the gap between business and IT. MDM initiatives are often approached as purely technical projects, which can result in a disconnect between IT and business users. By adopting BPMN, the methodology produces visual workflows that are easily understood by both business analysts and technical teams. These models help uncover redundancies, delays, and risks within healthcare operations, enabling targeted improvements and more efficient use of resources.

The implementation of this integrated, process-driven MDM framework allowed the healthcare organization to significantly enhance patient data accuracy, operational efficiency, and regulatory compliance, while fostering a sustainable governance culture. The methodology's detailed focus on stakeholder engagement, process transparency through BPMN, and continuous monitoring offers a replicable and highly recommended approach for addressing MDM challenges related to data quality and fragmentation.

VI. CONCLUSION

In this paper, we proposed a novel framework that demonstrates how addressing insufficient data governance, fragmented data, and stakeholder resistance through a structured approach leads to measurable improvements in data quality, operational efficiency, and regulatory compliance—findings that align with current academic and industry research. This study introduced a six-stage Master Data Management framework that integrates strategic alignment with robust data quality governance. The framework was developed by merging principles from established data governance practices with stakeholder-centered methodologies and Business Process Modeling

Notation. Each stage targets a critical dimension of MDM implementation—from high-level visioning to operational quality control and continuous improvement. The framework promotes a balanced approach between strategic oversight and day-to-day execution, enabling organizations to build MDM capabilities that are both technically robust and business-aligned. It emphasizes cross-functional ownership, process transparency, and sustained quality management—effectively addressing many of the root causes of MDM failure identified in prior literature.

What distinguishes our framework is its integration of BPMN modeling, stakeholder-driven governance, and data quality metrics into a unified, actionable methodology. This combination enables organizations to evolve beyond static data governance strategies and adopt a more dynamic, role-aware, and metrics-driven execution of MDM programs. Hospital managers found our methodology and results highly valuable, particularly in improving their understanding of how to reduce duplicate patient records, address data completeness and accuracy issues, and minimize delays in care delivery.

However, the study's primary limitation lies in its validation through a single case study within the healthcare sector. While illustrative, it may not capture the full range of complexities present in other industries. Additionally, the framework's effectiveness is influenced by several contextual factors, including organizational readiness, technology infrastructure, and stakeholder engagement, all of which can vary significantly. Future research should explore multi-site implementations, cross-sectoral comparisons, and deeper integration with advanced analytics and metadata management platforms to further validate and enhance the framework's generalizability and impact.

VII. AUTHOR'S CONTRIBUTION

Conceptualization: Benkherourou Chafika and Abdelhabib Bourouis.
Methodology: Benkherourou Chafika and Abdelhabib Bourouis.
Investigation: Benkherourou Chafika and Abdelhabib Bourouis.
Discussion of results: Benkherourou Chafika and Abdelhabib Bourouis.
Writing – Original Draft: Benkherourou Chafika.
Writing – Review and Editing: Benkherourou Chafika and Abdelhabib Bourouis.
Resources: Benkherourou Chafika and Abdelhabib Bourouis.
Supervision: Abdelhabib Bourouis.
Approval of the final text: Benkherourou Chafika and Abdelhabib Bourouis.

VIII. REFERENCES

[1] D. Loshin, "*Master Data Management"*. Burlington, MA: Morgan Kaufmann, 2008.

[2] D. Cervo and M. Allen, "*Master Data Management in Practice: Achieving True Customer MDM"*. Hoboken, NJ: John Wiley & Sons, 2011.

[3] M. H. Ofner, K. Straub, B. Otto, and H. Oesterle, "Management of the master data lifecycle: a framework for analysis, *Journal of Enterprise Information Management*, vol. 26, no. 4, pp. 472–491, Jul. 2013.

[4] F. Haneem, N. Kama, N. Taskin, D. Pauleen, and N. A. Abu Bakar, "Determinants of master data management adoption by local government organizations: An empirical study," *International Journal of Information Management*, vol. 45, pp. 25-43, Dec. 2019.

[5] M. Spruit and K. Pietzka, "MD3M: The master data management maturity model," Computers in Human Behavior, vol. 51, pp. 1068–1076, 2014.

[6] "Master Data Management: Build a Strong Process, Framework and Solution," Gartner. https://www.gartner.com/en/data-analytics/topics/master-data-management (accessed Jan. 03, 2023).

[7] Fu Q, Nicholson GL, Easton JM. "Understanding data quality in a data-driven industry context: Insights from the fundamentals". J Ind Inf Integr. 2024; 42:100729. doi:10.1016/j.jii.2024.100729.

[8] R. Vilminko-Heikkinen, "Establishing an Organization's Master Data Management Function: A Stepwise Approach," in 2013 46th Hawaii International Conference on System Sciences, 2013, pp. 1-10.

[9] D. Jha, "Orchestrating Clean Data: Master Data Management Across the Lead-to-Cash Lifecycle," JISEM Journal, 2025, doi: 10.52783/jisem.v10i36s.6726.

[10] A. Haug, A. M. Staskiewicz, and L. Hvam, "Strategies for Master Data Management: A Case Study of an International Hearing Healthcare Company," Information Systems Frontiers, vol. 25, pp. 1903-1923, 2023, doi: 10.1007/s10796-022-10323-z.

[11] F. Gualo, I. Caballero, and M. Rodriguez, "Towards a software quality certification of master data-based applications," Software Quality Journal, vol. 28, no. 3, pp. 1019-1042, 2020, doi: 10.1007/s11219-019-09482-1.

[12] R. Vilminko-Heikkinen and S. Pekkola, "Master data management and its organizational implementation: An ethnographical study within the public sector," Journal of Enterprise Information Management, vol. 30, no. 3, pp. 454-475, Apr. 2017.

[13] S. V. Kuznetsov and D. Koznov, "Master data management in an iterative approach," Ontology of Designing, vol. 11, no. 2, pp. 170-184, 2023, doi: 10.18287/2223-9537-2021-11-2-170-184.

[14] A. Haug, J. S. Arlbjørn, F. Zachariassen, and J. Schlichter, "Master data quality barriers: an empirical investigation," *Industrial Management & Data Systems*, vol. 113, no. 2, pp. 234–249, 2013, doi: 10.1108/02635571311303550.

[15] M. Al-Ruithe and E. Benkhelifa, "Analysis and classification of barriers and critical success factors for implementing a cloud data governance strategy," Industrial Management & Data Systems, vol. 113, pp. 223–232, 2017. doi: 10.1016/j.procs.2017.08.352

[16] M. Janssen, P. Brous, E. Estevez, L. S. Barbosa, and T. Janowski, "Data governance: Organizing data for trustworthy artificial intelligence," Government Information Quarterly, vol. 37, no. 3, 2020, Art. no. 101493, doi: 10.1016/j.giq.2020.101493.

[17] S. Hikmawati, P. Santosa, and I. Hidayah, "Improving data quality and data governance using master data management: A review," IJITEE (International Journal of Information Technology and Electrical Engineering), vol. 5, p. 90, 2021, doi: 10.22146/ijitee.66307.

[19] P. Lieponitis, "Master data management: its importance and reasons for failed implementations," Ph.D. dissertation, Faculty of Sheffield Hallam University, 2020. [Online]. Available: https://doi.org/10.7190/shu-thesis-00311.

[19] H. N. Prasetyo and K. Surendro, "Designing a data governance model based on soft system methodology (SSM) in organization," *Journal of Theoretical and Applied* Information Technology, vol. 78, no. 1, pp. 46-52, 2015.

[20] T. Biard, M. Bigand, and J.-P. Bourey, "Explicitation et structuration des connaissances pour la transformation de l'entreprise : Les apports de la méthode Praxeme," in Proc. Congrès International de Génie Industriel, 2013, pp. [page numbers].

[21] "Camunda Modeler: Process Modeling using BPMN," Camunda. [Online]. Available: https://camunda.com/platform/modeler/. Accessed: May 13, 2024.

[22] "Talend Master Data Management," Talend Inc. [Online]. Available: https://www.talend.com/fr/solutions/. Accessed: Feb. 20, 2024.

[23] S. R. Cheruku, O. Goel, and S. Jain, "A Comparative Study of ETL Tools: DataStage vs. Talend," J. Quantum Sci. Technol., vol. 1, no. 1, pp. 80-90, 2024, doi: 10.36676/jqst.v1.i1.11.

Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11 n.55, p. 212-221. September/October, 2025.

DOI: https://doi.org/10.5935/jetia. v.11 n.55.2658

RESEARCH ARTICLE OPEN ACCESS

NOVEL INSIGHTS ON THE COMPARATIVE STUDY BETWEEN LSTM AND TRANSFORMER MODELS FOR FINANCIAL TIME SERIES PREDICTION

Rabia Nichani[1], Laid Gasmi[2], Salheddine Kabou[3] and Nabil Laiche[4]

[1, 2] Laboratory of Mathematics Modeling and Applications, University of Adrar, Algeria

[3] Department of Mathematics and Computer Science, Higher Normal School of Bechar.

[4] Dynamical Systems and Control Laboratory. Oum El Bouaghi University, Algeria

[1]https://orcid.org/0009-0004-4755-7946, [2]https://orcid.org/0000-0001-8925-0089

[3]http://orcid.org/ 0000-0002-1423-7215, [4]http://orcid.org/0000-0003-3392-1300

Email : nicha.rabia@univ-adrar.edu.dz, lai.gasmi@univ-adrar.edu.dz, kabou.salaheddine@yahoo.fr, nlaiche2020@gmail.com

ARTICLE INFO

Article History

Received: September 3, 2025

Revised: September 30, 2025

Accepted: October 5, 2025

Published: October 31, 2025

Keywords:

Financial Time Series,
Forecasting,
Deep Learning,
LSTM, Transformer,
Optimizers.

ABSTRACT

This study investigates the impact of deep learning model architectures and optimization techniques on financial time series forecasting, with a focus on Transformer-based models and Long Short-Term Memory networks. Optimization methods such as Adaptive Moment Estimation and Adaptive Moment Estimation with Infinity Norm play a critical role in enhancing training efficiency and prediction accuracy, while architectural decisions determine each model's ability to capture both short-term and long-term dependencies in sequential data. Using datasets including stock prices from Microsoft and NVIDIA, the models are evaluated based on metrics such as Mean Absolute Percentage Error, Root Mean Squared Error, prediction variance, and training speed. The results demonstrate the complementary strengths of Transformers and Long Short-Term Memory Networks, underscoring the importance of tailored architectures and optimization strategies in deep learning-based financial forecasting.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Time series forecasting plays a pivotal role in diverse domains, including finance, supply chain management, and environmental science. In finance, predicting future trends in stock prices and market indices is particularly challenging due to the complex statistical properties of financial time series. These properties include non-stationarity, volatility clustering, and heavy tails, which defy the assumptions of traditional time series models and necessitate more advanced approaches [**1**,**2**].

Traditional methods such as ARIMA (Autoregressive Integrated Moving Average) and exponential smoothing have proven effective for stationary data, but they struggle with the dynamic, high-dimensional nature of financial data. Deep learning models, notably Long Short-Term Memory (LSTM) networks and Transformers, have emerged as powerful tools in addressing these challenges. LSTMs excel at capturing temporal dependencies and long-term relationships in sequential data, thanks to their gated architecture [**3**, **4**]. On the other hand, Transformers leverage self-attention mechanisms to model both local and global dependencies effectively, making them particularly suited for complex and irregular time series like those in financial markets [**5, 6**].

Despite their promise, applying LSTMs and Transformers to financial forecasting remains underexplored compared to their widespread adoption in fields like image recognition [**7**] and natural language processing [**8, 9**]. Existing studies have shown mixed results, highlighting gaps in understanding the suitability of these architectures for financial time series forecasting. For example, LSTMs have demonstrated robust performance in capturing sequential patterns, but their reliance on sequential processing can limit computational efficiency [**10,11**]. Transformers, while efficient for long sequences, face challenges such as higher computational complexity due to their self-attention mechanism [**12**].

This study aims to address these gaps by conducting a systematic comparison of LSTMs and Transformers in the context of stock price prediction for Microsoft (MSFT) and Nvidia Corporation (NVDA). Key contributions include exploring the impact of diverse optimizers (e.g., Adam, RMSprop, Adadelta) and evaluating performance based on metrics such as the Mean Absolute Percentage Error

(MAPE) and Root Mean Squared Error (RMSE). The results highlight the transformative potential of deep learning architectures in financial forecasting, offering insights into their strengths and limitations while paving the way for future innovations in hybrid modeling.

The structure of the remainder of this paper is as follows. After this brief introduction, Section 2 will discuss existing surveys that focus on deep learning models such as Convolutional Neural Networks (CNN), Long Short-Term Memory (LSTM), and Transformer models for financial time series forecasting. Section 3 presents the study of the most important statistical properties of the data. In Section 4, we compare the effects of optimizers on the prediction quality of the LSTM model and the Transformer model for two financial time series, MSFT and NVDA, using diverse metrics. Finally, we conclude in Section 5 by summarizing our findings.

II. BACKGROUND AND RELATED WORKS

The ability of Deep Learning (DL) models to analyze intricate patterns in financial time series data has gained significant attention in recent years [**13**]. Numerous studies and surveys have investigated the application of advanced DL architectures such as Convolutional Neural Networks, Multi-Layer Perceptrons, Long Short-Term Memory networks, and Transformer models in financial forecasting tasks. CNNs, known for their strength in feature extraction, have been adapted to handle time series data by effectively capturing local temporal dependencies. For instance, studies such as [**14**,**15**] have demonstrated the utility of CNNs in forecasting financial trends, particularly in tasks requiring multi-dimensional input representations. LSTMs, a variant of RNNs, have been widely adopted due to their ability to capture long-term dependencies and sequential patterns in financial datasets. These capabilities make LSTMs particularly effective for time series forecasting, where past data points significantly influence future predictions. In the following sections, we review the literature on time series prediction, with a specific focus on the application of LSTM and Transformer models for financial forecasting.

II.1 LSTM-BASED TIME SERIES PREDICTION

Long Short-Term Memory (LSTM) networks have been extensively utilized in financial time series prediction, with numerous studies [**16**] demonstrating their success in forecasting stock prices. To further enhance prediction performance, Bidirectional LSTM (BiLSTM) has been employed [**17**]. The application of Natural Language Processing (NLP) techniques has also led to the introduction of Sequence-to-Sequence (Seq2Seq) models [**18**], initially developed for machine translation tasks. These models, when integrated with LSTM structures and attention mechanisms, have shown significant improvements in financial time series forecasting [**19, 20**]. Moreover, [**11**] innovatively combined CNNs and LSTM to create a novel architecture known as DeepLOB.

This architecture outperformed LSTM-only in predicting future stock price movements from Limit Order Book (LOB) data. In a subsequent study, [**10**] further enhanced DeepLOB by incorporating Seq2Seq and attention models, enabling multi-horizon price prediction in a single forward pass. This approach reduced training time and enhanced performance, especially for long-term predictions, outperforming the original DeepLOB model. According to [**21**], the combination of CNN and LSTM demonstrated superior stability and performance in predicting LOB (Limit Order Book) price movements. Similarly, [22] utilized the Order Flow Imbalance (OFI) feature from the LOB to forecast minimum-price returns, where the CNN-LSTM architecture again exhibited superior results. Recently, hybrid models such as Transformer-CNN [**23**], ARIMA-LSTM, and ARIMA-XGBoost have gained attention for improving forecasting accuracy in financial time series prediction [**24**].

II.2 TRANSFORMER-BASED TIME SERIES PREDICTION

The Transformer model, initially introduced for Natural Language Processing (NLP) tasks, has since inspired the development of various advanced models such as SOTA, BERT, and GPT-3. These models have gained significant attention due to their effectiveness, and their application has increasingly extended to time series prediction. However, the canonical self-attention mechanism in Transformer models has a time and memory complexity of $O(L^2)$, which can be computationally prohibitive for long sequences. To address this, several adaptations of the original Transformer architecture have been proposed to enhance its efficiency, particularly in processing long time series for forecasting tasks [**6**], with the most recent advancement being the FEDformer model [**12**]. In the financial domain, the Transformer has been applied to various prediction tasks. For example, [**25**] combined a Temporal Fusion Transformer with Support Vector Regression (SVR) and LSTM for stock price forecasting, showing improved performance. Similarly, [**26**] employed a Transformer to predict Dogecoin prices, demonstrating its superior predictive ability over LSTM models.

These studies primarily focus on using Transformer-based methods for predicting stock prices based on OHLC (Open, High, Low, Close) data. However, there is still limited research on the application of Transformer models to Limit Order Book (LOB) data for prediction. While LSTM models have been widely used in financial time series forecasting and have been extensively tested across different datasets, the use of Transformers remains more restricted. This study aims to extend the application of Transformer-based models to a broader range of financial time series prediction tasks, with a particular focus on comparing their performance against LSTM models [2**7, 28**]. Using financial time series data from companies like Microsoft (MSFT) and NVIDIA (NVDA), we evaluated the models based on MAPE, RMSE, prediction variance, and training speed. The results reveal that Transformer and LSTM models possess complementary strengths, with Transformers excelling at capturing long-term dependencies, while LSTMs, though effective at managing short-term sequential patterns, are more limited when dealing with complex long-term relationships.

These findings align with previous studies, such as those by [**16**], which demonstrated the success of LSTMs in financial time series forecasting. However, our approach, which combines Transformers with specific architectural optimizations, outperforms traditional LSTM models in terms of MAPE and RMSE, while also reducing training time, which is critical for large-scale applications. One of the major contributions of this study lies in exploring the impact of architectural choices and optimization strategies. We show that customizing model architectures based on the specific characteristics of financial time series, along with adapting optimizers like Adam and Adamax, can significantly improve forecasting accuracy while optimizing training efficiency. This approach surpasses the limitations observed in previous studies, where models based solely on LSTMs or CNN-LSTM [**29,30**] architectures did not always achieve optimal performance on extended time horizons or complex datasets. This study advances knowledge in financial time series

forecasting by highlighting the crucial role of architecture and optimization choices in improving the performance of deep learning models. Our results contribute to expanding the capabilities of both Transformers and LSTMs, paving the way for more robust and efficient hybrid approaches in financial forecasting.

III. METHODOLOGY

We will first describe two commonly used recurrent network architectures (LSTM and Transformer), then explain their use in financial time series forecasting.

III. 1 LONG SHORT-TERM MEMORY (LSTM)

An LSTM unit is composed of cells containing an input gate, an output gate, and a forget gate. These three gates regulate the flow of information. With these features, each cell can remember desired values over arbitrary time intervals [**4**]. The structure of the LSTM is shown in Figure .1. The structure of the LSTM consists of three main gates: a forget gate (**3.1**), an input gate (**3.2**), and an output gate (**3.3**). The forget gate, which controls the retention of information, can be expressed as follows:

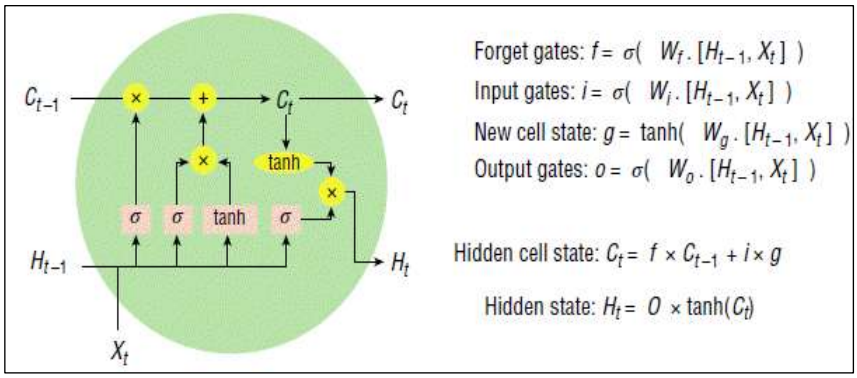


Figure 1: LSTM Diagram.
Source: Authors, (2025).

$$f_t = \sigma\big(W_t.[h_{t-1}, x_t] + b_f\big), \quad (Forgot\ Gate) \tag{3.1}$$
$$i_t = (W_i.[h_{t-1}, x_t] + b_i), \quad (Input\ Gate) \tag{3.2}$$
$$\tilde{c}_t = tanh(W_c.[h_{t-1}, x_t] + b_c), \quad (Cell\ State\ \ update) \tag{3.3}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t, \quad (New\ Cell\ State) \tag{3.4}$$
$$o_t = \sigma(W_0.[h_{t-1}, x_t] + b_0), \quad (Output\ \ Gate) \tag{3.5}$$
$$h_t = o_t \odot tanh(c_t). \quad (New\ Hidden\ \ State) \tag{3.6}$$

The output from the LSTM, as described in (3.6), is passed through fully connected (dense) layers for prediction:

$$z_1 = ReLU(W_1.ht + b_1), \quad (First\ Dense\ Layer) \tag{3.7}$$
$$y = W_2 . z_1 + b_2. \quad (Output\ Layer) \tag{3.8}$$

The model is trained by minimizing the Mean Squared Error (MSE), which is defined in (3.9):

$$Loss = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2. \tag{3.9}$$

Metrics such as Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are used to evaluate performance. The MAE, given by (3.10), and the MAPE, given by (3.11), are calculated as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i|. \tag{3.10}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100. \tag{3.11}$$

To prevent overfitting, dropout and early stopping are employed. Our baseline LSTM model consists of one LSTM layer with 200 units, followed by two dense layers with 50 and 1 neuron, respectively.

III. 2 TRANSFORMERS ARCHITECTURE

The architecture of the Transformer for financial time series prediction is shown in Figure. 2.

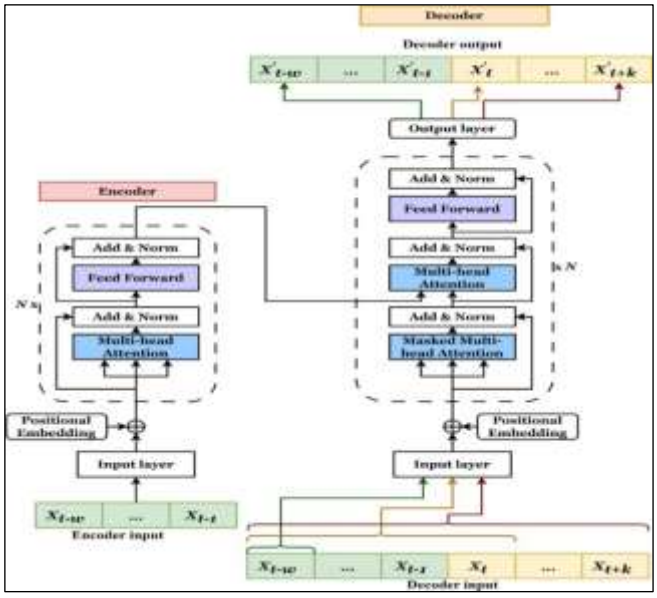


Figure 2: Architecture of the Transformer model for time series forecasting.
Source: Authors, (2025).

Thanks to attention mechanisms, distant elements can be connected directly, minimizing problems with gradients. Furthermore, Transformers solve the problem of network depth [**31**]. On the other hand, training, in practice, proves to require fewer epochs and, for each of them, the parallel treatment of the computation is much simpler to implement see [**32**]. The Transformer model leverages self-attention mechanisms to directly connect distant elements in sequential data, addressing the limitations of traditional RNNs. The self-attention mechanism is defined as:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V. \tag{3.12}$$

Where Q, K and V are the query, key, and value matrices, and d_k is the dimension of the key. Each Transformer block comprises:

- Layer Normalization:

$$LayerNorm(x) = \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}}.\gamma + \beta. \tag{3.13}$$

- Feed-Forward Network:

$$FFN(x) = ReLU(xW_1 + b_1)W_2 + b_2. \tag{3.14}$$

- Residual Connections:

$$Output = LayerOutput + Input. \tag{3.15}$$

The architecture includes global average pooling and dense layers for output prediction. Dropout is applied throughout to enhance generalization. This model efficiently captures temporal dependencies and exhibits strong performance in financial time series forecasting. Building an LSTM for prediction requires using the Keras method; First, it performs a series of steps to initialize the neural network. Second, a dense layer is used to load the densely connected neural network, and an LSTM layer is added to incorporate long short-term memory. Finally, dropout is applied to prevent overfitting. Predicting future inventories when loading a test set requires a series of very important procedures [**33**]. Ray et al **[31]** introduced an innovative hybrid ARIMA-LSTM model. This model leverages a random forest algorithm to determine the appropriate input lags for the LSTM component.

IV. STATISTICAL ANALYSIS OF THE DATASET

The datasets analyzed include the closing prices of Microsoft (MSFT) and Nvidia (NVDA) over distinct time periods. Understanding the statistical properties of these financial time series provides context to the modelling challenges posed by their inherent characteristics. Below are the key statistical summaries for each dataset (see Table 1):

Table 1: Statistical Properties of Microsoft (MSFT) and Nvidia (NVDA) Closing Prices.

Statistic	Microsoft (MSFT)	Nvidia (NVDA)
Mean Closing Price	52.77	4.19
Standard Deviation	78.67	8.78
Skewness	2.32	3.03
Kurtoris	7.53	12.66

Source: Authors, (2025).

These statistics highlight the highly volatile nature of financial markets, characterized by non-normal distributions with significant skewness and kurtosis. The high kurtosis further suggests the potential for extreme values, which must be considered in predictive modelling.

The analysis of Bollinger Bands reveals significant volatility in Microsoft's stock prices over time. The lower band, with an average of 49.64, indicates support levels, while the upper band, averaging 55.30, represents resistance levels. The moving average, which averages 52.47, reflects the central tendency of the stock prices. Extreme values in the percentage B (pctB), ranging from -0.54 to 1.49, highlight occasional deviations beyond the bands, suggesting anomalies or speculative activities. Overall, the dataset exhibits a general upward trend punctuated by periods of high variability, illustrating the dynamic nature of financial time series (see Figure 3).

Figure 3: MSFT closing prices with Bollinger Bands, showing resistance (upper band), support (lower band), and moving averag.
Source: Authors, (2025).

Similarly, the Bollinger Bands analysis for Nvidia Corporation (NVDA) stock demonstrates notable volatility. The lower band averages 3.79, marking support levels, while the upper band averages 12.97, indicating resistance levels. The moving average of NVDA stock prices is 4.19, representing its central tendency. Extreme pctB (percentage Bandwidth) values range from -0.72 to 4.96, indicating occasional deviations beyond the bands that may signal anomalies or significant market events. The dataset shows an overall upward trend interspersed with periods of high volatility, underscoring the dynamic and unpredictable nature of NVDA's stock price (see Figure **4**).

Figure 4. NVDA closing prices with Bollinger Bands, showing resistance (upper band), support (lower band), and moving averag.
Source: Authors, (2025).

In this study, we first establish classes that make it easier to convert unstructured financial time series data into a format suitable for training machine learning models, particularly those that require sequential inputs like LSTMs. This approach transforms our problem into a supervised learning problem. These classes provide consistency, efficiency, and scalability in dataset preparation by automating extraction, transformation, and windowing. Our proposed model is a sequential architecture consisting of an LSTM layer with 200 units, followed by two Dense layers with 50 and 5 neurons, respectively. The LSTM layer captures temporal dependencies in sequential data, while the dense layers reduce dimensionality to produce the final output. With 515,717 parameters (171,905 trainable), our model is trained for 25 epochs with a batch size of 32, minimizing MSE and using MAE and MAPE as performance metrics. Early stopping is employed to prevent overfitting, making the model well-suited for tasks such as financial time series prediction or multi-class classification (see Table 2).

Table 2: Summary of the Transformer model architecture with parameter details.

Layer (Type)	Output Shape	Parameters
lstm(LSTM)	(None, 200)	161,6
dense (Dense)	(None, 50)	10,05
dense_1 (Dense)	(None, 5)	255
Total parameters:		515,717
Trainable parameters:		171,905
Non-trainable parameters:		0
Optimizer parameters:		343,812

Source: Authors, (2025).

The Transformer Encoder uses layer normalization, multi-head attention, residual connections, and feed-forward networks with 1D convolutional layers and dropout for regularization. The build_transformer function stacks Transformer blocks, applies global average pooling, and ends with dense layers for output prediction. Key parameters include a head size of 128, four attention heads, a feed-forward dimension of 2, four Transformer blocks, and a dropout rate of 0.10. The model is trained with MSE loss using the Adam optimizer (learning rate: 0.001) and evaluated with MAE and MAPE.

Training runs for 25 epochs with early stopping to prevent overfitting. The hybrid architecture combines multi-head attention and 1D convolutional layers to process sequential data. It starts with an input layer, followed by layer normalization and multi-head attention to capture long-range dependencies. Convolutional layers and dropout are applied for feature extraction and regularization, while residual connections improve gradient flow. Global average pooling reduces dimensionality before dense layers produces the final output. With 51,617 parameters (17,205 trainable), the model is ideal for tasks like time series forecasting, balancing local and global dependencies (see Table **3**).

Table 3: Summary of the Transformer model architecture with parameter details.

Layer (Type)	Output Shape	Para.	Connected to
Input_layer_1	(None, 5, 1)	0	-
multi_head_attention	(None, 5, 1)	3,585	Input_layer_1
conv1d	(None, 5, 2)	4	multi_head_attention
conv1d_1	(None, 5, 1)	3	conv1d
conv1d_2	(None, 5, 2)	4	conv1d_1
conv1d_3	(None, 5, 1)	3	conv1d_2
dense_3	(None, 5)	1,285	conv1d_3
Total parameters:		51,617	
Trainable parameters:		17,205	
Non-trainable parameters:		0	
Optimizer pparameters:		34,412	

Source: Authors, (2025).

For our Transformer architecture, we base the design on the configuration recommended in the Keras documentation, which is originally tailored for classification tasks. However, we made a key modification: the activation function of the final output layer was changed from softmax to ReLU, and the loss function was switched to MSE to better align with the nature of the problem. In addition to this adjustment, we finetuned the hyperparameters based on experimental results to optimize model performance for this specific task.

V. EXPERIMENTATION RESULTS AN D EVALUATION

Various optimizers were utilized to train the neural network models. Adam combines Adagrad (Adaptive Gradient) and RMSProp, dynamically adjusting learning rates for each parameter [**28**]. Adamax, a variant of Adam, is effective for sparse gradients. RMSProp (Root Mean Square Propagation) is designed for non-stationary data, addressing diminishing learning rates. Adadelta dynamically adapts learning rates, eliminating the need for a fixed rate [**35**]. Adagrad (Adaptive Delta) adjusts learning rates based on the historical sum of squared gradients. Nadam (Nesterov-accelerated Adaptive Moment Estimation) combines Nesterov's accelerated gradient with Adam for faster convergence [**36**]. The performance of Long Short-Term Memory (LSTM) and Transformer models was evaluated using Microsoft (MSFT) and NVIDIA (NVDA) stock datasets. The choice of optimizer significantly impacted the models' performance. Three key metrics were employed to evaluate the models: RMSE, MAPE. The variance ratio was included as an additional evaluation metric.

V. 1. COMPARISON USING VARIOUS OPTIMIZERS FOR THE FIRST SERIES (MSFT)

This subsection presents the results of evaluating LSTM and Transformer models on the MSFT dataset using various optimizers. Lower MAPE values indicate better forecasting accuracy, while variance and RMSE provide additional insights into model performance and stability. The performance metrics for different optimizers, including MAPE, variance, RMSE, and inference times, are summarized in Tables 4,5 ,6, and 7, respectively.

Table 4: MAPE values for various optimizers on the MSFT training dataset.

MODEL	ADAM	ADAMAX	RMSPROP	ADADELTA	ADAGRAD	NADAM
LSTM	0.03382	0.03398	0.048708	0.031500	0.028346	0.026310
TRANSF.	0.02309	0.025436	0.030663	0.382570	0.021955	0.023650

Source: Authors, (2025).

Table 4 compares the LSTM and Transformer models for MSFT stock data forecasting, assessed using the MAPE. The results highlight the overall superior performance of the Transformer model. The Adagrad optimizer achieved the best MAPE for both models, with values of 0.028346 for LSTM and 0.021955 for the Transformer.

Table 5: Var values for various optimizers on the MSFT training dataset.

MODEL	ADAM	ADAMAX	RMSPROP	ADADELTA	ADAGRAD	NADAM
LSTM	0.111730	0.095078	0.133927	0.004376	0.026858	0.012855
TRANSF.	0.026778	0.035896	0.048417	0.668279	0.004472	0.027254

Source: Authors, (2025).

Table **5** illustrates variance values for different optimizers on the MSFT training dataset using LSTM and Transformer models. For LSTM, Adagrad yields a low variance (0.026858) and performs consistently well, while Adadelta achieves the lowest variance (0.004376) but fails for the Transformer with the highest variance (0.668279). Adagrad proves to be the best optimizer overall, achieving the lowest variance for the Transformer (0.004472) and maintaining stability across both models. Other optimizers show varying degrees of variance, with RMSprop having the highest for LSTM.

Table 6: RMSE values for various optimizers on the MSFT training dataset.

MODEL	ADAM	ADAMAX	RMSPROP	ADADELTA	ADAGRAD	NADAM
LSTM	6.92941	7.99667	6.77863	8.83112	15.01159	8.68687
TRANSF.	6.68984	6.82673	6.54173	129.54425	7.18775	6.95662

Source: Authors, (2025).

Table **6** compares the RMSE values of LSTM and Transformer models across various optimizers for the MSFT training dataset. The Transformer consistently outperforms LSTM, achieving lower RMSE values with Adam (6.68984), Adamax (6.82673), RMSprop (6.54173), Adagrad (7.18775), and Nadam (6.95662). However, LSTM performs better with Adadelta, achieving 8.83112 compared to the Transformer's significantly higher RMSE of 129.54425. This underscores the importance of optimizer selection, with Transformers showing superior and more stable performance overall.

Table 7: Inference times for different optimizers on the MSFT training dataset.

MODEL	ADAM	ADAMAX	RMSPROP	ADADELTA	ADAGRAD	NADAM
LSTM	20.74728	18.58163	22.88478	23.25408	22.49023	22.24550
TRANSF.	20.05767	21.41631	24.01012	25.90485	23.71948	23.52633

Source: Authors, (2025).

Table **7** shows that the LSTM model trains fastest with Adamax (18.58 seconds), followed by Nadam and Adam. For the Transformer, Adam is the optimal optimizer with the shortest training time (20.06 seconds), followed by Adagrad and Nadam. These results highlight that LSTM achieves faster training with specific optimizers, while the Transformer shows slightly longer but consistent training times.

V. 2. COMPARISON USING VARIOUS OPTIMIZERS FOR THE SECOND SERIES (NVDA)

Before analyzing the detailed performance of LSTM and Transformer models with various optimizers on the NVDA dataset, it is important to provide a brief overview of the evaluation approach. This subsection examines metrics such as MAPE, RMSE, variance, and inference times to assess and compare model performance. The results aim to highlight the impact of optimizer selection on predictive accuracy and computational efficiency in the context of time series forecasting applied to NVDA data. We begin by comparing the performance of each model using the metric, MAPE.

Table 8: MAPE values for various optimizers on the NVDA training dataset.

MODEL	ADAM	ADAMAX	RMSPROP	ADADELTA	ADAGRAD	NADAM
LSTM	0.098944	0.141417	0.8429064	0.845593	0.322159	0.059463
TRANSF.	0.045220	0.041582	0.033147	0.695001	0.047531	0.049984

Source: Authors, (2025).

Table **8** shows that the Transformer achieves the lowest MAPE (0.033147) with RMSprop, while LSTM performs best with Nadam (0.059463). For Nadam, the Transformer has a slightly higher MAPE (0.049984) than LSTM. Overall, the Transformer demonstrates greater robustness across optimizers.

Table 9**:** Var values for various optimizers on the NVDA training dataset.

MODEL	ADAM	ADAMAX	RMSPROP	ADADELTA	ADAGRAD	NADAM
LSTM	0.374856	0.141196	0.017813	0.961713	0.122696	0.027150
TRANSF.	0.049727	0.020455	0.056631	0.957959	0.027713	0.069556

Source: Authors, (2025).

Table **9** shows that for LSTM, RMSprop achieves the lowest variance (0.017813), while Adadelta displays the highest variance (0.961713). Adagrad and Nadam also perform well, with relatively low variance values. For the Transformer, Adamax (0.020455) and

Adagrad (0.027713) yield the lowest variance, while Adadelta (0.957959) again exhibits the highest variance. Both models show consistent performance with Adamax and Adagrad but struggle with Adadelta.

Table 10: RMSE values for various optimizers on the NVDA training dataset.

MODEL	ADAM	ADAMAX	RMSPROP	ADADELTA	ADAGRAD	NADAM
LSTM	11.372042	8.604392	3.768150	44.840153	34.692550	14.44011
TRANSF.	3.173453	2.974158	3.447670	38.953478	3.168120	2.851640

Source: Authors, (2025).

Table **10** highlights that the Transformer performs better than LSTM for the RMSE metric on the NVDA series. The Transformer achieves its lowest RMSE with Nadam (2.8516) and Adamax (2.974158), while LSTM's best RMSE is higher at 3.76815 with RMSprop. Overall, the Transformer consistently delivers lower RMSE values across most optimizers, demonstrating its superiority.

Table 11: Inference times for different optimizers on the NVDA training dataset.

MODEL	ADAM	ADAMAX	RMSPROP	ADADELTA	ADAGRAD	NADAM
LSTM	15.63792	16.28799	23.23479	20.133928	16.62776	18.66561
TRANSF.	16.15776	15.414918	25.02268	20.795882	16.29275	14.93296

Source: Authors, (2025).

Table 11 highlights the inference times for LSTM and Transformer models across various optimizers on the NVDA training dataset. The LSTM achieves the lowest inference time with Adam (15.63792), while the Transformer is fastest with Nadam (14.93296). Despite some variation, both models show comparable inference times across most optimizers, except for RMSprop, where both exhibit higher inference times, with the Transformer slightly slower (25.02268).

V.3. RESULTS SUMMARY AND EVALUATION

The comparative analysis of our proposed models, LSTM, and Transformer, across various optimizers demonstrates distinct advantages of the Transformer regarding accuracy and robustness, particularly with metrics like MAPE, variance, and RMSE. The Transformer consistently achieved lower MAPE and RMSE values in most scenarios, affirming its superiority in capturing complex temporal patterns in stock price forecasting. Among the optimizers, Adagrad and Adamax emerged as the most effective, ensuring stable performance and faster convergence. However, certain limitations were observed, such as the Transformer's occasional instability with optimizers like Adadelta.

On the other hand, LSTM models showed adaptability with optimizers such as Nadam and Adadelta in specific contexts, achieving competitive results. These findings highlight the importance of selecting the right optimizer to balance model complexity, computational efficiency, and prediction accuracy. Overall, the experimentation confirms that the Transformer model, when paired with optimizers like Adamax or Adagrad, offers a promising financial time series forecasting approach, with the potential to outperform traditional LSTM models in accuracy, stability, and scalability. Future optimizations in hyperparameter tuning and the integration of additional market features could further enhance the predictive capabilities of these models.

VI.CONCLUSION AND FUTURE WORK

This study addresses the challenges of forecasting financial time series, specifically focusing on stock price prediction using Transformer-based and LSTM-based models. The findings highlight the critical impact of optimizer choice on model performance. For instance, optimizers such as Adam and Adamax significantly enhance training efficiency and prediction accuracy, underscoring the need to align model architecture and optimizer selection with the dataset's characteristics and the specific requirements of the forecasting task. Moreover, Transformer-based models excel in capturing complex, long-term temporal dependencies, while LSTM models perform effectively in scenarios requiring adaptation to specific optimization techniques. These results emphasize the importance of a tailored approach to model selection and hyperparameter tuning to achieve optimal forecasting accuracy. From a practical perspective, these models hold substantial potential for improving decision-making in financial contexts, such as algorithmic trading, risk management, and investment strategy optimization. Their ability to identify both short- and long-term patterns in market data provides valuable tools for navigating volatile financial markets.

Future research should aim to deepen these findings by exploring more advanced model architectures, including hybrid approaches that combine the strengths of deep learning models (LSTM, GRU, and Transformer) with traditional models such as ARIMA, GARCH, or other statistical techniques. Such combinations could leverage the structural modelling capabilities of traditional methods while benefiting from the ability of deep learning models to capture non-linear and complex relationships. Additionally, expanding training datasets to include macroeconomic variables, such as inflation rates, interest rates, and GDP growth, could enhance model robustness and forecasting accuracy. Applying these models in real-world scenarios, considering factors like market volatility, transaction costs, and dynamic trading strategies, would provide valuable insights into their practical effectiveness. Finally, integrating explainable AI (XAI) techniques could improve the interpretability and reliability of these models, making them more suitable for operational use in financial forecasting.

VII.AUTHOR'S CONTRIBUTION

Conceptualization: Rabia Nichani, Laid Gasmi, Salheddine Kabou and Nabil Laiche.
Methodology: Rabia Nichani, Laid Gasmi, Salheddine Kabou and Nabil Laiche.

Investigation: Rabia Nichani, Laid Gasmi, Salheddine Kabou and Nabil Laiche.
Discussion of results: Rabia Nichani, Laid Gasmi, Salheddine Kabou and Nabil Laiche.
Writing – Original Draft: Rabia Nichani, Laid Gasmi, Salheddine Kabou and Nabil Laiche.
Writing – Review and Editing: Rabia Nichani, Laid Gasmi, Salheddine Kabou and Nabil Laiche.
Resources: Rabia Nichani, Laid Gasmi, Salheddine Kabou and Nabil Laiche.
Supervision: Rabia Nichani, Laid Gasmi, Salheddine Kabou and Nabil Laiche.
Approval of the final text: Rabia Nichani, Laid Gasmi, Salheddine Kabou and Nabil Laiche.

VIII.ACKNOWLEDGMENTS

The authors would like to thank the editor and the anonymous referees for their careful comments and valuable suggestions that led to a substantial improvement of the presentation of the paper.

IX.REFERENCES

[1] H. Wasserbacher and M. Spindler, “Machine learning for financial forecasting, planning, and analysis: Recent developments and pitfalls,” Digit. Finance, vol. 4, no. 1, pp. 63–88, 2022.

[2] C. Zhang, N. N. A. Sjarif, and R. Ibrahim, “Deep learning models for price forecasting of financial time series: A review of recent advancements: 2020–2022,”Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 14, no. 1, p. e1519, 2024.

[3] T. Fischer and C. Krauss, “Deep learning with long short-term memory networks for financial market predictions,” Eur. J. Oper. Res., vol. 270, no. 2, pp. 654–669, 2018.

[4] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, …, and I. Polosukhin, “Attention is all you need,” Adv. Neural Inf. Process. Syst., vol. 30, 2017.

[6] Q. Wen, T. Zhou, C. Zhang, W. Chen, Z. Ma, J. Yan, and L. Sun, “Transformers in time series: A survey,” arXiv preprint arXiv:2202.07125, 2022.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” Adv. Neural Inf. Process. Syst., vol. 25, 2012.

[8] T. Brown, B. Mann, N. Ryder, M. Subbiah, …, and M. Murray, “Language models are few-shot learners,” Adv. Neural Inf. Process. Syst., vol. 33, pp. 1877–1901, 2020.

[9] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” arXiv preprint arXiv:1810.04805, 2018.

[10] Z. Zhang and S. Zohren, “Multi-horizon forecasting for limit order books: Novel deep learning approaches and hardware acceleration using intelligent processing units,” arXiv preprint arXiv:2105.10430, 2021.

[11] Z. Zhang, S. Zohren, and S. Roberts, “Deeplob: Deep convolutional neural networks for limit order books,” IEEE Trans. Signal Process., vol. 67, no. 11, pp. 3001–3012, 2019.

[12] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, “Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting,” in Proc. Int. Conf. Mach. Learn., PMLR, pp. 27268–27286, 2022.

[13] G. Avinash, V. Ramasubramanian, M. Ray, R. K. Paul, …, and M. A. Iquebal, “Hidden Markov guided deep learning models for forecasting highly volatile agricultural commodity prices,” Appl. Soft Comput., vol. 158, p. 111557, 2024.

[14] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A survey of convolutional neural networks: Analysis, applications, and prospects,” IEEE Trans. Neural Netw. Learn. Syst., vol. 33, no. 12, pp. 6999–7019, 2021.

[15] S. Meng, A. Chen, C. Wang, M. Zheng, F. Wu, X. Chen, H. Ni, and P. Li, “Enhancing exchange rate forecasting with explainable deep learning models,” arXiv preprint arXiv:2410.19241, 2024.

[16] J. Cao, Z. Li, and J. Li, “Financial time series forecasting model based on CEEMDAN and LSTM,” Physica A, vol. 519, pp. 127–139, 2019.

[17] S. Siami-Namini, N. Tavakoli, and A. S. Namin, “A comparison of ARIMA and LSTM in forecasting time series,” in Proc. 17th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA), pp. 1394–1401, 2018.

[18] I. Sutskever, O. Vinyals, and Q. V. Le, “Sequence to sequence learning with neural networks,” Adv. Neural Inf. Process. Syst., vol. 27, 2014.

[19] X. Wen and W. Li, “Time series prediction based on LSTM-attention-LSTM model,” IEEE Access, 2023.

[20] X. Zhang, X. Liang, A. Zhiyuli, S. Zhang, R. Xu, and B. Wu, “At-LSTM: An attention-based LSTM model for financial time series prediction,” in IOP Conf. Ser.: Mater. Sci. Eng., vol. 569, no. 5, p. 052037, 2019.

[21] A. Tsantekidis, N. Passalis, A. Tefas, J. Kanniainen, M. Gabbouj, and A. Iosifidis, “Using deep learning to detect price change indications in financial markets,” in Proc. 25th Eur. Signal Process. Conf. (EUSIPCO), pp. 2511–2515, 2017.

[22] P. N. Kolm, J. Turiel, and N. Westray, “Deep order flow imbalance: Extracting alpha at multiple horizons from the limit order book,” Math. Finance, vol. 33, no. 4, pp. 1044–1081, 2023.

[23] M. Ranjbar and M. Rahimzadeh, “Advancing gasoline consumption forecasting: A novel hybrid model integrating transformers, LSTM, and CNN,” arXiv preprint arXiv:2410.16336, 2024.

[24] R. Nichani, L. Gasmi, N. Laiche, and S. Kabou, “Optimizing financial time series predictions with hybrid ARIMA, LSTM, and XGBoost models,” Stud. Eng. Exact Sci., vol. 5, no. 2, p. e11188, 2024.

[25] X. Hu, “Stock price prediction based on temporal fusion transformer,” in Proc. Int. Conf. Mach. Learn. Big Data Bus. Intell. (MLBDBI), pp. 60–66, 2021.

[26] S. Sridhar and S. Sanagavarapu, “Multi-head self-attention transformer for Dogecoin price prediction,” in Proc. 14th Int. Conf. Hum. Syst. Interact. (HSI), pp. 1–6, 2021.

[27] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA: MIT Press, 2016.

[28] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv preprint arXiv:1412.6980, 2014.

[29] C. Han, H. Park, Y. Kim, and G. Gim, “Hybrid CNN-LSTM based time series data prediction model study,” in Proc. IEEE/ACIS Int. Conf. Big Data Cloud Comput. Data Sci. Eng., pp. 43–54, 2022.

[30] D. Patnaik, N. J. Rao, B. Padhiari, and S. Patnaik, “Optimised hybrid CNN-LSTM model for stock price prediction,” Int. J. Manag. Decis. Mak., vol. 23, no. 4, pp. 438–460, 2024.

[31] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 770–778, 2016.

[32] G. H. Nayak, M. W. Alam, G. Avinash, R. R. Kumar, M. Ray, S. Barman, …, and J. Bisen, “Transformer based deep learning architecture for time series forecasting,” Software Impacts, vol. 22, 2024.

[33] F. Altché and A. de La Fortelle, “An LSTM network for highway trajectory prediction,” in Proc. 20th IEEE Int. Conf. Intell. Transp. Syst. (ITSC), pp. 353–359, 2017.

[34] S. Ray, A. Lama, P. Mishra, T. Biswas, S. S. Das, and B. Gurung, “An ARIMA-LSTM model for predicting volatile agricultural price series with random forest technique,” Appl. Soft Comput., vol. 149, p. 110939, 2023.

[35] N. Zhang, D. Lei, and J. F. Zhao, “An improved Adagrad gradient descent optimization algorithm,” in Proc. Chinese Autom. Congr. (CAC), pp. 2359–2362, 2018.

[36] J. Sharma, S. Soni, P. Paliwal, S. Saboor, P. K. Chaurasiya, …, and A. Afzal, “A novel long term solar photovoltaic power forecasting approach using LSTM with Nadam optimizer: A case study of India,” Energy Sci. Eng., vol. 10, no. 8, pp. 2909–2929, 2022.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 222-229. September/October, 2025.
DOI: https://doi.org/10.5935/jetia. v.11 n.55.2666

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

A DEEP LEARNING MODEL TO CLASSIFY FOREST FIRES BURNED AREAS USING SENTINEL-2 DATA IN ALGERIA.

OUAHAB Abdelwhab[1]

[1]Department of Mathematics and Computer Sciences, University of Ahmed Draia, Adrar, Algeria.

[1]https://orcid.org/0000-0003-0648-2947

Email: ouahab.abdelwhab@univ-adrar.edu.dz

ARTICLE INFO

Article History
Received: September 4, 2025
Revised: September 30, 2025
Accepted: October 6, 2025
Published: October 31, 2025

Keywords:
Deep learnig,
CNN,
Algeria,
Burned,
Forest fires.

ABSTRACT

Transfer learning involves using pre-trained CNN models, originally trained on large datasets like ImageNet, and fine-tuning them for specific tasks with smaller datasets. In this research, six pre-trained CNN models—VGG16, VGG19, DenseNet121, InceptionResNetV2, MobileNet, and MobileNetV2—were evaluated on a dataset comprising 30 plant species. The goal is to determine which transfer learning model performs best for plant species recognition. The forest fires in Algeria had a great impact on the public, economic, and environmental levels. Estimating the burned areas caused by this fire is essential. In this study, we suggest a new methodology based on deep learning using sentinel-2 images to classify the burned area that occurred in Algeria in 2022. This methodology uses a convolution neural network (CNN) to learn and classify the Sentinel-2 images into burned and unburned areas. The inputs of the proposed model are generated by calculating some spectral indices used in detecting the burned area. We measure the performance of the proposed model using accuracy, rappel, precision, and f1-score. The proposed method gives an accuracy of 0.97. We show that the proposed method has a high performance in detecting and classifying the burnt area.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Fire is one of the most important factors that lead to damage to the environment and climate change. Every year, many forests are exposed to fires in different countries of the world [1]. This damage affects the physical and chemical properties of the soil, biodiversity, soil fertility, air purity, and other environmental impacts [2]. It is difficult to assess those affected by fires and to identify precisely burned areas due to the large temporal and spatial variability and uncertainty of fire occurrence [3]. Currently, remote sensing techniques are widely used to monitor changes on the Earth's surface and assess the impact of these changes on the environment by providing rapid information to determine fire-damaged areas in an accurate and appropriate way [4], [5]. These techniques have a lot of benefits compared to traditional techniques, including a large coverage and multiple temporal and spatial scales [2]. This is due to its ability to produce reliable and fast results, as well as its capacity to rapidly identify burnt areas for post-fire rehabilitation activities [6]. It is difficult to accurately determine the extent of fire damage and the severity of the burn by using ground measurements only due to the high variability and uncertainty of forest fires spatially and temporally [7].

Satellite sensors that are multispectral, such as Sentinel-2 satellite and Landsat series utilize medium-to-high spatial resolution to offer dependable data with a high temporal resolution [8], [9]. Several recent studies have shown that various spectral bands are sensitive to significant changes in spectral radiance in response to burning. These spectral bands can detect burnt areas, burn severity, and vegetation changes with reliability. For instance, thermal bands, mid-infrared and near-infrared are sensitive to burn magnitude changes and are commonly used for fire effects studies in vegetated areas [7]. In this sense, spectral indexes have been used for detecting burnt areas such as the Normalized Burned Raio (NBR) index [10], the Normalized Difference Vegetation Index (NDVI) [11], and the Soil-adjusted vegetation index [12]. The issue with these methods is that they may not perform effectively under varying weather conditions. Furthermore, the utilization of indices to identify fire damage typically necessitates manual or semi-manual techniques, which involve setting thresholds that are dependent on the soil type and cannot be easily established [13]. In recent times, deep learning methods have

been utilized in various applications with significant achievements, and their popularity is increasing in diverse fields. The generality of deep learning architectures allows them to be used for a broad range of classification tasks. As a result, similar architectures are being increasingly utilized for multi-label classification [14], among of them estimation, classification, and detection of burnt area. One of the major challenges in developing a deep learning model for burned areas is the limited access to training data [13]. The availability of free data from the Sentinel-2 satellite had a great role in monitoring the Earth and fires [4]. In recent years, Algeria has witnessed a terrible increase in the number of forest fires. This rise can be related to the effects of global warming and change in weather patterns [8]. It could also be due to criminal acts. These fires caused the loss of property and lives. Therefore, it has become necessary to predict these fires, know their causes, and estimate the extent of the damage they have caused. In this study, we suggest a new technique based on deep learning to classify and estimate the burned area using Sentinel-2 images. This approach has a high capacity for estimating the burned area. The used images are provided before and after the fire. We will exploit the various data generated by the various indices for the detection of burned areas to feed the deep learning model.

II. MATERIALS AND METHODS

The Sentinel-2A and Sentinel-2B satellites were respectively launched and put into orbit, in 2015 and 2017. The Sentinel-2 images consist of 13 spectral bands that cover the Visible, Near-Infrared (NIR), and Short-Wave Infrared (SWIR) ranges, with spatial resolutions of 10, 20, and 60 meters [8]. Sentinel-2 is frequently updated with a temporal frequency of 5 days. Table 1 displays the spatial and spectral characteristics of the satellite images obtained by Sentinel-2.

Table 1: Sentinel-2 bands with their spatial and spectral characteristics.

Band number	Wavelength	Spectral name	Spatial Resolution
B1	443	Coastal Aerosol	60
B2	490	Blue	10
B3	560	Green	10
B4	665	Red	10
B5	705	Red EDG-1	20
B6	740	Red EDG-2	20
B7	783	Red EDG-3	20
B8	842	NIR	10
B8a	865	Narrow NIR	20
B9	945	Water Vapor	60
B10	1375	Cirrus	60
B11	1610	SWIR-1	20
B12	2190	SWIR-2	20
TCI	RGB	RGB	10

Source: Authors, (2025).

