Proceedings

ICTIS '25

INTERNATIONAL CONFERENCE ON

INNOVATIONS IN COMPUTING TECHNOLOGIES AND INFORMATION SCIENCES (ICTIS)

29TH & 30TH APRiL, 2025.

Preface

With immense pleasure, I am honored to present the Proceedings for the 2nd International Conference on Innovations in Computing Technologies and Information Sciences (ICTIS) 2025 as the Conference Chief Organizer. Hosted by the esteemed UET Peshawar, this conference marks a significant milestone in the realm of computing and information sciences.

The primary objective of this conference is to provide a dynamic platform for the convergence of scholars, researchers, academicians, and industry professionals from around the globe. Our focus on cutting-edge innovations in computing technologies and information sciences aims to foster interdisciplinary collaborations and propel advancements in these rapidly evolving fields.

The conference themes span a wide spectrum, including but not limited to Artificial Intelligence, Machine Learning, Data Science, Cybersecurity, Internet of Things (IoT), and Blockchain Technology. By facilitating dialogue and knowledge exchange on these topics, ICTIS 2025 seeks to catalyze innovation and drive positive impact across academia, industry, and society.

UET Peshawar, renowned for its academic excellence and commitment to research, serves as an ideal host for this prestigious event. With dedicated qualified faculty, and enthusiastic students, UET Peshawar continues to be a beacon of innovation in the region.

I extend my heartfelt gratitude to all the keynote speakers, session chairs, and authors whose invaluable contributions have enriched the conference proceedings. I would also like to express sincere appreciation to our technical and financial sponsors for their generous support, without which this event would not have been possible.

Lastly, I commend the diligent efforts of all my colleagues on the organizing committees for their unwavering commitment and dedication to the successful execution of ICTIS 2025. Together, we have created a platform that not only fosters intellectual discourse but also paves the way for groundbreaking innovations in computing technologies and information sciences.

Thank you to everyone who has contributed to making ICTIS 2025 a resounding success. Your passion and dedication inspire us to continue pushing the boundaries of knowledge and innovation.

Dr. Nasru Minallah

Chief Organizer

Organizing Committee

Honorary Chairs

Prof. Qaiser Ali, Vice Chancellor, University of Engineering and Technology, Peshawar, Pakistan

Prof. Syed Waqar Shah, Dean, Faculty of Electrical and Computer Engineering, University of Engineering and Technology, Peshawar, Pakistan

Program Chair

Prof. Laiq Hasan, University of Engineering and Technology, Peshawar, Pakistan

Organizing Chair

Dr. Nasru Minallah, University of Engineering and Technology, Peshawar, Pakistan

Technical Chairs

Dr. Nasir Ahmad, University of Engineering and Technology, Peshawar, Pakistan

Dr. Sohail Imran, IQRA National University, Peshawar, Pakistan

Dr. Khurram Khattak, University of Engineering and Technology, Peshawar, Pakistan

Dr. Safdar Marwat, University of Engineering and Technology, Peshawar, Pakistan

Dr. Sheeraz Ahmed, IQRA National University, Peshawar, Pakistan

Dr. Ihsan Ul Haq, University of Engineering and Technology, Peshawar, Pakistan

Dr. Arbab Masood, University of Engineering and Technology, Peshawar, Pakistan

Dr. Amjad Ali, University of Engineering and Technology, Peshawar, Pakistan

Dr. Amaad Khalil, University of Engineering and Technology, Peshawar, Pakistan

Dr. Muniba Ashfaq, University of Engineering and Technology, Peshawar, Pakistan

Dr. Yasir Afridi, University of Engineering and Technology, Peshawar, Pakistan

Dr. Abeer Irfan, University of Engineering and Technology, Peshawar, Pakistan

Dr. Athar Javed Sethi, University of Engineering and Technology, Peshawar, Pakistan

Dr. Madiha Sher, University of Engineering and Technology, Peshawar, Pakistan

Dr. Rehmat Ullah Khattak, University of Engineering and Technology, Peshawar, Pakistan

Engr. Sumayyea Salahuddin, University of Engineering and Technology, Peshawar, Pakistan

Industry Chairs

Engr. Mohsin Tariq, Machine Learning Engineer, Turing

Engr. Muhammad Zeeshan, Web & App Developer, NCBC UET Peshawar

Mr. Tufail Ahmad, Blockchain Developer, NCBC UET Peshawar

Keynote Speakers Sessions

Session1:

Title: **Smart Disaster Prevention and Emergency Management of Urban Major Infrastructure**
Keynote Speaker: **Dr. Yeshou Xu**, Lecturer at South East University, Nanjing China

Session2:

Title: **Constrained Multi-Objective Optimization of Mass Dampers for Vortex-Induced Vibration Control in Bridges**
Keynote Speaker: **Dr. Wen-Peng Xu**, Postdoctoral Fellow at South East university China

Session3:

Title: **Study on Mutli-Hazard Effects and Hybrid Simulation of Building Structures under Earthquake-Fire Conditions**
Keynote Speaker: **Dr. Ya Xin Wei,** Associate Professor at Tsinghua University, China

Session4:

Title: **Investigation of a Multi-Dimensional Seismic Isolation and Mitigation Device in Precast Wall Panel Structures**
Keynote Speaker: **Dr. Zhong Wei Hu,** Professor at Huaqiao University, China

CPD Sessions:

Session1:

Title: **Digital Twin & AI in Dentistry -**
Innovations in AI-Driven Assistive Technologies
Resource Person: **Dr. Shahzad Anwar,** Principal Investigator at Artificial Intelligence in Healthcare (AIH) Lab

Session2:

Title: **Exploring Solar Technologies**
Resource Person: **Engr. Saleem Barg,** CEO Barg Engineering Consultant

Contents

Modified Incremental Conductance-Based Active Power Curtailment for Voltage Regulation in Highly PV-Fed Distribution Grids

Abdullah Said
University of Engineering and Technology
Department of Power Engineering
Peshawar, Pakistan
16abelt0693@uetpeshawar.edu.pk

Dr. Gul Muhammad
University of Engineering and Technology
Department of Power Engineering
Peshawar, Pakistan
gk502@uetpeshawar.edu.pk

Muhammad Zakir
Dr. Khan Shaheed Degree College
Department of Computer Science
Kabal, Swat, Pakistan
engr.muhammadzakirr@gmail.com

***Abstract*—Increasing photovoltaic (PV) penetration in the distribution grid presents a challenge of voltage rise and even violation in low load and peak PV generation. This paper presents a modified incremental conductance (IC) method to limit the active power injection of PV systems into the distribution grid to mitigate voltage violations. A voltage-sensitive control loop is incorporated into the IC to enhance its performance. This algorithm enables dynamic adjustment of active power output based on real-time voltage measurements of the grid, which facilitates precise voltage regulation while maintaining maximum power point tracking (MPPT) under normal operating conditions, shifting the operating point away from MPP to limit active power injection and keeping voltage within limits. The proposed algorithm has reduced complexity, is economical, and can be implemented in existing inverters with the IC algorithm. The algorithm is validated through Simulink/MATLAB simulations. The setup consists of a distribution grid, penetrated with two solar PV-based distributed generator (DG) having a capacity of 100kW each.**

***Index Terms*—Modified Incremental Conductance, PV Penetration, Voltage Violation, Active Power Curtailment, MPPT**

I. Introduction

The increased PV penetration in the distribution grid has generated substantial benefits, such as decreased dependence on existing energy sources, reduced reliance on fossil fuel-based energy sources, and a reduction in greenhouse gas emissions [1]. An increase in Rooftop PVDG's integration to the grid has been observed around the world [2]. Despite their diverse benefits, high PV penetration in distribution grids has led to challenges such as voltage violation in low-voltage networks, mainly driven by bidirectional power flow [3]. Voltage violation mainly happens when peak PV generation does not coincide with peak load, and it is the main cause that puts a limit on PV penetration [4]. This rise in voltage is mainly due to the impedance of the feeder; in a greater impedance line, reverse power will lead to exceeding its regulatory limits [5]. Recent studies propose diverse mitigation strategies, such as feeder enhancement, on-load tap changing OLTC transformer, battery storage system BSS, active power curtailment APC, reactive power control RPC, machine learning-based solutions, or a combination of any two of these methods [6].

Studies in [7] suggest replacing the existing feeder with a conductor of high ampacity ratings and small impedance, this feeder will allow more PV penetration, as voltage violations depend on feeder impedance. On-load tap changers (OLTC) are another method that dynamically adjusts the turns ratio between transformer primary and secondary windings to regulate voltage, as presented in [8].

Battery energy storage systems (BESS) have emerged as an effective solution for mitigating voltage violations caused by high photovoltaic (PV) penetration in distribution grids. BESS can regulate voltage rise during peak PV generation and voltage drop during peak load periods through coordinated charging and discharging strategies [9]. These systems can effectively maintain voltage levels within acceptable limits, even at high PV penetration levels of 80-100% [10]. In [11], several autonomous PV inverter control techniques were compared, and an improved hybrid control strategy to mitigate voltage rise in LV grids was proposed. The authors of [12] developed a strategy for coordinated control of battery energy storage systems (BESSs) and OLTC-equipped transformers to mitigate the voltage deviation from its required values in PV-rich distribution networks. The above-mentioned methods of feeder enhancement, OLTC, BESS, or a combination of these, needs changes in the current existing system, which is not cost-effective and economical.

Other methods, such as reactive power control, have shown superior performance in voltage regulation, necessitating a lower demand for reactive power compensation [13]. A novel distributed Volt/Var control method is proposed to minimize the reactive power burden on substations and enhance the power factor in [14]. Additionally, a local voltage control strategy based on PV inverters' reactive power capability has been developed, utilizing a Q-V-based control curve with tuning parameters for faster response [15]. These strategies effectively mitigate voltage violations in distribution networks with high PV penetration.

Active power curtailment (APC) is an effective strategy for mitigating voltage violations caused by high photovoltaic (PV) penetration in distribution grids [6, 16]. APC can be implemented using smart PV inverters to reduce active power output

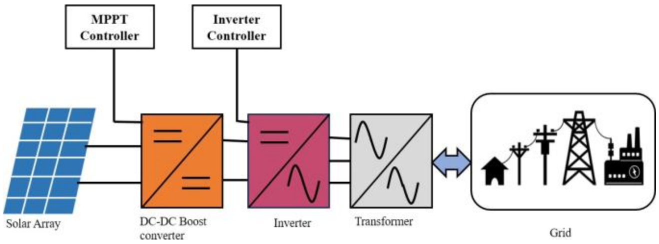


Figure 1: Block diagram of a PV fed microgrid integration with the distribution grid

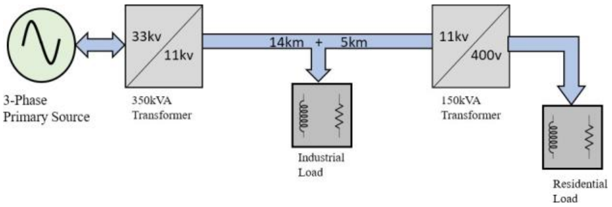


Figure 2: Block diagram of the distribution grid

based on control signals, improving voltage regulation and thermal feeder profiles [17]. Short-term PV power forecasts can be utilized to predict and prevent imminent overvoltage scenarios, allowing for adaptive APC implementation.

In this article, an incremental conductance method is modified to curtail active power flow in case of voltage violation, which mostly happens when the peak load on the feeder does not coincide with peak PV power generation. The study provides a simple and cost-effective solution for power control in case of a voltage rise. Next, Section II explains the model development, Section III presents and explains the proposed algorithm, and Section IV is devoted to results discussions, followed by the conclusions in Section V.

II. Model Development

The Simulations are performed on a MATLAB/SIMULINK environment, which offers a robust and adaptable platform for dynamic simulation. This platform facilitates the development of a comprehensive model that effectively replicates the empirical dynamics of the system under investigation. This investigation focuses on the development of a holistic model representing the grid-coupled Photovoltaic (PV) System. The system is modeled to mirror interaction between a PVDG and the distribution grid. The structural overview of the simulated system is shown in Figure 1. The distribution grid model developed for this case study is shown in Figure 2, contains a primary energy source, a three-phase transformer with a power rating of 350 kVA, a voltage rating of 33 kV/11 kV, and an efficiency of 99 %. The 11 kV ACSR (50 mm^2) feeder under study is 19 km long, which ends at 11/0.4 kV, 150 kVA transformer. The primary energy source impedance is 150 ohms, and the inductance is 0.75 H. This high source impedance will ensure voltage violation at 200 kW PV penetration.

The total peak load of the distribution network is 200 kVA, at the most economical 0.9 power factor means 180 kW and

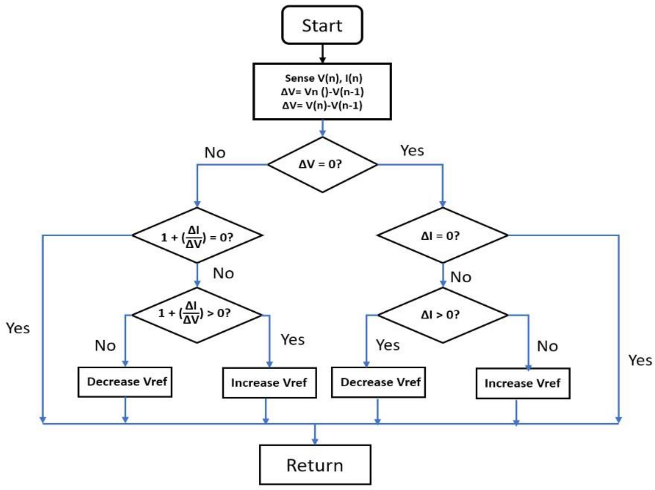


Figure 3: Traditional Incremental Conductance Flowchart

87.178 kVAR. For this study, the load is divided into two parts: The first load is located 14 km from the substation, on the 11 kV feeder, which is 104.1 kVA (94.5 kW and 43.59 kVAR). This represents the larger industrial or commercial load connected directly to the 11 kV feeder. The second load is along the 400 V lines tapped off at the end of an 11 kV feeder, representing smaller consumers connected at the LV level, is 95.97 kVA (85.5 kW and 43.59 kVAR).

A. PV System

Two 100 kW PV systems are used in the simulation, which can provide 200 kW of power. Each PV array is made from 66 parallel strings with 5 modules per string, and each module is made from 96 cells. The panels used in the simulation are SunPower SPR-305E-WHT-D. Each panel has a rated power of 305.26 watts, which gives us a total array power of 100.7 kW. SunPower solar panels are designed to degrade by only 0.55 % annually over their 25-year lifespan, ensuring long-term consistent performance. The array contains 330 PV panels, giving a total power of 100kW at the inverter output. The parameters of the panels used in the simulation are listed below in Table I.

Table I: Solar Panel Parameters from Datasheet

Module Name	SunPower SPR-305E-WHT-D
Maximum power	305.226 W
Open Circuit Voltage (**Voc**)	64.2 V
Short Circuit Current (**Isc**)	5.96 A
Voltage at MPP	54.7 V
Current at MPP	5.58 A

A DC-DC boost converter controls DC power flow from the PV array to the inverter by tracking MPP through incremental conductance in combination with an integral regulator [18]. The Incremental Conductance IC algorithm, as depicted in the flowchart in Figure 3, starts by sensing the voltage and current of the photovoltaic system and calculating the change in voltage (ΔV) and current (ΔI). The core principle of this algorithm is to compare the instantaneous conductance (I/V) with the incremental conductance (ΔI/ΔV). If ΔV is zero, it checks the sign of ΔI to adjust the reference voltage (Vref);

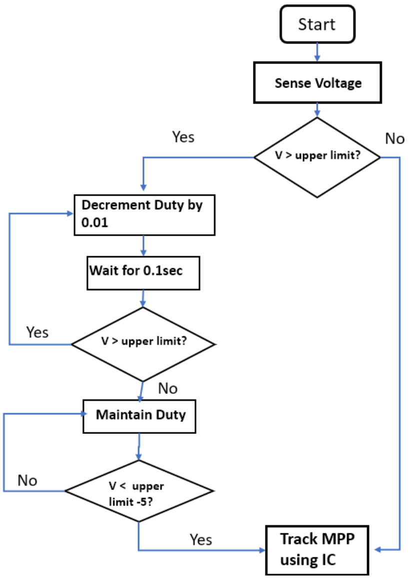


Figure 4: Flowchart of Modified Incremental Conductance method

a positive ΔI indicates moving towards higher power, so Vref is increased, and vice versa. If ΔV is not zero, the algorithm checks if the condition 1 + (ΔI/ΔV) = 0 is met, which is equivalent to ΔI/ΔV = -I/V (or ΔI/ΔV + I/V = 0). If this condition holds, the MPP is reached, and no change in Vref is needed. If 1 + (ΔI/ΔV) is not zero, the algorithm checks its sign to decide whether to increase or decrease Vref to reach the MPP.

To reduce output power oscillations, achieve optimal efficiency, minimize error, and improve MPP tracking speed, an integral regulator controller is combined with incremental conductance. It minimizes the steady-state error, which arises during tracking MPP, and ensures the operating point converges to a single point for a wide range of environmental conditions. The Integral Regulator integrates the error signal and adds a term based on the integral. The error is the difference between the conductance at the actual operating point and the maximum power point.

$$e = \frac{dI}{dV} + \frac{I}{V} \tag{1}$$

$$U_{IR} = e \cdot K_{pb} + \frac{K_{ib}}{s} \cdot e \tag{2}$$

The error signal, denoted as e, generated within the IC is processed by an IR controller to mitigate this discrepancy. The controller employs gains, represented by K_{pb} and K_{ib}, to achieve this regulation. Subsequently, the controller's output signal, U_{IR}, is given to PWM, which generates the appropriate pulse signals necessary for controlling the IGBT of the boost converter.

Dnew = Dprevious + U_{IR}, when $e > 0$
Dnew = Dprevious, when e = 0
Dnew = Dprevious - U_{IR} when $e < 0$

Three-phase three-level VSC inverter is used to convert DC power from the boost converter to 400 V, 50 Hz power for the grid. The inverter used can only supply active power, and the synchronization with the grid is done through the Current regulator, Voltage regulator, and phase-locked loop PLL method [5].

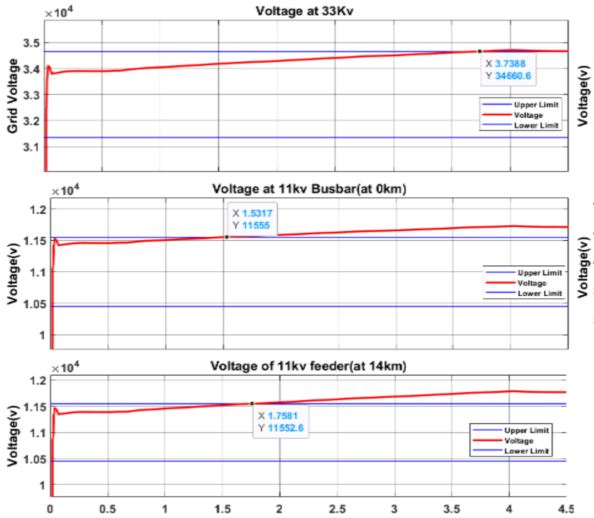
III. Proposed IC Algorithm

The algorithm addresses the issue of voltage violation by incorporating a voltage regulation component into the MPPT control. This provides a dynamic and adaptive approach to MPPT that considers the distribution network voltage constraints. RMS voltage of the grid at PCC is tested. If it is not within the acceptable range of ± 5 %, the algorithm shifts the operating point away from MPP to curtail the active power flow to the system, which eventually keeps the voltage within the limits.

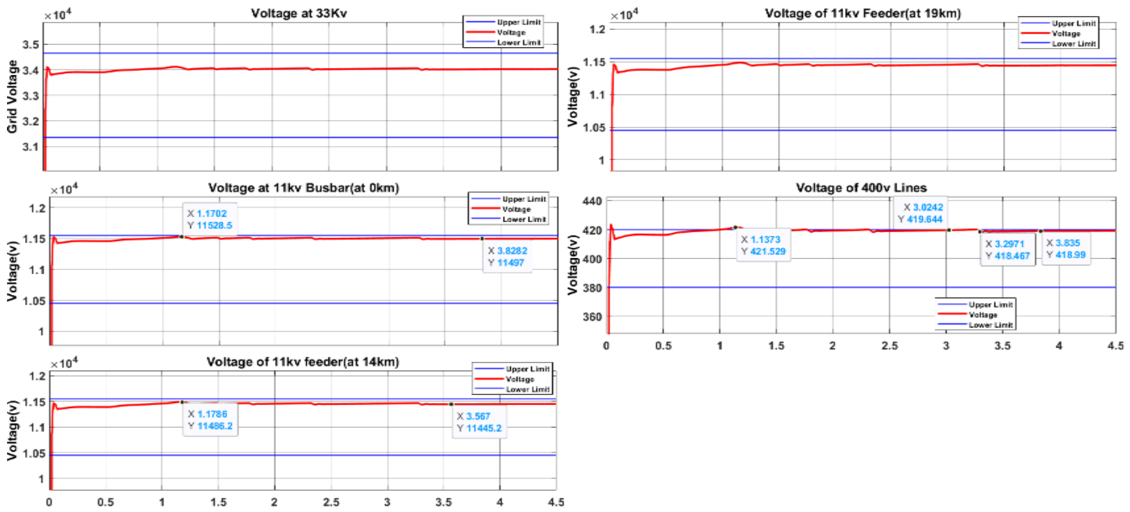
The flow chart of the algorithm is shown in Figure 4. If the sensed voltage exceeds the upper threshold (e.g., 420 V for the 400 V level), it indicates a potential voltage violation. This triggers the algorithm's voltage control mechanism. The algorithm recognizes that injecting more power into the grid at this point would exacerbate the voltage problem. Therefore, instead of seeking the MPP, it deliberately shifts the PV system's operating point away from the MPP, reducing the power output. This power injection reduction helps lower the grid voltage back within the regulatory limits. The amount by which the power is reduced is carefully controlled to ensure a balance between voltage regulation and power production. The algorithm doesn't drastically curtail power output; it makes gradual adjustments to match energy demand. Before the algorithm adjusts the PV system's operating point, it incorporates a short time delay. This delay serves two important purposes:

1) First, it prevents the algorithm from adversely reacting to transient voltage fluctuations. Power systems can experience brief voltage dips or spikes, and the control system needs to distinguish between these transient events and sustained voltage rises. The time delay acts as a filter, allowing the algorithm to respond only to sustained voltage violations.
2) Second, the delay helps to prevent oscillations in the PV system's operation. Without delay, the algorithm might continuously adjust the operating point back and forth, leading to instability. The delay allows the voltage to stabilize after an adjustment before further action is taken.

Once the algorithm determines that a sustained voltage violation has occurred and the time delay has elapsed, it adjusts the PV system's operating point. It achieves this by decrementing the duty cycle of the DC-DC converter connected to the PV array, which under normal conditions draws the maximum available power from it by utilizing the MPPT process. The duty cycle controls the amount of power extracted from the PV array. The algorithm effectively reduces the PV system's power output by reducing the duty cycle. The algorithm typically decrements the duty cycle in small, discrete steps of 0.01. This gradual adjustment allows for finer control



(a) With traditional IC algorithm



(b) With proposed IC algorithm

Figure 5: Voltage profile without and with the modified IC algorithm

over the PV power output and minimizes the risk of sudden voltage changes.

After adjusting the duty cycle, the algorithm enters a short stabilization period of 0.1 s. During this period, the duty cycle is held constant, and no further adjustments are made. This stabilization period is crucial for preventing oscillations and ensuring smooth operation of the PV system. Finally, the duty cycle is adjusted such that it maintains the voltage within the acceptable range for any voltage level (400 V, 11 kV, or 33 kV, etc.) with any given penetration limit and location.

IV. Simulation Results and Discussion

The proposed modified algorithm is simulated for PV integration at 400 V, 40 % of peak load, while all voltage levels

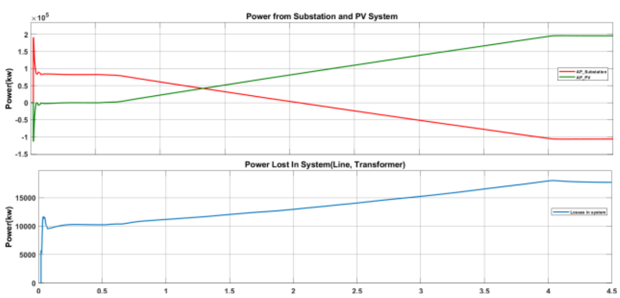


Figure 6: Power profile without the proposed IC method

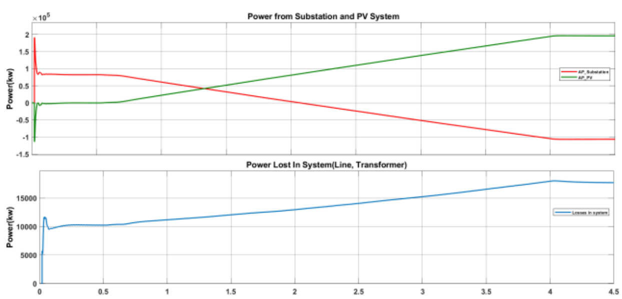
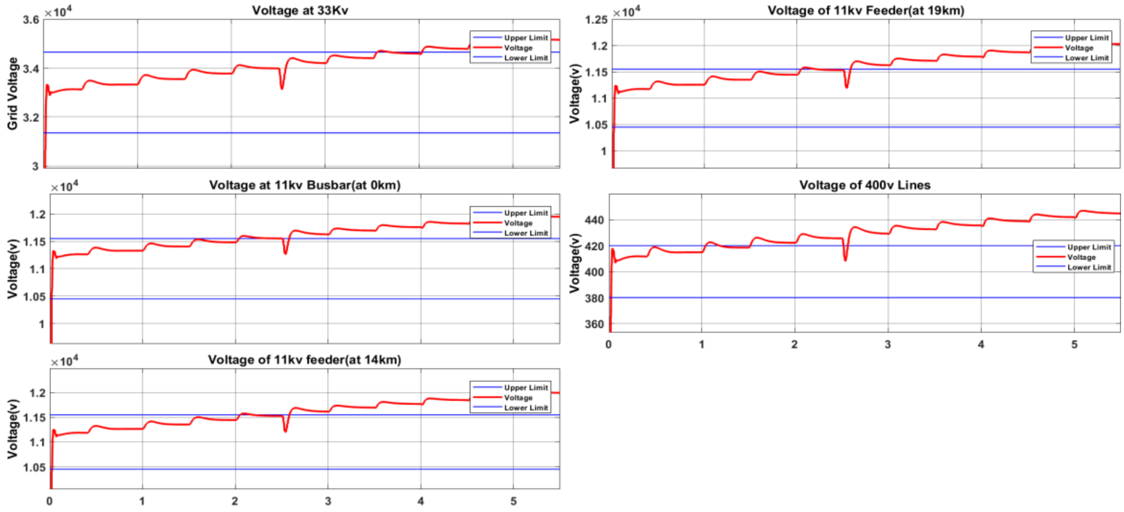
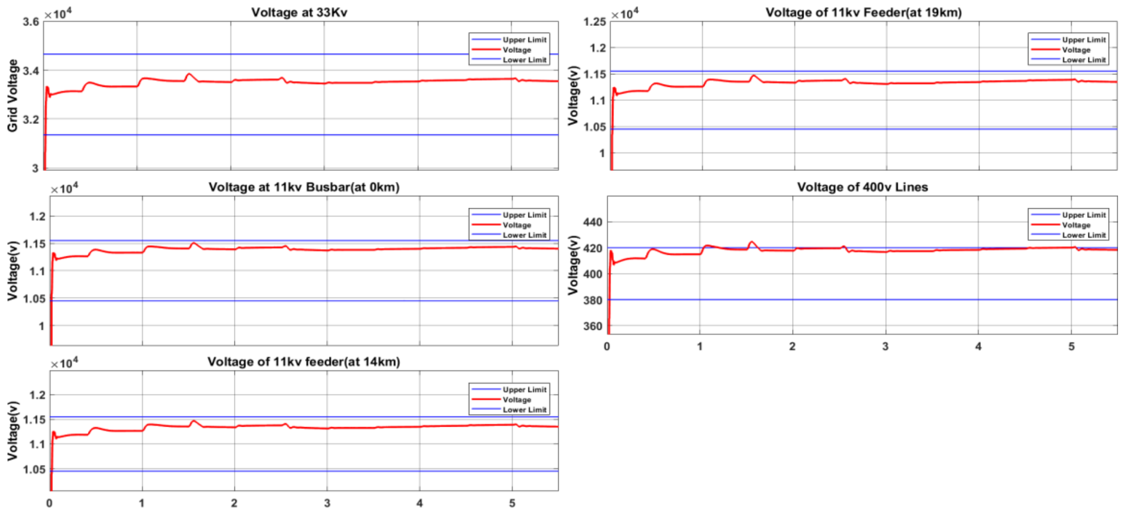


Figure 7: Power profile with the proposed IC method



(a) With traditional IC algorithm



(b) With proposed IC algorithm

Figure 8: Voltage profile for step PV input without and with the modified IC algorithm

of the whole infrastructure, 400 V, 11 kV, and 33 kV, are monitored to observe voltage violations.The PV output is linearly increased from 0 kW at 0.5 s to 200 kW at 4 s with a total simulation time of 4.5 seconds. The voltage curves monitored are shown in red lines, while the blue lines represent the upper and lower regulatory limits. First, the system is simulated with the traditional IC algorithm, without the proposed method, the results for which are shown in Figure 5(a). It is seen from the figure that voltage violations occurred at all three voltage levels —400 V, 11 kV (0 km, 14 km, and 19 km), and 33 kV—at different times due to the PV power exceeding the penetration limit, causing varied impacts on each voltage level. These violations occur sooner at lower voltage levels. For example, the 400 V line exceeds the regulatory limit first (at time t = 1.13 s), followed by the 11 kV feeder, where the voltage violation occur at t = 1.53 s, 1.75 s, and 1.79 s for distances 0 km, 14 km, and 19 km, respectively. Finally, the upper acceptable limit of 33 kV is crossed at t = 3.73 s.

In the next phase, our suggested algorithm is integrated, and the simulation is initialized. The results obtained are shown in Figure 5(b). The modified IC keeps all three voltage levels within the specified limits. This is achieved by keeping the penetrated PV power from causing reverse flow, as explained below.

Power from the main source is represented by the red line, and from the PV is shown in the green line for the power flow and power loss curves. Figure 6 shows the power flow to the load from the primary energy source, PV, and the other curve represents the power losses in the system with normal IC. The figure clearly shows that as power injection from PV increases, power from the main energy source decreases, and at t = 2.1 s, the PV power matched the load demand, and any further power from the PV will flow to the main energy source.

The power losses in the system are also directly related to the voltage rise as $P = \frac{V^2}{R}$ implying that power losses increases with voltage rise, and from the figure, it can be seen that power losses increase from 10 kW to 18 kW. The integration of the modified algorithm curtails active power injection into the system after the violation is detected, as shown in Figure 7 Power losses in the system are also limited from 10 kW to 11 kW.

The algorithm's effectiveness is again tested for stepped input from PV array, in which scenario, there is a high chance of voltage instability and even moderate levels of PV penetration can lead to significant voltage rises, especially during periods of low load and high solar irradiance. The modified incremental conductance algorithm mitigates this problem by proactively reducing PV power output when the voltage approaches the upper limit. This helps to prevent voltage violations and maintain grid stability, allowing for higher levels of PV integration. Figure 8 (a & b) compares the voltage profiles with and without the modified control for stepped input.

V. Conclusion

Most of the existing solutions to the voltage rise problem due the widespread deployment of PV generation are based on reactive power control. This paper presents a unique method for the mentioned by curtailing active power injection of PV to the distribution grid. This is solved by modifying incremental conductance, which tracks normal power when the feeder voltage is under regulatory limits and decreases the active power when the feeder voltage crosses regulatory limits. The operating point is moved away from MPP by decreasing duty, which shifts the operating point away from MPP and limits active power injection. This method is very simple and cost-effective as compared to other voltage regulation methods and has the advantage of being implementable in existing PV systems, as most of them use IC for MPP tracking.

VI. References

[1] A. Jäger-Waldau, "PV status report 2019," Publications Office of the European Union: Luxembourg, pp. 7-94, 2019.

[2] J. Tang, D. Cai, C. Yuan, Y. Qiu, X. Deng, and Y. Huang, "Optimal configuration of battery energy storage systems using for rooftop residential photovoltaic to improve voltage profile of distributed network," The Journal of Engineering, vol. 2019, no. 16, pp. 728-732, 2019.

[3] J. D. Glover, T. J. Overbye, and M. S. Sarma, Power system analysis & design. Cengage Learning, 2017.

[4] T. Aziz and N. Ketjoy, "PV penetration limits in low voltage networks and voltage variations," IEEE Access, vol. 5, pp. 16784-16792, 2017.

[5] M. Shafiullah, S. D. Ahmed, and F. A. Al-Sulaiman, "Grid integration challenges and solution strategies for solar PV systems: A review," IEEE Access, vol. 10, pp. 52233-52257, 2022.

[6] S. A. Aleem, S. S. Hussain, and T. S. Ustun, "A review of strategies to increase PV penetration level in smart grids," Energies, vol. 13, no. 3, p. 636, 2020.

[7] F. Shahnia, R. Majumder, A. Ghosh, G. Ledwich, and F. Zare, "Sensitivity analysis of voltage imbalance in distribution networks with rooftop PVs," in IEEE PES General Meeting, 2010: IEEE, pp. 1-8.

[8] M. Armendariz, D. Babazadeh, D. Brodén, and L. Nordström, "Strategies to improve the voltage quality in active low-voltage distribution networks using DSO's assets," IET Generation, Transmission & Distribution, vol. 11, no. 1, pp. 73-81, 2017.

[9] M. Zeraati, M. E. H. Golshan, and J. M. Guerrero, "Distributed control of battery energy storage systems for voltage regulation in distribution networks with high PV penetration," IEEE Transactions on Smart Grid, vol. 9, no. 4, pp. 3582-3593, 2016.

[10] N. M. A. Ramli, S. M. Hussin, D. M. Said, N. Rosmin, and A. Nawabjan, "Voltage regulation control using battery energy storage system in distribution network with high PV penetration strength," Jurnal Teknologi (Sciences & Engineering), vol. 83, no. 6, pp. 203-209, 2021.

[11] S.-B. Kim and S.-H. Song, "A hybrid reactive power control method of distributed generation to mitigate voltage rise in low-voltage grid," Energies, vol. 13, no. 8, p. 2078, 2020.

[12] T. Tewari, A. Mohapatra, and S. Anand, "Coordinated control of OLTC and energy storage for voltage regulation in distribution network with high PV penetration," IEEE Transactions on Sustainable Energy, vol. 12, no. 1, pp. 262-272, 2020.

[13] D. Almeida, J. Pasupuleti, and J. Ekanayake, "Comparison of reactive power control techniques for solar PV inverters to mitigate voltage rise in low-voltage grids," Electronics, vol. 10, no. 13, p. 1569, 2021.

[14] K. R. Babu and D. K. Khatod, "Smart inverter-based distributed volt/var control for voltage violation mitigation of unbalanced distribution networks," IEEE Transactions on Power Delivery, 2024.

[15] K. A. Khan, S. Shafiq, and M. Khalid, "A strategy for utilization of reactive power capability of PV inverters," in 2019 9th International Conference on Power and Energy Systems (ICPES), 2019: IEEE, pp. 1-6.

[16] A. M. Nour, A. Y. Hatata, A. A. Helal, and M. M. El-Saadawi, "Review on voltage-violation mitigation techniques of distribution networks with distributed rooftop PV systems," IET Generation, Transmission & Distribution, vol. 14, no. 3, pp. 349-361, 2020.

[17] C. N. A. Campas, M. M. Martinez, and H. F. R. Paredes, "PV active power curtailment for distribution network management," IEEE Latin America Transactions, vol. 20, no. 7, pp. 2010-2019, 2022.

[18] M. U. Akhtar and M. T. Iqbal, "Modeling and simulation of grid-tied Three-Phase PV system in Lahore, Pakistan," European Journal of Electrical Engineering and Computer Science, vol. 8, no. 1, pp. 23-32, 2024.

Feature-Level Fusion of CNN and Vision Transformer for Tomato Leaf Disease Identification

Afnan Ahmed
Department of Electrical Engineering
University of Engineering Technology
Peshawar, Pakistan
afnanahmed.eec@uetpeshawar.edu.pk

Sadiq Ali
Department of Electrical Engineering
University of Engineering Technology
Peshawar, Pakistan
sadiqali@uetpeshawar.edu.pk

Abstract— **Tomato leaf diseases pose a serious threat to crop yield and quality, necessitating timely and accurate detection for effective management. Traditional visual inspection methods are subjective, labor-intensive, and inefficient, highlighting the need for automated solutions. This study explores the use of transfer learning and fine-tuning of deep learning models, ResNet-50 and Vision Transformers (ViT), for tomato leaf disease detection. A novel hybrid model integrating ResNet-50 and ViT through feature-level fusion is proposed to enhance classification accuracy. While ResNet-50 and ViT achieved accuracies of 95.20% and 98%, respectively, the hybrid model outperformed both with 99.07% accuracy. These results demonstrate the effectiveness and scalability of the hybrid model for early disease detection, offering a promising solution to enhance crop health and agricultural productivity.**

Keywords— Leaf disease detection, Hybrid Model, Deep Learning Models

I. Introduction

Globally, among the top three broadly traded vegetables, one of them is Tomatoes, which play a crucial part in the international vegetable market [1]. As part of the daily increase in the demand for tomatoes, their production on the worldwide level plus the area of cultivation keeps growing. At the same time, preferences of consumers are shifting towards eco-friendly and high-quality products, which results in the increased need to improve the standards of food quality. However, the production and quality of tomatoes are often compromised due to various diseases that lead to a significant reduction in yield and, eventually, economic losses for poor farmers [2]. Plant diseases and insect damage are among the leading causes of agricultural losses globally. Estimations suggest that the losses in annual production are mostly due to pests and diseases, which have been substantial since the start of the 21st century [3]. On the other hand, plant diseases alone specifically contribute to the annual losses of 15% to 17% (approximately) of the total crop, which is a highly alarming figure! [4]. An estimated amount of 68% in the total annual production of the crop is lost due to factors such as pests, weeds, and plant leaf diseases [5], which causes a major economic setback. To address these challenges, there is a requirement to integrate advanced crop protection and enhancement strategies, utilizing and maximizing the latest global practices and emerging technologies. To ensure that the yield and quality of tomatoes are significant not only for food security but also for global economic trade. Conventionally, the identification of tomato leaf disease relied on manual visual inspection done visually by field workers, a method prone to subjectivity, inefficiency, and low accuracy. Considering these limitations, it is essential to develop efficient and automated methods for detecting tomato leaf diseases and pests using modern technology. In response to these challenges, we propose:

- Existing tomato leaf disease datasets often lack diversity, limiting model generalization and real-world accuracy. To overcome this, we enhanced the dataset using extensive data augmentation techniques, including variations in lighting, orientation, and background noise. This improved diversity allows the model to perform robustly under real-world conditions, significantly boosting the reliability of automated disease detection for effective crop health management.
- Traditional CNNs primarily capture local features, while Vision Transformers (ViTs) focus on global dependencies, limiting their standalone effectiveness in fine-grained disease detection. To address this, we fine-tuned pre-trained CNN and ViT models using transfer learning and proposed a hybrid CNN-Transformer model that combines local and global feature extraction. This fusion leverages the strengths of both architectures, significantly enhancing classification performance for tomato leaf disease detection.

II. Related Work

Many studies have explored plant disease detection, particularly in tomato leaves, but limitations persist. Traditional methods rely on principal component analysis, using a single sample leaf as a reference, and texture-based segmentation techniques [6][7], which struggle in real-time scenarios [8][9]. To address these challenges, [10] introduced PLPNet, which tackles intraclass variability and similarity, key factors in disease classification. Object detection techniques were emphasized for improving accuracy, particularly in cases where soil backgrounds obscure infected leaf edges. Building on this, [11] proposed TomatoDet, integrating Swin-DDETR's self-focus mechanism, Meta-ACON launch, and an improved bidirectional weighted feature pyramid network (IBiFPN) to enhance small-target disease identification and reduce false detections. Transformer-based approaches were further explored in [12] with the NanoSegmenter model, incorporating lightweight techniques such as quantization and sparse attention to optimize efficiency. It achieved a precision of 0.98, recall of 0.97, and mIoU of 0.95, with an inference speed of 37 FPS, making it viable for real-time agricultural applications. In [13], traditional image classification models were compared with the YOLO object detection framework, where optimized feature layers and attention mechanisms improved real-time decision-making for early pest detection and crop loss reduction. Beyond tomato crops, deep learning has been applied to plant disease detection more broadly. In [14], transfer learning was used to train a deep CNN on cassava disease images, achieving up to 98% accuracy for certain diseases. A comparison of machine learning models found that SVM outperformed others in four out of six disease categories. Lastly, [15] highlighted the importance of deep learning in food security, training a deep CNN on a public dataset of 14 crop species and 26 diseases. While the model achieved high accuracy, performance dropped under varying environmental conditions, underscoring the need for diverse training data. Despite limitations, increasing smartphone accessibility presents opportunities for AI-driven disease detection to aid farmers and enhance crop management.

III. PROPOSED METHODOLOGY

The proposed framework comprises four key stages: Data Acquisition (for collecting labeled disease images), Pre-processing (to standardize and normalize inputs), Augmentation (to increase dataset diversity and generalization), and Classification (using deep learning models). The complete workflow is illustrated in Fig. 1.

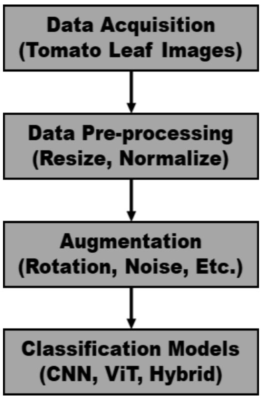


Fig. 1. Proposed Methodology

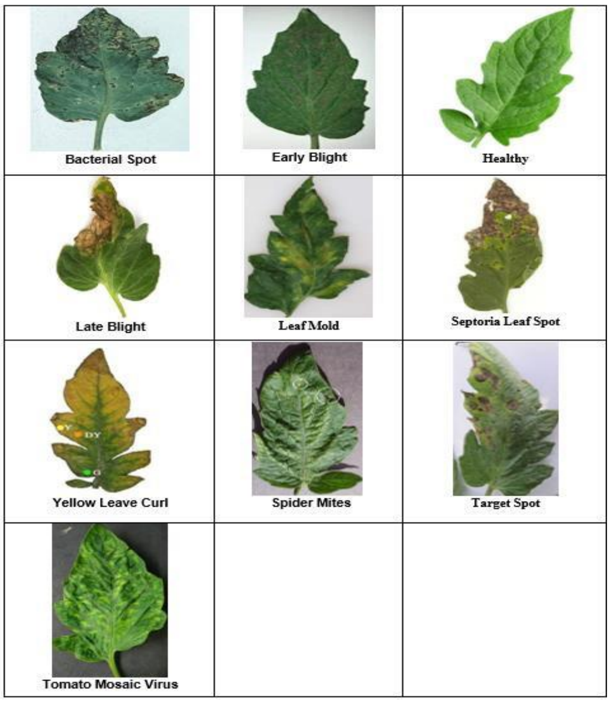
A. *Data Acquisition:*

The data set of tomato leaf disease consists of images collected from public sources and agricultural research databases that have captured the variations in leaf conditions due to various factors such as climate conditions, soil conditions, and different farming practices. The data set includes ten classes: Early Blight, Late Blight, Septoria Leaf Spot, Bacterial Spot, Target Spot, Leaf Mold, Yellow Leaf Curl Virus, Mosaic Virus, Powdery Mildew, and Healthy Leaves. The key features that the dataset focuses on are lesion texture, shape, color, and leaf structure for accurate classification. It helps in the early detection of disease, which enables timely interventions to reduce the losses of yield. This dataset strengthens the models of machine learning to efficiently diagnose the disease and to be scalable, which will support better crop management. Sample images can be seen in Table 1.

B. *Data Pre-processing*

After the construction of the dataset of Tomato Leaf Disease, the techniques of data pre-processing were applied to ensure the consistency and the suitability for deep learning applications. All the images were systematically renamed based on their class, appending an incremental numeric identifier for clarity (e.g., "EarlyBlight_1," "EarlyBlight_2"). To maintain a consistent input size for deep learning models, all the images were resized uniformly to 224×224 pixels. In the final step, the pixel values were normalized to a [0,1] range by dividing them by 255, enhancing the stability and performance of the model during training and evaluation

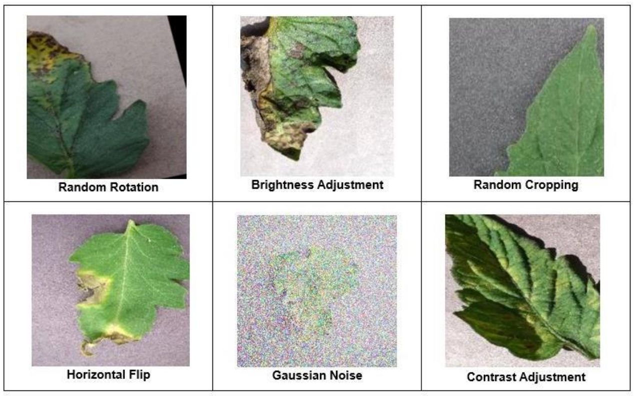
TABLE 1. Sample Pictures of tomato Leaf Disease dataset



C. *Augmentation*

To improve the robustness of the dataset of Tomato Leaf Disease, various techniques of data augmentation were applied. Random rotation (-30 to 30 degrees) and flipping horizontally led to enhancing the viewpoint diversity, while the adjustments in brightness and contrast were recorded for varying lighting conditions. The introduction of subtle variations and random cropping simulated zooming effects were done by Gaussian noise (mean 0, std 25). These augmentations collectively resulted in enhancing the diversity of the dataset, improving the adaptability of the model to the conditions of the real world. The Sample images of Augmentation can be seen below in Table 2.

TABLE 2. Sample Pictures of Augmentation



D. *Classification Model*

This study utilizes and implements a deep learning-based classification framework for the detection of Tomato Leaf Disease by the integration of ResNet-50, Vision Transformer (ViT-B_16), and a Feature-Level Fusion Hybrid Model. ResNet-50 (Figure 2) is a model that is fine-tuned to capture the patterns that are disease-specific, such as texture variations and discoloration, employing transfer learning to hold its general feature extraction while performing the processes to refine deeper layers for precise recognition of disease. The robustness of the model was enhanced with a custom classification head, along with dynamic learning rate scheduling, regularization, and data augmentation.

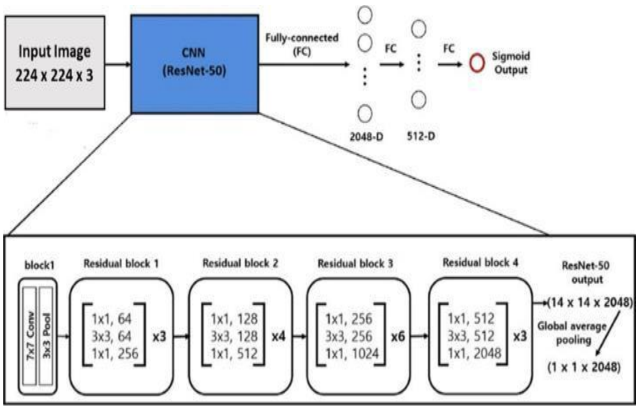


Fig. 2. ResNet-50 Architecture

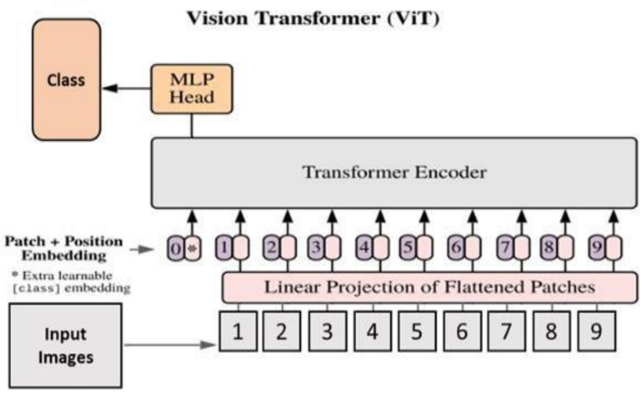


Fig. 3. Vision Transformer Architecture

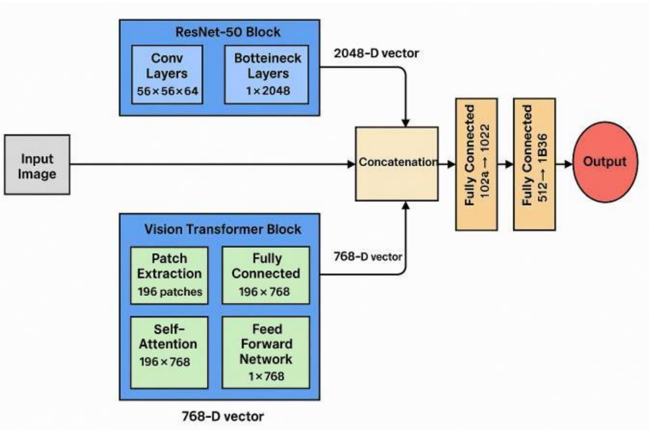


Fig. 4. Feature-Level Fusion Hybrid Model Architecture

It also prevents overfitting, ensuring reliable detection of disease [16]. ViT-B_16 (Figure 3) uses a mechanism of self-attention to capture global dependencies and intricate disease features. Fine-tuned on the dataset of tomato leaf, ViT-B_16 can effectively distinguish between healthy and diseased leaves, enabling intervention at an early stage. Task-specific layers and dynamic hyperparameter tuning further improve the accuracy of classification, making the model highly adaptable to complex disease patterns [17].