All data used in the study are provided by the Sentinel-2B satellite. We used two types of data from the same area: data before the fire and data after the fire. The research area is located in Northeastern of Algeria which had a fire in August 2022. This area is illustrated in Fig. 1. The post-fire and pre-fire were chosen close to each other in order to reduce the spectral variance as a result of climate change. The pre-fire image was provided on the 12th of August 2022, and the post-fire image was provided on the 21st of August 2022. Both of the two selected images are free of clouds and fog [14].

Figure 1: Location of the study area (Algeria).
Source: Authors, (2025).

II.1 THE PROPOSED METHOD

The proposed method uses Sentinel-2B images, before and after the fire, to classify their pixels into burned and unburned pixels. All bands are resampled to 20 m in order to make them the same size. In our tasks, Convolution Neural Network (CNN) is used to classify the whole image. We generate a vector of nine values V from each pixel. This vector is considered as input for CNN. The nine values are calculated using some of the most well-known indices that are sensitive to burning.
There are no single spectral indices that can be universally effective in all environments. The used indices are:

II.1.1 NBR (Normalized Burn Ratio)

The NBR [6] is an index that detects burnt areas by analyzing the variations in the way burnt green vegetation and unburned green vegetation absorb and reflect light. It is computed using two specific Sentinel-2 bands which are Band 8 and Band 12. The equation of NBR is computed as follows [15]:

$$\mathrm{NBR} = \frac{B_8 - B_{12}}{B_8 + B_{12}} \tag{1}$$

The NBR takes values between -1 and 1. If the value of this indices is close to one, this means that the vegetation is not burnt and is in good health, while if the indices is close to -1, this means that the vegetation may be burnt or its health is not good.

II.1.2 dNBR (differenced Normalized Burn Ratio)

The dNBR indices is an effective tool for identifying the extent of burned areas and the severity of the burns. It takes positive values for burned areas and negative values for unburned areas [6].

$$\mathrm{NBR} = \frac{B_8 - B_{12}}{B_8 + B_{12}} \tag{2}$$

II.1.3 NDVI (Normalized Difference Vegetation Index)

The NDVI is the most important indices. It is used to extract vegetation areas [16]. It can be calculated using the bands 4 and 8 as follows:

$$\mathrm{NDVI} = \frac{B_8 - B_4}{B_8 + B_4} \tag{3}$$

NDVI takes values from -1 to 1, positive values corresponding to vegetation areas.

II.1.4 GNDVI (Green Normalized Difference Vegetation

This is indices is used to determine stressed vegetation. It is very sensitive to chlorophyll variation. It is calculated using the green band (B_3) and the near infrared (B_8) as follows [17] :

$$\mathrm{GNDVI} = \frac{B_8 - B_3}{B_8 + B_3} \tag{4}$$

The proposed model contains 3 CNNs. The first CNN is used to feed vectors generated from pre-fire image. The second CNN is used to feed data from post-fire image. The last CNN is used to combine the last result of the convolution layers of CNN-1 and CNN-2. CNN-1 and CNN-2 are used to extract the features from images before and after the fire. CNN-1 and CNN-2 have similar layers. The structure of CNN-1, CNN-2 and CNN-3 is shown in Table 2. The flowchart of the proposed model is shown in Fig.2.

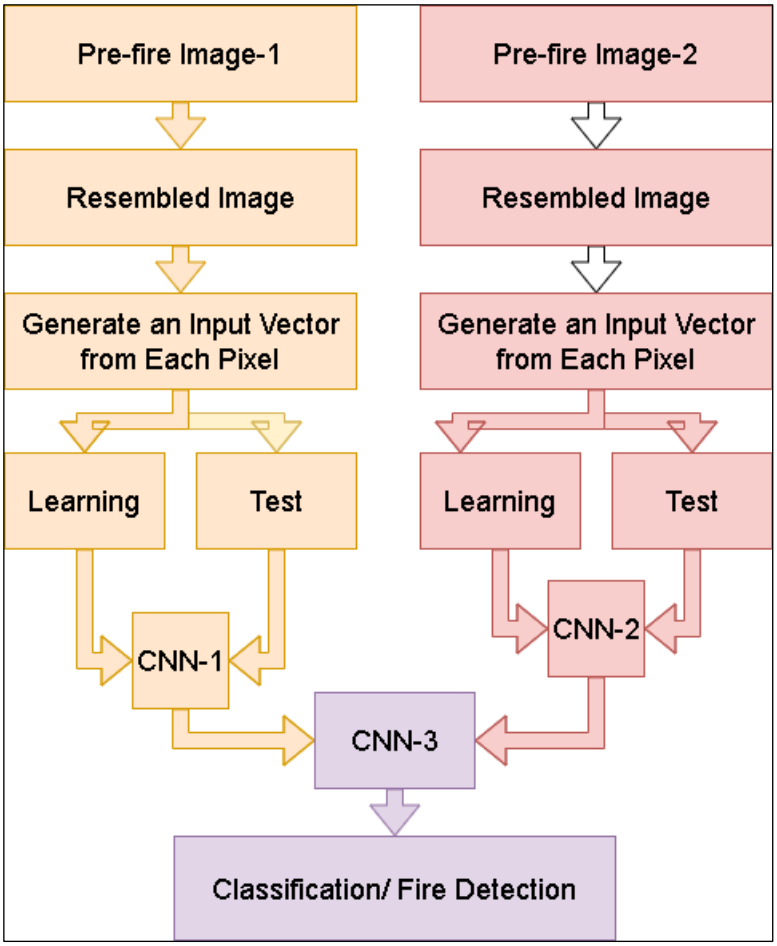


Figure 2: Flowchart of the proposed modal.
Source: Authors, (2025).

In order to perform the deep learning process, we select some pixels from the burned and unburned areas of the image under study. By using images before and after the fire, it is possible to discover in an automatic way the difference between the characteristics in the burned and unburned areas. After that, we calculate the various indices of fire detection corresponding to each pixel to form a vector of nine values that are used to feed the proposed model.

Table 2: The architecture of CNN-1, CNN-2 and CNN-3.

CNN-1 and CNN-2			
	Layer	**Kernel size**	**activation**
input	3*3*3		
Conv2D	16	3x3	relu
Conv2D	16	3x3	relu
MaxPooling2D		2x2	
Dropout(0.2)			
Conv2D	32	3x3	relu
Conv2D	32	3x3	relu
MaxPooling2D		2x2	
Dropout(0.2)			
Conv2D	64	3x3	relu
Conv2D	64	3x3	relu
MaxPooling2D		2x2	
Dropout(0.2)			
Flatten()			
CNN-3			
	Layer	Kernel size	activation
input	[FC1 FC2]		
Dense	64		relu
Dropout(0.2)			

Source: Authors, (2025).

The aim of using several indices instead of using one is to benefit from the characteristics of each index in detecting fire. We have determined the form of the input vector experimentally after a lot of tests and combinations. The structure of this vector is shown in Table 3.

Table 3: The structure of input vector.

NBR	NDVI	GNDVI	dNBR	NBR	NDVI	GNDVI	dNBR	dNBR

Source: Authors, (2025).

III. EXPERIMENT AND RESULTS

The experiment was carried out using TensorFlow and the Python 3.7 platforms. In our experiment, we selected 44418 pixels from the burned area and 44418 pixels from the unburned area. These pixels are used to train the proposed model. 66018 pixels are used to test the model. Half of them were taken from the burnt areas and the other half were taken from the unburned areas. Due to the lack of reference data in order to evaluate the validity of the proposed CNN model, a set of metrics are computed, which are the accuracy, precision, recall, F1-score, and loss function.

$$\text{Accurancy (A)} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{5}$$

$$\text{Precision (P)} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{6}$$

$$\text{Recall (R)} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{7}$$

$$\text{F1} - \text{Score (F1)} = 2 \times \frac{\text{P} \times \text{R}}{\text{P} + \text{R}} \tag{8}$$

TP : The number pixels of burned area classified as burned pixels, FP : The number pixels of burned area classified as unburned pixels. TN: The number pixels of unburned area classified as unburned pixels. FN : The number pixels of unburned area classified as burned pixels. The classified image is shown in figure 3.

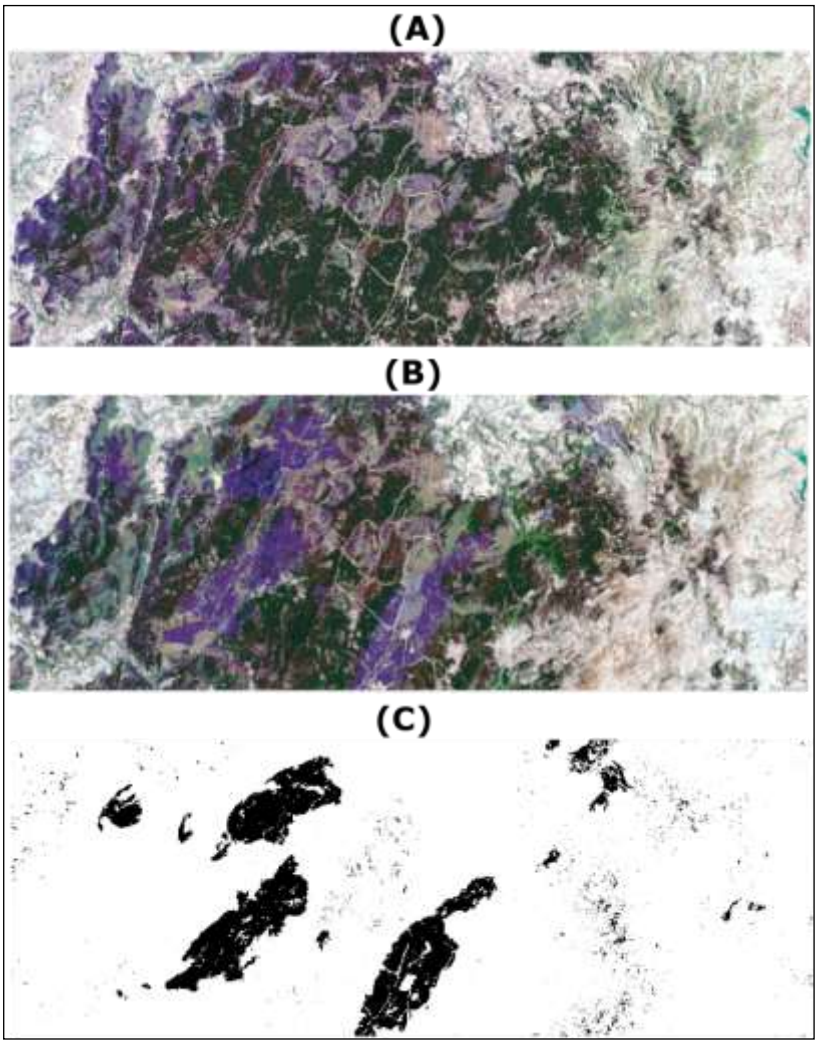


Figure 3: The sentinel-2 images of study area and its classification a) Image before the fire. b) Image after the fire c) the classified image using the proposed model.
Source: Authors, (2025).

Figure 4 and Figure 5 show the accuracy and loss functions for the proposed model that runs 100 iterations. It can be seen that the loss function decreases gradually and the accuracy function decreases gradually as the number of iterations increases. The accuracy values stabilize at 0.993. This indicates that the proposed model has a good performance.

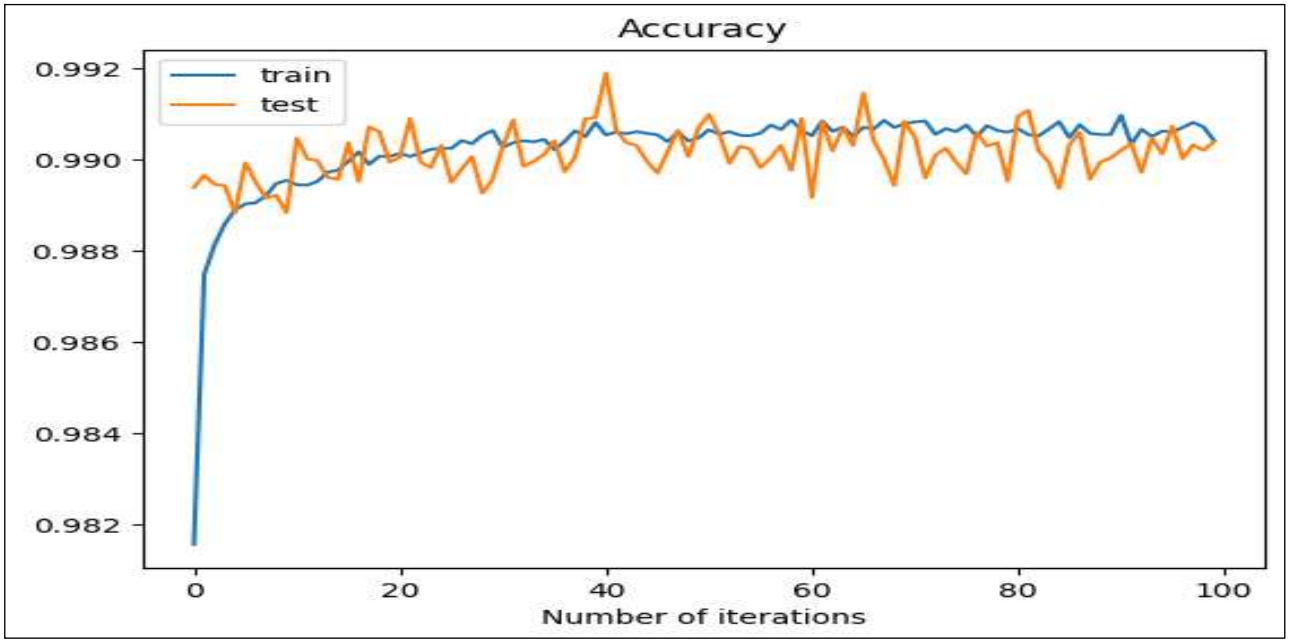


Figure 4: The accuracy functions of test and training abtained by the proposed model.
Source: Authors, (2025).

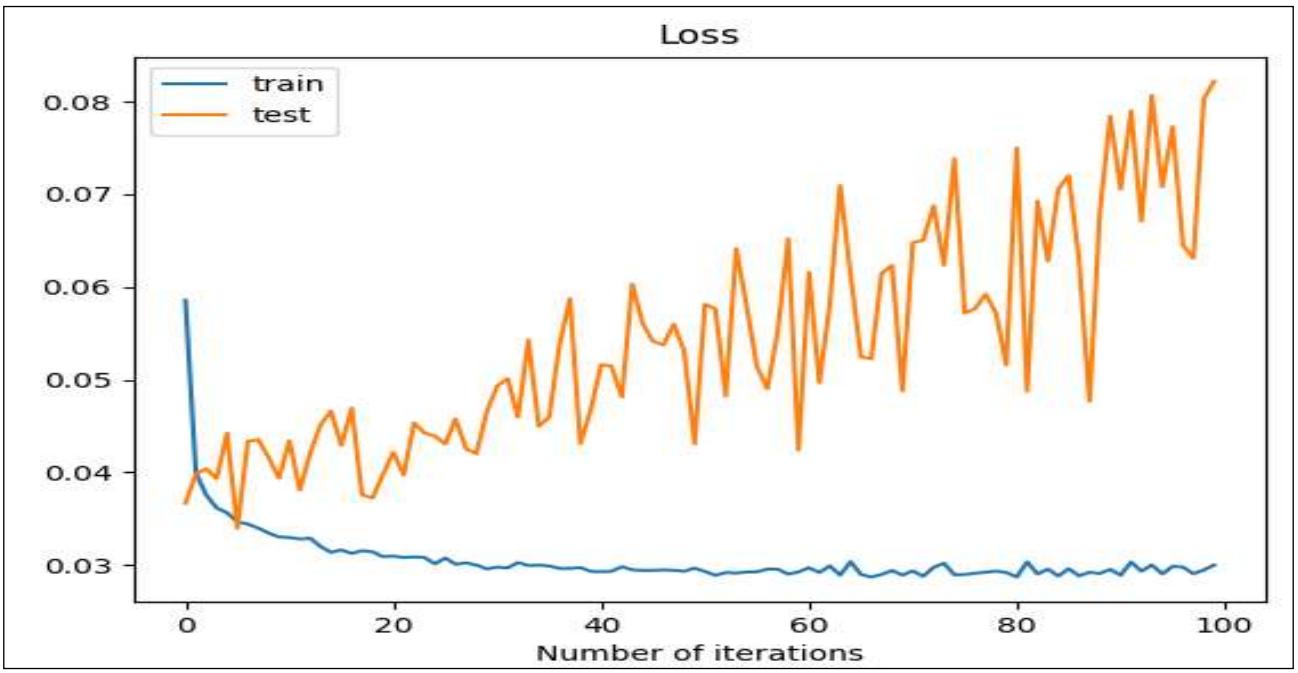


Figure 5: The loss functions of test and training abtained by the proposed model.
Source: Authors, (2025).

From Table 4, it can be seen that the precision is 0. 976. This can explain by the capacity of the proposed model to classify burned areas correctly. The value of accuracy is 0.98. This can be explained by the capacity of the proposed model to classify burned and unburned areas.

Table 4: Evaluation metrics.

Precision	Accuracy	Recall	F1-Score
0.976	0.98	0.97	0.97

Source: Authors, (2025).

Table 5 shows the confusion matrix obtained. In this matrix, the diagonal shows the number of pixels that are correctly classified. The others show the number of pixels that are incorrectly classified. It is noted that 96.7 % of burned pixels are correctly classified. The proposed model gives an error of 3.3% in estimating the burned areas.

Table 5: The confusion matrix.

Desired class % / Found class %	Burned pixels	Unburned pixel
Burned pixels	31919 (96.7%)	1090 (3.3%)
Unburned pixels	226 (0.69%)	32783 (99.31%)

Source: Authors, (2025).

IV. CONCLUSIONS

In the present work, we proposed a new methodology to classify the burned area using sentinel-2 images. The study area depends on the fire that happened in October 2022 in Algeria. We used a CNN with two inputs in order to enter inputs generated from images before and after the fire. The inputs are generated from the used data by computing some spectral indices, which are NDI, NBR, dNBR, and GNDVI. The performance of the proposed model is computed using accuracy, rappel, precision, and f1-score. The proposed method gives an accuracy of 0.97. This indicates that the proposed method has a high performance in classifying the burnt area.

V. AUTHOR'S CONTRIBUTION

Conceptualization: OUAHAB Abdelwhab.
Methodology: OUAHAB Abdelwhab.
Investigation: OUAHAB Abdelwhab.
Discussion of results: OUAHAB Abdelwhab.
Writing – Original Draft: OUAHAB Abdelwhab.
Writing – Review and Editing: OUAHAB Abdelwhab.
Resources: OUAHAB Abdelwhab.
Supervision: OUAHAB Abdelwhab..
Approval of the final text: OUAHAB Abdelwhab.

VI. REFERENCES

[1] T. Seyd, H. Mahdi, and C. Jocelyn, "Burnt-Net: Wildfire burned area mapping with single post-fire Sentinel-2 data and deep learning morphological neural network," Remote Sens., vol. 140, pp. 1–10, 2022.

[2] H. Luo and J. Wu An, "Assessment of the suitability of Sentinel-2 data for identifying burn severity in areas of low vegetation," J. Indian Soc. Remote Sens., 2022. [Online]. Available: https://doi.org/10.1007/s12524-022-01518-7

[3] X. Chen, J. E. Vogelmann, M. Rollins, D. Ohlen, C. H. Key, L. Yang, C. Huang, and H. Shi, "Detecting post-fire burn severity and vegetation recovery using multitemporal remote sensing spectral indices and field-collected composite burn index data in a ponderosa pine forest," Int. J. Remote Sens., vol. 32, no. 23, pp. 7905–7927, 2011.

[4] P. Zhang, X. Hu, and Y. Ban, "Wildfire-S1S2-Canada: A large-scale Sentinel-1/2 wildfire burned area mapping dataset based on the 2017–2019 wildfires in Canada," in Proc. IGARSS 2022 - IEEE Int. Geosci. Remote Sens. Symp., pp. 7954–7957, Jul. 2022, doi: 10.1109/IGARSS46834.2022.9884275.

[5] F. Filipponi, "Exploitation of Sentinel-2 time series to map burned areas at the national level: A case study on the 2017 Italy wildfires," Remote Sens., vol. 11, no. 6, pp. 622, Mar. 2019.

[6] R. Lorens, J. A. Sobrino, C. Fernández, J. Fernández-Alonso, and J. A. Vega, "A methodology to estimate forest fires burned areas and burn severity degrees using Sentinel-2 data: Application to the October 2017 fires in the Iberian Peninsula," Int. J. Appl. Earth Obs. Geoinf., vol. 95, pp. 102243, Jan. 2021.

[7] H. Farhadi, M. Mokhtarzade, and H. Ebadi, "Rapid and automatic burned area detection using Sentinel-2 time-series images in Google Earth Engine cloud platform: A case study over the Andika and Behbahan regions, Iran," Environ. Monit. Assess., vol. 194, pp. 369, Jun. 2022.

[8] M. Zanetti, "A high resolution burned area detector for Sentinel-2 and Landsat-8," in Proc. 10th Int. Workshop Analysis of Multitemporal Remote Sensing Images (MultiTemp), Shanghai, China, pp. 1–4, Jul. 2019, doi: 10.1109/Multi-Temp.2019.8866958.

[9] W. Chen, K. Moriya, T. Sakai, L. Koyama, and C. X. Cao, "Mapping a burned forest area from Landsat TM data by multiple methods," Geo. Hazards Risk, vol. 7, no. 1, pp. 384–402, Mar. 2014.

[10] A. R. Huete, H. Q. Liu, K. Batchily, and W. Leeuwen, "A comparison of vegetation indices global set of TM images for EOS-MODIS," Remote Sens. Environ., vol. 59, pp. 440–451, 1997.

[11] A. M. Al-Dabbagh and M. Ilyas, "Uni-temporal Sentinel-2 imagery for wildfire detection using deep learning semantic segmentation models," Geo. Nat. Hazards Risk, vol. 14, no. 1, pp. 1–15, 2023.

[12] O. Abdelwhab, "Multimodal convolutional neural networks for detection of COVID-19 using chest X-ray and CT images," Opt. Mem. Neural Netw., vol. 30, pp. 276–283, Jul. 2021.

[13] R. Zennir and B. Khallef, "Forest fire area detection using Sentinel-2 data: Case of the Beni Salah national forest Algeria," J. For. Sci., vol. 69, no. 1, pp. 1–10, 2023.

[14] K. Choudhary, W. Shi, and M. S. Boori, "Agriculture phenology monitoring using NDVI time series based on remote sensing satellites: A case study of Guangdong, China," Opt. Mem. Neural Netw., vol. 28, pp. 204–214, Mar. 2019.

[15] G. Navarro, I. Caballero, G. Silva, P. Parra, A. Vázquez, and R. Caldeira, “Evaluation of forest fire on Madeira Island using Sentinel-2A MSI imagery,” Int. J. Appl. Earth Obs. Geoinf., vol. 58, pp. 97–106, Feb. 2017.

[16] A. Craig, P. George, and F. Konstantinos, “Determining the use of Sentinel-2A MSI for wildfire burning and severity detection,” Int. J. Remote Sens., vol. 40, no. 3, pp. 905–930, Feb. 2018.

[17] M. Rahman, L. Di, E. Yu, L. Lin, C. Zhang, and J. Tang, “Rapid flood progress monitoring in cropland with NASA SMAP,” Remote Sens., vol. 11, no. 2, pp. 1–15, Jan. 2019.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 230-237. September/October, 2025.
DOI: https://doi.org/10.5935/jetia. v.11 n.55.2670

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

PERFORMANCE IMPROVEMENT IN REAL-TIME 3D RENDERING USING OPENGL FRAGMENT SHADER AND UNITY SURFACE SHADER

Eriq Muhammad Adams Jonemaro[1], Ivan Maulana Agusti[2], Herman Tolle[3], Muhammad Aminul Akbar[4] and Aryo Pinandito[5]

[1, 2, 3] Department of Informatics Engineering, Faculty of Computer Science, Brawijaya University, Malang, Indonesia.
[4, 5] Department of Information Systems, Faculty of Computer Science, Brawijaya University, Malang, Indonesia.

[1]https://orcid.org/0000-0001-8315-3210, [2]https://orcid.org/0009-0003-6453-3068, [3]https://orcid.org/0000-0002-1550-0513
[4]https://orcid.org/0000-0003-0794-7443, [5]0000-0002-8509-2383

Email: eriq.adams@ub.ac.id, ivanm63@student.ub.ac.id, emang@ub.ac.id, muhammad.aminul@ub.ac.id, aryo@ub.ac.id

ARTICLE INFO

Article History
Received: September 4, 2025
Revised: September 30, 2025
Accepted: October 7, 2025
Published: October 31, 2025

Keywords:
OpenGL fragment Shader,
Unity surface Shader,
Real-time 3D rendering,
GPU-based rendering,
Shader performance.

ABSTRACT

3D Objects with high geometric complexity are needed to create detailed visualization in video games. However, rendering 3D objects with high geometric complexity requires high computing resources, which can reduce the framerate. GPU-based rendering can make the rendering process efficient. This study uses comprehensive evaluation parameters to evaluate the performance of real-time 3D GPU-based rendering between Unity surface shader and OpenGL fragment shader, based on geometric complexity. The experiment was conducted by rendering various cube objects. The collected data, analyzed using statistical tests, showed that the OpenGL fragment shader had better overall performance in framerate, memory usage, and CPU usage, but more GPU Usage. Thus, this indicates that the OpenGL fragment shader relies more heavily on GPU processing power for rendering efficiency, making it better suited for desktop systems with dedicated GPUs, while Unity surface shaders may be more appropriate for systems with integrated graphics.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The video game industry has experienced rapid development in recent years [1] [2]. Nowadays, the video game industry drives in advancement of three-dimensional graphics technology [3]. Real-time 3D object rendering technology plays a significant role in creating a realistic visual experience in the world of video games [4]. Increasingly significant improvements in graphics hardware technology allow rendering 3D worlds with increasingly detailed visuals [5].

The more visual detail rendered increases the complexity of the geometry, resulting in greater computing resource requirements and longer rendering times [6]. High rendering time will reduce framerate during rendering [7]. A high Framerate can create an immersive and seamless experience [8]. GPU-based rendering can result in higher framerates compared to CPU-based rendering [9]. It can be implemented using various platforms' game engines and graphics API.

Unity is a popular and de facto game engine used by game developers [10]. OpenGL is a low-level graphics API widely used in graphics, CAD, and Virtual Reality [11] applications and can run on multiple platforms [12]. GPU-based rendering in the Unity platform can be implemented using Unity compute shader and Unity surface shader. It can also be implemented using the OpenGL compute shader and the OpenGL fragment shader on the OpenGL platform. Unity compute shaders and OpenGL compute shaders are programs designed for general-purpose computing (GPGPU) that are not limited to graphics computing [13] [14]. In contrast, Unity surface shaders and OpenGL fragment shaders are designed explicitly for graphics computing.

While existing studies have examined performance differences between rendering platforms, understanding shader-specific performance characteristics remains underexplored. Studies comparing Unity and OpenGL have mainly focused on GPGPU implementations through compute shaders [15] [16] for specialized applications such as particle systems and deformable object simulations using framerate as the primary performance metric. However, these investigations do not address graphics-specific shader programs—Unity surface shaders and OpenGL fragment shaders—which game developers routinely employ for rendering complex

lighting scenarios [17] [18]. Furthermore, previous comparative studies typically measured only framerate without considering other critical performance metrics such as CPU usage, GPU usage, and memory consumption. This single performance metric approach prevents developers from understanding performance-resource trade-offs, such as whether a higher framerate comes at the cost of increased CPU overhead or memory consumption, which are critical for matching rendering technology to specific hardware constraints and performance requirements.

Therefore, this study investigates the real-time performance of 3D object rendering between Unity surface shader and OpenGL fragment shader based on geometry complexity, with comprehensive evaluation parameters including framerate, CPU usage, GPU usage, and memory usage. The results provide comprehensive guidelines for game developers to determine the platform and rendering technology that matches their hardware constraints and performance requirements.

II. LITERATURE REVIEW

II.1 GRAPHICS API PERFORMANCE COMPARISONS

GPU-based rendering is widely used to make the rendering process more efficient with computational parallelization that utilizes the advanced multi-core processing of the GPU. The large selection of GPU-based rendering technologies and platforms makes it difficult for game developers to choose appropriate platforms and technologies. Therefore, many studies investigate rendering performance in various technologies and platforms. Several studies investigating the performance of OpenGL in various specific development platforms, such as the one conducted by [19], show that GLSL performs better than WebGL in augmented reality applications. Meanwhile, OpenGL ES and OpenGL performance using mobile GPU and desktop GPU are explored by [20].

Furthermore, some studies evaluate the performance between OpenGL and other graphics APIs, such as [21], where OpenGL has more efficient energy consumption than Vulkan. However, it has lower performance than Vulkan using Nvidia's Jetson Nano. A similar study [22] showed that Vulkan has a higher framerate than OpenGL ES in big.LITTLE architecture. Additionally, based on studies conducted by [23], OpenGL is faster than DirectX, and Vulkan uses integrated graphics, while Vulkan is the fastest for complex models, while DirectX is the fastest for simple models using discrete graphics.

These API evaluations reveal important performance differences across platforms and hardware configurations. However, these studies evaluate native API implementations rather than game engine shader programs. The study that evaluates the performance characteristics of the Unity surface shaders and OpenGL fragment shaders—the graphics-specific shaders commonly used for Material rendering—across varying geometric complexity with comprehensive performance metrics remains underexplored.

II.2 UNITY AND OPENGL RENDERING PERFORMANCE

In addition to the OpenGL performance evaluation study with the graphics API, several studies evaluate the performance of OpenGL and Unity, such as [24], showing that LibGDX, which utilizes OpenGL, has lower CPU and memory usage than Unity. Several studies were also conducted to evaluate the performance of Unity compute shader, where Unity compute shader has higher performance than CPU-based rendering in the simulation of deformable body volume [25] and volumetric deformable object simulation [26]. In addition, Unity compute shader also has a higher framerate than OpenGL compute shader [16]. However, in deformable object simulation, Unity compute shader has a lower framerate than OpenGL compute shader [15].

While these studies provide valuable insights into Unity and OpenGL performance, critical limitations prevent an understanding of graphics rendering shader performance. Studies [15][16][25][26] use compute shaders for physics simulation rather than graphics shaders for material rendering, and measure combined simulation plus rendering workloads without isolating rendering performance characteristics. Study [24] tested 2D sprites with aggregate CPU and RAM metrics but no 3D shader analysis or GPU profiling. Across all studies, none examines Unity surface shaders versus OpenGL fragment shaders. They do not systematically vary geometric complexity to characterize shader performance, nor provide comprehensive CPU, GPU, and memory profiling simultaneously for rendering workloads. These gaps necessitate an investigation of Unity surface shader and OpenGL fragment shader performance across varying geometric complexity with multi-dimensional performance metrics.

III. METHODOLOGY

The research methodology flow in this study is shown in Figure 1. This study begins by formulating problems and research objectives, focusing on analyzing the performance of Unity surface Shader and OpenGL fragment Shader using multiple test parameters. Then proceed with the design of experiments and development of benchmark applications. The benchmark application evaluates Unity surface shader and OpenGL fragment shader performance. Performance data collection is done through the MSI Afterburner monitoring application. Finally, the data that has been collected will be analyzed using inferential statistics.

III.1 BENCHMARK APPLICATION DEVELOPMENT

The procedure for rendering cube objects with varying amounts will be implemented to simulate the rendering of 3D objects with various levels of geometrical complexities. Two benchmark applications are developed: one implemented using Unity surface shader and the other implemented using OpenGL fragment shader. The Unity surface shader benchmark application was developed using Unity 2022.3.22f1 with the C# programming language. Meanwhile, the OpenGL fragment shader benchmark application was developed using OpenGL 3.3 and the C++programming language. Implementation of the rendering process in both benchmark applications is made as identical as possible by providing cube dimension constraints using 1x1x1 units using the Phong-lighting model and shadow mapping. The flow of the rendering implementation procedure of both benchmark applications can be seen in Figures 2 and 3, respectively. Generally, the rendering process in both benchmark applications begins with creating cube objects by defining vertices, normals, and

UVs. Afterwards, the cube object's mesh data is sent to the GPU, which will later perform lighting calculation, and finally, the object is rendered on the screen.

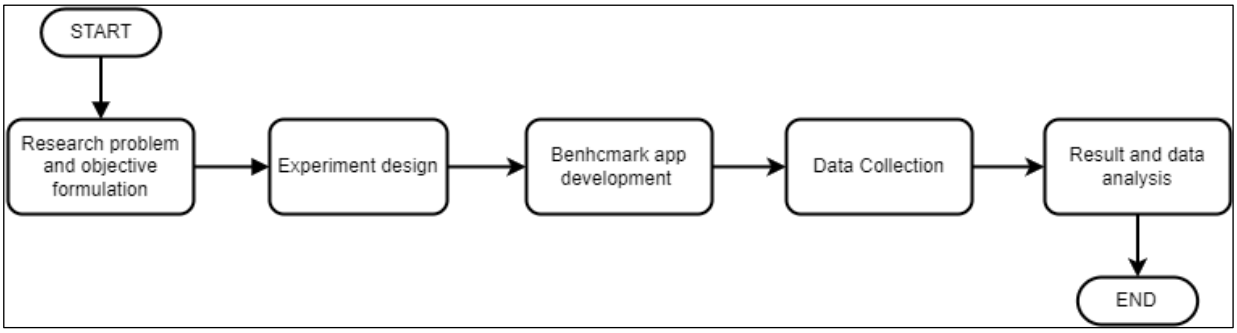


Figure 1: Research methodology flow.
Source: Authors, (2025).

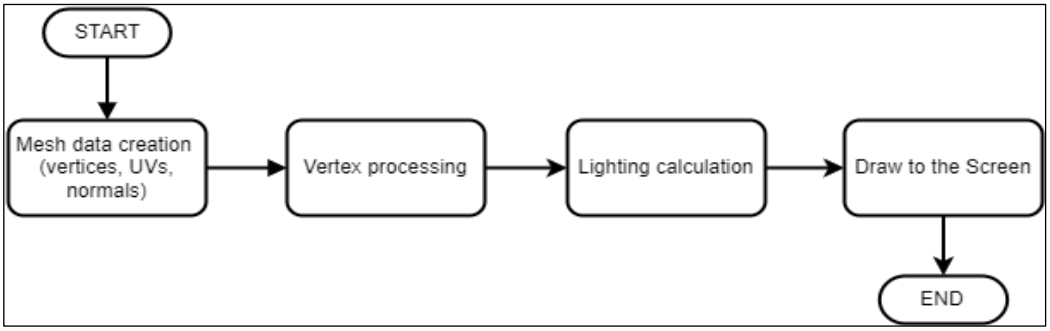


Figure 2: Rendering procedure using Unity surface shader.
Source: Authors, (2025).

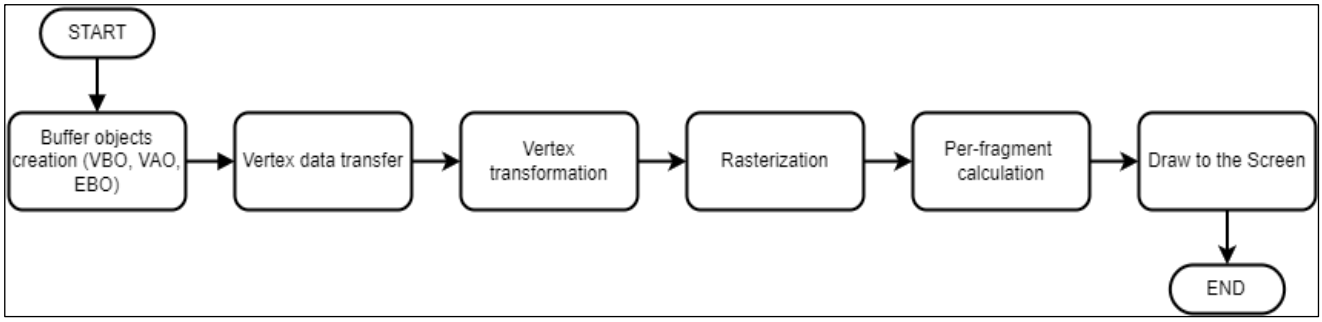


Figure 3: Rendering procedure using OpenGL fragment shader.
Source: Authors, (2025).

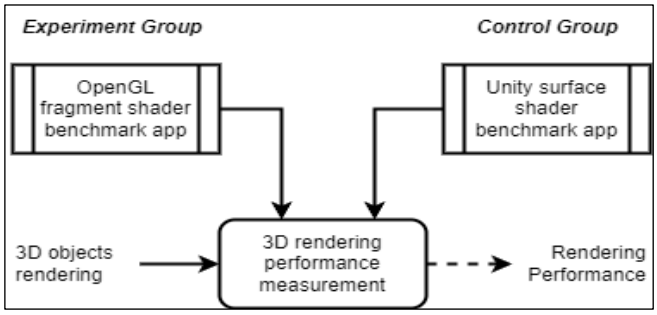


Figure 4: Experimental flow.
Source: Authors, (2025).

III.2 EXPERIMENT DESIGN

Figure 4 shows the experimental design of this study. 3D rendering performance will be compared using an OpenGL fragment shader and a Unity surface shader. Each shader is evaluated using a 3D cube object with a varying number of objects. This study used cube objects for evaluation, following established GPU benchmarking practices [27]. Using cubes for evaluation provide several advantages: (1) all cubes have the same shape and vertex count, ensuring consistent geometry; (2) complexity can be easily adjusted by adding more cubes (800 to 40,000); (3) cubes can be created programmatically without external 3D files, making the experiment reproducible on any system; and (4) both Unity and OpenGL can generate identical cubes. This controlled setup allows a clear comparison

of shader performance. The rendering performance in memory usage, CPU usage, GPU usage, and framerate is measured and analyzed statistically to compare shader performance in a desktop environment.

III.3 DATA COLLECTION

Data was collected by running two benchmark applications and the MSI Afterburner monitoring application. The tests involved varying the number of Cube objects from 800 to 40000, with two increments. Each test lasted for 1 minute on a Windows 11 64-bit operating system, using an Intel Core i5-12500H processor, NVIDIA RTX 3050 GPU, 1TB SSD, and 16GB of DDR4 RAM. The baseline resolution used for the tests was 1280x720 (HD). 720p was selected in this study for wide reproducibility in various hardware specifications, as aligned with this study [28]. Furthermore, this 720p is widely used in benchmark applications such as Unigine Superposition Benchmark [29].

III.4 DATA ANALYSIS

Log data obtained from the MSI Afterburner application will be processed and stored in spreadsheet format. The Shapiro-Wilk test is performed to determine whether the data is normally distributed. If the data is normally distributed, it will be used as a parametric test using the t-test, while the opposite will be a non-parametric test using the Mann Whitney U test. The t-test and the Mann-Whitney U test are used to determine if there is a significant difference between the performance of Unity surface shader and OpenGL fragment shader—all statistical analyses conducted using R Studio.

IV. RESULTS AND DISCUSSIONS

The experimental results from 50 trials for each shader show significant performance differences between the OpenGL fragment shader and the Unity surface shader in all evaluation parameters. A descriptive statistic of the performance measurements of the OpenGL fragment shader and the Unity surface shader is shown in Table 1. The normality test using the Shapiro-Wilk test shows that all performance data in all evaluation parameters are not in normal distribution (*p-value* <.05) (Tabel 2). Therefore, a non-parametric test with the Mann-Whitney U-test is used to determine significant differences in performance data in all evaluation parameters between the OpenGL fragment shader and the Unity surface shader. The results of the Mann-Whitney U-Test are shown in Table 3, where there are significant differences (*p-value* <.05) in all performance data from all evaluation parameters.

Based on the test results, OpenGL fragment shaders have an average frame rate 5.8 times higher than Unity surface shaders. In average memory usage, OpenGL fragment shaders are 4.4 times more efficient than Unity surface shaders. OpenGL fragment shaders also have an average CPU usage 1.6 times lower than Unity surface shaders. However, OpenGL fragment shaders have an average GPU usage 7.8 times higher than Unity surface shaders. The rendering comparison performance in framerate, CPU usage, GPU usage, and memory usage is shown in Figures 5, 6, 7, and 8, respectively.

Table 1: Descriptive statistics of OpenGL fragment shader and Unity surface shader performance measurements.

Shader	Parameter	Mean	SD	Min	Max	Median
Fragment Shader	CPU Usage (%)	13.06	1.74	11.00	18.00	13.00
Surface Shader	CPU Usage (%)	21.26	1.26	19.00	24.00	21.00
Fragment Shader	Framerate (FPS)	101.78	95.98	34.00	492.00	62.00
Surface Shader	Framerate (FPS)	17.46	29.57	4.00	181.00	7.00
Fragment Shader	GPU Usage (%)	31.12	5.82	26.00	52.00	27.50
Surface Shader	GPU Usage (%)	4.30	2.61	3.00	18.00	3.00
Fragment Shader	Memory Usage (MB)	110.00	1.07	107.00	111.00	110.00
Surface Shader	Memory Usage (MB)	485.20	146.83	237.00	715.00	492.00

Source: Authors, (2025).

Table 2: The Shapiro-Wilk test of performance data between the OpenGL fragment shader and the Unity surface shader.

Shader	Parameter	*p-value*
Fragment Shader	CPU Usage (%)	1.97×10^{-08}
Surface Shader	CPU Usage (%)	8.77×10^{-03}
Fragment Shader	Framerate (FPS)	4.00×10^{-09}
Surface Shader	Framerate (FPS)	4.40×10^{-12}
Fragment Shader	GPU Usage (%)	5.19×10^{-07}
Surface Shader	GPU Usage (%)	3.60×10^{-11}
Fragment Shader	Memory Usage (MB)	6.46×10^{-07}
Surface Shader	Memory Usage (MB)	1.56×10^{-02}

Source: Authors, (2025).

Table 3: Mann-Whitney U-Test results of OpenGL fragment shader and Unity surface shader performance measurements.

Parameter	*p-value*
CPU Usage (%)	2.52909×10^{-18}
GPU Usage (%)	2.00599×10^{-18}
Memory Usage (MB)	3.87951×10^{-18}
Framerate (FPS)	1.56527×10^{-14}

Source: Authors, (2025).

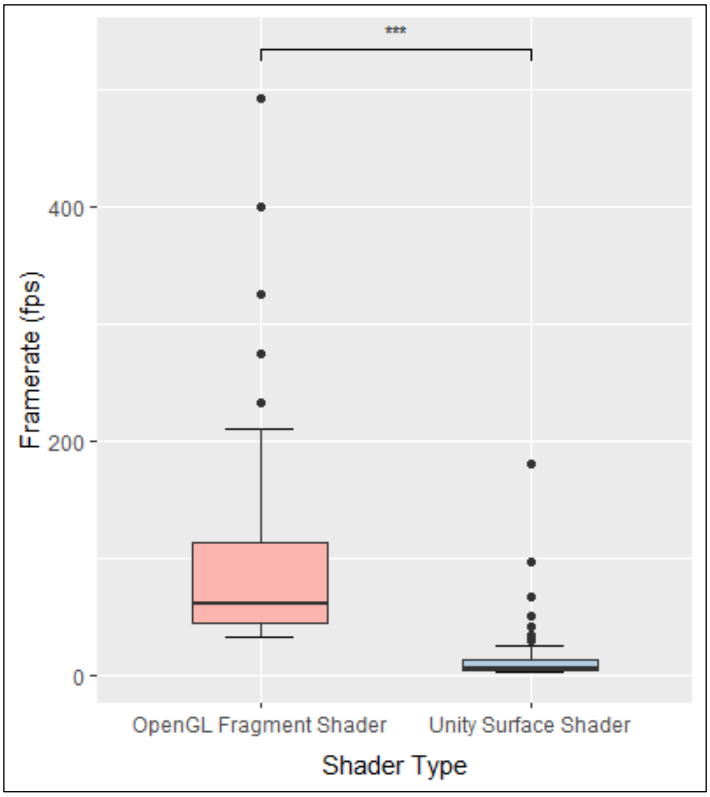


Figure 5. Comparison of the framerate in rendering between the OpenGL fragment shader and the Unity surface shader.
Source: Authors, (2025).

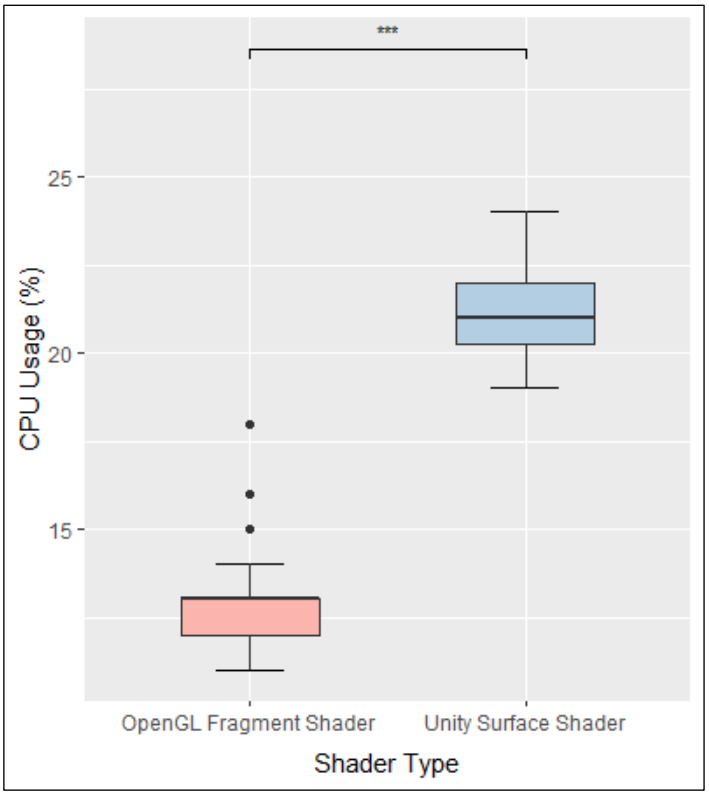


Figure 6: Comparison of CPU usage in rendering between the OpenGL fragment shader and the Unity surface shader.
Source: Authors, (2025).

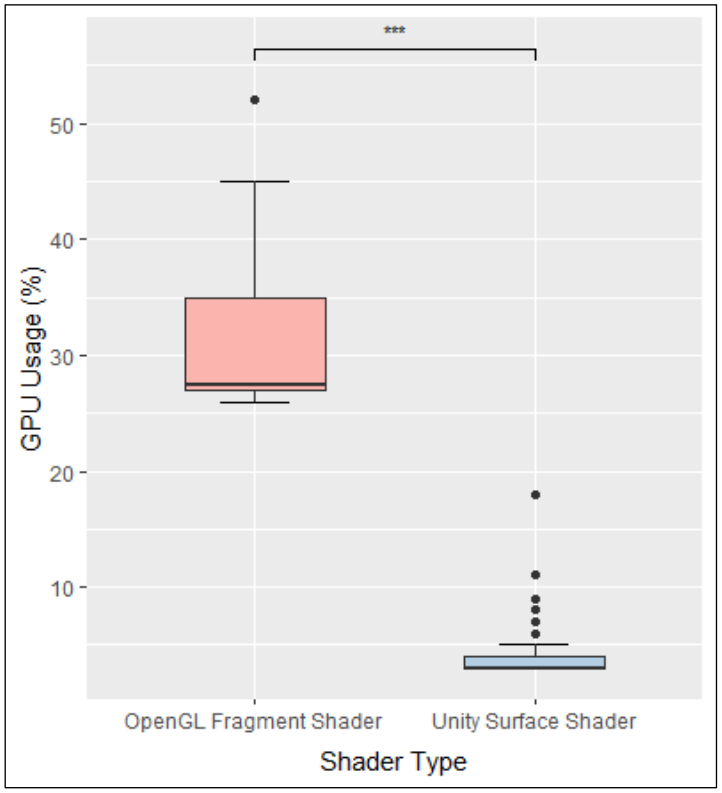


Figure 7: Comparison of GPU usage in rendering between the OpenGL fragment shader and the Unity surface shader.
Source: Authors, (2025).

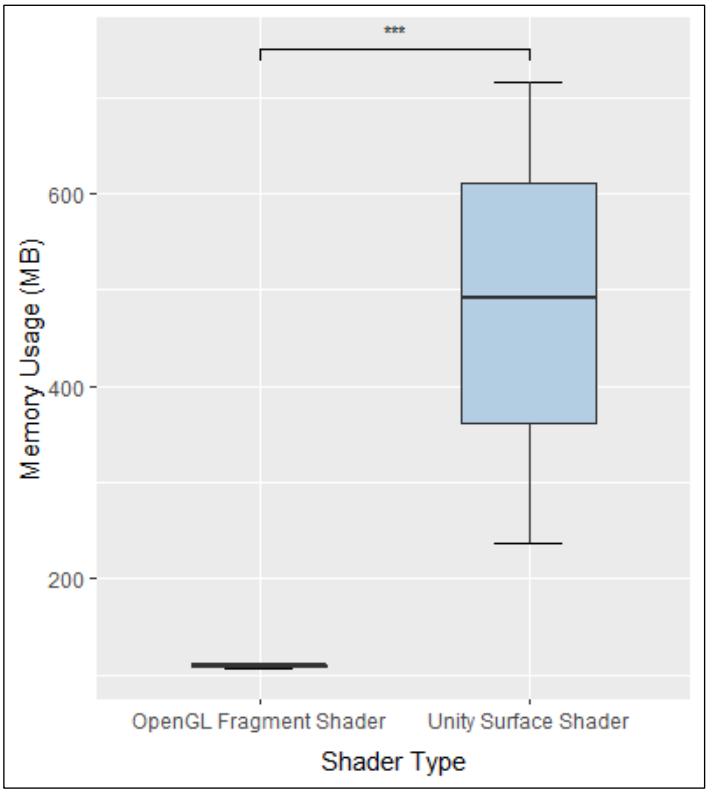


Figure 8: Comparison of memory usage in rendering between the OpenGL fragment shader and the Unity surface shader.
Source: Authors, (2025).

The higher GPU utilization of OpenGL fragment shaders leads to reduced CPU usage. Lower CPU and memory usage of OpenGL fragment shaders in this study aligns with the findings in [24]. Meanwhile, the higher frame rate of OpenGL fragment shaders in this study is also aligned with the experiments by [16]. However, in contrast to the findings obtained by [15], OpenGL compute shaders have lower frame rates than Unity compute shaders in particle simulations. These different findings indicate that performance characteristics depend highly on specific aspects of the rendering implementation.

The results suggest practical implications for platform selection. The higher GPU utilization in OpenGL fragment shaders indicates that they are better suited for desktop systems with dedicated GPUs. In contrast, Unity surface shaders are better suited for systems with integrated graphics. However, the higher GPU usage may impact power consumption, which is not measured in this study.

V. LIMITATIONS

Although this study provides valuable insights, this study has several limitations that must be addressed. First, cube objects used in evaluation may not fully represent real-world 3D models with various polygon counts and mesh complexities. Second, the experiment was conducted on a single desktop configuration, which can be extended to various configurations and platforms. Third, baseline 720p resolution may not scale linearly to higher and common used resolutions in modern applications such as 1080p or 4K. Additionally, this study focused on Phong lighting and shadow mapping, without evaluating advanced techniques such as physically-based rendering or global illumination. Framerate consistency, power consumption, particularly relevant for sustained workloads and mobile deployments, were not measured in the performance evaluation. Hence, addressing these limitations by evaluating real-world 3D models in various hardware configurations, platforms, and resolutions, while incorporating additional performance metrics, is needed. Thus, it can provide more comprehensive guidelines for shader technology selection.

VI. CONCLUSIONS

Based on the experimental results, OpenGL fragment shaders have better overall performance in framerate, lower memory usage, and lower CPU usage than Unity surface shaders. However, OpenGL fragment shaders required significantly more GPU usage, indicating that they rely more heavily on GPU processing power for optimal performance in their rendering process. Furthermore, OpenGL fragment shaders are particularly well-suited for desktop systems with dedicated GPUs and Unity surface shaders may be better suited for systems with integrated graphics. Thus, developers should take into account the capabilities of the target hardware, the performance requirements, and any specific resource limitations.Furthermore, these findings indicate that OpenGL fragment shaders can guide game developers in choosing the most appropriate rendering technology to build efficient games and graphics applications. However, further investigation must be extended using real-world 3D models across various hardware configurations, platforms, and resolutions. It should also include a wide range of performance metrics, such as frame consistency and power consumption, to provide more comprehensive guidelines for selecting shader technologies.

VII. AUTHOR'S CONTRIBUTION

Conceptualization: Eriq Muhammad Adams Jonemaro, Ivan Maulana Agusti, and Herman Tolle.
Methodology: Eriq Muhammad Adams Jonemaro, Ivan Maulana Agusti, and Herman Tolle.
Investigation: Eriq Muhammad Adams Jonemaro, Ivan Maulana Agusti.
Discussion of results: Eriq Muhammad Adams Jonemaro and Ivan Maulana Agusti.
Writing – Original Draft: Eriq Muhammad Adams Jonemaro and Ivan Maulana Agusti.
Writing – Review and Editing: Eriq Muhammad Adams Jonemaro and Aryo Pinandito.
Resources: Eriq Muhammad Adams Jonemaro, Ivan Maulana Agusti, and Muhammad Aminul Akbar.
Supervision: Aryo Pinandito.
Approval of the final text: Eriq Muhammad Adams Jonemaro, Ivan Maulana Agusti, Herman Tolle, Muhammad Aminul Akbar, and Aryo Pinandito.

VIII. ACKNOWLEDGMENTS

This research is part of the Scientific Article Publication Acceleration Program of the Faculty of Computer Science, Brawijaya University.

IX. REFERENCES

[1] H. Xiao, Y. Xia, and A. Baird, "Predicting digital product performance with team composition features derived from a graph network," *Decis. Support Syst.*, vol. 183, no. 114266, p. 114266, Aug. 2024.

[2] T. Kraemer, W. H. Weiger, and S. Heidenreich, "Do all stars shine the same? investigating the nonlinear effects of user and critic reviews on video game sales," *J. Bus. Res.*, vol. 188, no. 115034, p. 115034, Feb. 2025.

[3] Y. Lan, "Development and application of 3D modeling in game," *Academic Journal of Science and Technology*, vol. 7, no. 2, pp. 94–97, Sep. 2023.

[4] M. Casini, "Advanced digital design tools and methods," in *Construction 4.0*. Elsevier, 2022, pp. 263–334.

[5] R. M. Reffat, "A virtual platform for improving coordination and promoting cooperation on traffic safety," *Int. J. Transp. Sci. Technol.*, vol. 3, no. 1, pp. 43–62, Mar. 2014.

[6] B. Kuth, M. Oberberger, F. Kawala, S. Reitter, S. Michel, M. Chajdas, and Q. Meyer, "Real-time meshlet decompression," *Comput. Graph.*, no. 104292, p. 104292, Jul. 2025.

[7] N. Komodakis, C. Panagiotakis, and G. Tziritas, "3D visual reconstruction of large scale natural sites and their fauna," *Signal Process. Image Commun.*, vol. 20, no. 9-10, pp. 869–890, Oct. 2005.

[8] P. Wu, Y. Liu, H. Chen, X. Li, and H. Wang, "VR-empowered interior design: Enhancing efficiency and quality through immersive experiences," *Displays*, vol. 86, no. 102887, p. 102887, Jan. 2025.

[9] M. Yuan, L. Wang, and B. Ai, "Parallel terrain node processing and rendering: a GPU-accelerated approach," in *International Conference on Remote Sensing, Mapping, and Image Processing (RSMIP 2025)*, F. Tosti and R. Alvarez, Eds.SPIE, May 2025, p. 45.

[10] T. M. Moreira, J. G. de Faria, Jr, P. O. S. Vaz-de Melo, and G. Medeiros-Ribeiro, "Development and validation of an AI-Driven model for the la rance tidal barrage: A generalisable case study," *Appl. Energy*, vol. 332, no. 120506, p. 120506, Feb. 2023.

[11] Y. Liu, Z. Xie, and H. Liu, "Three-line structured light vision system for non-cooperative satellites in proximity operations," *Chin. J. Aeronaut.*, vol. 33, no. 5, pp. 1494–1504, May 2020.

[12] S. Liu, S. Tian, Z. Zhang, L. Liu, and T. Li, "Dynamic growth tomato inflorescence modeling with elastic mechanics data," *Smart Agric. Technol.*, vol. 11, no. 100884, p. 100884, Aug. 2025.

[13] D. Algis, B. Bramas, E. Darles, and L. Aveneau, "InteropUnityCUDA: A tool for interoperability between unity and CUDA," *Softw. Pract. Exp.*, vol. 55, no. 6, pp. 1127–1141, Jun. 2025.

[14] S. Park and N. Baek, "A shader-based ray tracing engine," *Appl. Sci. (Basel)*, vol. 11, no. 7, p. 3264, Apr. 2021.

[15] M. S. Kim, N.-J. Sung, Y.-J. Choi, and M. Hong, "Performance comparison of particle simulation using gpu between opengl and unity," *KIPS Transactions on Software and Data Engineering*, vol. 6, no. 10, pp. 479–486, 2017.

[16] M.-H. Choi, M. S. Kim, N.-J. Sung, Y.-J. Choi, and M. Hong, "Comparison in performance of parallel deformable object simulation between opengl and unity," *KIPS Trans. Softw. Data Eng*, vol. 6, pp. 479–486, 2017.

[17] A. Chrysanthakopoulou and K. Moustakas, "Real-time shader-based shadow and occlusion rendering in AR," in *2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, Mar. 2024.

[18] J. Zhu, Z. Wu, Q. Zhang, C. Liao, and Z. Huang, "WishGI: Lightweight static global illumination baking via spherical harmonics fitting," *ACM Trans. Graph.*, vol. 44, no. 4, pp. 1–12, Aug. 2025.

[19] M. Hamzaturrazak, E. M. A. Jonemaro, and A. Pinandito, "Performance analysis of 3D rendering method on web-based augmented reality application using WebGL and OpenGL shading language," in *Proceedings of the 8th International Conference on Sustainable Information Engineering and Technology*. New York, NY, USA: ACM, Oct. 2023.

[20] J. McCaffrey, "Exploring mobile vs. desktop opengl performance," pp. 337–352, 2012.

[21] O. Ferraz, P. Menezes, V. Silva, and G. Falcao, "Benchmarking vulkan vs OpenGL rendering on Low-Power edge GPUs," in *2021 International Conference on Graphics and Interaction (ICGI)*. IEEE, Nov. 2021.

[22] M. Lujan, M. McCrary, B. W. Ford, and Z. Zong, "Vulkan vs OpenGL ES: Performance and energy efficiency comparison on the big.LITTLE architecture," in *2021 IEEE International Conference on Networking, Architecture and Storage (NAS)*. IEEE, Oct. 2021.

[23] M. Y. Protsyk and T. O. Korotyeyeva, "Analysis the efficiency of programming tools for 3d rendering in video game development," *Scientific Bulletin of UNFU*, vol. 35, no. 3, pp. 158–165, May 2025. [Online]. Available: https://nv.nltu.edu.ua/index.php/journal/article/view/2762

[24] P. Kosidło, K. Kowalczyk, and M. Badurowicz, "Comparison of capabilities of the unity environment and LibGDX in terms of computer game development," *jcsi*, vol. 21, pp. 324–329, Dec. 2021.

[25] H. Va, J. Ma, and M. Hong, "Comparative study of iteration and coefficient influence on volume preservation in position-based dynamics simulation," in *2023 IEEE 6th International Conference on Knowledge Innovation and Invention (ICKII)*. IEEE, Aug. 2023.

[26] H. Va, M.-H. Choi, and M. Hong, "Efficient simulation of volumetric deformable objects in Unity3D: GPU-accelerated Position-Based dynamics," *Electronics (Basel)*, vol. 12, no. 10, p. 2229, May 2023.

[27] F. Fossum, "Real-time rigid body interactions," Master's thesis, Institutt for datateknikk og informasjonsvitenskap, 2011.

[28] J.-H. Nah, Y. Suh, and Y. Lim, "L-bench: An android benchmark set for low-power mobile gpus," Computers Graphics, vol. 61, pp. 40–49, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S009784931630111X

[29] D. Shergin, D. Vidiger, and A. Fofanova, "Superposition benchmark: innovative ssrtgi lighting in real time," in ACM SIGGRAPH 2017 Real Time Live!, ser. SIGGRAPH '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 25. [Online]. Available: https://doi.org/10.1145/3098333.3098335

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 238-249. September/October, 2025.

DOI: https://doi.org/10.5935/jetia. v.11 n.55.2673.

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

INVESTIGATION OF PARTIAL DEMAGNETIZATION FAULTS IN PERMANENT MAGNET SYNCHRONOUS MOTORS BASED ON EFFICIENCY AND LOSS MAPS ADVANCED ANALYSIS

Layachi Chebabhi[1], Toufik Tayeb Naas[2], Mohamed Zitouni[3], Monaem Elmnifi[4], and Abdelkarim Cherhabil[5]

[1] Laboratory of Applied Automation and Industrial Diagnostics (LAADI), Department of Mechanical Engineering in the Faculty of Science and Technology, Ziane Achour University of Djelfa, PO Box 3117, Djelfa 17000, Algeria.

[2,3] Laboratory of Gas Turbine Joint Research Team, Department of Mechanical Engineering in the Faculty of Science and Technology, Ziane Achour University of Djelfa, PO Box 3117, Djelfa 17000, Algeria.

[4] Department of Mechanical Engineering Technology, FSBEI "Belgorod State Technological University named after. V.G. Shukhov." 308012, Belgorod, st. Kostyukova 46. Russia.

[5] Laboratoire de Modélisation Simulation et Optimisation des Systèmes Complexes Réels, Department of Electronics and Communications, Ziane Achour University of Djelfa, PO Box 3047, Ain Chih, Djelfa, 17000, Algeria.

[1]https://orcid.org/0009-0002-4451-8533, [2]http://orcid.org/0000-0002-5539-6415, [3]https://orcid.org/0009-0003-3088-7431, [4]http://orcid.org/0000-0003-4074-1877, [5]http://orcid.org/0009-0006-1437-1453

Email: layachi.chebabhi@univ-djelfa.dz, t.naas@univ-djelfa.dz, mohamed.zitouni@gmail.com, monm.hamad@Yahoo.co.uk, Abdelkarim.cherhabil@univ-djelfa.dz

ARTICLE INFO

Article History

Received: September 5, 2025

Revised: September 30, 2025

Accepted: October 7, 2025

Published: October 31, 2025

Keywords:

Demagnetization,
Permanent Magnet Synchronous Motors,
Electromagnetic Analysis,
Efficiency Maps.

ABSTRACT

Demagnetization failure is a critical and prevalent fault in interior permanent magnet synchronous motors (IPMSMs), resulting from a variety of factors including thermal overheating, high electrical currents, and mechanical vibrations. Although the electromagnetic analysis and performance curves provides a deep insight into the machine behavior, it is not able to emulate the actual performance under various operating loads. To tackle these restrictions, advanced analytical tools such as efficiency maps are adopted to predict the behavior more accurately, which allows for enhanced performance monitoring. This paper aimed to incorporate electromagnetic analysis with efficiency maps to investigate the IPMSMs partial demagnetization fault, with emphasis on the magnetic flux residual density coefficient (Br) variations. The results indicated a significant degradation in the performance after exposure to the fault, with a 57% decrease in the generated torque, losses increased up to 35%, with a significant reduction in the range of high efficiency areas greater than 70% over the efficiency maps, and a deterioration in the efficiency especially at low speeds as the peak torque range is reduced. This study provides a comprehensive framework for evaluating the demagnetization faults effection, thus contributing to the elaboration of effective operation and maintenance strategies.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

IPMSMs are used in electromechanical power conversion technologies in high performance applications, dominating the sectors of traction systems for electric and hybrid vehicles [1-3], renewable energy systems for wind turbines as part of grid conversion systems [4,5], photovoltaic to mechanical energy conversion systems [6,7], advanced industrial automation platforms, Maglev high speed trains [8], electric aircraft and unmanned aerial vehicles (UAVs) [9,10], Distributed propulsion control systems [11], high resolution automated production lines, electric submarines, smart ships, and unmanned underwater vehicles (UUVs) [12-14].

The features of these machines include high power density, which allows reducing size and weight compared to other machines with the same power, constant speed range, low losses and efficiency up to 95% owing to the existence of internal magnets, the ability of operating in a wide range of speeds without significant losses and generate high torque which makes them suitable for rapid start up in electric vehicles [15], and the possibility of precise torque and speed control and compatibility with advanced control technologies such as direct torque control (DTC) and field oriented control (FOC) [16,17], Extended permanent magnet service lifetime and homogeneous flux distribution inside the rotor that eliminates magnet direct thermal conductivity, efficient dynamic performance and rapid response to load and transients that reduces noise and vibration. The operational reliability of permanent magnet synchronous motors (PMSM) systems is mainly related to the integrity of the permanent magnet magnetic field , despite the superior materials properties, they are susceptible to demagnetization effects under extreme operational conditions, which constitutes a critical failure mode that affects both operational safety and long term system performance.

Demagnetization is a complex degradation mechanism that involves an irreversible or quasi reversible decrease in the flux density of a permanent magnet. This failure occurs through a variety of methods including heat induced demagnetization when operating temperatures exceed the magnet's curie temperature [18] or approach critical thermal limits, demagnetization caused due to the exposure to adverse magnetic fields that exceed strength limits, and mechanical demagnetization caused due to the occurrence of structural stresses and vibrations. Industry researchs indicates that partial demagnetization, which can affect more than 10% of the magnetic field strength, leads to a 5-25% reduction in torque [19], a decrease in efficiency, and significant changes in the characteristics of electromagnetic torque. Whereas total demagnetization, although rare leads to a complete loss of torque capability and system failure. Predictive maintenance is considered an important direction in the PMSM maintenance field, where studies [20,21] have developed advanced methods and techniques for predicting failures before their occurrence through monitoring electrical and mechanical signals and operational indicators. These techniques aim to reduce costs and enhance operational efficiency through pre-planned maintenance.

Several designs of IPMSMs have been proposed for enhancing performance and improve the operational efficiency, where researchs [22-24] focused on improving the machines structures, permanent magnet geometry and its distribution inside the rotor to achieve higher power efficiency, speed and torque properties to reduce harmonic distortion and magnetic losses and to reduce vibration and torque ripples. Electromagnetic analysis contributes to the advanced analysis of the behavior and determination of areas of magnetic flux concentration precisely, the effect of magnetic saturation, and iron loss on electric motors [24-26]. Previous researches have investigated several important aspects of IPMSMs, focusing on analysis of various fault types and their direct impact on the performance and operating efficiency in [27-31]. The broken permanent magnets failure of and its impacts on the mechanical and electrical behavior of the PMSM were presented in [32], while researches [33,34] have focused on the analysis of acoustic noises and electromagnetic vibration generated during operation and their impact on overall PMSM performance. A demagnetization fault diagnosis method for PMSM based on machine learning (ML) was proposed by applying time frequency domain analysis and short time fourier transform (STFT) in detecting permanent magnet faults through stator phase current signal analysis [35].

In [36], the method of using acoustic signals for detection in the diagnosis of demagnetization and deflection faults of PMSM was investigated by feature conversion into 2D images, and fault detection and classification by a DeiT classifier was examined. The accuracy of the proposed method reaches 99.26%, demonstrating the reliability and effectiveness of the method. The thermal behavior of the interior permanent magnet synchronous motor (IPMSM) local demagnetization fault was discussed, adding temperature to the input signals to diagnose the demagnetization fault through back propagation neural network, predicting the healthy state of the permanent magnet, and determining the demagnetization ratio as an output signal [37]. LSTM and GRU neural network models were adopted for fault classification and early detection of demagnetization and short circuit in the PMSM based on the three current signals of different operating states of the motor, the two methods showed a high accuracy of about 98.7% in detecting faults [38]. A technique has been proposed that allows accurate localization of demagnetized permanent magnets in PMSM. This method adopts three toroidal yoke coils, which are used to measure the Back EMF. They are coiled in the stator core and through waveform analysis, the location of the fault can be determined [39].

A method for identifying and detecting the location of demagnetization faults in PMSM by analyzing the sub-current signal and calculating the pearson correlation coefficient between modeled and real currents was introduced, which provides an accurate and non-intrusive method for determining the location and level of fault [40]. A method that relies on the combination of neural convolutional networks (CNN) and gramian angular fields (GAF) transformation was proposed to classify and identify various demagnetization faults, based on the output torque signal with a model training accuracy of 98.33% and a test accuracy of 97.41% [41]. The impact of inclined magnetic fields on the partial demagnetization in NdFeB magnets is investigated through proposing a model that relies on an exponential function to predict this behavior [42]. In [43], a novel method has been proposed that is distinguished with its ability to detect and accurately locate the fault based on time signal analysis without requiring complex analysis, relying on the Back EMF constant using a new search coil to diagnose the partial demagnetization fault in IPMSMs, while in [44] a real time fault diagnosis method has been proposed based on Kurtosis analysis of the magnetic flux density variation in the air gap with the use of frozen permeability technique and equivalent magnetic circuit model (EMC).

A method for demagnetization fault diagnosis using multi-parameter estimation and Dual Extended Kalman Filter (Dual EKF) using a least squares algorithm is presented, this method allows real time magnetic flux monitoring , and accurate fault diagnosis with an error ratio less than 5% [45]. While current diagnostic approaches rely on existing post-failure detection methodologies, which limits their benefit in adaptive control and predictive maintenance strategies, experimental investigations have focused on failure injection and rapid aging tests, which provide valuable insights into demagnetization mechanisms but remain limited in scale due to time requirements and cost constraints. Recent analytical modeling approaches have used electromagnetic, mechanical, and thermal analyses to predict the effects of demagnetization, but these techniques typically involve simplifying assumptions, limiting their accuracy in complex engineering structures and different operational conditions. Although previous studies have focused on presenting electromagnetic analyses of IPMSMs and various curves of the performance evaluation for demagnetization faults, it is very important

to adopt advanced predictive analytics to evaluate the behavior under various operating conditions over a wide range of speeds. The aim of this study is to use predictive analysis for demagnetization faults by combining efficiency and loss maps as well as electromagnetic analysis of different magnetic fields for more thorough evaluation of the machines behavior.

The main contributions of this study include the enhancement of an analytical framework for demagnetization fault by relying on high resolution electromagnetic simulation results with a focus on the efficiency and loss map analysis, the implementation of detailed finite element modeling to simulate the demagnetization fault and compare it with a healthy IPMSM, and the quantification of efficiency map variations under the demagnetization.This analysis uses a pure approach using experimental data obtained from simulation results with the ANSYS Maxwell to solve and analyze electromagnetic fields. Efficiency maps are a effective tool for advanced performance characterization of the IPMSMs, providing a thorough analysis of efficiency and loss distributions across different operating speed and torque ranges. Finite Element Analysis (FEA) has the ability to simplify complex engineering models in a short duration and a superior ability to capture detailed electromagnetic phenomena. Which makes it a popular choice for application in many areas of electrical engineering.

II. APPROACHED DESIGN OF THE IPMSM

This study aims to establish a simulation model for advanced analysis of the partial demagnetization faults in IPMSMs through exploiting advanced predictive analysis techniques using efficiency and loss maps. The prototype of the proposed machine was created with the ANSYS RMxprt tool, under the following operating conditions: voltage of 220 V, speed of 1800 rpm, and power of 550 watts. For the core components, high grade materials were selected, with M19_24G for the stator, copper as a conductor, and NdFeB as a permanent magnet. After a thorough analysis of the 2D prototype, the evaluation of various curves, and advanced electromagnetic analysis, the acquired data and results were used to compute and create the efficiency maps across a wide range of speed and torque variation. To study the demagnetization faults with a particular focus on the residual magnetic flux density (Br), and comparing the results with healthy IPMSM to assess the impact of this fault on the overall system behavior. The ANSYS Maxwell approach is based on the application of the finite element technique as the basis for numerical analysis to simplify various geometrical models and translate them into solvable systems with high efficiency, with the ability to provide accurate and fast solutions to complex designs.

Table 1: Parameters of the presented IPMSM.

Parameters	Evaluated Value
Outer Diameter of Stator (mm):	120
Inner Diameter of Stator (mm):	75
Number of Slots:	24
Outer Diameter of Rotor (mm):	74
Inner Diameter of Rotor (mm):	26
Length of Stator Core (mm):	65
Type of Steel:	M19_24G
Conductor Type:	copper_75C
Number of Rotor Poles:	4
Rated Output Power (kW):	550
Rated Voltage (V):	220
Given Rated Speed (rpm):	1800
Operating Temperature (C):	75

Source: Authors, (2025).

Table 2: Parameters of the IPMSM Pole.

Name	Evaluated Value
D1	72mm
O1	3mm
O2	3mm
B1	4mm
Rib	3mm
HRib	3mm
Layers	1
Magnet Thickness	5mm
Magnet Width	40mm
Magnet Type	NdFeB (N38)

Source: Authors, (2025).

III. DEMAGNETIZATION FAULT IN THE IPMSM

Permanent magnets are relied on to generate the magnetic field required for the operation of the IPMSM. Demagnetization is a gradual or abrupt process that leads to the permanent magnets to lose their magnetic strength, which decreases the machine's efficiency. It occurs as a result of several complex electrical and physical factors. Among the causes of the demagnetization are thermal factors when permanent magnets are exposed to high temperatures, there occur alterations in the structure of the magnetic material which reduces the overall magnetic power and affects the internal magnetic fields leading to a reduction in the coercivity and increases the potential for irreversible demagnetization to occur. High electric currents, particularly at overloads or short circuits, lead to the

generation of strong magnetic fields that oppose the normal magnetic field of the permanent magnet and may exceed the strength of the magnetic resistance, causing a partial or complete demagnetization. Mechanical shocks and recurring high vibrations can affect the permanent magnet's structure, causing a gradual deterioration of its magnetic properties. Demagnetization can be achieved by adjusting the properties of the materials. Magnetic effects can be reduced or eliminated by assigning passive non-magnetic properties to the material used, such as setting the magnetic permeability value to transition to a non-magnetic state or by adjusting the permanent magnet field. The demagnetization fault can be monitored through indicators such as voltage, torque, currents, and magnetic field analysis of flux density.

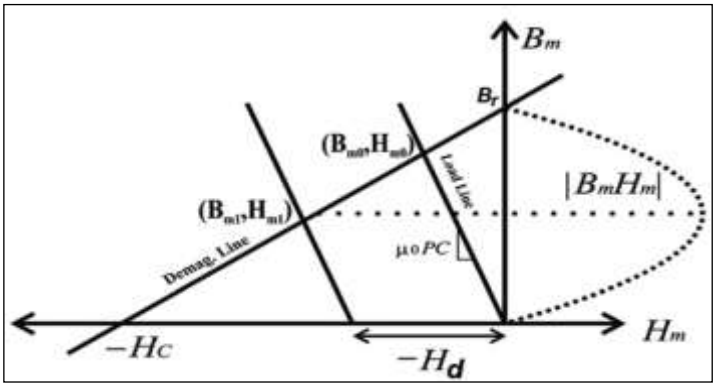


Figure 1: The characteristic of B-H demagnetisation of the permanent magnet.
Source: Taken from [46,47].

The no-load operating point is the intersection point of the B–H characteristic of the second quadrant (B-H demagnetization characteristics) at the load line and operating temperature, as illustrated in Figure 1. The load line slope relies on the permittivity coefficient PC [47] and is calculated by:

$$PC = -\frac{B_m}{\mu_0 H_m} = \frac{1}{f_{lkg}} \times \frac{A_g l_m}{A_m l_g} \tag{1}$$

where, the area of the air gap A_g equals the area of the magnet surface A_m, f_{lkg} is the different coefficient of the rotor leakage, l_g is the air-gap length. Therefore, the no-load operating (B_{m0}, H_{m0}) of the magnets are defined by:

From [46,47], the operating point of the no-load magnet (B_{m0}, H_{m0}), As illustrated in Figure 1, as follows:

$$B_{m0} = \frac{PC}{\mu_{rec} + PC} \times B_r$$
$$H_{m0} = \frac{B_r - B_{m0}}{\mu_{rec}\mu_0} \tag{2}$$

Where the magnet relative permeability is expressed as: $\mu_{\text{rec}} = \frac{B_r}{-\mu_0 H_c}$ [47].

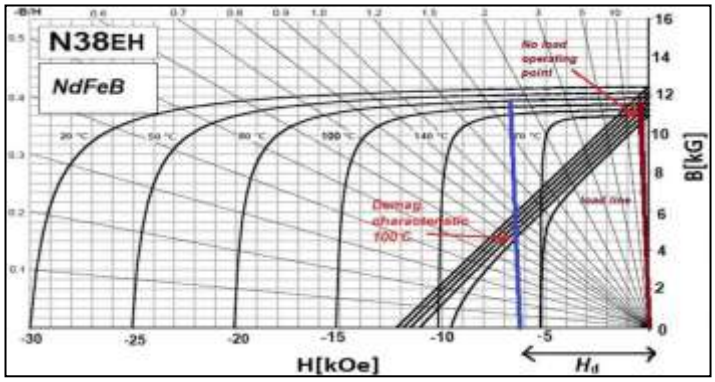


Figure 2: The NdFeB (N38EH) magnet demagnetisation B-H curves at various temperatures (1 kOe = 79.6 kA/m and 1 kG = 0.1 T).
Source: Taken from [48].

IV. FINITE ELEMENT ANALYSIS (FEA)

The finite element analysis (FEA) provides effectiveness in solving complex engineering systems. This method relies on numerical techniques to analyze complex engineering and physical systems that are difficult to solve with conventional analytical methods. The FEA divides geometric elements into a large number of small elements of a specific shape such as tetrahedrons and triangles connected to each other by a set of nodes, which facilitates and allows the application of mathematical equations on the elements to determine the overall system response in electrical machines. The machine parts are divided into very small elements,

which enables the magnetic fields to be separately modeled after being combined to obtain more accurate results and solve Maxwell's electromagnetic equations with high precision for each element as it contributes to solving the partial differential equations of magnetic fields for the flux density, eddy currents, mechanical ,and thermal in the electrical machine that is able to simplify the handling of heterogeneous materials and complex structures, as it contributes to accurately analysis of the flux, determine the parts exposed to magnetic saturation, anticipate the parts that are more prone to failures, and analysis of the dimensional and material variations effect based on simulation results leads to a reduction in the requirement to manufacturing many real models, which saves time and costs.

V. MATHEMATICAL MODELING OF THE IPMSM

A thorough analysis of the dynamic structure of the PMSM and the determination of the main factors that affect its functioning involves the use of scientific mathematical models that reveal the interdependence between different parameters such as torque, efficiency, loss, and power. These models provide the conceptual framework for the electromagnetic analysis and interpretation of experimental data extracted from the simulation. These equations are crucial for visualizing the mechanisms of power conversion and assessing the impact of external factors on the machine operational behavior. They allow for the analysis of the complicated interactions between magnetic fields and mechanical motion, enabling an accurate interpretation of the dynamic response of the machine in a variety of operating conditions. It also includes the evaluation of the design factors' impact on the operational properties.

Applying the Park transformation on the PMSM equation system, all current, voltage and flux vectors can be described in a rotor frame of reference [49-50]. Electrical equations of voltage in the d-q axis :

$$\begin{cases} V_d = R_s \cdot I_d + \frac{d\phi_d}{dt} - \omega_r \cdot \phi_q \\ V_q = R_s \cdot I_q + \frac{d\phi_q}{dt} + \omega_r \cdot \phi_d \end{cases} \tag{3}$$

Where $\omega_r = p.\Omega$ is the rotational electrical speed.

The magnetic flux in the d-q axis can be expressed by the formula below:

$$\phi_d = L_d \cdot I_d + \phi_f \tag{4}$$

$$\phi_q = L_q, I_q \tag{5}$$

Where ϕ_f is the field constant due to the permanent rotor excitation.

By replacing the Equations (4) and (5) in the Voltage Equation (3), this gives :

$$\begin{cases} V_d = R_s \cdot I_d + L_d \cdot \frac{di_d}{dt} - \omega_r \cdot L_q \cdot I_q \\ V_q = R_s \cdot I_q + L_q \cdot \frac{di_q}{dt} + \omega_r \cdot \left(L_d \cdot I_d + \phi_f\right) \end{cases} \tag{6}$$

Hence, the vector of the armature currents is expressed in the d-q axis as follows:

$$\begin{cases} \frac{dI_d}{dt} = -\frac{R_s}{L_d} \cdot I_d + \omega_r \cdot \frac{L_q}{L_d} \cdot I_q + \frac{1}{L_d} \cdot V_d \\ \frac{dI_q}{dt} = -\omega_r \cdot \frac{L_d}{L_q} \cdot I_d - \frac{R_s}{L_q} \cdot I_q + \frac{1}{L_q} \cdot V_q - \frac{1}{L_q} \cdot \phi_f \cdot \omega_r \end{cases} \tag{7}$$

Electromagnetic torque is defined by the partial derivative of the electromagnetic power with regard to the rotational angle of the rotor. This is given by :

$$T_e = \frac{dW_e}{d\theta_{geo}} = p \cdot \frac{dW_e}{d\theta} \tag{8}$$

Where W_e indicates the energy stored in the air-gap magnetic circuit, p is the number of pairs poles, θ_{geo} is the angular deviation of the moving part, expressed as follows: $\theta_{geo} = \frac{\theta_{\text{ele}}}{r}$.

The power expression is described as follows:

$$P(t) = \frac{3}{2}\left(V_d \cdot I_d + V_q \cdot I_q\right) \tag{9}$$

The d-q voltage Equation (3) are replaced in Equation (9) to obtain :

$$P(t) = \frac{3}{2}\left[R_s\left(I_d^2 + I_q^2\right)\right] + \frac{3}{2}\left[I_d \cdot \frac{d\phi_d}{dt} + I_q \cdot \frac{d\phi_q}{dt}\right] + \frac{3}{2}\left[\phi_d \cdot I_d - \phi_q \cdot I_q\right] \tag{10}$$

Where, the first expression (i.e., $\frac{3}{2}\left[R_s\left(I_d{}^2 + I_q{}^2\right)\right]$) represents the joule effect losses, the second term (i.e., $\frac{3}{2}\left[I_d \cdot \frac{d\phi_{\text{ft}}}{dt} + I_q \cdot \frac{d\phi_q}{dt}\right]$) represents the variation in stored magnetic energy, the last term (i.e., $\frac{3}{2}\left[\phi_d \cdot I_d - \phi_q \cdot I_q\right]$) represents the electromagnetic power transmitted from the stator to the rotor through the air gap.

The Equation of an electromagnetic torque is given by:

$$T_e = \frac{3}{2} p \cdot \left(L_d - L_q\right) \cdot I_d \cdot I_q + \frac{3}{2} \cdot p \cdot \phi_f \cdot I_q \tag{11}$$

Where ϕ_f is the permanent magnet flux linkage [Wb].
The machine rotor is considered to have sliding poles, i.e., $\left(L_d = L_q\right)$. Consequently, the expression (11) is simplified as follows:

$$T_e = \frac{3}{2} \cdot p \cdot I_q \cdot \phi_f \tag{12}$$

The conversion of electrical energy to mechanical energy in the PMSM is expressed by the following Equation:

$$J \cdot \frac{d\Omega}{dt} + f_r \cdot \Omega = T_e - T_r \tag{13}$$

By replacing the Equation (12) in Equation (13). this gives :

$$\frac{d\Omega}{dt} = \frac{3 \cdot p \cdot \phi_f}{2 \cdot J} \cdot I_q - \frac{f_r}{J} \cdot \Omega - \frac{1}{J} \cdot T_r \tag{14}$$

The overall efficiency η of the PMSM is defined as the ratio of the mechanical output power to the electrical input power:

$$\eta = \frac{P_{\text{out}}}{P_{\text{in}}} = \frac{T_e \cdot \omega_m}{P_{\text{in}}} \tag{15}$$

Where P_{out} is the output mechanical power [W], P_{in} is the input electrical power [W], T_e is the electromagnetic torque [Nm], ω_m is the angular speed [rad/s]

Electrical input power Equation:

$$P_{\text{in}} = \frac{3}{2} \cdot (v_d i_d + v_q i_q) \tag{16}$$

Where v_d, v_q is the voltage in the dq-axis, i_d, i_q is the current components in the dq-axis reference frame

Total losses Equation:

$$P_{\text{loss}} = P_{\text{Cu}} + P_{\text{Fe}} + P_{\text{mech}} + P_{\text{add}} \tag{17}$$

Where P_{add} is the additional losses (Stray, Harmonic, PWM)

The copper losses (stator winding losses):

$$P_{\text{Cu}} = 3 \cdot R_s \cdot (i_d^2 + i_q^2) \tag{18}$$

Where R_s is the resistance per phase of the stator winding [Ω]

The iron losses (core losses):

$$P_{\text{Fe}} = k_h \cdot f \cdot B^2 + k_e \cdot f^2 \cdot B^2 \tag{19}$$

Where k_h is the hysteresis loss coefficient, f is the frequency [Hz], k_e is the eddy current loss, B is the magnetic flux density in the iron core [T].

The mechanical losses (Friction & Windage):

$$P_{\text{mech}} = B_m \cdot \omega_m^2 + F \cdot \omega_m \tag{20}$$

Where B_m is the windage loss coefficient, F is the friction loss coefficient.

VI. RESULTS AND DISCUSSIONS

The curves shown in the Figures 3 to 6 reflect the dynamic variations of torque, stranded losses, and currents where the results are compared between the IPMSMs with demagnetization fault and the healthy IPMSM using Maxwell 2D software. The obtained results illustrated in Figures 7 and 8 indicate an analysis of the magnetic flux density distributions, magnetic flux intensity, and flux lines, which are intended to evaluate the electromagnetic behavior of the various IPMSM components. Based on the data extracted from this analysis, the efficiency and loss maps shown in Figures 9 to 10 were calculated and generated, which provide a thorough estimation of the motor's behavior under various operating conditions and the fault's impact on the overall system functioning.

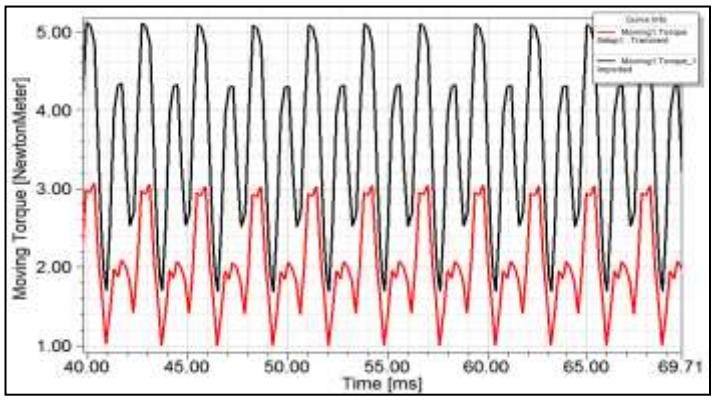


Figure 3: Torque curves variations of the IPMSMs.
Source: Authors, (2025).

In Figure 3, the black curve indicates the torque of the IPMSM in the healthy state, exhibiting a consistent and regular pattern in the range of 2 to 5 Nm and an average torque of approximately 3.7 - 3.8 Nm, with limited steady ripples indicating a balanced distribution of magnetic flux inside the rotor. In contrast, the red curve of the faulty IPMSM exhibits a sharp reduction in torque between 1 and 3 Nm, with an average torque of about 1.5 - 1.6 Nm, and the ripple in the red curve is noticeable and severe, indicating a perturbation in the magnetic flux causing an increase in mechanical vibrations and noise, which indicates the effect of demagnetization on the stability and operational efficiency of the IPMSM.

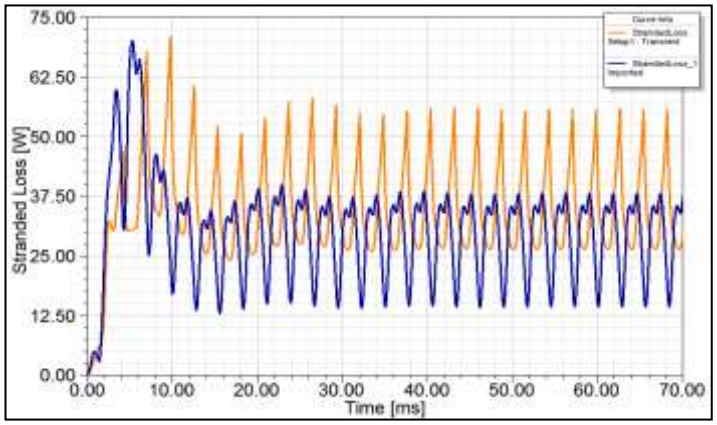


Figure 4: The IPMSMs stranded losses.
Source: Authors, (2025).

As illustrated in Figure 4, which shows the stranded losses curves of the IPMSM, the affected demagnetized machine (orange curve) exhibits a noticeable and continuous increase in losses compared to the healthy IPMSM (blue curve). While the losses in the healthy state settle within a steady range between 15 and 37 watts in the shape of periodic waves, reflecting the stability of the internal magnetic field and balanced currents, in the faulty state they exceed 50 watts with sharp oscillations, reflecting the disturbance of the magnetic field and increased compensating currents in the stator windings to compensate for the reduced flux. This increase in losses is directly attributed to the decrease in flux due to the permanent magnets degradation, which leads to an increase in the currents applied to the stator currents to compensate for the decrease in flux, thereby increasing the losses. These results confirm that continued monitoring of stranded losses curves is an effective means for early detection of demagnetization failure in the IPMSM, allowing for improved system reliability and failure avoidance.

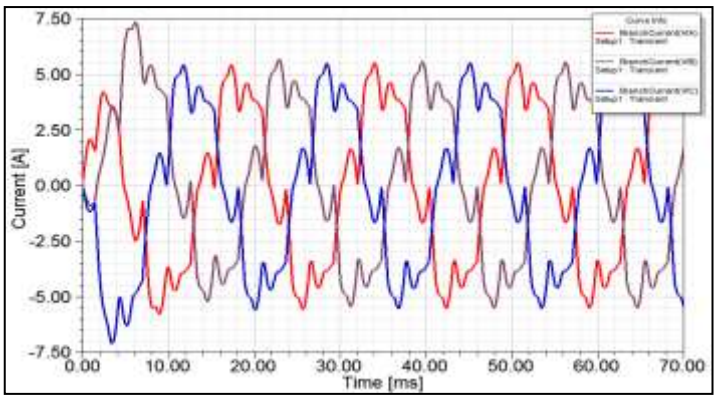


Figure 5: Healthy IPMSM current.
Source: Authors, (2025).

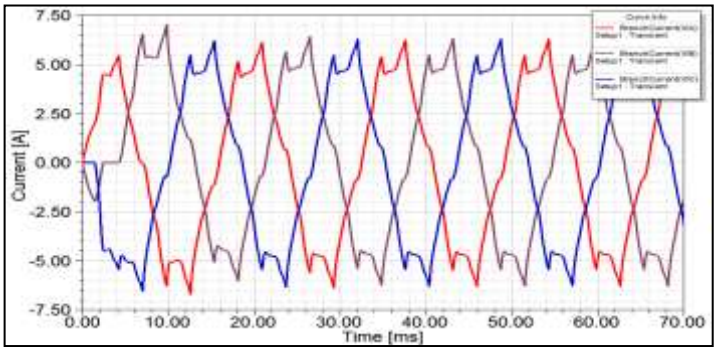
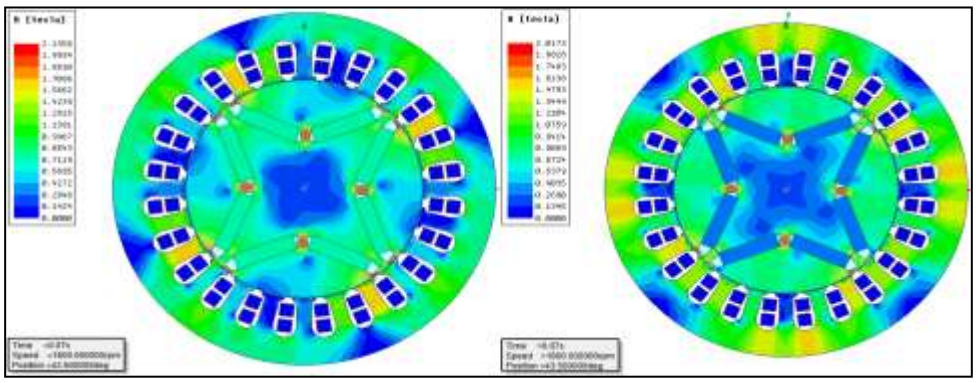


Figure 6: Current of IPMSM under demagnetization fault.
Source: Authors, (2025).

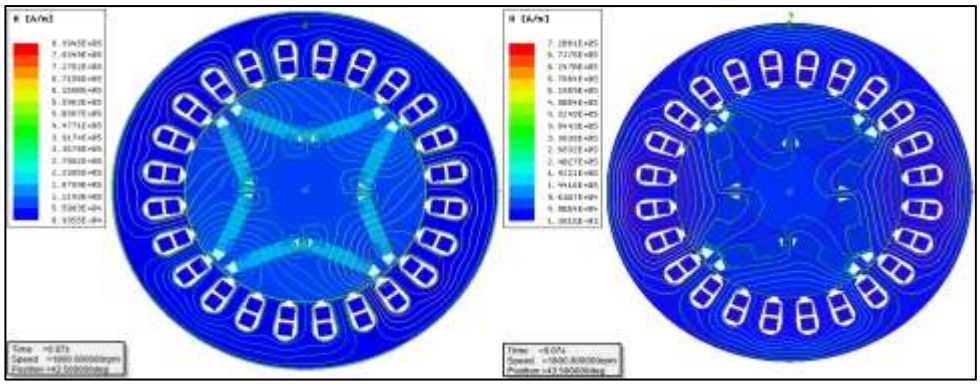
Figures 5 and 6 show a comparison of the branch current curves of the IPMSM in the cases of healthy operation and demagnetization failure, exhibiting a distinct difference in the shape and temporal behavior of the currents. In the healthy state, the currents are regularly alternating between the three branches, reflecting a balanced system and a steady magnetic flux inside the machine. On the other hand, the defective case reveals noticeable distortions in the waveform, with steeper slopes currents, an increase in the maximum current values, and relatively greater peaks and troughs compared to the healthy case, indicating that the system is injecting higher currents to compensate for the flux loss due to weakened magnets, as well as overlapping harmonic components and a slight misalignment of currents. Harmonic interference and a slight phase imbalance, which may affect the IPMSM operation and increase the torque ripple, as a results of the weak magnetic flux caused by the deterioration of the permanent magnets To compensate for this decrease in flux, the inverter is forced to drive higher currents, which increases the deformation in the current form. This negatively affects the efficiency and stability of the machine. Therefore, the branches currents analysis indicates a malfunction in the system and contributes to the early diagnosis of faults in the IPMSM.



(a) (b)

Figure 7: Distribution of the magnetic flux density: a) Healthy IPMSM, b) IPMSM under demagnetization fault.
Source: Authors, (2025).

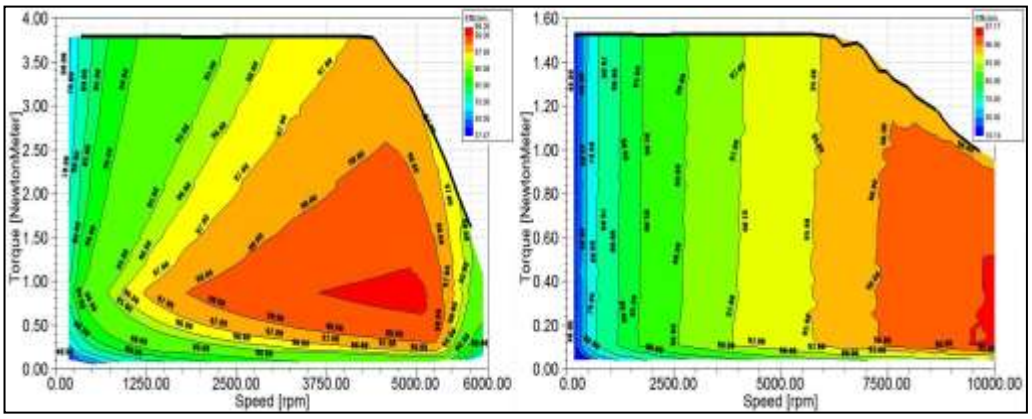
The magnetic flux density of the healthy IPMSM in Figure 7a exhibits a homogenous and uniform distribution, reflecting stable magnetic field behavior and high efficiency in power and torque conversion. In contrast, in the defective IPMSM in Figure 7b, the permanent magnet's magnetic flux density is significantly reduced to 0.4 Tesla while in the healthy IPMSM it reaches 0.8 Tesla. Irregular flux density distribution is observed in other machine parts with the appearance of regions of reduced flux between the air gap and the permanent magnets due to the loss of their effectiveness, while the magnetic saturation appears in the rotor between the magnets and the stator as well as more saturation in the stator, this disruption leads to a decrease in efficiency, power conversion, and increase in torque ripples.



(a) (b)

Figure 8: Distribution of the magnetic intensity and flux lines: a) Healthy IPMSM, b) IPMSM under demagnetization fault.
Source: Authors, (2025).

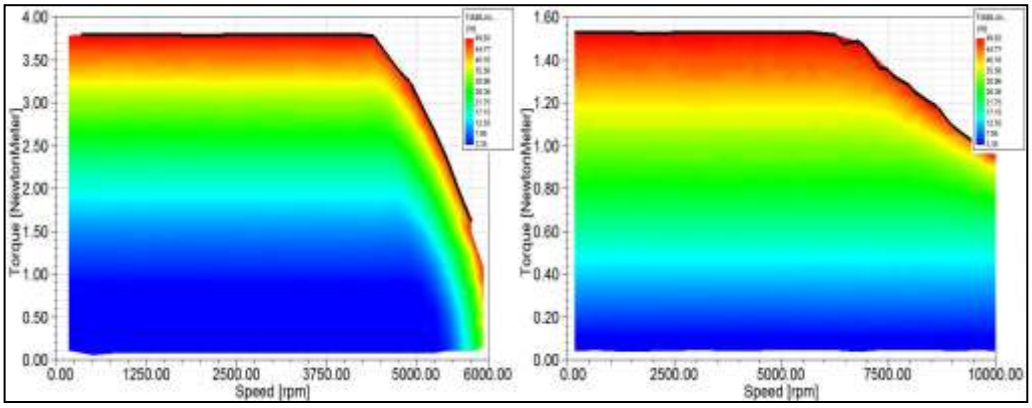
Figure 8 indicates that the magnetic intensity and the flux lines of a healthy IPMSM exhibit a homogeneous and uniform field distribution, with consistent and uniform magnetic flux lines, reflecting high efficiency and the strength of the magnetic connection between the rotor and the stator. In contrast, a demagnetization fault leads to a decrease in magnetic field strength in each permanent magnet as well as deformation and irregularity of the flux lines, causing magnetic imbalance inside the machine, reduced operating efficiency and torque instability, highlighting the importance of electromagnetic analysis in the early detection of this type of failure.



(a) (b)

Figure 9: Efficiency Maps: a) Healthy IPMSM, b) IPMSM under demagnetization fault.
Source: Authors, (2025).

The efficiency map of the IPMSM after the demagnetization failure shows an overall degradation in performance. In the case of a healthy IPMSM, there is a distinct concentration of high efficiency areas greater than 97% over an effective operational range between 1250 and 5500 rpm and torques up to 3.8 Nm, with a wide spread of high efficiency areas indicating a magnetic balance and consistent flux distribution inside the rotor. After the demagnetization, the highest recorded efficiency decreases to 97% and the highest efficiency above 96% is confined to a range of speeds between 7300 to 10000 rpm and a lower torque of 0.1 to 1.1 Nm. The efficiency decreases at lower speeds and over most areas of the map. The expansion of the speed range to 10000 rpm in case of failure is due to the torque drops drastically and attempting to compensate the loss through expansion of the operating range. The decrease in efficiency is attributed to the deterioration of magnetic flux due to the weakening effect of permanent magnets after the demagnetization, increased losses in the iron core and windings, torque deformation, and a reduction in power conversion.



(a) (b)

Figure 10: Loss maps: a) Healthy IPMSM, b) IPMSM under demagnetization fault.
Source: Authors, (2025).

The loss map of the IPMSM after the demagnetization, in Figure 10b, indicates a significant alteration in the performance characteristics compared to the healthy IPMSM. The faulty machine exhibits a sharp decrease in the torque producing ability, The maximum operating peak torque is reduced by 57 % from 3.8 Nm in the healthy machine to 1.55 Nm, in the healthy IPMSM in Figure 10a, the losses increase in a tight range between 3.2 and 3.8 torques. While in the case of partial demagnetization failure, the maximum operating range at 1.2 to 1.55 Nm , and the speed range extends to 10000 rpm, this comes at the expense of the ability to generate a higher torque, which reduces the overall efficiency.

The degradation in functioning is due to the reduction in flux resulting from the loss of magnetic material properties, which leads to lower magnetic flux density in the air gap and reduced electromotive force (EMF). As a result, the IPMSM requires higher currents to compensate for the lack of torque, which increases the copper losses. The absence of magnetization affects the positioning of the optimal operating point in the MTPA curve, resulting in inefficient operation outside the maximum efficiency area. On the other hand, Iron losses increases as the operational frequencies increase at higher speeds. Thus, it can be considered that partial demagnetization leads to a profound disruption of the torque and the efficiency in the IPMSM, which hinders the stability of operation in high torque applications and requires the adoption of strategies for early detection and compensation in intelligent control systems.

VII. CONCLUSIONS

In this paper, the effect of partial demagnetization faults on the IPMSM behavior is investigated by relying on the efficiency and loss maps compared to a healthy IPMSM. The results indicate that these faults deteriorate the efficiency of the machine under various operating conditions, particularly in the high torque and speed areas, with large ripples and a significant reduction in torque affecting the maximum safety threshold for operation. The results emphasize the importance of monitoring the condition of permanent magnets based on data extracted from the efficiency maps and the advanced electromagnetic analysis, for early detection of the demagnetization faults, which can lead to system degradation, increased losses, reduced efficiency, significant loss in power conversion effectiveness, the appearance of magnetic saturation in certain parts of the machine and inconsistent flux distribution, which emphasizes the importance of early detection of this failure type. The analysis of the magnetic field distribution reveals complex distortions in the magnetic flux density and intensity that affect the overall machine functioning and highlights the reliance on the advanced analysis based on efficiency maps for in-depth investigation of the impact of these failures. This contributes to the enhancement of predictive maintenance systems for electric motors. These systems increase reliability and reduce maintenance costs.

VIII. AUTHOR'S CONTRIBUTION

Conceptualization: Layachi Chebabhi, Toufik Tayeb Naas, Mohamed Zitouni, Monaem Elmnifi and Abdelkarim Cherhabil.
Methodology: Layachi Chebabhi, Toufik Tayeb Naas, Mohamed Zitouni, Monaem Elmnifi and Abdelkarim Cherhabil.
Investigation: Layachi Chebabhi, Toufik Tayeb Naas, Mohamed Zitouni, Monaem Elmnifi and Abdelkarim Cherhabil.
Discussion of results: Layachi Chebabhi, Toufik Tayeb Naas, Mohamed Zitouni, Monaem Elmnifi and Abdelkarim Cherhabil.
Writing – Original Draft: Layachi Chebabhi, Toufik Tayeb Naas, Mohamed Zitouni, Monaem Elmnifi and Abdelkarim Cherhabil.
Writing – Review and Editing: Layachi Chebabhi, Toufik Tayeb Naas, Mohamed Zitouni, Monaem Elmnifi and Abdelkarim Cherhabil.
Resources: Layachi Chebabhi, Toufik Tayeb Naas, Mohamed Zitouni, Monaem Elmnifi and Abdelkarim Cherhabil.
Supervision: Layachi Chebabhi, Toufik Tayeb Naas, Mohamed Zitouni, Monaem Elmnifi and Abdelkarim Cherhabil.
Approval of the final text: Layachi Chebabhi, Toufik Tayeb Naas, Mohamed Zitouni, Monaem Elmnifi and Abdelkarim Cherhabil.

IX. REFERENCES

[1] Djaloul Karboua, Toufik Mebkhouta, Youcef Chouiha, Abdelkader Azzeddine Bengharbi, B. ouadeh Douara, and Belgacem Toual, "Enhancing Performance of Permanent Magnet Synchronous Motor Drives Through Hybrid Feedback Linearization and Sliding Mode Control," ITEGAM- Journal of Engineering and Technology for Industrial Applications (ITEGAM-JETIA), vol. 10, no. 49, pp. 178–186, Jan. 2024, doi: https://doi.org/10.5935/jetia.v10i49.1222.

[2] M. O. Mustafa and G. S. Mohammed, "Design of Particle Swarm Optimization-Based PID Controller for High-Performance PMSM Speed Control," ITEGAM-Journal of Engineering and Technology for Industrial Applications (ITEGAM-JETIA), vol. 11, no. 53, pp. 216–224, Jan. 2025, doi: https://doi.org/10.5935/jetia.v11i53.2024.

[3] Touahria D, Bouchemha A, Kahla S. Design and simulation of hybrid system PV/PEMFC/Battery/Super Capacitor powered PMSM motor for electric vehicle. Sebha University Conference Proceedings 2025; 4 (1): 176–181. doi: https://doi.org/10.51984/sucp.v4i1.3976.

[4] Nouraldin NA, Chebaani M, Számel L, Abdelwahab SAM, Abdellatif WS. Experimental investigation of predictive control for PMSM-based wind turbine generation system. Computers and Electrical Engineering 2024; 119: 109554. doi: https://doi.org/10.1016/j.compeleceng.2024.109554.

[5] Számel L, Nouraldin NA, Chebaani M, Abdelwahab SAM, Abdellatif WS. Experimental investigation of predictive control for PMSM-based wind turbine generation system. SSRN 2024. doi: https://doi.org/10.2139/ssrn.4871457.

[6] Sonandkar S, Selvaraj R, Chelliah TR. PV powered improved quasi-Z-source inverter fed five-phase PMSM for marine propulsion systems. IEEE Transactions on Transportation Electrification 2024; 11 (1): 393–403. doi: https://doi.org/10.1109/tte.2024.3391615.

[7] Karthik K, Ponnambalam P. Design and implementation of time-based fault tolerance technique for solar PV system reliability improvement in different applications. Scientific Reports 2025; 15 (1): 7377. doi: https://doi.org/10.1038/s41598-025-91464-4.

[8] Zhao W, Cheng M, Ji J, Cao R, Du Y, Li F. Design and analysis of a new fault-tolerant linear permanent-magnet motor for maglev transportation applications. IEEE Transactions on Applied Superconductivity 2012; 22 (3): 5200204. doi: https://doi.org/10.1109/tasc.2012.2185209.

[9] Karboua D, Belgacem T, Khan ZH, Kellal C. Robust performance comparison of PMSM for flight control applications in more electric aircraft. PLoS ONE 2023; 18 (7): e0283541. doi: https://doi.org/10.1371/journal.pone.0283541.

[10] Zhang C, Zhang X, Zhao F, Gerada D, Li L. Improvements on permanent magnet synchronous motor by integrating heat pipes into windings for solar unmanned aerial vehicle. Green Energy and Intelligent Transportation 2022; 1 (1): 100011. doi: https://doi.org/10.1016/j.geits.2022.100011.

[11] Lamichhane A, Zhou L, Yao G, Luqman M. Modeling, control and power management of six-phase PMSM based shipboard MVDC distribution system. Energies 2020; 13 (16): 4229. doi: https://doi.org/10.3390/en13164229.

[12] Kim SA. A study on the predictive maintenance algorithms considering load characteristics of PMSMs to drive EGR blowers for smart ships. Energies 2021; 14 (18): 5744. doi: https://doi.org/10.3390/en14185744.

[13] Hong DK, Joo DS, Lee JY, Woo BC. Effects of the pole-slot combination on the PMSM of an integrated motor propulsor for an unmanned underwater vehicle considering its electric performance, noise and vibration. Int J Appl Electromagn Mech 2016; 52 (3–4): 1689–1695. doi: https://doi.org/10.3233/jae-162162.

[14] Chen W, Mao Z, Tian W. Water cooling structure design and temperature field analysis of permanent magnet synchronous motor for underwater unmanned vehicle. Applied Thermal Engineering 2024; 240: 122243. doi: https://doi.org/10.1016/j.applthermaleng.2023.122243.