To utilize the strengths of both architectures, a Feature-Level Fusion Hybrid Model (Figure 4) integrates ResNet-50 and ViT-B_16. ResNet-50 extracts fine-grained spatial details, while ViT-B_16 captures long-range dependencies that provide a comprehensive understanding of the symptoms of the disease. The fusion of feature representations through specialized layers enhances the accuracy of classification, which ensures the robustness in the identification of disease for effective protection of the crop

IV. Experimental Setup

The model was implemented using Kaggle's GPU environment for efficient training. The dataset was split into 70% for training, 15% for validation, and 15% for testing. Key hyperparameters included 15 epochs, a 0.0005 learning rate, and a batch size of 32. Images were resized to 224x224x3 for standardization. A custom learning rate scheduler optimized convergence, while overfitting detection determined epoch limits. Gamma adjustment and a step schedule-controlled learning rate decay. For multi-class classification, class-specific weighting and a hybrid loss function (Categorical Cross-Entropy) improved generalization. The Adam optimizer, enhanced with dynamic learning rate adjustments and gradient smoothing, ensured stable and efficient training, leading to strong classification performance [18].

V. Result and Analysis

For Tomato Leaf Disease Detection, all three models demonstrated consistent improvement over 15 epochs. ResNet-50 achieved a training accuracy increase from 61.23% to 97.03%, with validation accuracy rising from 81.87% to 94.67%, while training and validation losses decreased to 0.0075 and 0.0092, respectively (Figure 5). ViT-B_16 outperformed ResNet-50, with training accuracy improving from 65.74% to 99.63% and validation accuracy stabilizing at 97.60%, alongside a steady decline in training and validation losses to 0.0014 and 0.0026, respectively (Figure 6). The hybrid model exhibited the most robust performance, leveraging both architectures to enhance feature representation. It achieved a training accuracy increase from 78.66% to 99.97%, with validation accuracy stabilizing at 98.70%. Training and validation losses steadily declined to 0.0196 and 0.0516, respectively (Figure 7).

Fig. 5. Training and Validation Accuracy and Loss ResNet-50

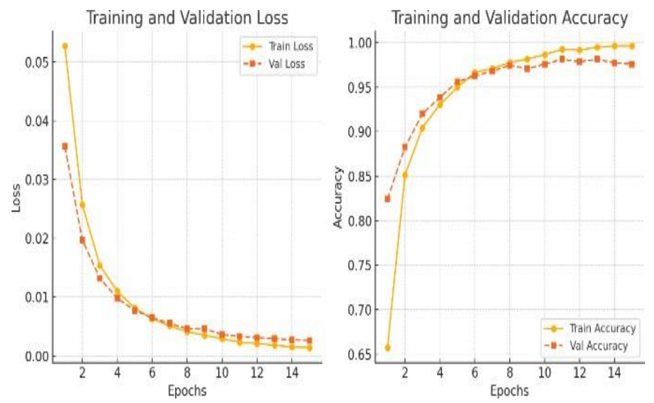


Fig. 6. Training and Validation Accuracy and Loss Vision Transformer

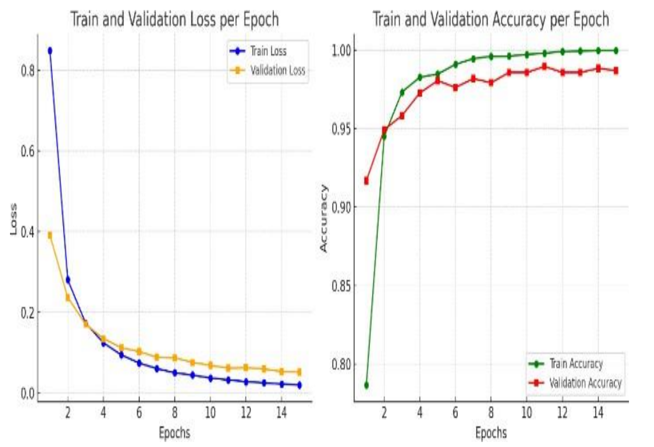


Fig. 7. Training and Validation Accuracy and Loss Hybrid Model

Performance metrics in Table 3 further validate these findings. ResNet-50 achieved an accuracy of 95.20% with a recall, precision, and F1-score of 95%, indicating reliable classification. ViT-B_16 demonstrated superior performance, achieving 98% across all evaluation metrics. The hybrid model outperformed both, achieving the highest accuracy of 99.07%, along with 99% recall, precision, and F1-score. These results confirm that integrating ResNet-50 and ViT-B_16 significantly enhances classification accuracy and stability, making it the most effective model for reliable tomato leaf disease detection.

To examine the output and results of our suggested hybrid model, we contrasted its classification metrics versus existing studies that utilized ResNet-50 and ViT models. As shown in Table 4, ResNet-50 in [19] achieved exceptionally high accuracy (99.97%), recall (99.87%), and precision (99.86%). However, it is important to note that this result was obtained without data augmentation, leading to overfitting and a lack of generalization. The dataset used in [19] remained the same in terms of having ten classes, but the absence of augmentation limited its robustness for real-world applications. In contrast, the ViT model in [20] achieved a significantly lower accuracy of 90.99%, with an average precision of 90.9% and recall of 89.3%. This drop in performance suggests that ViT struggled with feature extraction in comparison to CNN-based models, particularly when dealing with variations in disease symptoms across different tomato leaf samples. Our proposed hybrid model demonstrated superior generalization, achieving an accuracy of 99.07% with balanced precision, recall, and F1-score values of 99%. The inclusion of both CNN-based feature extraction and transformer-based attention mechanisms contributed to its enhanced classification performance. Unlike [19], our approach incorporated augmentation techniques to ensure dataset diversity, leading to a more reliable and generalized model for tomato leaf disease classification.

These results confirm that our hybrid approach effectively combines the strengths of convolutional networks and transformer architectures, making it a robust solution for real-world applications in precision agriculture

TABLE 3. Performance Metrics

Model	Recall	Precision	F1-Score	Accuracy
ResNet-50	95%	95%	95%	95%
ViT B_16	98%	98%	98%	98%
Hybrid Model	99%	99%	99%	99%

TABLE 4. Comparative Analysis

Model	Recall	Precision	F1-Score	Accuracy
[19]	99.87%	99.86%	99.88%	99.97%
[20]	89.3%	90.9%	90.7%	90.99%
Hybrid Model	99%	99%	99%	99.07%

VI. CONFUSION MATRIX

For Tomato Leaf Disease Detection, ResNet-50 achieved 95.2% accuracy but showed minor misclassifications, particularly between Early Blight and Septoria Leaf Spot (Figure 8). ViT-B_16 improved accuracy to 98.0%, enhancing class differentiation, though slight confusion remained in closely related diseases (Figure 9). The Hybrid model was successfully achieved the highest accuracy of 99.07%, with a classification that is near-perfect and has minimal misclassifications in diseases that are visually similar (Figure 10). The results highlighted the superior precision of the hybrid model, which makes it highly effective for the early detection of disease in precision agriculture.

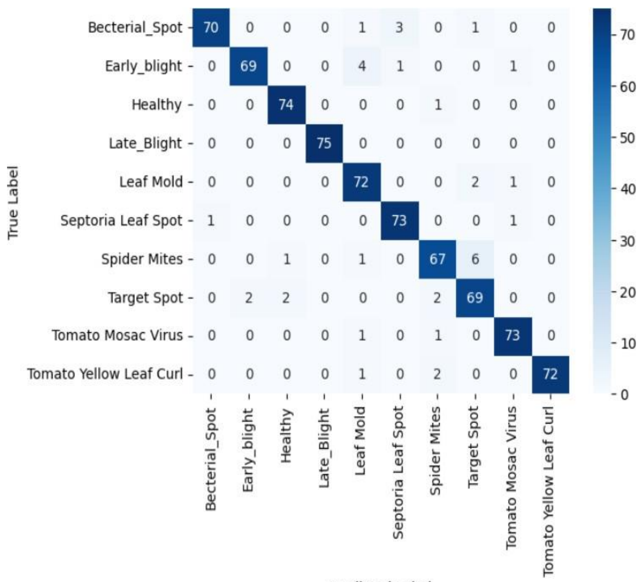


Fig. 8. ResNet-50 Confusion Matrix

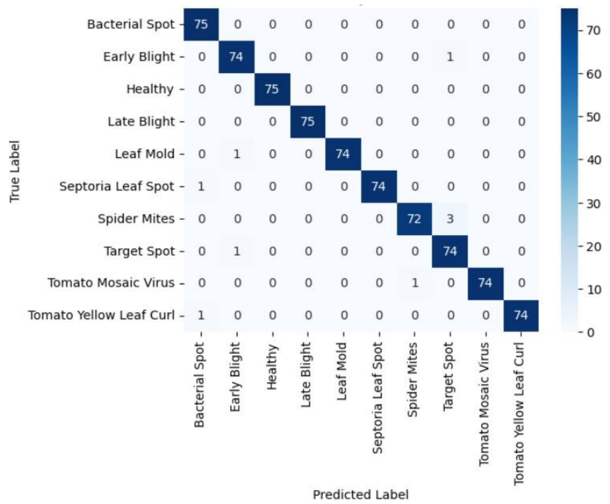


Fig. 9. Vision Transformer Confusion Matrix

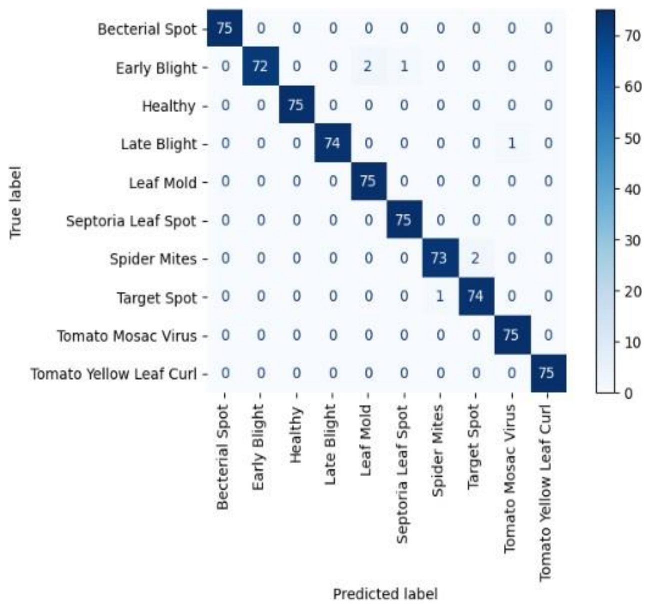


Fig. 10. Hybrid Model Confusion Matrix

VII. CONCLUSION

For Tomato Leaf Disease Detection, ResNet-50 achieved 95.2% accuracy but showed minor misclassifications, particularly between Early Blight and Septoria Leaf Spot (Figure 8). ViT-B_16 improved accuracy to 98.0%, enhancing class differentiation, though slight confusion remained in closely related diseases (Figure 9). The Hybrid model was successfully achieved the highest accuracy of 99.07%, with a classification that is near-perfect and has minimal misclassifications in diseases that are visually similar (Figure 10). The results highlighted the superior precision of the hybrid model, which makes it highly effective for the early detection of disease in precision agriculture.

REFERENCES

[1]. Li, T.; Cui, J.; Guo, W.; She, Y.; Li, P. The Influence of Organic and Inorganic Fertilizer Applications on Nitrogen Transformation and Yield in Greenhouse Tomato Cultivation with Surface and Drip Irrigation Techniques. *Water* 2023, *15*, 3546.

[2]. Yu, Y.H. Research Progress of Crop Disease Image Recognition Based on Wireless Network Communication and Deep Learning. *Wirel. Commun. Mob. Comput.* 2021, *2021*, 7577349.

[3]. H. Durmus, E. O. Günes and M. Kirci, " Disease detection on the leaves of the tomato plants by using deep learning," in 2017 6th International Conference on Agro-Geoinformatics, Fairfax, VA, USA, 2017.

[4]. C. Valenzuela, R. G. Baldovino, A. A. Bandala and E. P. Dadios,"PreHarvest Factors Optimization Using Genetic Algorithm for Lettuce," Journal of Telecommunication, Electronic, and Computer Engineering, vol. 10, 2018.

[5]. C.Valenzuela, A. B. Culaba and E. P. Dadios, "Identification of philippine herbal medicine plant leaf using artificial neural network," in IEEE 9th International Conference on Humanoid, Nanotechnology.

[6]. C. Valenzuela, J. C. V. Puno, A. A. Bandala, R. G. Baldovino, R. G.D. Luna, A. L. d. Ocampo, J. Cuello and E. P. Dadios, "Qualityassessment of lettuce using artificial neural network," in 2017 IEEE 9thInternational Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), Manila, 2017.

[7]. J. B. U. Dimatira, E. P. Dadios, F. Culibrina, J.-A. Magsumbol, J. Dela Cruz, K. Sumage, M. T. Tan and M. Gomez, "Application of fuzzy logic in recognition of tomato fruit maturity in smart farming," in IEEE Region 10 Conference (TENCON), Singapore, 2016.

[8]. C. Valenzuela, R. B. Baldovino, A. A. Bandala and E. P. Dadios, "Optimization of Photosynthetic Rate Parameters using Adaptive NeuroFuzzy Inference System (ANFIS)," in 2017 International Conference on Computer and Applications (ICCA), Doha, 2017.

[9]. J. Shijie, J. Peiyi, H. Siping and s. Haibo, "Automatic detection oftomato diseases and pests based on leaf

images," in 2017 Chinese Automation Congress (CAC), Jinan, 2017
[10]. Tang, Z., He, X., Zhou, G., Chen, A., Wang, Y., Li, L., & Hu, Y. (2023). A precise image-based tomato leaf disease detection approach using PLPNet. *Plant Phenomics*, *5*, 0042.
[11]. Wang, X., & Liu, J. (2024). An efficient deep learning model for tomato disease detection. *Plant Methods*, *20*(1), 61.
[12]. Liu, Y., Song, Y., Ye, R., Zhu, S., Huang, Y., Chen, T., ... & Lv, C. (2023). High-Precision Tomato Disease Detection Using NanoSegmenter Based on Transformer and Lightweighting. *Plants*, *12*(13), 2559.
[13]. Liu, J., & Wang, X. (2020). Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. *Frontiers in plant science*, *11*, 898.
[14]. Ramcharan, A., Baranowski, K., McCloskey, P., Ahmed, B., Legg, J., & Hughes, D. P. (2017). Deep learning for image-based cassava disease detection. *Frontiers in plant science*, *8*, 1852.
[15]. Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science*, *7*, 1419.
[16]. D. B. Ishrat Zahan Mukti, "Transfer Learning Based Plant Diseases Detection Using ResNet50," in *IEEE Access*, December 2019.
[17]. S. S. B. Amer Tabbakh, "A Deep Features Extraction Model Based on the Transfer Learning Model and Vision Transformer "TLMViT" for Plant Disease Classification," *IEEE Access*, May 2023.
[18]. R. O. Ogundokun, R. Maskeliunas and S. M. &. R. Damaševičius, "Improved CNN Based on Batch Normalization and Adam Optimizer," in *Computational Science and Its Applications – ICCSA 2022 Workshops*, July 2022.
[19]. Muslih, A. D. Krismawan, "Tomato Leaf Diseases Classification Using Convolutional Neural Networks with Transfer Learning ResNet-50," *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, May 2024.
[20]. U. Barman, P. Sarma, M. Rahman, V. Deka, S. Lahkar, V. Sharma, and M. J. Saikia, "ViT-SmartAgri: Vision Transformer and Smartphone-Based Plant Disease Detection for Smart Agriculture," *Agronomy*, February 4, 2024.

AI Vision for Health Care: Virtual Keyboard and Mouse Empowering Partially Disabled Patients

Sabeen Zulfiqar
Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
sabeenzulfiqar27@gmail.com

Ahad Abbas
Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
ahadabbas.x8@gmail.com

Mehroz
Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
mehrozkhan473@gmail.com

Engr. Muhammad Farooq
Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
muhammad.farooq@uetpeshawar.edu.pk

Ihsan Ul Haq
Department of Energy Engineering,
University of Hull
England, United Kingdom
ihsankhan97@gmail.com

***Abstract*— This paper presents a machine-learning based virtual keyboard and mouse system designed to empower individuals with physical disabilities. The system recognizes hand gestures through computer vision techniques, interprets gestures to simulate keyboard inputs and mouse controls. By using Convolutional Neural Networks (CNNs) and the YOLOv8 model, the system achieves real-time performance with a mean accuracy of 92%, enabling touchless human-computer interaction. The proposed solution uses commonly available hardware such as standard webcams, ensuring accessibility, affordability, and ease of deployment. This system enhances the usability of computing devices for individuals with motor impairments, providing an innovative, touchless alternative to traditional input methods. It also supports essential tasks like scrolling, clicking, and zooming through intuitive gestures. The framework is adaptable for diverse environments, ensuring ease of use across different uses.**

Keywords: Virtual Keyboard, Virtual Mouse, YOLOv8, PyAutoGUI, Gesture Recognition, Computer Vision, CNN, OpenCV, Tkinter.

I. INTRODUCTION

Traditional input devices like keyboard and mouse rely on fine motor skills and tactile input and remain out of reach for many users with physical impairments [1]. However, till date, we still lack an alternative device that can be effectively used in today's environment, which has witnessed a significant rise of the global touchless input market as reflected in this study's main objective of addressing touchless virtual input system based on hand gesture recognition application as an alternative to provide a fast and secure means of input while driving and providing a touchless interaction to the computer [2]. Combining machine learning algorithms with off-the-shelf consumer hardware such as standard webcams, the system is designed for accessibility, low-cost, and real-time performance, aiming to bring affordable human-computer interaction to people of all physical abilities [3].

Existing gesture-based assistive technologies are often limited in terms of accuracy, adaptability and cost [4]. However, despite their promising gesture recognition ability on such controlled experimental setup, the issues of varying ambient light, occlusions and individual differences hinder their usability for practical applications [5]. Though previous work has shown impressive gesture identification results in lab environments, they face fundamental issues that limit their performance during real-world application such as varying lighting conditions, occlusions, and diverse motion and hand scale between different users [6]. Other works have introduced hybrid systems combining several inputs modalities, however there are still gaps in the balancing process between computational efficiency and universal usability [7]. This motivates the need for an adaptable approach that uses state of the art machine learning techniques to increase robustness among heterogeneous spaces and user distributions [8].

Primary objective of this study is to design a virtual input system by integrating a virtual keyboard and mouse, enabling the precise mapping of hand movements to computer commands, including typing and pointer control. The system employs convolutional neural networks (CNNs), trained on custom-made datasets, to develop generalized capabilities. Real-time motion tracking is achieved through optical flow algorithms, ensuring accurate and responsive interaction in real-life scenarios [9]. The architecture of the system demands three performance characteristics: rapid response, speed, and all users with varying levels of motor function abilities [10]. The performance of the system is based on evaluating the accuracy of recognition of gestures, system response, speed, and more general, the basic usability of the system in user interface when the system is tested in real world settings [11].

II. LITERATURE REVIEW

Existing gesture and gaze detection systems have been explored extensively using machine learning techniques. Projects such as Google's G-board for mobile devices and research using computer vision models like Open Pose have demonstrated successful virtual input systems [2]. However, many of these systems require specialized hardware or high computational resources, limiting their accessibility [12]. This project will focus on using commonly available hardware such as standard webcams, combined with efficient deep learning algorithms like Convolutional Neural Networks (CNNs) for gesture recognition, ensuring usability in various environments [13].

Convolutional Neural Networks are a specific kind of deep learning model designed to recognize images. They perform exceptionally well at tasks like object classification, medical imaging analysis, and face detection because they are inspired by the way the human brain interprets visual information [6]. CNN makes predictions by dissecting a picture into several layers, identifying patterns like edges, forms, and textures, and then applying that information [9]. Convolutional Neural Networks are the cornerstone of modern image identification technologies if you're interested in AI and computer vision [14].

With the help of the Python module PyAutoGUI, you can automate keyboard and mouse operations so that your programs may interact with the screen in the same way that a human would [15]. It could enter text, click buttons, and even recognize screen objects. This makes it extremely useful for integrating AI models into useful, real-world applications, automating repetitive chores, and testing user interfaces [16].

The study concludes that the gesture-controlled system offers an innovative, touchless alternative to traditional input methods by seamlessly combining various ways to interact with technology [17].

The hand gesture-based control system simplifies tasks such as managing presentations, virtual drawing, and adjusting system zoom in, and zoom out through a touchless interface [18]. It enhances mobility and creativity, providing a more accessible control option for users, including those with physical limitations [19].

This project seeks to develop a virtual input system that uses hand gesture recognition. The goal is to achieve high accuracy and real-time performance while ensuring compatibility with standard webcams such as (1080P) [3]. Ultimately, the project aims to create a more inclusive computing environment that facilitates independent technology use for people with disabilities [20].

Machine learning based systems using hand gestures and other non-contact methods have proven to be effective alternatives to traditional keyboard and mouse. For example, solutions leveraging AI and computer vision for virtual keyboards allow users to perform typing tasks without physical interaction [1]. The use of machine learning algorithms, and hand gesture recognition enables systems to control virtual keyboard and mouse seamlessly, enhancing accessibility for users with disabilities [4].

Also enhancing hand gestures for controlling systems beyond typing, such as presentations, drawing, and zoom in zoom out gestures, powered by machine learning and real-time gesture recognition, significantly improve user engagement by offering freedom from traditional input methods [8].

Overall, the development of these systems provides efficient, contactless input solutions, furthermore explanation include in framework section below.

III. OUR FRAMEWORK

The input method for our framework employs custom gestures captured via a standard webcam (e.g, a built-in camera) to track hand movements during gesture input. In the virtual keyboard implementation, two subsystems process gesture input.

1. **Gesture Detection**: It identifies and localizes hand gestures in the video stream.
2. **Gesture Recognition**: Classifies detected gestures into predefined actions or characters [10].

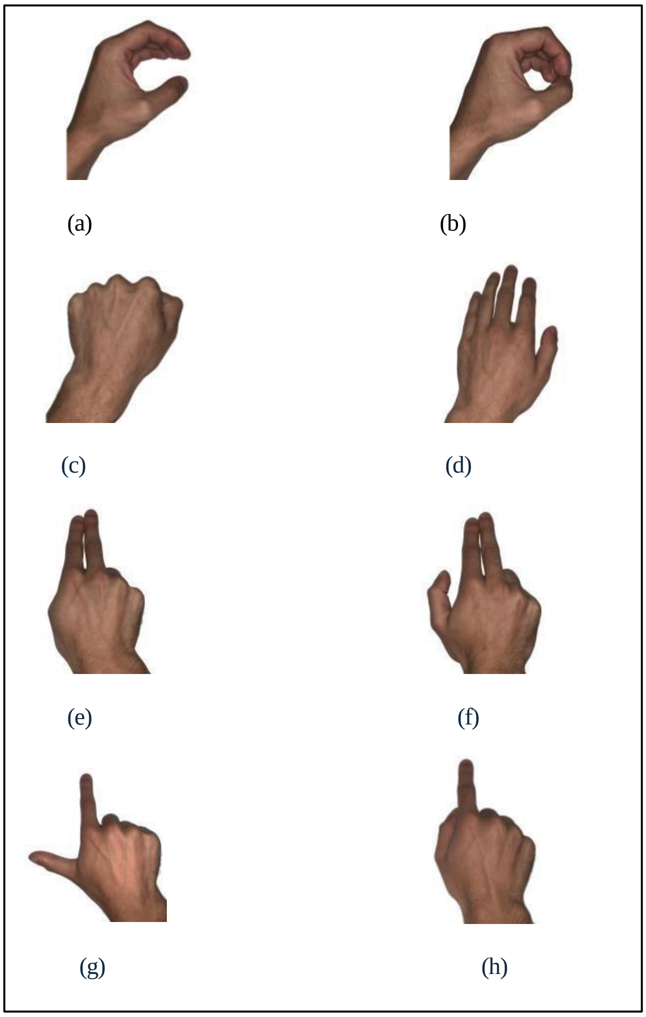


Figure 1: Hand Gestures for Virtual Keyboard and Mouse System

IV. GESTURE RECOGNITION

Gesture recognition demands the selection and detection of specific gestures to ensure interaction between humans and machines. Standard webcam detects the hand movement and gestures. We selected eight gestures providing different actions or task: zoom in, zoom out, left click, right click, pointing, clicking, scroll up, and scroll down. Concept behind the design of these gestures is to provide a method that can be easily detected by standard webcam and less effort needed for person with disabilities during typing. Figure.1 illustrates the hand gestures for virtual keyboard and mouse: (a) selected for zoom in, (b) selected for zoom out, (c) selected for scroll down, (d) selected for scroll up, (e) selected for right click, (f) selected for left click, (g) selected for pointing, and (h) selected for clicking. The eight gestures are also discussed below:

1. ZOOM OUT: Zoom out with left hand thumb and index fingers to form an O shape, same as for right hand. This gesture is selected to Zoom out, as shown in Figure 1(c) [18].
2. ZOOM IN: Zoom in with thumb and index finger formed into C shape. This hand action selected for certain task. Shown in Figure 1(a) [18].
3. SCROLL UP: For scrolling up open your hand as shown in Figure 1(d) such that the system detects the gesture.
4. SCROLL DOWN: For scrolling down close fingers to form a fist as shown in Figure 1(c) [17].
5. RIGHT CLICK: For clicking right with effortless special gesture selected for people with disabilities open index finger and middle finger and close other fingers and thumb as shown in Figure 1(e) this gesture is selected for click right [17].
6. LEFT CLICK: For click left gesture which is selected shown in Figure 1(f) open index finger or middle finger with thumb which can detected by system and disable person can effortlessly perform task [16].
7. POINTING: To perform pointing task simple gesture is selected as shown in Figure 1(g) open index finger with thumb and close other fingers [19].
8. CLICKING: click gesture is selected as shown in Figure 1(h) pointing index finger and close hand, same as for left hand [16].

Each gesture can be performed with small adjustments of the hand while adjusting the hand within the range of webcam. Gestures are selected to be simple, easy to memories, and can be performed with little adjustments of hands movement.

Gestures are chosen in such a way that people with disabilities can perform with less physical action. Gesture for Zoom in and zoom out are picked like each other for comfort and easy to understand. Similarly, Right click or Left click are most common task that performed by any user, so such kind of gesture are designed simple, require small arrangement of fingers and hand with less physical moment. In addition, scroll up and scroll down gesture are set in such a way that can be performed with minor changes in gesture or hand moment. These gestures provide the user seamless or touchless input to performed task. These gestures could be used to achieve the functions like virtual drawing, scrolling up and down. To develop an effective gesture recognition system, a systematic approach is necessary; the following section outlines the methodology employed in this study

V. SYSTEM METHODOLOGY

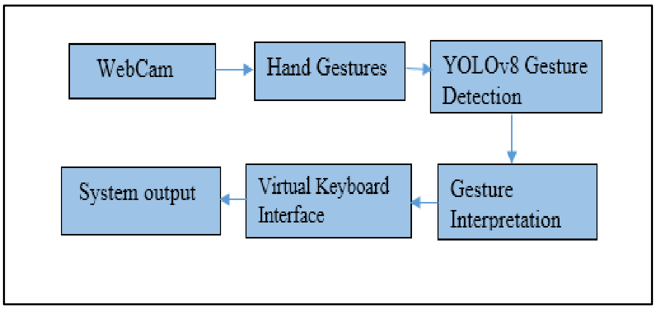


Figure 2: System Architecture block diagram

The system uses computer vision and machine learning to recognize motions precisely in real time. Hand gestures are recorded and tagged for training as part of the initial data collection and preparation process. The YOLO framework (YOLOv5 and YOLOv8) is utilized to guarantee speed and accuracy because it performs exceptionally well in real-time object detection [14]. The system methodology involves:

1. **Data Collection and Labelling:**
   We use an open-source dataset combined with our own created data, which were converted to YOLO's format using labelling [14].

2. **Model Development & Training:**
   After collecting and labelling data the model was trained using this combined dataset, enhancing it for accurate gesture recognition and real time performance.
   Training Setup: The dataset was split into training (87%), validation (6%), and testing (7%). Training used a batch size of 16, and 100 epochs. We monitored MAP (mean average precision).

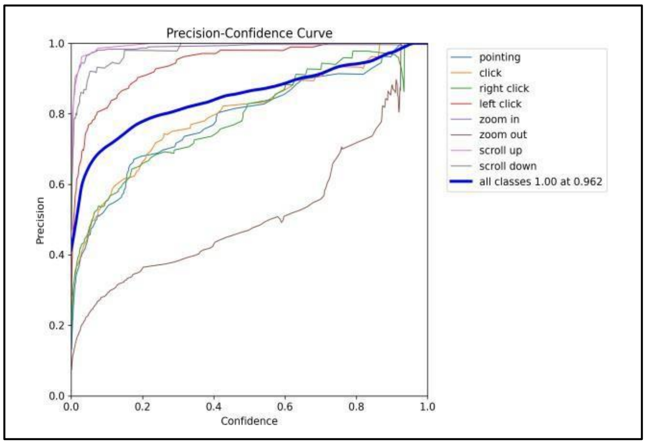


Figure 3: Precision of Model Result

Precision Curve: Figure (3) plot shows how correct (precisely) our model predicts during training and validation.

3. Real-Time Processing & Integration:

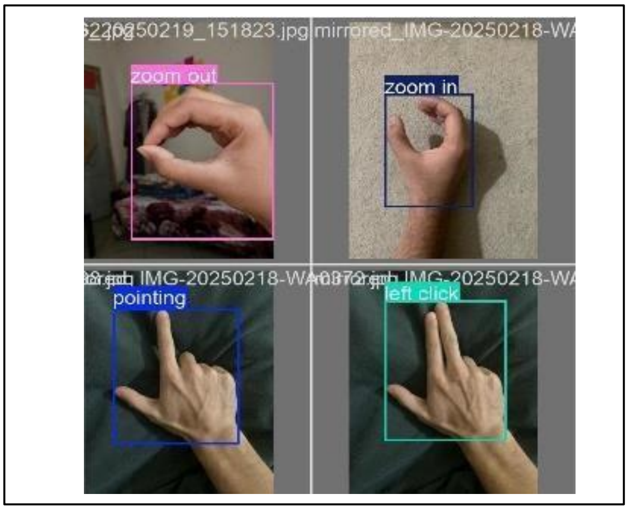


Figure 4: Validation and Testing Results

OpenCV captured webcam frames, which were resized, normalized, and denoised. The model was tested on real-world data with diverse backgrounds [5] shown in Figure (7).

4. Mouse Control Implementation:

a) Gesture to Mouse Mapping: Gestures were mapped to mouse actions:
Pointing: Cursor movement.
Left and right clicks.
Scroll & Zoom.
PyAutoGUI handled mouse events, with smoothing algorithms reducing cursor jitter [15].

b) Calibration & Testing: A calibration phase mapped physical hand movements to screen coordinates. Visual feedback improved user interaction, and latency was monitored for real-time responsiveness [10].

5. Virtual Keyboard Development:

a. UI & Framework: Figure (8) A virtual keyboard was built using Tkinter for its flexibility and modern interface. The layout was resizable and provided visual and auditory feedback [19].

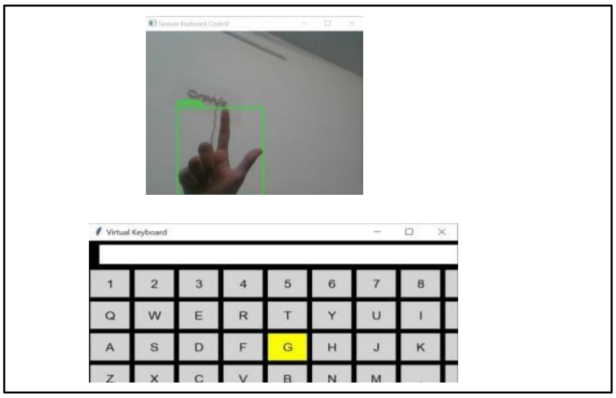


Figure 5: Pointing action of User

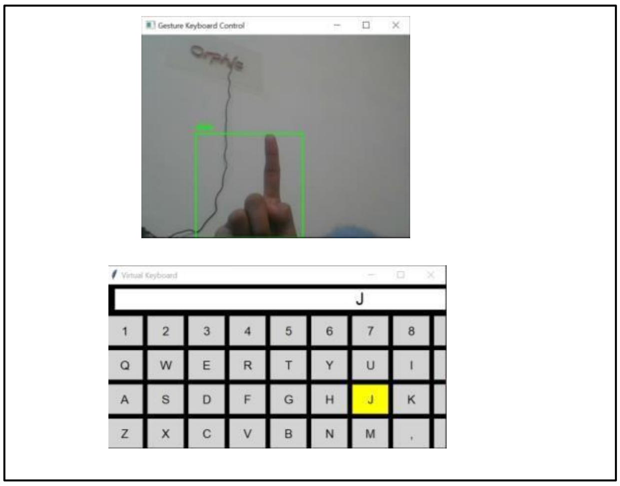


Figure 6: Clicking action of User

VI. RESULT DISCUSSION & ACCURACY

The virtual system developed can recognize eight gestures and interpret them into corresponding keyboard and mouse moments. The results implies that the proposed system is effective and easily accessible for individuals with physical disabilities. As compared to existing virtual system our system combines both mouse and keyboard and provides faster and accurate system to detect and translate gesture to mouse and keyboard action.

Model Performance during testing:

- mAP@50 = 98.6%: Excellent detection with a generous overlap requirement.
- mAP@50-95 = 91.4%: Strong detection across various overlap levels (stricter evaluation).
- Slightly lower MAP for "Right Click" (86.8%) and "Zoom Out" (90.1%).

The following graphs shows the test results:

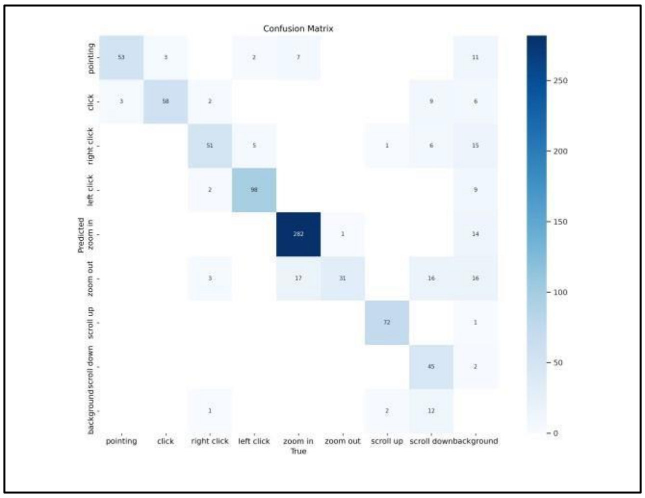


Figure 7: Model Performance Across Classes Showing Correct and Incorrect Predictions

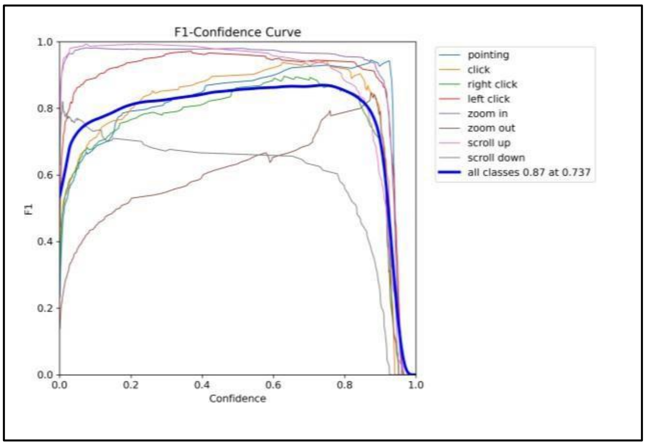


Figure 8: F1 - Curve Result

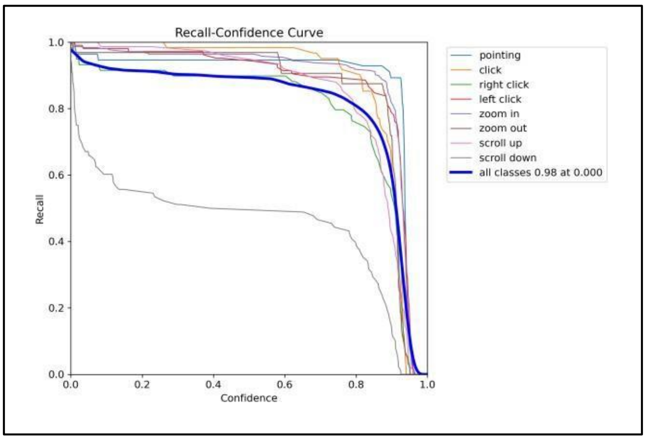


Figure G: Recall - Curve Result

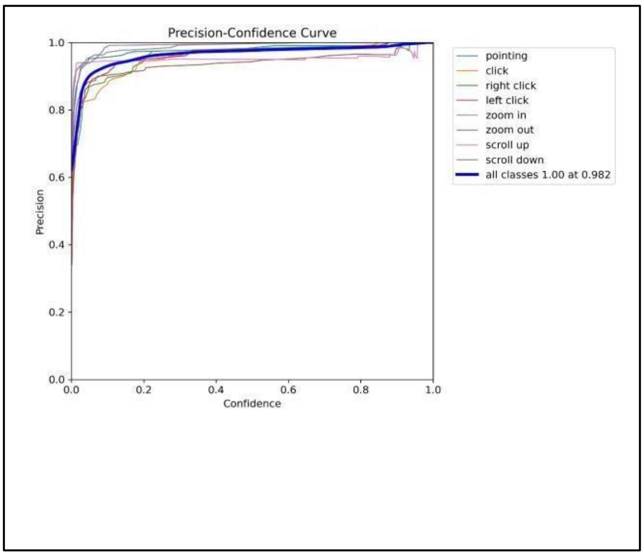


Figure 10: Precision - Curve Result

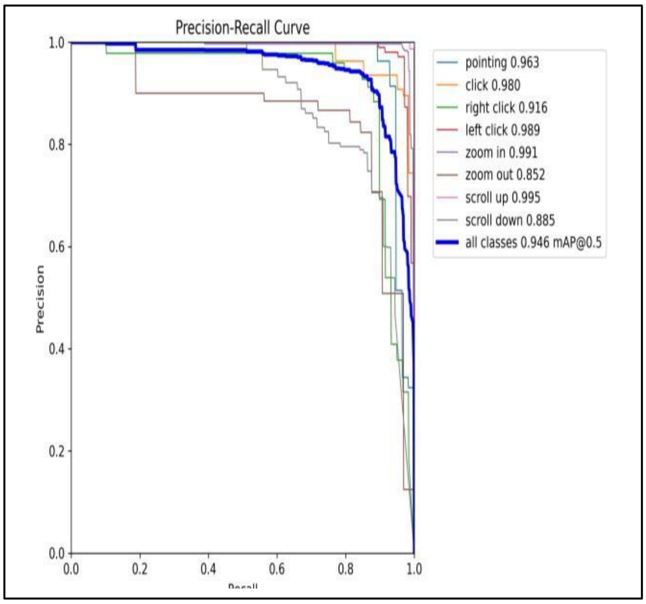


Figure 11: PR - Curve Result

Confusion matrix shown in Figure (7) displays model performance across classes, showing correct and incorrect predictions.

F1 Curve shown in Figure (8) is used to visualize balance between recall and precision for a model. It means how well the model performs in terms of completeness and correctness.

Recall Curve or R-Curve: Figure (9) plots either the model has seen everything or not.

Precision - Recall Curve or PR-Curve: Figure (11) plots precision against recall at different thresholds.

Precision curve: Figure (10) plots how correct (precisely) our model predicts while testing.

VII. CONCLUSION & FUTURE WORK

This work successfully implements an accessible and cost-effective virtual keyboard and mouse system using machine learning and gesture recognition. The system, powered by YOLOv8, achieves 92% accuracy in detecting eight predefined gestures, offering a reliable, touchless input method for users with physical disabilities. It is compatible with standard webcams and operates with minimal latency (~30-50 Ms per frame), ensuring smooth and responsive performance. Future work will focus on enhancing the system’s adaptability through improved lighting conditions, advanced gesture recognition, selecting of gestures for special letters and predictive text capabilities.

ACKNOWLEDGEMENTS

We extend our sincere gratitude to our fellow Fahad for his invaluable insights and collaborative spirit throughout this research. We deeply thank our department for their expert guidance, critical feedback, and unwavering support. Finally, we owe immense appreciation to our parents for their constant encouragement, motivation, and belief in our efforts. This work would not have been possible without their collective contributions.

REFERENCES

[1] A. L. D. K. Michael Roberts, "Accessibility Challenges in Input Devices," in *IEEE Transactions on Human-Machine Systems*, 2020.

[2] P. D. M. K. Tomasz Bednarz, "OpenPose: Real-Time Gesture Tracking," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019.

[3] H. L. J. G. Guangjun Jiang, "YOLO Architectures for Edge Devices," in *Springer Nature Machine Intelligence*, 2023.

[4] Y. L. Z. Z. Hao Wang, "Adaptive Gesture Recognition Under Dynamic Lighting," in *IEEE International Conference on Image Processing*, 2022.

[5] Q. C. W. Z. Yan Li, "Robust Gesture Recognition in Variable Lighting," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[6] J. C. M. L. Kai Sun, "Deep CNNs for Hand Gesture Classification," in *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

[7] C. G. L. S. Patricia Martinez, "Gaze-Gesture Hybrid Systems," in *IEEE Transactions on Systems, Man, and Cybernetics*, 2019.

[8] L. W. N. X. Zhiqiang Zhang, "YOLOv5 in Assistive Robotics," in *Springer Journal of Intelligent and Robotic Systems*, 2022.

[9] R. Szeliski, "Computer Vision Algorithms Using OpenCV," in *Springer*, 2021.

[10] R. M. S. C. James Davis, "Efficient ROI Extraction for Gesture Recognition," in *IEEE Sensors Journal*, 2022.

[11] J. E. R. T. Lisa Clark, "User-Centric Evaluation of Assistive Technologies," in *Springer Journal of Accessibility and Design for All*, 2021.

[12] M. B. E. D. Sarah Green, "Low-Cost Gesture Recognition Using Webcams," in *IEEE Sensors Journal*, 2021.

[13] S. D. R. G. A. F. Joseph Redmon, "YOLOv8: Real-Time Object Detection," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[14] M. G. T. W. Alex Brown, "YOLOv8 Performance Benchmarking," in *IEEE Transactions on Intelligent Transportation Systems*, 2023.

[15] J. D. E. W. Mark Smith, "PyAutoGUI for Assistive Automation," in *IEEE International Conference on Robotics and Automation*, 2020.

[16] K. S. B. J. Steven Thompson, "Click Gesture Design for Motor-Impaired Users," in *Springer Universal Access in the Information Society*, 2022.

[17] A. G. M. H. David Brown, "Scroll Gesture Recognition Using CNNs," in *IEEE Conference on Human Factors in Computing Systems*, 2023.

[18] D. T. E. M. Laura Perez, "Zoom Gestures for Touchless Control," in *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[19] L. M. P. T. Andrew King, "Predictive Text for Assistive Systems," in *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[20] J. E. R. T. Lisa Clark, "User-Centric Evaluation of Assistive Technologies," in *Springer Journal of Accessibility and Design for All*, 2021.

Cost-Effective Energy Management of A Microgrid Using A Hybrid Yellow Saddle Goatfish Optimization Algorithm

Asadullah Younas
Department of Electrical Engineering University of Engineering and Technology
Taxila, Pakistan
asadullahyounas1020@gmail.com

Tahir Mahmood
Department of Electrical Engineering University of Engineering and Technology
Taxila, Pakistan
tahir.mahmood@uettaxila.edu.pk

Muhammad Mansoor Ashraf
Department of Electrical Engineering University of Engineering and Technology
Taxila, Pakistan
mansoor.ashraf@uettaxila.edu.pk

***Abstract*—The increasing penetration of renewable energy sources in hybrid microgrids introduces challenges such as power fluctuations, system complexity, and high operational costs. This paper proposes an optimized energy management framework of microgrid integrating the Hybrid Yellow Saddle Goatfish Optimization Algorithm (HYSGA) with Sequential Quadratic Programming (SQP) to enhance system efficiency, stability, and cost-effectiveness. The proposed HYSGA approach optimally manages energy distribution among solar photovoltaic (PV) systems, battery energy storage systems (BESS), and the power grid while ensuring reliable and cost-effective operations. HYSGA quickly identifies near-optimal solutions for complex energy management problems, while SQP refines these solutions to enhance precision and convergence speed. Extensive simulations and comparative cost analyses validate the performance of the proposed framework. In a baseline scenario, the hybrid microgrid incurs an annual operational cost of $26,900, While Case I reduces this cost to $13,800 and achieve 49% savings. Further enhancement through HYSGA optimization lowers the cost to $13,430.08 and achieving a 50.118% savings. Additionally, comparative evaluations demonstrate that the HYSGA strategy outperforms traditional techniques such as Mixed-Integer Nonlinear Programming (MINLP) in terms of cost savings, computational efficiency, and solution accuracy. This study provides a comprehensive analysis of the research methodology, solution approach, and performance evaluation and ensuring clarity. The findings confirm that the HYSGA framework offers a scalable, computationally efficient, and economically viable solution for hybrid microgrid energy management. The proposed approach presents a promising pathway for improving energy efficiency and cost reduction in modern smart grid applications.**

Keywords—Microgrid, Energy Management, PV, BESS, HYSGA, SQP

I. INTRODUCTION

A microgrid combines power-consuming devices and small-scale energy sources to create a local power network. It operates independently, quickly responding to grid demands and can be customized for specific user needs. Benefits include improved reliability, reduced feeder losses, voltage support, efficient waste heat use, voltage sag correction, and uninterrupted power supply [1]. The U.S. Department of Energy defines MICROGRIDS through their definition of networks made from connected loads and distributed energy resources (DERs) and clear electrical boundaries. This single control unit demonstrates grid relationship capability while enabling on and off-grid operation throughout both modes [2]. Rising power demand necessitates renewables for energy expansion and sustainability. Microgrids facilitate decentralized management, enabling efficient coordination between consumers and generation units. By integrating local energy resources, they enhance resilience and reliability while supporting the transition to a greener energy landscape [3]. The information and control system of microgrids functions as energy management through delivering needed operational capabilities for generating distribution systems to supply power at minimal operational expenses [4].

Many different control methods exist to optimize grid-connectivity-based microgrids by enabling intelligent management of controllable energy resources and power loads [5]. The problem of sub-optimal microgrid energy management under stochastic conditions remains challenging for mixed-integer linear programming, linear programming and dynamic programming techniques due to their weak performance under high-dimensional systems [6]. These approaches demonstrate limitations in their capabilities of adapting to changes due to volatile control variables like load requirement shifts and power prices together with renewable energy pattern [7]. The energy management problems in microgrids receive solutions through the use and combination of metaheuristic techniques where particle swarm optimization (PSO) and genetic algorithm (GA) are integrated [8]. These optimization methods have long processing times, making them unsuitable for real-time applications. They lack the ability to learn and cannot retain knowledge for future tasks, leading to decreased computational efficiency as hourly load demands change [9]. The performance of these techniques gets negatively affected if they lack accurate models combined with appropriate state variable forecasting methods. The combination of metaheuristic methods with linear techniques happens frequently to achieve better complementary features [10].