[15] Yang Y, He Q, Fu C, Liao S, Tan P. Efficiency improvement of permanent magnet synchronous motor for electric vehicles. Energy 2020; 213: 118859. doi: https://doi.org/10.1016/j.energy.2020.118859.

[16] Yuan T, Wang D, Wang X, Wang X, Sun Z. High-precision servo control of industrial robot driven by PMSM-DTC utilizing composite active vectors. IEEE Access 2019; 7: 7577–7587. doi: https://doi.org/10.1109/access.2018.2890539.

[17] S. C. Vanamala and K. K. Danduprolu, “Speed Control of PMSM drive using Model Predictive Control Based Field Oriented Control,” ITEGAM- Journal of Engineering and Technology for Industrial Applications (ITEGAM-JETIA, vol. 11, no. 52, 2025, doi: https://doi.org/10.5935/jetia.v11i52.1243.

[18] Bilgin O, Kazan FA. The effect of magnet temperature on speed, current and torque in PMSMs. In: 2016 XXII International Conference on Electrical Machines (ICEM); 2016 Sep; pp. 2080–2085. IEEE. doi: https://doi.org/10.1109/icelmach.2016.7732809.

[19] Shi C, Peng L, Zhang Z, Shi T. Analytical modeling and analysis of permanent-magnet motor with demagnetization fault. Sensors 2022; 22 (23): 9440. doi: https://doi.org/10.3390/s22239440.

[20] Vlachou VI, Karakatsanis TS. Development of a fault-tolerant permanent magnet synchronous motor using a machine-learning algorithm for a predictive maintenance elevator. Machines 2025; 13 (5): 427. doi: https://doi.org/10.3390/machines13050427.

[21] Amira Amira Slimani, Amor Bourek, A. Ammar, Khoudir Kakouche, Wassila Hattab, and M. Bacha, “Artificial Neural Network-Based Deadbeat Predictive Current Control with Dead-Time Compensation for PMSMs,” ITEGAM- Journal of Engineering and Technology for Industrial Applications (ITEGAM-JETIA), vol. 11, no. 51, Jan. 2025, doi: https://doi.org/10.5935/jetia.v11i51.1456.

[22] S.-C. Wang, Y.-C. Nien, and S.-M. Huang, “Multi-Objective Optimization Design and Analysis of V-Shape Permanent Magnet Synchronous Motor,” Energies, vol. 15, no. 10, p. 3496, May 2022, doi: https://doi.org/10.3390/en15103496.

[23] M. Ayaz, Yiğit Karabulut, Serkan Aktaş, and Erkan Meşe, “Electromagnetic Design and Comparative Analysis of Compact PM Motors for Ventricular Assist Devices,” Physica Scripta, vol. 100, no. 8, pp. 085019–085019, Jul. 2025, doi: https://doi.org/10.1088/1402-4896/adf401.

[24] Y.-J. Lee, Y.-C. Chang, and Y.-L. Lim, “Design and Performance Evaluation of High-Efficiency SiC Interior Permanent Magnet Synchronous Motor Drive Systems,” 2022 IEEE Industry Applications Society Annual Meeting (IAS), pp. 1–6, Jun. 2025, doi: https://doi.org/10.1109/ias62731.2025.11061490.

[25] Chebabhi L, Naas TT, Zitouni M. Comparative study of 2D designs of 12/8 and 10/8 switched reluctance motors using ANSYS Maxwell. ITEGAM-JETIA 2025; 11 (52): 186–195. doi: https://doi.org/10.5935/jetia.v11i52.1604.

[26] Chebabhi L, Naas TT, Zitouni M, Ghibeche I, Benmessaoud T. Influence of choosing materials on 6/4 switched reluctance motor performance. Studies in Engineering and Exact Sciences 2024; 5 (1): 2391–2406. doi: https://doi.org/10.54021/seesv5n1-118.

[27] J. de las Morenas, L. M. Belmonte, and R. Morales, “Streamlined Bearing Fault Detection Using Artificial Intelligence in Permanent Magnet Synchronous Motors,” Machines, vol. 13, no. 5, p. 357, Apr. 2025, doi: https://doi.org/10.3390/machines13050357.

[28] A. H. Baharvand, S. Hossein Beigi Fard, A. H. Poursaeed, and M. Doostizadeh, “An optimized classifier chains-based deep learning framework for Inter-Turn Fault diagnosis in Permanent Magnet Synchronous Motors,” Applied Soft Computing, vol. 180, p. 113482, Jun. 2025, doi: https://doi.org/10.1016/j.asoc.2025.113482.

[29] A. Sergakis, M. Salinas, N. Gkiolekas, and K. N. Gyftakis, “A Review of Condition Monitoring of Permanent Magnet Synchronous Machines: Techniques, Challenges and Future Directions,” Energies, vol. 18, no. 5, p. 1177, Feb. 2025, doi: https://doi.org/10.3390/en18051177.

[30] Wu S, Ma G, Yao C, Sun Z, Xu S. Current sensor fault detection and identification for PMSM drives using multichannel global maximum pooling CNN. IEEE Transactions on Power Electronics 2024; 39 (8): 10311–10325. doi: https://doi.org/10.1109/tpel.2024.3395290.

[31] Wu Y, Zhang J, Xu Z, Wang S, Fu H. Feature extraction and applicability comparisons for fault detection of inter-turn short-circuited PMSM. IEEE Transactions on Instrumentation and Measurement 2024; 73: 1–10. doi: https://doi.org/10.1109/tim.2024.3509588.

[32] Goktas T, Zafarani M, Akin B. Discernment of broken magnet and static eccentricity faults in permanent magnet synchronous motors. IEEE Transactions on Energy Conversion 2016; 31 (2): 578–587. doi: https://doi.org/10.1109/tec.2015.2512602.

[33] Pindoriya RM, Gautam G, Rajpurohit BS. A novel application of pseudorandom-based technique for acoustic noise and vibration reduction of PMSM drive. IEEE Transactions on Industry Applications 2020; 56 (5): 5511–5522. doi: https://doi.org/10.1109/tia.2020.2997904.

[34] Wang YS, Guo H, Yuan T, Ma LF, Wang C. Electromagnetic noise analysis and optimization for permanent magnet synchronous motor used on electric vehicles. Engineering Computations 2021; 38 (2): 699–719. doi: https://doi.org/10.1108/ec-02-2020-0070.

[35] Pietrzak P, Wolkiewicz M. Demagnetization fault diagnosis of permanent magnet synchronous motors based on stator current signal processing and machine learning algorithms. Sensors 2023; 23 (4): 1757. doi: https://doi.org/10.3390/s23041757.

[36] Huang W, Du B, Li T, Cheng Y, Cui S. Detection of PMSM demagnetization fault and eccentricity fault based on acoustic images and DeiT classifier. IEEE Transactions on Energy Conversion 2025. doi: https://doi.org/10.1109/tec.2025.3550544.

[37] Meiwei Z, Weili L, Haoyue T. Demagnetization fault diagnosis of the permanent magnet motor for electric vehicles based on temperature characteristic quantity. IEEE Transactions on Transportation Electrification 2022; 9 (1): 759–770. doi: https://doi.org/10.1109/tte.2022.3200927.

[38] Lale T, Yüksek G. Identification and classification of turn short-circuit and demagnetization failures in PMSM using LSTM and GRU methods. Bull Pol Acad Sci Tech Sci 2025; 73 (1). doi: https://doi.org/10.24425/bpasts.2024.151958.

[39] Chen H, Gao C, Si J, Nie Y, Hu Y. A novel method for diagnosing demagnetization fault in PMSM using toroidal-yoke-type search coil. IEEE Transactions on Instrumentation and Measurement 2021; 71: 1–12. doi: https://doi.org/10.1109/tim.2021.3134991.

[40] Yu Y, Gao H, Chen Q, Liu P, Niu S. Demagnetization fault detection and location in PMSM based on correlation coefficient of branch current signals. Energies 2022; 15 (8): 2952. doi: https://doi.org/10.3390/en15082952.

[41] Zhang Q, Cui J, Xiao W, Mei L, Yu X. Demagnetization fault diagnosis of a PMSM for electric drilling tools using GAF and CNN. Electronics 2024; 13 (1): 189. doi: https://doi.org/10.3390/electronics13010189.

[42] Ruoho S, Arkkio A. Partial demagnetization of permanent magnets in electrical machines caused by an inclined field. IEEE Transactions on Magnetics 2008; 44 (7): 1773–1778. doi: https://doi.org/10.1109/tmag.2008.921951.

[43] Li Z, Zhao S, Yang X, Cao J, Li J. Partial demagnetization fault diagnosis of interior permanent magnet synchronous motor based on back electromotive force constant of a new search coil. IEEE Transactions on Instrumentation and Measurement 2025. doi: https://doi.org/10.1109/tim.2025.3585215.

[44] Ding S, Wang X, Li Y, Wu Z, Hang J. Online diagnosis of partial demagnetization fault in PMSMs based on kurtosis of air-gap flux density signal difference. IEEE Transactions on Power Electronics 2025. doi: https://doi.org/10.1109/tpel.2025.3580750.

[45] Xing Z, Wang X, Zhao W, Chen A, Yu S. Analysis and improvement of reliability of surface-mounted permanent magnet synchronous motors based on subdivision subdomain models. IEEE Transactions on Energy Conversion 2025. doi: https://doi.org/10.1109/tec.2025.3573783.

[46] Hanselman DC. Brushless permanent magnet motor design. The Writers' Collective; 2003.

[47] Hendershot JR, Miller TJE. Design of brushless permanent-magnet motors. Oxford University Press; 1995. Available: https://academic.oup.com/book/52967

[48] Dexter Magnetic Technologies. Available at: https://www.dextermag.com/products/permanent-magnets/neodymium-iron-boron-magnets/

[49] MAGNET SOP. Mathematical modeling and simulation of permanent magnet synchronous motor. Int J Adv Res Electr Electron Instrum Eng 2013; 2 (8).

[50] Roshen W. Iron loss model for permanent-magnet synchronous motors. IEEE Transactions on Magnetics 2007; 43 (8): 3428–3434. doi: https://doi.org/10.1109/tmag.2007.899687.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 250-259. September/October., 2025.

DOI: https://doi.org/10.5935/jetia. v11i55.2686

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

STRUCTURAL ANALYSIS OF UPPER ARM EXOSKELETON FOR BIOMEDICAL APPLICATIONS

Vijayakumar K[1], Bubesh Kumar D[2] and Sangeetha Krishnamoorthi[3]

[1, 2, 3] Department of Mechanical Engineering, Aarupadai Veedu Institute of Technology, Vinayaka Mission's Research Foundation (DU), Tamil Nadu, India.

[1]https://orcid.org/0000-0002-5964-982X, [2]https://orcid.org/0000-0002-4308-2793, [3]https://orcid.org/0000-0003-1427-8743

Email: k.vijaykumar86@gmail.com, bubeshkumar@avit.ac.in, sangeethas@avit.ac.in

ARTICLE INFO

Article History
Received: September 8, 2025
Revised: September 15, 2025
Accepted: October 1, 2025
Published: October 31, 2025

Keywords:
Rehabilitation,
Exoskeleton,
ANSYS,
Biomedical,
Paralytic.

ABSTRACT

Upper limb rehabilitation exoskeletons worn on the body have many features, such as compatibility with human structure, safety and training, etc. This paper aims to examine the structural strength of the exoskeleton. Using ANSYS Workbench 15, the static analysis was carried out on the human body's upper limb. The study of the various parameters showed that the design of the titanium exoskeleton satisfied the requirements of an exoskeleton for paralytic patients for a load up to 300N; for the stress value and deformation values for various parameters, a safe design of the exoskeleton has been achieved. The comparison of the exoskeleton structure with that of stainless steel, aluminium, and titanium shows that the titanium structure is most suitable for the exoskeleton since it is lightweight and Young's modulus is similar to that of human bone. The aluminium exoskeleton structure will shear when load is applied and the stainless exoskeleton structure is heavier than the human bone. It is concluded from the Finite element analysis carried out on workbench 15 for the three materials, namely, aluminium, titanium and steel. Titanium is suitable for exoskeleton structure construction since it satisfies design criteria as per biomechanical design considerations.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Millions of patients, around 15 million people all over the world, suffer from paralysis stroke. Out of these, 85% of the patients survive, and only 10% recover completely; the rest of the patients lose their motor functions and old age people have a nervous weakness; the exoskeleton can rehabilitate these patients to lead normal lives. In these exoskeleton robots, the human muscle power is combined with exoskeleton structure energy and gives a synergy effect [1]. It combines the fields of biomechanics, sensors, controllers, and software algorithms to ensure the smooth functioning of the exoskeleton device and improve the standard of living of paralytic patients and old-age people. The other applications of exoskeletons are in the military to help soldiers carry heavy loads like arms, ammunition, and personal belongings of the soldiers and in the labour-intensive construction industry. The upper limb exoskeleton patients carry out routine exercises that improve motor functions and help them to lead a normal life. The difficulty with exoskeletons is the interface between the human body and the exoskeleton structure; in this research, the ergonomics aspect of the exoskeleton is considered; the motion of the upper exoskeleton shoulder rotation should coincide with the human shoulder joint rotation for smooth functioning of the exoskeleton robot. The number of degrees of freedom of the exoskeleton is limited to the human arm, which has seven degrees of freedom. The controller of the exoskeleton robot is programmed with artificial intelligence to align the trajectory of the exoskeleton with human arm movements. Intensive physiotherapy is needed to improve the condition of the patients, and the number of physiotherapists is smaller to bridge the gap between the numbers of therapists. Exoskeletons are used to improve the patient's health [2]. Exoskeletons are used in repetitive and precise control of the motion, which are desirable characteristics for rehabilitation purposes.

The exoskeleton technology was originally designed for military applications as a wearable device on the human body, and it was designed to help soldiers carry heavy loads such as ammunition and luggage, and food soldiers it weighed around 150kg, it increased the capacity of soldier to carry more weight at the same time with little muscle power utilisation more loading carrying capacity of the soldier was increased. Firefighters and other people in the manufacturing and rescue operations also used it. Research in this area has been studied since the 1960s onwards [3]. Today medical rehabilitation of the exoskeleton robot is used to assist paralytic patients to undergo physiotherapy, rehabilitation, and also to do daily activities. With the scarcity of trained physiotherapists and the duration of the physiotherapy long and frequent therapy needed for quick recovery of the patient, the exoskeleton has become a cost-effective device for the treatment of paralytic patients, spinal cord injured patients, stroke patients, and rehabilitation of the injured persons [4].

However, exoskeletons are not widely used due to their heavy weight and the limitations they have to the human degrees of movement of the limbs. In this research, three types of materials were analysed, namely aluminium, steel, and stainless steel, and they are compared in terms of the structure of the exoskeleton suit [5]. Research has been carried out on the structural analysis of the exoskeleton; it was found titanium is light in weight, aluminium fails due to fatigue corrosion, and stainless steel is strong, but the weight of the exoskeleton is heavy when worn on the human body. Since the human bone young's modulus is 20GPa and the steel young's modulus is 210GPa [6], the increased weight of the exoskeleton will harm the structural bone of the humans. Less Young's modulus, such as titanium, can be used for the development of wearable exoskeletons for paralytic patients; titanium is also noncorrosive. It is an ideal material for exoskeletons manufactured for patients [7].

Types of Exoskeleton systems are upper limb, lower limb, and full body. Exoskeletons used for patients are classified further as rehabilitation, haptics, assistive devices, teleoperations, and power augmentation, depending on the purpose for which they are used. Further, exoskeletons are classified as active exoskeletons and passive exoskeletons; the design of exoskeletons for patients is active exoskeletons. They have to work on the limbs of the patients for rehabilitation and exercise [8]. The battery gives power to the BLDC motors. Passive exoskeletons are used in the military and construction industry as normal people wear them to increase their load-carrying capacity. The exoskeleton is designed to assist the upper limb; understanding the anatomy of the shoulder complex helps in designing the exoskeleton more efficiently [9].

The current research emphasises the increasing importance of upper limb exoskeletons in rehabilitation; nonetheless, a notable deficiency exists in the structural optimisation of materials for biomechanical compatibility and patient comfort. Many studies emphasise functional performance while neglecting the ergonomic interaction between the human body and exoskeleton design, which is essential for prolonged use. Furthermore, there is scant research that analyses material properties—such as Young's modulus and density—in connection to the qualities of human bone. The incorporation of AI for motion alignment has been investigated; nevertheless, its coordination with material flexibility remains insufficiently addressed. This study examines appropriate materials and ergonomic design criteria to remedy these deficiencies. This work introduces a novel comparative static structural analysis of upper limb exoskeletons utilising Finite Element Analysis (FEA) in ANSYS Workbench 15, with an emphasis on biomedical rehabilitation applications.

An essential novelty is the thorough assessment of three often utilised materials—aluminium, stainless steel, and titanium—under realistic stress circumstances that replicate human arm motions, especially for those with paralysis. The work distinctly highlights biomechanical compatibility by contrasting material characteristics, including Young’s modulus, with those of human bone, a frequently neglected requirement in exoskeleton design. This study incorporates both mechanical strength and physiological compatibility into the material selection process, contrasting with previous studies that primarily focused on load-bearing capability. The designation of titanium as the optimal material, owing to its superior strength-to-weight ratio and bone-like flexibility, enhances the value of future exoskeleton design. This comprehensive material-performance relationship offers a novel viewpoint in assistive biomechanics and exoskeletal device engineering.

II. METHODOLOGY

II.1 SHOULDER COMPLEX ANATOMY

The upper limb exoskeleton includes the hand, shoulder area, elbow, forearm, and wrist. The way exoskeleton moves depends on coordinated action of 3 bones on 4 joints [10]. The shoulder joint also known as glenohumeral joint. The sternoclavicular joint connects the shoulder region to the main part of the body. The acromioclavicular joint is formed at the point where the end of the clavicle connects with the acromion, which is a part of the shoulder blade. The sternoclavicular joint is a complex joint that involves parts of the clavicle, sternum, and cartilage from the first rib [11]. Because of this, the shoulder has a total of five degrees of freedom. Pneumatic powered Exoskeleton arm is designed and used for weight lifting of 20kgs. Basic hand biomechanics through pneumatic actuator is used to work with load [12].

If the exoskeleton structure limits the translational motion, then the range of motion of the shoulder joint is limited, so in this paper ergonomical design of the exoskeleton is considered during the construction of the robot upper arm shoulder system to avoid pain in the joint[13]. In this robotic arm total nine degrees of freedom. The glenohumeral joint, which is part of the shoulder complex, is a ball-and-socket joint that allows the humerus to rotate around the glenohumeral head in three different directions [14]. The sternoclavicular joint has two degrees of freedom, often referred to as shoulder elevation and depression, or retraction and protraction. These movements cause the glenohumeral joint to move in a straight line. So, altogether, the shoulder has five degrees of freedom [15]. The elbow complex can move in two directions, and the wrist joint also has 2 DOF. The axes around which the wrist moves pass through the capitate carpal bone [16].

When raising the arm above the shoulder, the motion of the shoulder girdle is involved, which includes the clavicle and scapula, plays a more important role than the movement of the arm below the shoulder. The shifting of the glenohumeral joint is crucial for lifting the arm high [17], [18]. If this shifting motion is restricted by outside forces, it can limit the range of motion of the shoulder joint and lead to discomfort or pain in the joint [19]. Because there are differences between individuals, the table below shows the usual

range of motion for the shoulder, elbow, and wrist joints [20], [21]. Human shoulder joint anatomy is shown in Figure 1. Table 1 depicts the human normal range of motion for the upper limb.

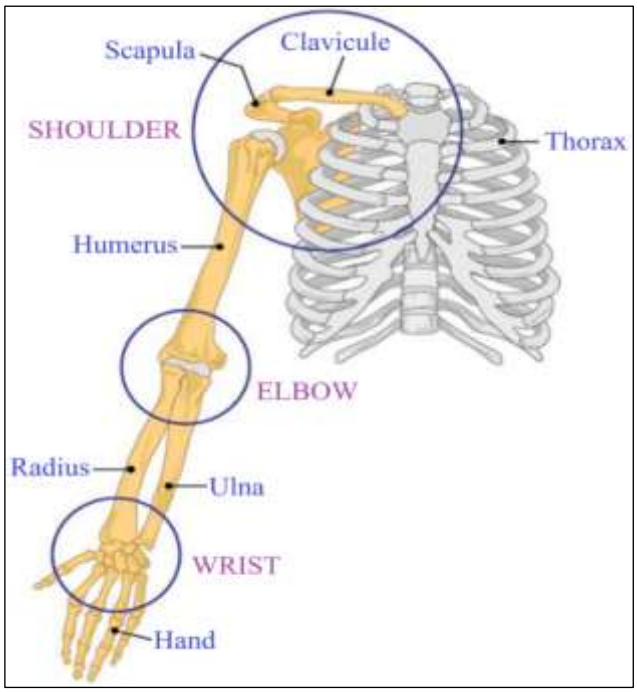


Figure 1: Shoulder joint anatomy.
Source: Authors, (2025).

Table 1: Normal range of motion for the shoulder, elbow, and wrist joints in degrees.

Upper arm description	Position	Angle
Shoulder abduction	From horizontal position	0-90
Shoulder adduction		0-90
Shoulder flexion	From vertical position	0-180
Shoulder extension	-	0-50
Internal rotation (medial)	With the arm at your side	0-90
External rotation (lateral)	-	0-90
Elbow flexion	With upper arm horizontal	0-160
Elbow extension	Until the arm is vertical at your side	145-0
Elbow pronation	-	0-90
Elbow supination	-	0-90
Wrist flexion	-	0-90
Wrist extension	-	0-70
Wrist radial deviation	-	0-25
Wrist ulnar deviation	-	0-65

Source: [22].

II.2 EXOSKELETON DESIGN

The exoskeleton mechanical design for 5 DOF was considered for the shoulder complex. The major difficulty with the exoskeleton is it has to coincide with the movement of the human joints for rehabilitation and exercise for paralytic patients [23]. The exoskeleton was designed for active joints to enhance the patient's muscle power [24]. Table 2 shows the range of motion for the exoskeleton (UMI) compared to humans. The 4 DOFs for the humans.

Table 2: Exoskeleton range of angular movement of arm

DOF - 4	Human	Caden	Exos	MGA	UMI
Ext/Flx	235	180	134	211	224
Abd/Add	172.5	181	141	137	141
Int/Extr	160	161	134	130	161
Elv/Dep	14	-	-	-	11

Source: Authors, (2025).

II.3 EXOSKELETON ROBOT

Modelling a Robotic exoskeleton is completely different from human anatomy; the interaction of the exoskeleton, which is motorised, and the human joints work on the muscle power, is more complex than the robotic joints [25]. In this research, the joints have been designed to work in tandem with the human joints without harming the muscles and range of motion of the human arm [26]. A comprehensive musculoskeletal model was studied analytically by the author an optimal design was suggested for the exoskeleton model. The figure shows the multibody model of the hand; it has two links and a joint at the centre, at the top, that is near the shoulder joint [27]. It is shown as a rigid support; the gravitational forces act at the centre of gravity of each link, and the angle made by each link is also shown in the model in Figure 2. Thus, the muscle activity of the exoskeleton is calculated, along with that of the exoskeleton robot. The exoskeleton biomechanical model was carried out in this research to arrive at the design, and it was modelled and analysed [28].

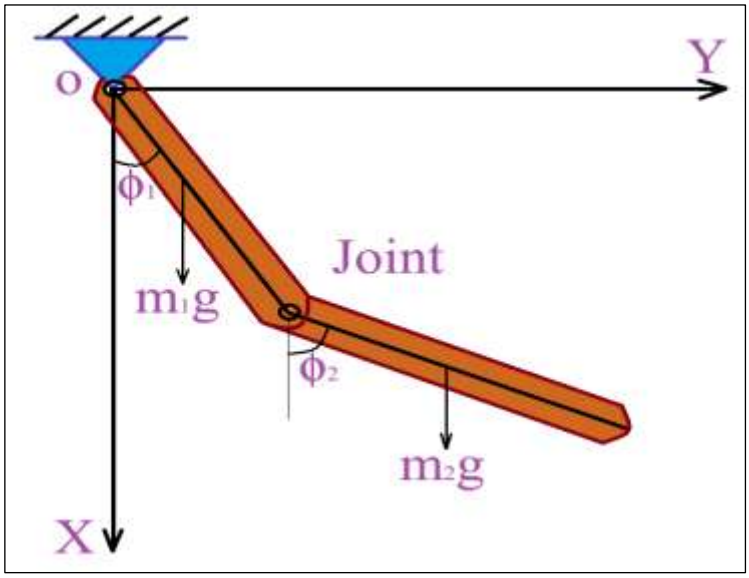


Figure 2: Biomechanical model human hand,
Source: Authors, (2025).

II.4 UPPER-LIMB EXOSKELETON DESIGN CHALLENGES

Exoskeleton design should be kinematically compatible with the patient's motion [29]. In this research, the design of the exoskeleton is such that the four muscles of the shoulder are not constrained. The exoskeleton's range of motion is different from the range of motion of the human anatomy. To achieve this human anatomy, more complex structures are needed; when the structure becomes complex, it becomes heavier, which will be difficult for paralytic patients [30]. In this research new design with lightweight material for the structure was developed in this research to improve the COR of the upper exoskeleton shoulder joint. The discomfort of wearing an exoskeleton suit for paralytic patients is taken care of in the design of the special belt arrangement. Care is taken in the design to align the physiological joints to align with the exoskeleton.

Since the exoskeleton is aligned with the shoulder and wrist, the physiological joints are unique due to the muscle arrangement on the bone and due to the contraction abduction of the muscles, which helps in the movement of the bones due to its complex constructions. As an outcome, the rotational joint axes tend to change their place amid movement [31]. Connections of exoskeletons on human limbs are not the author designs inflexible, special grips to avoid slippage. Thus, lightweight titanium metal is used for the construction of the exoskeletons. The glenohumeral joint is often called a ball-and-socket joint, created where the head of the humerus meets the glenoid cavity. Many studies have mainly focused on the glenohumeral joint when designing a three degrees of freedom (DOF) shoulder model.

II.5 DESIGN CRITERIA FOR EXOSKELETON

The exoskeleton robot should move the upper limb in a way that feels natural, like how a human body would move. At the same time, it should not be heavy, and the wearer should feel ergonomically comfortable.

II.5.1 Increasing loads and discomfort

Increasing load and discomfort on the exoskeleton will affect the human body; the exoskeleton reduces the load on the human, but prolonged use of the exoskeleton will increase the back pain of the wearer. In this research, an active exoskeleton is used to give more power to the paralytic patient and to reduce the workload on the muscles.

II.5.2 Centre of rotation of the shoulder complex

Centre of rotation for the shoulder complex (CR) changes with the motion of the arm motions. So, the exoskeleton is designed by taking into consideration the four muscle forces in the human shoulder joint and the position difference between them.

II.5.3 Wearability

The structures of the exoskeletons are designed similarly to the anatomy of humans, with flexibility, lightweight, and high strength. Those systems have been manufactured with less-weight materials such as titanium to reduce discomfort or fatigue. The materials used in the exoskeleton structure are aluminium, titanium, and stainless steel, which are compatible with human skin and do not slip.

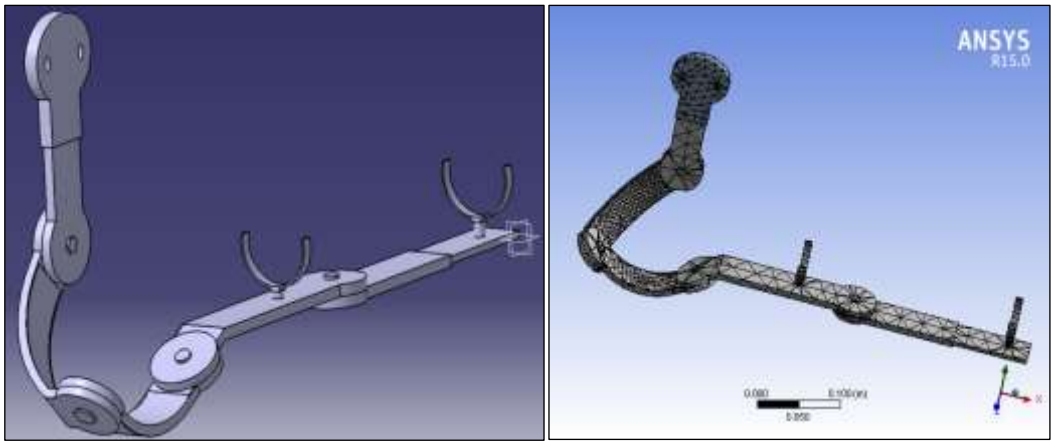
II.5.4 Stress analysis

A stress analysis was conducted to select the material for the exoskeleton structure. After analysing the exoskeleton structure, it is concluded that the titanium is better than the other two materials, aluminium and stainless steel. The stress analysis of the exoskeleton structure of the shoulder flexed 90° and internally rotated 90°.

III. RESULTS AND DISCUSSION

III.1 ANALYSIS OF UPPER ARM EXOSKELETON MODEL

The upper arm exoskeleton model was designed in CATIA modelling software, and then it was imported to Ansys workbench for analysis using different materials with different loading conditions. Figure 3 shows the CATIA exoskeleton model and Ansys meshed structure.

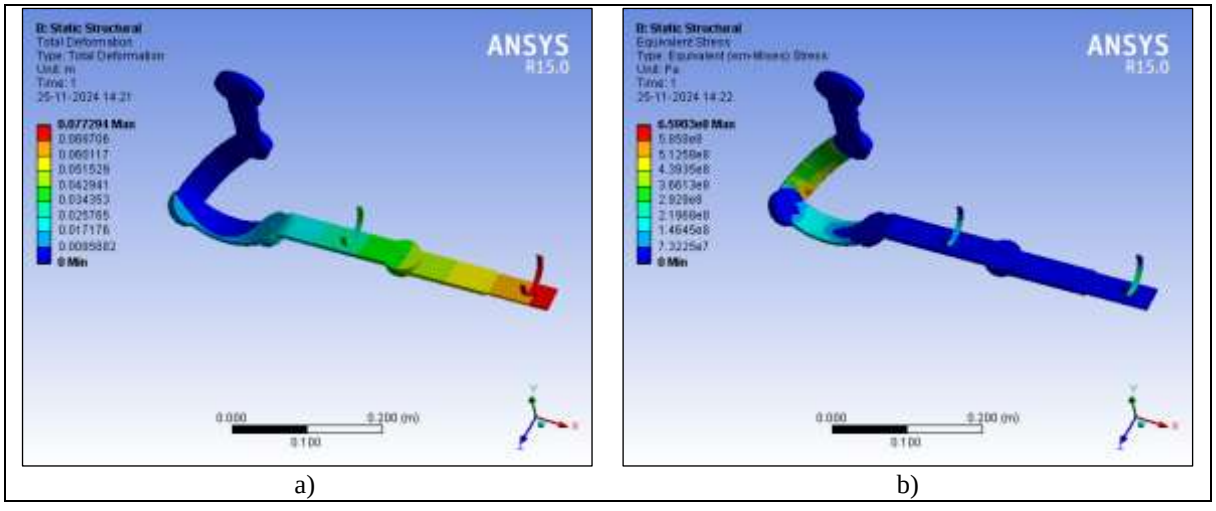


a) b)

Figure 3: a) CATIA model b) Meshed structure in Ansys.
Source: Authors, (2025).

III.2 FINITE ELEMENT ANALYSIS OF EXOSKELETON TITANIUM STRUCTURE

From the analysis, we found that the total deformation is high near the wrist of the exoskeleton, Equivalent von Mises stress is high near the shoulder joint, Maximum Principal Stress uniform deformation is found on the exoskeleton structure, Minimum principal stress deformation is high on the lower side of the shoulder joint, Equivalent elastic strain the deformation is more near the shoulder joint and safety factor is less in the upper part of the shoulder and deformation is found in the lower part of the shoulder. Normal stress deformation is moderate on the whole exoskeleton structure; directional deformation on the X axis is high in the edges, directional deformation on the Y axis is high on the shoulder joint, and Z directional deformation is high from the shoulder joint to the shoulder attachment. Figure 4 represents a different analysis of the exoskeleton arm structure using titanium material.



a) b)

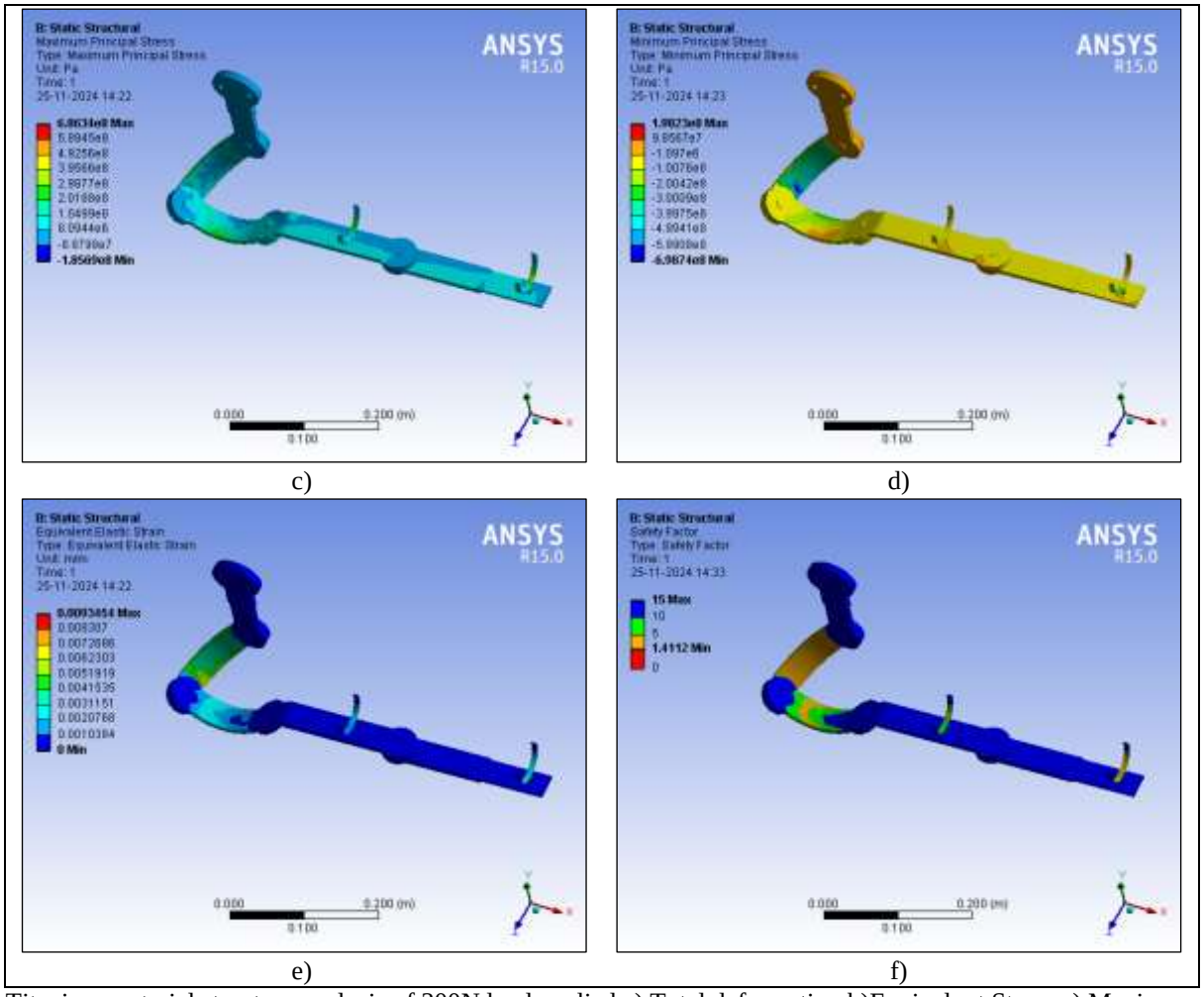
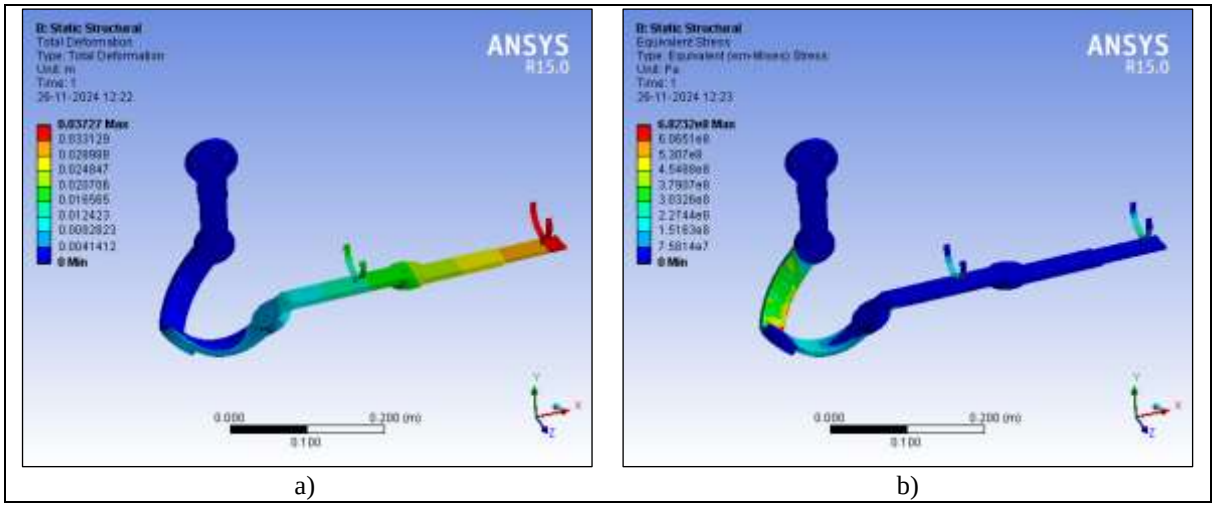


Figure 4: Titanium material structure analysis of 300N load applied a) Total deformation b)Equivalent Stress c) Maximum Principal Stress d) Minimum Principal Stress e)Equivalent Elastic Strain f) Safety factor.
Source: Authors, (2025).

III.3 FINITE ELEMENT ANALYSIS OF EXOSKELETON STEEL STRUCTURE

Figure 5. Shows different analyses of exoskeleton arm structure using steel material. For the steel structure, we found the total deformation is high near the wrist, Equivalent stress is high near the upper shoulder joint, Maximum Principal Stress uniform deformation is found on the exoskeleton structure, Minimum principal stress is high in whole of the structure except the upper shoulder joint, Equivalent elastic strain is high near the shoulder joint, and the safety factor is high near the upper shoulder joint. Normal stress is high in the exoskeleton edge near the shoulder joint, directional deformation about the x-axis is high from the shoulder to the shoulder joint, y-axis deformation about is high in the area surrounding the shoulder joint, and directional deformation about the z-axis deformation is high near the shoulder joint and the exoskeleton attachment near the human spinal cord.



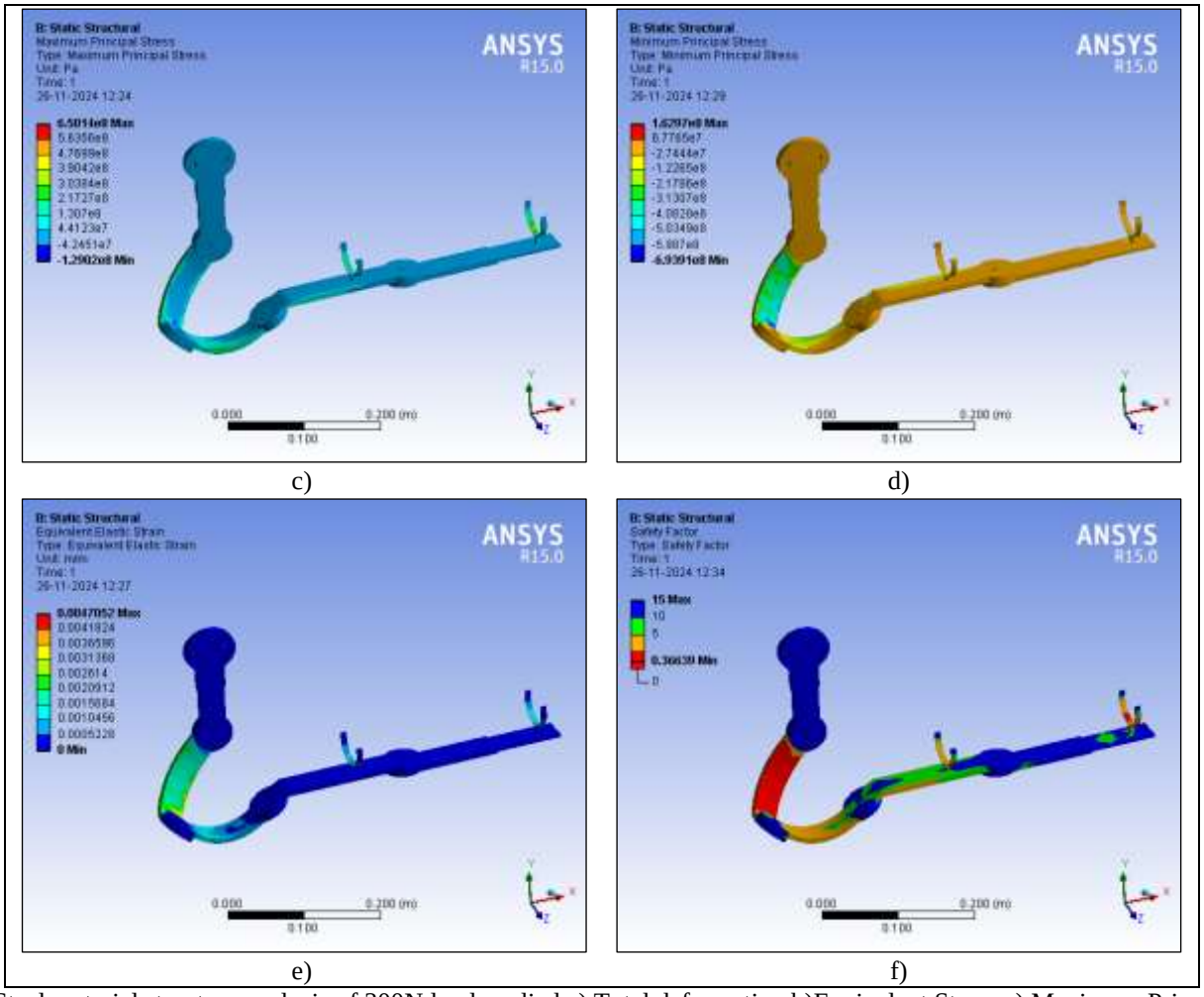
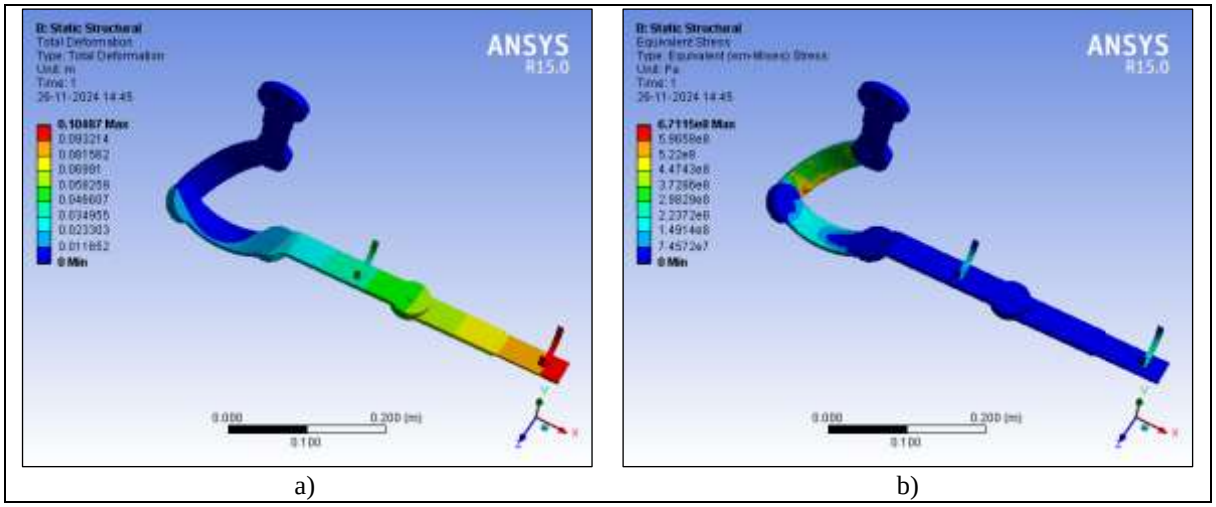


Figure 5: Steel material structure analysis of 300N load applied a) Total deformation b)Equivalent Stress c) Maximum Principal Stress d) Minimum Principal Stress e)Equivalent Elastic Strain f) Safety factor
Source: Authors, (2025).

III.4 FINITE ELEMENT ANALYSIS OF EXOSKELETON ALUMINIUM STRUCTURE

For the aluminium material, we found the total deformation is high near the wrist. Equivalent stress is high above the shoulder joint, Maximum Principal Stress the deformation is uniform on the exoskeleton structure, Minimum principal stress is high in the shoulder and elbow joint, and Equivalent elastic strain deformation is higher in the joint above and below. The Safety factor is high above the shoulder joint. Normal stress is minimal in the whole of the exoskeleton structure, directional deformation about the x-axis is high on the edge of the exoskeleton, Directional deformation about the y-axis is high in the shoulder and directional deformation z-axis is high near the shoulder joint and the exoskeleton joint at the centre. Figure 6. Shows a different analysis of the exoskeleton arm structure using aluminium material.



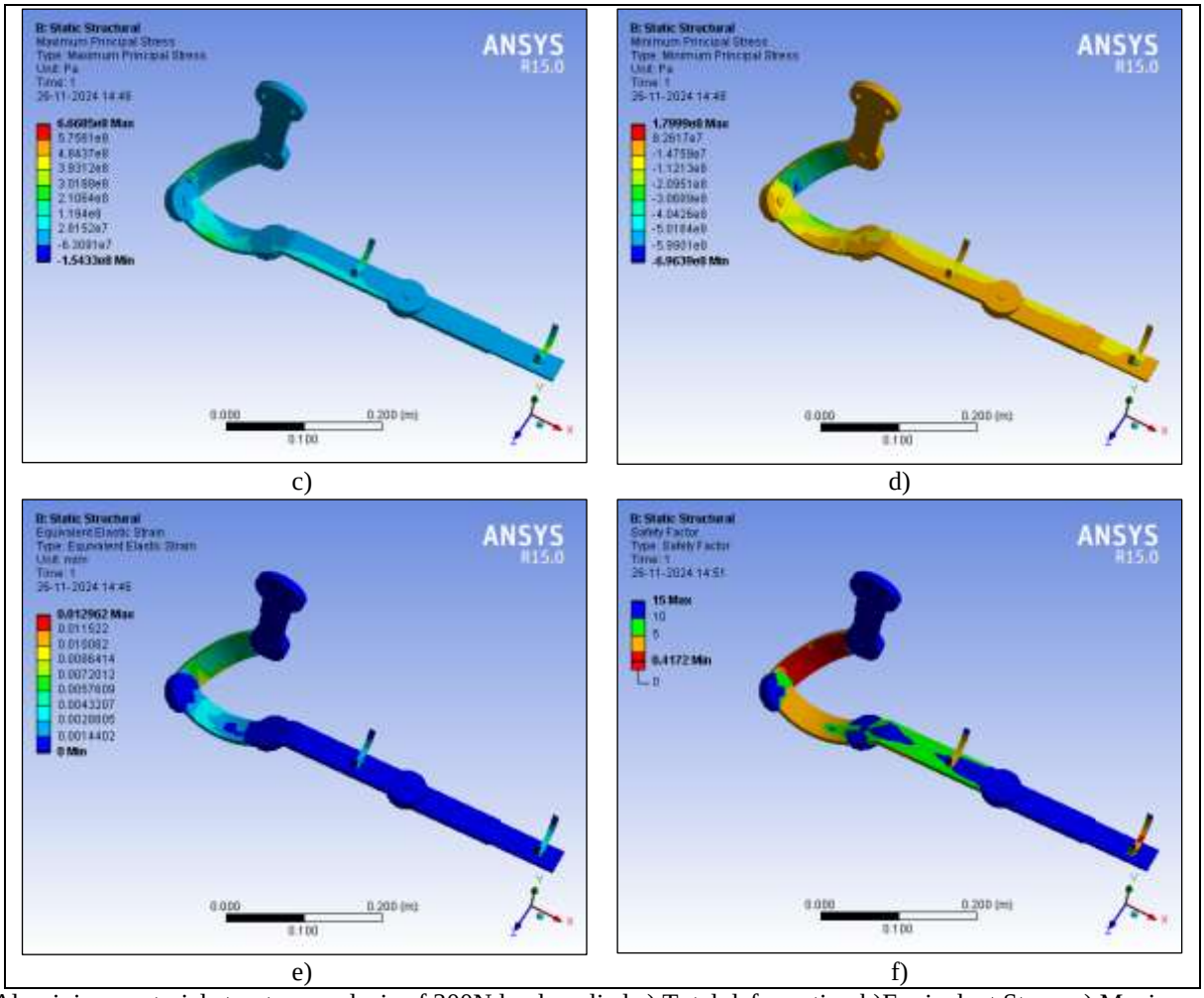


Figure 6. Aluminium material structure analysis of 300N load applied a) Total deformation b)Equivalent Stress c) Maximum Principal Stress d) Minimum Principal Stress e)Equivalent Elastic Strain f) Safety factor
Source: Authors, (2025).

III.5 COMPARISON SUMMARY:

Titanium Equivalent elastic strain, titanium body exoskeleton deformation is higher near the shoulder joint Aluminium Equivalent elastic strain deformation is higher in the joint above and below the shoulder joint. Steel Equivalent elastic strain is high near the shoulder joint, which results in high deformation near the shoulder joint and the stretching of the muscles near the shoulder joint. Design changes should be made near the shoulder joint to take the load falling on the shoulder. Titanium Equivalent Von-mises stress is high near the shoulder joint. Aluminium Equivalent stress is high above the shoulder joint. Steel Equivalent stress is high near the upper shoulder joint. Von-mises stresses predict when a material will yield when complex loading is applied on the exoskeleton shoulder. Low-yielding material is used in the exoskeleton structure. Titanium Maximum principal elastic strain deformation is high near the connecting plates of the shoulder joint. Steel's Maximum principal elastic strain is high around the shoulder joint, yielding on the exoskeleton structure, so the design is carried out in such a way that the exoskeleton structure does not yield compression and tension.

Titanium Maximum Principal Stress uniform deformation is found on the exoskeleton structure. Aluminium Minimum principal stress is high in the shoulder and elbow joint. Titanium Minimum principal elastic strain deformation is high on the whole exoskeleton structure. Aluminium Minimum principal elastic strain is high above and below the shoulder joint. Steel Minimum principal elastic strain is high on the shoulder joint forearm and near the shoulder attachment. Since strain is high near the shoulders, muscle elongation may occur near the shoulder joint. Titanium Minimum principal stress deformation is more on the lower side of the shoulder joint. Aluminium Minimum principal stress is greater in all the other portions of the exoskeleton than in the upper portion of the shoulder joint. Steel Minimum principal stress is greater in the whole structure than in the upper shoulder joint. Titanium Normal elastic strain deformation is moderate on the whole exoskeleton structure.

Aluminium Normal elastic strain is minimal in the whole of the exoskeleton structure. Steel Normal elastic strain is minimal in the entire exoskeleton. Since the strain is minimal, the normal strain did not have any adverse effect on the exoskeleton structure. Titanium Normal stress deformation is moderate on the whole exoskeleton structure. Aluminium Normal stress is minimal throughout the exoskeleton structure. Steel Normal stress is high in the exoskeleton edge near the shoulder joint. When a force acts perpendicular to the surface of an exoskeleton structure, it exerts normal stress. When a force acts parallel to the surface of an exoskeleton structure, it exerts a shear stress. Aluminium can be used since the normal stress is minimal; in titanium, the normal stress is moderate, and steel deformation occurs at the edges.

Titanium safety factor is lower in the upper part of the shoulder, and deformation is found in the lower part of the shoulder. The Aluminium Safety factor is high above the shoulder joint. The Steel Safety factor is high near the upper shoulder joint. The safety factor should be increased to avoid tearing of the shoulder muscle that can occur at the shoulder joint. Titanium Directional deformation X axis is high in the edges of the exoskeleton structure. Aluminium Directional deformation about the x-axis is high on the edge of the exoskeleton. Steel Directional deformation about the x-axis is high from the shoulder to the shoulder joint; since the directional deformation is safe on the x-axis, the deformation on the edges does not affect the exoskeleton structure. Titanium Total deformation is high near the wrist of the exoskeleton. Aluminium Total deformation is high near the wrist. Steel Total deformation is high near the wrist. The wrist bone is weak, and high deformation should be avoided to eliminate injury to the wrist.

Titanium Directional deformation y-axis deformation is high on the shoulder joint. Aluminium Directional deformation about the -axis is high in the shoulder. Steel Directional deformation about the y-axis is high in the area surrounding the shoulder joint. Since the directional deformation is high in the y-axis. Titanium Directional deformation of the z-axis is high from the shoulder joint to the shoulder attachment. The Aluminium Directional deformation z-axis is high near the shoulder joint and the exoskeleton joint. Steel Directional deformation about the z-axis is high near the shoulder joint and the exoskeleton attachment near the human spinal cord. The attachment near the spinal cord will affect the spinal cord joint, and the exoskeleton should be redesigned to minimise the effect. Aluminium Shear stress is minimal in all the portions of the exoskeleton. Steel Shear stress deformation is moderate on the whole exoskeleton structure. Aluminium fails due to shear; in this analysis, the sharer stress is minimal for aluminium, and it can be used.

IV. CONCLUSIONS

Upper-limb exoskeleton systems are used in rehabilitation. In this paper, an extensive analysis of upper-limb exoskeletons was carried out for different design parameters. Moreover, the authors have reviewed exoskeletons to modify their future design for more comfort and rehabilitation purposes. The new design challenges were also identified in the mechanical design of the exoskeleton shoulder joint and the biomechanical design aspects of the exoskeleton for human upper-limb anatomy. In this paper, based on biomechanical design considerations, three types of upper-hand exoskeleton materials were used. The comparison of the parameters was discussed, and it was found that the shoulder joints undergo maximum deformation, hence the muscles adjoining the shoulder joint supraspinatus, infraspinatus, teres minor and subscapularis undergo deformation; if the design has to take care of the deformation to keep the humeral head stable. The spinal cord muscles, rhomboids, and trapezius muscles connecting shoulder blades to the spine also undergo deformation due to the exoskeleton, which may result in pain and discomfort to the patients. It is concluded that with less Young's modulus than steel and more shear strength than aluminium, and it satisfies biomechanical and biocompatibility criteria, it is a better choice for upper-hand exoskeleton construction.