Energy management in microgrids employs various advanced optimization techniques. These include programming approaches like linear, non-linear, mixed-integer, and robust programming. Common metaheuristic algorithms are Particle Swarm Optimization (PSO), Genetic Algorithms, Artificial Bee Colony, and Bacterial Foraging. Intelligent methods, such as Fuzzy Logic and Evolutionary Algorithms, further enhance performance. Additionally, Dynamic Programming, Game Theory, Model Predictive Control, and Multi-Agent Systems are utilized to improve decision-making in managing energy resources effectively [11]. A microgrid integrates distributed generators and loads to function as a single controllable unit, operating both on and off the main grid. However, managing hybrid microgrids is complex due to renewable variability and cost-efficiency demands. The problem of hybrid Micro Grid energy management is addressed through the combination of hybrid yellow saddle goatfish optimization with sequential quadratic programming algorithm in this article. The total operating cost of Microgrid gets minimized through optimization of decision variables $C_{grid}(\tau)$ and $C_{bat}(\tau)$ [12]. The article follows a structured format which includes Section II for problem mathematical modelling after which Section III presents hybrid Yellow Saddle Goatfish optimization with Sequential Quadratic Programming Algorithm and Section IV shows simulation statistical results and graphical data before the study concludes in Section V.

II. Problem Formulation

A. Objective Function

The Cost minimization for residential microgrids contains two essential cost elements called $C_{grid}(\tau)$ and $C_{bat}(\tau)$ which are represented numerically as an optimization problem in this research work [12]. The main objective requires minimizing system costs related to energy expenditures across various nodes throughout different time periods. The mathematical representation of this formula appears as

$$f(x) = min\sum_{r=1}^{r=24}[C_{grid}(\tau) + C_{bat}(\tau)] \quad (1)$$

The operational cost optimization model described in (1) evaluates the full 24-hour expenses between battery storage and grid power. Equations (2) through (6) determine battery expenditures while also assessing device depth of discharge and charging state effects on battery operation duration and grid system communication.

$$C^{r}_{bat} = -(C_{DoD}\Delta\wp(\tau)\Delta\tau) \quad (2)$$

$$C_{DoD} = C_{ici}\, \left| \frac{1}{L(DOD_2)} - \frac{1}{L(DOD_1)} \right| \Delta\wp(\tau)\Delta\tau \quad (3)$$

$$DOD(\tau) = 1 - SOC(\tau) \quad (4)$$

$$L(DOD) = \alpha DOD^{-6} \quad (5)$$

$$C_{grid}(\tau) = -(G_r(\tau)\mathcal{P}_p(\tau).\Delta\tau - \upsilon G_r(\tau)\mathcal{P}_s(\tau).\Delta\tau) \quad (6)$$

The state space set (S) contains all possible system states at each time step τ according to (7). The state variable ($\mathfrak{s}_r$) in (8) represents system conditions at time by incorporating PV power output along with SOC values and grid prices and load requirements. Equation (9) presents the state space that encompasses all system states over the complete time period. Equation (10) determines net load demand after factoring in PV-generated power to reveal remaining load requirements. The SOC management system uses (11) and (12) to control battery energy delivery by calculating residual load requirements alongside forecasting future SOC values while respecting SOC restrictions [12].

$$S = \{S_r\} \quad (7)$$

$$\mathfrak{s}_r = \{\tau, \mathcal{P}^{PV}_r, SOC_r, G_r, \mathcal{P}_{l,r}\} \quad (8)$$

$$\mathcal{S} = \mathfrak{s}_0 \cup \mathfrak{s}_1 \cup \mathfrak{s}_2 \cup \dots \dots \cup \mathfrak{s}_{T-1} \quad (9)$$

$$\mathcal{P}^{NET}_{ld,r} = max((\mathcal{P}_{ld,r} - \mathcal{P}^{pv}_r), 0) \quad (10)$$

$$\mathcal{P}^{REM}_{ld,r} = max(\mathcal{P}^{NET}_{ld,r} - (SOC_r - SOC^{min}).E, 0) \quad (11)$$

$$SOC^{NEXT}_r = min(SOC^{max}, (max(\mathcal{P}^{pv}_r - \mathcal{P}_{ld,r}, 0) + max\left(\frac{(Soc_r - Soc^{min}).E - \mathcal{P}^{NET}_{ld,r}}{E}, 0\right)) \quad (12)$$

$$Action_{\mathfrak{s}_r} = \{-\mathcal{K}\Delta\wp, \dots, -\Delta\wp, 0, \Delta\wp, \dots, \mathcal{K}\Delta\wp\} \quad (13)$$

$$B_{ESS}(\mathcal{A}_r) = \begin{cases} -\frac{\mathcal{K}(\mathcal{A}_r)}{E}, & if\ \mathcal{A}_r = discharge \\ \frac{\mathcal{K}(\mathcal{A}_r)}{E}, & if\ \mathcal{A}_r = charge \end{cases} \quad (14)$$

Time τ updates the grid cost calculation through (15) by consolidating information about both remaining load requirements and battery energy storage system (BESS) functions. Equation (16) integrates the remaining load and BESS discharge into the cost computation through its consideration of the grid price multiplier. The calculation of battery cost in (17) considers net load requirements and State of Charge constraints to support grid pricing within battery operational boundaries. The opportunity cost from SOC limit deviations in (18) creates penalties for excessive charging or discharging. The reward function for states ($\mathfrak{s}_r$) and actions ($\mathcal{A}_\tau$) includes all cost components through (19).

$$C^{r}_{grid} = -(\mathcal{P}^{REM}_{ld,r} + B_{ESS}(\mathcal{A}_r).E).G_r \quad (15)$$

$$C^{r}_{p} = \mathcal{P}^{REM}_{ld,r} + B_{ESS}(a_{r.discharge}).E).G_r.\upsilon \quad (16)$$

$$C^{r}_{b} = \begin{cases} \mathcal{P}^{NET}_{ld,r}.G_r & if\ \mathcal{P}^{NET}_{ld,r} \le (SOC_r - SOC^{min}).E \\ else\ (SOC_r - SOC^{min}).G_r.E \end{cases} \quad (17)$$

$$C^{\tau}_{0} = \begin{cases} -((SOC_\tau + \mathcal{K}(\mathcal{A}_\tau) - SOC^{max}).G_r.E & if\ (SOC_\tau + \mathcal{K}(\mathcal{A}_\tau) > SOC^{max} \\ -(|\mathcal{K}(\mathcal{A}_\tau)| - (SOC_\tau + SOC^{max})).G_r.E & else\ if\ (SOC_\tau + \mathcal{K}(\mathcal{A}_\tau) < SOC^{min} \\ else\ 0 \end{cases} \quad (18)$$

$$\mathcal{R}(\mathfrak{s}_r, \mathcal{A}_r) = C^{r}_{grid} + C^{r}_{bat} + C^{r}_{b} + C^{r}_{0} \quad (19)$$

Equation (20) defines the value function ($\upsilon^{\pi}_{r}(\mathfrak{s})$), representing the cumulative reward starting from state ($\mathfrak{s}_r$) and action ($\mathcal{A}_r$), discounted over future time steps. This value function helps in evaluating the long-term effectiveness of different strategies by summing immediate rewards and future discounted rewards. It guides the optimization process towards maximizing overall performance and cost efficiency [12].

$$\mathcal{V}^{\pi}_{r}(\mathfrak{s}) = \mathcal{R}(\mathfrak{s}_r, \mathcal{A}_r) + \sum_{i=1}^{\infty} \Upsilon^{i}.\mathcal{R}(\mathfrak{s}_{r+1}, \mathcal{A}_{r+1}) \quad (20)$$

III. Yellow Saddle goatfish optimization algorithm

Zaldivar et al. [13] published the Yellow Saddle Goatfish Optimization Algorithm (YSGA) during 2018. The Yellow Saddle Goatfish exhibits an unusual cooperative behavior which serves as one of their most intriguing collaborative methods in fish. A hunting strategy sees all fish organized into smaller groups which evenly divide the entire exploitation zone. Each sub-population contains all fish working together in a dual role as Chaser and blocker to perform the hunt.

The YSGA mathematical model developed by the author employs objective function variables $C_{grid}(\tau)$ and $C_{bat}(\tau)$ to execute the cost-effective management of microgrids.

A. First Phase: Exploration

The goatfish population $\mathfrak{p}$ consists of x individuals $\{\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{p}_3 \dots, \mathfrak{p}_x\}$, uniformly distributed within a y-dimensional search space defined by upper boundary $\mathcal{B}^{high}$ and lower boundary $\mathcal{B}^{low}$. Each individual $\mathfrak{p}_i$ is represented as a vector of decision variables $\{\mathfrak{p}_i^1, \mathfrak{p}_i^2, \mathfrak{p}_i^3 \dots, \mathfrak{p}_i^x\}$. Equation (21) defines the initialization.

$$\mathfrak{p}_i^j = rand.(\mathcal{B}_j^{high} - \mathcal{B}_j^{low}) + \mathcal{B}_j^{low} \qquad (21)$$
$$i = 1,2, \dots, x; \quad j = 1,2, \dots, y$$

Where $rand$ is a random number among [0,1].

Equation (22) defines the squared error between μ_ℓ and the data points $\{\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{p}_3 \dots, \mathfrak{p}_h\}$ in cluster $\mathcal{C}_\ell$ using the population of goatfish P as the data set.

$$e(\mathcal{C}_\ell) = \sum_{\mathfrak{p}_g \in \mathcal{C}_\ell} \| \mathfrak{p}_g - \mu_\ell \|^2 \qquad (22)$$
$$g = 1,2, \dots, h; \quad \ell = 1,2, \dots, \kappa$$

The objective of k-means is to minimize the objective function, which is the total sum of squared errors across all k clusters, as expressed by (23).

$$E(\mathcal{C}) = \sum_{\ell=1}^{\kappa} e(\mathcal{C}_\ell) \qquad (23)$$

In a goatfish school, there is a sole chaser fish $\varphi_\ell \in \mathfrak{p}$ tasked with leading the pursuit, selected based on its fitness value. Within each group, the particle with the highest fitness value should be closest to the solution, a behavior demonstrated through Lévy flight, a non-Gaussian random process employing a Lévy stable distribution for random walks. The chaser fish employs random walks to alter its position, aiming to find concealed prey in crevices. Equation (24) determines the chaser fish's new position.

$$\varphi_\ell^{r+1} = \varphi_\ell^r + \alpha \oplus \text{Lévy}(\beta) \qquad (24)$$
$$0 < \beta \leq 2$$

The chaser fish's new and current locations are denoted by φ_ℓ^{r+1} and φ_ℓ^r respectively. α represents the step size, $\oplus$ denotes entry-wise multiplication and β is the Lévy index, which governs the shape of the probability distribution, particularly the tail. Equation (25) determines the value of β

$$\beta = 1.99 + \frac{0.001r}{r_{max}/10} \qquad (25)$$

Typically, each group disregards other sub-populations, focusing only on the group that captures the best prey. This conduct is expressed in (26).

$$S = \alpha \oplus \text{Lévy}(\beta) \sim \alpha\left(\frac{u}{|v|^{1/\beta}}\right)(\varphi_\ell^r - \varphi_{best}^r) \qquad (26)$$

In the given scenario, S represents the random step, while φ_{best}^r denotes the most successful chaser fish among all the clusters.

Using Γ as the Gamma function, the parameters σ_u and σ_v are formulated in (27)

$$\sigma_u = \left\{\frac{\Gamma(1+\beta)\sin\frac{\pi\beta}{2}}{\Gamma\left(\frac{1+\beta}{2}\right)\beta 2^{(\beta-1)/2}}\right\}^{1/\beta}, \quad \sigma_v = 1 \qquad (27)$$

On behalf of given statements, the revised location of the chaser fish outlined in (24) can be restated in (28).

$$\varphi_\ell^{r+1} = \varphi_\ell^r + S \qquad (28)$$

Consequently, the new position of the best chaser fish is computed using (29).

$$\varphi_{Best}^{r+1} = \varphi_{Best}^r + S' \qquad (29)$$

Where, S' is formulated in (30)

$$S' = \alpha\left(\frac{u}{|v|^{1/\beta}}\right) \qquad (30)$$

The strategy used by blocker fish $\Phi_g \in \mathfrak{p}$ when hunting is to encircle the corals in order to cut off escape routes for prey, while chaser fish attempt to capture the prey. Fish attempt to capture prey using spiral algorithm. Equation (31) calculates the new location of the blocker fish Φ_g^{r+1} based on the algorithmic spiral.

$$\Phi_g^{r+1} = \mathcal{D}_g \cdot e^{b\rho} . \cos 2\pi\rho + \varphi_\ell \qquad (31)$$

In this context, the distance between the blocker and the chaser fish along the spiral path is determined by the random number ρ, which falls within the range of [$\boldsymbol{a}$, 1]. To enhance exploitation, $\boldsymbol{a}$ linearly decreases from -1 to -2 as the iteration count increases. The parameter b is a fixed value that dictates the form and orientation of the spiral, for this method, b is set to 1. Equation (32) computes the distance $\mathcal{D}_g$ between the current position of the blocker fish Φ_g^r and the chaser fish in the cluster $\mathcal{C}_\ell$.

$$\mathcal{D}_g = |r . \varphi_\ell - \Phi_g^r|$$
$$\{\varphi_\ell, \; \Phi_g^r\} \in \mathcal{C}_\ell \qquad (32)$$

Where r is the random number among [-1,1].

After fully depleting the hunting area, the group shifts to a new location in search of more prey. The YSGA model incorporates an over-exploitation parameter (λ), and if no better solution is found after λ iterations, the hunt is deemed successful, prompting the goatfish in the cluster to relocate. This is described in (33).

$$\mathfrak{p}_g^{\tau+1} = \frac{\varphi_{Best} + \mathfrak{p}_g^\tau}{2} \qquad (33)$$

The Fig. 1 represents an optimization framework for a microgrid test system, integrating a Yellow Saddle Goatfish Optimization Algorithm (YSGA) with Sequential Quadratic Programming (SQP) [14]. It outlines exploration and exploitation phases to enhance energy management, optimize power distribution, and improve system efficiency. This hybrid approach ensures optimal decision-making for reliable

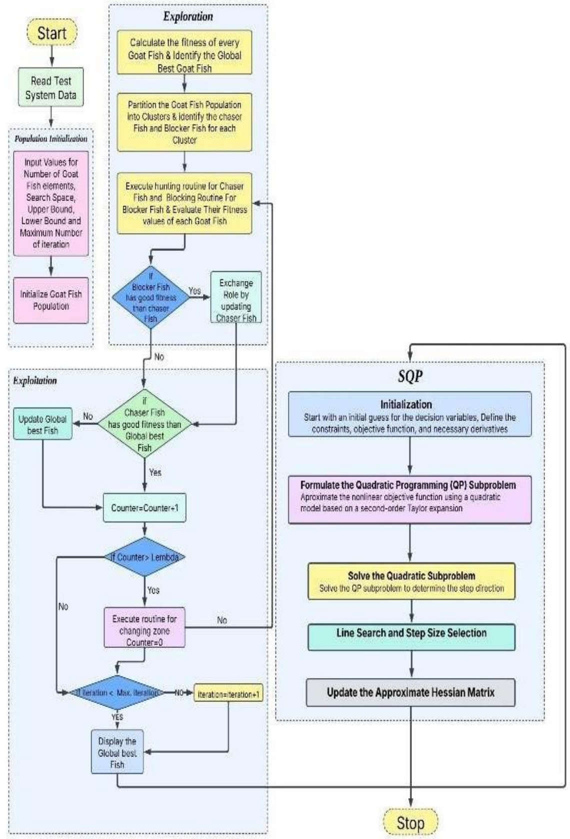


Fig. 1 Flow Chart of HYSGA with SQP

microgrid operations.

Table I defines the decision variables and constants used in the optimization process for the cost-effective energy management of microgrid.

TABLE I OPTIMAL DECISION VARIABLE VALUES FOR HYSGA

Parameter	***Value***
C^{r}_{total}	13430.08 \$/Year
Cost Saving	50.118%
E_{BESS}	12 KWh
C^{r}_{bat}	\$2280
η_c, η_d	89%
DOD(τ)	20% - 80%
Battery Lifetime	7 Years
C^{r}_{grid}	13430.08 \$/Year
Payback Period	2 Months
ROI	589%
T	24Hours/day
Δt	1 Hour
SOC_r	20% - 90%
$\mathcal{P}^{PV}_{r}$	5 KW
$\mathcal{P}_{l,r}$	40 KW
$\mathcal{P}^{NET}_{ld,r}$	15 - 30 KWh
$\mathcal{P}^{REM}_{ld,r}$	5 - 15 KWh
r_r	\$0.07 -\$0.22/KWh
υ	10.8%
CSC	\$7.55/Month

IV. RESULTS AND DISCUSSION

This section evaluates cost-effective energy management of Hybrid Microgrid through the combination of Hybrid Yellow Saddle Goatfish Optimization Algorithm (HYSGA) with Sequential Quadratic Programming (SQP) to assess its performance in energy management issues. The total operating cost of Microgrid gets minimized through optimization of decision variables $C_{grid}(\tau)$ and $C_{bat}(\tau)$. issue. HYSGA-SQP's effectiveness is also compared with other established optimization algorithms, highlighting its reliability in addressing optimization issues.

Figure 2 shows a hybrid microgrid with a Power Grid, PV array, BESS, and Load. It illustrates efficient power flow among these components. The PV array powers both the Load and charges the BESS. The BESS stores excess energy and supplies it when needed. The Grid supports the Load and BESS during low solar output. This setup ensures continuous and reliable power delivery. Bidirectional flow between BESS and the grid adds flexibility. The system enhances efficiency and supports renewable integration.

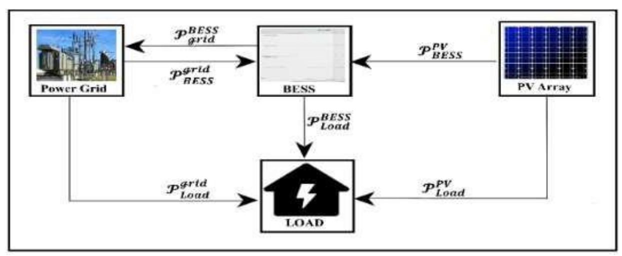


Fig.2 Microgrid Test System

Figures 3a and 3b emphasize key elements of cost-effective energy management in hybrid microgrids. Figure 3a indicates that as energy costs rise, savings decrease, highlighting the importance of utilizing low-cost renewables like solar and wind, as well as efficient energy storage for enhanced flexibility. Figure 3b shows a significant reduction in computational time with each optimization iteration, reflecting improved algorithm efficiency and faster convergence for real-time decision-making.

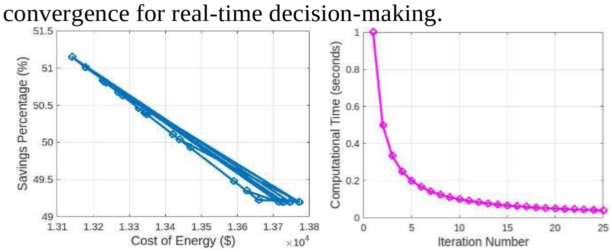


Fig. 3 Cost of Energy^{FIRST} vs Savings Percentage^{SECOND} Iteration Number vs Computational Time

Figure 4 shows the relationship between energy costs and savings percentage during multiple optimization iterations in hybrid microgrid management. The blue curve represents energy costs, while the red curve indicates savings. Their fluctuations reflect the algorithm's efforts to balance cost minimization and savings maximization, highlighting the complexity of achieving equilibrium. This dynamic behavior emphasizes the need for adaptive energy dispatch strategies based on load demand and renewable energy availability. Overall, the figure illustrates improved responsiveness and performance in hybrid microgrid operations.

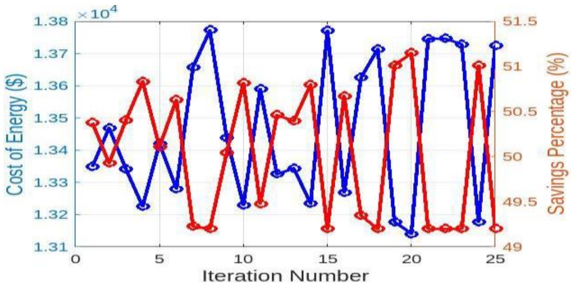


Fig. 4 Iteration Number vs Cost of Energy and Savings Percentage

Figure 5 illustrates the savings percentage during hybrid microgrid optimization, fluctuating between 49% and 51.5%. This variation reflects the algorithm's adaptive search for cost-effective solutions, typical of heuristic methods utilizing iterative energy dispatch adjustments. Despite minor oscillations, the overall high savings indicate a strong energy management strategy. The results highlight trade-offs in cost optimization while maintaining system stability and show the system's responsiveness to changing load and energy availability. Ultimately, the process aims for reliable and economical operation.

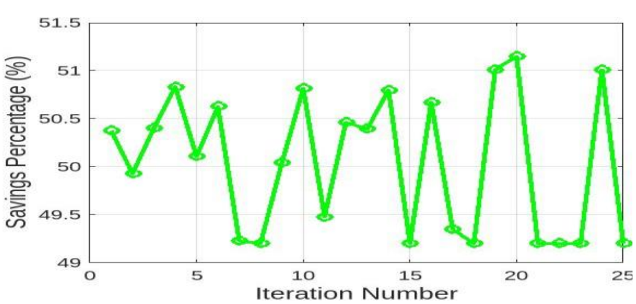


Fig. 5 Iteration Number vs Savings Percentage

Figure 6 shows the baseline load profile of a hybrid microgrid over 34 hours, highlighting daily demand fluctuations. These peaks and dips are crucial for demand-side management. During high demand, utilizing storage or renewables can reduce grid reliance, while low-demand periods are ideal for charging batteries or shifting loads. Understanding these patterns enhances forecasting and planning, aligning energy use with demand to improve cost-efficiency and reliability.

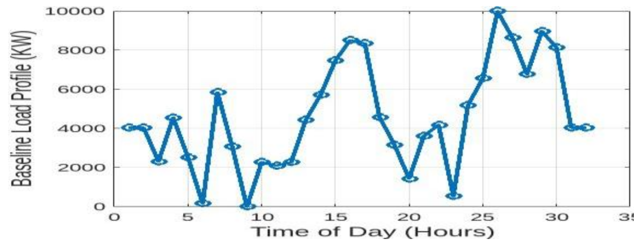


Fig. 6 Baseline Load Profile vs Time of Day

Figure 7 shows power output variations in a hybrid microgrid, highlighting dynamic energy usage. A dip near zero between the 15th and 20th points suggests load shedding or reliance on stored energy. These fluctuations help inform decisions to minimize generation or switch to cheaper sources. Analyzing these trends supports cost-effective energy planning and optimizes the use of renewables and battery storage, ensuring availability and reducing costs.

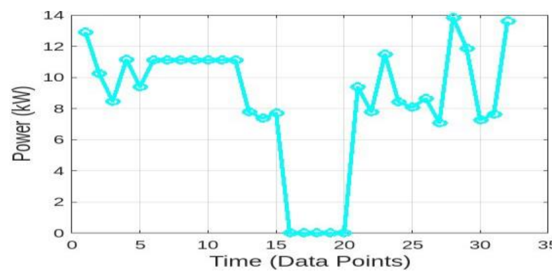


Fig. 7 Power vs Time (32 Data Points)

Figure 8 shows fitness values from different optimization runs for hybrid microgrid energy management, illustrating performance variability. Higher fitness indicates more effective energy strategies; run 7 achieves the highest value (13514.6), while run 16 shows the lowest (13358.0), reflecting less efficiency. This variation highlights the algorithm's stochastic nature. Multiple simulations promote convergence toward optimal solutions, improving decision-making for cost savings and system performance, underscoring the value of robust optimization in hybrid microgrid management.

Fig. 8 Fitness Value Vs Number of Runs

TABLE II ANNUAL ENERGY COST AND COST SAVINGS ASSOCIATED WITH BASELINE, CASE I, YSGF ALGORITHM AND PROPOSED HYSGF WITH SQP

Parameter	*Energy Cost ($/Year)*	*Cost Savings (%)*
Baseline	26900 [15]	-
CASE 1	13800 [15]	49% [15]
Proposed Hybrid YSGA with SQP	**13430.08**	**50.118%**

TABLE III RUN-WISE COST TO SAVINGS COMAPRISON FOR BASELINE AND CASE I FOR [15] V/S PROPOSED HYSGA

Run No	***Parameter***	***Baseline***	***CASE 1***	***HYSGA with SQP***
1	Energy Cost ($/Year)	26900	13800	13429.16
	Cost Savings (%)	-	49%	50.12%
2	Energy Cost ($/Year)	26900	13800	13434.32
	Cost Savings (%)	-	49%	50.10%
3	Energy Cost ($/Year)	26900	13800	13438.12
	Cost Savings (%)	-	49%	50.09%
4	Energy Cost ($/Year)	26900	13800	13385.76
	Cost Savings (%)	-	49%	50.25%
5	Energy Cost ($/Year)	26900	13800	13479.0
	Cost Savings (%)	-	49%	49.99%
6	Energy Cost ($/Year)	26900	13800	13422.12
	Cost Savings (%)	-	49%	50.16%
7	Energy Cost ($/Year)	26900	13800	13437.28
	Cost Savings (%)	-	49%	50.07%
8	Energy Cost ($/Year)	26900	13800	13473.96
	Cost Savings (%)	-	49%	49.95%
9	Energy Cost ($/Year)	26900	13800	13416.92
	Cost Savings (%)	-	49%	50.15%
10	Energy Cost ($/Year)	26900	13800	13412.32
	Cost Savings (%)	-	49%	50.18%
11	Energy Cost ($/Year)	26900	13800	13451.60
	Cost Savings (%)	-	49%	50.04%
12	Energy Cost ($/Year)	26900	13800	13412.72
	Cost Savings (%)	-	49%	50.19%
13	Energy Cost ($/Year)	26900	13800	13360.04
	Cost Savings (%)	-	49%	50.36%
14	Energy Cost ($/Year)	26900	13800	13426.12
	Cost Savings (%)	-	49%	50.11%
15	Energy Cost ($/Year)	26900	13800	13375.52
	Cost Savings (%)	-	49%	50.29%
16	Energy Cost ($/Year)	26900	13800	13438.68
	Cost Savings (%)	-	49%	50.09%
17	Energy Cost ($/Year)	26900	13800	13467.88
	Cost Savings (%)	-	49%	49.99%
18	Energy Cost ($/Year)	26900	13800	13495.20
	Cost Savings (%)	-	49%	49.91%
19	Energy Cost ($/Year)	26900	13800	13425.16
	Cost Savings (%)	-	49%	50.15%
20	Energy Cost ($/Year)	26900	13800	13472.12
	Cost Savings (%)	-	49%	49.94%

Table II compares energy management strategies for a microgrid, showing the baseline scenario costs $26,900, while Case I reduces this to $13,800 for 49% savings. Table III highlights cost-to-savings ratios over 25 iterations, with the HYSGA using SQP performing best, lowering costs to $13,430.08 and achieving maximum savings of 50.118%. These results confirm the algorithm's effectiveness and establish HYSGA with SQP as the optimal strategy.

V. CONCLUSION.

The proposed research introduces a cost-effective energy management system (EMS) for hybrid microgrids using Hybrid Yellow Saddle Goatfish Optimization Algorithm (HYSGA) with Sequential Quadratic Programming (SQP). It optimizes energy scheduling in stochastic conditions, considering battery degradation cost and Time of Use (ToU) grid tariffs costs in a single cost function. The performance of the Hybrid Yellow Saddle Goatfish Optimization Algorithm (HYSGA) with Sequential Quadratic Programming (SQP) is compared to the baseline scenario with the grid as the sole energy supplier. In the baseline case relying entirely on grid power importation incurs an annual energy cost of $26,900. The microgrid retrofit of Case I reduces the cost of energy to $13800 and achieves 49% energy cost savings. The proposed Hybrid Yellow Saddle Goatfish Optimization (HYSGA) with Sequential Quadratic Programming (SQP) algorithm reduces the cost of energy $13430.08 and achieves 50.118% energy cost savings. These results show the effectiveness of proposed Hybrid Yellow Saddle Goatfish Optimization Algorithm (HYSGA) with Sequential Quadratic Programming (SQP) effectively improves cost-effectiveness in the management of microgrid energy systems.

REFERENCES

[1] Meteab, W.K., Alsultani, S.A.H. and Jurado, F., 2023. Energy Management of Microgrids with a Smart Charging Strategy for Electric Vehicles Using an Improved RUN Optimizer. *Energies*, *16*(16), p.6038.

[2] Guo, C., Wang, X., Zheng, Y. and Zhang, F., 2022. Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning. *Energy*, *238*, p.121873.

[3] Mandal, S. and Mandal, K.K., 2020. Optimal energy management of microgrids under environmental constraints using chaos enhanced differential evolution. *Renewable Energy Focus*, *34*, pp.129-141.

[4] Yuan, Z.P., Xia, J. and Li, P., 2021. Two-time-scale energy management for microgrids with data-based day-ahead distributionally robust chance-constrained scheduling. *IEEE Transactions on Smart Grid*, *12*(6), pp.4778-4787.

[5] Prudhviraj, D., Kiran, P.B.S. and Pindoriya, N.M., 2019. Stochastic energy management of microgrid with nodal pricing. *Journal of Modern Power Systems and Clean Energy*, *8*(1), pp.102-110.

[6] Aghajani, G.R. and Heydari, I., 2021. Energy management in microgrids containing electric vehicles and renewable energy sources considering demand response. *J. Oper. Autom. Power Eng*, *9*(1), pp.34-48.

[7] Jordehi, A.R., Javadi, M.S. and Catalao, J.P., 2020. Energy management in microgrids with battery swap stations and var compensators. *Journal of Cleaner Production*, *272*, p.122943

[8] Charadi, S., Chakir, H.E., Redouane, A., El Hasnaoui, A. and El Bhiri, B., 2023. A Novel Hybrid Imperialist Competitive Algorithm–Particle Swarm Optimization Metaheuristic Optimization Algorithm for Cost-Effective Energy Management in Multi-Source Residential Microgrids. *Energies*, *16*(19), p.6896.

[9] Zhang, Y., Meng, F., Wang, R., Zhu, W. and Zeng, X.J., 2018. A stochastic MPC based approach to integrated energy management in microgrids. *Sustainable cities and society*, *41*, pp.349-362.

[10] Wamalwa, Fhazhil, and Ariane Ishimwe. "Optimal energy management in a grid-tied solar PV-battery microgrid for a public building under demand response." *Energy Reports* 12 (2024): 3718-3731.Parhizi, Sina, Hossein Lotfi, Amin Khodaei, and Shay Bahramirad. "State of the art in research on microgrids: A review." *IEEE access* 3 (2015): 890-925.

[11] Mansouri, S.A., Ahmarinejad, A., Nematbakhsh, E., Javadi, M.S., Jordehi, A.R. and Catalao, J.P., 2021. Energy management in microgrids including smart homes: A multi-objective approach. *Sustainable Cities and Society*, *69*, p.102852.

[12] Muriithi, Grace, and Sunetra Chowdhury. 2021. "Optimal Energy Management of a Grid-Tied Solar PV-Battery Microgrid: A Reinforcement Learning Approach" *Energies* 14, no. 9: 2700. https://doi.org/10.3390/en14092700.

[13] Zaldivar, Daniel, et al. "A novel bio-inspired optimization model based on Yellow Saddle Goatfish behavior." *Biosystems* 174 (2018): 1-21.

[14] Angalaeswari, S., and K. Jamuna. "Optimal energy management using sequential quadratic programming algorithm for stand alone PV system." *Int. J. Appl. Eng. Res* 12 (2017): 12250-12255.

[15] Wamalwa, Fhazhil, and Ariane Ishimwe. "Optimal energy management in a grid-tied solar PV-battery microgrid for a public building under demand response." *Energy Reports* 12 (2024): 3718-3

A Compact Slotted Micro-Strip Patch Antenna Operating at 28 GHz for 5G-IoT Applications

Ali Hassan
Electrical Engineering Department
HITEC University
Taxila, Pakistan
ali_3641@yahoo.com

N. Nizam-Uddin
Biomedical Engineering Department
HITEC University
Taxila, Pakistan
nizam.uddin@hitecuni.edu.pk

***Abstract*— This paper aims to present a compact slotted microstrip patch antenna for 5G-IoT applications operating at 28 GHz frequency. The antenna structure is modeled on an FR4 substrate with a compact size of 12 mm × 13 mm, (substrate height = 1.6 mm, Epsilon = 4.3, and Loss tangent = 0.02). The antenna comprises a patch on top of a dielectric substrate and a defected ground plane (DGS) on the bottom side. Slots and curves are incorporated in the patch radiator to achieve desired resonating frequency of 28 GHz. Simulation results demonstrate a return loss of –22 dB, a bandwidth of 4.64 GHz, a VSWR of 1.16, a gain of 3.2 dBi, and an efficiency of 60%. These attributes make the antenna appropriate for a range of 5G-IoT applications including smart cities, industrial IoT and autonomous systems where high data throughput and reliable connectivity are essential. The overall results depicts that the proposed design is a good candidate for deployment in 5G-enabled IoT ecosystems.**

***Keywords*—IoT, 5G, Patch Antenna, Millimeter Wave.**

I. Introduction

To keep up with the most recent market trends, the mobile industry is always creating new portable devices. Today's mobile devices are equipped with features like multimedia streaming and fast internet browsing thanks to the Long-Term Evolution (LTE) standard, commonly referred to as 4G. Modern 4G communication technology has made it possible for portable devices such as laptops and notebooks to use LTE bands like 700, 2400, and 2600 MHz. In light of this, a variety of LTE antenna designs have recently been published [1-2].

With the increasing demand for rich multimedia files, the existing communication regime is shifting toward connections with faster data rates. The increased processing power provided by nano-electronic devices and components supports this evolution. However, as a result of the congestion in the lower frequency bands now utilized by numerous communication networks, industry stakeholders are being compelled to investigate other spectrum pools. Unlicensed ultra-wideband (UWB) bands between 3.1 and 10.6 GHz are being explored for indoor use because of their shorter working ranges [3]. UWB antennas have drawn a lot of attention from both industry and academics due to its possible application in wireless transmission systems. To improve spectrum coverage and reduce interference from co-channel signals inside the UWB bandwidth, such as WLAN (IEEE-802.11a), WiMAX (IEEE-802.16), and other narrow-band technologies, these antennas are being improved utilizing a variety of techniques [4]. However, UWB antennas lack the versatility to cover higher frequencies, therefore they are only able to operate in the lower bands of 700–2600 MHz [5].

Fifth Generation (5G) communications have the potential to revolutionize mobile connectivity and facilitate the smooth integration of the Internet of Things (IoT). Reliable, low-latency communication with high bandwidth is becoming more and more necessary as the IoT is introduced into areas including smart cities, healthcare monitoring, industrial automation, and autonomous systems. The intricate traffic patterns of the IoT and the exponential growth in data transmission were not anticipated by the design of current 4G networks [6-7].

5G communication uses millimeter-wave (mm-Wave) bands from 20 to 300 GHz to meet this demand, initially allocating between 24 and 57 GHz. These bands are perfect for IoT use cases that demand extremely fast data transfer, extensive connection, and low power consumption since they provide large, unexplored spectrum pools. The development of 5G networks enabled by the IoT depends on the design of effective and compact antennas for the mm-Wave band [8].

This next generation of devices needs to be able to operate at mm-Wave frequencies, be multi-mode, and be energy efficient [9]. Antenna design for 5G has accelerated, especially for the 28 GHz and 38 GHz spectrum, which are essential for both 5G and IoT applications. It is necessary to overcome challenges including air absorption-induced propagation losses while maintaining high gain, efficiency, and a bandwidth greater than 1 GHz. For mm-Wave bands, printed antenna technology has become a popular choice because of its low profile, affordability, broad bandwidth, and simplicity of fabrication [10-12].

In this study, we examine and develop a 5G-optimized Slotted Microstrip Patch Antenna (SMPA). The proposed antenna is a viable option for deployment in 5G-IoT applications since it provides a compact and reliable solution for the 28 GHz band.

II. Methodology (Design of Antenna)

The methodology and designing process of the proposed antenna comprises multiple stages as shown in

Fig. 1. The antenna structure is modelled on an FR4 substrate with a compact size of 12 mm × 13 mm (substrate height=h= 1.6 mm, Epsilon = 4.3, and Loss tangent = 0.02). The antenna is designed and simulated in CST Studio Suite 2019. The estimated Length (L) and Width (W) of the antenna is calculated by using the Eq. 1-6 [13, 14].

$$W = \frac{c}{2 \times f_r} \times \sqrt{\frac{2}{\varepsilon_r + 1}} \quad (1)$$

where c is the velocity of light, f_r is the resonance frequency and ε_r is the Substrate's dielectric constant. The effective dielectric constant is given as:

$$\varepsilon_{reff} = \left(\frac{\varepsilon_r + 1}{2}\right) + \frac{\varepsilon_r - 1}{\left(2\sqrt{1 + 12\frac{h}{W}}\right)} \quad (2)$$

After calculating this the patch effective length (L_{eff}) is determined as:

$$L_{eff} = \frac{c}{2 \times f_r \times \sqrt{\varepsilon_{reff}}} \quad (3)$$

Now, to determine the amount by which the patch needs to be shortened, the following relation is used:

$$\Delta L = (0.412h)\,\frac{(\varepsilon_{reff} + 0.3)\left(\frac{W}{h} + 0.264\right)}{(\varepsilon_{reff} - 0.258)\left(\frac{W}{h} + 0.8\right)} \quad (4)$$

The final length of patch is calculated by:

$$L = L_{eff} - 2\Delta L \quad (5)$$

The dimensions of ground plane are calculated by using the following relations:

$$L_g = L + 6h \quad (6)$$

$$W_g = W + 6h \quad (7)$$

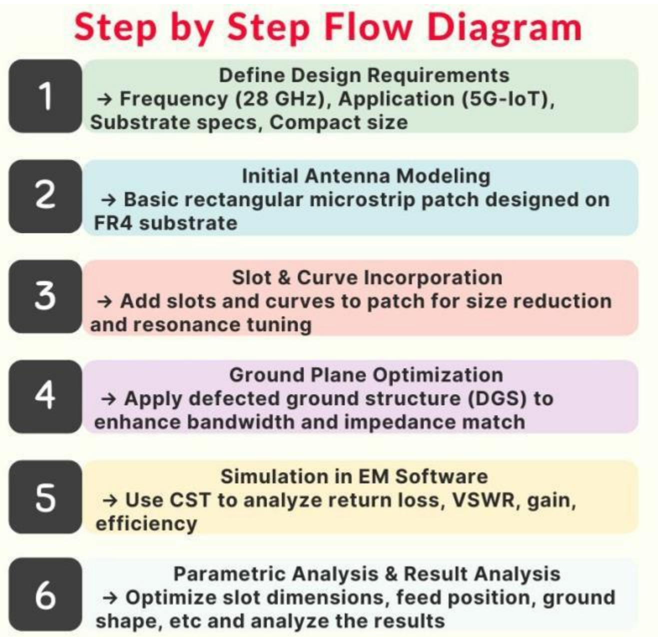


Fig. 1. Methodology Flow Diagram

The SMPA contains a patch on top of the dielectric substrate and a ground plane on the bottom side. Slots are inserted on the patch radiator to enhance the impedance matching. To increase the performance of the antenna we used the technique of DGS.

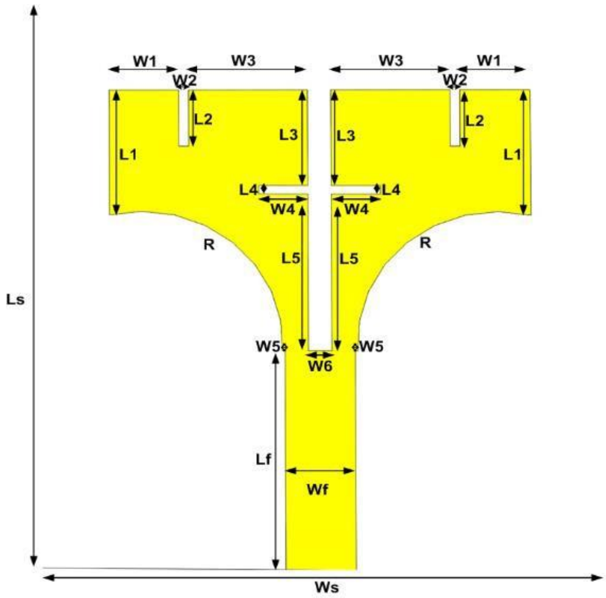


Fig. 2. Physical Layout of Patch

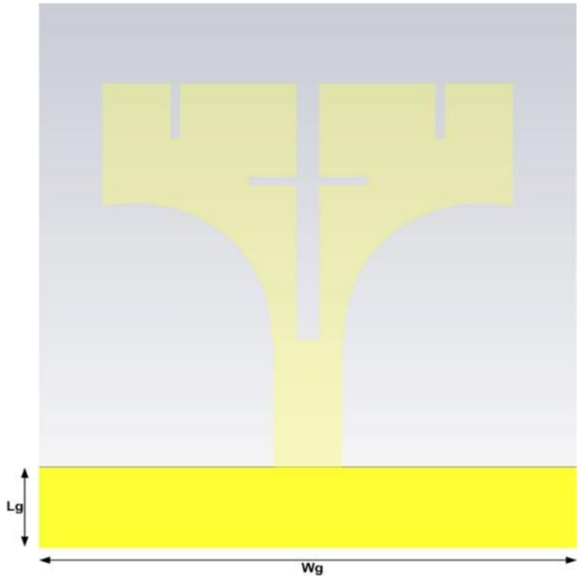


Fig. 3. Physical Layout of Ground

The Gain and Bandwidth are improved by using DGS. Figs. 2 and 3 presents the structure of the radiating patch and the ground plane of the proposed antenna. Table I shows the optimum dimensions of the proposed antenna.

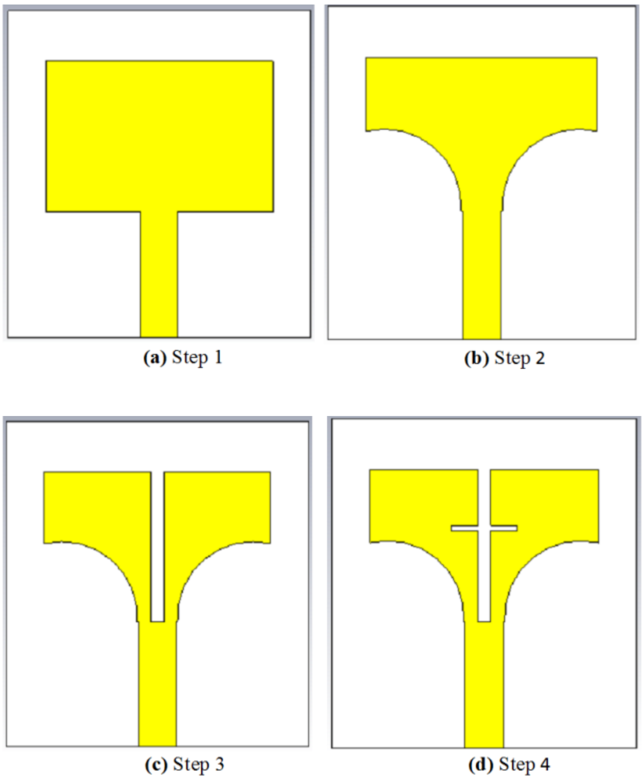
A. *Evolution of Design*

The design steps for the proposed antenna are illustrated in Fig. 4. In the first step, the general microstrip patch antenna does not show a fruitful response. In the second step, a curve with a radius (R = 3 mm) is introduced at both the bottoms of the patch which resulted in an improvement in the resonating frequency. In the third step a vertical slot of width W_6 is introduced in the patch to further improve the results. The resonating frequency is further improved in step four by introducing the horizontal slot of width W_4 and then finally the vertical slot of width W_2 is introduced so that our antenna can resonate at the desired frequency i.e. 28 GHz. The return loss of this whole process is presented in Fig. 4.

TABLE 1

OPTIMAL PARAMETERS OF THE DESIGNED ANTENNA

Parameters	Values (mm)	Parameters	Values (mm)
W_s	12	L_s	13
W_g	12	L_g	2.56
W_f	1.5	L_f	5
L_1	2.88	L_2	1.3
L_3	2.2	L_4	0.2
L_5	3.6	W_1	1.5
W_2	0.2	W_3	2.55
W_4	1.05	W_5	0.06
W_6	0.5	R	3



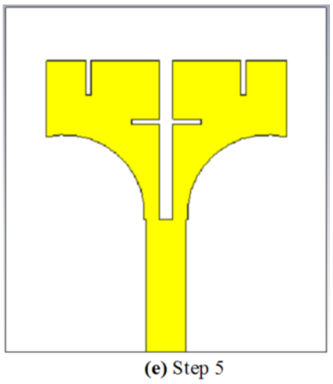


Fig. 4. Structure Evolution of the Proposed Antenna

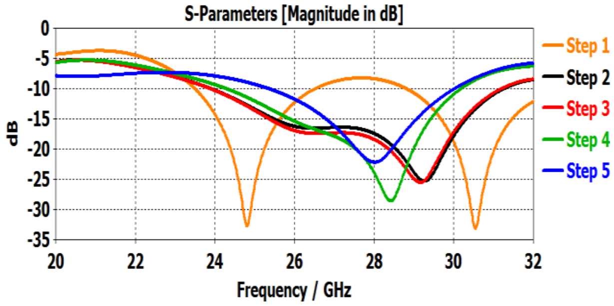


Fig. 5. The Return Loss of all Steps

III. OPTIMIZATION OF ANTENNA AND ANALYSIS OF RESULTS

In this section, the parametric study is performed to analyse the impact of design parameters on overall performance of the antenna. For this purpose, three parameters are selected which are the length of ground (L_g) vertical slot (L_2) and horizontal slot (W_4). These parameters contribute towards achieving the required resonant frequency, gain and impedance matching. The effect of changing each parameter is discussed below.

A. The Effect of Length of Ground (L_g)

By examining the variations in return loss (S_{11}) the influence of L_g is analysed. The L_g is varied from 2.56 mm to 8.4 mm. When L_g is set at 2.56 mm the proposed antenna showed promising results at desired 28 GHz frequency as shown in Fig. 6. As the value increases from 2.56 mm the antenna starts to resonate at the lower frequencies. So for the proposed design L_g = 2.56 mm is considered to be an optimum value for the length of a ground plane.

B. The Effect of Vertical Slot (L_2)

For an effective improvement in the overall performance of the microstrip patch antenna, the slots are embedded whichresults in the improvement in the bandwidth, reduction of antenna size, and improvement in radiation properties. Fig. 7 presents the variations in the return loss for choosing different values of the slot L_2. As the value increases from 1.3 mm the proposed antenna started to resonate at the lower frequencies.

C. The Effect of Horizontal Slot (W_4)

Fig. 8 illustrates the effect of horizontal slot W_4 for the different values on the return loss. As the value is decreased from 1.05 mm then the resonating frequency is shifted towards the higher frequency band with a return loss of around -42 dBi but the optimum value for W_4 is 1.05 mm.

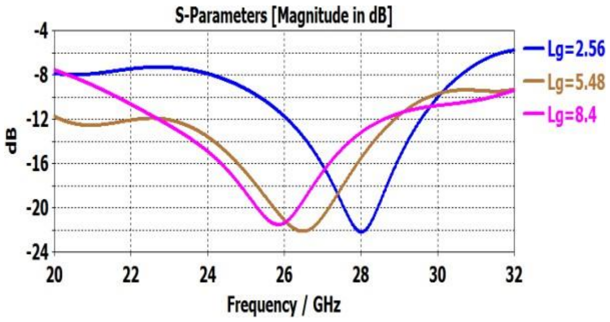


Fig. 6. Simulated Return Loss with Variations in terms of Length of ground

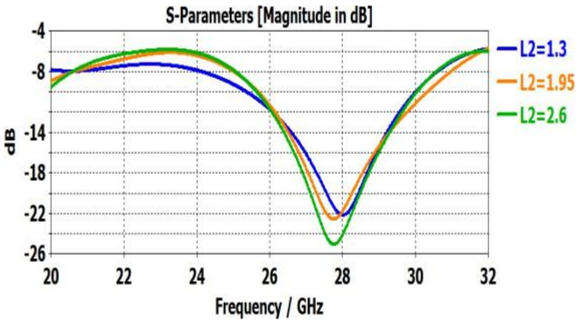


Fig. 7. Simulated Return Loss with Variations in terms of L_2

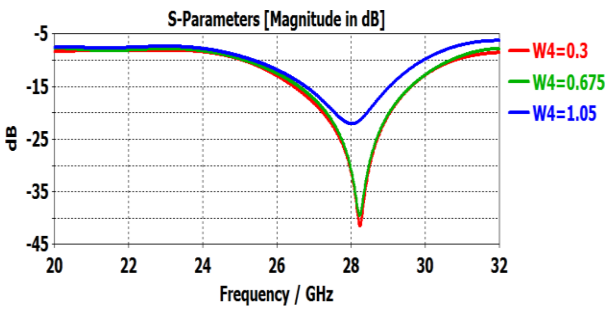


Fig. 8. Simulated Return Loss with Variations in terms of W_4

D. The Radiation Pattern and The Surface Current

The antenna pattern or radiation pattern is the graphical depiction of the antenna's radiation properties as a function of space. The 3D radiation pattern at 28 GHz is illustrated in Fig. 9. The directivity of the proposed antenna is 5.5 dB. Figs. 10 and 11 express the 2D radiation pattern of the proposed antenna. The surface current density is presented in Fig. 12. On the radiating element, a symmetrical distribution appears and is concentrated at the edges of the structure.