V. AUTHOR'S CONTRIBUTION

Conceptualization: Vijayakumar K, Bubesh Kumar D and Sangeetha Krishnamoorthi.
Methodology: Vijayakumar K, Bubesh Kumar D.
Investigation: Vijayakumar K, Bubesh Kumar D.
Discussion of results: Vijayakumar K, Bubesh Kumar D.
Writing – Original Draft: Vijayakumar K.
Writing – Review and Editing: Vijayakumar K, Bubesh Kumar D and Sangeetha Krishnamoorthi.
Resources: Vijayakumar K.
Supervision: Bubesh Kumar D and Sangeetha Krishnamoorthi.
Approval of the final text: Vijayakumar K, Bubesh Kumar D and Sangeetha Krishnamoorthi.

VI. REFERENCES

[1] Low KH, Xiaopeng Liu, Haoyong Yu., "Development of NTU wearable exoskeleton system for assistive technologies", IEEE Int. Conf. Mechatronics Autom., vol. 2, IEEE; n.d., p. 1099–106, 2005. https://doi.org/10.1109/ICMA.2005.1626705.

[2] Kim S, Nussbaum MA, Mokhlespour Esfahani MI, Alemi MM, Jia B, Rashedi E, "Assessing the influence of a passive, upper extremity exoskeletal vest for tasks requiring arm elevation: Part II – "Unexpected" effects on shoulder motion, balance, and spine loading", Appl Ergon 2018;70:323–30. https://doi.org/10.1016/j.apergo.2018.02.024.

[3] Carmeli E, Peleg S, Bartur G, Elbo E, Vatine J. HandTutor TM "Enhanced hand rehabilitation after stroke — a pilot study", Physiother Res Int 2011;16:191–200, 2011. https://doi.org/10.1002/pri.485.

[4] Nilsson M, Ingvast J, Wikander J, von Holst H, "The Soft Extra Muscle system for improving the grasping capability in neurological rehabilitation", IEEE-EMBS Conf. Biomed. Eng. Sci., IEEE, p. 412–7, 2012. https://doi.org/10.1109/IECBES.2012.6498090.

[5] JarrassÃ© N, Proietti T, Crocher V, Robertson J, Sahbani A, Morel G, et al. "Robotic Exoskeletons: A Perspective for the Rehabilitation of Arm Coordination in Stroke Patients", Front Hum Neurosci, 2014. https://doi.org/10.3389/fnhum.2014.00947.

[6] Bianchi M, Fanelli F, Conti R, Governi L, Meli E, Ridolfi A, et al. "Design and Motion Analysis of a Wearable and Portable Hand Exoskeleton", p. 373–7, 2017. https://doi.org/10.1007/978-3-319-46532-6_61.

[7] Stewart AM, Pretty CG, Adams M, Chen X. "Review of Upper Limb Hybrid Exoskeletons", IFAC-PapersOnLine, 50:15169–78, 2017. https://doi.org/10.1016/j.ifacol.2017.08.2266.

[8] Kim B, Deshpande AD, "An upper-body rehabilitation exoskeleton Harmony with an anatomical shoulder mechanism: Design, modeling, control, and performance evaluation", Int J Rob Res, 36:414–35, 2017. https://doi.org/10.1177/0278364917706743.

[9] Christensen S, Bai S, "Kinematic Analysis and Design of a Novel Shoulder Exoskeleton Using a Double Parallelogram Linkage. J Mech Robot",10, 2018. https://doi.org/10.1115/1.4040132.

[10] Murray D, Meldrum D, Moloney R, Campion A, Horgan F, Hardiman O, "The effects of a home-based arm ergometry exercise programme on physical fitness, fatigue and activity in polio survivors: protocol for a randomised controlled trial", BMC Neurol, 12:157, 2012. https://doi.org/10.1186/1471-2377-12-157.

[11] Kazerooni H. "Human-robot interaction via the transfer of power and information signals", IEEE Trans Syst Man Cybern, 20:450–63, 1990. https://doi.org/10.1109/21.52555.

[12] Vijayakumar K, Kumar DB, Kumar AS, Daniel A, Haripal JL. "Design and fabrication of Exoskeleton robotic arm", p. 020064, 2023. https://doi.org/10.1063/5.0111476.

[13] Kawamoto H, Suwoong Lee, Kanbe S, Sankai Y, "Power assist method for HAL-3 using EMG-based feedback controller", SMC'03 Conf. Proceedings, IEEE Int. Conf. Syst. Man Cybern. Conf. Theme - Syst. Secur. Assur. (Cat. No.03CH37483), vol. 2, IEEE; n.d., p. 1648–53, 2003. https://doi.org/10.1109/ICSMC.2003.1244649.

[14] Islam MR, Spiewak C, Rahman MH, Fareh R, "A Brief Review on Robotic Exoskeletons for Upper Extremity Rehabilitation to Find the Gap between Research Porotype and Commercial Type", Adv Robot Autom 2017;06. https://doi.org/10.4172/2168-9695.1000177.

[15] Young AJ, Ferris DP, "State of the Art and Future Directions for Lower Limb Robotic Exoskeletons", IEEE Trans Neural Syst Rehabil Eng;25:171–82, 2017. https://doi.org/10.1109/TNSRE.2016.2521160.

[16] Gopura RARC, Bandara DSV, Kiguchi K, Mann GKI, "Developments in hardware systems of active upper-limb exoskeleton robots: A review", Rob Auton Syst; 75:203–20, 2016. https://doi.org/10.1016/j.robot.2015.10.001.

[17] Bogue R, "Exoskeletons – a review of industrial applications", Ind Robot An Int J;45:585–90, 2018. https://doi.org/10.1108/IR-05-2018-0109.

[18] Stienen AHA, Hekman EEG, van der Helm FCT, van der Kooij H, "Self-Aligning Exoskeleton Axes Through Decoupling of Joint Rotations and Translations", IEEE Trans Robot ;25:628–33, 2009. https://doi.org/10.1109/TRO.2009.2019147.

[19] Maciejasz P, Eschweiler J, Gerlach-Hahn K, Jansen-Troy A, Leonhardt S, "A survey on robotic devices for upper limb rehabilitation", J Neuroeng Rehabil;11:3, 2014, https://doi.org/10.1186/1743-0003-11-3.

[20] Engín A, "On the biomechanics of the shoulder complex", J Biomech, 13:575–90, 1980. https://doi.org/10.1016/0021-9290(80)90058-5.

[21] Pan G, Fu H, Zhang X, Ma F, "Research on Bionic Mechanism of Shoulder Joint Rehabilitation Movement", p. 181–94, 2017. https://doi.org/10.1007/978-981-10-2404-7_15.

[22] Olsen BS, Søjbjerg JO, Nielsen KK, Væsel MT, Dalstra M, Sneppen O, "Posterolateral elbow joint instability: The basic kinematics", J Shoulder Elb Surg, 7:19–29, 1998. https://doi.org/10.1016/S1058-2746(98)90179-3.

[23] Andrews JG, Youm Y, "A biomechanical investigation of wrist kinematics", J Biomech, 12:83–93, 1979. https://doi.org/10.1016/0021-9290(79)90012-5.

[24] Copaci D, Flores A, Rueda F, Alguacil I, Blanco D, Moreno L, "Wearable Elbow Exoskeleton Actuated with Shape Memory Alloy", p. 477–81, 2017. https://doi.org/10.1007/978-3-319-46669-9_79.

[25] Prinold JA, Masjedi M, Johnson GR, Bull AM, "Musculoskeletal shoulder models: A technical review and proposals for research foci.", Proc Inst Mech Eng Part H J Eng Med, 227:1041–57. 2013. https://doi.org/10.1177/0954411913492303.

[26] Neu C., Crisco J., Wolfe S, "In vivo kinematic behavior of the radio-capitate joint during wrist flexion–extension and radio-ulnar deviation", J Biomech, 34:1429–38, 2013. https://doi.org/10.1016/S0021-9290(01)00117-8.

[27] Shilko E V., Dmitriev AI, Balokhonov RR, Romanova VA, "Multiscale Modeling and Computer-Aided Design of Advanced Materials with Hierarchical Structure", Phys Mesomech, 27:493–517, 2014. https://doi.org/10.1134/S1029959924050011.

[28] Naydenkin E V., Mishin IP, Ratochka I V., Straumal BB, Zabudchenko O V., Lykova ON, et al. "Mechanical Properties and Fracture of Ultrafine-Grained Near β Titanium Alloy under Three-Point Bending", Phys Mesomech, 27:736–46, 2024. https://doi.org/10.1134/S1029959924060109.

[29] Proietti T, Crocher V, Roby-Brami A, Jarrasse N, "Upper-Limb Robotic Exoskeletons for Neurorehabilitation: A Review on Control Strategies", IEEE Rev Biomed Eng, 9:4–14, 2016. https://doi.org/10.1109/RBME.2016.2552201.

[30] Lo HS, Xie SQ, "Exoskeleton robots for upper-limb rehabilitation: State of the art and future prospects", Med Eng Phys, 34:261–8, 2012. https://doi.org/10.1016/j.medengphy.2011.10.004.

[31] Yu H, Choi IS, Han K-L, Choi JY, Chung G, Suh J. Development of a upper-limb exoskeleton robot for refractory construction", Control Eng Pract, 72:104–13, 2018. https://doi.org/10.1016/j.conengprac.2017.09.003.

Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11 n.55, p. 260-267. September/October, 2025.
DOI: https://doi.org/10.5935/jetia. v.11 n.55.2690

RESEARCH ARTICLE OPEN ACCESS

MODELING THE ELASTIC BEHAVIOR OF CLAY SOILS REINFORCED WITH RECYCLED RUBBER FIBERS

Melik Bekhiti[1], Abdelhalim Bensaada[2], Ali Smaida[3]

[1,3]Civil Engineering Department, University of Djelfa, 17000, Djelfa. Algeria
[2]Civil engineering and environmental protection laboratory (LGPE), Civil Engineering Department, Faculty of Technology, University of Medea, 26000. Medea. Algeria

[1]https://orcid.org/0009-0008-2090-4601, [2]https://orcid.org/0009-0009-5280-9788, [3]https://orcid.org/0000-0002-4827-679X

E-mail: bensaada.abdelhalim@univ-medea.dz, m.bkhiti@univ-djelfa.dz, smaidaali@yahoo.fr

ARTICLE INFO

Article History
Received: September 9, 2025
Revised: September 30, 2025
Accepted: October 8, 2025
Published: October 31, 2025

Keywords:
Mixing laws,
Effective Young's modulus,
Scrap tire rubber fiber,
Clay soil.

ABSTRACT

This work use composite material models to forecast the modulus of elasticity of a composite made from scrap tire rubber fibers and clayey soils. The computed modulus of elasticity is juxtaposed with the reference modulus derived from experimental testing. Forecasting the effective mechanical characteristics of composites is essential in scenarios when testing is unfeasible, difficult, or expensive. The investigation encompasses many methodologies within the elasticity framework, including rheological models like Voigt, Reuss, Hirsch-Dougill, Popovics, Halpin-Tsai, Hashin, and the Bache & Napper–Christensen estimate. These models seek to forecast the effective Young's modulus of the composite system consisting of soil and rubber fibers. The highest disparities recorded are 10.66%, 12.71%, and 12.98% for both types of soil. Voigt, Hashin, and Bache calculations provide precise forecasts of the effective Young's modulus, demonstrating remarkable concordance with experimental data across various fiber volume fractions from 10% to 50%.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Used tires are abundant and concerning waste. Aggregates derived from grinding used tires are increasingly used in the field of civil engineering (geotechnical, hydraulic structures, lightweight concrete, asphalt concrete, etc.). Depending on the type of used tires, dimensions, and any separations and treatments, the physical and mechanical properties of these fibers may change. This type of material, such as fiber concrete, fiber plaster, or fiber-reinforced soils, as well as certain natural materials like bone tissues, presents a microstructure consisting of a matrix and a distribution of fibers oriented continuously in all directions (randomly) Fritsch [1]. Evaluating the effective behavior of composite using analytical methods requires a profound understanding of the various approaches existing in the abundant literature in this field. According to Gilormini et al [2], the selection of a model is influenced by various parameters, such as the geometry of the heterogeneous medium, the mechanical contrast between phases, and the volume fraction of reinforcements.

Various researchers have employed models to estimate composites, while others have created distinct models aimed at predicting the elastic properties of specific materials, such as dam concrete [3, 4], rubberized concrete [5 - 7], silica fume concrete [8], and concretes produced with different types of aggregates [9, 10]. The prediction of effective mechanical properties of composites using appropriately validated mixing laws is of great practical interest in all circumstances where tests are impossible, difficult, or expensive. Our study revolves around determining the effective Young's modulus of soil (matrix) reinforced with fibers from recycled rubber tires, comparing it to the effective Young's modulus of experimentally tested soil-fiber composites.

II. MATERIALS AND METHODS

II.1 MATERIALS USED

II.1.1 THE MATRIX PHASE (CLAY) AND ELASTIC YOUNG'S MODULUS

Two soils with distinct origins and physical characteristics were selected for analysis. The first sample was collected from the Ayaida (A) region in Oran, located in northwest Algeria. This soil has been associated with various structural issues in buildings. The second sample is Bentonite (B), provided by the Bental unit in Maghnia, also in northwest Algeria. Fig. 1 illustrates the molecular size distribution of both soils. Young's modulus (E), also known as the elastic modulus of soil, is a key parameter that reflects the stiffness of soil. It represents the ratio of stress to strain within the elastic range of the soil's behavior. This modulus is commonly used in assessing ground support systems and analyzing elastic deformations. The elastic modulus can be determined through either laboratory or in-situ testing methods, or it can be estimated using empirical correlations with other soil properties.

In laboratory settings, it is typically obtained from triaxial tests or indirectly through oedometer tests. In the field, it can be estimated using techniques such as the standard penetration test (SPT), cone penetration test (CPT), pressure meter test, or indirectly via the dilatometer test. Representative Young's modulus values for cohesive materials (in MPa), as presented by Obrzud and Truty [11], compiled from the works of Kezdi [12] and Prat et al [13]. Based on the Unified Soil Classification System (USCS), Ayaida soil is classified as low-plasticity clay (CL), with silica as its primary component. In contrast, Maghnia Bentonite is categorized as high-plasticity clay (CH). Young's modulus of Ayaida soil is 3.509 MPa, and Young's modulus of Maghnia Bentonite is 3.972 MPa.

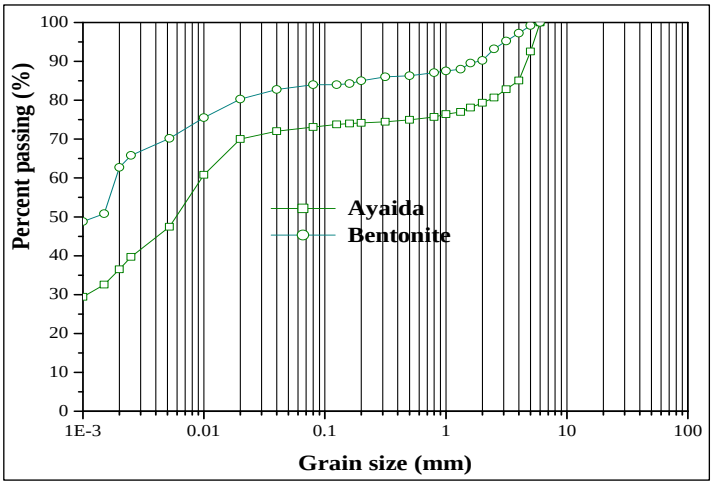


Figure 1: Grain size distributions for the studied materials.
Source: Authors, (2025).

II.1.2. THE REINFORCEMENT PHASE (RUBBER FIBERS) AND ELASTIC YOUNG'S MODULUS

The rubber fibers used in this work come from the processing of used tires by crushing. The steel fibers are separated from the powder magnetically. Fig.2 Photograph showing scrap tire rubber fiber. In addition to fine rubber particles, the powder contains textile fibers. The principle of this experimental campaign is to make test pieces of composite soils and their mixtures with different identical fiber contents (10%, 20%, 25% and 50%) of the same water content and density (homogeneous and isotropic material). The materials are packaged in hermetically sealed bags and are stored at the ambient temperature of the test room (~20°C). The rest time necessary to ensure homogeneous distribution of water within the sample for all composite materials .

Extensive research has been conducted on the Young's modulus of rubber tires. For example, Humphrey et al [14] examined tire chips sourced from three different suppliers by performing compressibility tests under vertical and horizontal loading and unloading conditions at low stress levels. Young's modulus values range from 1.2 to 5.1 MPa, while the average Poisson's ratio falls between 0.20 and 0.32. Triaxial tests were conducted on tire chips ranging in size from 0.08 to 2.01 inches (2 to 51 mm), as reported by Bressette [15], Ahmed [16], Masad et al [17], and Wu et al [18].

Wu et al [18] conducted compression loading tests, including loading-unloading sequences in which the confining pressure (σ_3) was gradually reduced from the initial consolidation pressure while maintaining a constant vertical load (σ_1). The initial tangent modulus derived from the stress-strain curves—comparable to Young's modulus—ranged from 0.3 to 2.5 MPa, with higher values observed under increased confining pressures. Additionally, Hernández et al [19] reported nominal mechanical properties of truck tire rubber, with Young's modulus varying from 1.97 MPa (measured at 100% strain) to 22.36 MPa. In the present study, the average Young's modulus (Ea) of waste tire rubber was found to be 2.214 MPa. The Poisson's ratio for scrap tire rubber is assumed to be 0.45.

Figure 2: Scrap tire rubber fiber.
Source: Authors, (2025).

II.2 METHODS

II.2.1 YOUNG'S MODULUS RESULTS FROM EXPERIMENTAL TESTING OF THE TWO SOILS

The test was conducted on all compacted samples following the ASTM D2435 [20] standard. The test was conducted using a conventional Oedometer with a diameter of 50 mm and a thickness of 20 mm. The load is applied in stages kept constant, successively increasing according to a defined program (0; 0.013; 0.025; 0.051; 0.102; 0.408; 0.815; 1.630) MPa. Variations in the height of the specimen are measured during the test as a function of the duration of application of the load. The principle of this experimental campaign is to make test pieces of composite soils and their mixtures with different fiber contents (0%, 10%, 20%, 25%, and 50%) identical with the same water content and density (homogeneous material and isotropic). The materials are packaged in hermetically sealed bags and are stored at the ambient temperature of the test room (~20°C). The rest time is necessary to ensure a homogeneous distribution of water within the sample for all composite materials. The composite soil is placed in a rigid envelope, a variable pressure is applied using a piston, and the subsidence observed after stabilization is measured, ASTM D 2435. The oedometric module noted E' varies according to the pressures corresponding to the stress interval (Sanglerat).

$$\Delta\sigma = \sigma_1 - \sigma \quad , \quad E' = \frac{1+e}{c_c} \frac{\Delta\sigma}{\log\left(1+\frac{\Delta\sigma}{\sigma}\right)} \tag{1}$$

With e the vacuum index and σ the stress characterize the initial state.

$$E = E'\left(1-\frac{2\upsilon^2}{\upsilon}\right) \tag{2}$$

Where (υ) is the Poisson's ratio of the materials. We can admit that this formula applies to soils as a first approximation. The Poisson's ratio of soils is poorly known and few determinations have been made. However, it seems that U =0.33 constitutes a reasonable estimate. We will therefore write, adopting this value:

$$E = \frac{2}{3}E' \tag{3}$$

Table (1) compiles the Young's modulus of experimentally measured soil-fiber composite (E_c^{exp}). These first tests make it possible to trace the reference curves to then compare them with tests on soil – rubber fiber composites with different volume fractions. Bibliographic data for different experimentally tested soils are compiled in Table 2. The mechanical data are E_m the Young's modulus of the matrix, E_a that of the rubber fibers and E_c^{exp} the Young's modulus of the soil measured experimentally. As well as different volume fractions for soil-fiber composites. In our work we tested two types of clay soils with Young's modulus (3.509 and 3.972) MPa. Each type of matrix soil is mixed with reinforcement of rubber fibers having Young's modulus (2.214) MPa. Which corresponds to a contrast ratio between the two phases varying from (Ea/ Em) varying from 0.56 to 0.63.

Table 1: the modulus of elasticity of soil-fiber composite measured experimentally.

Ref	σ(MPa)	Δσ (MPa)	e	c_c	E_c^{exp} (MPa)
A1	1.63	0.518	0.595	3.99	3.509
A2			0.285	6.64	3.185
A3			0.550	6.98	3.295
A4			0.689	12.62	3.221
A5			0.357	7.31	3.203
B1			0.799	2.66	3.972
B2			0.773	5.65	3.472
B3			0.698	5.32	3.455
B4			0.853	7.64	3.396

Source: Authors, (2025).

Table 2: Experimental characteristics and volume fraction for soil-fiber composites.

Ref.	V_a	V_m	E_a(MPa)	E_m(MPa)	$\mathbf{E_c^{exp}}$ (MPa)
A_1	0	1	2.214	3.509	3.509
A_2	0.1	0.9			3.185
A_3	0.2	0.8			3.295
A_4	0.25	0.75			3.221
A_5	0.5	0.5			3.203
B_1	0	1		3.972	3.972
B_2	0.1	0.9			3.472
B_3	0.2	0.8			3.455
B_4	0.25	0.75			3.396
B_5	0.5	0.5			3.316

Source: Authors, (2025).

II.2.2 CHOICE OF ANALYTICAL MODELS

As noted by Gilormini and Bréchet [2], model selection is influenced by several factors, notably the geometry of the heterogeneous medium, the mechanical contrast between the matrix phases (soils), and the volume fraction of the reinforcements (rubber fibers). Remember that in our case, we are studying a heterogeneous composite material (soil-rubber fibers from worn tires) composed of two homogeneous phases: the fibers (or the inclusion; phase a) of volume fraction (Va), embedded in the soil (or matrix, phase m). Each of these two phases presents a linear, homogeneous and isotropic elastic behavior. The two phases are assumed to be perfectly stuck together (matrix/inclusion bond, perfect adhesion is assumed), which justifies the biphasic representation. Used tire rubber fibers are evenly distributed in the soil matrix.

The multi-phase description of composite materials made up of an elastic matrix. The determination of the effective properties then consists in this case of defining the homogeneous equivalent soil-fiber behavior based on the characteristics of the soil and the rubber fibers of the worn tires. The two mixing laws of Voigt and Reuss are simple and incapable of determining with acceptable precision the modulus of elasticity of the composite because they do not take into account the morphology of the material, the discontinuity of the reinforcement, nor its orientation and of the true nature of matrix/reinforcement interface, these two terminals are enriched by combining them with each other. Two combinations proposed by Larrard [21]. The effective elastic modulus for both models (Voigt and Reuss) is given by equations (4) and (5). The biphasic models of Hirsch-Dougill, Popovics , Halpin-Tsai, Hashin & Bache and Napper-Christensen originally designed for particle composites Nielsen et al [22], propose effective elastic modulus of the composite by equations (6), (7), (8), (9) and (10) successively.

$$E_{CVoigt} = E_m V_m + E_a V_a \tag{4}$$

$$E_{C\,\mathrm{Re}uss} = \frac{E_m}{V_m} + \frac{E_a}{V_a} \tag{5}$$

$$E_{CHirsh-Dougill} = \frac{1}{2}\left(\frac{1}{E_{CVoigt}} + \frac{1}{E_{C\,\mathrm{Re}uss}}\right) \tag{6}$$

$$E_{CPopovics} = \frac{1}{2}\left(E_{CVoigt} + E_{C\,\mathrm{Re}uss}\right) \tag{7}$$

$$E_{CHalpin-Tsai} = \frac{3}{8}E_{CVoigt} + \frac{5}{8}E_{C\,\mathrm{Re}ss} \tag{8}$$

$$E_{CHalpin} = \frac{(E_a + E_m) + (E_a - E_m)V_a}{(E_a + E_m) - (E_a - E_m)V_a} E_m \quad (9)$$

$$E_{CHalpin} = E_m^{V_m} E_a^{V_m} \quad (10)$$

Where E_m represents the modulus of elasticity of the soil matrix, E_a is the modulus of elasticity of the fiber phase, E_C is the modulus of elasticity of soil-fibers, V_m is the volume fraction of the soil as matrix and V_a is the volume fraction of the rubber fibers of the reinforcement phase. E_C Takes the indices of the composite model used for estimation of the elastic modulus. For example, E_C Voigt expresses the elastic modulus of soil-fibers estimated using the Voigt composite model.

III. RESULTS OF PREDICTIVE APPROACHES AND DISCUSSION

III.1 IMPACT OF FIBER CONTENT ON THE DRY DENSITY OF TWO CLAY SOILS

Based on the results depicted in Fig. 3, it is evident that the dry density of fiber-reinforced soils gradually decreases with the increasing content of fibers for both soil types A and B. This reduction is attributed to the decrease in the average unit weight of solids within the mixture. The test results indicate that the inclusion of fibers leads to a decrease in the maximum dry density in both soils A and B, primarily due to the lower density of the added fibers. The decline in the mixture's density can be attributed to several factors. Firstly, the rubber fibers have a smaller specific gravity compared to the soil. In this study, the dry density of clay B and A, and rubber fibers were 1.98 g/m³, 1.93 g/m³, and 1.20 g/cm³, respectively. Secondly, the flexibility of rubber fibers contributes to enhanced compaction efficiency, resulting in a reduced maximum dry density of the soil and fiber mixture.

Thirdly, the lower optimum moisture content of clay and fibers, influenced by the poor water absorption of rubber fibers compared to clay. For instance, an increase in the content of tire rubber fibers from 10% to 50% leads to a reduction in density by 26.67% and 26.88% for soil A and B, respectively. These findings align with previous studies [23, 24] indicating that, while the optimum water content remains unchanged, the maximum dry unit weight of the mixture decreases by about 0.3 g/cm³ with the addition of 20% rubber by weight . Bekhiti et al [25] also observed a decrease in dry density with an increase in waste tire rubber fiber content. The highest dry density value, 1.29 g/cm³, was achieved with a 30% PVC waste aggregate content. Interestingly, this maximum dry density is 17.83% lower than the unreinforced samples [26].

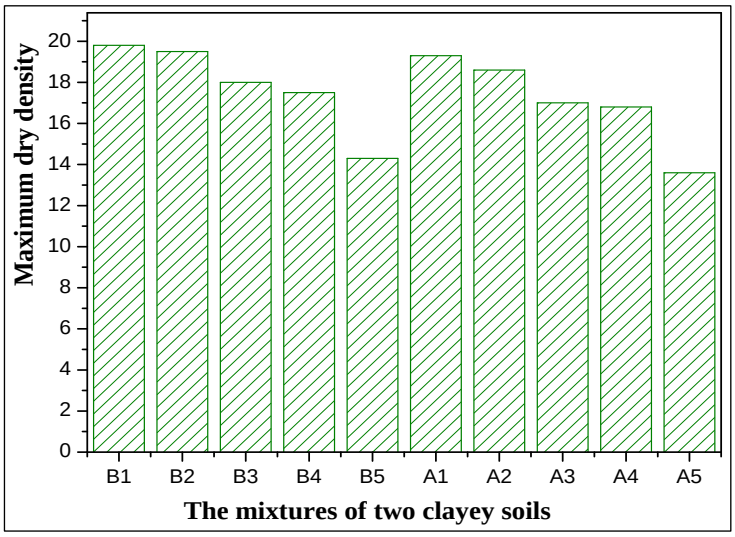


Figure 3: Variation of dry density of different mixtures.
Source: Authors, (2025).

III.2 COMPARISON OF ANALYTICAL MODELS WITH EXPERIMENTAL RESULTS.

Table 3 shows our results from the application of analytical approaches and those obtained by experimental method. To identify the most accurate model for calculating the effective Young's modulus, the differences between the predicted values from various models and the experimental results were computed (see Table 4). The average relative differences between the measured experimental E_c^{Exp} Young's modulus and the numerical predictions E_c^{Cal} are calculated by the relation below:

$$E_{Cart} = \left| \frac{E_C^{Cal} - E_C^{Exp}}{E_C^{Exp}} \right| x100 \quad (11)$$

Figs. 4 and 5 give the evolution of the effective Young's modulus obtained by the different predictive approaches processed and the experimental results as a function of the volume fraction of the rubber fibers (reinforcement) (Va) which takes the values: 0-0.1-0.2-0.25 and 0.50 for a Young's modulus is 2.214 MPa.

If we examine the results presented in Table 4 we notice that the maximum difference is reached with the combined models of Reuss and Hirsch is worth 93.66% and 88.08% respectively, this is a great value. The two models are not capable of predicting the effective behavior of the composites (soil-fiber). As for the Popovics and Halpin-Tsai models, they gave percentages close to half, for example, with 10% fibers, this gave a difference of 42.21% and 54.29% for soil A, as for the ground B; this gave 46.83% and 54.58% respectively. The Voigt, Hashin, and Bache approximations demonstrate a minimal deviation range, spanning from 0.00% to 10.66% for Voigt, 0.00% to 12.717% for Hashin, and 0.00% to 12.988% for Bache, considering two types of soils, A and B. These three models prove effective in predicting the composite materials (soil-rubber fibers) effective Young's modulus.

Table 3: Effective modulus of elasticity (MPa): Comparative analysis of various analytical models against experimental results.

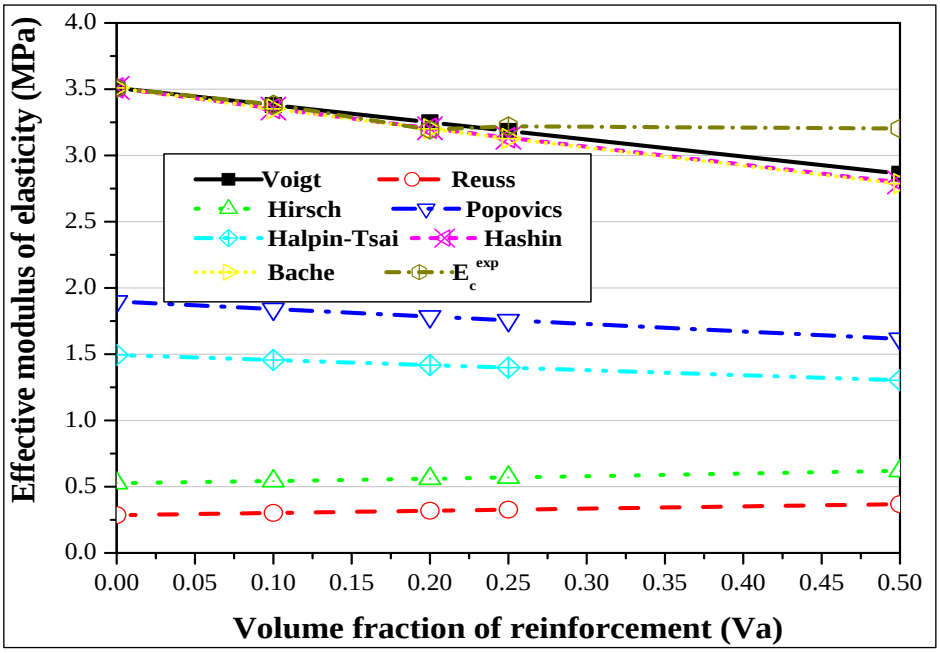
Ref.	E_{CVoigt}	E_{CReuss}	$E_{CHirsch}$	$E_{CPopovics}$	$E_{CHalpin-Tsai}$	$E_{CHashin}$	E_{CBache}	E_C^{exp}
A_1	3.510	0.281	0.526	1.898	1.495	3.600	3.601	3.601
A_2	3.380	0.301	0.544	1.840	1.455	3.353	3.352	3.386
A_3	3.250	0.320	0.560	1.783	1.419	3.204	3.200	3.194
A_4	3.190	0.331	0.568	1.757	1.400	3.132	3.128	3.222
A_5	2.860	0.370	0.618	1.614	1.301	2.797	2.788	3.203
B_1	3.971	0.251	0.474	2.111	1.646	3.973	3.973	3.971
B_2	3.801	0.271	0.491	2.035	1.594	3.751	3.748	3.471
B_3	3.620	0.291	0.510	1.957	1.540	3.544	3.535	3.454
B_4	3.530	0.301	0.521	1.916	1.511	3.446	3.433	3.397
B_5	3.101	0.351	0.580	1.723	1.380	2.983	2.966	3.315

Source: Authors, (2025).

Table 4: Percentage deviations in Young's modulus between analytical model predictions and experimental results.

Ref.	E_{CVoigt}	E_{CReuss}	$E_{CHirsch}$	$E_{CPopovics}$	$E_{CHalpin-Tsai}$	$E_{CHashin}$	E_{CBache}
A_1	0.000	91.880	84.980	45.940	57.420	0.000	0.000
A_2	6.110	90.530	82.950	42.210	54.290	5.300	5.120
A_3	1.370	90.340	83.000	45.850	56.970	2.730	2.880
A_4	1.110	89.860	82.330	45.480	56.580	2.720	2.920
A_5	10.660	88.500	80.670	49.580	59.110	12.720	12.990
B_1	0.000	93.660	88.080	46.830	58.540	0.000	0.000
B_2	9.340	96.170	85.830	41.420	54.110	8.080	7.920
B_3	4.790	91.560	85.210	43.380	55.430	2.600	2.290
B_4	4.020	91.110	84.630	43.550	55.440	1.440	1.060
B_5	6.720	89.390	82.480	48.060	57.420	0.000	10.580

Source: Authors, (2025).



Figurre 4: Comparison of analytical and experimental results for effective elastic moduli as a function of the volume fraction of the reinforcement in Ayaida clay.
Source: Authors, (2025).

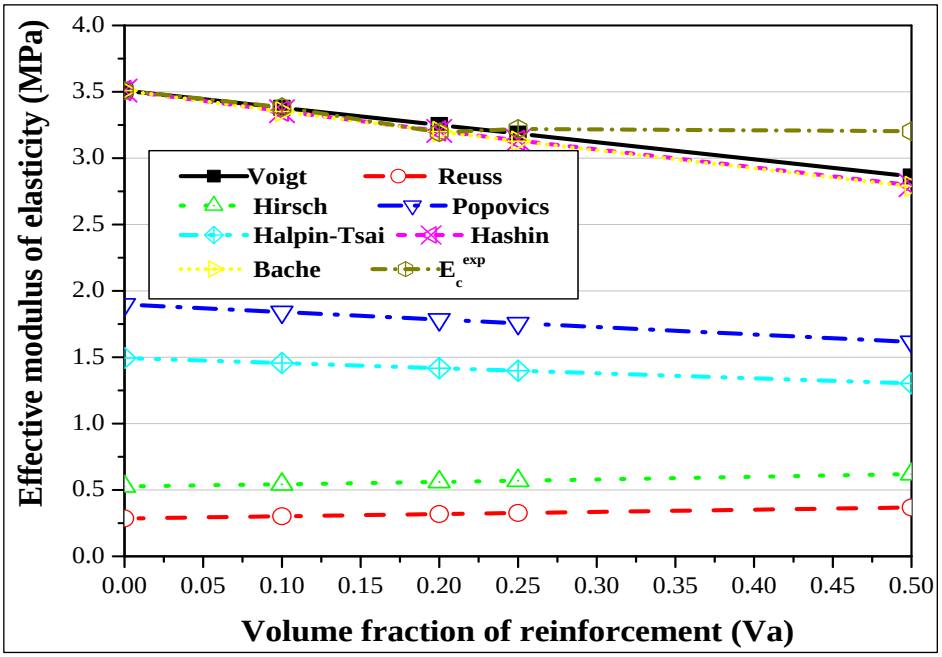


Fig.ure 5: Comparison of analytical and experimental results for effective elastic moduli as a function of the volume fraction of the reinforcement in Bentonite of Meghnia.
Source: Authors, (2025).

IV. CONCLUSIONS

Our comparative study allows us to conclude that:

- The increase in fiber content results in a reduction of the dry density in both varieties of clay soil, mostly due to the reduced specific gravity and unit weight of the fibers.
- The elevation of rubber fiber content results in a reduction of the modulus of elasticity in soils A and B.
- The results obtained from predictive approaches for calculating the modulus of elasticity in two clayey soils reinforced with rubber fibers from worn tires indicate that the Voigt, Hashin, and Bache models are in closer agreement with experimental results, showing a maximum deviation of 10.66% for Voigt, 12.717% for Hashin, and 12.988% for Bache in the case of soil A. In soil B, the maximum differences are 6.72%, 0%, and 10.58% with the same models when using 50% fibers. The Voigt and Hashin models exhibit superior accuracy, particularly for clayey soils incorporating fibers.

This study involves predicting the effective mechanical properties of composites using well-validated mixing laws and is of significant practical interest in situations where testing is impossible, difficult, or expensive. In this study, some of these composite models provide accurate calculations and closer values for the moduli of elasticity. They can also calculate the modulus of elasticity for clayey soil reinforced with waste rubber fibers, providing a feasible alternative to prior models with requisite assumptions. Future studies could investigate some of the composite models not examined in this paper for clay soil reinforced with rubber fibers to forecast the modulus of elasticity of elasticity and compare the results with experimental data. Future studies could develop composite models specifically tailored for soils, rather than relying on models primarily designed for concrete, to precisely forecast the modulus of elasticity for clay composites augmented with rubber fibers derived from recycled tires.

V. AUTHOR'S CONTRIBUTION

Conceptualization: Melik Bekhiti, Abdelhalim Bensaada, and Ali Smaida
Methodology: Author One and Author Two.
Investigation: Author One and Author Two and Author Three.
Discussion of results: Author Two and Author Three.
Writing – Original Draft: Author One and Author Two.
Writing – Review and Editing: Author One and Author Two.
Resources: Author Two and Author Three.
Supervision: Author Two and Author Three.
Approval of the final text: Author One, Author two and Author Three.

VI. REFERENCES

[1] A. Fritsch, *Inter-prétentation multi-échelle de l'élasticité et de la résistance de l'os et des biomatériaux en hydroxyapatite, vitrocéramique ou titane: une approche basée sur la micromécanique des milieux continus*, Ph.D. dissertation, ENPC, 2009.

[2] P. Gilormini and Y. Bréchet, "Mechanical properties of heterogeneous media: which material for which model? Which model for which material," *Model. Simul. Mater. Sci. Eng.*, vol. 7, pp. 805–816, 1999, doi: 10.1088/0965-0393/7/5/312.

[3] I. B. Topçu, "Alternative estimation of the modulus of elasticity for dam concrete," *Cem. Concr. Res.*, vol. 35, no. 2, pp. 2199–2202, 2005, doi: 10.1016/j.cemconres.2004.08.010.

[4] J. Vilardell, A. Aguado, L. Agullo, and R. Gettu, "Estimation of the modulus of elasticity for dam concrete," *Cem. Concr. Res.*, vol. 28, no. 1, pp. 93–101, 1998, doi: 10.1016/S0008-8846(98)00124-1.

[5] I. B. Topçu, "Assessment of the brittleness index of rubberized concretes," *Cem. Concr. Res.*, vol. 27, no. 2, pp. 177–183, 1997, doi: 10.1016/S0008-8846(96)00199-8.

[6] I. B. Topçu and N. Avcular, "Analysis of rubberized concrete as a composite material," *Cem. Concr. Res.*, vol. 27, no. 8, pp. 1135–1139, 1997, doi: 10.1016/S0008-8846(97)00115-4.

[7] I. B. Topçu, "Investigation of rubberized concrete by composite material rules," in *Proc. Kütahya Structural Mechanics Conf.*, Kütahya, Turkey, 1994, pp. 71–80.

[8] D. Mostofinejad and M. Nozhati, "Prediction of the modulus of elasticity of high strength concrete," *Iran. J. Sci. Technol.*, vol. 29, no. B3, pp. 311–321, 2005. https://www.researchgate.net/publication/237812225

[9] P. Simeonov and S. Ahmad, "Effect of transition zone on the elastic behavior of concrete," *Cem. Concr. Res.*, vol. 25, no. 2, pp. 304–310, 1995, doi: 10.1016/0008-8846(94)00124-H.

[10] A. U. Nilsen and P. J. M. Monteiro, "Concrete: A three phase material," *Cem. Concr. Res.*, vol. 23, no. 1, pp. 147–151, 1993, doi: 10.1016/0008-8846(93)90145-Y.

[11] R. Obrzud and A. Truty, *The Hardening Soil Model – A Practical Guidebook*, Z Soil, PC 100701 Report, rev. Jan. 31, 2012.

[12] A. Kezdi, *Handbook of Soil Mechanics*. Amsterdam, Netherlands: Elsevier, 1974.

[13] M. Prat, E. Bisch, A. Millard, P. Mestat, and G. Cabot, *La modélisation des ouvrages*. Paris, France: Hermes, 1995.

[14] Y. D. N. Humphrey, M. M. Sandford, G. Cribbs, and W. P. Manion, "Shear strength and compressibility of tire chips for use as retaining wall backfill," *Transp. Res. Rec.*, no. 1422, pp. 29–35, 1993.

[15] T. Bressette, *Used Tire Material as an Alternative Permeable Aggregate*. Sacramento, CA: California Dept. of Transportation, Office of Transportation Laboratory, Rep. No. FHWA/CA/TL-84/07, 1984.

[16] I. Ahmed, *Laboratory Study on Properties of Rubber–Soils*, Purdue Univ., Indiana, Joint Highway Research Project, Rep. No. FHWA/IN/JHRP-93/4, 1993, doi: 10.5703/1288284314210.

[17] E. Masad, R. Taha, and C. H. T. Papagiannakis, "Engineering properties of tire/soil mixtures as lightweight fill material," *Geotech. Test. J.*, vol. 19, no. 3, pp. 297–304, 1996, doi: 10.1520/GTJ10355.

[18] W. Y. Wu, C. C. Benda, and R. F. Cauley, "Triaxial determination of shear strength of tire chips," *J. Geotech. Geoenviron. Eng.*, vol. 123, no. 5, pp. 479–482, 1997, doi: 10.1061/(ASCE)1090-0241(1997)123:5(479).

[19] O. F. Hernández and G. Barluenga, "Fire performance of recycled rubber-filled high-strength concrete," *Cem. Concr. Res.*, vol. 34, no. 1, pp. 109–117, 2004, doi: 10.1016/S0008-8846(03)00253-9.

[20] ASTM D2435, *Standard Test Method for One-Dimensional Consolidation Properties of Soils*, ASTM International, 2005.

[21] F. de Larrard, "Une approche de la formulation des bétons légers de structure," *Bull. Liaison Lab. Ponts Chaussées*, no. 195, pp. 39–47, 1995.

[22] H. G. Nielsen and E. L. Chen, "Young's modulus of composites filled with randomly oriented short fibers," *J. Mater.*, vol. 3, no. 2, pp. 352–358, 1968.

[23] H. Seda, J. C. Lee, and J. A. H. Carraro, "Beneficial use of waste tire rubber for swelling potential mitigation in expansive soils," in *Geo-Denver Conf.*, ASCE, 2007, pp. 1–9, doi: 10.1061/40916(235)5.

[24] F. J. Dunham and J. A. Carraro, "Effects of compaction effort, inclusion stiffness and rubber size on the shear strength and stiffness of expansive soil–rubber (ESR) mixtures," in *Geo-Congress*, 2014, pp. 3635–3644, doi: 10.1061/9780784413272.352.

[25] M. Bekhiti, H. Trouzine, and M. Rabehi, "Influence of waste tire rubber fibers on swelling behavior, unconfined compressive strength and ductility of cement stabilized bentonite clay soil," *Constr. Build. Mater.*, vol. 208, pp. 304–313, 2019, doi: 10.1016/j.conbuildmat.2019.03.011.

[26] L. Karboua, M. Bekhiti, and R. Zaitri, "Optimizing the mechanical behavior of clay using cement and PVC aggregates with the application of a mixture design method," *Constr. Build. Mater.*, vol. 208, pp. 304–313, 2023, doi: 10.1016/j.conbuildmat.2023.133343.

Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11 n.55, p. 268-275. September/October., 2025.

DOI: https://doi.org/10.5935/jetia. v11i55.2712

RESEARCH ARTICLE OPEN ACCESS

INVESTIGATION ON PERFORMANCE EVALUATION OF CONTAINERIZED MICROSERVICE FOR EDGE COMPUTING ENVIRONMENT

Pranjit Kakati[1], Abhijit Bora[2]

[1,2]Assam Don Bosco University, Assam

[1]https://orcid.org/0009-0001-3312-7847, [2]http://orcid.org/0000-0002-7754-639X

E-mail: kpranjit05@gmail.com, abhijit.bora0099@gmail.com

ARTICLE INFO

Article History
Received: October 18, 2025
Revised: October 20, 2025
Accepted: October 21, 2025
Published: October 31, 2025

Keywords:
Microservice,
Containerized technology,
Edge computing,
Performance metrics.

ABSTRACT

Microservices have emerged as one of the most popular software development architectures among developers, industries, and academia due to their advantages over traditional monolithic architectures. Many large technology companies are transitioning from monolithic systems to microservice-based architectures. Containerization technologies offer scalable and efficient deployment solutions. This study proposes a methodology for deploying a microservice designed for greenhouse-related computations using Docker and Kubernetes containerization method. The experimental architecture is presented, and the deployed microservice is evaluated using key performance metrics, including Response time(RT), Throughput (Th), and Hits per second (HpS), under varying user loads. Statistical methods are applied to analyze the microservice performance results. The findings conclude that the quality aspects of deploying microservices through the containerization method are stable up to a specific stress level within an edge computing environment.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

With the fast development of technologies in the communication and information sector, the past decade has witnessed significant advancements in smart living, data transmission, and related fields such as surveillance, navigation, sensors, infotainment, networking, and low-power circuits[1]. Consequently, data traffic across internetwork and processing of data in the cloud has surged tremendously. Although, cloud infrastructure has been leveraged to manage this demand, it faces challenges like high RT, network congestion, and heavy storage burdens. These limitations are particularly critical for applications that are delay-sensitive such as real-time video streaming, augmented reality and autonomous driving which require substantial data processing and low-latency performance [2]. Edge computing emerges as a solution to the increased RT in cloud, where edge computing enabling processing of data at the proximity of the network for downstream data as a representative of services in the cloud and upstream data in representation of services of smart devices/IoTs [3].

The emerging paradigm Edge computing integrates networking, computation, and storage capabilities nearer to end users by shifting them from remote cloud servers to the edge of the network being used. Edge computing has enormous potential to deliver applications, more intelligent services and with enhanced experiences of the users in the of IoT and 5G context [4]. Edge computing is ideal for real-time processing, decision making locally and low latency applications. However, resource constraints and heterogeneity in edge demands a flexible, lightweight and scalable system that support easy and fast development and deployment among service providers. To deploy the functions of cloud computing on the edge, it is required to adopt the architecture and technologies that are suitable for available resources in edge [5]. In this context, microservice applications come up with characteristic like loose coupling, scalability, fine granularity, fault tolerance, low maintenance cost etc [6]. An emerging architectural style called microservices was created to overcome the drawbacks of conventional monolithic software structures. All of an application's components are closely connected and function as a single process in monolithic systems. While this approach can simplify initial development, it often leads to

significant challenges in maintenance, scalability, and system evolution as the application grows in complexity. To get around these issues, the microservice design suggests breaking up an application into a number of tiny, loosely connected, and autonomously deployable services. Each microservice is designed to carry out a certain business task and runs in its own process [7]. These services communicate with one another through lightweight mechanisms, typically HTTP-based APIs or messaging queues. The ability to scale each components independently as per developer's requirement is one of the major advantage of microservices instead of scaling the program as a whole. This approach also enhances resilience of the system, because the system as a whole is not always impacted when one service fails. Additionally, microservices provides the facility of continuous integration and delivery by allowing development teams to independently build, test, deploy, and update services. Thus, microservices promote modularity, scalability and agility in development of application, making them an attractive choice for modern, cloud, and edge native applications.

Container technology is one of the recent advancements in cloud services. It is a container based virtualization where the strategy of machine based virtualization has been upgraded. Here, hardware infrastructure is broken down into virtual entities and different containers used the shared operating system[8]. Containerization offers the advantages of resource allocation and isolation, much like server virtualization technologies. Encapsulation of lightweight, executable package containing codes, setting and system libraries in containerization is the main advantages of the technology [9]. The Docker container was used in this experiment. The management of microservice containers in manual mode with starting and stopping them, monitoring them, balancing of load, and updating them is difficult and error prone for a project. Kubernetes is a container cluster management system that will automatically orchestrate all containers. It is an open source platform used to automate the maintenance, scaling and deployment of containerized microservice applications.

Running a microservice application in Kubernetes involves several stages: development, containerization, deployment, monitoring, and scaling. Each microservice is developed independently and remains loosely coupled using the developer's preferred programming language along with necessary dependencies and settings. To containerize the application, a dockerfile is written for each microservice, which is then used to build Docker images. These containers can be locally built and tested to ensure initial correctness. Once containerization is complete, Kubernetes manifests are created for each microservice. These typically include: (a) Deployment which specifies the Docker image and the number of replicas (pods) to run, (b) Service which exposes the application internally within the cluster or externally to users, and (c) ConfigMaps/ Secrets which store data related to configuration and sensitive information required by the microservices.

At the core of execution, Kubernetes schedules Pods, which are the smallest units of deployment that encapsulates containers. Every deployment ensures that the designated quantity of pod replicas are operational, thereby enabling fault tolerance and scalability. The Kubernetes scheduler dynamically places pods on available nodes, considering resource availability and workload balancing. In a Kubernetes cluster, each pod has a unique IP address that enables inter-Pod communication via APIs or services. To allow external access, services that expose Pods either within the cluster or to external clients. Thus, Kubernetes manages the scheduling, scaling, and orchestration of containerized microservices, providing a robust and flexible platform for running distributed applications[10]. With the increased adaptation of containerization and orchestration technologies in the management and scaling of large applications, it is imperative to evaluate the different performance parameters of microservice execution under varying workloads. Evaluation with an increase in concurrent user threads provides insights into the system's ability to handle high load situations. Performance evaluation will not only help with responsive benchmarking but also help identify limitations and bottlenecks in microservice deployment through Docker and Kubernetes.

II. RELATED WORK

Microservices is now a widely used development and deployment architectural style by different production organizations and is a topic of research and experiments by academics and rehearses. Different aspects of microservice like its architectures, its scaling and deployment, performance evaluation etc. are being studied and researched around the globe.*Vayghan et al.* discussed about architecture of microservices, its deployment, maintainability and scalability through Kubernetes. The author stated that Kubernetes increases application availability through self-healing and failure recovery. They examined the availability of Kubernetes applications by considering the default configuration [10]. Rossi et al. presents on Multilevel Elastic Kubernetes (me-kube) as Kubernetes extension that introduces a hierarchical architecture to control elasticity of application based on microservice.

The author proposes novel proactive and reactive policies of hierarchical control on queuing theory[11]. *Amaral etal.* discuss about the two models for implementation of microservice architecture through containers - master-slave and nested-container. The author experimented in these two models to compare performance of CPU and network running benchmark[12]. *Čilić et al.*presents analysis of edge orchestration platform for Quality of Service (QoS) in terms of their workflow. According to their research, Kubernetes and its distributions may be able to schedule resources on the edge of a network effectively [13]. *Medhi et al.* presents a study on SOAP based service on windows communication. The authors proposes a design and implementation of service oriented prototype research on services of automated teller machine using technology of Windows communication foundation to investigate and predict some aspects of reliability of web services. Here the authors presentnovel prototype architecture, procedures of testing, HTTP transactions and reliability analysis under massive user load[14]. *Aitlmoudden et al.* in their paper proposes a framework based on.

Microservice for scalable data analysis which contain IoT integration and Microservices in agriculture domain. The framework maintains scalability and fault tolerance while offering speedier data processing and agriculture data analysis activities. The authors also address the obstacles like data integration, development of analytical microservices on agriculture and ensuring data security[7]. *Qu et al.* proposed a microservice architecture designed to deliver scalable as well as extensible services within large-scale distributed systems. Their research investigated several microservice deployment strategies on an edge computing platform with Docker containers. Furthermore, the authors conducted a performance analysis of microservices and examined the interference effects arising from concurrent microservice execution under benchmark scenarios [6]. *Huang et al.* proposes a cloud and edge computing framework

with integration of docker container and Kubernetes. They deploy a model for machine learning (Inception V3) in the edge computing platform. The authors conducted an experiment to prove the feasibility of their design and idea [15] .

III. OBJECTIVE AND METHODOLOGY

The primary objective of the proposed experiment is to deploy a containerized orchestration microservice for scalable computation related to greenhouse applications in the edge computing environment. Figure 1 shows the experimental microservice deployment arrangement in Docker and Kubernetes. The key performance parameters of microservice execution are evaluated against various user load conditions. The experimental setup was implemented on an edge server equipped with an Intel Core i3 processor clocked at 2.20 GHz and 8 GB of RAM. The software tools and configurations used included Docker Desktop version 4.39.0 for containerization, Kubernetes CLI (*kubectl*) for orchestration management, Eclipse IDE version 1.34.0, Apache Tomcat 9.0, Maven 3.9.9 for microservice development, and Apache JMeter 5.6.3 for performance testing.This configuration provided a controlled environment for evaluating the microservices while simulating realistic computational loads relevant to greenhouse applications. In the proposed study, a microservice has been developed to compute the equilibrium temperature inside a greenhouse. The equilibrium temperature is estimated on the basis of energy balance approach that considers the balance between incoming solar radiation and outgoing thermal radiation. The equation for equilibrium temperature is shown in Equation 1.

$$T = \left(\frac{S.(1-a)}{4.\sigma.\varepsilon}\right)0.25 \quad [1]$$

Here, T is the equilibrium temperature of the greenhouse (Kelvin),Sis the incoming solar radiation flux (W/m²),ais the albedo or reflectivity of the greenhouse surface (dimensionless, fraction of solar energyreflected), σ presents Stefan Boltzmann constant (5.67×10^{-8} $Wm^{-2}K^{-4}$), ε is the emissivity of the greenhouse surface (dimensionless, accounts for radiative heat loss). In Equation 1, the greenhouse receives solar radiation, of which a fraction is reflected and the rest is absorbed. Hence, the effective absorbed energy is S·(1−a). The factor of 4 in the denominator accounts for the redistribution of the incoming solar energy over the total surface area of the greenhouse.