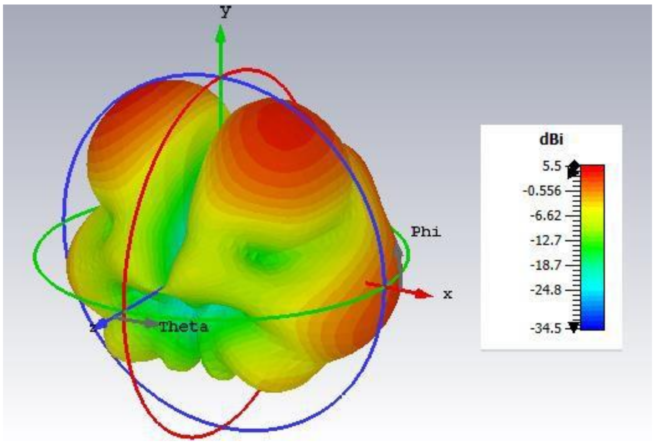


Fig. 9. Radiation Pattern (3D)

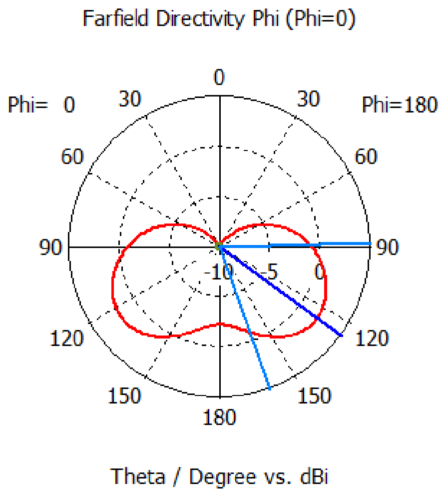


Fig. 10. Radiation Pattern E-Plane (2D)

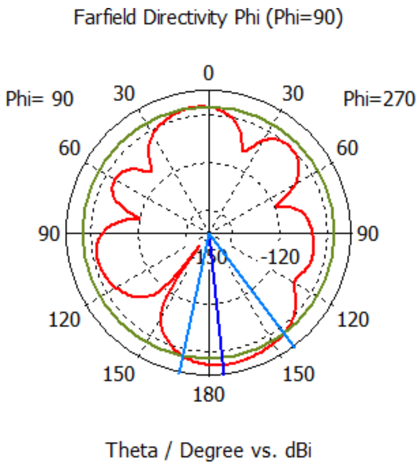


Fig. 11. Radiation Pattern H-Plane (2D)

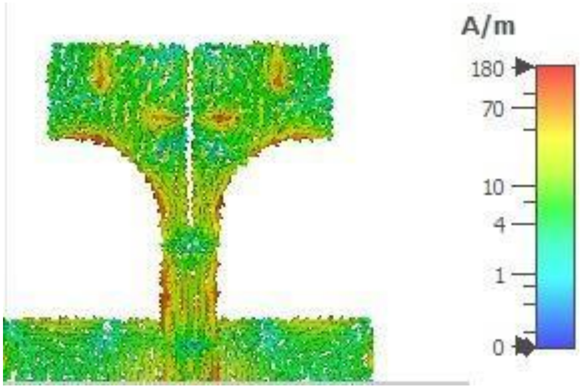


Fig. 12. Surface Current Density

E. The Voltage Standing Wave Ratio (VSWR)

VSWR expresses the amount of mismatch between an antenna and the feed line connecting to it. More power will be delivered to the antenna if the VSWR is small. The minimum VSWR is 1 and a value of VSWR under 2 is considered good for most applications. The VSWR graph of our proposed antenna is given in Fig. 13 and it is clear that the value is 1.16 at the operational frequency.

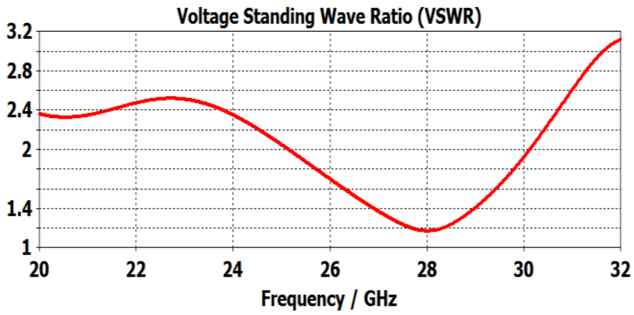


Fig. 13. The VSWR

F. The Radiation Efficiency and The Gain

The variation in efficiency and simulated gain of the proposed antenna is presented in Fig. 14 and Fig. 15. At the operating frequency, the designed antenna has a radiation efficiency of around 60 percent and a gain of 3.2 dBi.

Finally, in Table II a comparison between the reference antennas and the proposed antenna is presented to signify the performance of this design. Our antenna is smaller in size and has a higher bandwidth (25.36-30 GHz) as compared to the reference antennas.

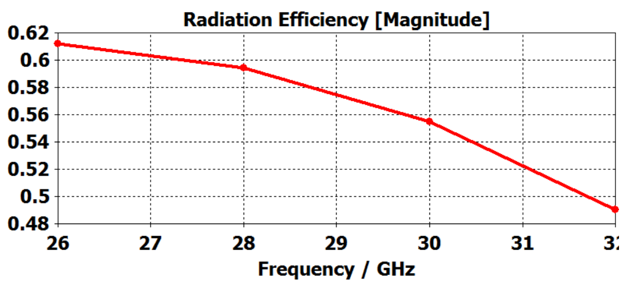


Fig. 14. The Radiation Efficiency

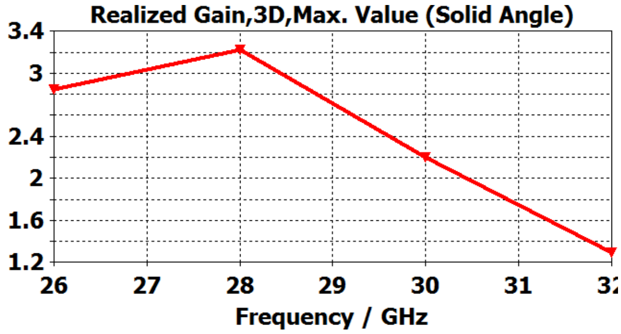


Fig. 15. The Realized Gain

IV. CONCLUSION

The design and simulation of a novel slotted microstrip patch antenna for 5G-IoT applications resonating at 28 GHz is presented in this study. The proposed antenna is compact in size with a return loss of -22 dB at 28 GHz. The antenna exhibits an efficiency of 60 %, a VSWR of 1.16, a gain of 3.2 with a workable bandwidth of 4.64 GHz making it good choice for 5G-IoT applications. In future the proposed antenna design will be fabricated to get measured results on the Rogers substrate and also single antenna element will be extended to a MIMO structure.

TABLE II
COMPARISON: PROPOSED ANTENNA VS REFERENCE ANTENNAS

Ref.	**Dimensions (mm^2)**	**Frequency of Operation (GHz)**	**Operational Bandwidth (GHz)**
[15]	8 × 8.489	28	1.43
[16]	15 × 8	28	0.280
[17]	55 × 110	28	1.06
[18]	19 × 19	28	1
[19]	10 × 10	28	2.94
[20]	12 × 14	28	2.55
Proposed	**12 × 13**	**28**	**4.64**

REFERENCES

[1] Chung M.A, Lee M.C, Hsu C.C, Lin C.W, "Multi-Band Coupled-Fed Antenna for 4G LTE, Sub-6G, and WLAN Frequency Bands in Various Electronic Devices", *IEEE Access* vol. 12, pp. 45398 – 45422, 2024.

[2] W. Hong and Chow-Yen Desmond Sim, "Microwave and Millimeterwave Antenna Design for 5G Smartphone Applications," *John Wiley Sons*, 2022

[3] A. N. Uwaechia and N. M. Mahyuddin, "A comprehensive survey on Millimeter Wave Communications for fifth-generation wireless networks: Feasibility and challenges," *IEEE Access*, vol. 8, pp. 6236762414, 2020.

[4] M. Koripi, "5G vision and 5G standardization," *SSRN Electronic Journal*, 2021.

[5] X. Wang, L. Kong, F. Kong, F. Qiu, M. Xia, S. Arnon, and G. Chen, "Millimeter Wave Communication: A comprehensive survey," *IEEE Communications Surveys and amp; Tutorials*, vol. 20, no. 3, pp. 16161653, 2018.

[6] R. Dangi, P. Lalwani, G. Choudhary, I. You, and G. Pau, "Study and Investigation on 5G Technology: A Systematic Review," *Sensors*, vol. 22, no. 1, p. 26, Dec. 2021.

[7] W. Hong, ”Solving the 5G mobile antenna puzzle: Assessing Future Directions for the 5G Mobile Antenna Paradigm shift,” *IEEE Microwave Magazine*, vol. 18, no. 7, pp. 86102, 2017.

[8] J. Martens, ”‘What Is My Measurement Equipment Actually Doing?: Implications for 5G, Millimeter-Wave, and Related Applications,” *IEEE Microwave Magazine*, vol. 23, no. 1, pp. 18-30, Jan. 2022.

[9] S. Y. Fatah, E. K. Hamad, W. Swelam, A. M. Allam, M. F. Abo Sree, and H. A. Mohamed, ”Design and implementation of UWB slot-loaded printed antenna for microwave and Millimeter Wave Applications,” *IEEE Access*, vol. 9, pp. 2955529564, 2021.

[10] Y. Luo, L. Zhu, Y. Liu, N. -W. Liu and S. Gong, ”Multiband Monopole Smartphone Antenna With Bandwidth Enhancement Under Radiation of Multiple Same-Order Modes,” *IEEE Transactions on Antennas and Propagation,*, vol. 70, no. 4, pp. 2580-2592, April 2022.

[11] A. Desai, T. Upadhyaya, and R. Patel, ”Compact wideband transparent antenna for 5G Communication Systems,” *Microwave and Optical Technology Letters*, vol. 61, no. 3, pp. 781786, 2018.

[12] K. Yekeh Yazdandoost and I. Laakso, ”EMF Exposure Analysis for a compact multi-band 5G antenna,” *Progress In Electromagnetics Research M*, vol. 68, pp. 193201, 2018.

[13] C. A. Balanis, ”Antenna theory analysis and design,” Hoboken, New Jersey Wiley, 2016.

[14] B. Cheng, Z. Du, and D. Huang, ”A differentially fed broadband multimode microstrip antenna,” *IEEE Antennas and Wireless Propagation Letters*, vol. 19, no. 5, pp. 771775, 2020.

[15] AbdulGuddoos S. A. Gaid, et al. “Design and Analysis of a Low Profile, High Gain Rectangular Microstrip Patch Antenna for 28 GHz Applications.” *Cogent Engineering*, vol. 11, no. 1, 3 Mar. 2024. 2024.

[16] Li, Yitong. “A Microstrip Patch Antenna for 5G Mobile Communications.” Journal of Physics. Conference Series, vol. 2580, no. 1, 1 Sept. 2023, pp. 012063–012063, 2024.

[17] H. M. Marzouk, M. I. Ahmed, and A.-E. H. Shaalan, ”Novel dual-band 28/38 GHz MIMO antennas for 5G Mobile Applications,” *Progress In Electromagnetics Research C*, vol. 93, pp. 103117, 2019.

[18] Y. Jandi, F. Gharnati, and A. Oulad Said, ”Design of a compact dual bands patch antenna for 5G applications,” 2017 *International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)*, 2017.

[19] J. Saini and S. K. Agarwal, ”T and L slotted patch antenna for future mobile and wireless communication,” *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2017.

[20] M. N. Hasan, S. Bashir, and S. Chu, ”Dual band Omnidirectional Millimeter Wave Antenna for 5G communications,” *Journal of Electromagnetic Waves and Applications*, vol. 33, no. 12, pp. 15811590, 2019.

Blind Vision: A Computer Vision Based Assistive Technology for the Visually Impaired

Haseeb Tahir
Electrical Communication Engineering
University of Engineering and Technology, Peshawar
Pakistan
hasebmalik2@gmail.com

Salman Aziz
Electrical Communication Engineering
University of Engineering and Technology, Peshawar
Pakistan
m.salmanaziz12@gmail.com

Jan sher Khan
Electrical Communication Engineering
University of Engineering and Technology, Peshawar
Pakistan
jansher.koh@gmail.com

***Abstract*—Developing tools for visually impaired individuals presents significant challenges, and computer assisted solutions are still in their infancy. These technologies strive to facilitate independent navigation without external assistance. Existing systems primarily utilize computer vision but often target specific tasks, failing to meet all essential needs comprehensively. This research introduces a practical approach and framework that amalgamates various technologies—object identification, speech recognition, text-to-speech, OCR, and speech-to text—within the Blind Vision project. Our proposed approach is building an affordable wearable gadget that includes the concept of computer vision to interpret any data surrounding the user, regardless of alignment or lighting circumstances. The device uses a suitable camera and Raspberry Pi to capture ambient content and vocalize it in a regional language for blind people. In addition, a sensor counts the things in its field of view and tells the user how far they are from the closest object at eye level. The architecture combines methods for artificial speech, neural networks, and image manipulation. It is noted that the combined accuracy of the object detection and character recognition using optical characters methods is a score of 8. This innovative approach addresses several critical issues faced by visually impaired individuals. Furthermore, we conduct a comparative analysis of existing navigation methods, highlighting areas for further research and enhancement.**

Keywords—Visually Impaired people, Object Recognition, YOLOv8, Text to Speech, Raspberry pi, Optical character Recognition (OCR).

I. INTRODUCTION

While having good vision is essential for carrying out daily chores, many people find it difficult to do so because of visual impairments. Like all others, people with visual impairments have a wide range of pursuits, yet a sizable fraction of them are unemployed because of inadequate learning resources. There aren't enough blind-specific educational institutions and groups in place right now to educate the sizable population that needs them. Braille is a partially translated form of English speech that was created to help people with impaired vision read. However, evaluations performed in blind campuses reveal numerous pupils find it challenging to master Braille, and there are inadequate learning materials available for such a large cohort. Additionally, many visually impaired individuals rely on others for understanding their surroundings and navigating while eliminating obstructions. In order to overcome these obstacles and improve environmental awareness and self-learning, a solution that uses cutting-edge technology for object detection and obtaining text is desperately needed. Abhilash Parthasarathy et al. [1] developed an integrated system where a camera captures printed text documents for digitization, which are then processed using an OCR module. However, this technology can only handle photos with a high quality of documents. Abhilash Mishra et al. [2] observed that earlier research has brought to light a number of important findings and suggested an improved removal of features and TTS (text to-speech) approach for people with visual impairments that makes use of OCR (optical character recognition) technology. This technology turns recognized text into audible voice by comparing it to a template. Nevertheless, this method is limited to brightly lit spaces and does not combine text extraction with object detection in the scene. Yingli et al. [3] created a camera-based adaptive text comprehension structure designed to assist blind individuals read written labels and packaging of goods from portable items. The main objective of this system is to acquire text from things. Hermes et al. [4] suggested TactoBook, a computer program that creates Braille versions of eBooks. This approach assumes users are fluent in Braille and is restricted to eBooks. Nagarjun et al. [5] demonstrated a Raspberry Pi-based vision-based information recognition device that speaks words extracted from photos. However, this system does not include object recognition. Yudong Luo et al. [6] proposed "Finger-eye," a refreshable, compact text reading device that translates text into Braille or audio, mimicking the natural reading experience of Braille displays or books. However, this system requires users to know Braille. Hemalatha et al. [7] presented an assistance gadget for the deaf, silent, and blind that is built on a Raspberry Pi. This device converts images to speech using Tesseract software, but it struggles with hazy pictures and inadequate lighting. Vignesh et al. [8] used an optical character reader and a concealed Markov model to create a self-learning read and write kit for the blind; however, this kit is devoid of object recognition features. Shruti B. Hatti et al. [9] suggested B-LIGHT, an auditory tool for the blind that creates machine-encoded text from images of printed or typed text using OpenCV and OCR. The only things this system can process are texts and subtitles, introduced an Intelligent Reading

Device for the Blind individuals that employs a Raspberry Pi. This system aims to improve accessibility by integrating advanced technologies to support text and object recognition in various lighting conditions. Nikhil Mishra et al. [10] suggested a technique for employing a Raspberry Pi, which is intended to read tags and data on posters of different dimensions, to convert picture text to speech. However, this method necessitates a rather large font size for the text, making it unsuitable for recognizing smaller text. J. Regina et al. [11] created a voice-assisted text recognition device that reads text aloud by using an OCR analyzed finger-mounted lens to gather data from printed paperwork. Because of the limitations of the camera, this technology is only applicable to printed materials and books. In summary, while many products exist to assist each person with visual impairments have restrictions, such as focusing solely on text recognition, being restricted to English, or lacking comprehensive functionality. No single product currently addresses all required objectives. Richard et al. [12] designed a system using haptic technology for multicolor reading, tactile visuals, and visually impaired accessibility and positioning. Text on volumes or boards is not supported by this system. In order to aid with movement and avoid collisions, the suggested system includes a device that is worn that interprets text in front of the individual, recognizes and counts a variety of nearby objects, and calculates the separation between the nearest item at your level of vision. The product costs 68 USD as a one-time development expense and requires minimal maintenance, making it suitable for long-term use, even in developing countries. This innovative system stands out due to its multifunctionality and effectiveness.

II. Proposed Methodology

A key benefit is its accommodation for visually impaired individuals who may not be fluent in English. The device incorporates language translation capabilities, allowing the result will be readable in several local dialects not just English.

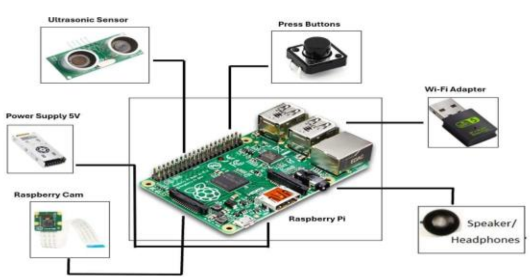


Fig. 1. Component Diagram

The hardware setup, shown in Fig 1, includes the following components:

The Raspberry Pi Lens Unit the (second version) Without Ultraviolet Filtering (NoIR), is a high-end 8MP camera with a Sony IMX219 image sensor that can capture 3280 x 2464 pixels in still photos and 640 x 480 pixels in 60/90-second videos. It has a constant focusing lens and can be used for CCTV security, ultraviolet the use of photography, darkroom photography, and crop development monitoring.

Raspberry Pi: 40-pin expanded GPIO, four separate connections USB 2.0, combined movie segment and stereotypical outcome, bid High-Definition Multimedia Interface, Camera Serial Interface to interface lens, connection for LCD display, memory card slot for operating system also for saving the information, a mini portable connection of power. Quad Core Processing unit running at 900megahertz with a single GB RAM (Broadcom BCM2837 Arm7).

Supersonic distance sensor SEN-15569: includes input potential of five volts, an ac- curacy of 3 mm, a range of measurement of 20 mm to 4500 mm in principle and 20mm to 800mm in fact, calculating degree of less than 14, a working charge of less than 15 mA, and a running frequency of 40 Hz. Supersonic distance sensor is a versatile instrument.

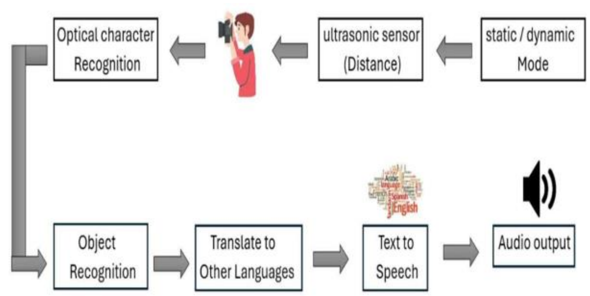


Fig. 2. Block Diagram of the proposed system

Fig. 2 shows the component schematic of the suggested system. This system uses a sensor with ultrasound capabilities to determine how close an object is, then notifies the user of that distance.

The detector communicates using ultrasonic energy, and it measures the distance by measuring the interval between the wave's emission and receipt. The sensor measures by determining how far it traveled by measuring the duration of the interval between sending and getting the audio wave. Using the sound's velocity—roughly 762 miles per hour in the air.

Raspberry Pi camera module captures an image of the NoIR camera's environment to the person using it. This image is then processed by the Optical Character Recognition (OCR) and classification of items. OCR extracts text from the captured image and saves it as a text file, utilizing the Google Cloud Vision API. With the help of robust machine learning models, this REST API provides a user-friendly interface for PUT, GET, POST, and DELETE data queries via HTTP requests. Once the camera module captures the image, the Google Cloud Vision API receives it, and the recognized text is saved in a .txt file. The system automatically sets up the necessary Google Cloud Library credentials for the user.

The aim of object recognition is to identify every element in a picture. It is implemented using Google Cloud Vision API. Every data point in the three-dimensional space created by the process correlates to a single image from a certain cluster. Related images like "cat" and "dog" are closer together than dissimilar ones like “car” or “traffic signal”. S. Ottaviano et

al. [13] used t-Distributed Stochastic Neighbor Embedding (t- SNE) reduction of dimension approach to convert every graphic into a three-dimension vector to display this complex database.

YOLOv8, the newest member of the YOLO-based real-time object identification model family, was developed by Ultralytics and offers state-of-the-art features. YOLO's single-pass processing capacity allows it to quickly and accurately recognize and classify a wide range of objects in a scene. In addition to being speedier and more accurate than previous YOLO versions, the YOLOv8 model offers a well-thought-out architecture for training models that can object detection, object tracking, instance seg- mentation and image classification.

The mentioned integrated approach ensures the system is both multifunctional and effective in assisting visually impaired users with reading text and navigating their surroundings.

The formula determines probabilities pij of items xi and xj in the following way, assuming that the acquired image contains N number of large objects, i.e., x1 to xN.

$$P_{j/i} = \frac{\exp exp\left(-|x_i - x_j|^2 / 2\sigma_i^2\right)}{\Sigma_{k \neq i} \exp exp\left(-|x_i - x_k|^2 / 2\sigma_i^2\right)} \qquad (1)$$

With yi ∊ Rw, t-SNE develops a fresh map of decreased characteristics y1…..yM from these probabilities, reflecting the parallels as precisely as possible. Here, “w” is 3 when a three-dimensional vector is created. It determines the correlation (qij) between both locations (yi and yj) in the newly generated map using a similar method as follows:

$$q_{ij} = \frac{\left(1 + |y_i - y_j|^2\right)^{-1}}{\Sigma\ k \neq l\left(1 + |y_k - y_l|^2\right)^{-1}} \qquad (2)$$

After the reduction is complete, Liu et al. [14] employ a straightforward K-means technique that takes advantage of the Euclidean distance to detect objects. A specific color is used to link each cluster. Additionally, each cluster's name can be easily found by selecting the most prevalent label among its points. The photo is saved in the local folder after it has been captured by the camera. Once the saved image has been retrieved, the address is then sent to the algorithm for object recognition. It indicates which file name should contain the object list. A .txt file having the list of identified items is the output.

The user must first set up his login credentials and create a Google project account to access this Google Cloud Library function. The product does this automatically be- fore you use the feature. A normal Google REST API can be used to access the Google Language Translation API. It does multilingual text translation. Programmatic references are made to the source and destination text files. The output file that was specified contains the translated text.

III. EXPERIMENTAL RESULTS

The experimental configuration for the suggested system equipped by a person is seen in Fig. 2 below. The device known as the Raspberry Pi, No Infrared filter lens, battery, Press buttons, resistors, jumper cables, a board for experimenting and ultrasonic device are all intended to be accommodated within. It is a small, light, and easy-to-use device that may be used on the head. An overview of the actions the product took to achieve the project's goals is shown below:

1. Firstly, the camera module captures the photo of the environs of an individual.

2. To prevent accidents with any obstructions, the ultrasound detector determines how close it is to the nearby object in the user's field of vision and communicates that information to the user via vocal output.

3. The picture undergoes additional processing after the OCR and object detection algorithms are called. This stage's result is stored in a text document.

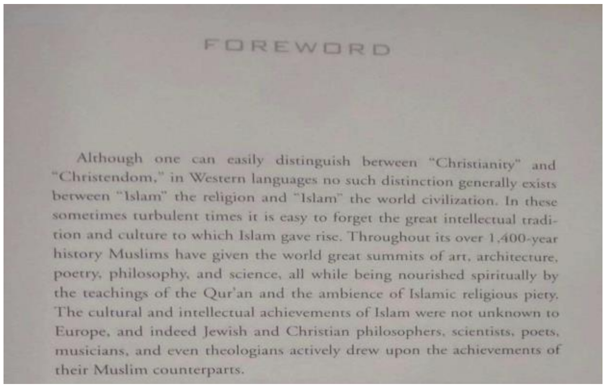
4. Using the local language script, the result taken from last stage then converted to a local tongue of the region to save in the form of .txt document.

5. Following the conversion of document through Text-to-Speech segment, the visually handicapped can hear the final result through headphones or earbuds.

Fig. 2. Setup Proposed Experimentally

A variety of images with diverse font styles, dimensions, shades, foreground colors, angles, and orientations were used to evaluate the OCR and object identification unit. Below are some examples of both object recognition and OCR raw photographs along with the related generated pictures.

Figure 3 (a) displays the book image that was obtained and sent into the OCR,



FOREWORD

Although one can easily distinguish between "Christianity" and "Christendom," in Western languages no such distinction generally exists between "Islam" the religion and "Islam" the world civilization. In these sometimes turbulent times it is easy to forget the great intellectual tradition and culture to which Islam gave rise. Throughout its over 1,400-year history Muslims have given the world great summits of art, architecture, poetry, philosophy, and science, all while being nourished spiritually by the teachings of the Qur'an and the ambience of Islamic religious piety. The cultural and intellectual achievements of Islam were not unknown to Europe, and indeed Jewish and Christian philosophers, scientists, poets, musicians, and even theologians actively drew upon the achievements of their Muslim counterparts.

Fig. 3. (a) Input Image

```
FOREWORD
 Although one can easily distinguish between "Christianity"
and “Christendom,” in Western languages no such distinction
generally exists between “Islam” the religion and “Islam” the
world civilization. In these sometimes turbulent times it is
easy to forget the great intellectual tradition and culture
to which Islam gave rise. Throughout its over 1,400-year
history Muslims have given the world great summits of art,
architecture, poetry, philosophy, and science, all while
being nourished spiritually by the teachings of the Quran and
the ambience of Islamic religious piety. The cultural and
intellectual achievements of Islam were not unknown to Europe
, and indeed Jewish and Christian philosophers, scientists,
poets, musicians, and even theologians actively drew upon the
achievements of their Muslim counterparts.
```

Fig. 3. (b). Output of the image containing text

Whereas Fig 3 (b) displays the text output that was produced as an output of the detected text. A sample picture provided as input to the object identification module is displayed in Fig. 4(a), and the text output produced with the objects identified is shown in Fig. 4(b).

Fig. 4. (a)

```
The objects around you are
car
street
city
urban area
mode of transport
transport
town
metropolitan area
vehicle
downtown
```

Fig. 4. (b)

The Confusion matrix makes it simple to calculate the performance parameters for the test cases. The metrics listed below have been established using this organized matrix:

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$\text{Accuracy} = \frac{TP + TN}{FN + FP + TP + TN} \tag{4}$$

$$Specificity = \frac{TN}{FP + TN} \tag{5}$$

$$Senstivity = \frac{TP}{FN + TP} \tag{6}$$

Which means; FN is False Negative, FP is False Positive, TP is True Positive and TN is True Negative.

400 photos that were taken in different spots and under varied situations were analyzed for object identification and OCR. The results of measurements of all parameters are displayed below.

Accuracy	Sensitivity	Specificity
83%	87%	81%

Table 1. OCR Performance

Accuracy	Sensitivity	Specificity
93.7%	98%	96%

Table 2. Object Recognition Performance

Method Used	Precision	Recall
Yingli et al. [3]	59.07	68.33
Hemalatha et al. [7]	53.56	69.94
Method proposed	66.40	86.13

Table 3. Related Study

IV. **CONCLUSION**

In this article, we first evaluated and compared a few fully functional current systems. It has been discovered that obstacle identification for blind navigation greatly benefits from image processing. Through theoretical analysis and experimental research, we discovered that some aiding systems are useful for specific tasks. This poll has shown us that there isn't a mechanism in place to aid blind people with all their fundamental needs. We also spoke with a few blind people, and they expressed dissatisfaction with the current arrangements. They still have certain issues; thus, they require a more advanced system with a variety of features. as each of the current systems has a specific purpose. Thus, a new framework that integrates navigation with other fundamental assistive technology for the blind is suggested. The efforts to help blind people while resolving the issues are still ongoing. Blind people may find that this new system helps to resolve some of their main issues

REFERENCES

[1] S.Parthasarathy, M.S.Sonam, K.R.Arun et al. "A smart reader for visually impaired people using raspberry pi", International Journal of Engineering Science and Computing, 2016, vol, issue 3, pp. 1997-2001.

[2] Abhilash Mishra et al. "Improved feature extraction and TTS method based blind OCR", International Journal of Advance Engineering and Research Development, 2018, vol. 5, is- sue 4, pp. 1858-1864.

[3] Yingli Tian, Aries Arditi et al. "Portable Camera-Based Assistive Text and Product Label Reading From Hand-Held Objects for Blind fromons", IEEE/ASME Transactions on Mech- atronics, June 2014, vol. 19, issue 3, pp. 808-817.

[4] Hermes Hernández et al. "A portable ebook reader for the blind", Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE,31 Aug.- 4 Sept. 2010, pp. 2107-2110.

[5] Nagarjun R S, Nishanth M, NIthin D, Veena S Murthy, "Vision based text recognition using raspberry pi", International Journal of Computer Applications, National Conference on Power Systems & Industrial Automation, 2015, pp. 1-3.

[6] Yudong Luo, Jose Cordero, "Finger-eye: A wearable text reading assistive system for the blind and visually impaired", IEEE International Conference on Real-time Computing and Robotics, 6-10 June 2016, pp. 125-128.

[7] R. Hemalatha, Surendhar K, Suresh N R et al. "Raspberry Pi based assistive device for deaf, dumb and blind people", International Journal of Scientific Research Engineering & Tech- nology, May 2017, vol. 2, issue 2, pp. 167-174.

[8] Vignesh Babu T.N, Sri Hari Kavin N, "Automated Read-Write kit for Blind using Hidden Markov Model and Optical Character Recognition", International Conference on Applied and Theoretical Computing and Communication Technology, 29-31 Oct. 2015, pp. 828-831.

[9] Shruti B. Hatti, Shruti S. Salimath, Vijayalaxmi I. Byakod, Shivaleela Kanede, "B-Light: A reading aid for the blind people using OCR and OpenCV", International Journal of Scientific Research Engineering & Technology, May 2017, vol. 6, issue 5, pp. 546-548.

[10] Nikhil Mishra, "Image Text to Speech Conversion using Raspberry Pi & OCR Techniques", International Journal for Scientific Research and Development, vol. 5, issue 08, 2017, pp. 523-525.

[11] J.Rejina Parvin et al. "Voice assisted text reading system for visually impaired persons using TTS method", IOSR Journal of VLSI and Signal Processing, May-June 2016, vol. 6, issue 3, ver. III, pp. 15-23.

[12] Richard J. Adams, Ryo Kitada, "Designing haptic assistive technology for individuals who are blind or visually impaired", IEEE Transactions on Haptics, July-Sept. 2015, vol. 8, issue 3, pp. 258-278.

[13] L. Ottaviano, and S. Occhinegro," Optical Character Recognition - A survey", International Journal of Pattern Recognition and Artificial Intelligence, June 1991, vol. 5, issue 1.

[14] Liu Xumin, Guan yong "Research on k - means Clustering Algorithm" Third International Symposium on Intelligent Information Technology and Security Informatics, IEEE 2010, pp. 63-67.

Design of Multiband MIMO Antenna for 5G Applications

Mehwish Mehreen
Telecommunication Engineering
University of Engineering and Technology, Taxila.
Pakistan
mehwishmehreen43@gmail.com

Muhammad Abdul Basit
Telecommunication Engineering
University of Engineering and Technology, Taxila.
Pakistan
abdul.basit@uettaxila.edu.pk

Yasar Amin
Telecommunication Engineering
University of Engineering and Technology, Taxila.
Pakistan
yasar.amin@uettaxila.edu.pk

Abstract-- **This study presents a novel antenna design that integrates a bow-tie slot antenna with four Linear Tapered Slot Antennas (LTSAs) on a single aperture, catering to multiple frequency bands for 5G applications. The antenna's design features a coplanar waveguide transmission line feeding the bow-tie slot antenna, while a MIMO configuration is implemented for the LTSAs in the mm-wave band. The proposed antenna exhibits a bi-directional radiation pattern in the microwave range, achieving a maximum gain of 6.70 dB, and an end-fire radiation pattern for each LTSA in the mm-wave band. Notably, the design successfully covers several microwave bands and the n258 band in the mm-wave range, encompassing both lower and high bands of 5G. With its compact and efficient design, this antenna offers a promising solution for multi-band applications, rendering it suitable for various wireless communication systems. The proposed design's versatility and performance make it an attractive option for future 5G implementations.**

Keyword--Linear Tapered Slot Antennas; Bow-Tie Slot Antennas; Coplanar waveguide; Bi-directional; MIMO Antenna.

I. Introduction

The increasing demand for high-speed, low-latency, and broadband communication systems in technologies such as 5G, IoT, and next-generation radar has highlighted the importance of bow-tie slot antennas. Their planar structure and compatibility with printed circuit boards (PCBs) make them highly suitable for compact and integrated systems. Additionally, they are less prone to fabrication errors and can be designed using various materials, including flexible substrates for wearable and biomedical applications. In military and defense applications, bow-tie slot antennas are widely used in ground-penetrating radar (GPR) and stealth communication, benefiting from their ultra-wideband (UWB) characteristics.

Various bow-tie antenna designs have been explored in the literature. Compared to conventional microstrip antennas, the designs presented in [1],[2] and [3] offers a wider frequency spectrum and exhibit either uni-directional or bi-directional radiation patterns. To further enhance bandwidth, two small metal conductors are incorporated within the bow-tie slot antenna, as demonstrated in [4]. The impact of rounding the corners on antenna performance is examined in [5], where it is observed that rounded corners improve return loss. For dual- or triple-band operation, tuning stubs have been introduced in [6],[7], while [8],[9] explore methods to improve spectral efficiency.

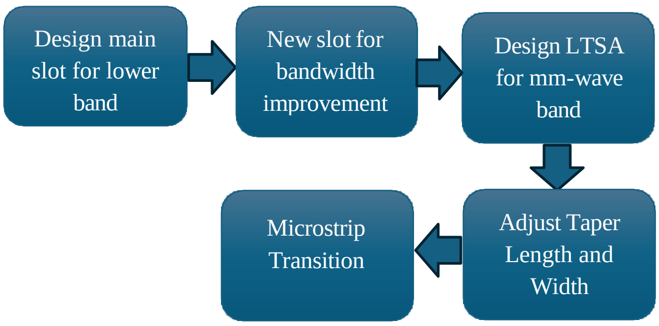
To extend the range and capacity of wireless communication, MIMO antennas play a crucial role. By receiving synchronous inputs and transmitting multiple data streams, MIMO technology enhances bandwidth, spectral efficiency, and signal reliability. The integration of sub-6 GHz and millimeter-wave (mm-w) technologies has become a key trend in the development of 5G communication systems. To support the miniaturization of communication terminals and base stations, there is a growing demand for compact dual-band antennas that operate across both sub-6 GHz and mm-w frequencies [10], [11]. Moreover, to address the increasing need for high data rates and efficient spectrum utilization, MIMO technology has been integrated into mm-w antenna systems. Traditionally, dual-band functionality is achieved by employing separate antennas for different frequencies, simplifying MIMO implementation. However, such MIMO antenna configurations tend to occupy significant space. To overcome this challenge, researchers have explored structure reusing techniques for realizing dual-band antennas [12].

To leverage the advantages of both the sub-6 GHz microwave band and the mm-w band of 5G, our proposed design integrates a bow-tie slot antenna for the microwave range and four longitudinally tapered slot antennas (LTSAs) for the mm-w range. The LTSAs are positioned along the edges of the substrate, radiating in an end-fire pattern to form a four-element MIMO array. Since the bow-tie slot antenna exhibits bi-directional radiation, covering both the upper and lower hemispheres, the side space is effectively utilized for embedding the LTSAs. This approach enables coverage of both the sub-6 GHz and mm-w bands, enhancing the system's performance for 5G applications.

The rest of this paper is structured as follows: Section II describes the structure of the proposed antenna, while Section III presents a parametric study. The impact of supporting surfaces beneath the antenna is analyzed in Section IV. Section V covers fabrication and radiation field measurements, and finally, the conclusion is presented in Section VI.

II. METHODOLOGY

Bow-tie slot antenna is fabricated on a substrate Rogers RO4003C with a thickness of 0.2mm (εr = 3.38 and tan δ = 0.0027). The dimension of the proposed antenna is 120 mm × 60 mm. A bow-tie slot antenna is designed from two right-angle triangles sots and their corners are rounded as shown in Fig 1. These right-angle triangles are fed by coplanar waveguide transmission line connected through SMA (Sub Miniature Version A) connector. There is another slot in between the Right-angle triangle arms. This additional slot helps the antenna to resonate on another frequency and hence a band around 5.6 GHz is also covered the arm of the right-angle triangle slot antenna which is perpendicular to the feeding line is L = 20.8 mm. The width of all the slots is same which is denoted by W and is 5 mm. The antenna is fed by coplanar waveguide which is connected through SMA conductor of characteristic impedance 50 Ω. The quarter wavelength transformer is constructed between antenna and CPW fed line. The length of longer side of the quarter wave transformer is 29.5 mm and shorter side is 19.1 mm. Width of the quarter wave transformer is 2 mm. The spacing between both quarter wave transformer slots is 4 mm. Fig. 1 shows the model of the bow- tie slot antenna which is fed by CPW fed-line. Total length(U) of LTSA slot antenna is 17 mm and width (d) are 3mm. The LTSA fed by the microstrip feed line contains a circular transition both in the microstrip as well as in the slotted LTSA. The geometry of single element LTSA can be seen in Fig. 2. The bow-tie antenna is excited by CPW feed line and LTSA antennas are excited through microstrip feed line. Table 1 shows all the values used for both antennas. Following is the flow diagram for the design of Bow-Tie slot antenna with 4 LTSA antennas.



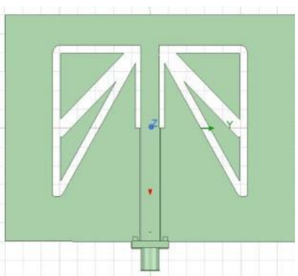


Fig. 1. Bow-Tie Slot Antenna with Co-Planar Waveguide Fed Line

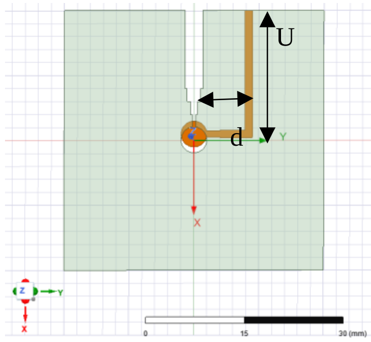


Fig. 2.Single Element of Linear Tapered Slot Antenna

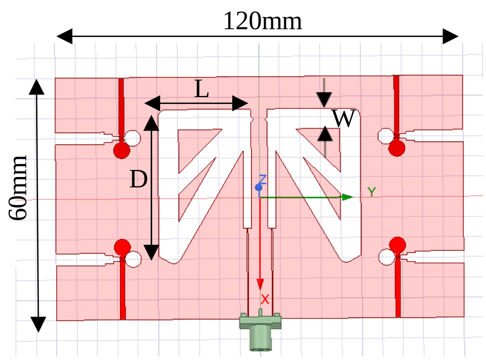


Fig.3.Final Model of Bow-Tie Antenna with 4 Linear Tapered Slot Antennas

Table 1. Parametric List of Bow-Tie Slot Antenna with 4 LTSA Antennas

Variables	Value(mm)
L	20.8
W	5
D	34.37
U	17
d	3
r	2

Table 2. Comparison Between Antennas Presented in Introduction

Ref.	Size	Max Gain	Bandwidth	MIMO Capability
1.	52mm × 46mm	6.5 dB	15.84 GHz	No
2.	20.5mm × 25mm	4.39 dB	2.85 GHz	No
4.	30mm × 26mm	--	12.9 GHz	No
5.	70mm × 80mm	8.5 dB	12 GHz	No
This work	60mm × 120mm	6.7 dB	8.5 GHz	Yes

III. RESULTS AND DISCUSSION

The combined model of proposed antenna containing bow-tie antenna and 4 LTSAs is shown in Fig. 3. Fig. 4 shows the S-parameter of the proposed antenna. It can be seen from the result that the bowtie antenna resonates at 2.45 GHz as well as covers the band from 3.4GHz to 5.8 GHz. The resonance points are affected by changing in the arm length of bow-tie antenna. If length of the arm increases the resonance shifts towards the lower frequency. The main radiator in the proposed antenna is the horizontal and diagonal arms. The first and third resonances are affected by the length of the horizontal and diagonal arms of bow-tie antenna. Quarter wavelength transformer deployed middle resonance which is installed between the bow-tie antenna and CPW fed line. The slot bow-tie antenna also depends upon the W of the antenna. If the width of the antenna increases then resonance shifts toward the higher frequency and vice versa. The result shows that antenna covers several standard bands in the microwave range.

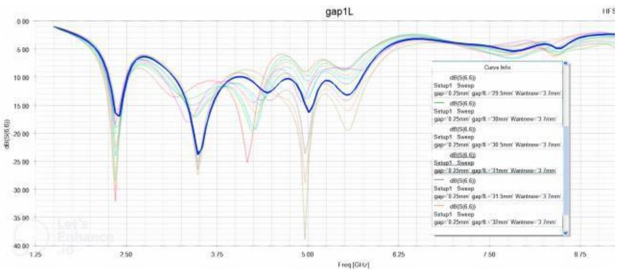


Fig. 4. S-Parameter of Bow-tie slot antenna

Fig 5 shows the s-parameters of the same structure when operated at the mm wave band. The bowtie antenna connected at port 1 will not be excited as it operates on the microwave bands. The four LTSA are excited through individual ports. Figure 5 shows the simulated s-parameter of port 2. It is seen that the LTSA covers the frequency band between 25 to 27.5 GHz which is designated as n258 band. This band is allocated for high-capacity, short-range communications, often used in dense urban environments and fixed wireless access (FWA) applications.

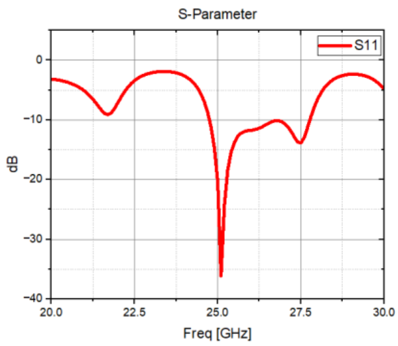


Fig. 5. Improved S-parameter at milli-meter wave band

Fig. 6 shows the gain plot of the system in theta is 90 degrees and phi is 0 degree. End-fire radiation pattern of single LTSA antenna system in YZ plane is shown. The radiation pattern equally radiates in both upper and lower hemi-sphere. When the two right LTSA antennas are excited, end-fire radiation pattern occurs in right side. Similarly, when two left LTSA antennas are excited, end-fire pattern appears in the opposite directions.

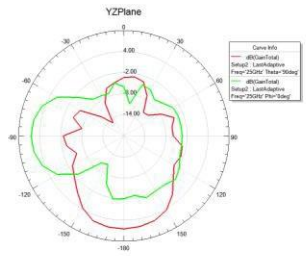


Fig. 6. Radiation Pattern at 25 GHz

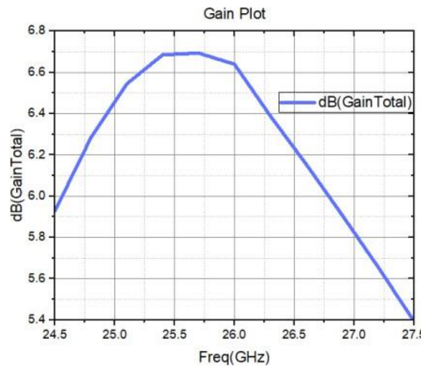
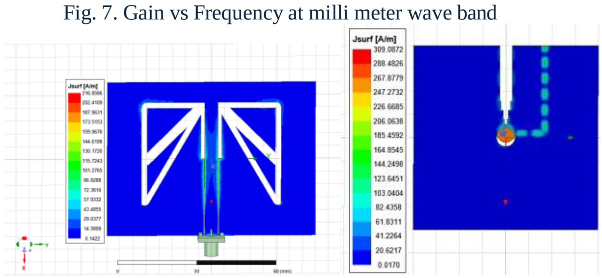


Fig. 7. Gain vs Frequency at milli meter wave band



(a) (b)

Fig.8. (a) Current Distribution of bow-tie slot Antenna (b) Current Distribution of Linear Tapered Slot Antenna

Fig. 7 shows the gain plot at millimeter wave band. The maximum gain of 6.7 dB at 26 GHz is achieved. In Fig. 8, current distribution of bow-tie slot antenna and Single MIMO antenna at 25GHz is shown. It is clear from both figures that antennas are properly excited. In Fig. 9, the isolation between MIMO ports is shown. It can be seen that the isolation is better than -22dB throughout the frequency range. Fig. 10 shows the Envelope Correlation Coefficient (ECC) of MIMO array. The graph shows that ECC value is better than 0.01. Thus, the antenna is a good candidate for MIMO applications.

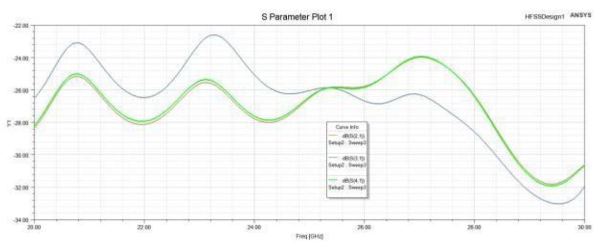


Fig. 9. Isolation between MIMO ports

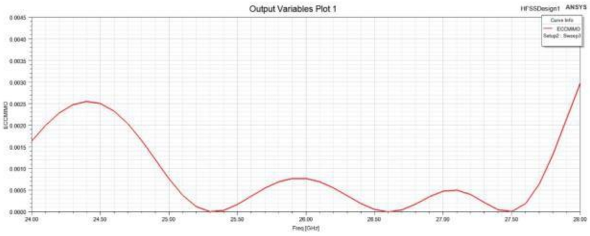


Fig. 10. Envelope Correlation Coefficient of MIMO Array

IV. CONCLUSION

The proposed antenna with 4 LTSA antennas and bow-tie slot antenna is presented. The antenna operates well in the microwave and millimeter wave range covering several standard frequency bands. The antenna radiates bi-directionally in the microwave range, and in the end-fire direction at mm wave range where each LTSA covers each quadrant in the azimuth plane. The maximum gain of 6.70 dB is achieved in the microwave range. The LTSA antennas are arranged in a MIMO configuration which makes the system more reliable and also enhances the spectral efficiency. The bands in sub-6Ghz range include n79, n96 bands, Wi-Fi band in the microwave range, and n258 band in the mm wave range.

REFERENCES

[1] J. Yeo, Y. Lee and R. Mittra, “Wideband slot antennas for wireless communications,” IEEE Proc. Microw. Antennas Propag., vol.151, no.4, pp.351-355, Aug. 2004.

[2] E. A. Soliman, et al, “Bow-tie slot antenna fed by CPW,” Electron Lett., vol.35, no.7, pp.514-515, Apr. 1999.

[3] A. A. Eldek, A. Z. Elsherbeni, and C. E. Smith, “Characteristics of bow-tie slot antenna with tapered tuning stubs for wideband operation,” Progress In Electromagnetics Research, vol.49, pp.53-69, 2004.

[4] Y.-L. Chen, C.-L. Ruan, and L. Peng, “A novel ultra-wideband bow- tie slot antenna in wireless communication systems,” Progress In Electromagnetics Research, vol.1, pp.101-108, 2008.

[5] S. W. Qu, and C. L. Ruan, “Effect of round corners on bowtie antennas,” Progress In Electromagnetics Research, vol.57, pp.179-195, 2006.

[6] L.-C. Tsai, “A dual bow-tie-shaped CPW-fed slot antenna for WLAN applications,” Progress In Electromagnetics Research, C, vol.47, pp. 167-171, Apr. 2014.

[7] J. H. Yoon, and Y. C. Lee,” Modified bow-tie slot antenna for the 2.4/5.2/5.8 GHz WLAN bands with a rectangular tuning stub,” Microw. Opt. Technol. Lett., vol.53, no.1, pp.126-130, Jan.