The greenhouse, in turn, emits heat as longwave radiation according to the Stefan–Boltzmann law, which is proportional to $\sigma\varepsilon T^4$. At thermal equilibrium, the absorbed solar energy equals the emitted radiation, giving rise to the above expression for the greenhouse temperature. In the proposed study, the Docker is employed for containerization of the developed microservice, enabling the encapsulation of the microservice and its dependencies into a portable and lightweight container. Kubernetes is used for orchestration, providing automated deployment, scaling of application and management of the containerized microservices. The computational task selected for the experiment involves determining the equilibrium temperature of soil inside a greenhouse, a process that can benefit from the scalable and distributed computation capabilities of a microservice-based architecture.

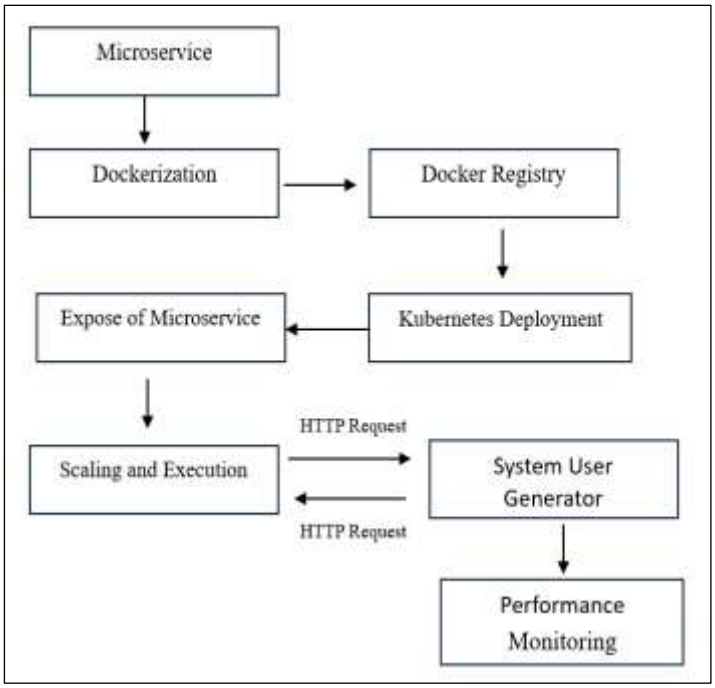


Figure 1: Architecture of experimental microservice deployment arrangement in Docker and Kubernetes.
Source: Authors, (2025).

To evaluate the performance and responsiveness of the deployed architecture, three key performance metrics were considered: (a) RT is a metric that calculates how long it takes to issue a request and get a response; (b) Th, which shows how many requests are handled in a certain amount of time; and (c) HpS, representing the frequency of incoming requests handled by the system. The above mentioned metrics were selected to provide a comprehensive evaluation of both the efficiency and scalability of the deployed microservice under varying load conditions.

IV. EVALUATION OF OVERALL ASSESSMENT

Apache JMeter is used to monitor microservice execution. The metrics considered provides a comprehensive view of the system's behaviour under different load conditions. In the experimental setup, the load on the system was progressively increased by simulating concurrent users (threads) starting from 100 threads, with increments of 100 threads per iteration, continuing up to 3000 threads, or until the system became non-responsive. The load testing was performed using a performance testing tool configured to mimic real-world usage patterns by sending consecutive HTTP requests to the Kubernetes-managed microservices.

Each test iteration involved the capture of the aforementioned performance metrics across all active microservices. The system exhibited stable performance up to 3600 concurrent threads. However, at 3700 threads, the local Kubernetes cluster on which the application was hosted began to fail. The cluster stopped responding and exhibited behaviour indicative of resource saturation or failure in orchestrating container pods due to excessive load. The performance degradation observed beyond this point demonstrates the limitations of the local execution environment, highlighting the need for either hardware resource enhancement or migration to a more scalable production-grade Kubernetes cluster. Table 1 presents variations in performance metrics with increasing trend of concurrent users from 100, 200, 300, 400, and up to 3600 users until the system begins to fail.

Table 1: Variations of performance metrics with increasing trend of concurrent users.

S/N	Nos. of Active Threads	Th	HpS	RT
1	100	238174	1242	121.077
2	200	547243	2849	71.540
3	300	450533	2350	128.667
4	400	10996132	4366	12.516
5	500	10973484	4211	1.418
6	600	9552119	3887	72.843
7	700	10261131	4438	100.643
8	800	10113415	4171	89.645
9	900	11203628	4454	18.429
10	1000	10674796	4233	65.864
11	1100	10364628	4137	74.580
12	1200	11700000	4490	1.771
13	1300	11200000	4321	8.957
14	1400	11700000	4777	132.535
15	1500	11700000	4649	124.609
16	1600	10200000	3986	48.365
17	1700	11900000	4573	235.226
18	1800	10500000	4146	288.839
19	1900	9770220	4099	26.853
20	2000	10929928	4189	35.462
21	2100	9739428	3937	197.728
22	2200	9966200	4258	91.593
23	2300	10900000	4515	79.752
24	2400	9834536	4007	222.614
25	2500	10500000	4066	296.796
26	2600	12300000	4756	108.855
27	2700	9966948	4034	739.483
28	2800	11100000	4081	25.784
29	2900	8617092	3895	1765.557
30	3000	8437656	4084	558.533
31	3200	7159768	1908	120.252
32	3300	5259496	1355	323.741
33	3400	7746748	3069	307.188
34	3500	5663908	2284	747.250
35	3600	2383128	1085	293.261
34	3500	5663908	2284	747.250
35	3600	2383128	1085	293.261

Source: Authors, (2025).

Figure 2 illustrates the comparative behavior of three critical performance metrics including RT, Th, and HpS with increasing numbers of users (active threads). The RT remains relatively stable with only minor fluctuations at low user loads. However, as the number of active threads increases around 2500 - 3500, the RT sharply increases with noticeable spikes. This behavior indicates that as the system approaches saturation, it experiences resource contention and queuing delays, resulting in degraded responsiveness. In the case of Th and HpS increase sharply with the initial increase in active threads, reaching a relatively steady state in the mid-range of user loads up to ~2400 threads. Eventually, Th and HpS show a downward trend as user count grows further, suggesting that the system cannot handle additional requests efficiently once critical resource limits are approached. Overall, the comparative analysis reveals that while the system initially scales well with increasing user load, its performance metrics deteriorate beyond a certain threshold, marking the saturation point. This behaviour is crucial for understanding the system’s scalability, capacity planning, and performance

bottlenecks. In this experiment the configuration used in Apache Jmeter are – 100 nos. of threads, 30 seconds as ramp-up period, loop to be continued for 60 seconds. Active threads over time, RT, Th and HpS are the performance metric that has been used while executing microservice. 60 numbers of sample data generated by Jmeter are considered for evaluation of performance. The experimental results for each performance metric are graphically represented to illustrate the system's behaviour under load. Figure3, Figure 4, Figure 5 and Figure 6 presents the variations of increasing active users (threads) with time, RT with respect to increasing concurrent users, Th (requests per second processed successfully) with time and HpS - the total number of requests received by the system per second respectively.

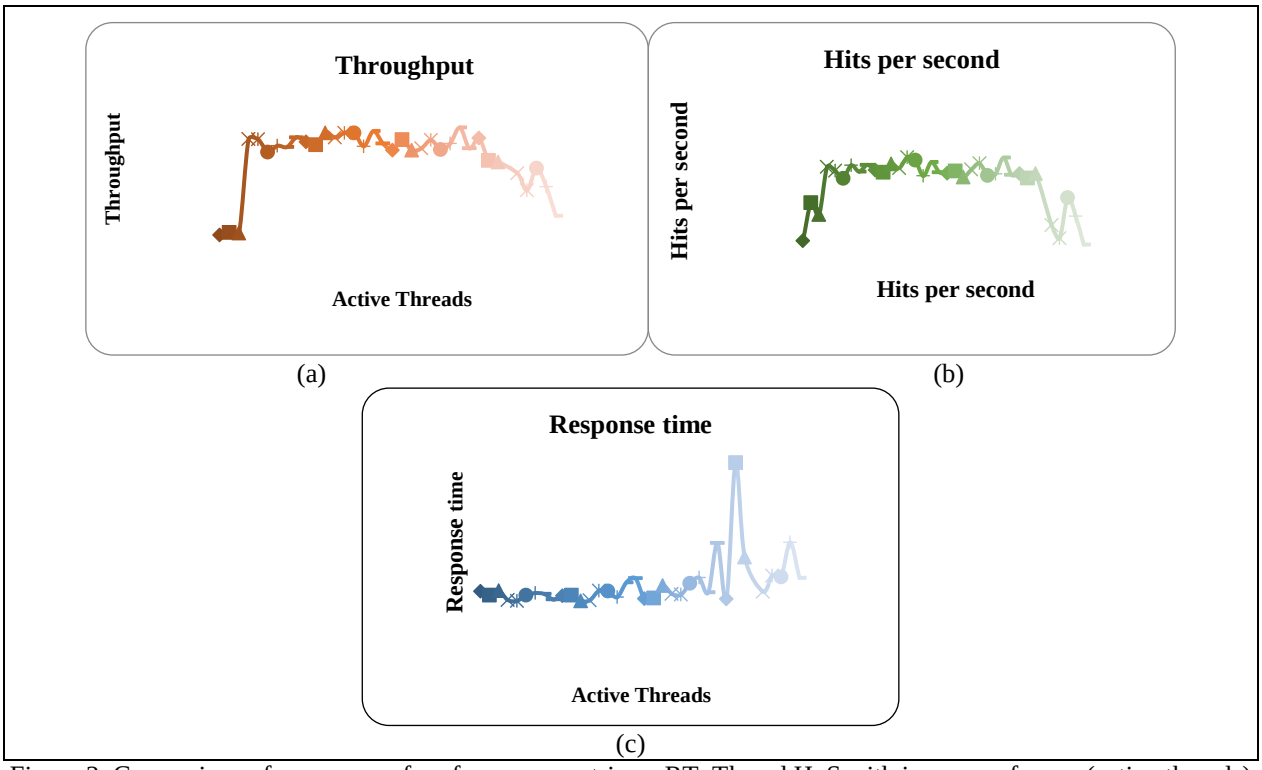


Figure 2. Comparison of responses of performance metrics – RT, Th and HpS with increase of users (active threads)
Soucer: Authors, (2025).

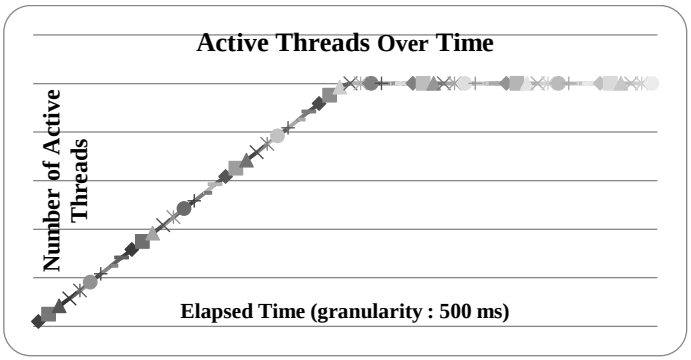


Figure 3: variations of increasing Active users (threads) with time.
Soucer: Authors, (2025).

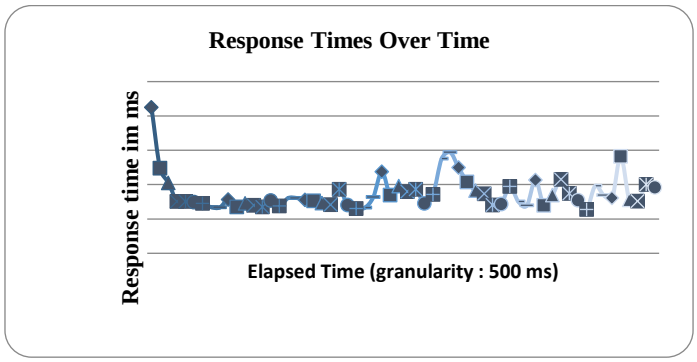


Figure 4: RT over time with respect to increasing concurrent users.
Soucer: Authors, (2025).

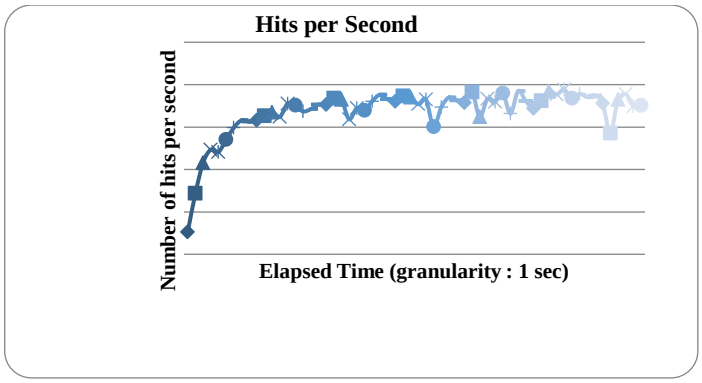


Figure 5: HpS with respect to increasing concurrent users and time
Soucer: Authors, (2025).

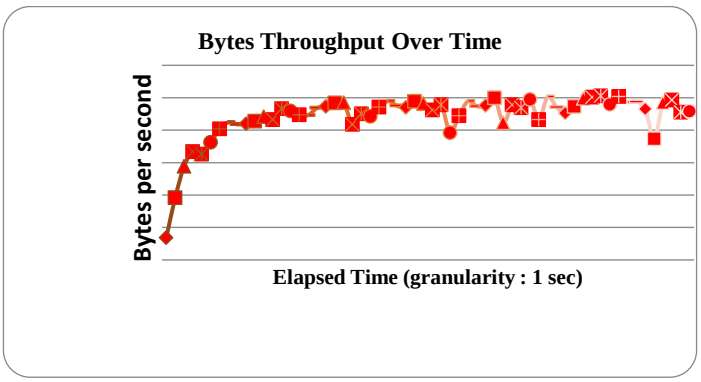


Figure 6: Th over time.
Soucer: Authors, (2025).

These figures collectively provide insights into how the system scales and where its performance bottlenecks begin to manifest. The results serve as a basis for performance tuning and infrastructure planning for high-load microservice deployments in Kubernetes environments.

Figure 7: Relations among the performance metrics including active users, RT, Th and HpS.
Soucer: Authors, (2025).

The correlation among the performance metrices for 100 consecutive request is shown in Figure 7. 60 records are chosen as a data sample for a thread group of 100 users. Table 2 displays the evaluation of the descriptive statistics for the data sample.

Table 2: Descriptive statistics of Th, HpS and RT.

Parameters	Th	HpS	RT
Mean	8850369.733	3367.916667	0.87815331
Standard Deviation	1493271.115	579.4468225	0.25459349
Median	9345600	3547	0.792436217

Soucer: Authors, (2025).

To visually evaluate the distribution nature of the recorded data sample with respect to RT, Th, and HpS, we employ a normal probability plot. Figures 7, 8, and 9 display the normality plots for RT, Th, and HpS respectively.

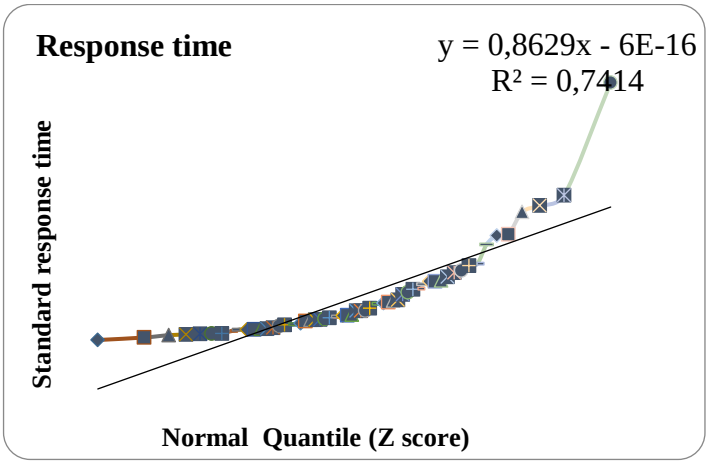


Figure 7: Plot of normality for RT (msec).
Soucer: Authors, (2025).

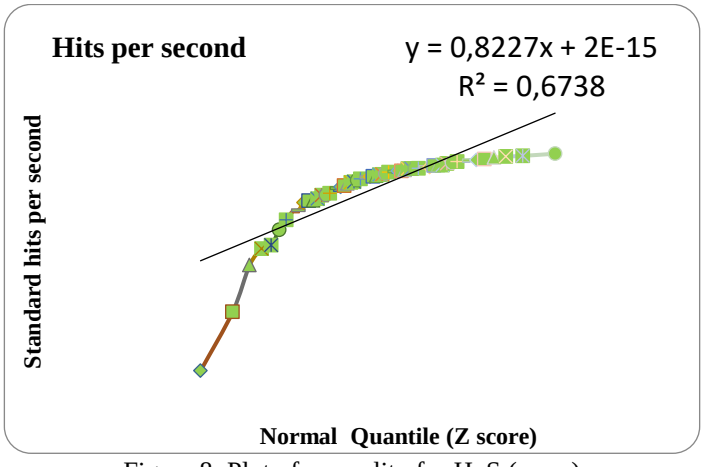


Figure 8: Plot of normality for HpS (msec)
Soucer: Authors, (2025).

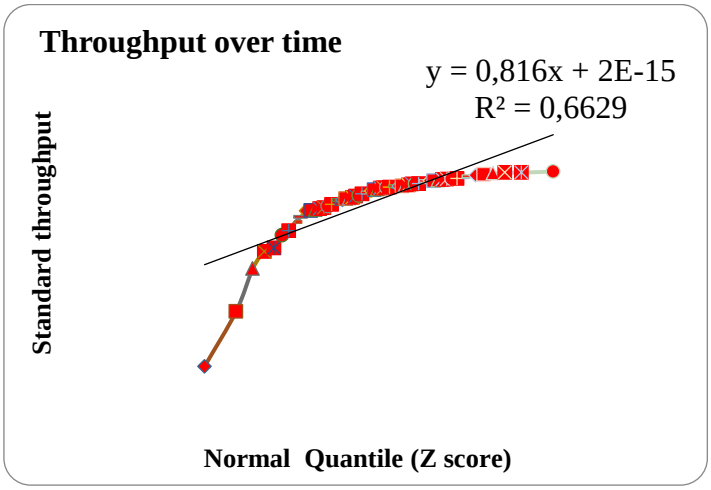


Figure 9: Plot of normality for Th (msec).
Soucer: Authors, (2025).

The normal probability plot revealed a reasonably linear trend with a regression equation y=0.862x, R^2= 0.741 for RT, y=0.822x, R^2= 0.673 for HpS and y=0.816x, R^2= 0.662 for Th. This indicates that the RT data approximately follows a normal distribution, although some deviations exist, likely due to skewness or variability at higher load conditions.

To observe the Goodness of Fit, we evaluate the performance parameters with Anderson Darling (AD) Test. This test is a statistical method used to evaluate whether a given dataset follows a specific probability distribution. It is a type of goodness-of-fit test, to measures how well the observed data align with the theoretical distribution being tested. To further analysis the distribution of data set for RT, Th and HpS, we evaluate through AD test as shown in Equation 2 [14]. The hypothesis H_0 is set that the data of all the performance parameter follows a normal distribution.

$$AnD = N - \left(\frac{1}{N}\right)\sum_{i=1}^{N}(2i-1)[(ln(F(Y_i)) + ln(1 - F(Y_{N+1-i}))] \quad [2]$$

where, N is the number of samples, F() is the Cumulative Distribution Function, Y_i is the i^{th} sample in ascending order, Y_{N+1-i} is the i^{th} sample in descending order. Since the AD statistic depends on the sample size, a modified form is often used to improve accuracy for small samples. It is shown in Equation 3.

$$AnD^* = (AnD) * (1 + 0.75/N + 2.25/N) \quad [3]$$

From the computed AD statistic, a p-value is obtained. p-value is the probability of sample data that is used to check how well the sample follows as specific probability distribution. We have calculated the p-value of AD test using Equation 2 and 3, respectively. The hypothesis of AD test will be rejected if the p-value is smaller than α, the significance level and fail to reject if p-value is higher than than α. Here in this experiment, we use significance level (α) as 0.05 (5%).The AD normality test was applied to the datasets taken for the experiment corresponding to RT, Th, and HpS. The computed p-values for RT, Th, and HpS are found to be 4.521, 8.715, and 1.090 respectively. Since these p-values are significantly higher than the commonly accepted significance threshold (α = 0.05), the test provides strong evidence against the null hypothesis. As such, it can be included that the null hypothesis that the data follows a normal distribution is rejected. This indicates that the distributions of the observed performance metrics deviate from normality. This may be due to workload variations and system resource constraints at higher user loads.

V. CONCLUSION

This study demonstrates the deployment of microservices in Docker and Kubernetes for scalable computation within an edge computing environment. In addition, a performance evaluation of the deployed microservice applications was carried out, focusing on three key metrics including RT, Th, and HpS. The evaluation was conducted using Apache JMeter, which generated test samples to simulate varying user loads and operational conditions. The experimental results offer important insights on how microservice running in Docker and Kubernetes behave. Specifically, the results highlight how the system responds under different levels of concurrency and high user loads. By analyzing the collected performance data, it was possible to identify performance trends, detect bottlenecks, and determine the scalability limits of the deployed architecture. These findings contribute to a deeper understanding of microservice performance in containerized and orchestrated environments. Moreover, the outcomes serve as a reference for optimizing microservice architectures, fine-tuning Kubernetes resource allocation, and guiding future improvements in system design. Ultimately, this ensures higher efficiency, reliability, and scalability in edge computing deployments.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Pranjit Kakati. Abhijit Bora
Methodology: Pranjit Kakati. Abhijit Bora.
Investigation: Pranjit Kakati. Abhijit Bora.
Discussion of results: Pranjit Kakati. Abhijit Bora.
Writing – Original Draft: Pranjit Kakati. Abhijit Bora.

Writing – Review and Editing: Pranjit Kakati. Abhijit Bora.
Resources: Pranjit Kakati. Abhijit Bora.
Supervision: Pranjit Kakati. Abhijit Bora.
Approval of the final text: Pranjit Kakati. Abhijit Bora

VII. REFERENCES

[1] Md. D. Hossain et al., "The role of microservice approach in edge computing: Opportunities, challenges, and research directions," ICT Express, Jun. 2023, doi: 10.1016/j.icte.2023.06.006.

[2] K. Cao, Y. Liu, G. Meng, and Q. Sun, "An Overview on Edge Computing Research," 2020, Institute of Electrical and Electronics Engineers Inc.doi: 10.1109/ACCESS.2020.2991734.

[3] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge Computing: Vision and Challenges," IEEE Internet Things J, vol. 3, no. 5, pp. 637–646, Oct. 2016, doi: 10.1109/JIOT.2016.2579198.

[4] F. Liu, G. Tang, Y. Li, Z. Cai, X. Zhang, and T. Zhou, "A Survey on Edge Computing Systems and Tools," Proceedings of the IEEE, 2019, doi: 10.1109/JPROC.2019.2920341.

[5] P. Sha, S. Chen, L. Zheng, X. Liu, H. Tang, and Y. Li, "Design and Implement of Microservice System for Edge Computing," in IFAC-PapersOnLine, Elsevier B.V., 2020, pp. 507–511. doi: 10.1016/j.ifacol.2021.04.137.

[6] Q. Qu, R. Xu, S. Y. Nikouei, and Y. Chen, "An Experimental Study on Microservices based Edge Computing Platforms," Apr. 2020, [Online]. Available: http://arxiv.org/abs/2004.02372

[7] O. Aitlmoudden, M. Housni, N. Safeh, and A. Namir, "A Microservices-based Framework for Scalable Data Analysis in Agriculture with IoT Integration," International Journal of Interactive Mobile Technologies, vol. 17, no. 19, pp. 147–156, 2023, doi: 10.3991/ijim.v17i19.40457.

[8] 2020 International Conference on Computer Communication and Informatics (ICCCI). IEEE, 2020.

[9] Y. Huang, R. Zong, K. Cai, and Y. Mao, "Design and implementation of an edge computing platform architecture using docker and kubernetes for machine learning," in ACM International Conference Proceeding Series, Association for Computing Machinery, Mar. 2019, pp. 29–32. doi: 10.1145/3318265.3318288.

[10] L. Abdollahi Vayghan, M. A. Saied, M. Toeroe, and F. Khendek, "Deploying Microservice Based Applications with Kubernetes: Experiments and Lessons Learned," in IEEE International Conference on Cloud Computing, CLOUD, IEEE Computer Society, Sep. 2018, pp. 970–973. doi: 10.1109/CLOUD.2018.00148.

[11] F. Rossi, V. Cardellini, and F. Lo Presti, "Hierarchical scaling of microservices in Kubernetes," in Proceedings - 2020 IEEE International Conference on Autonomic Computing and Self-Organizing Systems, ACSOS 2020, Institute of Electrical and Electronics Engineers Inc., Aug. 2020, pp. 28–37. doi: 10.1109/ACSOS49614.2020.00023.

[12] M. Amaral, J. Polo, D. Carrera, I. Mohomed, M. Unuvar, and M. Steinder, "Performance evaluation of microservices architectures using containers," in Proceedings - 2015 IEEE 14th International Symposium on Network Computing and Applications, NCA 2015, Institute of Electrical and Electronics Engineers Inc., Jan. 2016, pp. 27–34. doi: 10.1109/NCA.2015.49.

[13] I. Čilić, P. Krivić, I. Podnar Žarko, and M. Kušek, "Performance Evaluation of Container Orchestration Tools in Edge Computing Environments," Sensors, vol. 23, no. 8, Apr. 2023, doi: 10.3390/s23084008.

[14] S. Medhi, A. Bora, and T. Bezboruah, "Investigations On Some Aspects of Reliability of Content Based Routing SOAP based Windows Communication Foundation Services," International Journal of Information Retrieval Research, vol. 7, no. 1, pp. 17–31, Jan. 2017, doi: 10.4018/ijirr.2017010102.

[15] Y. Huang, R. Zong, K. Cai, and Y. Mao, "Design and implementation of an edge computing platform architecture using docker and kubernetes for machine learning," in ACM International Conference Proceeding Series, Association for Computing Machinery, Mar. 2019, pp. 29–32. doi: 10.1145/3318265.3318288.

Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.11 n.55, p. 276-292. September/October, 2025.

DOI: https://doi.org/10.5935/jetia. v.11 n.55.2713

RESEARCH ARTICLE OPEN ACCESS

COMBINED EFFECT OF GLASS POWDER AND RECYCLED SAND ON FRESH AND HARDENED PROPERTIES OF STEEL FIBER-REINFORCED SELF-COMPACTING CONCRETE

Souhila Fergani[1], Said Zaouai[2], Rachid Rabehi[3] and Ali Abbache[4]

[1, 3, 4] Faculty of Civil Engineering, University of Science and Technology Houari Boumediene (USTHB), Bab Ezzouar, 16111 Algiers, Algeria.
[2] Laboratory of Eco-Materials: Innovations & Applications (EMIA), Civil Engineering Department, University of Djelfa, 17000, Algeria.

[1]https://orcid.org/0009-0005-7086-8141, [2]https://orcid.org/0009-0009-5936-1074
[3]https://orcid.org/0000-0002-9398-7291, [4]https://orcid.org/0009-0002-6257-4336

Email: sfergani@usthb.dz, said.zaouai@univ-djelfa.dz, rachid.rabehi@usthb.edu.dz, ali.abbache@usthb.edu.dz

ARTICLE INFO

Article History
Received: September 19, 2025
Revised: October 3, 2025
Accepted: October 8, 2025
Published: October 31, 2025

Keywords:
Self-compacting concrete,
Waste glass powder,
Recycled sand,
Steel fibers,
Response surface methodology.

ABSTRACT

This study investigates the synergistic effects of waste glass powder and recycled sand on steel fiber-reinforced self-compacting concrete through systematic mixture optimization. Waste glass powder (WGP) and recycled sand (RS) from construction demolition waste were evaluated as sustainable replacements for cement and natural aggregates. The experimental program examined WGP replacement levels of 5-20%, RS substitution of 15-60%, and hooked-end steel fiber incorporation at 0.3-1.2% by volume, maintaining water-to-binder ratio of 0.42 with optimized superplasticizer dosage. Fresh properties were assessed through standardized workability tests including slump flow, V-funnel, and L-box evaluations, while hardened performance encompassed compressive and flexural strength development alongside durability parameters. Design-Expert 13 software facilitated response surface methodology optimization, revealing complex material interactions that challenge conventional additive assumptions. The optimal sustainable formulation comprising 15% waste glass powder, 15% recycled sand, and 0.9% steel fibers achieved balanced performance characteristics. This mixture demonstrated compressive strengths of 53.4 MPa at 28 days, progressing to 64.8 MPa at 90 days, enhanced flexural capacity of 8.2 MPa, and reduced water absorption from 4.2% to 3.8%. Fresh properties remained within acceptable self-compacting concrete criteria despite substantial waste material incorporation. The research demonstrates that waste-derived materials enhance concrete performance when integrated through systematic design approaches, advancing sustainable construction technologies.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Concrete has served as the backbone of construction since its earliest use in ancient civilizations, where it enabled the development of durable and enduring structures. Over the centuries, it has evolved into one of the most essential building materials worldwide, underpinning advancements in infrastructure and architectural design. Its significance lies in a rare combination of properties: high compressive strength, adaptability to diverse applications, and cost-effectiveness. Capable of being molded into almost any form while ensuring long-term durability, concrete remains indispensable across all sectors of construction and public works—from monumental dams and extensive highway networks to bridges, tunnels, and modern high-rise buildings.

The foundational development of Portland cement during the 19th century marked a turning point, granting concrete unprecedented consistency and strength. Since then, global concrete production has soared, now exceeding 10 billion tonnes annually. This pervasive use, while instrumental in economic growth and societal development, comes with critical environmental consequences. Cement manufacturing alone contributes around 7–8% of global anthropogenic CO_2 emissions [1], stemming from both the energy-intensive firing

process and chemical decarbonation of limestone. If current trajectories continue, global concrete demands are projected to approach 20 billion m³ by 2050, potentially doubling the associated carbon footprint unless mitigated [2]. In addition to emissions, material consumption poses sustainability challenges. The relentless extraction of natural sand and gravel for concrete depletes riverbeds and coastal areas, threatening ecosystems and undermining resource availability for future generations [3]. Despite this, no material currently matches concrete's aggregate performance, resilience, or cost-effectiveness for large-scale infrastructure. Thus, the pathway forward is not to replace concrete, but to reimagine it—making it greener and more circular [4].

Self-compacting concrete (SCC) represents such an evolution. Pioneered in Japan in the late 1980s, SCC flows under its own weight into formworks, even in fully reinforced zones, eliminating the need for vibration. This innovation streamlines construction, improves surface finish, enhances safety, and increases productivity [5]. However, SCC typically requires high binder content and chemical admixtures, diminishing its potential sustainability benefits and imposing economic constraints [6]. To address these limitations, the research community has turned toward sustainable supplementary materials and recycled aggregates that can partially replace key constituents of SCC [7], [8]. Among these recycled materials are those that replace coarse and/or fine aggregates with various other materials to improve or maintain their physical mechanical properties. These include plastic aggregates, recycled concrete aggregates, and rubber aggregates. Other materials that replace cement are limestone powder, pozzolans, silica fume, ground sand, metakaolin, and glass powder.

Waste glass powder (WGP) is increasingly recognized as an effective supplementary cementitious material (SCM). When finely ground (typically <75 μm), glass powder reacts pozzolanically with calcium hydroxide to form additional calcium silicate hydrate (C–S–H), thus enhancing matrix density and durability while mitigating alkali-silica reaction risks common in coarser glass additions [9–11]. For glass powder with a diameter of less than 75 μm the pozzolanic reaction is largely privileged over ASR [12]. WGP inclusion (10–30% replacement) has demonstrated reductions in permeability, shrinkage, and chloride ingress, with many studies confirming that SCC's fresh and hardened properties remain satisfactory or even improved [13], [14].

On the other hand, the use of recycled sand from construction and demolition waste offers a sustainable alternative to depleting natural sand reserves. Although recycled sand has higher water absorption and angularity, judicious mix design can enable up to 50% replacement without compromising SCC's performance [15]. Such substitutions also support local resource reuse, reducing both material transport impacts and landfill burdens [16]. Additionally, studies have shown that recycled aggregates can be successfully used in percentages of up to 50 wt.% with water absorption values below 6.0% [17].

Despite these two areas receiving attention separately, few studies have pursued a combined assessment of WGP, recycled sand, and steel fiber reinforcement in SCC. The inclusion of steel fibers further addresses concrete's brittleness, enhancing flexural toughness, post-cracking ductility, and impact resistance. For instance, even at low volume fractions, fibers significantly influence mechanical behavior and durability, though they can reduce fresh workability if not properly dosed [18], [19].

While binary combinations—such as WGP with fibers or recycled materials with fibers—have been explored, the simultaneous integration of all three sustainable constituents in SCC remains largely underinvestigated. Recent studies have demonstrated that the combined use of recycled aggregates and waste glass powder can enhance both mechanical performance and durability while contributing to resource efficiency [20]. Similarly, research on fiber-reinforced SCC highlights the need for mixture designs that comply with EFNARC criteria to maintain adequate workability while achieving improved toughness and ductility [21]. These findings underscore the necessity of a holistic approach that considers the coupled effects of WGP, RS, and SF to develop optimized, sustainable SCC mixtures for structural and infrastructure applications.

Moreover, real-world deployment necessitates evidence of practicality. SCC formulations that leverage locally available recycled materials could drastically reduce embodied carbon, transportation emissions, and waste management costs while bolstering structural resilience. Such approaches align with circular economy goals and regulatory pushes for low-carbon construction [22], with recent life cycle assessment studies revealing that incorporating recycled supplementary cementitious materials as cement substitutes can achieve substantial embodied carbon reductions—up to 56% for ground granulated blast-furnace slag, 21% for calcined marine clay, and 16% for waste glass powder—while maintaining structural performance [23].

This This study addresses a key research gap by evaluating self-compacting concrete incorporating waste glass powder as a partial cement replacement, recycled sand as fine aggregate, and steel fiber reinforcement. The work investigates fresh properties (flowability, passing ability, segregation resistance), mechanical performance (compressive and flexural strength), and durability through water absorption tests. Synergistic interactions among these constituents were analyzed using response surface methodology to optimize mixture proportions that preserve self-compacting behavior while enhancing mechanical response. The findings provide a practical framework for implementing waste-derived materials in SCC without compromising structural integrity.

II. MATERIALS AND METHODS

II.1 MATERIALS

II.1.1 Cement

Portland cement conforming to CPA CEM I/52.5 specification was employed throughout this study. The chemical composition, mineralogical characteristics, and physical properties of the cement are presented in Tables 1, 2, and 3, respectively. The specific surface area was determined using the Blaine method, while chemical analysis was conducted through X-ray fluorescence spectroscopy (XRF). Mineralogical composition was calculated using Bogue's equations based on the chemical analysis results [24].

Table 1: Chemical composition of cement used (%).

SiO_2	CaO	Al_2O_3	Fe_2O_3	SO_3	MgO	K_2O	SO_3	Na_2O	LOI
20,87	62,36	5,1	3,42	2,92	1,88	0,68	2,92	0,11	1,8

Source: Authors, (2025).

Table 2: Mineralogical composition of cement used (%).

Mineralogical composition (%)			
C_3S	**C_3A**	**C_2S**	**C_4AF**
62,36	07,06	14,85	12,14

Source: Authors, (2025).

Table 3: Physical properties of cement used.

Properties	**CEM I/B**
Specific density (kg/m^3)	3096
Apparent density (kg/m^3)	1026
Fineness (cm^2/g)	3912
Initial setting time (min)	185
Final setting time (min)	245

Source: Authors, (2025).

II.1.2 Sand

Natural alluvial sand with particle size distribution of 0/5 mm was used as fine aggregate. The sand was sourced from the Oued M'zi region, located north of Laghouat city (Algeria). Prior to use, the sand was washed and dried to remove organic impurities and ensure consistent moisture content. The sand exhibited a fineness modulus of 2.28, classifying it as medium sand according to established grading standards. Physical properties including specific gravity, water absorption, and particle size distribution were determined in accordance with relevant ASTM standards. Table 4 shows the properties of the natural sand used, and Figure 1 presents its visual appearance.

Figure 1: Natural alluvial sand (0–5 mm) used as reference fine aggregate.
Source: Authors, (2025).

Recycled sand was produced through controlled crushing and processing of construction and demolition waste concrete. The source concrete specimens, measuring 0.40 × 0.40 × 0.15 m³, were initially prepared with a cement content of 400 kg/m³ using natural alluvial sand and quartz-based aggregates. After a standard 28-day water-curing period, the hardened concrete was mechanically crushed using a jaw crusher, followed by secondary impact crushing to achieve the desired particle size distribution. The crushed material was subjected to systematic screening to separate the recycled sand fractions with a 0–5 mm particle size distribution, comparable to that of the reference natural sand in accordance with ASTM C33 [25] and EN 12620 [26]. Due to its higher water absorption capacity and modified surface texture compared to natural sand, the recycled sand required careful mix design adjustments to maintain the workability of the self-compacting concrete matrix. Physical and mechanical properties of recycled sand were determined according to standard procedures: specific gravity and water absorption following ASTM C128 [27] (equivalent to EN 1097-6) [28]; fineness modulus by sieve analysis according to ASTM C136 [29] (equivalent to EN 933-1) [30]; and bulk density (apparent volume mass) in accordance with ASTM C29 [31] (equivalent to EN 1097-3) [32]. The results are presented in Table 4, and Figure 2 shows the recycled sand obtained from this process.

Figure 2: Recycled sand (0–5 mm) produced from processed construction and demolition waste concrete.
Source: Authors, (2025).

II.1.3 Gravel

The coarse aggregate used in this study was crushed limestone gravel sourced from the "Zaccar" quarry, south of Djelfa. Two size fractions, 3/8 mm and 8/16 mm, were utilized. As presented in Table 4, both fractions exhibited similar absolute densities (2.66 g/cm³) and low water absorption values (2.3% and 2.2%), indicating limited porosity and good dimensional stability. The Los Angeles abrasion coefficients (25.3% and 25.1%) confirm adequate mechanical resistance. The angular morphology of the particles, resulting from the crushing process, is expected to improve aggregate interlocking and enhance the mechanical performance of the concrete.

Table 4: Physical-mechanical properties of gravel and different types of sand.

Property	**Sand**		**Gravel**	
	Natural	**Recycled**	**Natural**	
	$(0/5)_N$	$(0/5)_R$	$(3/8)_N$	$(8/16)_N$
Apparent volume mass (g/cm^3)	1.52	1.22	1.34	1.39
Absolute density mass (g/cm^3)	2.63	2.49	2.66	2.66
Degree of absorption (%)	1.14	7.7	2.3	2.2
Fineness modulus	2.29	2.23	–	–
Sand equivalent (%)	87.9	–	–	–
Coefficient Los Angeles (%)	–	–	25.3	25.1

Source: Authors, (2025).

II.1.4 Steel Fibers (SF)

MEDAFAC steel fibers with undulated geometry were incorporated into the self-compacting concrete mixtures to enhance mechanical performance and post-cracking behavior. The fibers feature a wavy configuration along their length, providing enhanced mechanical anchorage within the concrete matrix compared to straight fibers.

The physical and geometric properties of the MEDAFAC steel fibers are presented in Table 5. The fibers have a length of 50 mm, width of 2 mm, and density of 7.6 g/cm³. The undulated geometry improves bond characteristics and load transfer capacity through enhanced mechanical interlocking with the surrounding cement matrix.

Table 5: Physical properties of MEDAFAC steel fibers.

Property		**Figure 3 :** Steel fiber used
Absolute density (g /cm^3)	7.60	
Tensile strength	1200 MPa	
Geometry of steel fiber used	D_f = 2 mm L_f = 50 mm Undulated	
Aspect ratio (L_f/D_f)	25	

Source: Authors, (2025).

II.1.5 Mixing Water

Potable tap water meeting drinking water standards was used for all concrete mixtures. The water was maintained at a temperature of 20 ± 1°C during mixing operations to ensure consistent hydration kinetics and workability characteristics across all experimental batches.

II.1.6 Superplasticizer

A third-generation superplasticizer based on polycarboxylate ether technology (MEDAPLAST SP 40) was employed to achieve the required flow characteristics for self-compacting concrete. The admixture exhibited the following properties: pH of 8.2, specific gravity of 1.20 ± 0.01, chloride ion content of <1 g/L, and solid content of 40%. The superplasticizer dosage was optimized through fresh property trials to maintain consistent performance across all mix variations while ensuring compatibility with the sand recycled materials.

II.1.7 Waste Glass Powder (WGP)

Waste glass powder was produced from post-consumer clear glass bottles collected from local recycling facilities. The glass was initially sorted and cleaned to remove labels, caps, and other contaminants. Following cleaning, the glass underwent a two-stage crushing process: primary crushing using a jaw crusher to reduce particle size to approximately 5 mm, followed by ball milling to achieve the target fineness. The grinding process was controlled to ensure 95% of particles passed through a 75 µm sieve, consistent with established recommendations for pozzolanic reactivity optimization [33], [34]. This fine particle size promotes the pozzolanic reaction between the amorphous silica in glass powder and calcium hydroxide produced during cement hydration, while minimizing the risk of deleterious alkali-silica reaction. The detailed chemical composition of the waste glass powder is summarized in Table 6, indicating that the material

contained approximately 72.6% SiO_2, 13.24% Na_2O, 10.42% CaO, and trace amounts of other oxides. The high silica content and amorphous structure contribute to the pozzolanic properties of the material. Physical properties including specific gravity (2.62 g/cm³) and particle size distribution were characterized using standard laboratory procedures. The processed glass powder was stored in sealed containers to prevent moisture absorption and maintain consistent properties throughout the experimental program. Prior to use, the material was verified to be free from organic contaminants and exhibited uniform color and texture characteristics. Figure 4 shows the Waste Glass Powder used.

Table 6: Chemical composition of waste glass powder (WGP).

Material	SiO_2	Fe_2O_3	Al_2O_3	CaO	TiO_2	MgO	Na_2O	K_2O	LOI
Glass Powder	72.6	0.4	1.5	10.42	0.04	1.2	13.24	0.5	0.55

Source: Authors, (2025).

Figure 4: Waste glass powder (≤75 μm).
Source: Authors, (2025).

III. EXPERIMENTAL PROGRAM

III.1 MIXTURE DESIGN APPROACH

The experimental investigation employed an I-optimal mixture design to systematically evaluate the combined effects of waste glass powder (WGP), recycled sand (RS), and steel fibers (SF) on the fresh and hardened properties of self-compacting concrete. This design methodology was selected based on its superior predictive capability and efficiency in capturing complex interactions among mixture constituents while minimizing prediction variance across the experimental domain. Unlike conventional factorial designs, mixture designs inherently account for the dependency among components where the sum of constituent proportions equals unity, mathematically expressed as:

$$\sum_{i=1}^{n} X_i = 1 \quad (1)$$

Where Xi represents the proportion of the i-th variable component [35]. Unlike conventional factorial designs that treat factors independently, mixture designs inherently account for the constraint that variable proportions must sum to unity, providing a more realistic framework for concrete mixture optimization within the defined experimental space.

The I-optimal criterion was specifically chosen over D-optimal or A-optimal alternatives due to its focus on maximizing prediction accuracy rather than parameter estimation precision or average prediction variance. This approach distributes design points optimally across the constrained mixture space, providing balanced information for model estimation while ensuring efficient exploration of both main effects and interaction terms [36]. The criterion is particularly advantageous when the primary objective is developing predictive models for practical applications rather than purely statistical inference.

The experimental framework maintained several parameters constant to isolate the effects of the three studied variables: total binder content at 400 kg/m³, water-to-binder ratio of 0.42, and superplasticizer dosage at 2.45% by cement mass. This approach enabled clear attribution of performance variations to the specific influence of WGP, RS, and SF combinations.

III.2 EXPERIMENTAL MATRIX AND FACTOR RANGES

The experimental program comprised forty-four concrete mixtures generated using Design-Expert®13 software (Stat-Ease, Inc.), incorporating both unique factorial combinations and strategically placed replicate points for pure error estimation. Factor levels were established through comprehensive literature review and preliminary laboratory trials to ensure both scientific rigor and practical feasibility for SCC production [37].

The selected variable ranges were defined as follows:

Waste Glass Powder (WGP): Cement replacement levels of 0%, 5%, 10%, 15%, and 20% by mass. This range reflects optimal balance between pozzolanic enhancement and potential early-age strength reduction. Literature evidence indicates that finely ground

glass powder (< 75 μm) exhibits significant pozzolanic activity while minimizing alkali-silica reaction risks at replacement levels up to 20% [38], [39]. Higher replacement ratios were avoided to prevent excessive early-age strength loss and delayed expansion concerns.

Recycled Sand (RS): Natural sand replacement at 0%, 15%, 30%, 45%, and 60% by mass. These levels represent practical scenarios for recycled aggregate utilization in SCC applications. The upper limit of 60% reflects the maximum feasible replacement that maintains acceptable fresh and hardened properties when supported by appropriate mix design adjustments [40], [41]. This range balances sustainability objectives with performance requirements.

Steel Fibers (SF): Volume fractions of 0%, 0.3%, 0.6%, 0.9%, and 1.2%. These dosages align with established recommendations for fiber-reinforced SCC, where fibers enhance post-cracking behavior without severely compromising workability [42–44]. The 1.2% upper limit prevents fiber balling phenomena and ensures uniform dispersion during mixing operations.

All mixtures maintained consistent water-to-cement ratio of 0.42 and cement content of 400 kg/m³, resulting in water content of 168 kg/m³. Polycarboxylate-based superplasticizer (MEDAPLAST SP 40) was incorporated at 2.45% by cement mass to achieve required flowability while preserving mixture stability.

The experimental matrix was generated using Design-Expert®13 software (Stat-Ease, Inc.), which provided specialized algorithms for mixture design and response surface methodology. The I-optimal algorithm strategically distributed the forty-four experimental points throughout the constrained mixture space, including replicated configurations to enable pure error estimation for lack-of-fit testing. This software-based approach enabled systematic optimization of mixture proportions, construction of polynomial regression models, and visualization of response surfaces, providing a robust analytical framework for examining interactions among WGP, RS, and SF while ensuring reliable detection of curvature effects and interaction terms essential for accurate response surface modeling.

III.3 MIX DESIGN AND PROPORTIONING

The mixture proportioning methodology was designed to maintain consistent baseline conditions while systematically varying the three target parameters. The following components remained constant across all formulations:

- Total binder content: 400 kg/m³ (cement + WGP)
- Water-to-binder ratio: 0.42 (considering WGP as active binder)
- Superplasticizer dosage: 2.45% by cement mass
- Total aggregate volume: maintained on absolute volume basis

WGP replaced cement on a direct mass substitution basis while maintaining total binder content at 400 kg/m³. This approach recognizes the pozzolanic contribution of finely ground glass powder to long-term strength development. RS replacement followed mass-for-mass substitution of natural sand according to designated percentages. Steel fibers were incorporated volumetrically, with corresponding reduction in 8/16 mm coarse aggregate mass to maintain constant paste volume and mixture balance. The 3/8 mm gravel fraction remained constant at 270.3 kg/m³ throughout all formulations.

This systematic proportioning strategy ensured that observed variations in concrete properties could be directly attributed to WGP, RS, and SF influences rather than inconsistencies in overall mixture composition, providing a reliable foundation for evaluating individual and interactive effects on SCC performance.

III.4 COMPLETE EXPERIMENTAL MATRIX

Table 7 presents the comprehensive experimental matrix detailing the proportions of WGP, RS, and SF for each mixture, along with calculated quantities of all constituents. The table includes both the percentage levels of variable components and absolute quantities (kg/m³) of all mixture constituents, ensuring complete reproducibility of the experimental program.

III.5 SPECIMEN PREPARATION AND CURING

All concrete mixtures were prepared in a 100-L vertical-shaft pan mixer under controlled laboratory conditions (20 ± 2 °C, 65 ± 5% RH) following a standardized sequence to ensure batch consistency. Cement, waste glass powder (WGP), and aggregates were first dry-blended for 2 minutes, after which steel fibers were gradually added and mixed for 1 minute to prevent clustering. Approximately two-thirds of the mixing water was then introduced and mixed for 2 minutes, followed by the remaining water pre-dissolved with a polycarboxylate superplasticizer and mixed for 3 minutes until a cohesive, self-compacting consistency was achieved.

Fresh concrete was cast into steel molds without external vibration, relying on its inherent self-compacting ability. After 24 hours under laboratory conditions, specimens were demolded and cured in water at 20 ± 2 °C until testing. This preparation and curing procedure ensured reproducibility and provided a robust basis for assessing the effects of WGP, recycled sand (RS), and steel fibers (SF) on both fresh and hardened properties of self-compacting concrete.

Table 7: Experimental matrix : proportions of WGP, RS, and SF with fixed binder and aggregate contents.

Mix No.	WGP	RS	SF	Cement	Sand		Gravel		WGP	Water
					Natural	Recycled	3/8	8/16		
	(%)	(%)	(%)	(kg/m³)	(kg/m³)	(kg/m³)	(kg/m³)	(kg/m³)	(kg/m³)	(kg/m³)
1	0	0	0	400.0	820.0	0.0	270.3	540.6	0.0	168.0
2	5	0	0	380.0	820.0	0.0	270.3	540.6	20.0	168.0
3	10	0	0	360.0	820.0	0.0	270.3	540.6	40.0	168.0
4	15	0	0	340.0	820.0	0.0	270.3	540.6	60.0	168.0
5	20	0	0	320.0	820.0	0.0	270.3	540.6	80.0	168.0
6	0	15	0	400.0	697.0	123.0	270.3	540.6	0.0	168.0

7	0	30	0	400.0	574.0	246.0	270.3	540.6	0.0	168.0
8	0	45	0	400.0	451.0	369.0	270.3	540.6	0.0	168.0
9	0	60	0	400.0	328.0	492.0	270.3	540.6	0.0	168.0
10	0	0	0.3	400.0	820.0	0.0	270.3	532.6	0.0	168.0
11	0	0	0.6	400.0	820.0	0.0	270.3	524.6	0.0	168.0
12	0	0	0.9	400.0	820.0	0.0	270.3	516.6	0.0	168.0
13	0	0	1.2	400.0	820.0	0.0	270.3	508.6	0.0	168.0
14	10	30	0.6	360.0	574.0	246.0	270.3	524.6	40.0	168.0
15	10	30	0	360.0	574.0	246.0	270.3	540.6	40.0	168.0
16	10	30	1.2	360.0	574.0	246.0	270.3	508.6	40.0	168.0
17	10	0	0.6	360.0	820.0	0.0	270.3	524.6	40.0	168.0
18	10	60	0.6	360.0	328.0	492.0	270.3	524.6	40.0	168.0
19	0	30	0.6	400.0	574.0	246.0	270.3	524.6	0.0	168.0
20	20	30	0.6	320.0	574.0	246.0	270.3	524.6	80.0	168.0
21	5	15	0	380.0	697.0	123.0	270.3	540.6	20.0	168.0
22	15	45	0	340.0	451.0	369.0	270.3	540.6	60.0	168.0
23	0	15	0.3	400.0	697.0	123.0	270.3	532.6	0.0	168.0
24	0	45	0.9	400.0	451.0	369.0	270.3	516.6	0.0	168.0
25	5	0	0.3	380.0	820.0	0.0	270.3	532.6	20.0	168.0
26	15	0	0.9	340.0	820.0	0.0	270.3	516.6	60.0	168.0
27	10	15	0.3	360.0	697.0	123.0	270.3	532.6	40.0	168.0
28	10	45	0.9	360.0	451.0	369.0	270.3	516.6	40.0	168.0
29	5	15	0.6	380.0	697.0	123.0	270.3	524.6	20.0	168.0
30	15	45	0.6	340.0	451.0	369.0	270.3	524.6	60.0	168.0
31	5	45	0.3	380.0	451.0	369.0	270.3	532.6	20.0	168.0
32	15	15	0.9	340.0	697.0	123.0	270.3	516.6	60.0	168.0
33	5	30	0.9	380.0	574.0	246.0	270.3	516.6	20.0	168.0
34	15	30	0.3	340.0	574.0	246.0	270.3	532.6	60.0	168.0
35	7.5	22.5	0.45	370.0	635.5	184.5	270.3	528.6	30.0	168.0
36	12.5	37.5	0.75	350.0	512.5	307.5	270.3	524.6	50.0	168.0
37	20	60	0	320.0	328.0	492.0	270.3	540.6	80.0	168.0
38	20	0	1.2	320.0	820.0	0.0	270.3	508.6	80.0	168.0
39	0	60	1.2	400.0	328.0	492.0	270.3	508.6	0.0	168.0
40	20	60	1.2	320.0	328.0	492.0	270.3	508.6	80.0	168.0

Source: Authors, (2025).

III.6 TESTING PROCEDURE

The experimental program was structured to comprehensively evaluate the fresh and hardened properties of steel fiber-reinforced self-compacting concrete incorporating waste glass powder (WGP) and recycled sand (RS).

Fresh concrete workability was assessed immediately after mixing using standardized SCC tests. The slump flow test (EN 12350-8) determined filling ability by measuring the average spread diameter in two perpendicular directions after lifting the cone. The V-funnel test (EN 12350-9) measured viscosity and segregation resistance through the flow time of concrete passing the V-shaped apparatus. Passing ability was evaluated by the J-ring test (EN 12350-12), comparing slump flow and J-ring flow diameters, and confirmed by the L-box test (EN 12350-10) through calculation of the blocking ratio (H_2/H_1).

All tests were completed within 15 minutes of mixing under controlled laboratory conditions (20 ± 2 °C, 65 ± 5% RH). Each measurement was conducted in triplicate, and mean values were reported when the variation among results was below 5%. This systematic protocol ensured reliable characterization of SCC workability in compliance with EFNARC guidelines [45]. Figure 5 depicts the fresh property testing setup used in the experimental program.