[8] S. Soltani, P. Lotfi, and R. D. Murch, “A port and frequency reconfigurable MIMO slot antenna for WLAN applications,” IEEE Trans. Antennas Propag., vol. 64, no. 4, pp. 1209–1217, Apr. 2016.

[9] K. Ding, C. Gao, D. Qu, and Q. Yin, “Compact broadband MIMO antenna with parasitic strip,” IEEE Antennas Wireless Propag. Lett., vol. 16, pp. 2349–2353, 2017.

[10] N. Hussain, A. Abbas, S. M. Park, S. G. Park, and N. Kim, “A compact tri-band antenna based on inverted-L stubs for smart devices,” Comput. Mater. Continua, vol. 70, no. 2, pp. 3321–3331, 2022.

[11] H. T. Chattha, M. K. Ishfaq, B. A. Khawaja, A. Sharif, and N. Sheriff, “Compact multiport MIMO antenna system for 5G IoT and cellular handheld applications,” IEEE Antennas Wireless Propag. Lett., vol. 20, no. 11, pp. 2136–2140, Nov. 2021.

[12] Y. Yao, X. Cheng, C. Wang, J. Yu, and X. Chen, “Wideband circularly polarized antipodal curvedly tapered slot antenna array for 5G applications,” IEEE J. Sel. Areas Commun., vol. 35, no. 7, pp. 1539–1549, Jul. 2017.

[13] G. Casula, P. Maxia, A. Fanti, and G. Mazzarella, “A low-cost dual-band CPW-fed printed LPDA for wireless communications,” IEEE Antennnas and Wireless Propagation Letters, vol. 15, pp.1333-1336, 2016.

[14] C. Thajudeen, A. Hoorfar, and F. Ahmad, “Measured complex permittivity of walls with different hydration levels and the effect of power estimation of TWRI target returns,” Progress In Electromagnetics Research B, vol.30, pp.177-199, 2011.

[15] G. A. Casula. G. Montisci, and G. Mazzarella, “A wideband PET inkjet-printed antenna for UHF RFID,” IEEE Antennas and Wireless Propagation Letters, vol. 12, pp. 1400-1403, 2013.

Deep Learning-Based Solar Panels Detection and Segmentation in Remote Sensing Imagery

Dr. Nasru Minallah
National Center in Big Data and Cloud Computing (NCBC), University of Engineering and Technology (UET), Peshawar.
Department of Computer System Engineering, UET Peshawar

n.minallah@uetpeshawar.edu.pk

Tufail Ahmad
National Center in Big Data and Cloud Computing (NCBC), University of Engineering and Technology (UET), Peshawar.
Department of Computer Science and IT, UET Peshawar

tahmad@uetpeshawar.edu.pk

Hadia Khan
National Center in Big Data and Cloud Computing (NCBC), University of Engineering and Technology (UET), Peshawar.
Department of Computer System Engineering, UET Peshawar

hadiakhan@uetpeshawar.edu.pk

Mahmood Ali Khan
National Center in Big Data and Cloud Computing (NCBC), University of Engineering and Technology (UET), Peshawar.
Department of Computer System Engineering, UET Peshawar

mahmoodali@uetpeshawar.edu.pk

***Abstract*— As global energy needs shift towards solar power, effective monitoring of solar panel installations is of paramount importance. This study develops an advanced automated detection system to accurately identify solar panels using high-resolution satellite imagery, enhancing the tracking and analysis of solar energy infrastructure. By employing deep learning-based object detection with the YOLOv8 algorithm, we created a comprehensive dataset of satellite images sourced from various locations, meticulously annotated for precise solar panel identification. The YOLOv8n model achieved impressive performance metrics, including a precision of 91% and a recall of 87%, reflecting its robustness in diverse environmental conditions. The model's Mean Average Precision (mAP) scores reached 89% at an intersection over union (IoU) threshold of 0.50 and 72% across varied thresholds. However, challenges remain, such as reduced performance in low-resolution images and detection errors caused by obstructions like shadows. Importantly, integrating geospatial data enhances the model's functionality, allowing for accurate mapping of solar panels. Future research will focus on enhancing the YOLOv8 model for better performance in challenging conditions and integrating diverse data sources, such as environmental factors, to refine detection accuracy. The model's adaptability also presents opportunities for implementation in various geographical regions, especially in emerging markets where solar energy adoption is growing. Ultimately, this research contributes to global efforts in renewable energy adoption and addresses climate change challenges.**

Keywords: Solar Panels, Satellite Imagery, YOLO Model, Automation, Sustainable Energy.

I. INTRODUCTION

The growing global demand for energy, integrated with the urgent need to mitigate climate change, has emphasized an essential problem: the insufficient monitoring and management of solar energy resources. As countries seek to transition from fossil fuels to renewable energy[1], the deployment of solar panels has rose rapidly, yet the mechanisms to accurately track and assess these installations remain inadequate. This gap poses risks not only to energy efficiency but also to effective urban planning, investment strategies, and policy-making initiatives aimed at enhancing solar adoption. Addressing this challenge is crucial for maximizing the potential of solar energy as a feasible and sustainable power source.

Recent advancements in solar monitoring technologies have yielded some progress. Traditional methods, including manual inspections and aerial photography[2], have faced limitations due to their labour-intensive nature and prone to technologies, such as satellite imagery and machine learning, while new technologies, such as satellite imagery and machine learning, show great promise in improving detection accuracy, however, significant hurdles still exist in scaling these technologies and adapting them to varying environmental conditions. This indicates that while progress has been made, optimal solutions tailored for comprehensive solar panel detection are still needed.

In response to these challenges, this study develops an advanced automated detection system utilizing cutting-edge object detection techniques, specifically the YOLO (You Only Look Once) framework. By employing high-resolution satellite imagery and deep learning models, this research aims to achieve accurate identification and mapping of solar panel installations across diverse geographies. The methodology focuses on training the YOLO model with curated datasets, enhancing its efficiency and accuracy in real-world applications.

The YOLO algorithm [3] operates through a grid-based prediction system, where the image is divided into a grid, and each grid cell is responsible for predicting potential objects within its area. This simultaneous prediction means that unlike other methods that might require separate stages for bounding box and class prediction, YOLO accomplishes both in one pass. Such a unified architecture not only allows for very fast processing but also makes YOLO[4] ideal for applications requiring real-time object detection, such as autonomous driving and video surveillance. In this study, this means that as we apply YOLO to our satellite imagery dataset, we can expect timely and precise localization of solar panels, making it an ideal solution for monitoring efforts in diverse environmental conditions.

Traditional object detection methods typically involve multiple stages such as region proposal generation [5], feature extraction, and separate classification steps. In contrast, YOLO performs all these tasks in a single pass through the neural network, making it significantly faster for real-time detection. For instance, while traditional methods like Selective Search[6] generate potential object bounding boxes by grouping similar pixels, YOLO predicts both bounding boxes and class probabilities simultaneously for each grid cell in an image. This unified approach not only enhances computational efficiency but also generally requires less computational power, making YOLO particularly well-suited for real-time applications.

One of the main objectives of this study is to ensure that decision-makers, such as policymakers and utility companies, have access to reliable data. By accurately tracking solar panel installations, stakeholders can make informed decisions about energy management and investments in renewable energy infrastructure. Additionally, our system will help identify underperforming or damaged solar panels, facilitating better maintenance strategies and ultimately maximizing the effectiveness of solar energy.

Previous studies have not comprehensively tackled the entire process of detecting solar panels and segmenting them in a given area. This study highlights the absence of a unified approach that combines detection, segmentation which this study aims to address. This research seeks to utilise these state-of-the-art models to improve accuracy and efficiency in identifying solar installations. This research also seeks to revolutionize solar panel monitoring by combining satellite imagery and advanced object detection techniques, we aim to provide efficient and accurate tracking of solar installations. This effort will contribute significantly to making solar energy a more reliable and sustainable power source for the future.

II. RELATED WORK

The rapid adoption of solar energy as a sustainable energy solution has necessitated the development of advanced systems for detecting and mapping solar panels. Accurate detection and mapping are critical for optimizing energy utilization, monitoring installations, and facilitating energy planning. Recent advancements in deep learning, particularly in object detection models like YOLO (You Only Look Once), have shown promise in automating these tasks. This literature review examines key studies in the field, identifies research gaps, and highlights the contributions of this paper to advancing automated solar panel detection and mapping.

Several studies have explored the use of deep learning for solar panel detection, with varying approaches and architectures. For instance, [7] proposed an end-to-end deep learning framework using a rotated object detection architecture, achieving a mean average precision (mAP) [8] of 83.3% on a diverse dataset of solar power plants. While this study demonstrated the feasibility of deep learning for solar panel detection, it focused on rotated object detection rather than the YOLO architecture, which is the focus of our research. Similarly, [9] compared the performance of seven deep convolutional neural networks (DCNNs) for extracting photovoltaic (PV) arrays from high-resolution remote sensing images. The study identified structural factors that enhance feature extraction, such as separable convolution and attention mechanisms, but did not explore the YOLO architecture in depth. These studies collectively underscore the importance of network architecture in achieving high accuracy, but they leave room for further exploration of YOLO-based models, particularly YOLOv8, which is the focus of our work.

The detection of small-scale and decentralized solar installations presents unique challenges, as highlighted in [10] this study utilized a U-Net model to detect small solar energy systems (SES) from aerial images, achieving promising results on datasets from Sweden and Germany. However, the study also identified research gaps, such as the need for architectural improvements and the integration of 3D building data to estimate solar panel tilt accurately. These challenges are particularly relevant to our research, as we aim to detect solar panels in both urban and rural environments, where small-scale installations are common.

Additionally, [11] highlighted the limitations of existing data sources for small-scale solar PV detection and the heterogeneity in performance evaluation strategies across studies. The authors emphasized the need for standardized evaluation methods and the challenges posed by distribution shifts between training and test data. Our research addresses these gaps by proposing a comprehensive evaluation strategy and leveraging multiple custom datasets to ensure robust performance across diverse environments.

The integration of geospatial data into solar panel detection systems is another area that has received limited attention in the literature. While studies like [12] have demonstrated the effectiveness of deep learning in challenging environments, they have not extensively explored the use of geospatial data to enhance detection accuracy. TEMCA-Net, which achieved high precision and recall rates in mining areas, focused primarily on texture and contextual features but did not integrate external geospatial data. Our research contributes to this area by incorporating geospatial data to improve the model's ability to map solar panel installations accurately, particularly in complex environments.

The YOLO architecture has gained significant attention in recent years for its efficiency and accuracy in object detection tasks. [13] demonstrated the effectiveness of YOLOv7 in detecting solar panel defects, achieving a mAP of 85.9% on a dataset of infrared images. Similarly, [14] combined YOLOv8 with Particle Swarm Optimization (PSO) to achieve an mAP of 94% for fault detection in PV cells. These studies highlight the potential of YOLO-based models in solar panel-related tasks but do not extensively explore their application for large-scale solar panel detection and mapping. Our research builds on these findings by evaluating the performance of YOLOv8 in detecting solar panels across diverse geographical settings, addressing challenges such as low-resolution images and obstacles.

Despite the progress made in automated solar panel detection, several research gaps remain. First, there is a lack of standardized evaluation strategies, as highlighted in [11] The heterogeneity in performance evaluation across studies makes it difficult to compare different models and approaches. Second, the integration of geospatial data into detection systems has not been extensively explored, limiting the ability of models to map solar panel installations accurately. Third, while YOLO-based models have shown promise, their application for large-scale solar panel detection and mapping has not been fully investigated. Our research addresses these gaps by proposing a standardized evaluation strategy, integrating geospatial data, and leveraging YOLOv8 for large-scale detection and mapping.

In conclusion, this literature demonstrates significant advancements in automated solar panel detection using deep learning techniques. However, challenges remain in detecting small-scale installations, integrating geospatial data, and standardizing performance evaluation. Our research contributes to addressing these challenges by leveraging the YOLOv8n model, integrating geospatial data, and proposing a comprehensive evaluation strategy. The findings highlight the potential of automated detection systems in improving solar energy management and monitoring, paving the way for better energy planning and resource allocation in the pursuit of sustainable energy goals.

III. METHODOLOGY

The methodology outlines the systematic approach used to develop an automated solar panel detection system using the YOLOv8 model. The process encompasses dataset collection, pre-processing, model architecture, training, evaluation, and fine-tuning. Below is a detailed and organized methodology also Fig. *2* shows a flow diagram of the methodology:

A. Dataset Collection and Preparation

1) Pretrained Dataset

A publicly available dataset comprising 8,474 satellite and aerial images with varying resolutions of 0.1m, 0.3m, and 0.8m was utilized for the initial model training. To ensure consistency, the images were pre-processed by resizing them to a uniform size of 640x640 pixels. Data augmentation techniques, including shear transformation, horizontal flipping, and vertical flipping, were applied to enhance dataset diversity. Following pre-processing, the dataset was divided into training, validation, and test sets, resulting in 17,791 training images, 1,697 validation images, and 845 test images.

2) Custom Datasets

To further refine the model, three custom datasets were curated. The first dataset, Custom Dataset-1, focused on the Hayatabad region, containing 950 high-resolution satellite images depicting urban solar panel installations. This dataset was split into 760 training images (80%) and 190 test images (20%). The second dataset, Custom Dataset-2, comprised 250 satellite images capturing rural solar panel installations in the Mohmand region, divided into 200 training images (80%) and 50 test images (20%). The third dataset, Custom Dataset-3, combined images from both the Hayatabad and Mohmand regions, totaling 1,200 images. This dataset was divided into 960 training images (80%) and 240 test images (20%) to improve model generalization across diverse environments.

3) Ground Truth Data Collection

To ensure high-quality annotations, panel regions of interest (ROI) were manually identified and extracted from high-resolution satellite images. Shapefiles, a geospatial vector data format, were employed to define the geometric boundaries of solar panels, allowing for precise annotations. Additionally, georeferencing techniques were applied to align solar panel locations with real-world geographic coordinates, ensuring accurate localization.

4) Google Earth Satellite Images

To verify and refine solar panel locations, extracted shapefiles were overlaid onto Google Earth satellite images. Panel mask extraction techniques were utilized to isolate solar panels from background noise, thereby improving dataset quality. Bounding boxes were manually drawn around detected panels to serve as labelled training data. Finally, a structured dataset.yaml file was created to define class names, image paths, and dataset configurations, ensuring seamless integration with the YOLO model.

B. Data Augmentation

Data augmentation was applied to enhance model generalization, prevent overfitting, and improve robustness to variations in lighting, orientation, and scale.

1) Augmentation Techniques

Various augmentation techniques were employed to increase dataset diversity. Images were randomly rotated at different angles to simulate various orientations of solar panels. Scaling adjustments ensured the model could detect solar panels at different resolutions. Horizontal and vertical flips were applied to introduce variability in panel placement. Additionally, brightness and contrast adjustments were made to simulate different lighting conditions, further improving the model's adaptability.

C. Model Architecture and Configuration

1) YOLOv8n Model

The YOLOv8n (You Only Look Once version 8 nano) model was selected for its balance between speed and accuracy. The model was configured with the following parameters: the number of classes (nc) was set to 1 for detecting solar panels, while scaling constants included a depth of 0.33, width of 0.25, and a maximum of 1024 channels. The model consisted of 225 layers and contained 3,157,200 parameters, with a computational complexity of 8.9 GFLOPs.

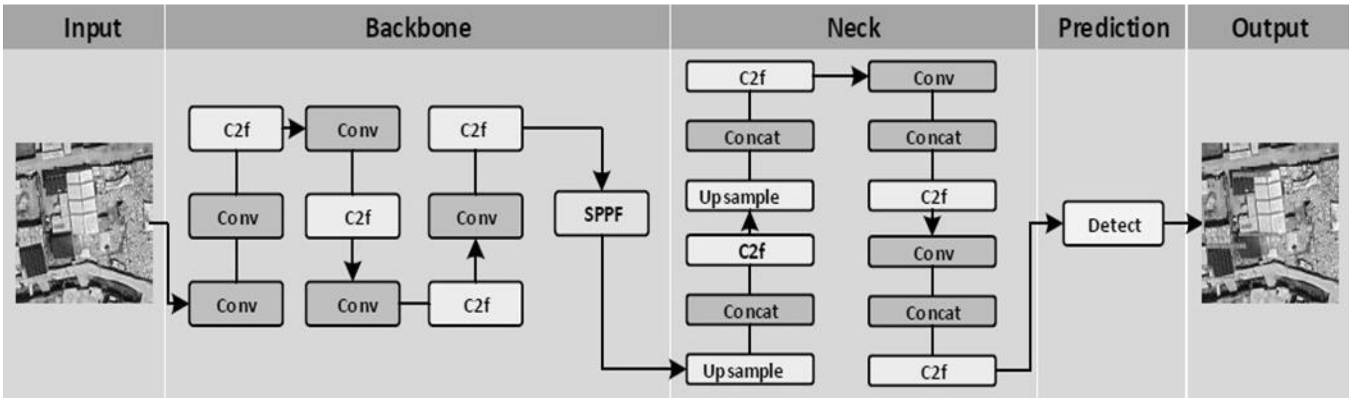


Fig. 1. YOLO block diagram

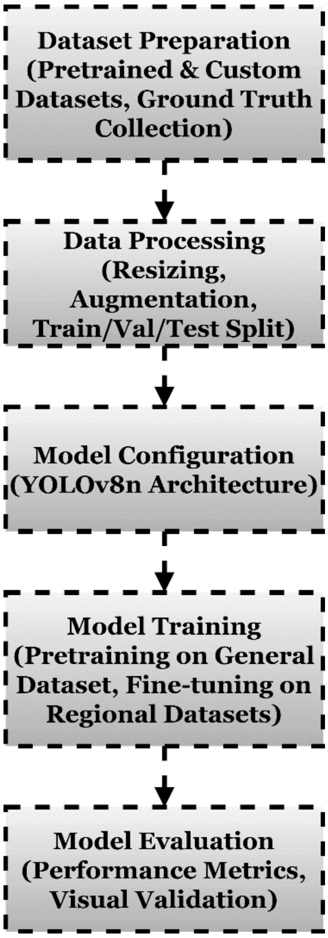


Fig. 2. Flow Diagram of the YOLOv8-based Solar Panel Detection

2) *Model Architecture*

As shown in Fig. *1*. the model's backbone consists of convolutional layers and C2f modules for feature extraction, followed by Spatial Pyramid Pooling-Fast (SPPF) to aggregate contextual features. The head performs upsampling and concatenation to combine high-level and low-level features, followed by detection layers that output bounding boxes and confidence scores.

3) *Model Training*

a) Pretraining

The model was initially trained on the pretrained dataset of 8,474 images to learn basic patterns and features of solar panels. This phase ensured that the model could generalize across diverse environments.

b) Fine Tuning

The pretrained model was fine-tuned on the custom datasets (Hayatabad, Mohmand, and combined) to adapt to specific regional characteristics. The fine-tuning process used the following hyperparameters: training was conducted for 50 epochs with an image size of 640x640 pixels, a batch size of 16, and workers set to 0 to allow data loading to be handled by the main process.

4) Model Evaluation

a) Evaluation Metrics

The model's performance was evaluated using standard object detection metrics. Precision was measured at 91%, representing the proportion of correctly identified solar panels. Recall was recorded at 87%, indicating the proportion of actual solar panels detected. The mean average precision at an IoU threshold of 0.5 (mAP@50) was 89%, while the mean average precision over a range of IoU thresholds (mAP@50-95) was 72%.

b) Visual Inspection

The model's detections were visually inspected on test images, showing accurate bounding boxes and minimal false positives. The model performed consistently across different image conditions, including varying lighting and resolutions.

IV. RESULTS AND DISCUSSION

The results and discussion section present the performance of the YOLOv8n model in detecting solar panels from high-resolution satellite imagery. This section is divided into two main parts: quantitative results (based on evaluation metrics) and qualitative results (based on visual inspection of detected solar panels). The discussion highlights the significance of the results, challenges encountered, and potential areas for improvement.

A. Quantitative Results

The performance of the YOLOv8n model was evaluated using standard object detection metrics, including precision, recall, mean average precision (mAP), and inference time. The results demonstrate the model's effectiveness in detecting solar panels across diverse environments.

Precision (91%):

Precision measures the proportion of correctly identified solar panels out of all detected objects. A high precision value of 91% indicates that the model has a low rate of false positives, meaning it rarely misclassifies non-solar panel objects (e.g., rooftops, reflective surfaces) as solar panels. This is critical for practical applications, as false positives can lead to inaccurate energy assessments and planning.

Recall (87%):

Recall measures the proportion of actual solar panels that were successfully detected. A recall value of 87% indicates that the model missed only 13% of solar panels in the dataset. This is particularly important for large-scale solar energy monitoring, as missed detections can result in underestimating solar energy potential.

Mean Average Precision (mAP@50: 89%, mAP@50-95: 72%):

The mAP@50 metric evaluates the model's accuracy at an Intersection over Union (IoU) threshold of 0.5, achieving a high score of 89%. This indicates that the model accurately localizes solar panels within the bounding boxes.

The mAP@50-95 metric, which averages precision over a range of IoU thresholds (0.5 to 0.95), achieved a score of 72%. This more stringent metric reflects the model's ability to handle varying levels of detection difficulty, such as partially obscured or overlapping solar panels.

Inference Time:

The model demonstrated fast inference times, making it suitable for real-time applications. This is particularly advantageous for large-scale solar panel detection, where processing thousands of satellite images efficiently is essential.

Fig. *3* illustrates the training and validation performance of the YOLOv8n model through various loss and metric curves. The train/box_loss and val/box_loss curves show a steady decline, indicating that the model effectively learned to localize solar panels with increasing accuracy over epochs. Similarly, the train/cls_loss and val/cls_loss curves demonstrate a reduction in classification errors, reflecting the model's ability to distinguish solar panels from other objects. The train/dfl_loss and val/dfl_loss curves, which represent the distribution focal loss, also show consistent improvement, suggesting that the model refined its predictions for bounding box distributions.

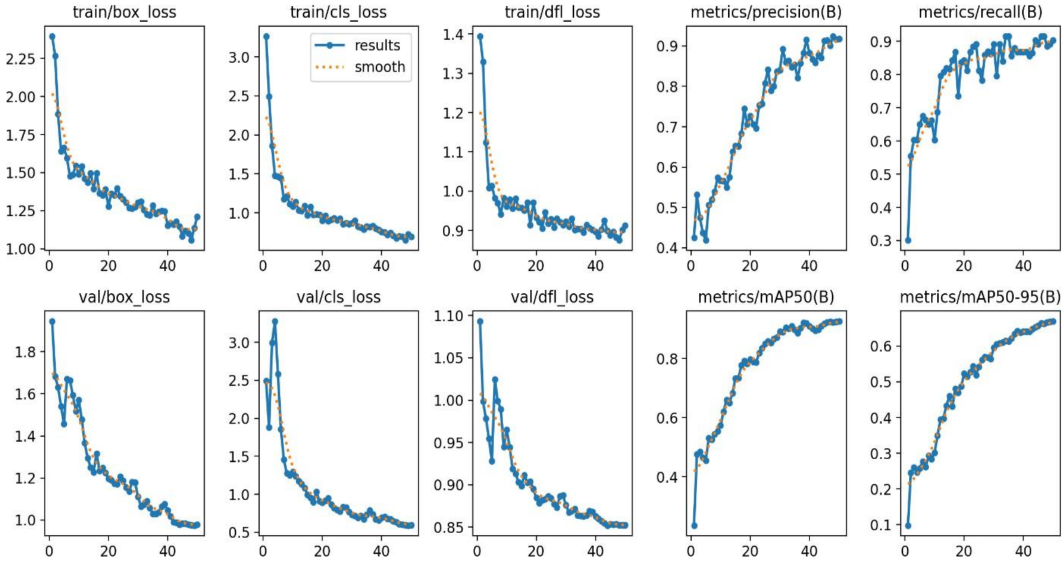


Fig. 3. Performance curves.

In terms of evaluation metrics, the metrics/precision(B) curve reaches a high value of 0.9, confirming the model's low false positive rate, while the metrics/recall(B) curve achieves 0.8, indicating a strong ability to detect most solar panels in the dataset. The metrics/mAP50(B) curve, peaking at 0.6, reflects the model's accuracy in localizing solar panels at an IoU threshold of 0.5. These results align with the quantitative findings, where the model achieved 91% precision, 87% recall, and 89% mAP@50, demonstrating its effectiveness in detecting solar panels across diverse environments. The consistent decline in losses and improvement in metrics underscore the model's robustness and suitability for real-time solar panel detection applications.

B. *Qualitative Results*

The model's detections were visually inspected on test images, showing accurate bounding boxes and minimal false positives. The model performed consistently across different image conditions, including varying lighting and resolutions.

1) *Custom Dataset-1(Hayatabad Region)*

Fig. *4* below presents images from the custom dataset-1 including images from Hayatabad region. The model effectively identifies solar panels, with well-defined bounding boxes indicating accurate detection. The images highlight various orientations and placements of solar panels within urban environments, illustrating the model's ability to adapt to different settings. The model's high precision is evident, as most detected panels are clearly outlined without misclassifications of non-solar panel objects.

2) *Custom Dataset-2 (Mohmand Region)*

Fig. *5* presents images from the custom dataset-2 in Mohmand region, where solar panels are often installed in open areas with fewer obstructions. The bounding boxes are well-localized, and the confidence scores are high, indicating reliable detection.

The model's performance in rural environments highlights its ability to generalize across diverse geographical settings. In rural areas, solar panels may be distributed across larger areas with varying terrain, making accurate detection more challenging. The model's success in this context underscores its adaptability and potential for large-scale deployment in both urban and rural regions.

3) *Custom Dataset-3(Mohmand_Hayatabad)*

Fig. *6* illustrates the model's detections on the combined dataset, which includes both urban and rural environments. The detections are accurate, with minimal false positives and missed detections.

The combined dataset showcases the model's ability to generalize across diverse environments, making it a versatile tool for large-scale solar panel detection. The consistent performance across different settings (urban and rural) highlights the model's robustness and its potential for practical applications in solar energy monitoring and mapping.

Fig. 4. Predicted images of custom dataset -1

Fig. 5. Predicted images of custom dataset -2

C. Key Observation and Interpretations:

Visual inspection of the model's predictions on test images provides insights into its real-world performance. The following observations were made based on the detection outputs:

- **Clear and Well-Localized Bounding Boxes:** The detected solar panels were enclosed within accurate bounding boxes, demonstrating good localization.
- **Minimal False Positives:** The model effectively distinguished between solar panels and similar-looking objects, reducing misclassifications.
- **Consistent Performance Across Different Image Conditions:** The model performed well on images with varying lighting, resolutions, and panel orientations.

Fig. 6. Predicted images of custom dataset -3

While other studies have explored various architectures, such as rotated object detection and deep convolutional neural networks (DCNNs), they did not delve deeply into the YOLO architecture[15].This research focuses specifically on YOLOv8, which has shown superior performance in detecting small and normal-sized target object. This study employs the YOLOv8 model, achieving impressive performance metrics, including a precision of 91% and a recall rate of 87%.

In contrast, other models such as Mask R-CNN have shown varying performance levels, with some studies reporting lower accuracy in similar detection tasks. This positions YOLOv8n as a strong contender in the field of image segmentation for solar panels.

D. Limitations of YOLOv8n

- **Challenges with Low-Resolution Images:**

One of the significant limitations of YOLOv8n is its performance in low-resolution images. The model struggles to maintain accuracy when the quality of the input images is compromised, which can be a common issue in remote sensing imagery.

- **Generalizability Issues:**

While YOLOv8n performs well on specific datasets, its effectiveness may diminish when applied to different datasets or environments. This lack of generalizability can limit its practical applications in diverse geographical settings.

E. Areas for Improvement

Future research should focus on improving the robustness of YOLOv8n in challenging conditions, such as low-resolution imagery or obstructed views. Techniques like data augmentation or the integration of additional contextual information could enhance the model's adaptability.

V. CONCLUSION

In conclusion, this research demonstrates the transformative potential that deep learning techniques, specifically the YOLOv8 model, hold for solar panel detection using satellite imagery. Through careful training and fine-tuning processes, we achieved notable results, with the model exhibiting a precision of 91% and a recall rate of 87%. These metrics underscore its capability to identify solar panels accurately while keeping false positive rates low. Furthermore, the Mean Average Precision (mAP) scores of 89% at an intersection over union (IoU) threshold of 0.50 and 72% across a range of thresholds attest to the model's robustness under varying conditions. Looking ahead, there are several promising avenues for future research stemming from this work. Enhancements to the YOLOv8 model could enable it to perform better in challenging scenarios, such as low-light environments or areas with significant physical obstructions. Moreover, integrating data from diverse sources—such as environmental conditions and ground surveys—could further refine detection accuracy and expand its applicability. The adaptability of this model also opens up opportunities for implementation in various geographical regions, particularly in emerging markets where the shift towards solar energy is gaining momentum. By developing scalable detection frameworks applicable across different contexts, we can significantly contribute to global efforts in increasing renewable energy adoption and tackling climate change.

ACKNOWLEDGMENTS

This research was conducted under the umbrella of the National Centre of Big Data and Cloud Computing (NCBC).

REFERENCES

[1] D. S. Vijayan *et al.*, "Advancements in Solar Panel Technology in Civil Engineering for Revolutionizing Renewable Energy Solutions—A Review," *Energies*, vol. 16, no. 18, 2023, doi: 10.3390/en16186579.

[2] G. Tanda and M. Migliazzi, "Infrared thermography monitoring of solar photovoltaic systems: A comparison between UAV and aircraft remote sensing platforms," *Therm. Sci. Eng. Prog.*, vol. 48, p. 102379, Feb. 2024, doi: 10.1016/j.tsep.2023.102379.

[3] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A Review of Yolo Algorithm Developments," *8th Int. Conf. Inf. Technol. Quant. Manag. ITQM 2020 2021 Dev. Glob. Digit. Econ. COVID-19*, vol. 199, pp. 1066–1073, Jan. 2022, doi: 10.1016/j.procs.2022.01.135.

[4] J. Huang, K. Zeng, Z. Zhang, and W. Zhong, "Solar panel defect detection design based on YOLO v5 algorithm," *Heliyon*, vol. 9, no. 8, Aug. 2023, doi: 10.1016/j.heliyon. 2023.e18826.

[5] T. Kong, A. Yao, Y. Chen, and F. Sun, "Hypernet: Towards accurate region proposal generation and joint object detection," presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 845–853.

[6] J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, "Selective search for object recognition," *Int. J. Comput. Vis.*, vol. 104, pp. 154–171, 2013.

[7] C. Wallace, I. Corley, and J. Lwowski, "Solar Panel Mapping via Oriented Object Detection," 2025, *arXiv*. doi: 10.48550/ARXIV.2502.03592.

[8] P. Henderson and V. Ferrari, "End-to-end training of object class detectors for mean average precision," presented at the Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part V 13, Springer, 2017, pp. 198–213.

[9] L. Li, N. Lu, H. Jiang, and J. Qin, "Impact of Deep Convolutional Neural Network Structure on Photovoltaic Array Extraction from High Spatial Resolution Remote Sensing Images," *Remote Sens.*, vol. 15, no. 18, p. 4554, Sep. 2023, doi: 10.3390/rs15184554.

[10] Â. Frimane, R. Johansson, J. Munkhammar, D. Lingfors, and J. Lindahl, "Identifying small decentralized solar systems in aerial images using deep learning," *Sol. Energy*, vol. 262, p. 111822, 2023, doi: https://doi.org/10.1016/j.solener.2023.111822.

[11] W. Hu *et al.*, "What you get is not always what you see—pitfalls in solar array assessment using overhead imagery," *Appl. Energy*, vol. 327, p. 120143, 2022.

[12] M. Tan, W. Luo, J. Li, and M. Hao, "TEMCA-Net: A texture-enhanced deep learning network for automatic solar panel extraction in high groundwater table mining areas," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 17, pp. 2838–2848, 2023.

[13] A. Haeruman, "AI-Based PV Panels Inspection using an Advanced YOLO Algorithm," presented at the Renewable Energy: Generation and Application, Aug. 2024, pp. 230–237. doi: 10.21741/9781644903216-30.

[14] Q. B. Phan and T. T. Nguyen, "A novel approach for PV cell fault detection using YOLOv8 and particle swarm optimization," presented at the 2023 IEEE 66th International Midwest Symposium on Circuits and Systems (MWSCAS), IEEE, 2023, pp. 634–638.

[15] R. Khanam, T. Asghar, and M. Hussain, "Comparative Performance Evaluation of YOLOv5, YOLOv8, and YOLOv11 for Solar Panel Defect Detection," presented at the Solar, MDPI, 2025, p. 6.

Disease Detection Using Wrist Pulse Analysis

Farman Ali
Department of Computer Systems Engineering, University of Engineerning and Technology Peshawar, Peshawar, Pakistan
20pwcse1872@uetpeshawar.edu.pk

Sudais Rehman
Department of Computer Systems Engineering, University of Engineerning and Technology Peshawar, Peshawar, Pakistan
20pwcse1888@uetpeshawar.edu.pk

Adnan Ayaz
Department of Computer Systems Engineering, University of Engineerning and Technology Peshawar, Peshawar, Pakistan
20pwcse1865@uetpeshawar.edu.pk

Arbab Masood Ahmad
Department of Computer Systems Engineering, University of Engineerning and Technology Peshawar, Peshawar, Pakistan
arbabmasood@uetpeshawar.edu.pk

Abstract

Early-stage detection of diseases is very crucial for treatment and management. Detection of diseases involves invasive and expensive medical techniques. However, the latest developments in non-invasive methods are highly successful in identifying illnesses. The wrist pulse has historically been a vital tool for disease detection. Traditional Chinese medicine makes extensive use of this technology, which shows great promise in diagnosing a wide range of illnesses. This study offers a thorough analysis of research on the analysis of wrist pulse methods and their uses in the identification of diseases. It investigates the physiological underpinnings of wrist pulse analysis, particularly the connection between underlying medical disorders and the characteristics of wrist pulses. The study also explores how to utilize wearable pulse detectors and machine learning algorithms to improve the precision and effectiveness of wrist pulse analysis. Here we are using a dataset of 300 samples for different diseases and analyzing it using MATLAB and then training some ensemble learning algorithms on it. We've achieved 80+ accuracies almost for all algorithms and the accuracy can be improved if the dataset is expanded by providing more samples and extracting some other features.

Keywords: Wrist Pulses, Wrist Pulse Analysis, Disease Detection, Pulse Signal Processing and analysis, Machine Learning.

I. INTRODUCTION

Looking, listening, inquiring, and feeling wrist pulses are some of the methods that are used in traditional Chinese medicine (TCM), which has been practiced for many centuries. The most important component of TCM, the wrist pulse diagnosis is simple, easy and has no negative side effects. In Western medicine, the body's blood flow is used to explain wrist pulse diagnosis. For example, the pulse of a person may feel different when they have lung cancer due to alterations in their lungs. Although it requires deep knowledge and experience, TCM doctors typically use three fingers to diagnose the pulse at several wrist sites. Researchers have improved the accuracy, precision and dependability of pulse diagnosis over time by utilizing advanced technology like Artificial Intelligence & Machine Learning. Based on the way blood flows through the body, this technique has grown in importance during medical examinations. Smart gadgets can now be used to monitor pulses at home because of advancements in wearable technology. To aid in the diagnosis of some illnesses, we have devised a method for analyzing wrist pulse data. [1]

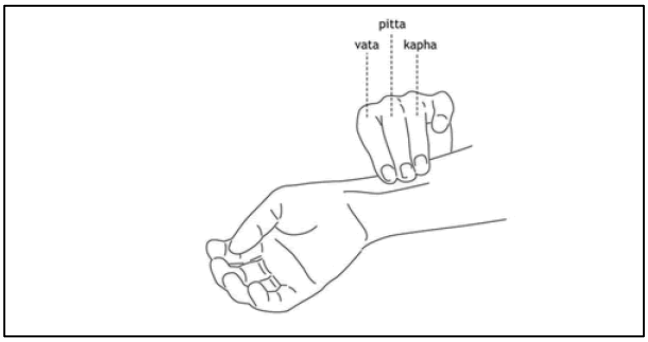


Figure: 1.1, Different parts of a wrist pulse

A) Wrist pulse and its parts

The wrist pulse is the most fundamental aspect of traditional Chinese medicine (TCM), and is made up of three separate points called Cun, Guan, and Chi. These points provide unique insights about an individual's health status. The doctors carefully examine the wrist pulse's strength, rhythm, and quality at these specific points of the wrist. Pulse of the deep layer is represented by Chi, the pulse of the medium layer by Guan, and the pulse of the shallow layer by Cun. Practitioners can detect minute changes by palpating these areas, which could point to underlying medical issues or bodily imbalances.[2]

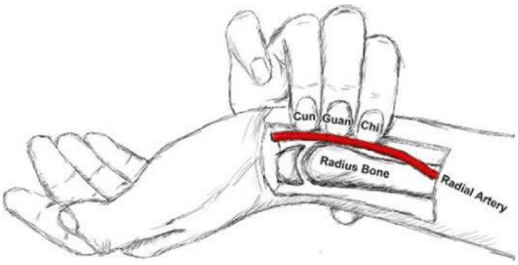


Figure: 1.2, Illustration of different parts of wrist pulse

B) The Chi or Vata

In Ayurvedic medicine, Vata is a mix of Ether and Air. It's important for two main things: movement in the body and mind, and communication. It helps muscles move, nerves work and keeps our thoughts and feelings flowing. When Vata is out of balance, it can cause restlessness, anxiety, and other problems. So, keeping Vata in check is important for staying healthy and feeling good. The Vata component possesses certain distinct characteristics such as irregular rhythm patterns, 80-90 bpm, fast, feeble, cold with lighter weight. [3]

C) The Guan or Pitta

Pitta is a mix of Fire and Water in Ayurvedic medicine, and it's connected to functions like digestion and metabolism. Pitta gets out of balance, it can cause problems with vision, digestion, and metabolism, especially as people get older. Pitta also shows up in the pulse as a regular rhythm with a strong beat, usually around 70-80 beats per minute. It's also linked to heat and moisture in the body. Kapha, on the other hand, is a mix of Earth and Water, and it's all about the body's structure and fluids. When Kapha is imbalanced, it can cause issues like congestion, poor taste and smell, back pain, and weight gain, especially in childhood.[3]

D) The Cun or Kapha

The Kapha shows up in the pulse as a steady rhythm, beating about 50-60 times per minute, and move slowly and smoothly. In Ayurveda, Vata, Pitta, and Kapha control our body, mind, soul, and spirit. Checking the pulse at the wrist is a special method in Ayurveda that helps understand a person's health and what might be causing any issues. The pulse wave, which is like a pressure signal, can be checked at different places like the foot, wrist, or neck. But usually, it's checked at the wrist because it's easy and gives accurate results. This pressure signal is created by the heart squeezing and relaxing, which pushes blood through the arteries. By looking at this signal, we can learn a lot about a person's health, both the good parts and any problems they might have.[3]

E) Machine Learning Models

1. XGBoost: XGBoost (Extreme Gradient Boosting) is an optimized implementation of gradient boosted decision trees designed for speed and performance. It includes regularization techniques to prevent overfitting and can handle missing values efficiently. Hyperparameters for XGBoost were optimized using Randomized SearchCV [4].

2. LightGBM: LightGBM (Light Gradient Boosting Machine) uses tree-based learning algorithms and is known for its efficiency and scalability, capable of handling large datasets with high-dimensional features. To determine the optimal splits, LightGBM uses a histogram-based approach, which speeds up training and uses less memory. Randomized Search CV was used for hyperparameter tuning.

3. Gradient Boosting Classifier: This method for regression and classification issues incorporates new models to address the shortcomings of preexisting models in a step-by-step manner. It can be computationally demanding, despite its effectiveness. To maximize model performance, Randomized Search CV was used for hyperparameter adjustment.

4. Hist Gradient Boosting Classifier: Using histogram-based techniques, this streamlined variant of the Gradient Boosting Classifier speeds up training and prediction. Because it can handle high-dimensional data more effectively than conventional gradient boosting techniques, this classifier is especially well-suited for huge datasets. Randomized Search CV was used for hyperparameter tweaking.

5. Stacking Classifier: We used a stacking classifier, which mixes many base models to increase predicted accuracy, to significantly improve classification performance. A meta-classifier that learns to generate the final prediction is fed the outputs of a variety of machine learning models that serve as base learners in the stacking framework. By utilizing the advantages of many classifiers, this ensemble strategy lessens overfitting and enhances generalization.

II. METHODOLOGY

A) Data Collection

The dataset used in this study consists of 300 wrist pulse signal samples total, divided into five classes—normal (100 samples), pancreatitis (54 samples), diabetes mellitus (DBU) (77 samples), appendicitis (35 samples), and acute appendicitis (54 samples)—make up the dataset used in this study. For additional investigation, the pulse signals were stored in MATLAB (.mat) format. Basic statistical features, frequency domain characteristics, rolling statistics, extra features like autocorrelation and entropy, and advanced statistical features generated from rolling windows were among the features that were taken out of the time series data. Then, to improve classification accuracy, these features were fed into machine learning models

B) Feature Extraction

Important characteristics were extracted, including waveform morphology, frequency, pulse amplitude, and statistical factors. The techniques of normalization and noise filtering were used for preprocessing the signals. Features like mean, median, mode, standard deviation, min, max, skewness, kurtosis in time domain and in the frequency domain with other features like rolling mean, rolling median and rolling mode features were extracted and used for the prediction. The autocorrelation features were also used to analyze the periodicity and temporal dependencies in the wrist pulse signals. The rolling mean and Standard Deviation are the most important features because they reflect the fundamental structure and variability of the pulse, which is often the clearest indicators of the health conditions.

C) Machine Learning Models:

We've used the gradient boosting algorithms for the classification, such as Hist Gradient Boosting Classifier, LightGBM, XGBoost, and Gradient Boosting Classifier. The dataset were divided into training and testing sets as according to the dataset division rules and the performance indicators like accuracy, precision, recall, and F1-score were calculated.

D) Model Architecture:

In this study we employed the ensemble learning architectures, especially the Gradient Boosting-based Models, to detect the diseases from the wrist pulse signals. We've used Hist Gradient Boosting Classifier, LightGBM, XGBoost, and Gradient Boosting Classifier, these models are based on decision tree ensembles and are optimized for speed and performance.

E) *Model Validation & Training:*

To avoid overfitting and guarantee robustness a stratified 10-fold cross validation technique was used to train the classifiers on the specific retrieved features. The actions listed below were taken.

1. Data Splitting: The dataset was divided into training and testing sets, with 80% of the data used for training and 20% for testing.

2. SMOTE: The Synthetic Minority Over-sampling Technique (SMOTE) was applied to address class imbalance by generating synthetic samples for minority classes.

3. Hyperparameter Tuning: We've used the Randomized Search CV to perform hyperparameter tuning for each classifier, and optimized the parameters such as the learning rate, maximum depth, and the number of estimators.

4. Model Training: The classifiers were trained on the training set using the best hyperparameters identified during tuning.

5. Model Evaluation: The performance of each classifier was evaluated on the testing set using metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

F) *Visualizations* of Feature Importance:

The feature importances graphs for each model are displayed in Figures 2.1- 2.3, to determine which characteristics, have the most predictive power.

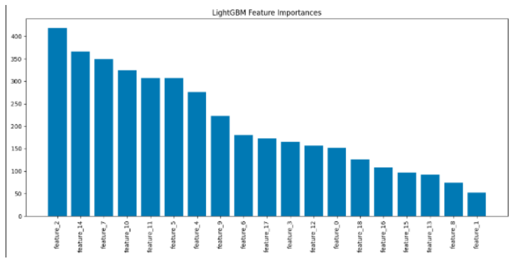


Figure 2.1 LightGBM Featuer Importances

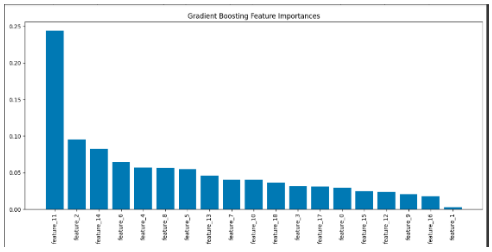


Figure 2.2 Gradient Boosting Feature Importances

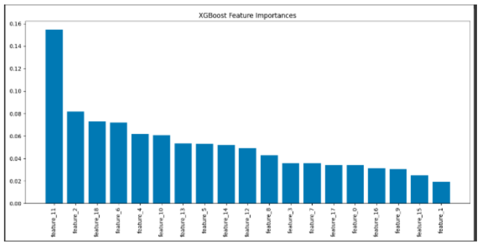


Figure 2.3 XGBoost Feature Importances

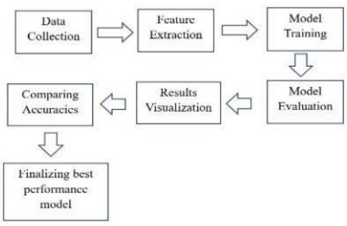


Figure 2.4 Methodology flow diagram

III. RESULTS & DISCUSSION

The performance of the four classifiers—XGBoost, LightGBM, Gradient Boosting Classifier, and Hist Gradient Boosting Classifier has been assessed as according to the above discussed methodology and the results are shown. Table 1 presents a summary of the findings, as illustrated below.

Classifier	Accuracy	Precision	Recall	F1-Score
XGBoost	0.83	0.84	0.83	0.89
LightGBM	0.80	0.79	0.80	0.80
Gradient Boosting Classifier	0.78	0.82	0.78	0.88
HistGradientBoostingClassifier	0.81	0.83	0.81	0.81
Stacking Classifier	0.80	0.83	0.80	0.80

Table 1: Results of different classifiers used

XGBoost: The XGBoost had the highest accuracy and F1-score, among the others we've used in this project indicating its resilience and efficacy in differentiating between various medical problems. Its superior performance was influenced by its capacity to deal with missing variables and carry out regularization to avoid overfitting.

The XGBoost emerging as the best performer in this study, it demonstrates the great effectiveness of ensemble approaches such as XGBoost, LightGBM, GradientBoostingClassifier, and HistGradientBoostingClassifier in biomedical signal classification. To give a better understanding of the classifiers' capabilities, the following sections will go into greater details and graphs for each classifier about the particular outcomes and feature importance studies.

True/Pred	0	1	2	3	4
0	9	0	1	1	0
1	0	18	0	3	2
2	3	3	21	1	1
3	2	0	1	11	0
4	1	0	1	0	21

Table 2: XGBoost Confusion Matrix

LightGBM: LightGBM also did well achieving a high accuracy and F1-score,. It was experienced that it is appropriate for this classification challenge because of its memory-efficient design and histogram-based technique for identifying splits. It did well, however, performed somewhat a little worse as compare to the XGBoost.

True/Pred	0	1	2	3	4
0	10	0	1	0	0
1	1	19	0	1	2
2	3	3	21	1	1
3	2	0	1	11	0
4	1	0	2	0	21

Table 3: LightGBM Confusion Matrix

Gradient Boosting Classifier: Gradient Boosting Classifier showed good performance but was computationally intensive and slow. Although it also performed well but its accuracy and F1-score were slightly lower than those of XGBoost and LightGBM, making it less suitable for large datasets or real-time applications.

True/Pred	0	1	2	3	4
0	10	0	1	0	0
1	1	20	0	2	0
2	3	5	19	1	1
3	3	0	1	10	0
4	3	0	1	0	19

Table 4: Gradient Boosting Confusion Matrix

Hist Gradient Boosting Classifier: This improved Gradient Boosting Classifier showed great accuracy and F1-score, as compare with XGBoost. The faster training and prediction times of this model were made possible by its histogram-based methodology, which made it a good choice for large datasets.

True/Pred	0	1	2	3	4
0	9	1	1	0	0
1	1	19	0	1	2
2	2	3	22	1	1
3	2	0	1	11	0
4	2	0	1	0	20

Table 5: Confusion matrix of Hist Gradient Boosting Classifier

Stacking Classifier: The best models from the previous phase of the training are used to build a stacking classifier. It uses a meta-estimator (here the logistic regression) to aggregate the predictions from several base estimators used.

True/Pred	0	1	2	3	4
0	10	0	1	0	0
1	1	19	0	1	2
2	4	3	20	1	1
3	2	0	1	11	0
4	2	0	1	0	20

Table 6: Stacking Classifier Confusion Matrix

Results & Discussions: The experimental results here demonstrated the effectiveness and easiness of using machine learning classifiers for computerized wrist pulse diagnosis. XGBoost and Hist Gradient Boosting Classifier showed the best performance, achieving high accuracy and F1-scores. The feature importance analysis provided valuable insights into which different aspects of the pulse signals were most indicative of specific health conditions.

The results suggest that computerized/automated pulse diagnosis using Machine Learning can enhance diagnostic accuracy and provide a standardized method for analyzing pulse signals. This approach can potentially reduce the subjectivity associated with traditional pulse diagnosis and support clinicians in making more informed decisions.

This method also provide very easy way of diagnoses the pulse with no side effects.

Below are the tables for different classifiers with their different accuracies and results.

1. **LightGBM Results**
   Accuracy: 80.00%

Class	*Precision*	*Recal*	*F1-score*	*Support*
Normal processed	*0.56*	*0.91*	*0.69*	*11*
Disease_AA_processed	*0.86*	*0.83*	*0.84*	*23*
Disease_A processed	*0.84*	*0.72*	*0.78*	*29*
Disease_DBU_processed	*0.85*	*0.79*	*0.81*	*14*
Disease_P processed	*0.86*	*0.83*	*0.84*	*23*
Accuracy			*0.80*	*100*
Macro Avg	*0.79*	*0.81*	*0.79*	*100*
Weighted Avg	*0.82*	*0.80*	*0.80*	*100*

Table 7: LightGBM results output

2. ***XGBoost Results***
   Accuracy: 83.00%

Class	*Precision*	*Recal*	*F1-score*	*Support*
Normal processed	*0.64*	*0.82*	*0.72*	*11*
Disease_AA_processed	*0.87*	*0.87*	*0.87*	*23*
Disease_A processed	*0.85*	*0.79*	*0.82*	*29*
Disease_DBU_processed	*0.79*	*0.79*	*0.79*	*14*
Disease_P processed	*0.91*	*0.87*	*0.89*	*23*
Accuracy			*0.83*	*100*
Macro Avg	*0.81*	*0.83*	*0.82*	*100*
Weighted Avg	*0.84*	*0.83*	*0.83*	*100*

Table 8: XGBoost results output

3. ***Gradient Boosting Results***
   Accuracy: 78.00%

Class	*Precision*	*Recal*	*F1-score*	*Support*
Normal processed	*0.50*	*0.91*	*0.65*	*11*
Disease_AA_processed	*0.80*	*0.87*	*0.83*	*23*
Disease_A processed	*0.86*	*0.66*	*0.75*	*29*
Disease_DBU_processed	*0.77*	*0.71*	*0.74*	*14*
Disease_P processed	*0.95*	*0.83*	*0.88*	*23*
Accuracy			*0.78*	*100*
Macro Avg	*0.78*	*0.79*	*0.77*	*100*
Weighted Avg	*0.82*	*0.78*	*0.79*	*100*

Table 9: Gradient Boosting results output

4. ***Hist Gradient Boosting Results***
   Accuracy: 81.00%

Class	*Precision*	*Recal*	*F1-score*	*Support*
Normal processed	*0.56*	*0.82*	*0.67*	*11*
Disease_AA_processed	*0.83*	*0.83*	*0.83*	*23*
Disease_A processed	*0.88*	*0.76*	*0.81*	*29*
Disease_DBU_processed	*0.85*	*0.79*	*0.81*	*14*
Disease_P processed	*0.87*	*0.87*	*0.87*	*23*
Accuracy			*0.81*	*100*
Macro Avg	*0.80*	*0.81*	*0.80*	*100*
Weighted Avg	*0.83*	*0.81*	*0.81*	*100*

Table 10: Hist Gradient Boosting results output

5. ***Stacking Classifier Results***
   Accuracy: 80.00%

Class	*Precision*	*Recal*	*F1-score*	*Support*
Normal processed	*0.53*	*0.91*	*0.67*	*11*
Disease_AA_processed	*0.86*	*0.83*	*0.84*	*23*
Disease_A processed	*0.87*	*0.69*	*0.77*	*29*
Disease_DBU_processed	*0.85*	*0.79*	*0.81*	*14*
Disease_P processed	*0.87*	*0.87*	*0.87*	*23*
Accuracy			*0.80*	*100*
Macro Avg	*0.80*	*0.82*	*0.79*	*100*
Weighted Avg	*0.83*	*0.80*	*0.80*	*100*

Table 11: Stacking Classifier's results output

From the above results and all of the results output for the classifiers used for the prediction of diseases it is cleared that the XGBoost and Hist Gradient Boosting Classifier showed best performance on the given dataset with accuracies as 83.00% and 81.00% respectively.

IV. Recommendations for Future Research

• **Expansion of the dataset:** To increase the classifiers' robustness and generalizability, it would be very helpful to gather a bigger and more varied dataset that covers a greater variety of illnesses and medical variances.

• Adding Additional Physiological Signals: To improve precision and offer a deeper evaluation of cardiovascular health, one should combine the wrist pulse signals with other physiological data, such as blood pressure and ECG.

• Deep Learning: Trying the Deep Learning models such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs) for the analysis of the pulse signal will be very helpful to increase the detection accuracy and precision. More intricate linkages and patterns in the data might be captured by using these models.

• Real-time Applications: One improvement in this research could be Creating a real-time computerized pulse diagnostic application. It would provide the accessibility to examine the deployment's practical features, such as patient acceptability, ease of use, and interaction with current healthcare systems.

V. ACKNOWLEDGEMENT

We sincerely thank all those who helped us in making this project a success, especially the faculty and staff at UET Peshawar, for their essential support and direction during this study.

VI. REFERENCES

[1] Zhichao Zhanga, Yuan Zhanga, Lina Yaob, Songc, Anton Kosd "A sensor-based wrist pulse signal processing and lung cancer recognition,"*Jornal of Biomedical Informatics 79(2018),* pp. 107–116.

[2] Shihang Wang, Zhinan Zhang, Zhijian Chen, Deqing Mei, and Yancheng Wang,"Development of Pressure Sensor Based Wearable Pulse Detection Device for Radial Pulse Monitoring", Micromachines 2022, 13, 1699.

[3] Suguna GC, Veerabhadrappa ST "A review of wrist pulse analysis" *Biomedical Research 2019,* vol. 30, no. 4, pp. 538–545, May 17, 2019.

[4] Zhang, X. et al. (2024). Classification of Tumor Types Using XGBoost Machine Learning Model: A Vector Space Transformation of Genomic Alterations. Journal of Translational Medicine.

[5] Sharoni Narang, Saumya P S K, Omkar Batwal, Mrunal Khandagale, "Ayurveda based Disease Diagnosis using Machine Learning", International Research Journal of Engineering and Technology (IRJET), Volume 5, March 2018.

[6] Jie Wang Yirun Zhu, Zhiyong Wu, Yunlin Zhang, Jian Lin, Tao Chen, Huicong Liu, Fengxia Wang, Lining Sun "Wearable multichannel pulse condition monitoring system based on flexible pressure sensor arrays", Microsystems & Nanoengineering (2022)

[7] Wangmeng Zuo, Member, IEEE, Peng Wang, and David Zhang, Fellow, IEEE, "Comparison of Three Different Types of Wrist Pulse Signals by Their Physical Meanings and Diagnosis Performance", IEEE journal of biomedical and health informatics, vol. 20, no. 1, january 2016.

[8] Bhaskar Thakker, Anoop Lal Vyas, "Support Vector Machine for Abnormal Pulse Classification", International Journal of Computer Applications (0975-8887), Volume 22-No7,May 2021.

[9] N. N. Khaire and Y. V. Joshi, "Diagnosis of disease using wrist pulse signal for classification of pre-meal and post-meal samples," in *Proc. 2015 Int. Conf. Ind. Instrum. Control (ICIC)*, Pune, India, May 28–30, 2015,

[10] H. Pogadadanda, U. S. Shankar, and K. R. Jansi, "Disease diagnosis using Ayurvedic pulse and treatment recommendation engine," in *Proc. 7th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, Mar. 2021, doi: 10.1109/ICACCS51430.2021.9441843.

[11] B. Thakker and A. Vyas, "Support vector machine for abnormal pulse classification," *Int. J. Comput. Appl.*, vol. 22, no. 7, pp. 1–5, May 2011, doi: 10.5120/2597-3610.

AmpliStride: From Signal to Stride A Breakthrough for Leg Paralysis Rehabilitation

Anum Khan
Department. of Computer Systems Engineering
University of Engineering and Technology, Peshawar Pakistan
20pwcse1879@uetpeshawar.edu.pk

Urooj Qaiser
Department. of Computer Systems Engineering
University of Engineering and Technology, Peshawar Pakistan
20pwcse1889@uetpeshawar.edu.pk

Mehmood Khan
Department. of Computer Systems Engineering
University of Engineering and Technology, Peshawar Pakistan
20pwcse1897@uetpeshawar.edu.pk

Arbab Masood Ahmad
Department. of Computer Systems Engineering
University of Engineering and Technology, Peshawar Pakistan
arbabmasood@uetpeshawar.edu.pk

***Abstract*— Foot drop is a condition in which a patient fails to lift a foot due to neuro-muscular disorder of the lower body. Many assistive devices are available, but they have some limitations. For this purpose, we offer a solution: Muscle Signal Amplification and Transmission System (MSATS). The system picks muscle signal from the healthy leg and after necessary processing and amplification transmits it to the muscle stimulator worn on the affected leg. The system stimulates the muscle with proper timing according to the gait cycle. This project aims to improve the quality of life of those afflicted with foot-drop by assisting their mobility and independence.**

Keywords: MSATS; Paralysis; Technology; Muscle; Strength.

I. INTRODUCTION

Foot drop, a disorder suffered by millions of individuals globally, results in dysfunction of the lower limbs and weakening of muscles. Many people around the world suffer from this disability, that interferes with their mobility and independence. The challenges faced by people suffering from foot drop due to paralysis or other neuronal disorders necessitate the discovery of novel solutions. MSATS is an advancement in functional walking. Unlike other solutions the proposed method makes use of the signal picked from healthy leg, processing and transmitting it to the disabled leg. On the disabled leg the receiving circuit produces the necessary delay and applies electrical stimulus to the muscle at the right time.

The Muscle Signal Amplification & Transmission System (MSATS) is poised to be the first solution set to transform foot drop management. The objective of this project is to leverage advanced technology to close gaps in existing assistive devices, offering new options to restore mobility and independence [1]. Signal processing systems form the backbone of MSATS, offering accurate signal capture, noise removal, and transmission efficiency. The inclusion of functional electrical stimulation (FES) is a targeted stimulation element, enhancing muscle response and mobility.

Foot-drop is a debilitating condition caused by any of the many possible disorders of the nerves or muscles of the lower body and results in loss of muscle strength and motor skills in the lower limbs. Existing assistive technology rarely can effectively handle this disability. This project seeks to fill this gap by providing an alternative muscle signal amplification transmission systems (MSATS) that will significantly improve the mobility and quality of life of people with paraplegia. The background to our project focuses on managing foot drops through neurologic signal amplification (MSATS), which involves understanding the condition, its impact on individuals, and limitations of current assistive technology [2].

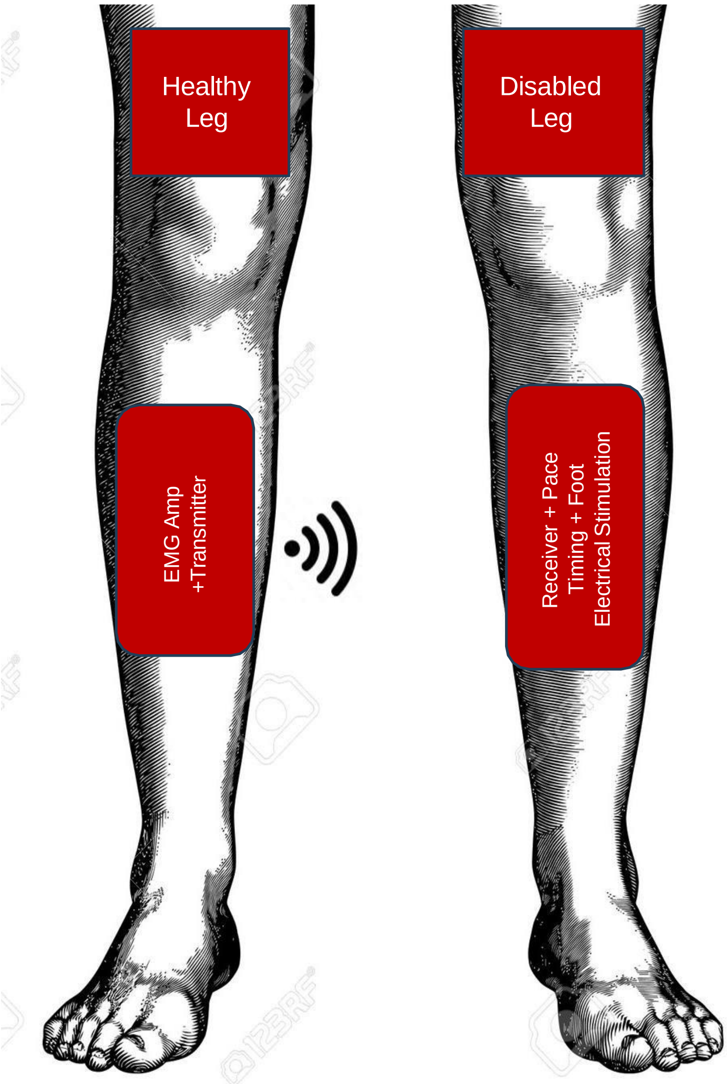


Figure 1: A signal is captured, conditioned, transmitted, delayed, and then used for Functional Electrical Stimulation (FES).

Amplistride is a groundbreaking technological innovation in rehabilitation from leg paralysis, offering an end-to-end solution with the integration of the most advanced technologies to restore mobility and enhance the lives of paralysis patients. Amplistride fundamentally employs a sophisticated neural interface mechanism, seamlessly translating neural signals into real-time representations of user intent, thus bridging the essential communication gap between the brain and paralyzed limbs. Adaptive learning algorithms employed ensure a personalized and progressively optimized rehabilitation process, responding to the diverse set of challenges presented by different types and magnitudes of leg paralysis.

Biomechanical feedback systems prevent compensatory movements, ensuring natural and effective strides. Virtual reality incorporation offers an interactive rehabilitation setting, engaging the user and not merely involving them, but also positively affecting their psychological condition. Preliminary clinical trials suggest the efficacy of Amplistride, with improvements in muscle strength, joint range of motion, and mobility in general. As a holistic and state-of-the-art solution, Amplistride is set to revolutionize disability rehabilitation, an important step towards restoring independence and enhancing the quality of life for patients with this disabling condition [6].

II. CHALLENGES AND ADVANCEMENTS IN FOOT DROP ASSISTIVE TECHNOLOGY

*A. Impact on individuals***:**

Foot drop significantly impairs mobility and can cause difficulties in independent daily activities. Loss of foot control affects gait, causing instability, falls, and decreases confidence in walking. Existing assistive devices such as braces, orthotics, or electrical stimulation systems partially address the issue but often present limitations in function, comfort, and effectiveness

B. Limitations of current assistive technology:

Current assistive foot control technologies have drawbacks. Braces and orthotics provide support but can restrict natural movement or lack flexibility in positions. Electrical stimulation systems have shown promise but often struggle with accuracy, flexibility, and user control.

C. Technological Prospects:

MSATS incorporates state-of-the-art technologies such as wearable devices, signal processing systems and wireless transmission mechanisms. Improvements in these areas not only benefit from the use of walking but also pave the way for innovation in related areas [4].

III. AMPLISTRIDE FOR SOCIAL GOOD

Integrating mobility through the MSATS may enhance the long-term reduction of disability associated health care costs, assist in socializing the walking disabled and improve their overall wellbeing.

A. Longevity:

Develop sustainable, user-friendly solutions and ensure implementation and durability in addressing foot drop challenges. Longevity helps to keep the lives of those affected moving forward.

In conclusion, the value of our project is that it can change the lives of people with foot drop by providing effective, innovative and accessible solutions that go beyond limitations already inherent in assistive technology. This mandate is in accordance with the mandate to improve health technology to promote the well-being of people with mobility impairment.

B. Accessibility:

Make MSATS affordable and highly accessible so that sophist icated assistive technology is accessible to people with different socioeconomic status and healthcare systems.

C. Adaptability:

Design MSATS adaptable for different user needs so that it can be adjusted for unique needs by utilizing adjustable stimulation level, sensitivity, and device to suit the degree of severity of foot drop.

D. Integration with Rehabilitation Programs:

Collaborate with physical therapists and healthcare providers to integrate MSATS into rehabilitation programs, optimizing recovery results and enhancing long-term mobility of patients with foot drop. Changes that are capable of successfully implementing diversity.

IV. METHODOLOGY

The research approach employed in this research study is intended to counteract the problems brought about by leg paralysis by developing and implementing the Muscle Signal Amplification and Transmission System (MSATS). Leg paralysis is a disabling condition that impacts millions of individuals worldwide, and it is a severe impairment of muscle strength and motor function in one or both legs. Existing assistive technologies are at times inadequate for individuals with leg paralysis, hence the necessity for innovative Solutions for enhancing mobility and quality of life.

The method described here is a systematic design, development, and testing of the MSATS system. It consists of several major components, including the signal capture module, amplification algorithm, transmission mechanism, user interface, and FES module. Each of these components is systematically designed and integrated to form an integrated system to restore mobility and independence to leg paralysis patients [11]. In the research process, rigorous evaluation and validation methods are employed for the analysis of the performance and functionality of the MSATS prototype. Laboratory testing in controlled environments allow testing of the signal capture, amplification, and transmission processes using EMG and FES modules. In addition, user trials on leg paralysis patients provide an insight into the usability and effectiveness of the MSATS system in real-world usage.

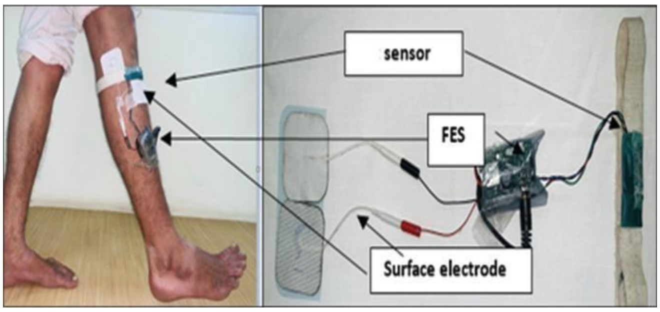


Figure 2: Wearable FES system for foot drop correction with sensors and surface electrodes.

While the proposed methodology presents a comprehensive approach to addressing the challenges of leg paralysis, it is essential to acknowledge potential limitations and avenues for future research. By embracing a multidisciplinary approach and leveraging advances in technology, the MSATS system holds the promise of significantly enhancing the quality of life for individuals living with leg paralysis.

A. Integrating with EMG signals:

FES can be triggered based on decoded EMG signals to facilitate coordinated limb movements. For example, when EMG signals indicate the intention to move a specific muscle group, corresponding FES electrodes can be activated to stimulate those muscles, initiating movement [12].

B. Closed loop control:

Implementing a closed-loop control system that combines EMG signal decoding with FES activation allows for real-time adjustments in stimulation parameters. By continuously monitoring EMG signals and adjusting FES parameters accordingly, the system can adapt to changes in muscle activity and user intent, optimizing movement control.

C. Muscle Conditioning & Rehabilitation:

FES can also be used for muscle conditioning and rehabilitation. By selectively stimulating paralyzed muscles electrically, FES prevents muscle atrophy, maintains strength in the muscles, and promotes neuroplasticity, thus improving overall long-term mobility results for paralyzed patients.

D. User feedback and Adjustment:

The integration of user feedback mechanisms in the system enables users to give feedback on FES intensity, timing, and effectiveness. The feedback channel enables personalization of the adjustment of FES parameters to suit individual preference as well as maximize user comfort and performance.

E. Technological Innovation:

Technological innovation is at the core of MSATS. The building blocks are signal processing systems, which provide accurate signal acquisition, minimal noise, and secure transmission. Adding muscle responsiveness and general mobility, the integration of Functional Electrical Stimulation (FES) provides a differentiating stimulation feature.

F. Potential Impact:

The possible influence of MSATS on the rehabilitation process for limb paralysis is investigated in detail. The main topic of debate is how the system, by utilizing cutting-edge technologies, has the potential to completely transform walking aid. Through the restoration of muscle control and mobility, MSATS presents a promising opportunity to greatly improve the quality of life for those who suffer from limb paralysis, promoting self-sufficiency and general welfare [16].

G. Future Directions:

This section examines the problems of cost and accessibility while acknowledging the difficulties that MSATS has experienced. The significance of enduring effectiveness is examined, guaranteeing consistent advantages for prolonged durations. There is thought to be room for MSATS to be applied to a wider range of paralysis kinds and demographic groupings. The necessity for continued innovation and addressing new requirements in rehabilitation are highlighted as future research priorities.

IV. RESULTS

The device that we are implementing will take EMG signals from healthy leg through EMG module and transmit it to the affected leg through esp32. Then we will analyze the value and will give stimulus to the affected leg.

Normally when we take forward step the EMG module gives value approximately greater than 1 as shown in figure 3.

Figure 3: EMG signals, First Step

So, by analyzing values of some we conclude that when we take step then the resultant values are greater than 1 so when the receiver side esp32 receives value greater than 1 then we will give stimulus to the paralyzed leg and the patient will walk like a normal person.

These reading are obtained through gait cycle phases i.e. Mid-stance.

A. Gait cycle:

Normal gait consists of two phases:

i. Stance Phase: The stance phase occupies 60% of the total gait cycle, during which some part of the foot is in contact with the ground. It is further divided into five sub-phases:

1. Initial contact (heel strike)
2. Loading response (foot flat)
3. Mid-stance
4. Terminal stance (heel off)
5. Pre-swing (toe off)

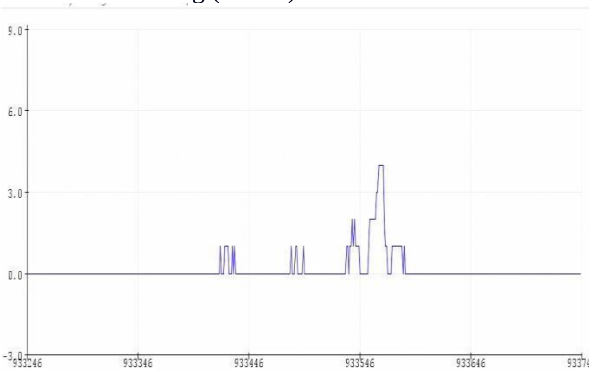


Figure 4: EMG signals, Second Step

ii. Swing Phase: The swing phase occupies 40% of the total gait cycle, during which the foot is not in contact with the ground and the bodyweight is borne by the other leg and foot. It is further divided into three sub-phases:

1. Initial swing
2. Mid-swing
3. Late swing

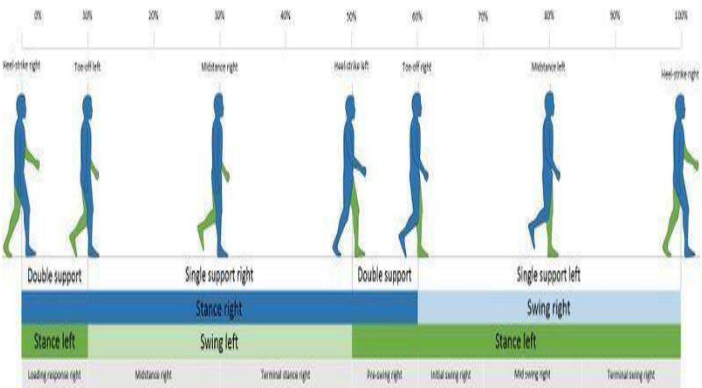


Figure 5: Breakdown of Gait Cycle Sub Phases

B. Circuit Diagram:

This project uses two ESP32 microcontrollers—one as a transmitter and the other as a receiver—to wirelessly transmit EMG signals from a healthy leg to an affected leg. The transmitter collects EMG signals from the healthy leg using an EMG sensor, processes the data, and sends it via Wi-Fi. The receiver then interprets these signals and triggers appropriate muscle stimulation or movement in the affected leg, assisting rehabilitation and mobility. As shown in figure 6.

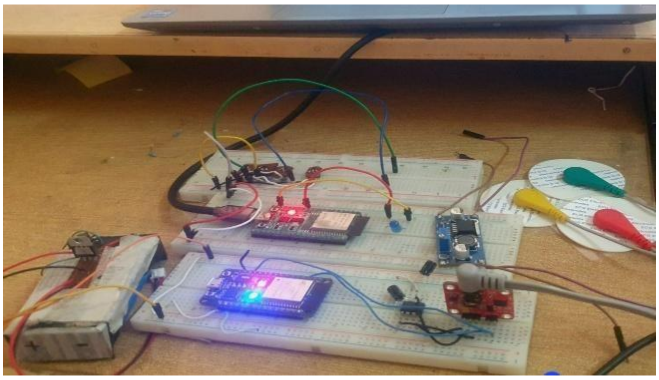


Figure 6: Circuit diagram

V. CONCLUSION

The conclusion of the MSATS study offers a set of impressive points that summarize the importance, potential contribution, and future dimension of the system. A sequence of topics is addressed in this complete overview without individual headings: The critical importance of Amplistride's use of electromyography (EMG) signals in the treatment of limb paralysis is summarized in the first sentence of the conclusion. It highlights the accuracy realized using EMG signals, the master key to Amplistride's success in enabling the intermediary communication between the user's will and the system. The investigation then proceeds to the potentially groundbreaking effect of Amplistride on recovery from limb paralysis. This includes the discussion of how new technologies combined with customized strategies can greatly improve the lives of people who are coping with paralysis of the legs.

The three objectives of increased mobility, increased independence, and increased overall well-being are emphasized [17].

The conclusion emphasizes the need for long-term innovation and cooperation and the acknowledgment of the dynamic nature of the field. It reasserts the significance of ongoing research and development of prospects in EMG-based technology. Synergies between researchers, engineers, and physicians are key in maintaining these innovations in the field of rehabilitation.

This promise to make an effective contribution in patient care is also upheld in the conclusion, which acknowledges the dynamic character of EMG-based technology. It is a call to researchers and practitioners to push further the envelopes of what EMG-based technology can achieve.

Key concepts can be solidified at the conclusion with the seamless incorporation of visual elements such as graphs and photographs. These photographs can show tangible outcomes, such as increases in joint range of motion, muscle strength, or other relevant parameters. The final thoughts could inspire more study, creation, and application of EMG-based technology. There is room for continued innovation and growth in the field of rehabilitation as future directions for research or possible technological advances are hinted at. This essay serves as an example of how important simulation settings are for researching medical problems that affect people's daily lives. Physiotherapy and medicinal interventions have been used to address foot drop issues in a number of ways. This work presents a proposed model of the robotic foot that may mimic and simulate the foot-drop case, taking into account its key characteristics. Additionally, this model allowed for the examination of various foot-drop situations without requiring direct interaction with actual patients [18].

REFERENCES

1. Knutson, J. S., Hansen, A. H., & Clark, T. W. (2012). "Foot and ankle electromyography during gait with partial functional electrical stimulation." *Journal of Biomechanics*, 45(2), 298-305.
2. Sabut, S. K., Sikdar, C., Kumar, R., & Mahadevappa, M. (2010). "Functional electrical stimulation of dorsiflexor muscle: effects on dorsiflexor strength, plantarflexor spasticity, and motor recovery in stroke patients." *NeuroRehabilitation*, 27(4), 331-339.
3. Kesar, T. M., Perumal, R., Jancosko, A., Reisman, D. S., Rudolph, K. S., Higginson, J. S., & BinderMacleod, S. A. (2011). "Novel patterns of functional electrical stimulation have an immediate effect on dorsiflexor muscle function during gait for people poststroke." *Physical Therapy*, 91(12), 17111721.
4. Everaert, D. G., Thompson, A. K., Chong, S. L., & Stein, R. B. (2010). "Does functional electrical stimulation for foot drop strengthen corticospinal connections?" *Neurorehabilitation and Neural Repair*, 24(2), 168-177.
5. Bethoux, F., Rogers, H. L., Nolan, K. J., Abrams, G. M., Annaswamy, T. M., Brandstater, M., ... & Fredrickson, E. (2014). "The effects of peroneal nerve functional electrical stimulation versus anklefoot orthosis in patients with chronic stroke: a randomized controlled trial." *Neurorehabilitation and Neural Repair*, 28(7), 688-697.
6. Thrasher, T. A., Popovic, M. R., & Collins, D. F. (2005). "A comparison of different functional electrical stimulation paradigms on the activation of dorsiflexor muscles during gait in individuals with drop-foot." *Neurorehabilitation and Neural Repair*, 19(4), 377-385.
7. Kottink, A. I., Tenniglo, M. J., de Vries, W. H., Hermens, H. J., & Buurke, J. H. (2012). "Effects of an implantable two-channel peroneal nerve stimulator versus conventional walking device on spatiotemporal parameters and kinematics of hemi paretic gait." *Journal of Rehabilitation Medicine*, 44(1), 51-57.
8. https://assets.newatlas.com/dims4/default/8bd3b6c/2147483647/strip/true/crop/921x614+0+233/resize/1200x800!/quality/90/?url=http%3A%2F%2Fnewatlas-brightspot.s3.amazonaws.com%2Farchive%2FReggie_Edgerton_patient_photo_moving_legs_no_caption2016.jpg
9. https://media.springernature.com/full/springerstatic/image/art%3A10.1038%2Fs41467-018-05282-6/MediaObjects/41467_2018_5282_Fig1_HTML.png
10. https://imageio.forbes.com/blogsimages/jvchamary/files/2016/08/walkagain31200x800.jpg?height=474&width=711&fit=bounds https://www.google.com/url?sa=i&url=https%3A%2F%2Fwww.cbsnews.com%2Fnews%2Fbrai ntraining-helps-8-paralyzed-people-regain-movement%2F&psig=AOvVaw2XLRIUyr5S3q6syjtA4H8p&ust=1703354446050000&source=images&cd=vfe&opi=89978449&ved=0CBIQjRxqFwoTCOD9pfrPo4MDFQAAAAAdAAAAABAD
11. Chou, L. W., Palmer, J. A., Binder-Macleod, S., & Knight, C. A. (2013). "Motor unit rate coding is severely impaired during forceful and fast muscular contractions in individuals post stroke." *Journal of Neurophysiology*, 109(12), 2947-2954
12. Hara, Y., Obayashi, S., Tsujiuchi, N., & Muraoka, Y. (2013). "A Feasibility Study of a Wireless Electromyography Data Acquisition System for Home Rehabilitation." *Biomedical Engineering Online*, 12(1), 56.
13. Farina, D., & Aszmann, O. C. (2014). "Bionic Limbs: Clinical Reality and Academic Promises." *Science Translational Medicine*, 6(257), 257-261.
14. Popovic, M. R., Curt, A., Keller, T., & Dietz, V. (2001). "Functional Electrical Stimulation for Grasping and Walking: Indications and Limitations." *Spinal Cord*, 39(8), 403-412.
15. Micera, S., Crea, S., & Vitiello, N. (2015). "Wearable Technologies for Neurorehabilitation." *Handbook of Neuroengineering*, 765-790.
16. Lopes, P., & Baud-Bovy, G. (2017). "Virtual Reality for Rehabilitation: Techniques and Applications." *IntechOpen*.

17. Dobkin, B. H. (2004). "Strategies for Stroke Rehabilitation." *Lancet Neurology*, 3(9), 528-536.
18. Vujaklija, I., Farina, D., & Aszmann, O. C. (2016). "New Trends in Prosthetics: A Journey from Amputation to Bionic Limbs." *IEEE Signal Processing Magazine*, 33(5), 28-39. 19. https://www.physio-pedia.com/File:Gait_cycle.jpg

Enhancing Driver Identification with a Crow Search-Optimized Stacking Ensemble

Anwar Mehmood Sohail
Department of Computer System Engineering
University of Engineering and Technology
Peshawar, Pakistan
anwarmehmood.sohail@gmail.com

Khurram Shehzad Khattak
Department of Computer System Engineering
University of Engineering and Technology
Peshawar, Pakistan
khurram.s.khattak@gmail.com

Zawar Hussain Khan
College of Computer Science and Engineering
University of Hail
Hail, Saudi Arabia
z.hussain@uoh.edu.sa

Ahmad Mustafa
Department of Computer System Engineering
University of Engineering and Technology
Peshawar, Pakistan
cadet.mustafa2011@gamil.com

***Abstract*— Driver identification systems are incredibly vital when it comes to enhancing vehicle security and providing personalized experiences for drivers. Conventional methods of driver identification rely on individual machine learning models that cannot sufficiently perform due to their inability to generalize across diverse driving behaviours. In the present study, we propose a novel stacking ensemble framework optimized with the Crow Search Algorithm (CSA), addressing the aforementioned limitations. The CSA-optimized stacking ensemble integrates the predictivity of various base models, including Logistic Regression (LR), Naïve Bayes (NB), Random Forests (RF), and K-Nearest Neighbour (KNN), with a meta-learner aimed at improving accuracy and robustness. The CSA optimizes the hyperparameters of the ensemble to guarantee maximal performance. Experimental evaluations on driving dataset showed that identification accuracy, precision, and recall efficiency were significantly improved compared to other methods. The proposed framework could be used in several applications including intelligent transportation systems and automotive cybersecurity.**

Keywords—crow search algorithm, driver identification, stacking optimization, stacking ensemble, intelligent transportation system

I. INTRODUCTION

Driver identification is referred to as driver fingerprinting, wherein the driver identity is established of individual driving the vehicle [1, 2]. It aims at recognizing or verifying the identity of the driver based on classification or identification of the driver's driving behavior. Behavioral human beings exhibit a wide variety that manifests differently when driving and propels driving styles as individualistic digital fingerprints. These driving patterns are generally affected by broader characteristics such as age, gender, education, experience, and so on [1]. Drivers differ in their maximum and minimum speeds on the same stretch of road. Some may have their foot constantly on the gas or brake pedal from dealing with heavy traffic or some minor bumps on the road, while others may not. It is such distinct driving behaviors that would help identify a driver. This area of research investigates specific driving behavior, and prosperously advanced due to the multitude of sensors and the expanding data being generated by present-day vehicles.

A. Research Context

Driver identification is important in many ways. The first aspect relates to safety; strong driver identification systems prevent unauthorized or indiscreet access to the vehicle, diminishing any risk of accidents [3]. Secondly, the driving experience can be personalized by adjusting the driver's environment and vehicle settings according to each driver's driving style [4]. Moreover, driver identification systems are vital for security applications such as anti-theft, access and verification, which are additional layers of security for the vehicle [2]. Last but not least, driver identification assists in the technological improvement of intelligent transportation systems (ITS).

Driver identification is very challenging and complicated task which has become topical in the past few years [1]. Traditional techniques using human effort have been mostly visual inspection of identification documents and verification by a driver's license, which are tedious and prone to false results. Advanced technologies have definitely thrown up machine learning and data-driven approaches as techniques that can do a lot towards solving the problem of driver identification [1]. All such techniques will extract additional value from large volumes of data in vehicles' sensors to create much better-performing and much more accurate driver-identification systems [5, 6].

Conventional techniques to identify a driver depend on behavioral data acquired from sensors embedded in an automobile. This data usually has information like how much an individual speeds up, reduces pressure on the brake pedal, the angle used for steering control, and how well does one maintain in their lane. Each of those variables was determined by various factors such as driving experience, mood, and road conditions, thus making accurate identification a tricky task. Conventional techniques for handling such aspects include Machine learning techniques such as Support Vector Machines (SVM) and Decision Trees (DT). But these single models, in general, fail to generalize on diverse driving behaviors especially when the data is noisy or contradictory.

The ensemble learning methods have arisen as a strong solution to overcome limitations posed by single-model learners. Among the most attractive methods are stacking ensembles, which mix the outputs of several models by a meta-learner. This hierarchical approach allows for better generalization and robustness. Nevertheless, the efficacy of stacking ensembles largely hinges upon optimum model selection and hyperparameter tuning. This is computationally expensive and can lead to suboptimal solutions if done manually or using traditional grid search techniques.

B. Scope and Significance

In this work, we propose stacking ensemble algorithm which is optimized with Crow Search Algorithm (CSA) for overcoming these challenges. The CSA mimics the intelligent foraging behavior of crows and is well suited for optimizing complex problems. By maintaining a balance between exploration and exploitation, CSA efficiently traverses the hyperparameter space for configurations that maximize predictive performance of the ensemble model. This study has demonstrated the validity of the proposed framework for improving driver identification accuracy, thus in principle allowing it to be integrated into real-world ITS applications.

This research contributes mainly to the following:

- Development of a Recursive Feature Elimination (RFE) approach that balances dimensionality reduction with classification accuracy.
- Establishing a new mechanism for dynamic model selection via CSA for the automated optimization of stacking ensemble configurations.
- Designing a novel stacking ensemble framework with optimized base models, aiming for both high accuracy and scalable implementations for driver identification tasks.
- Verification of the effectiveness of this framework through extensive experimentation.

The paper initiates its discussion by reviewing techniques for driver identification in Section II. Proposed methodology is presented in Section III, followed by an elaborate discussion concerning the experimental setup in section IV. The last section of the research will be devoted to significant findings and future prospects for investigation in Section V.

II. Related Work

The problem of identifying drivers has garnered great attention from researchers, who come up with numerous solutions, from the traditional to contemporary machine learning techniques. This section discusses some of the major contributions in the areas of single-model approaches, ensemble learning techniques, and optimization algorithms.

A. Single Model Approaches

Driver identification research has traditionally focused on single-model strategies. The earlier approaches used SVM, DT, and k-nearest neighbours (KNN) algorithms to classify drivers based on behavioral data [7]. As presented in [7], a novel driver identification solution is proposed that employs machine learning algorithms to analyze data from a 3-axis accelerometer. The paper illustrated the importance of raw sensor signal feature extraction for identification accuracy and employed a data transformation algorithm based on some existing methods. The performance of four basic classification algorithms—KNN, random forests (RFs), multilayer perceptron (MLP), and Adaboost—was evaluated in conjunction with a multi-model ensemble algorithm for performance enhancement. The results showed that the RFs algorithm provided the best performance among all in terms of identification accuracy and computational speed. The paper also discussed the need for data calibration to reduce biases introduced during sensor installation.

With dynamic human behavior models enhancing vehicle security and preventing theft, the research in [8] presented a driver identification system. The authors propose the methodology to capture and analyze driving performance data with driving features pertaining to steering, acceleration, and braking that lie at the core of individual driving styles. SVM classifiers were used to distinguish different drivers in this work, reaching over 85% success in identifying valid ones.

Various data preprocessing methods were compared in the study, concluding that Fast Fourier Transform (FFT) performed better than Principal Component Analysis (PCA) and Independent Component Analysis (ICA) for processing human dynamic behavior. The system was designed to be easily integrated into vehicles while requiring minimal additional hardware. Therefore, the driver was not distracted by any security measures. All in all, this research work has demonstrated the feasibility of using behavioral biometrics for real-time vehicle security applications and provided a highly effective deterrent to vehicle theft.

The research paper [9] presented the new method of drunk driving detection with RF algorithm-based feature selection. Data collected from a driving simulator helped identify critical driving behavior features, namely accelerator depth, speed, and distance to the center of the lane, which were decisive parameters for accurate detection. The road geometric characteristics were modeled as a dummy variable so the classifier could adapt to varied road conditions without building multiple models. RF classifier performance was compared against Linear Discriminant Analysis (LDA), SVM, and AdaBoost models, and RF and AdaBoost were found to have equal performance with 81.48% accuracy. Results indicated that accelerator depth feature played an outsized role in improving model's accuracy as compared to other features. The ultimate conclusion was that the proposed method consisting of an efficient integration of feature selection with geometric encoding improves drunk driving detection, which in turn is expected to reduce alcohol-impairment traffic accidents.

The SVM is widely applied because of its strength in dealing with high-dimensional data and investigating non-linear relations. However, the generalization of such models over unseen data is often poor, especially where driving behavior may have great diversity or where a lot of noise is present in the input. Neural networks, especially the multi-layer perceptron, have also been applied and thus proved their capability of learning complex patterns, yet overfitting and high computational costs in training on limited data affect their performance.

B. Ensemble Learning Techniques

Stacking ensembles represent a more advanced type of ensemble method that combines the outputs of different base models through a meta-learner. This hierarchical solution provides flexibility and generalization abilities. Stacking proves to be the best in driver identification from previous studies, but the absence of systematic tuning of its parameters and architecture most often hinders its power.

In [10], Lane-changing Intention (LCI) recognition modeling in a connected environment was illustrated by a driving simulator. Key differences between driver behavior in connected versus non-connected environments were analyzed, showing that LCI times are longer in the connected scenarios. The authors employed a stacking ensemble learning framework integrating RF, SVM, LSTM, and Bi-LSTM algorithms for LCI prediction, achieving an accuracy rate of 98.24%.

The authors of [11] developed a framework that exploits the realistic driving dataset for recognizing lane change maneuvers (LCMs) of preceding vehicles. In a stacking ensemble framework, performance levels of up to 98.25% accuracy were achieved using the combination of RF, SVM, LSTM, and attention-based Bi-LSTM classifiers. The accuracy of the model's predictions has greatly increased; the response time has accordingly been significantly reduced, which is particularly significant in ensuring safety in advanced driver assistance systems and in autonomous driving.

Ensemble methods emerged as a powerful alternative to single models. The use of Bagging, Boosting, and RF methods combines predictions from various base models to improve accuracy and reduce variance. For example, RF works on the diversity of DT for a significant gain in performance. Boosting algorithms such as AdaBoost and Gradient Boosting Machines (GBM) sequentially train weak learners by focusing on their misclassifications, thereby improving the predictive power of the model. These methods, however, do suffer from a few shortcomings due to their dependence on model-specific structures and the opportunity for overfitting when applied to highly complex datasets.

C. Optimization Algorithms

Optimization algorithms have played significant roles in improving machine learning models' effectiveness by the automation of hyperparameter tuning and feature selection. Evolutionary algorithms, including Genetic Algorithms (GA) and Particle Swarm Optimization (PSO), have found widespread use in this area. While the GA optimizes model parameters by iteratively evolving the solutions using natural selection principles, PSO uses social behavior of birds for efficient navigation through search space.

The paper [12] provided a new methodology to optimize stacking ensemble methods with Ant Colony Optimization (ACO). The stacking ensemble combined the predictions from multiple classifiers, and the problem of selecting an optimal configuration had its challenges. ACO-Stacking constructed the configurations of the ensemble by iteratively building the base classifier and choosing the meta classifiers by employing the ACO mechanism. The experimental results generated over 18 datasets suggest that ACO-Stacking provided better results than most of the ensemble techniques studied, namely Bagging, AdaBoost, RF, Stacking, and GA-Ensemble. Ten different learning algorithms were taken as base classifiers; the method showed improvements in accuracy across multiple datasets.

In paper [13], two novel methods were proposed to optimize stacking ensemble configurations using the Artificial Bee Colony (ABC) algorithm. The authors introduced two types of ABC-based stacking: ABC-Stacking1, where bees optimize the base-level classifiers with the meta-classifier fixed, and ABC-Stacking2, where optimization of both base- and meta-classifiers occurred simultaneously. They tested their approach on ten different benchmark datasets, that had been demonstrated with an improved predictive capability over existing ensemble techniques: Bagging, Boosting, RF, and Stacking. ABC-Stacking exhibited performances comparable to or better than other optimization-based methods reported in the literature, namely GA-Stacking and ACO-Stacking. The authors concluded that ABC-Stacking is an efficient classifier configuration selection scheme that can be further refined and applied to complex datasets.

In-depth and comparative analysis of metaheuristic algorithms such as GA, ACO, and PSO for optimizing stacking ensemble configurations was provided by [14]. The different algorithms approach the problem of classifier selection and parameter optimization from different perspectives. The experimental results on different datasets found GA-Stacking to be the most accurate, while ACO-Stacking was the most flexible and efficient. The authors suggested that the improvement of PSO-based stacking would be a potential future path. The authors concluded that metaheuristic selection of stacking ensembles could greatly aid in enhancing predictive accuracy while automating configuration selection.

The research paper [16] by Priyadharshini and Ukrit is focused on improving road safety using driving style classification. Collision rates are largely dependent on driver awareness, and hence, the authors have proposed a methodology that includes feature reconstruction through clustering and stacking classifiers whereby the classifier is optimized by ABC algorithm. They used data from On-Board Diagnostics (OBD) to eliminate manual data labeling. Accuracy of the proposed stacking classifier exceeded 98%, which showed a clear advantage over various traditional machine learning methods. The significance lies in understanding driving behavior in ITS applications such as advanced driving assistance systems (ADAS), which form a huge part of this paper. With this, the paper also provides interesting insights into ensemble learning applications for real-time driving style appraisal.

The recent attraction of nature-inspired algorithms, such as CSA, can be attributed to their balanced exploration and exploitation. The CSA imitates the intelligent behavior of crows foraging for food within their environment; hence it has shown good performance in solving complex optimization problems from different areas, including feature subset selection and parameter tuning. However, CSA has not been applied much to ensemble learning, which is the gap addressed by this study.

In conclusion, while major advances in real-time driver identification have been made through machine learning and ensemble methods, existing works often fail in terms of generalization and robustness. Building upon these findings, this study puts forth a CSA-optimized stacking ensemble framework, offering a novel solution to overcome these limitations.

III. METHODOLOGY

This section outlines the methodology employed for driver identification based on driving behavior analysis, with an optimized stacking ensemble model using the CSA for base model selection as shown in Fig. 1.

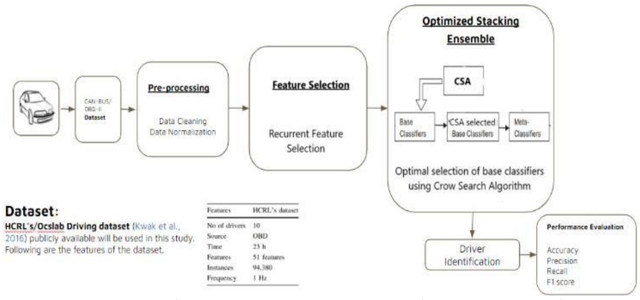


Fig. 1. Overview of the Proposed Methodology for Driver Identification

The ensembles are built on these models. We selected RF, KNN, SVM, Logistic Regression (LR), XGBoost, and Naïve Bayes (NB) due to their complementary strengths in working with structured data. Each of the base models has been trained in isolation on the dataset so as to learn different patterns in driver behavior. A LR model is chosen as the meta-learner and is utilized to aggregate the predictions of the base models into the final output. The meta-learner thus serves as a decision-making layer that strengthens the individual weaknesses of base models. The hyperparameters of the base models are optimized using CSA. Optimization, with the least burden of computation, provides the ensemble an option to attain the best possible performance. The proposed system is illustrated in Fig. 1.

A. Data Description

Driver identification is mainly dependent on behavioral information that is collected by using an instrumented vehicle consisting of several sensors like accelerometers, gyroscopes, GPS, and onboard diagnostics (OBD). These capture important driving patterns such as acceleration, braking, steering angle, and speed. To strengthen the system robustness, information from diverse driving styles is collected using many drivers. Preprocessing is done on the collected data by removing noise and outliers that affect data quality. The present study uses behavior driving dataset from Hacking and Countermeasures Research Lab (HCRL) [17]. This dataset has 51 features obtained from in-vehicle CAN-bus data and OBD-II. This set contains the driving data of about ten drivers moving on three different given types of roads in Seoul such as highway and parking lots and amounting to a total distance of 23 kilometers.

B. Data Preprocessing

Ensuring the quality and reliability of a dataset used for driver identification requires the data-preprocessing step. Typically, it entails several tasks during preprocessing:

1) Handling Missing Values: Missing data points can occur due to sensor malfunctions or data transmission issues. Techniques such as interpolation or imputation are used to fill missing values.
2) Normalization: Normalization is used to standardize values of all variables to a common range, improving the performance of the machine learning model, since this is relevant due to the differing scales of the features.

Preprocessing means cleaning the data, establishing relevant consistency, and defining what kinds of features will ultimately be extracted for modeling.

C. Feature Selection

In the quest of improving accuracy and efficiency in classification using the stacking ensemble model, feature selection has been performed. The idea is to minimize the size of the feature set by selecting relevant features while reducing redundancy.

A heatmap serves effectively as a visualization tool for studying relationships and correlations of the selected features in the dataset. Once the heatmap has been interpreted, it will be possible to see redundant features or those that are highly correlated and can confer multicollinearity that will interfere with the stability and interpretability of the model. The heatmap would thus be used as demonstrated in Fig 2. to supplement feature importance and multicollinearity assessment. The use of a heatmap makes it easy to see the features that show high correlation, thus ensuring that the selection of independent variables contributes positively to model performance

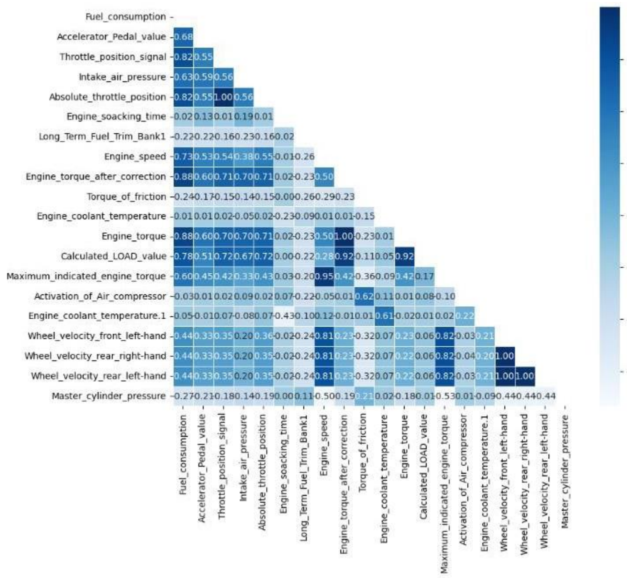


Fig. 2. Correlation Heatmap of Selected Features

Recursive Feature Elimination is the basic technique for feature selection in this study. RFE iteratively removes the least important features and constructs a model with the remaining subset of features. Feature importance can be assessed based on their contribution to the predictive performance of any machine-learning model (SVM or RF). The process of RFE involves the following steps:

Step 1: An initial model is trained using the complete feature set.