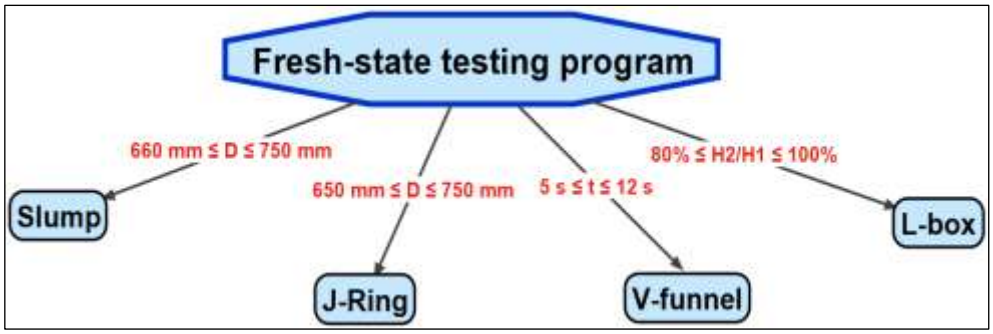


Figure 5: Fresh-state testing program for self-compacting concrete workability assessment.
Source: Authors, (2025).

The hardened performance of the mixtures was evaluated through a comprehensive program combining mechanical and durability assessments. Compressive strength was measured on 150 mm cubes at 28, 56, and 90 days in accordance with EN 12390-3. Specimens were demolded after 24 h and cured in water at 20 ± 2 °C until testing. For each age and mixture, the mean value of three specimens was reported to ensure statistical reliability.

Flexural strength was determined at 28 days using 100 × 100 × 400 mm prisms under a four-point bending setup following ASTM C78, with a controlled loading rate of 1.0 ± 0.3 MPa/min to maintain uniformity across batches. Durability was assessed at 90 days through capillary water absorption following ASTM C642. Oven-dried specimens (105 °C, constant mass) were exposed to unidirectional water ingress, and mass gain was recorded at specified intervals for 64 h. The sorptivity coefficient was calculated from the linear relationship between cumulative absorption and the square root of time.

This integrated methodology provided a robust basis for quantifying the effects of waste glass powder, recycled sand, and steel fibers on the mechanical and durability performance of self-compacting concrete. Figure 6 shows the testing setup and procedures for hardened property evaluation.

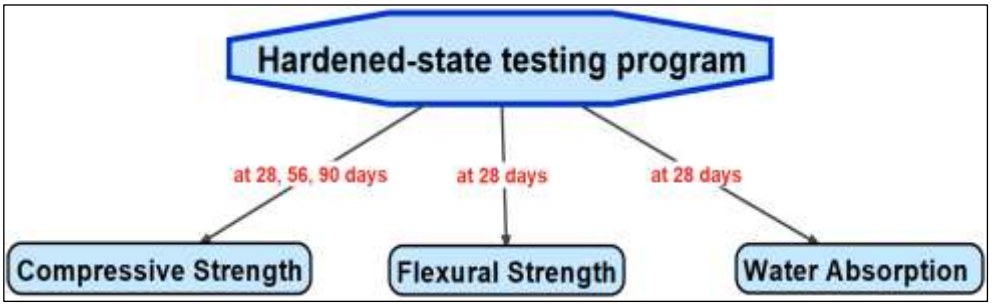


Figure 6: Hardened property evaluation setup for mechanical and durability characterization.
Source: Authors, (2025).

IV. RESULTS AND DISCUSSIONS

IV.1 MODEL DEVELOPMENT AND VALIDATION

The response surface methodology, implemented through the I-optimal mixture design, facilitated the development of statistically robust regression models describing the effects of waste glass powder (WGP), recycled sand (RS), and steel fibers (SF) on both the fresh and hardened properties of SCC. Quadratic polynomial functions were generally sufficient to capture the observed experimental trends, while cubic terms were introduced in specific cases—such as V-funnel and L-box tests—to better represent nonlinear flow behavior under restrictive conditions.

The predictive reliability of the models was demonstrated by coefficients of determination (R^2) ranging from 0.85 to 0.99, with adjusted R^2 values remaining closely aligned, confirming the absence of overfitting. Analysis of variance (ANOVA) further validated the statistical significance of all regression terms ($p < 0.0001$), while lack-of-fit tests were non-significant ($p > 0.05$), establishing the adequacy of the selected models.

Model adequacy was reinforced through diagnostic evaluations. Normal probability plots of residuals (Figure 7a and 7b) showed data points closely following a straight line, indicating normally distributed errors, while predicted versus experimental plots (Figure 8a and 8b) revealed strong correlations across all responses ($R^2 > 0.93$). These diagnostics confirm that the models not only meet statistical requirements but also capture the practical behavior of SCC mixtures under varying proportions of industrial by-products and fibers.

The complete regression equations, expressed in coded variables, are provided in Table 8 (fresh properties) and Table 9 (hardened properties). These tables consolidate the model forms, coefficients, and statistical indices, offering both a compact summary and a practical predictive framework. Collectively, the models provide a reliable tool for optimizing SCC mixtures, enabling engineers to balance workability, stability, and mechanical performance when incorporating recycled and supplementary materials.

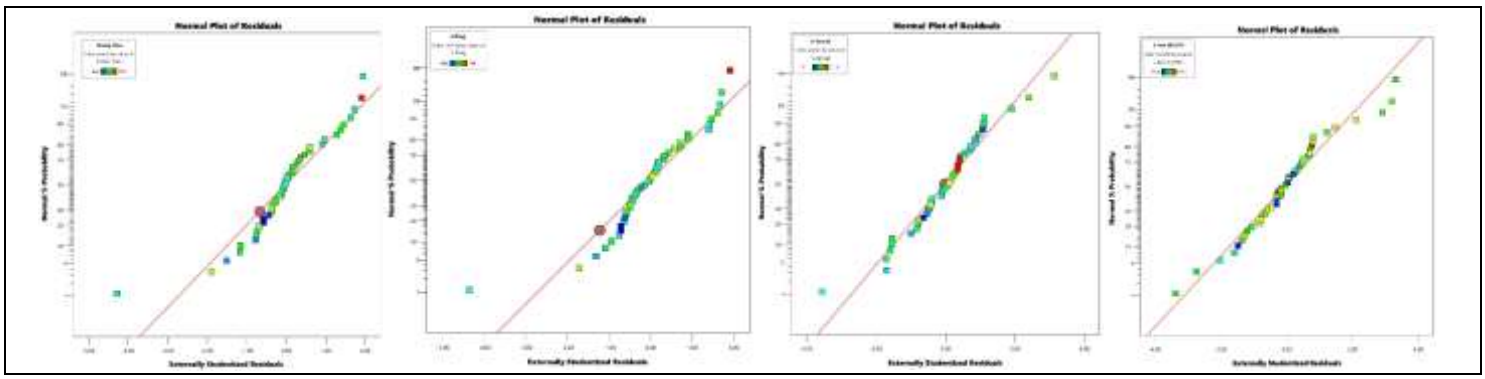


Figure 7a: Model validation diagnostics: Normal probability plots of residuals demonstrating adequate model fit for all response variables (fresh-state).
Source: Authors, (2025).

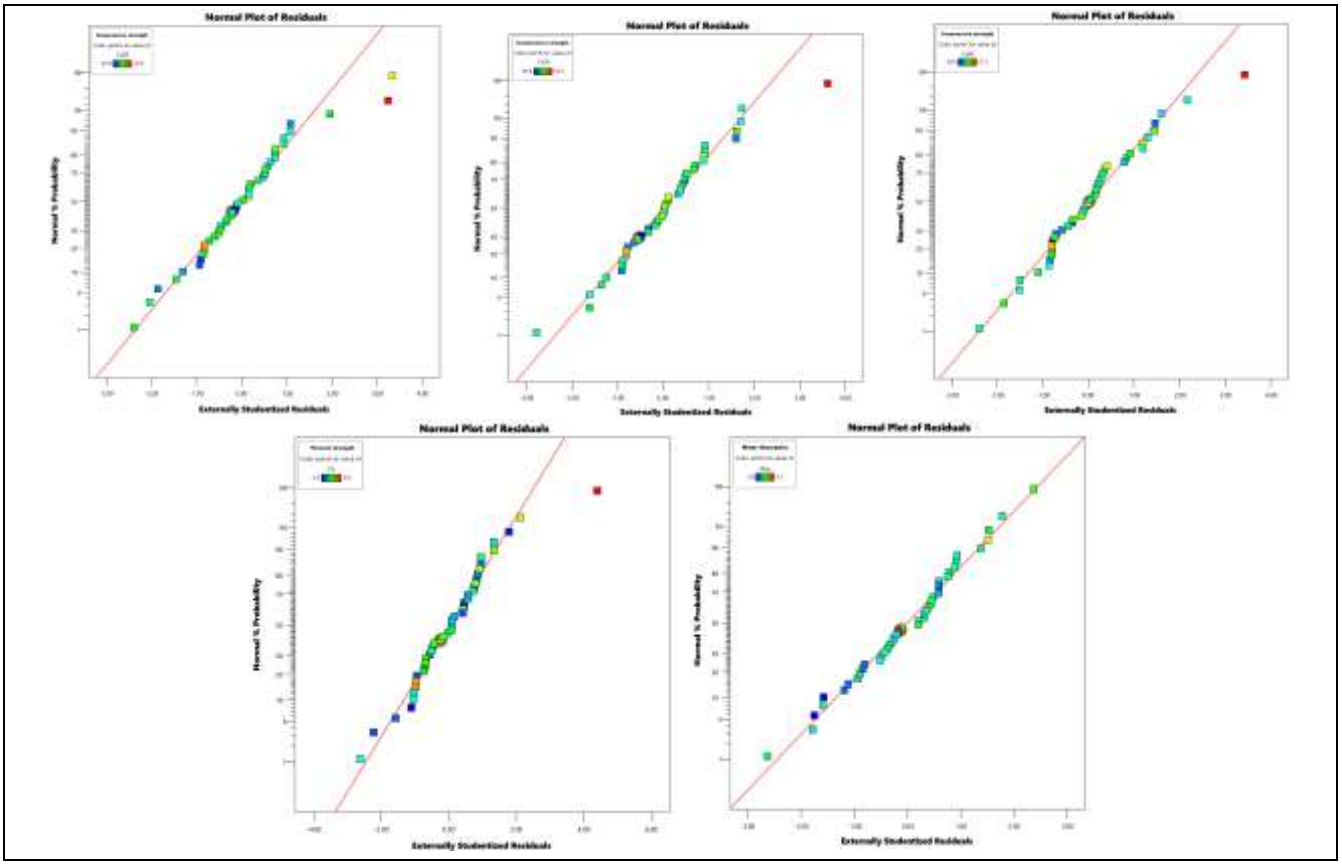


Figure 7b: Model validation diagnostics: Normal probability plots of residuals demonstrating adequate model fit for all response variables (hardened-state).
Source: Authors, (2025).

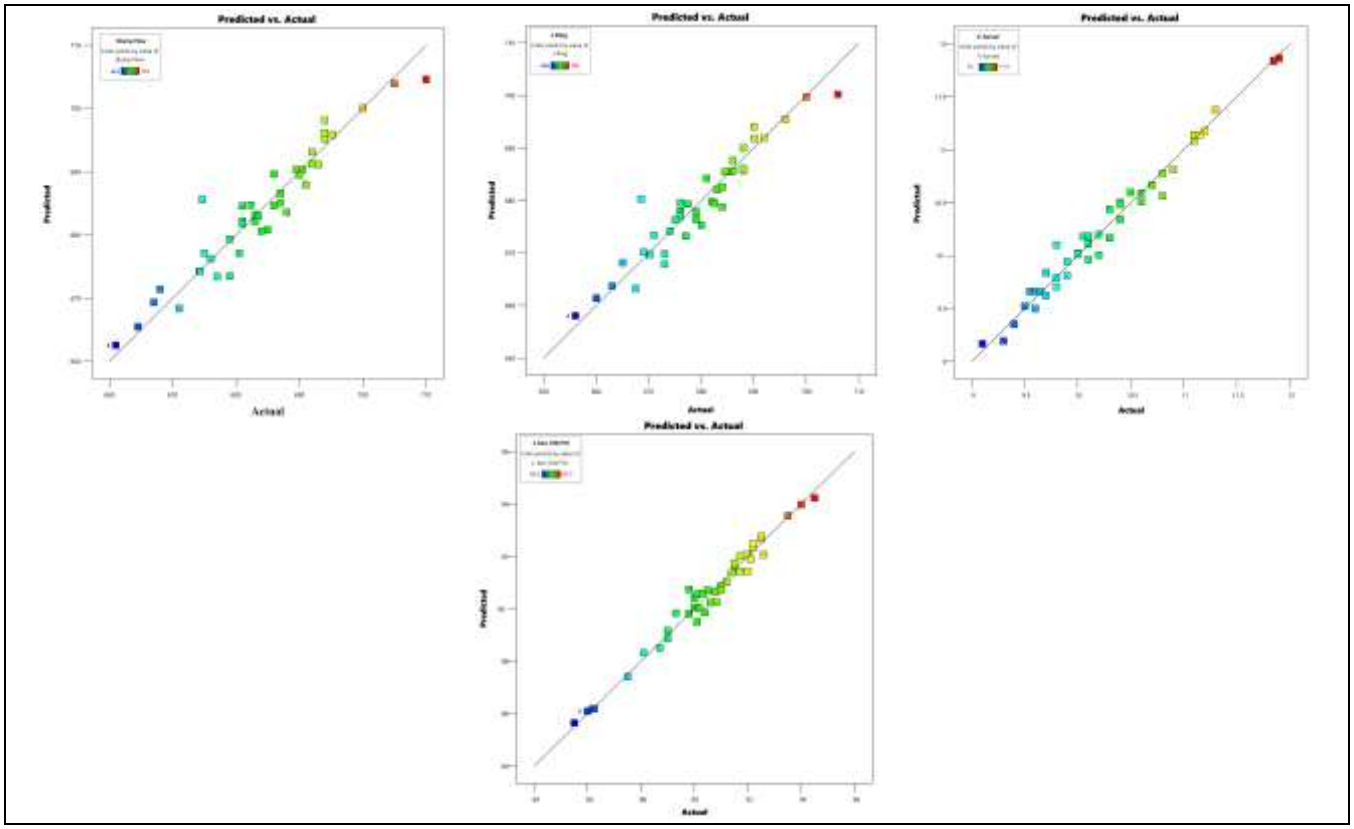


Figure 8a: Predicted versus actual correlations: (a-i) Model accuracy assessment across all measured properties with corresponding R^2 values (fresh-state).
Source: Authors, (2025).

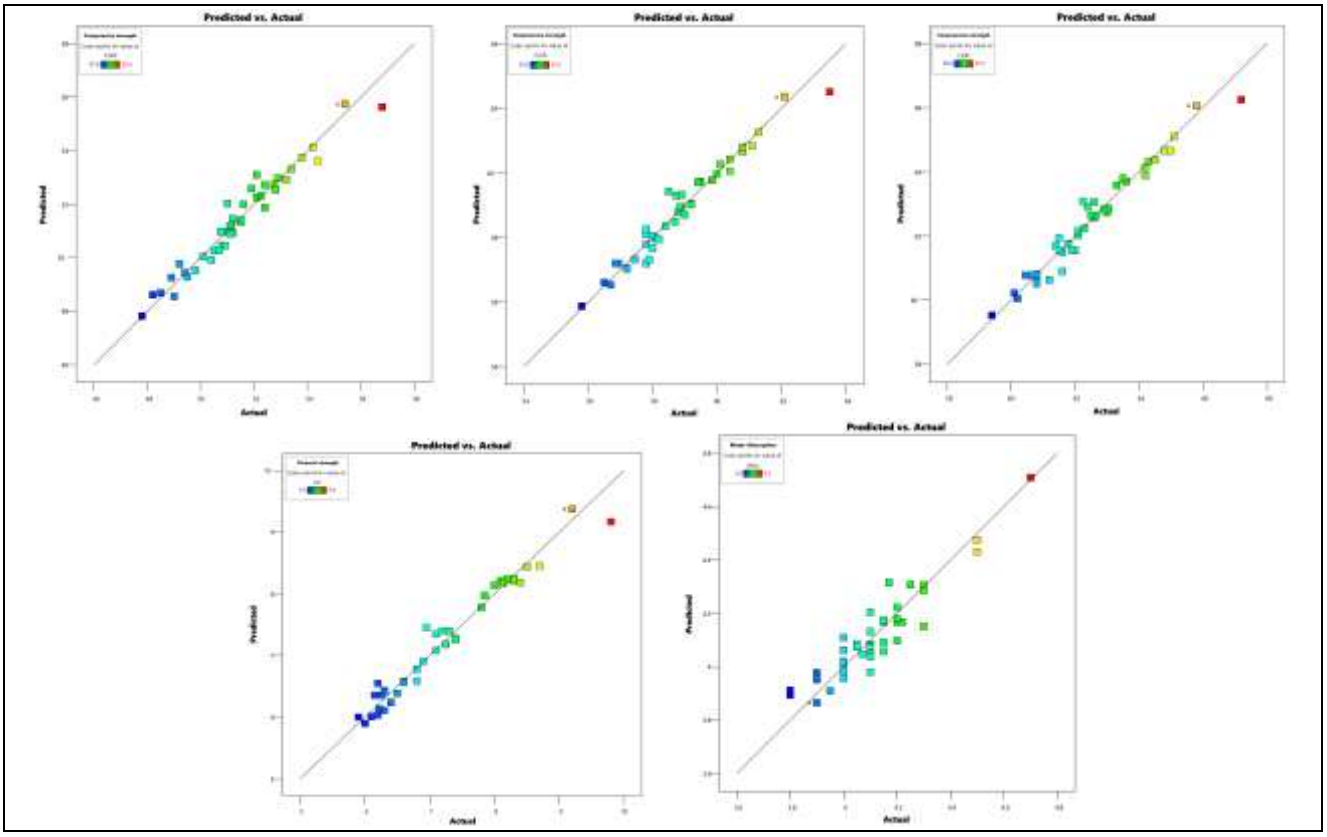


Figure 8b: Predicted versus actual correlations: (a-i) Model accuracy assessment across all measured properties with corresponding R^2 values (hardened-state).
Source: Authors, (2025).

Table 8: Regression models for the fresh-state properties of self-compacting concrete (SCC) incorporating waste glass powder (WGP), recycled sand (RS), and steel fibers (SF).

Response	Model Type	Regression Equation (coded variables)	F-value	R^2	Adj. R^2	Adeq. Precision
Slump Flow	Quadratic	$Y = 687.95 + 4.14A – 4.53B – 7.66C + 5.47BC – 10.34A^2$	47.20	0.9259	0.9063	26.82
J-Ring	Quadratic	$Y = 673.2 – 3.95A – 6.14B – 9.45C + 4.82BC – 8.90A^2$	41.71	0.9169	0.8950	25.32
V-Funnel	Cubic	$Y = 9.96 – 0.35A + 0.30B + 0.76C – 0.18AC – 0.29BC + 0.29A^2 + 0.34B^2 + 0.23C^2$	15.82	0.9507	0.9377	28.45
L-Box	Cubic	$Y = 0.91 – 0.01A – 0.02B – 0.05C + 0.02BC – 0.03C^2$	56.34	0.9021	0.8794	23.65

Source: Authors, (2025).

A = WGP (%), B = RS (%), C = SF (%); All models showed statistical significance ($p < 0.0001$).

Table 9: Regression models for the hardened-state properties of self-compacting concrete (SCC) incorporating waste glass powder (WGP), recycled sand (RS), and steel fibers (SF).

Response	Model Type	Regression Equation (coded variables)	F-value	R^2	Adj. R^2	Adeq. Precision
Compressive Strength (28d)	Quadratic	$Y = 52.03 – 0.36A – 0.52B + 3.26C – 0.33BC – 0.35B^2 – 0.33C^2$	26.29	0.9	0.8735	22.45
Compressive Strength (56d)	Cubic	$Y = 59.31 + 0.38A - 0.36B + 3.08C - 0.14AB - 0.70AC + 0.01BC - 0.12A^2 - 0.06B^2 - 0.22C^2$	31.78	0.9618	0.9315	27.43
Compressive Strength (90d)	Cubic	$Y = 63.06 + 0.63A – 0.38B + 3.20C – 0.65AC + 3.90ABC – 0.27C^2$	33.64	0.9638	0.9352	25.69
Flexural Strength	Cubic	$Y = 7.35 + 0.15A – 0.13B + 1.79C – 0.12AC$	36.30	0.9664	0.9397	30.64
Water Absorption	Quadratic	$Y = 4.02 – 0.047A + 0.16B – 0.16C – 0.057BC + 0.13B^2$	25.53	0.8715	0.8375	21.34

Source: Authors, (2025).

Coded variables: A = (WGP - 10)/10, B = (RS - 30)/30, C = (SF - 0.6)/0.6

IV.2 FRESH PROPERTIES

The fresh properties of SCC exhibited clear sensitivity to variations in WGP, RS, and SF. The following subsections present the individual test results, supported by response surface analyses, to elucidate the influence of mixture parameters on flowability, passing ability, and segregation resistance.

IV.2.1 Slump Flow

The slump flow test demonstrated that steel fiber (SF) content exerted the strongest influence on the workability of SCC mixtures. Regression analysis (Table 8) confirmed its dominant role, accounting for 21.9% of the model sum of squares ($F = 100.28$, $p < 0.0001$). Each 0.3% increment in fiber dosage consistently reduced slump flow diameter by approximately 25–30 mm. This reduction is primarily attributed to increased interparticle friction and mechanical blocking caused by fiber interlocking, a phenomenon widely reported in previous investigations [42], [46], [47]. From a practical perspective, this finding underscores the trade-off between the mechanical benefits of fibers and their adverse effect on flowability, an important consideration in the design of SCC for heavily reinforced structures.

A significant RS × SF interaction ($F = 38.73$, $p < 0.0001$) further revealed that the combined incorporation of recycled sand (RS) and fibers mitigated flow reduction compared to the expected additive effect of each factor alone. This compensating mechanism suggests that the angular particle geometry of RS may facilitate more efficient packing and particle rearrangement, thereby offsetting part of the flow restriction induced by fibers. Such synergy not only enhances mixture stability but also offers practical advantages when designing SCC with multiple waste-derived constituents.

Waste glass powder (WGP) also showed a notable influence on slump flow. At moderate replacement levels (≈12–15%), WGP improved flowability due to its filler effect and smooth particle morphology, which enhanced paste rheology and reduced internal friction. However, beyond this threshold, a significant quadratic effect (coefficient ≈ –10.34) was observed, leading to reduced flowability as a consequence of increased surface area and higher water demand. This nonlinear behavior confirms the presence of an optimal WGP content, beyond which the benefits are outweighed by rheological limitations. The trend aligns with earlier studies that have identified similar dosage thresholds for glass powder in SCC systems.

Figure 9 presents the response surface illustrating how waste glass powder and recycled sand jointly influence slump flow when steel fiber content remains fixed at 0.6%. The data indicate that flowability reaches its peak values (690-710 mm) within a relatively narrow window: WGP replacements between 5-12% paired with moderate recycled sand usage (15-30%). qWhen both materials exceed these ranges simultaneously, slump flow decreases noticeably. This pattern reflects how each material contributes to mixture thickening through different mechanisms - WGP through increased surface area at higher dosages, and recycled sand through its inherently higher water absorption and angular particle texture. Such findings underscore the critical balance required when incorporating multiple waste streams into self-compacting concrete formulations.

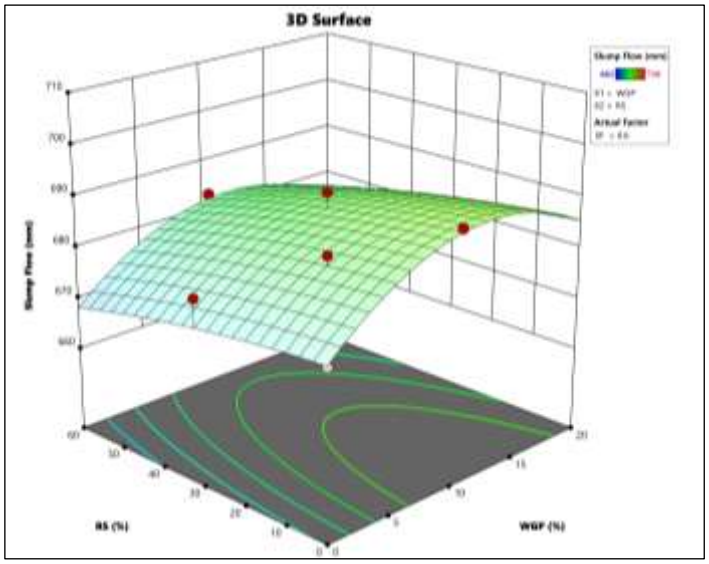


Figure 9: Interactive effects of waste glass powder and recycled sand on concrete flowability represented through three-dimensional response surface modeling.
Source: Authors, (2025).

The contour plot (Figure 10) offers a different perspective on the same data, making it easier to identify specific operating windows for mixture design. Two distinct zones emerge from this analysis: an optimal region where slump flow exceeds 700 mm, and areas where performance deteriorates rapidly.

The high-performance zone appears quite narrow, confined to combinations where WGP stays below 10% while recycled sand ranges between 15-30%. Beyond these boundaries, particularly when WGP content rises above 15%, flowability drops sharply regardless of the recycled sand level. Similarly, recycled sand proportions exceeding 45% consistently undermine workability, even with minimal glass powder addition.

These patterns suggest that each material imposes its own threshold beyond which the benefits reverse into drawbacks. For practical mixture design, this mapping helps engineers avoid problematic combinations while identifying the sweet spot for achieving both sustainability goals and workability requirements.

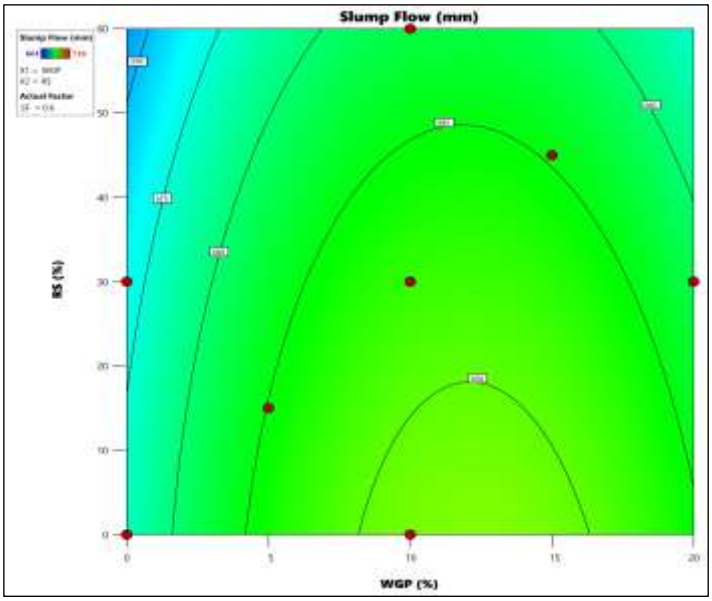


Figure 10: Contour representation of slump flow response revealing optimal mixture proportions for waste glass powder and recycled sand combinations.
Source: Authors, (2025).

IV.2.2 V-Funnel Flow Characteristics

The V-funnel test required a cubic regression model to capture the nonlinear flow behavior of SCC mixtures, achieving an excellent fit (R^2 = 0.9507). Flow times ranged from 8.2 s in the fiber-free control mix to 12.5 s at the highest fiber dosage (1.2%), clearly demonstrating the strong influence of steel fibers on mixture viscosity and segregation resistance. This marked increase reflects a transition from nearly Newtonian to distinctly non-Newtonian flow behavior, as fiber networks progressively develop within the fresh matrix. These results corroborate the observations of Roussel (2006), who reported similar rheological transitions in fiber-reinforced SCC under increasing shear.

The cubic model was crucial for identifying threshold effects, particularly at low fiber dosages where small additions caused disproportionately large increases in flow resistance. This suggests the presence of a critical fiber content at which percolation networks begin to form, fundamentally altering flow dynamics. Waste glass powder (WGP) exerted a secondary but measurable influence, reducing flow times by roughly 8–12% at moderate replacement levels due to enhanced particle packing and improved lubrication. However, replacement levels beyond 15% reversed this benefit, likely due to increased specific surface area and water demand. Recycled sand (RS) consistently elevated flow times, surpassing EFNARC limits (>12 s) when replacement exceeded 45%, primarily as a result of higher water absorption and the angular particle morphology typical of recycled aggregates.

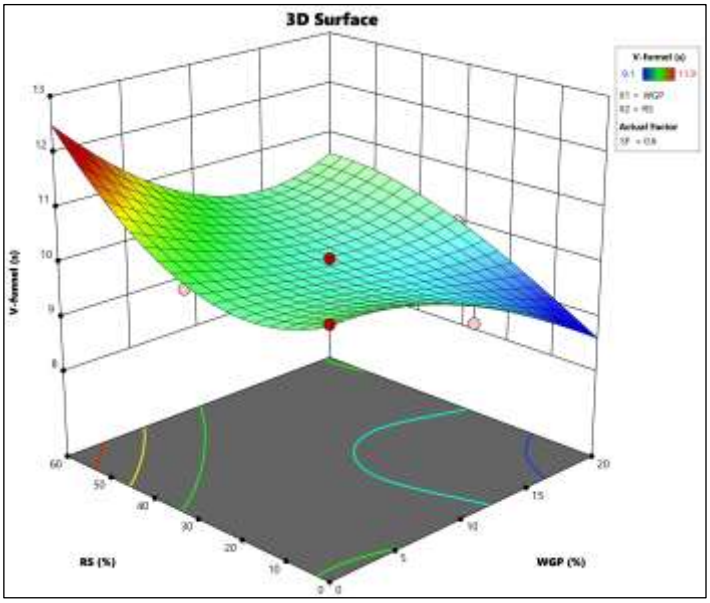


Figure 11: Response surface of V-funnel flow time showing nonlinear influence of steel fiber content.
Source: Authors, (2025).

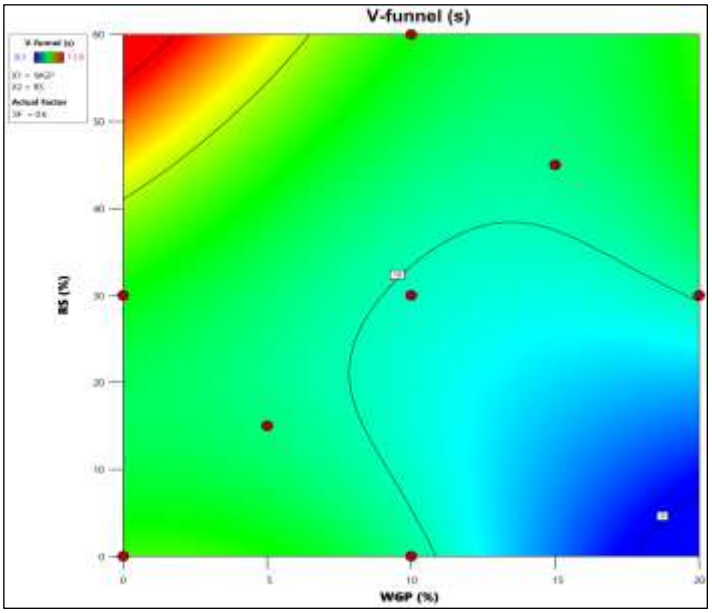


Figure 12: Contour map illustrating critical fiber thresholds required to maintain acceptable flow characteristics.
Source: Authors, (2025).

IV.2.3 J-Ring Performance

Passing ability is a key performance criterion for SCC, particularly in heavily reinforced elements. Both J-ring and L-box tests identified steel fiber dosage as the main factor restricting flow, though each test revealed different aspects of this limitation.

The J-ring model ($R^2 = 0.9169$) showed that each 0.3% increment of steel fibers reduced the J-ring spread by 15–20 mm compared with unconfined slump flow, reaching a maximum difference of approximately 50 mm at 1.2% fiber content. This reduction reflects the tendency of fibers to align around obstacles, forming bridging zones that hinder flow. Similar reductions were reported by Libre et al. (2011) in their investigation of hooked-end fibers.

Recycled sand further amplified these restrictions due to its angular shape and higher interparticle friction. Mixtures containing more than 60% RS consistently fell below recommended J-ring performance limits regardless of fiber dosage. Interestingly, moderate WGP levels (10–12%) mitigated this effect by improving paste lubrication and cohesion, leading to slightly higher J-ring spreads.

IV.2.4 L-Box

L-box results followed a similar trend but were even more sensitive to fiber content. Blocking ratios decreased from 0.95 in the control mix to 0.82 at 1.2% fiber dosage, approaching the EFNARC critical threshold of 0.80. The quadratic trend indicates that flow restriction accelerates at higher fiber contents, most likely due to intensified fiber–fiber interactions and network formation.

Critical combinations were observed when high fiber content was paired with elevated RS levels. For example, mixtures containing 1.0% SF and 45% RS produced blocking ratios of 0.79, slightly below EFNARC acceptance criteria. This confirms the multiplicative rather than additive nature of flow restrictions, consistent with the findings of Khayat et al. (2019), who reported similar compounded effects in fiber-reinforced SCC.

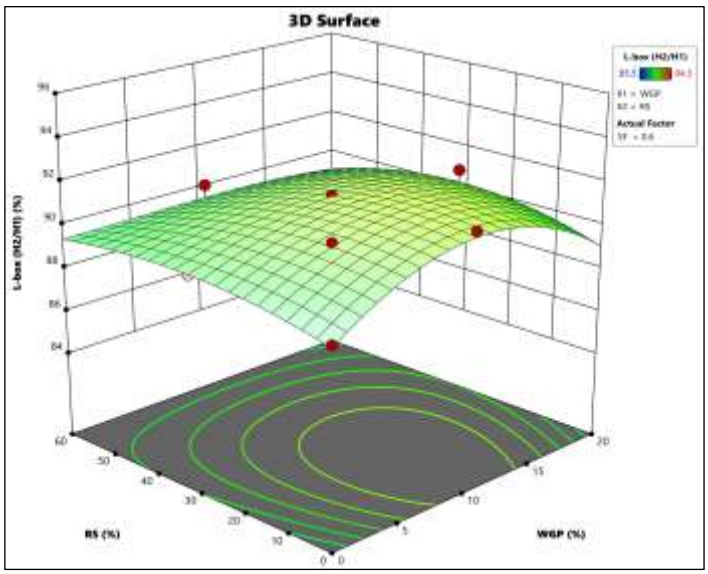


Figure 13: L-box blocking ratio response surface highlighting critical performance boundaries for fiber-reinforced SCC mixtures.
Source: Authors, (2025).

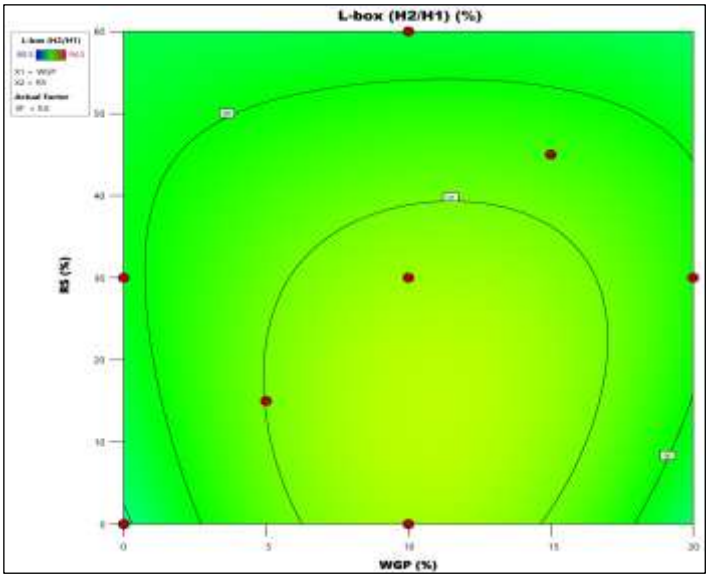


Figure 14: Optimization contours identifying acceptable design regions for sustainable SCC containing recycled materials.
Source: Authors, (2025).

IV.3 HARDENED PROPERTIES

IV.3.1 Compressive Strength Development

Compressive strength development reflected the combined effects of mechanical reinforcement from steel fibers and chemical enhancement from pozzolanic reactions. Early-Age Strength (28 Days), At 28 days, steel fiber content was the dominant factor (R^2 = 0.9000), with each 0.3% increment producing a 2.8–3.2 MPa strength gain. This enhancement is attributed to the crack-bridging capacity of fibers, which maintain load transfer across developing microcracks and delay crack propagation. These results align with the findings of Yazıcı et al. (2007), who reported comparable strength gains in high-performance concrete with hooked-end fibers.

WGP exhibited a slight negative coefficient (-0.36) at this age, reflecting the slower reactivity of glass powder compared to cement hydration. This early-age strength reduction has been reported by Shi et al. (2005) and is attributed to the time required for secondary C–S–H formation. RS replacement showed a predictable negative influence (-0.52 MPa per 15% replacement), consistent with the weaker interfacial transition zones typically associated with recycled aggregates (Evangelista and de Brito, 2007).

Intermediate-Age Development (56 Days), By 56 days, the pozzolanic contribution of WGP became evident, with positive coefficients (+0.38 MPa) and visible strength improvements. The optimal mix (15% WGP, 50% RS, 1.0% SF) reached 59.7 MPa, a 24% increase compared with the control. This demonstrates a synergistic interaction between fiber reinforcement and secondary hydration, which progressively densifies the matrix. Long-Term Strength (90 Days), At 90 days, the WGP effect increased further (+0.63 MPa), with the optimal mixture achieving 64.8 MPa, representing a 28% improvement over control. These results confirm the substantial pozzolanic contribution of WGP and align with the findings of Rashad (2014), who documented significant long-term strength gains in glass powder concrete.

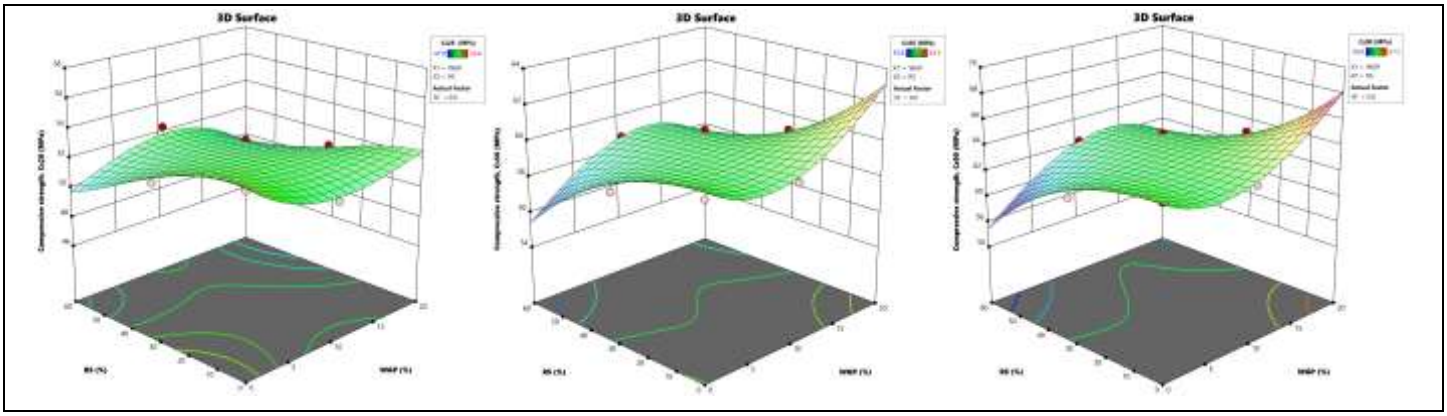


Figure 15: Response surfaces showing early-age fiber dominance, 56-day WGP activation, and 90-day strength improvement trends.
Source: Authors, (2025).

IV.3.2 Flexural Strength

Flexural strength analysis achieved excellent model accuracy (R^2 = 0.9276). Steel fiber content was the most significant factor (+1.79 MPa), with the optimal mix achieving 8.7 MPa at 28 days—a 38% increase over the control mixture. This demonstrates the ability of hooked-end fibers to transform brittle failure into pseudo-ductile behavior, consistent with the mechanism described by Naaman (2003).

WGP provided a minor positive contribution (+0.15), while RS showed minimal negative impact (-0.13). The limited influence of these materials indicates that fiber reinforcement can compensate for matrix weakening effects, supporting the use of multiple sustainable constituents without compromising structural performance.

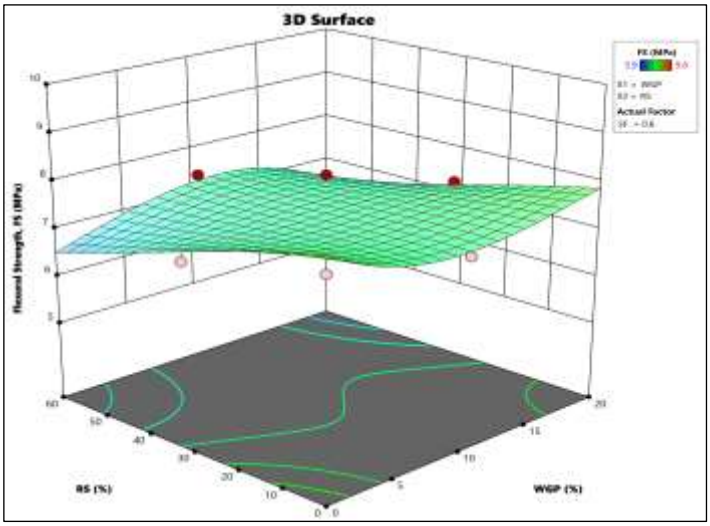


Figure 16: Response surface highlighting the dominant influence of steel fibers on flexural performance.
Source: Authors, (2025).

IV.3.3 Water absorption

Water absorption testing at 90 days (R^2 = 0.8715) revealed RS as the primary factor influencing permeability. Each 15% replacement increased absorption by 0.15–0.20% [48], due to the higher porosity and microcracking typical of recycled aggregates (Tam et al., 2018). WGP had a beneficial effect (-0.047 per 15% replacement) through matrix densification driven by pozzolanic reaction products. Steel fibers also reduced absorption (-0.164) by controlling crack formation and limiting pathways for fluid ingress, confirming the findings of Banthia and Gupta (2006). The optimal mix achieved 3.8% water absorption at 90 days, 15% lower than the control, placing it in the "excellent" category according to ACI 201.2R guidelines.

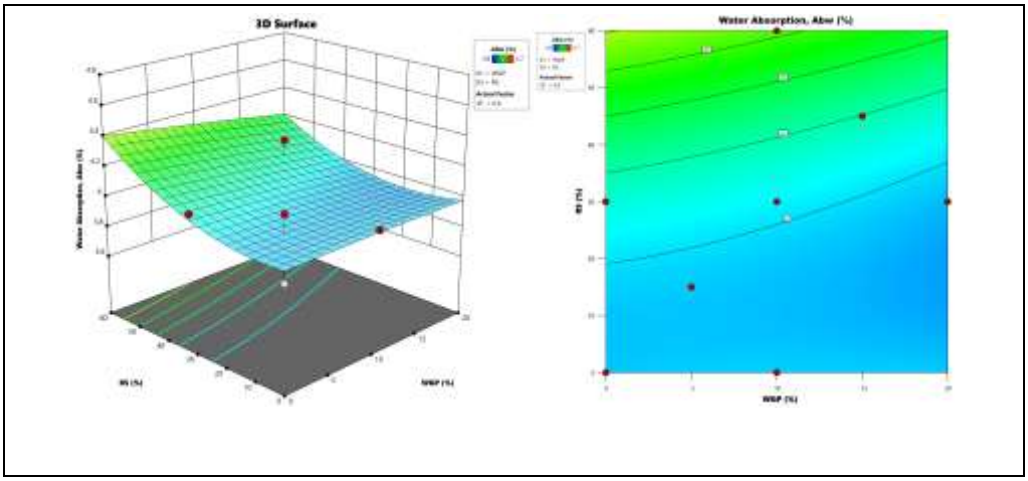


Figure 17: Response surface and optimization contours for durability performance.
Source: Authors, (2025).

V. CONCLUSIONS

This study systematically investigated the combined influence of waste glass powder (WGP), recycled sand (RS), and steel fibers (SF) on the fresh and hardened performance of self-compacting concrete (SCC) through response surface methodology. Results confirmed that partial cement replacement with WGP up to 15% enhanced workability and long-term strength through filler and pozzolanic effects, while RS replacement up to 50% maintained flowability within EFNARC limits when mix adjustments accounted for its higher absorption. Incorporation of 1.0% SF by volume markedly increased flexural strength and improved ductility without compromising self-compaction. Synergistic interactions between RS and SF produced compensatory effects that stabilized fresh properties and improved mechanical response, highlighting the importance of integrated mix design rather than isolated material contributions. The optimized mixture (15% WGP, 15% RS, 0.9% SF) achieved high compressive strength (53.4 MPa at 28 days, 64.8 MPa at 90 days), superior flexural performance, and reduced water absorption, reflecting improved matrix densification. The developed regression models ($R^2 > 0.92$) offer reliable predictive tools for practical mixture optimization. Overall, the findings demonstrate that carefully proportioned waste-derived constituents can deliver sustainable, high-performance SCC, supporting circular economy objectives while meeting modern durability and workability requirements.

VI. AUTdHOR'S CONTRIBUTION

Conceptualization: Souhila Fergani, Said Zaouai, Rachid Rabehi and Ali Abbache.
Methodology: Souhila Fergani, Said Zaouai, Rachid Rabehi and Ali Abbache.
Investigation: Souhila Fergani, Said Zaouai, Rachid Rabehi and Ali Abbache.
Discussion of results: Souhila Fergani, Said Zaouai, Rachid Rabehi and Ali Abbache.
Writing – Original Draft: Souhila Fergani, Said Zaouai, Rachid Rabehi and Ali Abbache.
Writing – Review and Editing: Souhila Fergani, Said Zaouai, Rachid Rabehi and Ali Abbache.
Resources: Souhila Fergani, Said Zaouai, Rachid Rabehi and Ali Abbache.
Supervision: Souhila Fergani, Said Zaouai, Rachid Rabehi and Ali Abbache.
Approval of the final text: Souhila Fergani, Said Zaouai, Rachid Rabehi and Ali Abbache.

VII. REFERENCES

[1] M. Sandanayake, G. Zhang, S. Setunge, Estimation of environmental emissions and impacts of building construction – A decision making tool for contractors, Journal of Building Engineering 21 (2019) 173–185. https://doi.org/10.1016/j.jobe.2018.10.023.

[2] M. Schneider, The cement industry on the way to a low-carbon future, Cement and Concrete Research 124 (2019) 105792. https://doi.org/10.1016/j.cemconres.2019.105792.

[3] M. Bendixen, L.L. Iversen, J. Best, D.M. Franks, C.R. Hackney, E.M. Latrubesse, L.S. Tusting, Sand, gravel, and UN Sustainable Development Goals: Conflicts, synergies, and pathways forward, One Earth 4 (2021) 1095–1111. https://doi.org/10.1016/j.oneear.2021.07.008.

[4] C.R. Rusnak, Sustainable Strategies for Concrete Infrastructure Preservation: A Comprehensive Review and Perspective, Infrastructures 10 (2025) 99. https://doi.org/10.3390/infrastructures10040099.

[5] H. Okamura, M. Ouchi, Self-Compacting Concrete, ACT 1 (2003) 5–15. https://doi.org/10.3151/jact.1.5.

[6] Said Mohammed Mostafa Aljamala, N.A. Safiee, N.A. Mohd Nasir, F.N.A. Abdul Aziz, Performance of Recycling Aggregate Self-Compacting Concrete Incorporating Supplementary Cementitious Materials: An Overview, JST 33 (2025). https://doi.org/10.47836/pjst.33.S4.08.

[7] L. Prasittisopin, W. Tuvayanond, T.H.-K. Kang, S. Kaewunruen, Concrete Mix Design of Recycled Concrete Aggregate (RCA): Analysis of Review Papers, Characteristics, Research Trends, and Underexplored Topics, Resources 14 (2025) 21. https://doi.org/10.3390/resources14020021.

[8] R. Rabehi, M. Rabehi, M. Omrane, Physical-mechanical and fresh state properties of self-compacting concrete based on different types of gravel reinforced with steel fibers: Experimental study and modeling, Construction and Building Materials 390 (2023) 131758. https://doi.org/10.1016/j.conbuildmat.2023.131758.

[9] C. Shi, Y. Wu, C. Riefler, H. Wang, Characteristics and pozzolanic reactivity of glass powders, Cement and Concrete Research 35 (2005) 987–993. https://doi.org/10.1016/j.cemconres.2004.05.015.

[10] A.A. Aliabdo, A.E.M. Abd Elmoaty, A.Y. Aboshama, Utilization of waste glass powder in the production of cement and concrete, Construction and Building Materials 124 (2016) 866–877. https://doi.org/10.1016/j.conbuildmat.2016.08.016.

[11] X. Jiang, R. Xiao, Y. Bai, B. Huang, Y. Ma, Influence of waste glass powder as a supplementary cementitious material (SCM) on physical and mechanical properties of cement paste under high temperatures, Journal of Cleaner Production 340 (2022) 130778. https://doi.org/10.1016/j.jclepro.2022.130778.

[12] M. Mejdi, W. Wilson, M. Saillio, T. Chaussadent, L. Divet, A. Tagnit-Hamou, Investigating the pozzolanic reaction of post-consumption glass powder and the role of portlandite in the formation of sodium-rich C-S-H, Cement and Concrete Research 123 (2019) 105790. https://doi.org/10.1016/j.cemconres.2019.105790.

[13] S. Ramírez-Arellanes, F. Montejo-Alvaro, H. Cruz-Martínez, H. Rojas-Chávez, J.M. Mendoza-Rangel, V.A. Franco-Luján, Pozzolanic Assessment of Recycled Waste Glass for Use as a Supplementary Cementitious Material, Construction Materials 5 (2025) 59. https://doi.org/10.3390/constrmater5030059.

[14] A.M. Matos, T. Ramos, S. Nunes, J. Sousa-Coutinho, Durability Enhancement Of SCC With Waste Glass Powder, Mat. Res. 19 (2016) 67–74. https://doi.org/10.1590/1980-5373-MR-2015-0288.

[15] E. Pahsha, P.S. Nair, R. Gupta, V. Agrawal, Sustainable development of self-compacting concrete incorporating granite waste and recycled concrete aggregate: Evaluation of strength, durability, and microstructure, Sustainable Chemistry and Pharmacy 47 (2025) 102136. https://doi.org/10.1016/j.scp.2025.102136.

[16] N. Garcia-Troncoso, L. Li, Q. Cheng, K.H. Mo, T.-C. Ling, Comparative study on the properties and high temperature resistance of self-compacting concrete with various types of recycled aggregates, Case Studies in Construction Materials 15 (2021) e00678. https://doi.org/10.1016/j.cscm.2021.e00678.

[17] M. Contreras Llanes, M. Romero Pérez, M.J. Gázquez González, J.P. Bolívar Raya, Construction and demolition waste as recycled aggregate for environmentally friendly concrete paving, Environ Sci Pollut Res 29 (2022) 9826–9840. https://doi.org/10.1007/s11356-021-15849-4.

[18] J. Ahmad, Z. Zhou, A.F. Deifalla, Steel Fiber Reinforced Self-Compacting Concrete: A Comprehensive Review, Int J Concr Struct Mater 17 (2023) 51. https://doi.org/10.1186/s40069-023-00602-7.

[19] M. Zhang, J. Chen, J. Liu, H. Yin, Y. Ma, F. Yang, Fracture Behavior of Steel-Fiber-Reinforced High-Strength Self-Compacting Concrete: A Digital Image Correlation Analysis, Materials 18 (2025) 3631. https://doi.org/10.3390/ma18153631.

[20] S. Thakur, U. Jhakal, P. Anand, P.N. Bhowmik, Synergistic use of recycled aggregates and waste glass powder for sustainable concrete: mechanical properties, durability performance, and microstructural insights, Asian J Civ Eng (2025). https://doi.org/10.1007/s42107-025-01483-9.

[21] A. Mardani, D. Hatungimana, N. Mardani, J. Assaad, H. El-Hassan, Feasibility of steel fiber-reinforced self-compacting concrete containing recycled aggregates – compliance with EFNARC guidelines, International Journal of Sustainable Engineering 18 (2025) 2538858. https://doi.org/10.1080/19397038.2025.2538858.

[22] S. Barbhuiya, B.B. Das, D. Adak, K. Kapoor, M. Tabish, Low carbon concrete: advancements, challenges and future directions in sustainable construction, Discov. Concr. Cem. 1 (2025) 3. https://doi.org/10.1007/s44416-025-00002-y.

[23] Z. Zhang, Y. Lei, J.Y.R. Liew, M. Liu, G. Wong, H. Du, Embodied carbon saving potential of using recycled materials as cement substitute in Singapore's buildings, Npj Mater. Sustain. 2 (2024) 27. https://doi.org/10.1038/s44296-024-00032-w.

[24] R.H. Bogue, The Chemistry of Portland Cement. Second Edition, Soil Science 79 (1955) 322.

[25] A. C33, ASTM C33 standard specifications for concrete aggregates, ASTM Standard Book (2003).

[26] B. EN, 12620, Granulats Pour Bétons Hydrauliques et Mortiers (2008).

[27] A. ASTM, Standard test method for relative density (specific gravity) and absorption of coarse aggregate, ASTM West Conshohocken, PA (2015).

[28] T. EN, 1097-6, 2013, Tests for Mechanical and Physical Properties of Aggregates-Part 6 (2013) 1–12.

[29] P. American Society for Testing and Materials, ASTM C136/C136M-19: Standard Test Method for Sieve Analysis of Fine and Coarse Aggregates, in: ASTM, 2019.

[30] B. En, Tests for Geometrical Properties of Aggregates-Part 1: Determination of Particle Size Distribution-Sieving Method, Czech Office for Standards, Metrology and Testing, Prague (2012) 933–1.

[31] P. ASTM, C29/C29M-17a. 2017, Standard Test Method for Bulk Density ("Unit Weight") and Voids in Aggregate (n.d.).

[32] B. EN, 1097-3: 1998, 1998. Tests for Mechanical and Physical Properties of Aggregates. Determination of Loose Bulk Density and Voids, British Standards Institution.

[33] H. Du, K.H. Tan, Properties of high-volume glass powder concrete, Cement and Concrete Composites 75 (2017) 22–29. https://doi.org/10.1016/j.cemconcomp.2016.10.010.