Step 2: Features are ranked based on their importance scores, which can be derived from model coefficients or feature importance metrics.

Step 3: The least important features are removed, and the model is retrained on the reduced feature set.

Step 4: Steps 2 and 3 are repeated iteratively until optimal subset based on performance metrics such as accuracy and feature importance scores are selected.

RFE helps in identifying the most relevant features that contribute significantly to driver identification, thereby reducing model complexity and improving generalization.

The following 20 features were selected by the RFE method:

- Engine_soaking_time
- Fuel_consumption
- Accelerator_Pedal_value
- Throttle_position_signal
- Intake_air_pressure
- Absolute_throttle_position
- Engine_soacking_time
- Long_Term_Fuel_Trim_Bank1
- Engine_speed
- Engine_torque_after_correction
- Torque_of_friction
- Engine_coolant_temperature
- Engine_torque
- Calculated_LOAD_value
- Maximum_indicated_engine_torque
- Activation_of_Air_compressor
- Engine_coolant_temperature
- Wheel_velocity_front_left-hand
- Wheel_velocity_rear_right-hand
- Wheel_velocity_rear_left-hand
- Master_cylinder_pressure

D. Stacking Ensemble Classification

Due to the synthesis of many ensemble methods, stacking is said to be a very powerful technique. It has two levels: base learning and meta-learning. Base learning involves training the selected set of classifiers with the training dataset. The base learners create a dataset for the meta-learners. The selection of base learner for the classification task is the critical part. The basic structure of the stacking model is shown in Fig. 3. The process operates from original data being passed onto level 1 models. The predictions made by different based models of level 1 are then passed on to level 2. The challenge of this setup is the optimum selection of machine learning models to act as the base model. Machine learning classifier selection at base learning is undertaken by using CSA.

The stacking ensemble method harnesses the strengths of various models, thereby enhancing the classification accuracy and robustness of the driver's identification system employing a stacked structure.

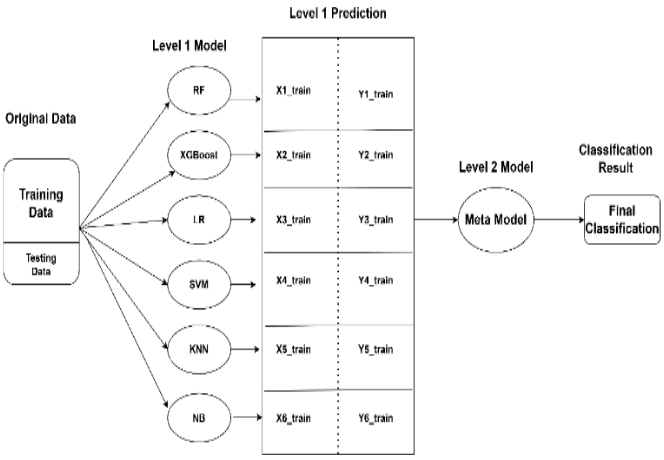


Fig. 3. Architecture of the Stacking Ensemble Model

E. Stacking Optimization Using CSA

The selection of base models becomes a key step in building a working stacking ensemble classifier. Therefore, in this study, CSA is deployed for selecting the combination of base models that provides the best overall performance from a selection of candidate classifiers.

1) CSA

CSA is a population-based metaheuristic optimization algorithm inspired by the intelligent foraging behavior of crows. Developed by Askarzadeh in 2016 [18], it has found its place in a variety of applications ranging from engineering to machine learning and data science. The algorithm considers mainly the cognitive capabilities of crows and their memory-aided foraging behavior and strategies against resources that are prone to being stolen. As for its title, it is a nature-oriented metaheuristic algorithm inspired by the intelligent foraging behavior of crows as depicted in Fig. 4. The algorithm maintains a population of candidate solutions (crows), which it iteratively improves through two main mechanisms:

- **Exploration**: Crows search for new solutions by moving randomly within the search space.
- **Exploitation**: Crows attempt to move toward the best solution found so far, with a probability of deception to prevent being followed by others.

2) CSA Implementation for Base Model Selection

The CSA-based base model selection process involves the following steps:

- Initializing crows swarm in d-dimensional randomly.
- A fitness function is used to evaluate each crow, and its value is put as an initial memory value. Each crow stores its hiding place in its memory variable m_i.

- Crow updates its position by selecting a random another crow, i.e. x_j and generating a random value. if this value is greater than Awareness Probability AP, then crow x_i will follow x_j to know m_j.
- Crow updates its position by selecting a random other crow i.e. x_j and following it to know m_j. Then new x_j is calculated as follows:

$$x_{i,iter+1} = \begin{cases} x_{i,iter} + r_i \times fl_{i,iter} \times (m_{j,iter} - x_{i,iter}) & r_j \geq AP_{j,iter} \\ a\ random\ position & otherwise \end{cases} \quad (1)$$

Where $AP_{j,iter}$ refers to crow j awareness probability, iter refers to iteration number, r_i, r_j refers to random numbers, $fl_{i,iter}$ is the crow i flight length to denote crow j memory.

- Updating memory

$$m_{i,iter+1} = \begin{cases} x_{i,iter} & f(x_{i,iter+1}) \leq f(m_{i,iter}) \\ m_{i,ier} & otherwise \end{cases} \quad (2)$$

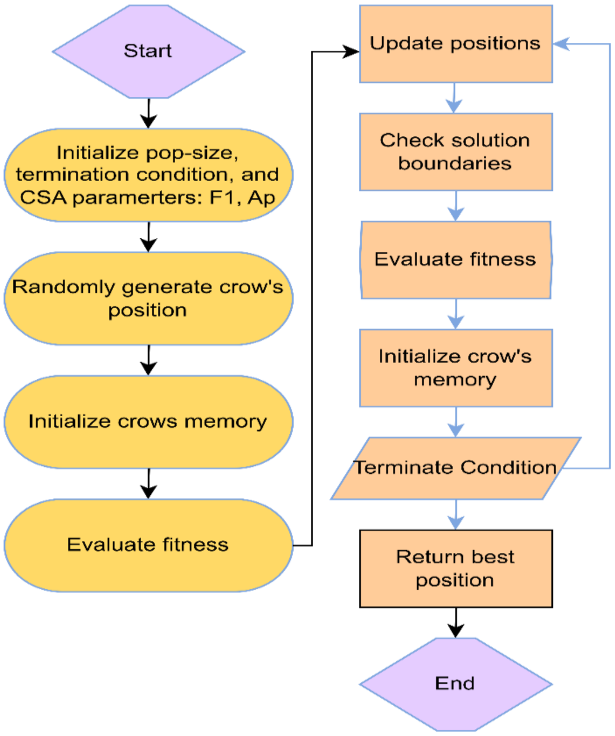


Fig. 4. Crow Search Algorithm Architecture

The following machine learning algorithms were chosen by CSA for the stacking ensemble:

- Random Forest Classifier
- Logistic Regression
- KNN Classifier
- Gaussian Naïve Bayes

Meanwhile, XGBoost and SVM were excluded as can be seen in Fig. 5.

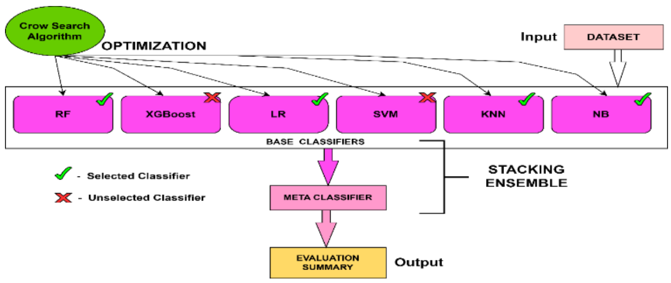


Fig. 5.Stacking Ensemble Framework with CSA for Classifier Selection

By applying CSA, the optimal combination of base models is selected, leading to an improved stacking ensemble classifier with higher predictive performance.

The performance of the proposed driver identification system is evaluated using standard metrics, including:

- **Accuracy**: The ratio of correctly identified drivers to the total number of test samples.
- **Precision**: The ratio of true positives to the sum of true positives and false positives.
- **Recall**: The ratio of true positives to the sum of true positives and false negatives.
- **F1-score**: The harmonic means of precision and recall, providing a balanced measure of the classifier's performance.

IV. Results

We carried out all our experiments using Google Colab. The framework has been implemented in Python using Scikit-learn and other renowned machine-learning packages. The training and optimization procedures were made faster by using parallel processing, so as to make fair use of compute resources. The design of the experimental settings was done keenly in order to achieve a realistic evaluation of the suggested framework, thus justly validating its merits. The proposed method therefore achieved the following results:

- **Accuracy**: 0.9843
- **Precision**: 0.9843
- **Recall**: 0.9843
- **F1 Score**: 0.9843

Six machine learning models have been compared closely in Table 1. The evaluation criteria include Accuracy, Precision, Recall, and F1 Score. The Optimized Stacking Model is seen to be the best performing model according to all measures, with scores of 98.43% to suggest near-perfect capability in classification and an equally high ability to discriminate both positive and negative cases with efficacy. Stacking Ensemble also runs strong with an accuracy of 95.83% and comparable Precision, Recall, and F1 Score. This model integrates multiple base models' predictions, showcasing the strong efficacy of ensemble methods in improving classification accuracy. Random Forest and K-Nearest Neighbor show moderate to good performance, with accuracy values around 90.69% and 85.73%, respectively. Their Precision, Recall, and F1 Scores are also reasonably high, indicating a balanced performance in terms of correctly identifying both positive and negative cases.

TABLE I COMPARATIVE PERFORMANCE OF MACHINE LEARNING MODELS BASED ON CLASSIFICATION METRICS

Metric	Random Forest	Logistic Regression	K-Nearest Neighbor	Naive Bayes	Stacking Ensemble	Optimized Stacking Model
Accuracy (%)	90.69	57.39	85.73	31.99	95.83	98.43
Precision	0.9106	0.5745	0.8570	0.4048	0.9582	0.9843
Recall	0.9069	0.5739	0.8573	0.3199	0.9583	0.9843
F1 Score	0.9062	0.5669	0.8569	0.3015	0.9582	0.9843
Training Accuracy	0.9112	0.5762	0.9148	0.3196	0.9755	-
Test Accuracy	0.9069	0.5739	0.8573	0.3199	0.9583	-

Logistic Regression demonstrates relatively lower performance, with an accuracy of 57.39%. This suggests that the model may not be well-suited for the given classification task, as it struggles to correctly classify instances. Naive Bayes exhibits the lowest performance among all models, with an accuracy of 31.99%. This indicates that the model's assumptions may not align well with the data, leading to poor classification accuracy. The Precision, Recall, and F1 Score are also low, further confirming its limited effectiveness.

In order to reduce the possibility of overfitting, particularly in view of metrics that seem near perfect, several strategies were enacted during the course of the experiments. First, we implemented k-fold cross-validation with $k = 5$: The dataset is partitioned into k equally sized folds, k -1 of which are used for training and the remaining fold for validation in each iteration. Thus, every data point was used for training and validation in turn, presenting an overall view of the model's performance. Another strategy was to ensure that the dataset was balanced to avoid bias towards the majority class for improved generalization across instances. Great care was taken to avoid data leakage in those scenarios, fitting transformations like normalization and encoding only on training data and applying them independently to the validation and testing set. These combined strategies ensured minimal risk for overfitting, showing consistent performances throughout cross-validation folds as well as a separate hold-out test set.

V. CONCLUSION

This study presents a CSA-optimized stacking ensemble framework for driver identification, demonstrating significant performance improvements over traditional methods. By leveraging the complementary strengths of ensemble learning and CSA-based optimization, the framework addresses key challenges in driver identification. Previous studies, such as [10] and [11], demonstrated the advantages of integrating multiple classifiers, including RF, SVM, LSTM, and Bi-LSTM, achieving high accuracy levels of 98.24% and 98.25%, respectively. In this study, the proposed framework further improved predictive performance, achieving an accuracy of 98.43%, surpassing the results of existing models. While the current study focuses on validating the framework using a controlled dataset, its potential for real-world ITS applications is significant. Future work will involve deploying the framework in real-world fleet environments to assess its adaptability and robustness in diverse conditions. This deployment will enable us to evaluate the framework's ability to handle unseen drivers, varying environmental conditions (e.g., weather, traffic patterns), and different vehicle types. Additionally, integrating the framework with dynamic datasets collected from connected vehicles and smart infrastructure will help enhance its scalability and real-time applicability. Another potential direction is incorporating edge computing techniques to enable on-device processing, ensuring low latency and efficient operation in real-time ITS systems. These efforts will contribute to a more comprehensive evaluation of the framework's performance and its practical usability in large-scale ITS deployments.

REFERENCES

[1] Sohail, A.M., et al., *A Comprehensive Review: Analysis of Machine Learning, Deep Learning, and Large Language Model Techniques for Revolutionizing Driver Identification.* Authorea Preprints, 2024.

[2] Tseng, P.-Y., P.-C. Lin, and E. Kristianto, *Vehicle theft detection by generative adversarial networks on driving behavior.* Engineering Applications of Artificial Intelligence, 2023. **117**: p. 105571.

[3] Sánchez, S.H., R.F. Pozo, and L.A.H. Gómez, *Driver identification and verification from smartphone accelerometers using deep neural networks.* IEEE Transactions on Intelligent Transportation Systems, 2020. **23**(1): p. 97-109.

[4] Balan, G., et al., *An Improved Deep Learning-Based Technique for Driver Detection and Driver Assistance in Electric Vehicles with Better Performance.* International Transactions on Electrical Energy Systems, 2022. **2022**(1): p. 8548172.

[5] Khattak, K.S., et al., *Sensing technologies for traffic flow characterization: From heterogeneous traffic perspective.* Journal of Applied Engineering Science, 2022. **20**(1): p. 29-40.

[6] Zeb, A., et al., *On-board diagnostic (OBD-II) based cyber physical system for road bottlenecks detection.* J. Eng. Sci. Technol, 2022. **17**(2): p. 906-922.

[7] Li, Z., et al., *Driver identification in intelligent vehicle systems using machine learning algorithms.* IET Intelligent Transport Systems, 2019. **13**(1): p. 40-47.

[8] Qian, H., et al., *Support Vector Machine for Behavior-Based Driver Identification System.* Journal of Robotics, 2010. **2010**(1): p. 397865.

[9] Li, Z., et al., *Random forest–based feature selection and detection method for drunk driving recognition.* International Journal of Distributed Sensor Networks, 2020. **16**(2): p. 1550147720905234.

[10] Zhang, H., S. Gao, and Y. Guo, *Driver lane-changing intention recognition based on stacking ensemble learning in the connected environment: A driving simulator study.* IEEE Transactions on Intelligent Transportation Systems, 2023.

[11] Zhang, H., et al., *Stacking-based ensemble learning method for the recognition of the preceding vehicle lane-changing manoeuvre: A naturalistic driving study on the highway.* IET Intelligent Transport Systems, 2022. **16**(4): p. 489-503.

[12] Chen, Y. and M.L. Wong. *Optimizing stacking ensemble by an ant colony optimization approach.* in *Proceedings of the 13th annual conference companion on Genetic and evolutionary computation*. 2011.

[13] Shunmugapriya, P. and S. Kanmani, *Optimization of stacking ensemble configurations through artificial bee colony algorithm.* Swarm and Evolutionary Computation, 2013. **12**: p. 24-32.

[14] Gupta, A. and A.R. Thakkar. *Optimization of stacking ensemble configuration based on various metahueristic algorithms*. in *2014 IEEE International Advance Computing Conference (IACC)*. 2014. IEEE.

[15] Onan, A., *Particle swarm optimization based stacking method with an application to text classification.* Academic Platform-Journal of Engineering and Science, 2018. **6**(2): p. 134-141.

[16] Priyadharshini, G. and M.F. Ukrit, *RETRACTED ARTICLE: Stacking optimized with artificial bee colony for driving style classification by feature reconstruction from OBD II data.* Soft Computing, 2023. **27**(1): p. 591-603.

[17] Kwak, B.I., J. Woo, and H.K. Kim. *Know your master: Driver profiling-based anti-theft method*. in *2016 14th Annual Conference on Privacy, Security and Trust (PST)*. 2016. IEEE.

[18] Askarzadeh, A., *A novel metaheuristic method for solving constrained engineering optimization problems: crow search algorithm.* Computers & structures, 2016. **169**: p. 1-12.

Enhancing Predictive Business Process Monitoring in Call Centers through Multimodal Data Fusion and Heterogeneous Time-Aware LSTM-Based Multi-Task Learning

Farhat Mehmood1*, Afshan Ishaq2
1Department of computer science, University of Gujrat Hafiz Hayat campus, Gujrat Pakistan
2Department of Electronics, University of Engineering and Technology Abbottabad Campus.

Abstract*— The optimization of call center operations and the enhancement of customer service are greatly supported by predictive business process monitoring. Traditional methods often overlook valuable multimodal data, such as conversations occurring in contact centers, because they typically rely on sequence data from business IT systems. This limitation hinders a complete understanding of business processes. In this study, we introduce a unique time-aware LSTM-based framework for predictive business process monitoring, which leverages both IT system data and dialogue data from contact centers. Our approach combines multiple data sources to improve the accuracy of forecasting ongoing business activities. To address challenges related to multi-task learning and to better utilize the rich information embedded in various data types, we propose a heterogeneous multi-task learning architecture called Heterogeneous Multi-gate Mixture-of-Experts (H-MMoE). Experimental results show that our method outperforms established baseline models such as Transformer, CNN, and standard LSTM. These findings demonstrate the potential of time-aware LSTM models to improve process monitoring, optimize workflows, and drive operational success in call center environments.

***Keywords*—H-MMoE, Deep learning, Optimization, , multitask learning, Call center.**

I. Introduction

In the swiftly advancing customer service sector, contact centers have emerged as essential centers for data acquisition and corporate communication, presenting considerable growth opportunities across several industries [1]. Using cutting-edge technologies including contact centers, databases, communication systems, and computer networks will enable companies to efficiently interact with customers therefore providing timely information. This skill helps companies to react fast to consumer needs, hence improving customer satisfaction, loyalty, and general competitiveness [2]. Call centers create enormous volumes of conversation data during service engagements, frequently including facial expressions and mood signals with great emotional weight. These encounters are intimately related to business results since, for example, a bad attitude of a consumer can influence the possibility of a successful sales transaction. Still, much study is needed to completely grasp and measure how such attitudes affect corporate activities [3].

Business process monitoring and prediction have advanced to the point where ongoing process instances can be analyzed to predict future states. Key prediction tasks include forecasting future activities, process outcomes, remaining time, and business process cycle time. Traditional methods typically rely on process models, such as transition systems, for these predictions. Modern AI methods, particularly deep learning models, have proven to outperform traditional approaches in many predictive contexts [4]. Despite this progress, limited research has focused on the unique role of customer sentiment in contact center settings. Since call center conversation data often holds crucial information for predicting outcomes or time estimates, this gap in research is significant. Moreover, most existing studies use event log sequence data, which may not be sufficient for complex real-world situations, leading to inaccurate predictions. Many of these techniques also only consider single-task predictions and neglect the interdependence among several tasks. Leveraging shared information across related tasks, multi-task learning has promise to increase prediction accuracy [5]. Many multi-task learning methods are limited in that the varied and heterogeneous data produced in practical applications does not always match the presumption of homogeneous feature representations across tasks [6].

This study suggests a new predictive paradigm for tracking company processes in contact center environments in order to solve these difficulties. Incorporating multimodal data inside a heterogeneous multi-task learning framework helps the method increase prediction accuracy and dependability. The suggested method allows thorough analysis by aggregating process sequence data from enterprise information systems with speech data from contact centers. A new architecture termed H-MMoE is presented to solve the difficulties of multi-task learning in different environments. This design gives advantage of the complementing characteristics of several data modalities. The proposed heterogeneous multi-task learning framework, which combines several data modalities, shows a notable improvement in experimental outcomes over conventional benchmarks including models employing a single data modality or isomorphic multi-task learning systems. The results of this study underline the need to include consumer comments into predictive models for corporate operations in call centers. The outcomes underline the possibilities of multimodal data fusion and artificial intelligence-driven approaches to improve operational efficiency, process optimization, and decision-making.

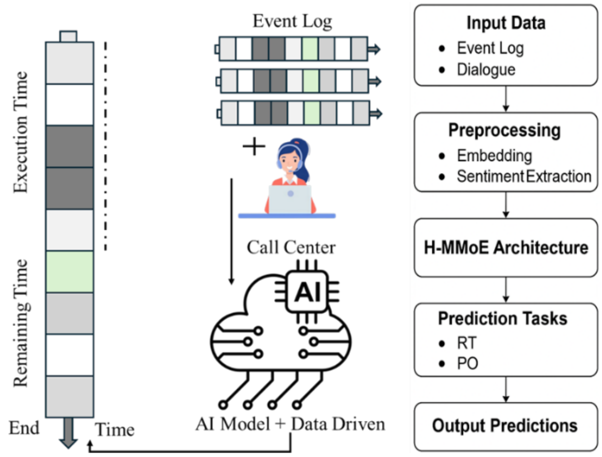


Fig. 1. Structure for business process projection.

To improve predictive business process monitoring, this study integrates multimodal data sources such event sequence logs and customer discussion sentiment information into a heterogeneous multi-task learning architecture. Using uncertainty-based loss weighting, the Heterogeneous Multi-gate Mixture-of-Experts (H-MMoE) model can handle varied feature representations and dynamically balance multiple prediction tasks. This research is new in combining sequential business data with DialogueRNN sentiment analysis to improve process outcome and remaining time predictions. Task-specific contributions are automatically adjusted during training by the dynamic loss weighting approach, improving model adaptability without operator intervention. The proposed approach outperforms single-modal models in predictive business process monitoring, according to extensive trials on real-world call centre datasets. This study highlights the importance of customer sentiments in call centre predictive algorithms. This research aims to show how AI-driven methods and multimodal data fusion can increase operational efficiency, process outcomes, and decision-making. This study uses event log data and sentiment-rich customer chat data in a multi-task learning framework to improve predictive business process monitoring.

The novelty of this research lies in the development of the Heterogeneous Multi-gate Mixture-of-Experts (H-MMoE) model, which integrates heterogeneous expert networks (LSTM, CNN, Transformer) with dynamic loss weighting and sentiment analysis via DialogueRNN. Unlike prior work limited to isomorphic data or static task handling, our approach dynamically adapts to task uncertainty and multimodal input variations, resulting in superior accuracy in both classification and regression tasks. The structure of this article is organized as follows: Section 2 provides a review of the relevant literature, Section 3 details the proposed methodology, Section 4 presents and analyzes the experimental results, and Section 5 concludes the study and outlines potential future directions.

II. Related Work

Recent advancements in business process prediction have been substantial, driven by the increasing complexity of operational data and the growing demand for efficient decision-making tools. Traditional methods often rely on sequential data from organizational information systems, focusing on either continuous variable such as process duration or discrete outcomes like process completion status. However, the integration of multimodal data, such as customer interactions from call centers, has been insufficiently explored, despite its potential to significantly improve prediction accuracy. Moreover, while sentiment analysis has progressed as a technique for interpreting emotional expressions in text, its application in business process forecasting remains limited. Similarly, multi-task learning has shown promise in leveraging common representations for multiple related tasks. However, current approaches typically rely on isomorphic models, which do not support diverse input modalities. This section reviews key developments in multi-task learning, multimodal data processing, sentiment analysis, and business process prediction, highlighting existing solutions and their limitations. It also lays the groundwork for our proposed methodology. An overview of the current literature in this field is presented in Table 1

TABLE I. Related Work

Ref	Dataset	Deployed Model	Contribution	Limitation
[7]	Process instance logs	Annotated Transition System	Predicted remaining process time	Limited to predicting remaining time.
[8]	Process instance logs	Random Petri Nets	Predicted remaining execution time	Focuses only on remaining execution time, lacks multimodal data integration.
[9]	Process instance logs	Queue Model	Predicted remaining time under queuing effect	Considers only queue effects without addressing multimodal information.
[10]	Process instance variables	Non-parametric Regression Model	Predicted remaining time using instance variables	Limited to time prediction; does not handle outcome or activity predictions.
[11]	Process instance logs	Decision Tree, Random Forest	Predicted process outcomes	Predicts outcomes but lacks sentiment and multimodal data analysis.
[12]	Process instance logs	Clustering Method	Outcome prediction using activity sequences	Predicts outcomes based on sequence data but overlooks multimodal integration.

[13]	Process sequence data + attributes	LSTM	Execution instance outcome prediction	Handles sequence data but lacks sentiment and multimodal capability.
[14]	Event sequence data	Hierarchical Attention Mechanisms (LSTM)	Improved time prediction with additional attributes	Focuses on next activity prediction; does not integrate heterogeneous data.
[15]	Hospital process logs	LSTM	Identified key events for next activity prediction	Restricted to hospital scenarios, lacks multimodal capabilities.
[16]	Process sequence data	CNN	Predicted next event in process models	Predicts next activity but neglects time and outcome prediction.
[17]	Process sequence data	Transformer	Predicted next activity	Lacks consideration of multimodal and sentiment-rich data.
[18]	Movie reviews	BiLSTM	Predicted next activity, timestamp, and duration	Restricted sentiment analysis in reviews, not applied to business processes.
[19]	Financial text	TextCNN	Improved next activity and time prediction	Limited to text data; no multimodal fusion.
[20]	Stack Overflow and stock news	BERT	Classified sentiment polarity	Restricted to text data and specific domains; no multimodal applications.
[21]	Dialogue data	DialogueRNN (GRU-based)	Classified sentiment polarity	Focuses only on dialogue sentiment without integrating other data modalities.
[22]	Smart city data	Multimodal (text + visual features)	Identified sentiments in posts	Domain-specific (restaurant/product); limited generalizability to business process prediction.
[23]	Multimodal datasets	Collaborative Learning	Identified sentiment orientation in stock news	Lacks detailed task-specific results and evaluation metrics.
[24]	Business process datasets	Multi-task Learning	Analyzed sentiment in dialogues	Limited to isomorphic features; struggles with heterogeneity in real-world data.
[25]	Mixed datasets	MMoE (Multi-gate Mixture-of-Experts)	Activity and time prediction for multitasking. Balance task relevance and variability	Assumes isomorphic input features, limiting use in heterogeneous multimodal data scenarios.

The recent advancements in predictive business process monitoring are summarized in Table 1 of the reviewed literature. Prominent models in this area include LSTM, CNN, and Transformer, which are utilized for tasks such as estimating remaining time, predicting process outcomes, and determining next actions. While deep learning methods outperform traditional approaches in terms of accuracy, they often fail to incorporate sentiment-rich and multimodal data effectively. Traditional systems, on the other hand, typically rely on sequence data from process logs. Emerging approaches in multimodal learning and sentiment analysis show promise but have yet to be widely adopted in commercial operations. Multi-task learning models like MMoE are particularly effective at handling task interrelationships, but they truly excel when dealing with real-world data heterogeneity. In order to increase prediction accuracy and hence the practical application of these models, this review emphasizes the need of include multimodal data and addressing various job needs.

III. Proposed Methodology

The pseudocode offers a disciplined and orderly framework to properly manage event logs in corporate process monitoring. It improves the contextual awareness of events by adding other factors like activity cost and consumer attitude, therefore transcending conventional definitions. By using this approach, detailed insights into both completed and ongoing processes are gained, with metrics being calculated at both the trace-level and prefix-level. Apart from the regular chores like projecting remaining time and process results, this approach adds fresh prediction activities including tracking sentiment trends and approximating average cost. Leveraging multimodal data guarantees a more complete analysis and helps to enhance forecasting accuracy and decision-making in commercial settings. Emphasizing modularity and scalability, the pseudocode lets future seamless incorporation of new features or prediction goals possible.

Pseudocode for Preliminaries

Here is a refined and readable version of the provided pseudocode for better understanding:

Pseudocode for Business Process Prediction with Multimodal Data

// Input:

1. **Input:**
   - Event log L containing process traces
   - Prediction tasks T = {Remaining Time (RT), Process Outcome (PO)}
2. **Initialize:**
   - Output predictions P = {} **// Store predictions for each task**
3. **For each trace σ in L do:**
   - **// Step 1: Define each event structure**
   - For each event e in trace σ:
     Define e = (act, cid, t, d, cost, sentiment), where:
     • **act**: Executed activity name
     • **cid**: Unique process instance ID
     • **t**: Timestamp of the event
     • **d**: Customer dialogue feature
     • **cost**: Associated cost of the activity
     • **sentiment**: Sentiment score from customer dialogue
   - **// Step 2: Sort events and compute statistics**
   - Sort all events in σ by t
   - Compute:
     • Duration(σ) = t_last - t_first
     • TotalCost(σ) = sum of all event costs in σ
4. **Trace representation:**
   σ = [e1, e2, ..., e|σ|] with |σ| = total events

5. **Group traces into cases:**
   Each case c = (cid, σ)
6. **Build Event Log L**
   L = {σ1, σ2, ..., σ|L|}
7. **Prefix Extraction for Ongoing Cases:**
   For each trace σ, extract prefix hd_k(σ) = [e1, e2, ..., ek] where k ∈ [1, |σ|]
   - **// Step 3: Compute metrics on prefix**
   - For each hd_k(σ) compute:
     • **Remaining Time (RT) = t_last - t_k**
     • **Process Outcome (PO) = last activity in σ**
     • **Average Cost = Σ(cost_e1 to cost_ek) / k**
     • **Sentiment Trend = aggregate sentiment scores of prefix**
8. **Define Targets for Each Prefix:**
   - f_r(σ, k) = RT(hd_k(σ))
   - f_o(σ) = PO(hd_k(σ))
   - f_avgCost(σ, k), f_sentTrend(σ, k)
9. **Apply Prediction Model:**
   - Predict all tasks
   - Append results to P
10. **Return final predictions P**

RT = Remaining Time
PO = Process Outcome
H-MMoE = Heterogeneous Multi-gate Mixture-of-Experts

- The **pseudocode** involves processing the event log LL, extracting key information from each trace, and predicting various aspects such as the remaining time, process outcomes, average cost, and sentiment trends.
- **Multimodal data** such as customer dialogues and sentiments are integrated to improve prediction accuracy.
- The method is modular, allowing for the inclusion of additional metrics or prediction tasks in the future.

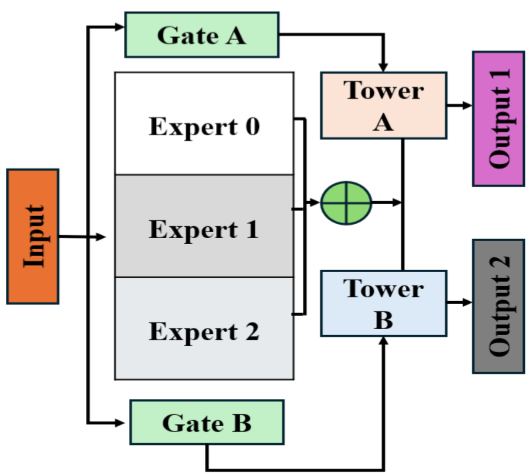


Fig. 2. Frasmework for Multitask learning.

In this section, we will discuss our novel method for predicting business process outcomes and remaining time. Our approach incorporates a variety of components, including an architecture for heterogeneous multi-task learning, data preprocessing, sentiment analysis of customer dialogues, and sequence data encoding. The process begins with data preprocessing and feature encoding, as illustrated in **Figure 3**. The key steps involved in our methodology are outlined below:

1. **Sequence Data Encoding**:
   - We convert sequence data taken from corporate IT systems using an embedding layer. For prediction chores, this approach transforms high-dimensional, sparse features into dense, linked vectors that are simpler to handle.
2. **Sentiment Analysis**:
   - We apply the DialogueRNN model to effectively capture the emotional subtleties of consumer contacts. By analysing the sentiment conveyed in the discussions inside call centre contacts, this approach helps us to grasp consumer underlying feelings.
3. **Heterogeneous Multi-Task Learning**:
   - We next feed our heterogeneous multi-task learning model the encoded features including sentiment analysis results and sequence data. This model lets us simultaneously anticipate process outcomes (considered as a classification problem) and remaining time (formulated as a regression task), hence enabling predictive analysis on several tasks concurrently.

Using both conventional sequence data and the emotional insights from consumer encounters, this integrated technique helps us to reasonably forecast business process results and time estimations.

A. Dataset Description and Pre-processing

Following business process prediction literature, event logs from enterprise systems are prepared during data preparation. This phase creates "prefix logs." to project unfinished process situations. These logs show partial process execution traces. These prefix logs allow the model to forecast process outcomes and durations based on event sequences up to a certain point. Each complete trace is divided into prefixes, each representing a process stage after an event. This method provides dynamic process result forecasting before the entire process is complete A trace has three prefixes after the first, second, and third occurrences. This prefix log, which has three sections, is used to train and evaluate the predictive model. Before creating the prefix log, event data is partitioned into training and testing sets to avoid prefix overlap. This divide ensures a thorough model evaluation and reduces data leakage.
The prefix log is created by processing all event log instances. Events are recalled progressively for each instance, creating incomplete traces. The outcome (final activity) and remaining time (the difference between the current and final event timestamps) are determined for each prefix. Prefixes are systematically compiled into the prefix log for analysis. This careful approach ensures that each sample is evaluated independently, preventing any interference between forecasts and ensuring the integrity of the predictive model.

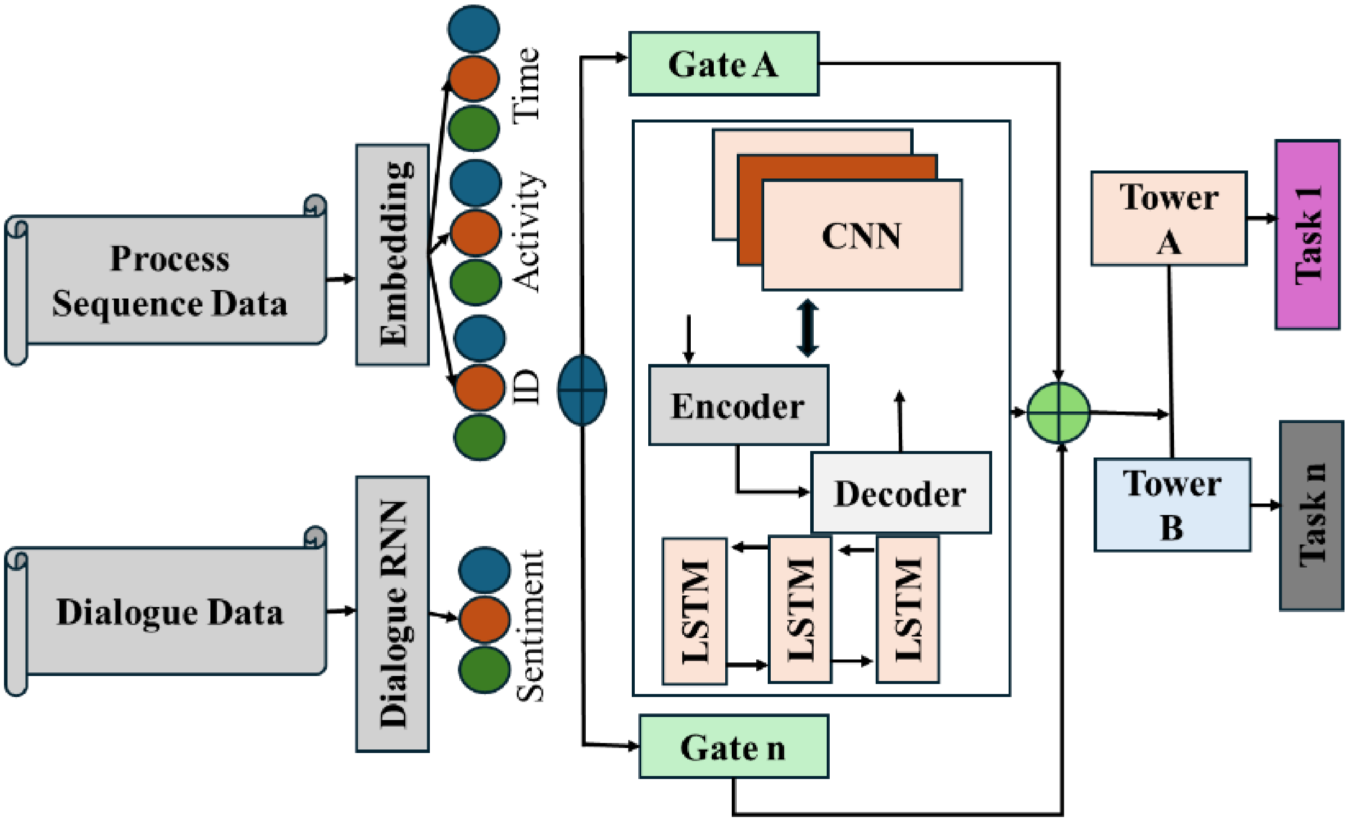


Fig. 3. Proposed Methodology Framework.

B. Feature Representation and Encoding

The prefixes in the preprocessed log need to be encoded into feature vectors of a specific size before the deep learning model could be trained. In the log, the attributes are classified as either categorical (qualitative) or numeric (quantitative), and each of these different types of attributes requires a different set of encoding techniques. In order to translate input activities into sparse, high-dimensional feature vectors, one-hot encoding is in the beginning applied to category attributes. Further encoding of these vectors is performed with the help of an embedding layer to solve the sparsity and loss of relational information. Normalization of numeric timestamp attributes is accomplished using min–max scaling in order to guarantee uniformity across all values.

The activity encoding procedure employs the skip-gram algorithm, with training depending on the simultaneous occurrence of actions in traces. For a trace $L= [x1, x2,...,xn]$ of length n, the likelihood of noticing m elements surrounding the t-th action is calculated as;

$$P(x_{t-m}, \ldots, x_{t-1}, x_{t+1}, x_{t+m|}x_t) = \prod_{-m\leq j\leq m, j\neq 0} P(x_{t+j|}x_t) \quad Eq\ (1)$$

During training, this research aims to increase the likelihood of events surrounding specific behavior as much as possible. This is expressed as.

$$e(x) = -\sum_{-m\leq j\leq m, j\neq 0} LogP(x_{t+j|}x_t) \quad Eq\ (2)$$

The approach efficiently represents predicted tasks' activities and temporal linkages using one-hot encoding, embedding layers, and skip-gram modeling.

C. Analysis of Sentiments

Capturing sentiment in call center discussions is challenging due to the interactive, brief, and contextually ambiguous nature of the conversations. To address these challenges, the **DialogueRNN** model is specifically designed for dialogue contexts. GRU networks underpin its Emotion Representation modules, Global State, and Party State. The Party State module tracks speaker and listener emotions. Listener and Speaker GRUs are crucial components. These elements analyze the conversation's mood and emotions. The Global State module integrates contextual information to help comprehend conversation flow. Encoding both the present and prior dialogue utterances gives a complete picture of the conversation's progression. The Emotion Representation module classifies conversation sentiment by analysing the speaker's words and context. This module pulls important emotional insights from speech, allowing the model to reliably represent sentimental changes throughout the exchange.

D. Dynamic Loss Weighting for Heterogeneous Multi-task Models

Designed to improve multi-task learning for commercial process prediction, the H-MMoE (Heterogeneous Mixture of Experts) architecture. It enhances prediction efficiency and accuracy over several tasks by combining dynamic loss weighting, heterogeneous expert networks, and transformer-based gating. Important elements of this construction consist in:

1. **Heterogeneous Expert Networks**:

- **LSTM (Long Short-Term Memory)**: Perfect for sequential data it captures temporal dependencies in event traces. Using input, forget, and output gates, LSTMs build models of the connections between events across time.
- **CNN (Convolutional Neural Networks)**: Uses spatial representations derived from process data to extract local information. CNNs concentrate in spotting trends in localised data areas.
- **Transformer Networks**: Use attention processes to calculate event linkages regardless of distance. Transformers capture long-range data dependencies well.

2. **Transformer-based Gating**:

Based on multi-head attention, the gating mechanism effectively distributes information tailored for each expert network. The model becomes more flexible and able of managing complicated input interactions by dynamically changing weights for expert outputs. This lets the model excel in predictions tailored for tasks.

3. **Dynamic Loss Weighting (UWL)**:

Using data distribution and performance measures for every subtask, the Uncertainty Weighting Loss (UWL) method adjusts task-specific loss weights during training. UWL dynamically changes the weights unlike stationary loss weighting techniques, hence enhancing training efficiency and lowering the need for hand-held hyperparameter adjustment. By more skillful handling of task trade-offs, this also maximizes the balance between chores. These methods used together guarantee strong generalization, diverse feature extraction, and improved prediction performance in the H-MMoE architecture. The dynamic character of the gating and loss mechanisms adds even more to the adaptability and flexibility of the model in challenging, real-world corporate process prediction activities.

The loss function is expressed as:

$$Loss\,(W, \sigma 1, \sigma 2) = \frac{1}{2\sigma_1^2} L1(W) + \frac{1}{2\sigma_2^2} L2(W) + log\sigma 1 + log\sigma 2 \qquad Eq\,(3)$$

The task-specific losses are where L $1(W)$ and $L\ 2(W)$; learnable parameters, or $\boldsymbol{\theta}$ 1 and $\boldsymbol{\theta}$ 2, reflect the uncertainty associated with each task. Regularizing to avoid overfitting are the logarithmic terms. Through constant weight updating during training, the UWL method helps the model to adjust to shifting data distributions. This lowers computing resource use and helps the model to attain improved performance. Combining heterogeneous experts, adaptive gating, and dynamic loss weighting guarantees that the H-MMoE model can efficiently manage the challenges of multi-task learning, hence enhancing both representation and prediction capability.

E. Decoding of Output

This study compares activity and time prediction tasks. With the SoftMax activation function at the output layer, activity prediction is a multi-class classification challenge. This loss function uses categorical cross-entropy to minimize the difference between expected and actual labels. Time prediction is a regression task with a fully connected output layer and the Mean Absolute Error (MAE) loss function to quantify the difference between anticipated and actual results. These specialized methods do both prediction objectives accurately and efficiently.

$$Loss(x) = -\sum_{i=1}^{c} yi * logfi(x) \qquad Eq(4)$$

The actual label for the i-th category by yi, the total number of categories by c, the projected probability for the i-th category by $fi(x)$, all define the input trace. This method reduces the discrepancy between the expected and real labels, thereby guaranteeing accurate classification.

IV. EXPERIMENTAL RESULTS

This study assessed the efficacy of the suggested approach by means of a real-world event log, therefore contrasting its performance with modern techniques. This section provides details about the dataset, experimental setup, evaluation metrics, baseline methods, and results for the two prediction tasks: predicting the outcome of the process and the remaining time.

A. Dataset

The BPIC_2016 dataset [20] was used in this research, and its statistics are shown in Table 2. This dataset contains customer behavior data collected by a Dutch government agency over eight months. It includes information from sources such as message data, call center logs, website click data, and complaint records. To prepare the dataset, cases with missing information were removed, and features such as the day of the week, time, event duration, and elapsed time were engineered. Each conversation was transformed into a "question event" and recorded under the message attribute. This dialogue data was then integrated with the process sequence data using unique case identifiers.

TABLE II. DATASET STATISTICS

Events	Dataset Entity
Attributes of Dialogue	Messages
Events Dialogue	Question
Activities	29
Mean cases duration in days	8.3
Median cases duration in day	0.979
Mean events per case	7.749
Events	221977
Trace variants	21948
Cases	28645

Table 3 provides a summary of the dataset that was produced consequently, which was restructured according to timestamps and stored in XES format.

TABLE III. SAMPLES OF DEPLOYED DATASET

Case ID	Activity	Timestamp	Message
	question	6/08/2023 13:10	What is the deadline for submitting the revenue issue form?
912	question	6/08/2023 13:13	Can you tell me when I can expect to get my unemployment benefits?
	home	6/08/2023 11:10	-
	taken	6/08/2023 11:12	-
	mijn_cv	6/08/2023 11:13	-
	mijn_cv	6/08/2023 11:13	-
	mijn_beri	6/08/2023 11:25	-
	mijn_rech	6/08/2023 11:25	-
	home	6/08/2023 11:26	-
	question	6/08/2023 9:44	The page has a script that has been running for some time.
15771	question	6/08/2023 9:58	What is the deadline for submitting the revenue issue form?
	taken	6/08/2023 20:51	-
	home	6/08/2023 21:24	-
	taken	6/08/2023 21:25	-

B. *Experimental Settings*

Implementation Framework: The proposed models were developed using TensorFlow and Keras, selected for their powerful deep learning capabilities. PM4Py, a specialized process mining library, was used to analyze event logs and manage internal sequence data representations. Dialogue preprocessing, including sentiment scoring, was carried out using the DialogueRNN framework, which effectively captures conversational context and emotional nuances.

Data Partitioning: An 80-20 temporal split was applied to divide the event log into training and test datasets. This method ensures that the test set contains instances that occur after the training set in time, thus preventing information leakage and better reflecting real-world scenarios. Two separate data samples were prepared for model training: one containing only information system sequence data, and another combining sequence data with dialogue data for enhanced multimodal analysis.

Evaluation Metric: Predictive model performance was assessed across a broad spectrum of criteria. While the mean absolute error (MAE) gauged the mistake in estimating remaining process time, accuracy and dependability of process outcome predictions were evaluated using precision (P), recall (R), and F1 Score. For both classification and regression tasks, this mix of measures guarantees a complete evaluation of the models.

Hyperparameter Optimization: To improve model performance, the hyperparameters indicated in Table 4 were optimized using a grid search method. Based on literature reviews and experimental data, key parameters like activation function, number of epochs, batch size, and learning rate were deliberately chosen. To identify the ideal values for every job, the grid search investigated many setups carefully. Since it guarantees accurate probability distributions for predictions and fits multi-class classification problems, the SoftMax activation function was selected.

TABLE IV. HYPERPARAMETER OF THE DEPLOYED PROPOSED MODEL

Parameter	Value
Optimizer	Nadam
Layers (CNN, LSTM, Transformer)	2, 2, 1
Learning Rate	0.01
Epochs	50
Loss Functions	Precision, Recall, F1, MAE

Batch Size	128
Dropout	0.2
Activation Function	ReLU, SoftMax
Transformer heads	4

C. Comparative Approaches

The developed H-MMoE model was compared to multiple state-of-the-art baseline business process prediction methodologies to assess its efficacy. These methods use specialized machine learning and deep learning techniques. Process models with temporal data were proposed in [26] to forecast process completion times. The author in [27] used probabilistic modeling to analyze the behavior of ongoing processes based on historical data. A researcher in [40] introduced a process monitoring technique based on stationarity assumptions, enabling predictions in stable process environments. In [16], running cases were transformed into spatial data representations, and Convolutional Neural Networks (CNNs) were applied for activity prediction. The researcher in [14] used Long Short-Term Memory (LSTM) networks to predict future activities and their durations by converting process traces into feature vectors. The author in [17] applied Transformer models to capture long-range dependencies in event logs, effectively modeling time-series data. In [4], a multi-task learning approach was developed using Transformer networks to predict event attributes and remaining time through shared feature representations. A researcher in [3] proposed a hierarchical Transformer model that combines information at different granularities using weighted representations, leading to improved predictions of next activities and remaining time. Finally, reference [28] employed BERT with transfer learning to build a multi-task prediction framework for forecasting activities and process outcomes. By benchmarking against these various methods, the proposed model demonstrates its superiority by integrating multimodal data and employing an advanced multi-task learning architecture.

V. Results & Discussion

The challenge of forecasting business process outcomes was analyzed using a Sankey diagram to evaluate the effectiveness of the proposed model. Figure 4 compares the actual and predicted results, highlighting two process outcomes: "question" and "home." In the diagram, nodes represent process activities, and the connecting lines show the flow and case counts between activities. The algorithm performed better for behaviors like "question," mostly because dialogue data improved prediction. However, "taken" predictions exhibited bigger errors because to the more diversified data distribution, illustrating the difficulty of this activity.

A. Comparative Analysis

Table V presents a comparison of the performance between single-modal and multimodal data techniques across baseline models. The multimodal approach showed higher prediction accuracy, highlighting the advantage of incorporating dialogue data, unlike single-modal methods that were limited to specific tasks.