[34] M.S. Hassani, J.C. Matos, Y. Zhang, E.R. Teixeira, Green Concrete with Glass Powder—A Literature Review, Sustainability 15 (2023) 14864. https://doi.org/10.3390/su152014864.

[35] B. Sara, A. Mhamed, B. Otmane, E. Karim, Elaboration of a Self-Compacting mortar based on concrete demolition waste incorporating blast furnace slag, Construction and Building Materials 366 (2023) 130165. https://doi.org/10.1016/j.conbuildmat.2022.130165.

[36] B.C. Mendes, L.G. Pedroti, C.M.F. Vieira, J.M.F.D. Carvalho, J.C.L. Ribeiro, C.M.M.D. Souza, Application of mixture design of experiments to the development of alkali-activated composites based on chamotte and waste glass, Construction and Building Materials 379 (2023) 131139. https://doi.org/10.1016/j.conbuildmat.2023.131139.

[37] Stat-Ease, (n.d.). https://www.statease.com/ (accessed August 25, 2025).

[38] H. Singh, R. Siddique, Long term durability assessment of self-compacting concrete made with crushed recycled glass and metakaolin, Construction and Building Materials 400 (2023) 132656. https://doi.org/10.1016/j.conbuildmat.2023.132656.

[39] M.H. Kumar, N.R. Mohanta, S. Samantaray, N.M. Kumar, Combined effect of waste glass powder and recycled steel fibers on mechanical behavior of concrete, SN Appl. Sci. 3 (2021) 350. https://doi.org/10.1007/s42452-021-04353-6.

[40] M. Nili, H. Sasanipour, F. Aslani, The Effect of Fine and Coarse Recycled Aggregates on Fresh and Mechanical Properties of Self-Compacting Concrete, Materials 12 (2019) 1120. https://doi.org/10.3390/ma12071120.

[41] B. Sara, A. Mhamed, B. Otmane, E. Karim, Elaboration of a Self-Compacting mortar based on concrete demolition waste incorporating blast furnace slag, Construction and Building Materials 366 (2023) 130165. https://doi.org/10.1016/j.conbuildmat.2022.130165.

[42] S. Yehia, A. Douba, O. Abdullahi, S. Farrag, Mechanical and durability evaluation of fiber-reinforced self-compacting concrete, Construction and Building Materials 121 (2016) 120–133. https://doi.org/10.1016/j.conbuildmat.2016.05.127.

[43] A. Khaloo, E. Molaei Raisi, P. Hosseini, H. Tahsiri, Mechanical performance of self-compacting concrete reinforced with steel fibers, Construction and Building Materials 51 (2014) 179–186. https://doi.org/10.1016/j.conbuildmat.2013.10.054.

[44] T. Ahmad Wani, S. Ganesh, Study on fresh properties, mechanical properties and microstructure behavior of fiber reinforced self compacting concrete: A review, Materials Today: Proceedings 62 (2022) 6663–6670. https://doi.org/10.1016/j.matpr.2022.04.666.

[45] Bibm, et al., The European Guidelines for Self-Compacting Concrete: Specification, Production and Use, 563, Eur. Guidel. Self-Compact. Concr. 22, (n.d.). http://www.efnarc.org/pdf/SCCGuidelinesMay2005.pdf.

[46] P.A. Sivanantham, G.G. Prabhu, G.G. Vimal Arokiaraj, K. Sunil, Effect of Fibre Aspect-Ratio on the Fresh and Strength Properties of Steel Fibre Reinforced Self-Compacting Concrete, Advances in Materials Science and Engineering 2022 (2022) 1–8. https://doi.org/10.1155/2022/1207273.

[47] V. Corinaldesi, G. Moriconi, Mechanical and thermal evaluation of Ultra High Performance Fiber Reinforced Concretes for engineering applications, Construction and Building Materials 26 (2012) 289–294. https://doi.org/10.1016/j.conbuildmat.2011.06.023.

[48] R. Rachid, A. Mohamed, R. Mohamed, O. Mohammed, Effect of additions on the self-compacting concrete's absorption, J. Eng. Exact Sci. 9 (2023) 16058–01e. https://doi.org/10.18540/jcecvl9iss6pp16058-01e.

Journal of Engineering and Technology for Industrial Applications

ITEGAM-JETIA

Manaus, v.11 n.55, p. 293-303. September/October, 2025.
DOI: https://doi.org/10.5935/jetia. v.11 n.55.2714

ISSN ONLINE: 2447-0228

RESEARCH ARTICLE OPEN ACCESS

COMPARATIVE ANALYSIS OF YOLOv3, MOBILENET-SSD, AND EFFICIENTDET FOR REAL-TIME PERSON DETECTION IN LOW-RESOLUTION IMAGES

Mhd. Idham Khalif[1] and Raden Deiny Mardian[2]

[1, 2] Electrical Engineering, Universitas Trisakti, Tomang, West Jakarta, Indonesia

[1]https://orcid.org/0009-0002-6911-9678, [2]https://orcid.org/0000-0001-5090-4604

Email: idham.khalif@trisakti.ac.id, deiny_wp@trisakti.ac.id

ARTICLE INFO

Article History
Received: September 19, 2025
Revised: October 7, 2025
Accepted: October 9, 2025
Published: October 31, 2025

Keywords:
Real-time detection,
Person detection,
Low-resolution camera,
ESP32-CAM

ABSTRACT

Object detection using low-resolution cameras is increasingly important for resource constrained and low-cost systems. This study evaluates the performance of three pre-trained machine learning models from the COCO dataset YOLOv3, MobileNet-SSD, and EfficientDet for human detection at three image resolutions: 160×120 (QQVGA), 320×240 (QVGA), and 640×480 (VGA). Low resolutions pose challenges due to limited visual and spatial details, which can reduce detection accuracy. Results show that YOLOv3 achieves the highest accuracy and consistency across all resolutions but demands higher computational resources, reflected in increased CPU usage, higher latency, and lower frame rates (FPS). MobileNet-SSD provides a balanced trade-off between accuracy and efficiency, making it suitable for real-time applications in power-constrained environments. EfficientDet, although the most lightweight in computation, shows lower detection accuracy. This study emphasizes empirical evaluation under real-world deployment constraints, following the zero-shot performance approach, where pre-trained models are applied directly to low-resolution edge data without retraining. The findings offer practical insights into the trade-offs among accuracy, efficiency, and latency, guiding model selection for IoT and edge computing applications in resource-limited settings. This work contributes to understanding suitable model choices for low-resolution human detection in real-world scenarios.

Copyright ©2025 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

The advancement of Artificial Intelligence (AI), Machine Learning (ML), and the Internet of Things (IoT) has accelerated rapidly, underpinning digital transformation across various sectors. These technologies support intelligent applications in embedded systems, automation, surveillance, and real-time decision-making, aligning with Indonesia's national strategy to accelerate digital transformation as outlined in Presidential Regulation No. 82 of 2023 [1].

Machine Learning enables systems to learn patterns from data and make decisions without explicit programming, supporting applications such as image classification, natural language processing, and anomaly detection. In IoT and edge computing, ML facilitates object detection, camera-based surveillance, and environmental monitoring, where decision intelligence contextualizes real-time data to support automated responses [2]. Its impact extends into economics, law, and healthcare, improving prediction, risk management, legal review, and personalized diagnosis [3].

In digital image processing, ML has significantly advanced automatic object detection and recognition, using techniques such as segmentation, classification, and feature extraction [4]. Popular models like You Only Look Once (YOLO), Single Shot MultiBox Detector (SSD), and Faster R-CNN offer different trade-offs between speed and accuracy [5]. Recent studies emphasize that low-resolution image recognition requires robust datasets and optimization strategies; for instance, YOLO can achieve real-time maritime object detection on edge computing when appropriately optimized [6], [7].

Low-resolution images (below 720p), such as QQVGA (160×120), QVGA (320×240), and VGA (640×480), challenge detection due to loss of fine details, yet they remain widely used in IoT devices and edge computing for power efficiency and bandwidth constraints [8-10]. Applying ML models under such constraints is challenging because most models require high-quality data and computational resources [11], [12]. Nevertheless, edge computing combined with low-resolution inputs can reduce inference latency and computational load while enabling real-time detection [12], [13].

Building on these insights, this study compares YOLOv3, MobileNet-SSD, and EfficientDet for real-time person detection using low-resolution images from an ESP32-CAM processed on an ARM-based edge computing. Evaluation metrics include average confidence, CPU utilization, frames per second (FPS), inference latency, and detection success, with lighting intensity as an additional factor. Unlike studies introducing new architectures, this work emphasizes empirical evaluation under resource-constrained conditions, reflecting real-world deployment and zero-shot performance scenarios [14].

II. MATERIAL AND METHODS

II.1 DATASET

The availability of datasets is a fundamental requirement for developing and applying computer vision-based methods [14]. The dataset used in this study is COCO (Common Objects in Context), which is one of the standard and widely adopted datasets for object detection tasks. The models employed in this research were pre-trained on COCO, with a specific focus on the person class to support the evaluation of human detection. Out of the 80 object categories available in COCO, this study utilized only the annotations for the person class, totaling 66,808 instances, in line with the research objective of emphasizing human detection in low-resolution images.

Figure 1: Sample dataset for person class images from the COCO dataset.
Source: [15].

Figure 1 shows a sample from the COCO dataset in the person class. This dataset provides a wide variety of poses, sizes, and backgrounds that reflect real world conditions, making it highly useful for training and evaluating person detection models in the context of edge computing and low-resolution cameras. By using only the person category, this study focuses on analyzing the performance of ML models for real-time human detection, reducing the complexity of multiclass classification and enabling more accurate evaluation of efficiency and accuracy under low-resolution conditions. The dataset also allows testing model robustness against variations in lighting, viewing angles, and object density, which are critical factors in implementing IoT-based camera systems.

II.2 MODEL PRE-TRAINED MACHINE LEARNING (ML)

The ML models used in this study are pre-trained models, which have been trained beforehand and can be implemented directly without the need for retraining. Pre-trained models not only unify semantic representations across multiple tasks and languages but also demonstrate high-level capabilities approaching human intelligence [16]. With this approach, the training and initial testing phases are unnecessary, allowing the study to focus directly on the implementation and evaluation of model performance in real time on edge computing with low-resolution inputs.

II.2.1 YOLOv3

YOLOv3 (You Only Look Once Version 3) is one of the most popular and widely used real-time object detection models in various computer vision applications. The model employs the Darknet-53 architecture as its backbone, consisting of 53 convolutional layers designed to balance accuracy and inference speed. As a one-stage detection model, YOLOv3 processes the entire image in a single inference, in contrast to two-stage approaches such as Faster R-CNN, making it highly efficient for real-time applications. YOLOv3 achieves high detection accuracy with fast inference time, making it suitable for real-time environments [17]. Previous studies have applied YOLOv3 for real-time object detection tasks, reporting strong performance; for example, with an input resolution of 512×512

pixels, the model achieved an accuracy of 74.1% [18] In the context of this study, YOLOv3 is used in the form of a pre-trained model trained on the COCO dataset. Despite its larger model size than the lightweight version, YOLOv3 remains relevant for mid-range edge computing devices due to its ability to detect objects with good precision.

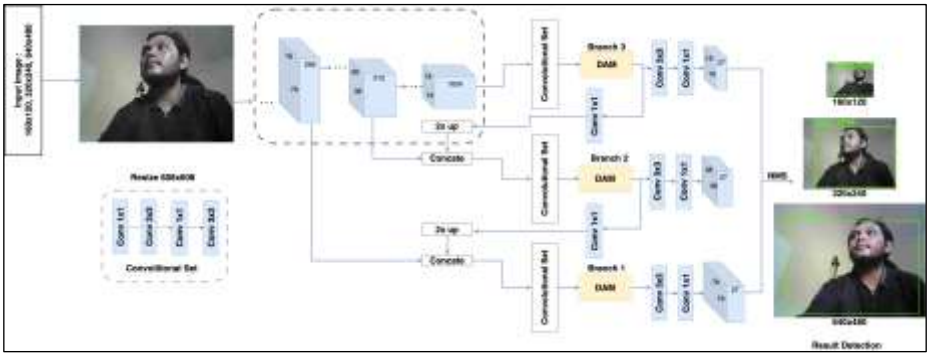


Figure 2: Architecture and result of YOLOv3 for human detection.
Source: Authors, (2025).

Figure 2 is illustrates the architecture and result of YOLOv3 applied in this study to detect human objects, starting from input images to final output results. The input images, originally in various resolutions such as 160×120 pixels, 320×240 pixels, or 640×480 pixels, are first uniformly resized to 608×608 pixels as part of a standard pre-processing step. These images are then processed by the network backbone, which extracts hierarchical features using a series of convolutional blocks composed of 1×1 and 3×3 filters, forming the basic convolutional sets. The network leverages the Feature Pyramid Network (FPN) principle by utilizing feature maps at three different scales—76×76 pixels, 38×38 pixels, and 19×19 pixels —to enable detection of objects at multiple sizes. Feature maps from deeper layers (low-resolution but rich in context) are up sampled and con-catenated with those from shallower layers (high-resolution but rich in spatial detail), enhancing the net-work's ability to extract comprehensive semantic and spatial information.

Following this multi-scale feature fusion, three detection branches (Branch 1, Branch 2, Branch 3) are employed, each equipped with convolutional layers and a Dilated Attention Module (DAM). The DAM helps the network focus on the most relevant areas, thereby improving the accuracy of bounding box pre-dictions, object confidence scores, and class probabilities. Finally, all predicted bounding boxes from the three branches are refined using Non-Maximum Suppression (NMS) to remove overlapping detections and retain the most relevant boxes for each detected object. The resulting output is a clear visualization of the detected human objects, along with the average confidence values for person detection using YOLOv3. Performance evaluations are based on confidence scores, mean Average Precision (mAP), and Precision-Recall (P-R) metrics for object detection [19].

II.2.2 MobileNet-SSD

MobileNet-SSD is the first TensorFlow-based deep learning model designed specifically for mobile devices. Due to its simpler and more computationally efficient architecture compared to other transfer learning models, such as VGG and ResNet, the name "Mobile" reflects its suitability for mobile applications [20]. In this study, human detection in images is performed using the MobileNet architecture combined with SSD (Single Shot MultiBox Detector) as the backbone of the object detection system. This combination enables the model to detect both the presence and location of humans in real-time, even on devices with limited computational resources. MobileNet-SSD, advantages in efficiency and speed make it highly suitable for various applications, including camera-based security surveillance systems, traffic monitoring, and augmented reality, all of which require fast and responsive human detection.

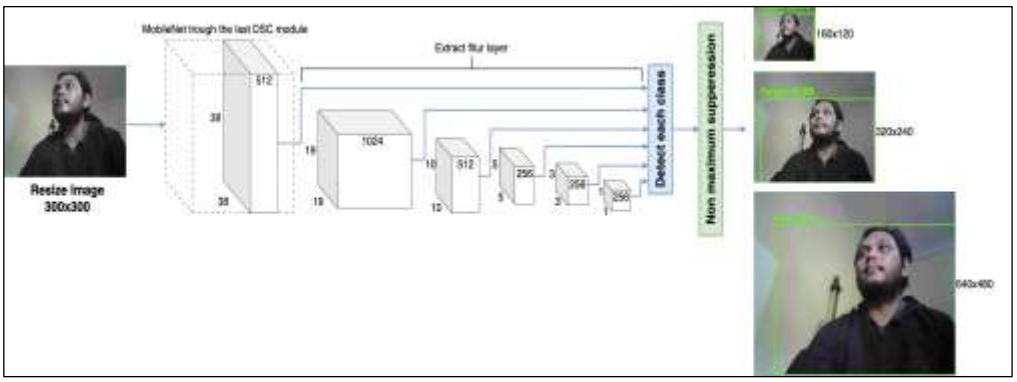


Figure 3: Architecture and results of MobileNet-SSD for human detection.
Source: Authors, (2025).

Figure 3, illustrates the workflow and results of the MobileNet-SSD architecture for performing object detection on devices with computational limitations, such as edge computing-based cameras. Similar to previous studies, three input resolutions—160×120 pixels, 320×240 pixels, and 640×480 pixels—are used. The process begins with resizing the input image to 300×300 pixels, which is then forwarded to the MobileNet-SSD backbone to extract visual features. MobileNet-SSD employs Depthwise Separable Convolution (DSC) up to the final block to produce a spatial feature representation of 38×38×512.

Furthermore, features from several layers (multiscale feature maps) are extracted at various spatial sizes—19×19, 10×10, 5×5, 3×3, and 1×1—with feature depths of 1024, 512, and 256 channels. This strategy enables MobileNet-SSD to perform object detection at multiple scales, from small to large objects, by utilizing high-resolution feature maps for spatial details and low-resolution feature maps for semantic context. Each of these features is then sent to the detection block (Detect Each Class), which predicts the bounding box location and object class directly in one step, without the region proposal stage used in Faster R-CNN. The detection results from all layers are combined and filtered through a Non Maximum Suppression (NMS) process to eliminate redundancy and retain only predictions with the highest confidence. The figure on the right shows the final result of human object detection.

II.2.3 EfficientDet

EfficientDet is a scalable and efficient object detection architecture that achieves state-of-the-art accuracy with significantly fewer parameters and FLOPs compared to previous detectors [21]. This model introduces two key innovations: BiFPN (Bidirectional Feature Pyramid Network), which enables efficient multi-scale feature fusion by allowing easy and weighted feature aggregation across different lev-els, and compound scaling, which systematically scales input resolution, network depth, and feature width in a balanced manner. These innovations allow EfficientDet to maintain high detection accuracy while substantially reducing computational demands, making it highly suitable for deployment in re-source-constrained environments, such as edge computing and IoT-based systems.

EfficientDet's design emphasizes both performance and efficiency, enabling it to operate effectively on devices with low power consumption and limited camera resolution. This makes it ideal for real-time ap-plications such as visual monitoring, autonomous vehicles, smart city infrastructure, and intelligent surveillance systems. Its ability to adapt to various hardware constraints without sacrificing detection quality has made it a preferred choice in many practical implementations. Moreover, EfficientDet supports a range of model sizes (D0 to D7), giving developers flexibility to choose the appropriate balance between speed and accuracy based on application requirements. This scalability ensures that EfficientDet can be optimized for both lightweight mobile devices and high-performance servers. Overall, EfficientDet represents a significant advancement in object detection, combining cutting-edge accuracy with practical efficiency for modern AI systems.

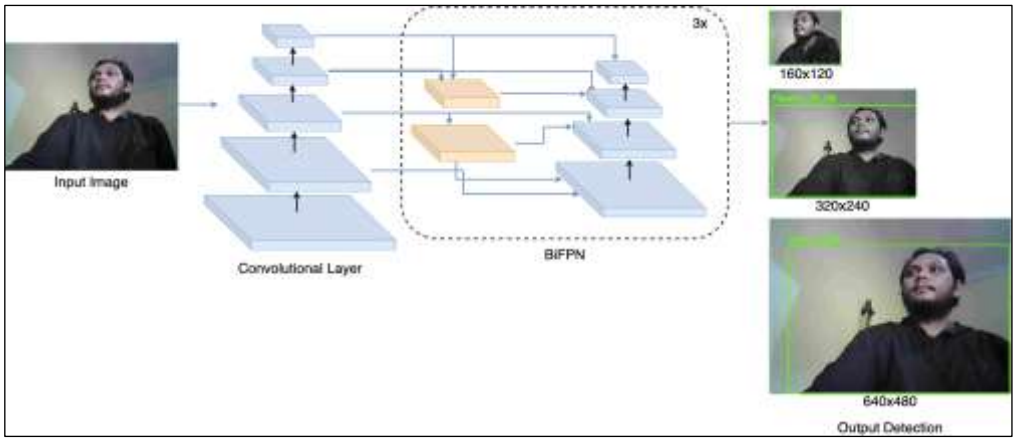


Figure 4: Architecture and results of EfficientDet for human detection.
Source: Authors, (2025).

Figure 4 shows the EfficientDet architecture and its results, demonstrating the model's ability to detect human objects efficiently and scalably through a combination of a lightweight backbone and an intelligent feature fusion mechanism. The process begins with an input image, which is then processed through several convolutional layers using the EfficientNet backbone to extract features at various resolution levels. These features are subsequently combined through the Bi-directional Feature Pyramid Network (BiFPN), a key component of the EfficientDet architecture that enables efficient feature fusion from top to bottom and vice versa.

BiFPN selects the most relevant feature information by weighting important information paths, resulting in a richer feature representation without significantly increasing computational complexity. After several iterations of BiFPN, the enriched features are used by the detection head to generate predictions of the locations and classes of objects in the image. In the architecture figure, the final result demonstrates successful object detection even when the input comes from a low-resolution camera, making EfficientDet highly suitable for applications in edge computing monitoring systems and IoT devices with limited power and computational resources.

II.3 MEASUREMENT AND HARDWARE SETUP

In this study, measurements and testing were carried out using real-time scenarios with low-resolution camera devices built on the ESP32-CAM. This device was chosen because it can transmit images directly via Wi-Fi, making it suitable for edge computing monitoring scenarios with limited resources. The camera captures images continuously, which are then analyzed using a lightweight ML-based object detection model to identify human presence in the test environment. Testing was conducted under various lighting conditions and perspectives to evaluate detection performance in situations that reflect real-world conditions. The system architecture is designed to maintain computing and power efficiency while providing accurate and responsive detection results in real time. The results of these tests are used to evaluate the effectiveness of the model in detecting objects from low-resolution images and to assess the stability of data transmission and overall system power efficiency.

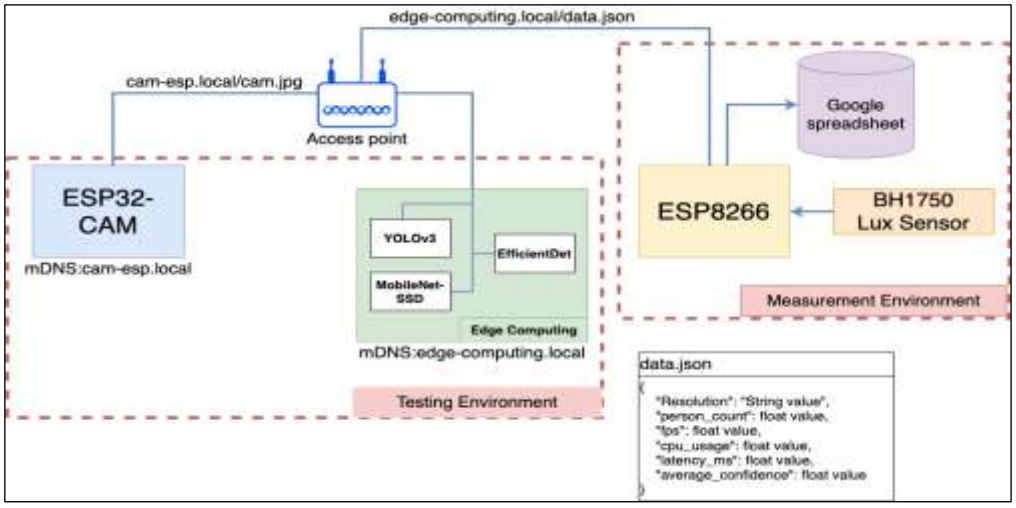


Figure 5: Architecture testing and measurement.
Source: Authors, (2025).

Figure 5, illustrates the architecture used in this study, consisting of two environments: testing and measurement. All devices are connected to a single access point within the same local network. In the testing environment, ESP32-CAM devices are configured to capture images in three resolutions: 160×120 (QQVGA), 320×240 (QVGA), and 640×480 (VGA), all in .JPG format. Edge computing devices are set up using the Multicast Domain Name System (mDNS) protocol, which allows dynamic device naming without relying on a unicast DNS server. mDNS simplifies device access even when IP addresses change across networks.

Captured images are continuously sent to the edge computing system, where they are processed using the selected ML models to detect human objects. Detection results are formatted in .JSON to facilitate communication between the testing and measurement systems. In the measurement environment, an ESP8266-based microcontroller measures ambient light intensity (lux) around the camera to evaluate the impact of lighting on model performance. Measurement data are transmitted to Google Spreadsheet via an internet connection in a time-series format for further analysis. Figure 6 is a show the setup of the testing device using ESP32-CAM and the measurement device using an ESP8266 microcontroller with a BH1750 lux sensor.

Figure 6: Camera and measurement device.
Source: Authors, (2025).

III. RESULT AND DISCUSSION

Testing was conducted in a controlled environment, specifically in a closed room with artificial lighting. The measurements showed that light intensity in the room ranged from 40 to 145 lux and was evenly distributed. Each test session lasted 30 minutes, with data collected periodically every minute. This procedure was repeated for each test scenario.

III.1 AVERAGE CONFIDENCE

During the deployment of object detection models, a confidence score threshold is set to reduce false positives and ensure that each predicted object is associated with an acceptable level of certainty [22]. This threshold is compared across ML models at various resolutions to assess stability and accuracy. In addition, tests measure human detection performance over a 30-minute period under resource constrained conditions.

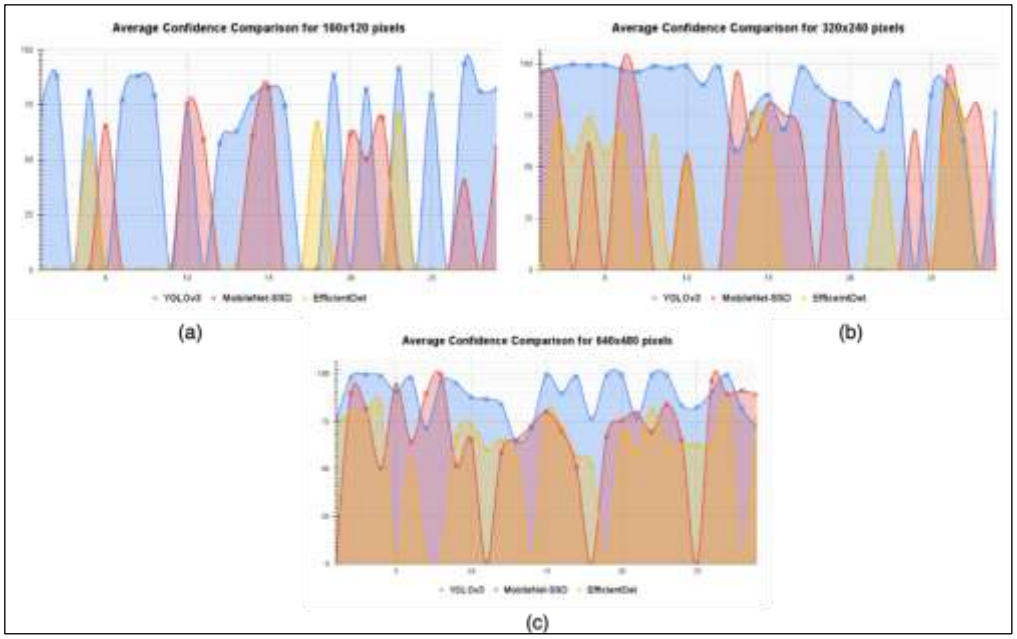


Figure 7: Comparison of average confidence values for each ML model: (a) 160×120 pixels (QQVGA). (b) 320×240 pixels (QVGA). (c) 640×480 pixels (VGA).
Source: Authors, (2025).

Figure 7a shows the results at 160×120 pixels (QQVGA). YOLOv3 achieved the highest average confidence (50.42%) and detection success rate (63.33%) during 30 minutes of testing. MobileNet-SSD followed with 20.7% confidence and a 33.33% success rate, while EfficientDet scored the lowest at 6.52% confidence and 10% success. These results indicate that YOLOv3, with its deep architecture and multiscale detection, can still recognize human objects reasonably well at very low resolutions. In contrast, MobileNet-SSD struggles due to limited feature extraction, while EfficientDet's lightweight design restricts its ability to detect small or low-quality objects. This highlights the importance of aligning model complexity with input quality, especially in low-resolution edge deployments.

At 320×240 pixels (QVGA), as shown in Figure 7b, YOLOv3 again led with 81.77% confidence and a 93.33% success rate. MobileNet-SSD achieved 44.04% confidence and 56.67% success, while EfficientDet reached 28.2% confidence and 43.33% success. YOLOv3's consistent performance stems from its robust feature extraction, making it effective even under varying lighting or object positions. MobileNet-SSD remains practical for real-time tasks but exhibits fluctuations in detection reliability. EfficientDet's lower performance reflects its trade-off between efficiency and feature depth, suggesting the need for model adaptation when applied to moderate-resolution inputs.

At 640×480 pixels (VGA), shown in Figure 7c, YOLOv3 maintained top performance with 88.9% confidence and a 100% success rate. MobileNet-SSD followed with 65.87% confidence and 86.67% success, while EfficientDet achieved 54.77% confidence and 80% success. These results confirm that YOLOv3 consistently delivers high detection accuracy across all resolutions, making it a strong candidate for applications that prioritize precision. MobileNet-SSD demonstrates reliable performance at medium and high resolutions, offering a good balance between accuracy and computational efficiency. In contrast, EfficientDet, while resource efficient, shows limited detection capabilities, particularly at lower resolutions. Model selection for edge compu-ting systems must therefore consider this trade-off: YOLOv3 for accuracy, EfficientDet for resource efficiency, and MobileNet-SSD as a middle ground. For detailed metrics, refer to Table 1.

Table 1: Summary of detection confidence metrics across models and resolutions.

Metric (Resolution)	YOLOv3 (%)	MobileNet-SSD (%)	EfficientDet (%)
Average (160 × 120)	50.42	20.07	6.52
Average (320 × 240)	81.77	44.04	28.2
Average (640 × 480)	88.9	65.82	54.77
Min (160 × 120)	00.00	00.00	00.00
Min (320 × 240)	00.00	00.00	00.00
Min (640 × 480)	64.62	00.00	00.00
Max (160 × 120)	93.46	82.92	70.7
Max (320 × 240)	99.92	98.56	85.16
Max (640 × 480)	99.87	99.73	83.98
Detection Success (160 × 120)	63.33	33.33	10.00
Detection Success (320 × 240)	93.33	56.67	43.33
Detection Success (640 × 480)	100.00	86.67	80.00

Source: Authors, (2025).

III.2 CPU USAGE

CPU Usage measures processor workload during inference, indicating computing power used in real-time detection. High usage risks delays, especially on edge computing device. Low usage improves efficiency and stability. Analyzing CPU usage is vital for model feasibility. In cloud systems, it ensures service quality [23].

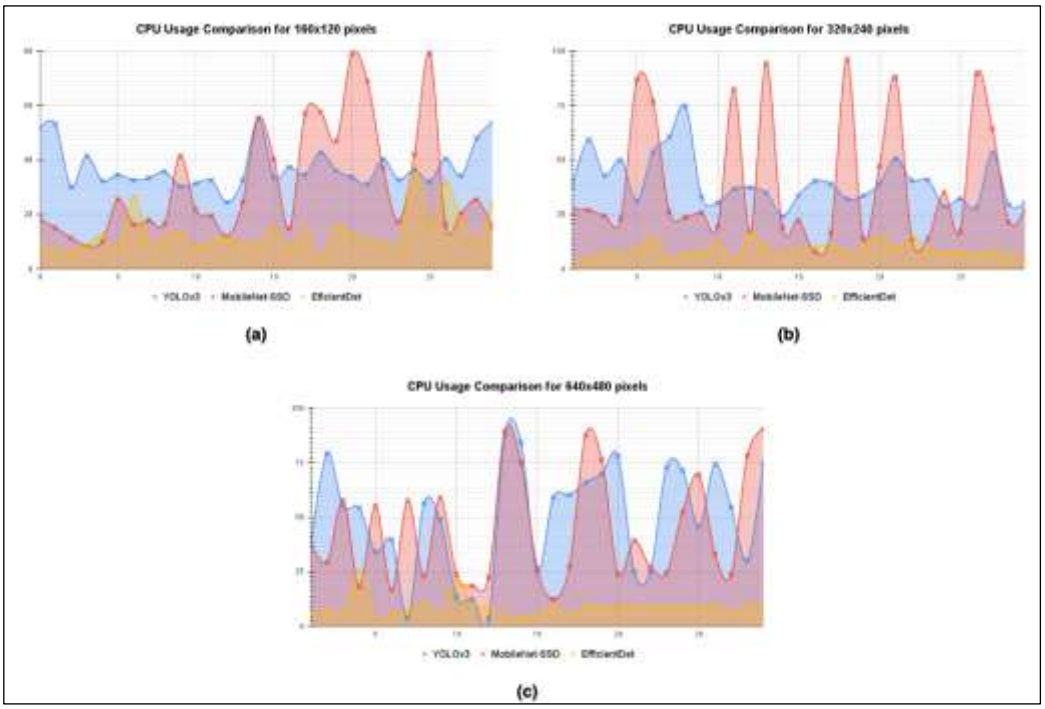


Figure 8: Comparison of CPU usage values for each ML model: (a) 160×120 pixels (QQVGA). (b) 320×240 pixels (QVGA). (c) 640×480 pixels (VGA).
Source: Authors, (2025).

Figure 8a shows the average CPU usage at 160×120 resolution. YOLOv3 recorded the highest usage (37.28%), followed by MobileNet-SSD (31.04%) and EfficientDet (13.23%). These results reflect the architectural complexity of each model. Despite the low resolution, YOLOv3 demands significant resources due to its multiscale detection and modules such as the Feature Pyramid Network and Non Maximum Suppression. MobileNet-SSD also incurs a moderate load from its multi-scale feature extraction, while EfficientDet remains the most efficient thanks to its lightweight structure and adaptive BiFPN-based feature fusion. At 320×240 resolution Figure 8b, YOLOv3 and MobileNet-SSD show increased CPU usage—40.46% and 39.43%, respectively—while EfficientDet remains steady at 13.23%. The rise in resource consumption for YOLOv3 and MobileNet - SSD is attributed to their deeper architectures and larger parameter sets, which increase the per-frame computational load. EfficientDet's consistently low CPU usage highlights its efficiency, largely due to EfficientNet and BiFPN. However, this efficiency may come at the cost of reduced detection accuracy, which must be considered in application design.

At 640×480 resolution Figure c, YOLOv3 peaks at 50.7% CPU usage, MobileNet-SSD at 43.58%, and EfficientDet drops further to 8.47%. Higher resolutions amplify computational demands, especially for YOLOv3, which employs complex processing to maintain detection precision. MobileNet-SSD also shows increased load but remains more efficient than YOLOv3. EfficientDet consistently consumes minimal CPU, though with potential trade-offs in accuracy. Overall, the results indicate that more accurate and complex models require significantly greater computational resources. YOLOv3 offers strong detection performance but at a high CPU cost, making it suitable for more powerful systems. EfficientDet delivers excellent efficiency with moderate accuracy, making it ideal for resource-constrained environments. MobileNet-SSD strikes a balance between detection quality and CPU load. Detailed CPU comparisons are provided in Table 2.

Table 2: Summary of CPU usage metrics across models and resolutions.

Metric (Resolution)	YOLOv3 (%)	MobileNet-SSD (%)	EfficientDet (%)
Average (160×120)	37.28	31.04	13.23
Average (320×240)	40.46	39.43	13.23
Average (640×480)	50.7	43.58	8.47
Min (160×120)	24.4	8.5	3.4
Min (320×240)	24.2	8.5	3.4
Min (640×480)	3.5	12.03	3.5
Max (160×120)	55.1	79.1	37.8
Max (320×240)	74.7	96.03	16.00
Max (640×480)	87.9	90.04	80.00

Source: Authors, (2025).

III.3 FRAME PER SECOND (FPS)

FPS indicates the number of frames a model processes per second, reflecting both inference speed and system efficiency. Object detection performance is commonly evaluated using FPS and FLOPs [24]. A higher FPS enhances video quality, and to achieve smooth playback at 10 FPS, processing must be completed within 100 ms [25].

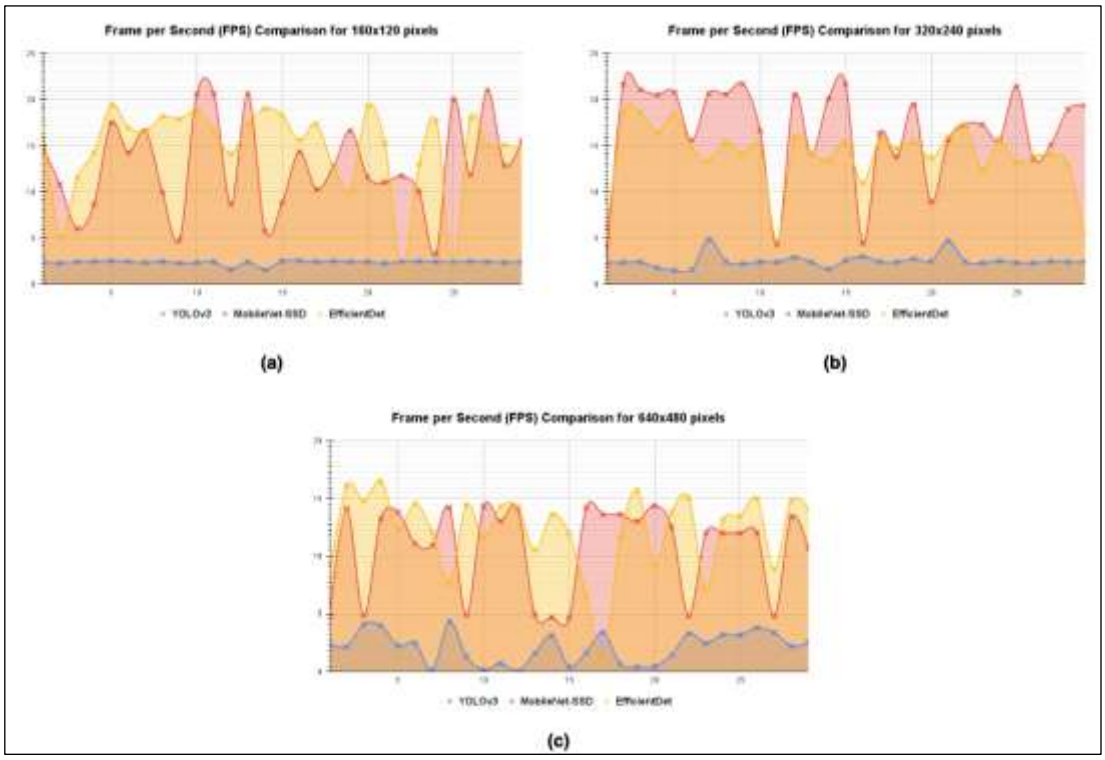


Figure 9: Comparison of FPS values for each ML model: (a) 160×120 pixels (QQVGA). (b) 320×240 pixels (QVGA). (c) 640×480 pixels (VGA).
Source: Authors, (2025).

Figure 9a compares the Frames Per Second (FPS) performance of YOLOv3, MobileNet-SSD, and EfficientDet at a resolution of 160×120. EfficientDet achieved the highest FPS (14.96), followed by MobileNet-SSD (12.98), while YOLOv3 lagged significantly (2.37). These results highlight EfficientDet's superior efficiency on low-resolution inputs due to its lightweight architecture and feature fusion mechanism. Although MobileNet-SSD is also efficient, its performance is slightly lower, likely due to its mul-ti-scale layer complexity. In contrast, YOLOv3's low FPS suggests that it is unsuitable for real-time applications on resource constrained devices, despite its high accuracy.

At 320×240 resolution (Figure 9b), MobileNet-SSD recorded the best performance (16.07 FPS), followed by EfficientDet (13.84 FPS) and YOLOv3 (2.52 FPS). These results confirm MobileNet-SSD's advantage in real-time processing, particularly for mid-resolution scenarios. While EfficientDet remains competitive, YOLOv3's relatively low FPS again underscores its heavier computational load, limiting its applicability in time critical edge computing systems.

At 640×480 resolution (Figure 9c), MobileNet-SSD maintained the lead with 12.27 FPS, followed by EfficientDet (10.54 FPS) and YOLOv3 (2.11 FPS). The overall decrease in FPS across all models re-flects the increased processing demands at higher resolutions. YOLOv3 consistently delivered the lowest FPS, though it may still be applicable in scenarios where real-time video output is not required. Overall, MobileNet-SSD and EfficientDet demonstrate better scalability in edge computing deployments. Detailed performance metrics are presented in Table 3.

Table 3: Summary of FPS metrics across models and resolutions.

Metric (Resolution)	YOLOv3 (FPS)	MobileNet-SSD (FPS)	EfficientDet (FPS)
Average (160×120)	2.37	12.98	14.96
Average (320×240)	2.52	16.07	13.84
Average (640×480)	2.11	12.27	10.54
Min (160×120)	1.54	3.31	1.52
Min (320×240)	1.51	2.06	4.67
Min (640×480)	0.01	1.27	4.67
Max (160×120)	2.56	21.02	19.45
Max (320×240)	4.87	21.71	18.60
Max (640×480)	4.34	16.5	14.40

Source: Authors, (2025).

III.4 LATENCY

Latency plays a crucial role in evaluating the performance of inference models, as it measures the time required to generate output from input. In the context of human detection, accuracy alone is insufficient; models must also operate in a timely manner. Deep learning models often face stringent latency constraints [26]. Lower latency significantly enhances the feasibility of deploying models in real-time systems, such as video streaming, which demands instantaneous responses, or edge computing, where processing must occur near the data source without significant delays. Therefore, latency optimiza-tion is a key factor in the design and selection of models for applications that require high speed and responsiveness.

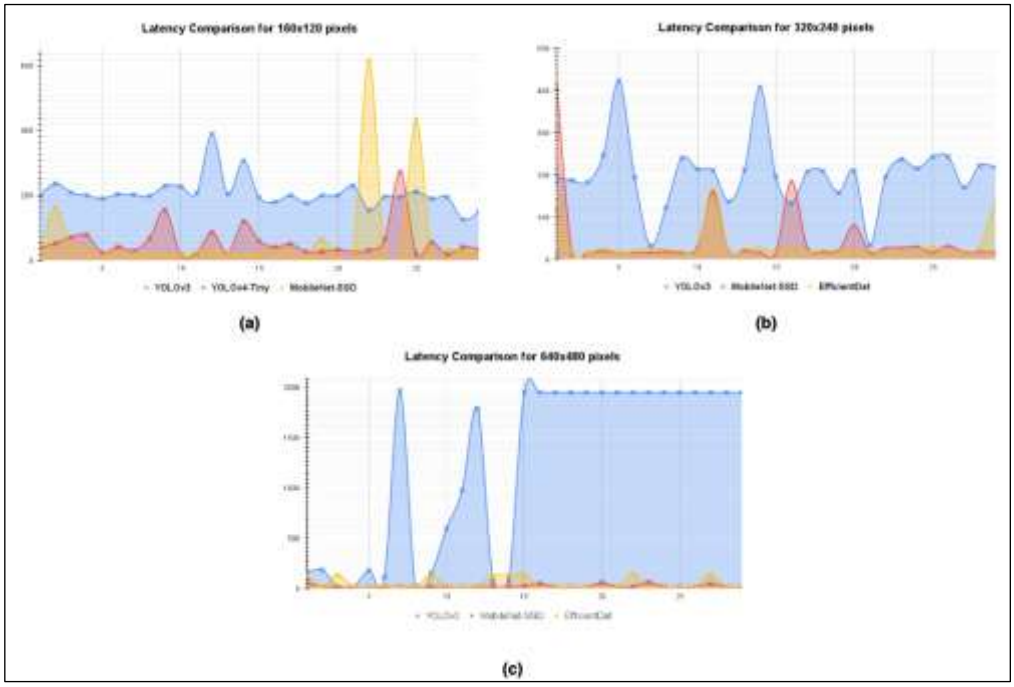


Figure 10: Comparison of latency values for each ML model: (a) 160×120 pixels (QQVGA). (b) 320×240 pixels (QVGA). (c) 640×480 pixels (VGA).
Source: Authors, (2025).

Figure 10a presents the latency results at 160×120 resolution, defined as the average time required to generate a detection output from a single image. MobileNet-SSD achieves the best performance with a latency of 55.71 ms, enabling efficient near real-time processing. EfficientDet follows with 63.01 ms, which remains suitable for responsive applications. In contrast, YOLOv3 records the highest latency at 207.85 ms, reflecting its heavier computational load. These results indicate that for low-resolution tasks requiring fast responses, MobileNet-SSD and EfficientDet are more suitable than YOLOv3.

At 320×240 resolution (Figure 10b), EfficientDet leads with the lowest latency of 35.46 ms an improvement compared to its performance at lower resolution—demonstrating strong scalability. MobileNet-SSD follows with 58.42 ms, maintaining consistent results. YOLOv3 again shows the highest latency at 202.24 ms, reinforcing its limitations for time-critical tasks. This resolution-dependent shift emphasizes the importance of evaluating detection models across different input sizes for optimal de-ployment.

At 640×480 resolution (Figure 10c), MobileNet-SSD achieves the lowest latency at 24.86 ms, followed by EfficientDet at 53.35 ms. YOLOv3, however, shows a dramatic increase to 1194.34 ms far beyond real-time thresholds. This trend confirms that although YOLOv3 may provide high accuracy, it is not latency-efficient, particularly at higher resolutions.

Overall, MobileNet-SSD performs best at low and high resolutions, while EfficientDet excels at medium resolution. Both models are better suited for real-time edge computing applications that demand fast response and energy efficiency. Detailed latency comparisons are presented in Table 4.

Table 4: Summary of latency metrics across models and resolutions.

Metric (Resolution)	YOLOv3 (ms)	MobileNet-SSD (ms)	EfficientDet (ms)
Average (160×120)	207.85	55.71	63.01
Average (320×240)	202.24	58.42	35.46
Average (640×480)	1194.34	25.86	53.35
Min (160×120)	126.31	16.77	15.96
Min (320×240)	33.21	16.07	16.19
Min (640×480)	17.22	16.42	15.96
Max (160×120)	390.13	272.35	615.36
Max (320×240)	424.4	438.1	170.43
Max (640×480)	1972.02	73.01	145.21

Source: Authors, (2025).

V. CONCLUSIONS

This study successfully applied and compared three machine learning models—YOLOv3, MobileNet-SSD, and EfficientDet—for human detection using ESP32-CAM devices with edge computing, based on Ubuntu ARM running on virtual machines. Testing was conducted at three resolutions: 160×120 pixels (QQVGA), 320×240 pixels (QVGA), and 640×480 pixels (VGA), under controlled lighting conditions (40–150 lux). The results show that YOLOv3 is superior in accuracy and detection consistency but has weaknesses in processing efficiency, with high CPU usage, latency, and low FPS. MobileNet-SSD offers a balance between accuracy and efficiency, making it an ideal choice for real-time systems with limited resources. EfficientDet is the most lightweight in terms of computation, suitable for edge computing de-vices with limited power, although at the cost of lower accuracy. This study recommends selecting mod-els based on system requirements: YOLOv3 for high accuracy, MobileNet-SSD for balanced performance, and EfficientDet for resource-constrained systems. Further research is suggested to evaluate these models on specialized edge computing devices such as NVIDIA Jetson, as well as to explore custom-trained lightweight models and assess their performance on edge computing such as NVIDIA Jetson or Coral TPU.

VIII. REFERENCES

[1] Government of Indonesia, "Presidential Regulation Number 82 of 2023 on the Acceleration of Digital Transformation and Integration of National Digital Services," 2023. [Online]. Available: https://peraturan.bpk.go.id/Details/279013/perpres-no-82-tahun-2023

[2] P. M. Jeyanthi, D. H. Polay, and T. Choudhury, "The Rise of Decision Intelligence: AI That Optimizes Decision-Making," in Decision Intelligence Analytics and the Implementation of Strategic Business Management, P. M. Jeyanthi, T. Choudhury, D. Hack-Polay, T. P. Singh, and S. Abujar, Eds. Cham, Switzerland: Springer, 2022. doi: 10.1007/978-3-030-82763-2_7.

[3] A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019. doi: 10.1056/NEJMra1814259.

[4] D. B. Olawade, O. Z. Wada, A. O. Ige, B. I. Egbewole, A. Olojo, and B. I. Oladapo, "Artificial intelligence in environmental monitoring: Advancements, challenges, and future directions," Hygiene and Environmental Health Advances, vol. 12, p. 100114, 2024. doi: 10.1016/j.heha.2024.100114.

[5] B. Bhavya Sree, V. Yashwanth Bharadwaj, and N. Neelima, "An inter-comparative survey on state-of-the-art detectors—R-CNN, YOLO, and SSD," in Intelligent Manufacturing and Energy Sustainability, A. Reddy, D. Marla, M. N. Favorskaya, and S. C. Satapathy, Eds., vol. 213, Smart Innovation, Systems and Technologies. Singapore: Springer, 2021, pp. 555–567. doi: 10.1007/978-981-33-4443-3_46.

[6] P. Dandekar, S. S. Aote, and A. Raipurkar, "Low-resolution face recognition: Review, challenges and research directions," Computers and Electrical Engineering, vol. 120, p. 109846, 2024. doi: 10.1016/j.compeleceng.2024.109846.

[7] D. Heller, M. Rizk, R. Douguet, A. Baghdadi, and J.-P. Diguet, "Marine objects detection using deep learning on embedded edge devices," in Proc. 2022 IEEE International Workshop on Rapid System Prototyping (RSP), 2022, pp. 1–7. doi: 10.1109/RSP57251.2022.10039025.

[8] R. Sunkara and T. Luo, "No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects," in Lecture Notes in Computer Science, vol. 13715. Springer, 2023. doi: 10.1007/978-3-031-26409-2_27.

[9] J. Diez-Tomillo, I. Martinez-Alpiste, G. Golcarenarenji, A. Muñoz, D. Moreno, and R. Alvarez, "Efficient CNN-based low-resolution facial detection from UAVs," Neural Computing and Applications, vol. 36, pp. 5847–5860, 2024. doi: 10.1007/s00521-023-09401-3.

[10] C. Xie, D. J. Pagliari, and A. Calimera, "Energy-efficient and privacy-aware social distance monitoring with low-resolution infrared sensors and adaptive inference," in Proc. 2022 17th Conference on Ph.D. Research in Microelectronics and Electronics (PRIME), 2022, pp. 181–184. doi: 10.1109/PRIME55000.2022.9816801.

[11] P. Mittal, "A comprehensive survey of deep learning-based lightweight object detection models for edge devices," Artificial Intelligence Review, vol. 57, p. 242, 2024. doi: 10.1007/s10462-024-10877-1.

[12] C. Avasalcai and S. Dustdar, "Edge computing: Use cases and research challenges," in Edge Computing: Use Cases and Research Challenges, B. Vogel-Heuser and M. Wimmer, Eds., Springer, 2023, pp. 125–142. doi: 10.1007/978-3-662-65004-2_5.

[13] L. Xu, Y. Zhao, Y. Zhai, W. Zhang, and N. Liu, "Small object detection in UAV images based on YOLOv8n," International Journal of Computational Intelligence Systems, vol. 17, p. 223, 2024. doi: 10.1007/s44196-024-00632-3.

[14] M. Desai, H. Mewada, I. M. Pires, and S. Roy, "Evaluating the performance of the YOLO object detection framework on COCO dataset and real-world scenarios," Procedia Computer Science, vol. 251, pp. 157–163, 2024. doi: 10.1016/j.procs.2024.11.096.

[15] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common objects in context," in Computer Vision – ECCV 2014, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., Lecture Notes in Computer Science, vol. 8693, Cham: Springer, 2014, pp. 740–755. doi: 10.1007/978-3-319-10602-1_48.

[16] Y. Lin, N. Ding, Z. Liu, and M. Sun, "Pre-trained models for representation learning," in Pre-trained Models for Representation Learning, Z. Liu, Y. Lin, and M. Sun, Eds. Cham: Springer, 2023, pp. 127–167. doi: 10.1007/978-981-99-1600-9_5.

[17] A. Vijayakumar and S. Vairavasundaram, "YOLO-based object detection models: A review and its applications," Multimedia Tools and Applications, vol. 83, pp. 83535–83574, 2024. doi: 10.1007/s11042-024-18872-y.

[18] Y. Zhou, H. Qian, and P. Ding, "Lite-YOLOv3: A real-time object detector based on multi-scale slice depthwise convolution and lightweight attention mechanism," Journal of Real-Time Image Processing, vol. 20, p. 123, 2023. doi: 10.1007/s11554-023-01379-4.

[19] Y. Zhou, H. Qian, and P. Ding, "Lite-YOLOv3: A real-time object detector based on multi-scale slice depthwise convolution and lightweight attention mechanism," Journal of Real-Time Image Processing*, vol. 20, p. 123, 2023. doi: 10.1007/s11554-023-01379-4.

[20] M. A. B. Abbass and Y. Ban, "MobileNet-based architecture for distracted human driver detection of autonomous cars," Electronics, vol. 13, no. 2, 2024. doi: 10.3390/electronics13020365.

[21] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020, pp. 10778–10787. doi: 10.1109/CVPR42600.2020.01079.

[22] L. Peng, H. Wang, and J. Li, "Uncertainty evaluation of object detection algorithms for autonomous vehicles," Automotive Innovation, vol. 4, no. 3, pp. 241–252, 2021. doi: 10.1007/s42154-021-00154-0.

[23] R. Estrada, I. Valeriano, and X. Aizaga, "CPU usage prediction model: A simplified VM clustering approach," in Complex, Intelligent and Software Intensive Systems, L. Barolli, Ed. Cham: Springer, 2023, pp. 210–221. doi: 10.1007/978-3-031-35734-3_21.

[24] W. Chen, J. Luo, F. Zhang, Y. Liu, C. Li, and J. Chen, "A review of object detection: Datasets, performance evaluation, architecture, applications and current trends," Multimedia Tools and Applications, vol. 83, pp. 65603–65661, 2024. doi: 10.1007/s11042-023-17949-4.

[25] D. Xu, H. Tao, J. Yu, and C. Xiao, "Real-time multi-camera video stitching based on improved optimal stitch line and multi-resolution fusion," in Image and Graphics (ICIG 2017), Y. Zhao, X. Kong, and D. Taubman, Eds., Lecture Notes in Computer Science, vol. 10668. Cham: Springer, 2017. doi: 10.1007/978-3-319-71598-8_12.

[26] H. She, Y. Luo, Z. Xiang, W. Liang, and Y. Xie, "Accurate latency prediction of deep learning model inference under dynamic runtime resource," in Neural Information Processing, B. Luo, L. Cheng, Z.-G. Wu, H. Li, and C. Li, Eds. Cham: Springer, 2024, pp. 495–510. doi: 10.1007/978-981-99-8126-7_39.