The proposed model reduced the Mean Absolute Error (MAE) by up to 7.47% for remaining time prediction and improved the F1 score by 0.4% to 3.4% for process outcome prediction compared to baseline techniques

.

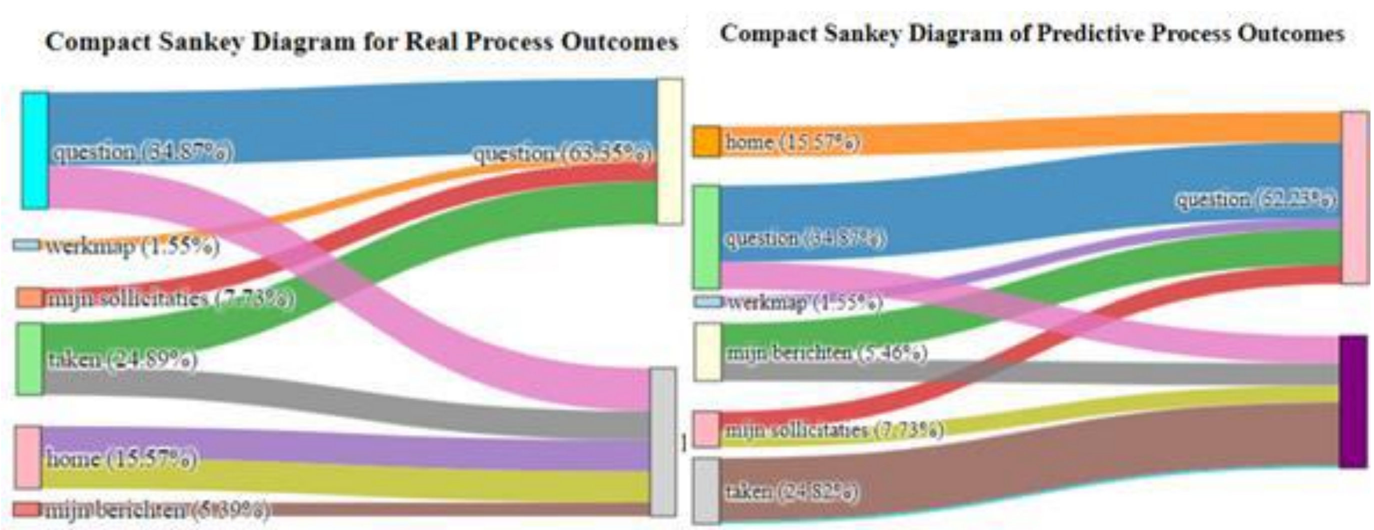


Fig. 4. Visual Illustration of Proposed Model Results.

The use of LSTM and Transformer networks produced better results than CNN, highlighting the importance of capturing sequential dependencies in business process data. This study's findings demonstrate that heterogeneous multi-task learning is both robust and effective in handling a wide range of feature representations.

TABLE V. COMPARISON WITH EXISTING LITERATURE

Ref	Remaining Time (MAE)	Process Outcome Precision	Process Outcome Recall	F1
[26]	**Single modal**: 3.108	0.486	0.209	0.293
	Multimodal: -	-	-	-
[27]	**Single-modal**: -	0.492	0.309	0.379
	Multimodal: -	-	-	-
[36]	**Single model**: 1.553	0.718	0.716	0.717
	Multimodal: 1.553	0.719	0.754	0.736
[17]	**Single model**: 1.001	0.832	0.634	0.720
	Multimodal: 0.977	0.726	0.723	0.724
[16]	**Single model**: 1.652	0.702	0.695	0.699
	Multimodal: 1.588	0.743	0.723	0.733
[14]	**Single model**: 1.037	0.695	0.708	0.708
	Multimodal: 0.983	0.748	0.716	0.732
[4]	**Single model**: 1.044	0.725	0.708	0.716
	Multimodal: 0.966	0.738	0.745	0.742
[3]	**Single model**: 1.022	0.736	0.722	0.729
	Multimodal: 0.970	0.741	0.729	0.735

Table V demonstrates the superiority of our H-MMoE model, particularly in multimodal settings. For example:

- In **Remaining Time prediction**, our model achieved an MAE of **0.966**, outperforming the baseline Transformer (**1.001**) and CNN (**1.652**) models.
- For **Process Outcome prediction**, H-MMoE attained a **Precision of 0.738**, **Recall of 0.745**, and **F1-score of 0.742**, whereas the best baseline (from [17]) had an F1-score of **0.720**.
- These metrics reflect improvements of **up to 7.47% in MAE** and **3.4% in F1-score**, demonstrating the value of combining heterogeneous data and dynamic task balancing.

B. *Performance of H-MMoE*

The H-MMoE model outperformed its isomorphic MMoE counterparts. As shown in Figure 5, process outcome predictions saw F1 score improvements of 4.8% to 6.8%, while the remaining time prediction achieved MAE reductions ranging from 8.71% to 34.47%. These improvements are attributed to the dynamic loss weighting mechanism, Transformer-based gating units, and heterogeneous expert subnets, which adaptively filter and prioritize shared features. The results demonstrate that H-MMoE effectively addresses the limitations of isomorphic models by utilizing multimodal data and task-specific representations.

C. *Ablation Study*

To evaluate the impact of individual components within H-MMoE, ablation experiments were conducted by removing key modules:

- **Without Dynamic Loss Weighting:** Replacing dynamic weighting with arithmetic means resulted in a 0.8% decrease in F1 and a 3.94% increase in MAE for process outcome and remaining time predictions, respectively.
- **Without Transformer Gate:** Using fully connected layers instead of Transformer gates led to a 1.6% decrease in F1 and a 2.48% increase in MAE.
- **Without Heterogeneous Experts:** Replacing heterogeneous subnets with isomorphic subnets caused a 1.9% decrease in F1 and a 2.59% increase in MAE.

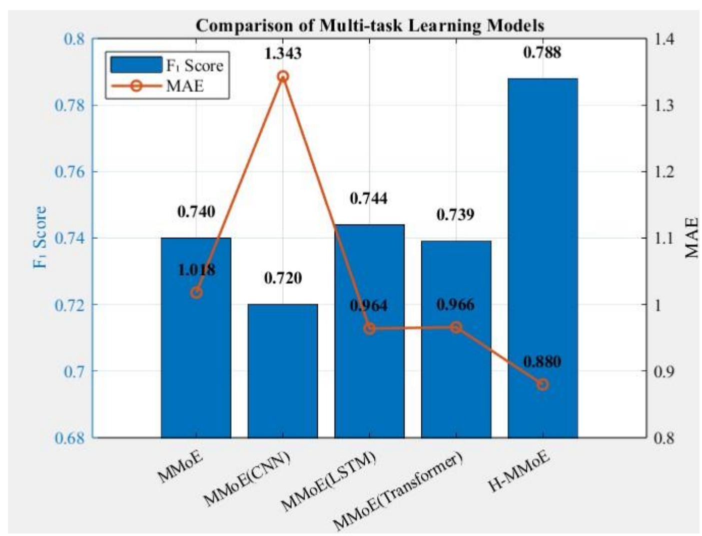


Fig. 5. Multitask learning Results comparison.

D. *Discussion and Comparison with Existing Studies*

The experimental results indicate that the proposed H-MMoE model outperforms traditional models such as CNN, LSTM, and Transformer when applied to predictive business process monitoring. For example, in comparison to the Transformer-based baseline [17], which achieved an F1-score of 0.720 and MAE of 1.001, our H-MMoE achieved an F1-score of 0.742 and a reduced MAE of 0.966. This improvement is largely due to the integration of multimodal data and dynamic task weighting. Similarly, compared to [14] (LSTM-based), which reported an F1 of 0.708 and MAE of 1.037, our model improved results by ~3.4% in F1 and ~7% in MAE. These comparisons demonstrate the superiority of our architecture in capturing both temporal dependencies and sentiment context. The ablation study further validated that removing components like dynamic loss weighting or the heterogeneous experts led to noticeable performance degradation, proving their critical contribution. Unlike single-modal or static-task frameworks in [3], [4], and [16], H-MMoE supports adaptive learning from multiple data streams, which is more aligned with real-world operational dynamics in call centers.

E. Influence of DialogueRNN

Figures 6 and 7 show the results of comparing the impact of DialogueRNN on sentiment analysis with that of BERT, BiLSTM, and TextCNN. With an F1 score of 0.735 and an MAE of 0.983, DialogueRNN outperformed all other models. Other models, including TextCNN, showed notable performance drops, with F1 decreasing by 3.3% and MAE increasing by 4.57%. In sentiment analysis for call center scenarios, DialogueRNN's superior performance is evident, aided by its ability to track participant states and retain contextual discussion information.

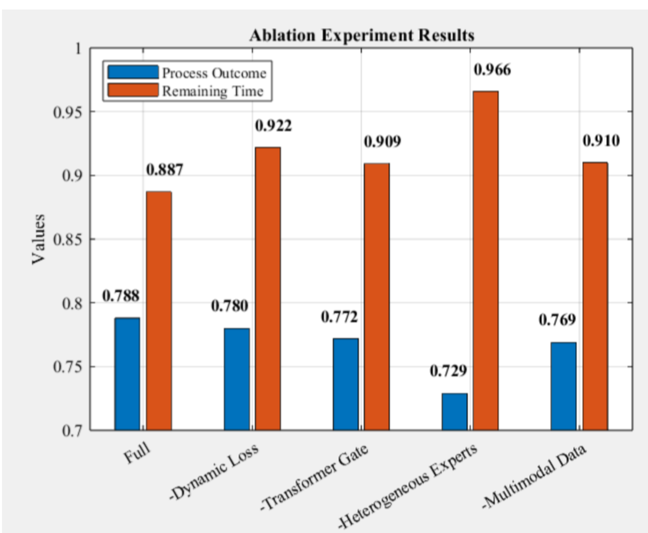


Fig. 6. Ablation Experimental Results.

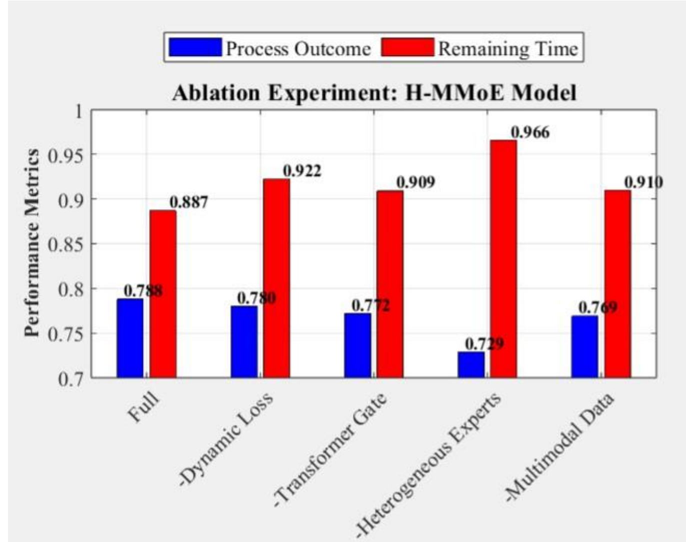


Fig. 7. H-MMoE Model Ablation Experimental Result

Figure 6 shows the results of the ablation experiment comparing the performance of the full H-MMoE model with its variants (e.g., without dynamic loss, transformer gate, heterogeneous experts, and multimodal data) for predicting process outcomes and remaining time. It highlights the effectiveness of the complete model and demonstrates how each component contributed to improved predictions. The performance measures of the H-MMoE model in both its whole form and ablated variations are shown in figure 7. Whereas the red bars reflect the MAE for estimating remaining time, the blue bars show the F1 scores for estimating process outcomes.

VI. Conclusion

An new predictive approach for call center business process monitoring utilising the H-MMoE model is presented in this study. We overcome the limits of single-modality approaches by merging corporate process sequence data with customer dialogue data. The suggested H-MMoE architecture outperformed LSTM, CNN, and Transformer in process result and duration prediction. These findings show that multimodal data and a field-specific framework improve prediction accuracy and understanding of complicated relationships. Dynamic loss weighting and Transformer-based gating units helped H-MMoE handle task unpredictability and improve shared feature extraction. The model performed better with dialogue data, especially for sentiment-sensitive tasks like "question." The study also found forecasting issues with different data distributions, indicating further improvements. The findings suggest that AI-driven methods could transform business operations and decision-making. To improve prediction, future studies may include audio elements like silent periods and sound pressure changes. The platform could also contain real-time predictions and adaptive learning, providing unique workflow optimization potential in dynamic business situations

REFERENCES

[1] ABBASI, Mostafa, et al. A Review of AI and Machine Learning Contribution in Predictive Business Process Management (Process Enhancement and Process Improvement Approaches). arXiv preprint arXiv:2407.11043, 2024.

[2] ZHENG, Ruixuan, et al. Prediction of steelmaking process variables using K-medoids and a time-aware LSTM network. Heliyon, 2024.

[3] AL OLAIMAT, Mohammad; BOZDAG, Serdar; ALZHEIMER'S DISEASE NEUROIMAGING INITIATIVE. TA-RNN: an attention-based time-aware recurrent neural network architecture for electronic health records. Bioinformatics, 2024, 40.Supplement_1: i169-i179.

[4] LIU, Ying, et al. TABLE: Time-aware Balanced Multi-view Learning for stock ranking. Knowledge-Based Systems, 2024, 303: 112424.

[5] CUI, Jin, et al. A novel sparse Gaussian process regression with time-aware spatiotemporal kernel for remaining useful life prediction and uncertainty quantification of bearings. Structural Health Monitoring, 2024, 14759217241282876.

[6] DENG, Lingfei, et al. Time-Aware Attention-Based Transformer (TAAT) for Cloud Computing System Failure Prediction. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024. p. 4906-4917.

[7] LIU, Xu, et al. LMGD: Log-Metric Combined Microservice Anomaly Detection through Graph-based Deep Learning. IEEE Access, 2024.

[8] ANG, Gary; LIM, Ee-Peng. Temporal Implicit Multimodal Networks for Investment and Risk Management. ACM Transactions on Intelligent Systems and Technology, 2024, 15.2: 1-25.

[9] ZONG, Chang, et al. Stock Movement Prediction with Multimodal Stable Fusion via Gated Cross-Attention Mechanism. arXiv preprint arXiv:2406.06594, 2024.

[10] BAO, Wuzhida, et al. Data-driven stock forecasting models based on neural networks: A review. Information Fusion, 2024, 102616.

[11] XUAN, Yang, et al. An Enhanced Bidirectional Transformer Model with Temporal-Aware Self-Attention for Short-Term Load Forecasting. IEEE Access, 2024.

[12] WANG, Peipeng, et al. MADMM: microservice system anomaly detection via multi-modal data and multi-feature extraction. Neural Computing and Applications, 2024, 1-19.

[13] NASARUDIN, Nurul Athirah, et al. A review of deep learning models and online healthcare databases for electronic health records and their use for health prediction. Artificial Intelligence Review, 2024, 57.9: 249.

[14] MARTÍNEZ-AGÜERO, Sergio, et al. Multimodal Interpretable Data-Driven Models for Early Prediction of Antimicrobial Multidrug Resistance Using Multivariate Time-Series. arXiv preprint arXiv:2402.06295, 2024.

[15] U. M. Ramirez-Alcocer, E. Tello-Leal, G. Romero, and B. A. Macías-Hernández, "A deep learning approach for predictive healthcare process monitoring," Information, vol. 14, no. 9, p. 508, 2023.

[16] W. Ni, G. Zhao, T. Liu, Q. Zeng, and X. Xu, "Predictive business process monitoring approach based on hierarchical transformer," Electronics, vol. 12, no. 6, p. 1273, 2023.

[17] J. Wang, J. Huang, X. Ma, Z. Li, Y. Wang, and D. Yu, "MTLFormer: Multi-task learning guided transformer network for business process prediction," IEEE Access, 2023.

[18] LIU, Cong, et al. Discovering hierarchical multi-instance business processes from event logs. IEEE Transactions on Services Computing, 2023.

[19] DUAN, Ganglong; YAN, Shunfei; ZHANG, Meng. A Hybrid Neural Network Model for Sentiment Analysis of Financial Texts Using Topic Extraction, Pre-Trained Model, and Enhanced Attention Mechanism Methods. IEEE Access, 2024.

[20] SURAJ, D., et al. Deciphering Product Review Sentiments Using BERT and TensorFlow. FMDB Transactions on Sustainable Computing Systems, 2023, 1.2: 77-88.

[21] JUN GU, Wen, et al. Predicting stock prices with finbert-lstm: Integrating news sentiment analysis. In: Proceedings of the 2024 8th International Conference on Cloud and Big Data Computing. 2024. p. 67-72.

[22] DUONG, Anh-Quang, et al. Residual Relation-Aware Attention Deep Graph-Recurrent Model for Emotion Recognition in Conversation. IEEE Access, 2024.

[23] ZHAO, Yuchen; BARNAGHI, Payam; HADDADI, Hamed. Multimodal federated learning on iot data. In: 2022 IEEE/ACM Seventh International Conference on Internet-of-Things Design and Implementation (IoTDI). IEEE, 2022. p. 43-54.

[24] Haq IU, Ahmed M, Assam M, Ghadi YY, Algarni A. Unveiling the future of oral squamous cell carcinoma diagnosis: an innovative hybrid AI approach for accurate histopathological image analysis. IEEe Access. 2023 Oct 20;11:118281-90.

[25] Wang Y, Wang Y, Shi C, Cheng L, Li H, Li X. An edge 3D CNN accelerator for low-power activity recognition. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. 2020 Jul 21;40(5):918-30.

[26] ASHRAFI, Negin, et al. Effect of a process mining based pre-processing step in prediction of the critical health outcomes. arXiv preprint arXiv:2407.02821, 2024.

[27] BHARADIYA, Jasmin Praful. Machine learning and AI in business intelligence: Trends and opportunities. International Journal of Computer (IJC), 2023, 48.1: 123-134.

[28] CHEN, Hang; FANG, Xianwen; FANG, Huan. Multi-task prediction method of business process based on BERT and transfer learning. Knowledge-Based Systems, 2022, 254: 109603.

[29] MEHDIYEV, Nijat; EVERMANN, Joerg; FETTKE, Peter. A novel business process prediction model using a deep learning method. Business & information systems engineering, 2020, 62: 143-157.

Experimental Design–Based Optimization of Football Manufacturing: A Case Study of Anwar Khawaja Industries

Muhammad Nauman
Department of Industrial Engineering
University of Engineering and Technology
Peshawar, Pakistan
engr.muhd.nauman@gmail.com

Misbah Ullah
Department of Industrial Engineering
University of Engineering and Technology
Peshawar, Pakistan
misbah@uetpeshawar.edu.pk

Muhammad Junaid
Department of Industrial Engineering
University of Engineering and Technology
Peshawar, Pakistan
muhammadjunaid@uetpeshawar.edu.pk

***Abstract*— This study aims to investigate the critical factors influencing the weight and quality of footballs bladder during the manufacturing process, with a focus on optimizing production at Anwar Khwaja Industries (Pvt) Limited, Sialkot. This research employs the Definitive Screening Design (DSD) to identify and quantify the impact of key variables, including material composition and process parameters, on the final product's performance. Among the factors analyzed, Calcium Carbonate (CaCO3) emerged as the most significant factor, demonstrating a strong effect on the response variable. Additionally, interactions between Sulphur–CaCO3, Zinc Oxide–BHT, and CaCO3–BHT were found to be critical in determining football quality, durability, and cost–efficiency. Statical analysis, including regression modeling and ANOVA, underscores these relationships but also reveals model limitations.**

This study also addresses model accuracy concerns, reporting an R–squared value of 52.2%, while the low adjusted R2 (19.4%) and predicative R2 (0.0%) indicate limited generalizability. To address multicollinearity concerns, the factor reduction technique was applied, improving the reliability of experimental findings. The study emphasizes the role of advanced statistical techniques in optimizing manufacturing processes to maintain Pakistan's global leadership in football production.

Keywords*— *manufacturing, football bladder, definitive screening design (dsd), anova

I. INTRODUCTION

This chapter provides a comprehensive overview of football manufacturing, tracing its evolution from basic, rigid designs to the high–quality, technologically advanced footballs used today. Football's global popularity is unparalleled, making it one of the most accessible sports worldwide. Despite regional naming variations, the core essence of the game, is kicking a ball to score, has remained unchanged for centuries.

Although football as sport spans back over 3,000 years, its production in Pakistan began in the late 19th and early 20th centuries during British colonial times. A British solider in Sialkot brought a punctured football to a local cobbler for repairs. The cobbler not only mended it but also created a replica, resulting in consistent football production. This event marked the beginning of football manufacturing in Sialkot [1], [2], [3].

For over 30 years, Pakistan has been the leading exporter, particularly through Sialkot, a globally recognized hub for FIFA-certified footballs and high–quality sports goods. While some firms in the region manage their own brands and supply top global companies, outdated technology has made Pakistan vulnerable to competition, especially from East Asia (particularly China) [4]. One significant barrier to technological adoption in Pakistan's football industry is labor resistance, driven by fears of wage reductions. As a result, China has strengthened its position, securing major supply contracts and surpassing Pakistan in market influence [5], [6].

Labor standards and process innovations are central concerns in football manufacturing. China's large workforce and technological advancements have attracted global firms, mirroring trends in industries like electronics manufacturing. Sialkot, Pakistan, remains a major production hub, housing approximately 1,000 football manufacturing units and employing over 60,000 people. It supplies more than 70% of the world's footballs, exporting over 37 million footballs for the 2018 FIFA World Cup. The intricate manufacturing process, from sourcing raw materials to stitching panels. It demonstrates the craftsmanship behind every football.

The continuous advancement of football manufacturing is driven by the pursuit of innovative materials and improved production techniques. Manufacturers explore cutting–edge technologies to enhance durability, performance, and sustainability. This includes integrating advanced manufacturing processes and rigorous quality control measures to meet evolving industry standards. Every football undergoes precise testing to ensure compliance with FIFA's stringent criteria before reaching the market.

Several studies have examined factors influencing football weight, bladder composition, and material properties, highlighting their impact on performance and sustainability. Recent advancements in football design focus on materials, sensor technology, and aerodynamics. Smart footballs embedded with sensors provide real–time performance data, while advanced materials like graphene enhance durability and flight stability [7], [8]. Aerodynamic studies using CFD simulations optimize panel designs to improve flight characteristics [9], [10].

The weight of a football plays a crucial role in player performance and safety. Studies highlight the importance of maintaining standardized weights to ensure consistency and reduce injury risks. Emerging technologies, such as 3D printing [11], enable the creation of customized football designs tailored to specific needs. Additionally, energy–harvesting innovations are paving the way for sustainable, smart footballs [8], [12]. These advancements are transforming modern football manufacturing, driving improvements in performance, safety, and accessibility. Quality assurance extends beyond materials and production methods to include rigorous testing for strength, durability, rebound properties, and water resistance [13], [14]. These tests simulate real–game conditions, ensuring consistent performance. Compliance with FIFA's quality standards is essential for certification, serving as a benchmark of excellence in the global sports industry.

Despite Pakistan's prominence in global football production, especially through firms such as Anwar Khawaja Industries (Pvt.) Ltd., the lack of systematic process optimization and over-reliance on traditional manufacturing methods result in variability in product quality, particularly the weight of football bladders. This inconsistency not only compromises product performance but also reduces manufacturing efficiency and increase production costs. The absence of robust experimental approaches to identify and control critical process parameters has created a significant gap in quality assurance and process optimization.

This study aims to identify and optimize critical factors influencing the weight and quality of football bladders in the manufacturing process at Anwar Khawaja Industries, Sialkot, with a focus on enhancing product performance, consistency, and cost-efficiency through a structured experimental approach. A key factor in determining the quality of a football bladder is material selection, as the bladder is typically made from latex, butyl and synthetic rubber [15], with a different composition which influence the weight and quality of football bladder. The quality of these materials plays a crucial role in the bladder's durability, air retention and overall performance. High quality materials ensure that the bladder maintains proper inflation, provides consistent bounce, and can withstand the stresses of regular use, ultimately contributing to improved product quality and performance.

II. METHODOLOGY

This research followed a systematic methodology to ensure a structured and reliable approach, as illustrated in Fig 1. The methodology comprised the following steps:

Fig 1. Research Methodology Flowchart

A. *Defining the research problem*

The research focused on optimizing the production process of latex-based football bladders within a specific weight range (395 to 450 grams). The aim was to identify key factors influencing product weight and enhance process efficiency using experimental design and statistical analysis.

B. *Literature review and variable selection*

A thorough literature review was conducted to assess current knowledge and practices in polymer blending, latex-based products, and experimental optimization. This review facilitated the selection of seven critical variables deemed influential in the quality and weight of football bladders, including

Sulphur, Zinc Oxide, ZDEC, Latex, Calcium Carbonate, BHT, and Carbon Black.

C. *Experimental design*

Experimental design using Definitive Screening Design (DSD) processes have developed extensively to boost industrial and scientific procedures by discovering key factors that influence the response variable. Developed by Bradley Jones and Chris J. Nachtsheim in the early 2010s. This design offers generous corrections over conventional screening designs, as it can assess the main effects, two–factor relations, and quadratic effects with a lowered number of experimental runs [16], [17].

Unlike fractional factorial designs, DSDs grant evaluations of major relations without being confounded by second–order effects, making them a competent tool for experimental optimization.

D. *Data collection*

Primary data were collected from the manufacturing department of Anwar Khwaja Industries Ltd., focusing on the latex-based football bladder production line. Each stage of the process was carefully observed, especially during the formulation and mixing stages, to ensure accurate and consistent data recording.

A total batch size of 300 kg was used, and all input variable concentrations were recorded meticulously across production batches.

Seven input variables were identified as having a considerable influence on the quality of the football bladders, with their respective mixing concentrations recorded as shown in Table I.

TABLE I. PERCENTAGE OF FACTORS FOR MANUFACTURING FOOTBALL BLADDER

S. No.	Factor	Percentage
1	Sulphur	1 – 3
2	Zinc–Oxide	2 – 8
3	ZDEC (Accelerator)	0.5 – 1
4	Latex	70 – 95
5	Calcium Carbonate (CaCO3)	10 – 30
6	BHT	0.2 – 0.5
7	Carbon Black	1 – 10

E. *Statistical analysis*

Before analysis, data preprocessing was conducted to ensure data quality and suitability for modeling. A manual inspection was performed to verify consistency in measurement units and eliminate any transcription errors from factory logs. As a result, the dataset was found to be complete, with no missing entries.

To facilitate effective comparison across variables with different scales, all continuous input variables were normalized using min-max scaling, transforming their values to a standard range of (0, 1). This step was critical in ensuring that no single variable disproportionately influenced the regression model due to magnitude differences.

Subsequently, the factors were encoded using orthogonal polynomial contrasts, in accordance with the requirements of the DSD methodology. This encoding strategy preserves the ability to estimate the main effects, quadratic terms, and two-factor interactions, while maintaining orthogonality and interpretability within the model structure.

A DSD was employed to assess the effect of seven experimental factors on the response variable (bladder weight). The design allows the estimation of:

1) Main effects for each factor
2) Quadratic effects for detecting curvature.
3) Limited two-factor interactions

Table II provides a summary of the experimental design structure used in this study.

TABLE II. DESIGN SUMMARY

Factors:	7	Replicates:	1
Base runs:	17	Total runs:	17
Base blocks:	1	Total blocks:	1

The design involved seven independent variables (factors), each with two levels (low and high). Each experimental condition was tested once, resulting in a total of 17 runs to evaluate the main effects and possible interactions of all factors. The design included only one base block, indicating that all the runs were conducted in a single block without any external variations.

The response variable (RV) — representing the bladder weight — was modeled using second-order regression with interactions. The resulting regression model is expressed as follows:

$$RV = 5268 - 72.0(Sulphur) + 46.7(Zinc\ Oxide) + 1(ZDEC) - 23.52(CaCO3) - 4.26(Latex) - 4550(BHT) - 3.6(Carbon\ Black) - 0.015(Sulphur^2) + \cdots\cdots + 2.6(BHT * Carbon\ Black) \quad (1)$$

This model includes all main effects, quadratic terms, and two-factor interactions that were statistically significant or practically relevant to the optimization of the bladder weight.

III. RESULTS AND DISCUSSION

This chapter depicts the consequences gained from the experimental analysis done using the Definitive

Screening Design (DSD) method. The study wants to discover the critical factors affecting the response variable (RV) in the manufacturing procedure at Anwar Khwaja Industries (Pvt.) Limited, Sialkot. The conclusions are conferred in relation to their statistical significance and practical consequences [18].

A. *Model Summary and Statistical Analysis*

The regression analysis accomplished on the experimental data yielded an R–squared (R^2) value of 52.2%, with an adjusted R^2 of 19.4%. The predictive R^2 was found to be 0.0%, suggesting potential overfitting or overlooking factors affecting the response variable. The model had a standard deviation (S) of 4.4697, indicating a reasonable deviation in response.

The Analysis of Variance (ANOVA) results are evaluated and listed in Table III, the overall model was marginally meaningful with a p–value of 0.0640, demonstrating that the selected factors and interactions impact the response variable but not strongly. The linear terms had a p–value of 0.161, implying that the individual factors alone might not be sufficient in explaining the response variability.

B. *Significant Factors*

Among the main effects, Calcium Carbonate ($CaCO_3$) was found to be statistically significant (p=0.029), suggesting a strong impact on the response variable. Other linear factors such as Sulphur, Zinc Oxide, BHT, Latex, and Carbon Black did not show significant effects individually.

Several interaction effects were found to be statistically significant at the 5% level:

1) Sulphur * Calcium Carbonate ($CaCO_3$) (p=0.023)
2) Zinc Oxide * BHT (p=0.018)
3) Calcium Carbonate ($CaCO_3$) BHT (p=0.020)
4) Zinc Oxide * ZDEC (Accelerator) (p=0.050)

These interactions suggest that the effect of one factor on the response variable is dependent on the presence of another factor, emphasizing the importance of considering combined factor influences.

C. *Interpretation of Regression Coefficients*

The regression equation in uncoded units provides an estimate of the response variable based on factor levels. The coefficient values indicate the direction and magnitude of each factor's influence.

The high coefficient of BHT (–4550) suggests that its presence has a substantial negative impact on RV when considered alone. However, its interaction with Calcium Carbonate ($CaCO_3$) and Sulphur yielded significant positive contributions.

D. *Residual Analysis and Model Adequacy*

The diagnostic analysis revealed five unusual observations (Obs 17, 22, 52, 59, and 67) with large residuals. These data points may represent experimental variability or unidentified factors affecting the response. The presence of high variance inflation factors (VIFs), particularly for Sulphur (34.07), Zinc Oxide (17.54), and BHT (15.38), indicates potential multicollinearity, which could affect the stability of the regression coefficients.

E. *Discussion & Recommendation*

The findings indicate that Calcium Carbonate ($CaCO_3$) plays a critical role in the process, both as an individual factor and in interaction with Sulphur and BHT. The high multicollinearity among certain factors suggests that some variables might be redundant or highly correlated, necessitating further refinement in factor selection.

The following recommendations can be implemented for the desired results.

1) Increase $CaCO_3$ content if the objective is to enhance the stiffness of the bladder and cost reduction.
2) Decrease $CaCO_3$ if the objective is to enhance the flexibility and impact resistance.
3) Increase Sulphur (up to 3.5%) and maintain BHT (0.3%) if durability is the objective.
4) Increase BHT (0.5%) and maintain Sulphur (2.5%) if flexibility and air retention are the objective.

Additionally, the non–significant predictive R^2 value (0.00%) suggests that the current model may not be fully generalizable for predicting future outcomes. Further experiments or the addition of surplus method variables might be necessary to enhance the accuracy of model.

TABLE III. ANALYSIS OF VARIANCE

Source	DF	Adj SS	Adj MS	F–Value	P–Value
Model	35	1113.03	31.801	1.59	0.064
Linear	7	221.58	31.655	1.58	0.161
Sulphur	1	48.44	48.440	2.42	0.126
Zinc Oxide	1	35.49	35.485	1.78	0.189
ZDEC (Accelerator)	1	0.27	0.270	0.01	0.908
Calcium carbonate ($CaCO_3$)	1	100.50	100.500	5.03	0.029
Latex	1	29.09	29.086	1.46	0.233
BHT	1	41.48	41.479	2.08	0.156
Carbon Black	1	0.14	0.142	0.01	0.933
Square	7	246.11	35.158	1.76	0.116
Sulphur * Sulphur	1	0.11	0.111	0.01	0.941
Zinc Oxide * Zinc Oxide	1	44.08	44.079	2.21	0.144
ZDEC (Accelerator) * ZDEC (Accelerator)	1	22.37	22.370	1.12	0.295
Calcium carbonate ($CaCO_3$) * Calcium carbonate ($CaCO_3$)	1	51.19	51.190	2.56	0.116
Latex * Latex	1	24.80	24.804	1.24	0.270
BHT * BHT	1	32.13	32.133	1.61	0.210
Carbon Black * Carbon Black	1	50.45	50.452	2.53	0.118

Source	DF	Adj SS	Adj MS	F–Value	P–Value
2–Way Interactions	21	724.58	34.504	1.73	0.057
Sulphur * Zinc Oxide	1	17.18	17.185	0.86	0.358
Sulphur * ZDEC (Accelerator)	1	0.00	0.004	0.00	0.988
Sulphur * Calcium carbonate ($CaCO_3$)	1	110.47	110.470	5.53	0.023
Sulphur * Latex	1	12.35	12.352	0.62	0.435
Sulphur * BHT	1	33.98	33.975	1.70	0.198
Sulphur * Carbon Black	1	0.06	0.058	0.00	0.957
Zinc Oxide * ZDEC (Accelerator)	1	80.77	80.768	4.04	0.050
Zinc Oxide * Calcium carbonate ($CaCO_3$)	1	0.04	0.045	0.00	0.963
Zinc Oxide * Latex	1	1.94	1.941	0.10	0.757
Zinc Oxide * BHT	1	119.53	119.535	5.98	0.018
Zinc Oxide * Carbon Black	1	11.61	11.610	0.58	0.449
ZDEC (Accelerator) * Calcium carbonate ($CaCO_3$)	1	0.11	0.110	0.01	0.941
ZDEC (Accelerator) * Latex	1	0.48	0.479	0.02	0.878
ZDEC (Accelerator) * BHT	1	0.32	0.324	0.02	0.899
ZDEC (Accelerator) * Carbon Black	1	48.75	48.751	2.44	0.124
Calcium carbonate ($CaCO_3$) * Latex	1	5.34	5.344	0.27	0.607
Calcium carbonate ($CaCO_3$) * BHT	1	116.18	116.181	5.82	0.020
Calcium carbonate ($CaCO_3$) * Carbon Black	1	0.04	0.042	0.00	0.964
Latex * BHT	1	13.00	12.997	0.65	0.424
Latex * Carbon Black	1	22.48	22.480	1.13	0.294
BHT * Carbon Black	1	0.78	0.779	0.04	0.844
Error	51	1018.90	19.978		
Total	**86**	**2131.93**			

The findings are consistent with those of industrial research on the mixing of rubber and polymers [19]where the performance of the material is greatly affected by component interactions. BHT and calcium carbonate interact to improve the material qualities of the football bladder, as shown in Fig. 2, by the significance of BF and DF interactions.

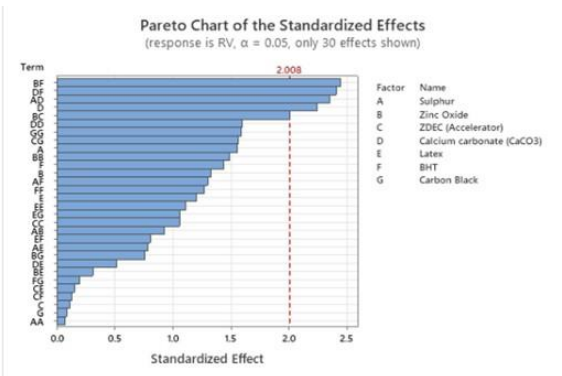


Fig 2. Pareto Chart of Standardized Effects

Additionally, previous investigations revealing the GG's (Carbon Black) influence on material durability and elasticity which supports the existence of the GG as a significant component.

In the production of football, these concepts support the claim that increasing the proportion of these elements may improve material performance, durability, and impact resistance. Additional research might increase the percentage of these principal factors to more effectively achieve the desired response variable.

Fig. 3 shows the residual plots, which give vital information about the suitability of the current model. The Normal Probability Plot (top–left) addresses that residuals closely follow the straight–line pattern, which shows that the errors are normally distributed. This normality assumption is essential for verifying the strength of statistical conclusions drawn from the current model [18].

The random scatter of points without any clear pattern in the Residuals vs. Fits plot (top–right) shows that the residuals establish a constant variance and no regular biasness. This means that the experimental model suitably tells the deviation in the response variable without having any major errors/issues. There are some outliers presents, which might indicate minor variability due to experimental states or material randomness.

The normality assumption in the Histogram of Residuals (bottom–left) is shown by a balanced distribution centered around zero. This regularity confirms that the experimental model correctly predicts the response without meaningful variations.

The residuals change randomly over the experimental runs is shown in the Residuals vs. Observation Order plot (bottom–right), with no apparent trend. This randomness means that there is no time–dependent errors in the experiment.

These residual plots support the experimental model acquired using Definitive Screening Design (DSD) and strengthen its robustness in forecasting the response variable in football manufacturing. Further improvement of key factor relations, such as BF, DF, and CG, could lead to higher material performance and improved manufacturing processes.

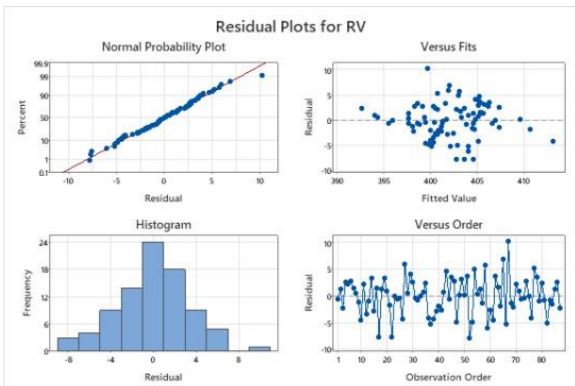


Fig 3. Residual Plots for RV

The correlation matrix in the sensitivity analysis represents the relationship between the selected factors (Sulfur, Zinc Oxide, ZDEC, Calcium Carbonate, Latex, BHT, Carbon Black) and the resultant variable are evaluated and shown in Fig 4.

Each value in the matrix ranges from –1 to 1, where the closeness of values to 1 indicates a strong positive correlation, meaning that an increase in one variable tends to increase the resultant variable while closeness of values to –1 indicates a strong negative correlation, indicating that an increase in one variable tends to decrease the resultant variable. A value near 0 suggests little to no correlation.

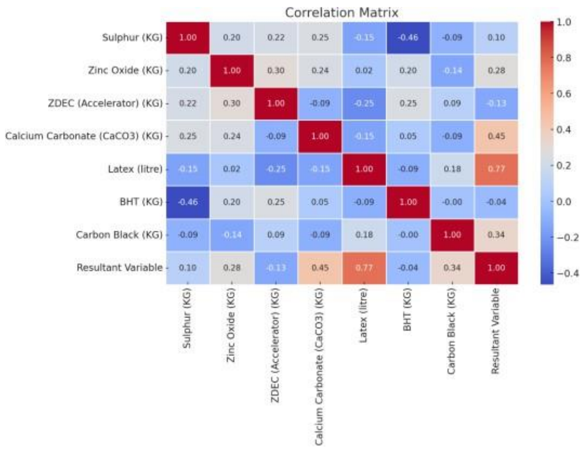


Fig 4. Correlation matrix of selected factors

TABLE IV. SELECTED FACTORS CORRELATION WITH RV AND ITS EFFECTS

Factor	Correlation with Resultant Variable	Effect
Calcium Carbonate ($CaCO_3$)	Positive (Strong)	Increasing the output
Latex	Positive (Moderate)	Increasing the output
BHT	Positive (Weak to Moderate)	Slightly increases the output
Sulfur (S)	Positive (Weak)	Slightly increases the output
Zinc Oxide (ZnO)	Negative (Weak)	Slightly decreases the output
ZDEC (Accelerator)	Negative (Moderate)	Decreases the output
Carbon Black	Negative (Moderate)	Decreases the output

Table IV presents the correlation between the response variable and the input parameters. High positive correlations indicate key factors affecting the resultant variable, while weaker correlations indicate non–significance.

By analyzing these relationships, it can determine the impact of factors on response variable. This information is necessary for optimizing processes and making data–driven decisions. This matrix helps in understanding dependencies, identifying key contributors, and guiding future experimental designs. Sensitivity analysis using correlation helps in prioritizing variables for further study.

IV. CONCLUSION

This study, conducted for Anwar Khwaja Industries (Pvt.) Limited, Sialkot, utilizes Definitive Screening Design to identify key factors influencing football quality. Statistical analysis highlights Calcium Carbonate ($CaCO_3$) as the most significant factor, with a p–value of 0.029, indicating a strong impact on the manufacturing process.

Additionally, significant interaction effects were observed, particularly among Sulphur–$CaCO_3$, Zinc Oxide–BHT, and $CaCO_3$–BHT, demonstrating that combined variable influences can outweigh individual effects. This underscores the complexity of the manufacturing process.

The R–squared value (52.2%) suggests the model explains a moderate portion of response variation, but the low adjusted R^2 (19.4%) and predictive R^2 (0.0%) indicate limitations in generalizability, requiring further model refinement.

Multicollinearity was identified among Sulphur, Zinc Oxide, and BHT, complicating parameter estimation and prediction. The Factor Analysis technique was applied to mitigate these issues and enhance experimental reliability [20].

Residual analysis revealed potential undocumented factors, necessitating further research to determine whether variations stem from external influences or measurement errors. This study highlights the need for a more robust experimental design to account for internal and external variability in football manufacturing.

This research lays a solid foundation for finding the key factors affecting the weight and quality and highlights the potential models and processes which may improve the response variable.

V. RECOMMENDATIONS FOR FUTURE WORK

The current study has a profound impact on finding the key factors resulting in direct effects on the response variable (quality). To enhance the application of the current findings, the following recommendations may result in an increase of strength and applicability.

1. Incorporation of additional factors.
2. Enhancing the levels of current factors.
3. Larger and updated datasets

REFERENCES

[1] Jalandhar and India, "Global Value Chains, Local Clusters and Corporate Social Responsibility: A Comparative Assessment of the Sports Goods Clusters in Sialkot," 2009.

[2] D. Atkin, A. Chaudhry, S. Chaudry, A. K. Khandelwal, and E. Verhoogen, "Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan*," *Q J Econ*, vol. 132, no. 3, pp. 1101–1164, Aug. 2017, doi: 10.1093/qje/qjx010.

[3] M. Zaheer, "Knowledge on the Fingertips: A Study of Football Production in Pakistan," vol. 13, no. 2, pp. 122–133, doi: 10.24312/1991130217.

[4] T. Raza, "Measuring Technology Differences Across Football Manufacturers in Sialkot."

[5] M. A. Tanveer, S. Shaheer, and H. Rizvi, "Declining Market Share of Pakistan in Football Industry," 2011. [Online]. Available: www.ajbms.org

[6] K. Nadvi, "Labour standards and technological upgrading: competitive challenges in the global football industry," 2011.

[7] S. , W. D. and G. R. Lee, "Innovative Materials for Football Manufacturing: Enhancing Performance and Durability. Materials & Design.," 2019.

[8] R. , S. T. and B. J. Adams, "Energy Harvesting in Smart Footballs: A Review of Technologies and Applications. Renewable and Sustainable Energy Reviews.," 2021.

[9] S. K. Vinod, R. V Menon, and S. Babu, "Analysis of footballs to study their overall performance using design parameters and explain the knuckling effect using CFD," 2021.

[10] J., M. L. and C. R. Taylor, "Aerodynamic Optimization of Football Design Using Computational Fluid Dynamics (CFD)," *Journal of Sports Engineering and Technology.*, 2021.

[11] S. Hong, J. E. Goff, and T. Asai, "The Aerodynamics of New Design Soccer Balls Using a Three-Dimensional Printer," *Applied Sciences (Switzerland)*, vol. 14, no. 9, May 2024, doi: 10.3390/app14093932.

[12] S. Harnadi and S. Supangkat, "Smart Football: A Systematic Review for Constructing Optimum Team Formation in Football," Apr. 2023, pp. 1–6. doi: 10.1109/ICISS59129.2023.10291603.

[13] Inside FIFA, "FIFA Quality Marks for Soccer Balls," https://inside.fifa.com/innovation/standards/footballs/fifa-quality-programme-for-footballs.

[14] Net World Sports, "FIFA QUALITY MARKS FOR SOCCER BALLS GUIDE," https://www.networldsports.com/buyers-guides/fifa-quality-marks-explained?srsltid=AfmBOoq2VzoKwwOCLwg9Het5X_gA7qxaygRYq5OuEXgZ9tgvDCeT1kxH&utm_source=chatgpt.com.

[15] Victeam Sport, "The Difference Between a Latex, Butyl and Synthetic Rubber (SR) Bladder," https://www.victeamsports.com/the-difference-between-a-latex-butyl-and-synthetic-rubber-sr-bladder/.

[16] M. Weese, P. Ramsey, and D. Montgomery, "Analysis of definitive screening designs: Screening vs prediction," *Appl Stoch Models Bus Ind*, vol. 34, Apr. 2017, doi: 10.1002/asmb.2297.

[17] R. Hayasaka, J. HÄnisch, and P. Cayado, "DSDApp: An Open-Access Tool for Definitive Screening Design," *J Open Res Softw*, vol. 12, pp. 1–8, 2024, doi: 10.5334/JORS.462.

[18] D. C. Montgomery, *Design and analysis of experiments*. John Wiley & Sons, Inc., 2013.

[19] N. R. and L. D. K. J. Draper, "Using definitive screening designs in response surface modeling," *Journal of Quality Technology*, vol. 45, no. 1, pp. 1–11, 2023.

[20] Muhammad Hassan, "Factor Analysis – Steps, Methods and Examples," https://researchmethod.net/factor-analysis/?utm_source=chatgpt.com.

Design and Development of 1-DOF Assisted Physiotherapeutic Device for Rehabilitation of Frozen Shoulder

Atiqa Saleem
Department of Biomedical Engineering
HITEC University
Taxila Cantt, Pakistan
atiqa.saleem@hitecuni.edu.pk

Shehla Inam
Department of Biomedical Engineering
Riphah International University
Islamabad, Pakistan
shehla.inam@riphah.edu.pk

Seemab Zakir
Department of Biomedical Engineering
Riphah International University
Islamabad, Pakistan
seemabzakir2@gmail.com

Leezish Gulshad
Department of Biomedical Engineering
Riphah International University
Islamabad, Pakistan
leesh.khan@hotmail.com

Faisal Amin
Department of Biomedical Engineering
Riphah International University
Islamabad, Pakistan
faisal.amin@riphah.edu.pk

Aleena Naweed
Department of Biomedical Engineering
HITEC University
Taxila Cantt, Pakistan
aleenanaweed@gmail.com

***Abstract*— The growing prevalence of limb-related issues has increased the demand for rehabilitation services. A key focus has been on restoring upper limb functionality and controllability, which are essential for patients to reintegrate into society and improve their quality of life. In this project, a prototype for an assisted physiotherapy device was designed to rehabilitate frozen shoulders. The device supports both external and internal rotation of the shoulder, allowing movement from 0 to 90 degrees. Its adjustable speed control lets patients move their shoulders based on their personal needs and abilities. Additionally, the device provides real-time feedback on the angle, speed, and force applied during rehabilitation. Experimental results show that this prototype is both practical and effective in rehabilitating shoulder rotation.**

Keywords—Assisted device, 1 DOF, Frozen shoulder, Rehabilitation, Upper limb

I. Introduction

Adhesive shoulder capsulitis, or frozen shoulder, is a condition where excessive scar tissue or adhesions form across the glenohumeral joint, causing pain, stiffness, and limited movement. It can develop either spontaneously (primary or idiopathic adhesive capsulitis) or following shoulder surgery or trauma (secondary adhesive capsulitis) [1]. In a study [2], a novel servomotor joint mobility device for treating frozen was compared to standard therapy. Patients using the device showed significant improvement in shoulder range of motion (ROM- flexion 36 %, abduction 51 %, internal rotation 81 %, and external rotation 88 %) and a 62% decrease pain (measured by VAS). These results outperformed the control group where ROM improvements were more modest, indicating the device's potential as a supplementary treatment for frozen shoulder.

Another study developed a game-based app to help increase patient engagement with home exercises following medical treatment for frozen shoulder. The app, compatible with mobile platforms aimed to improve joint mobility and function through rehabilitation exercises, with prototype testing to ensure it was both user-friendly and motivating. This allowed healthcare professionals to remotely monitor patient progress [3]. In a similar approach, a mobile using Unity 3D was created to simulate frozen shoulder movements, enabling tele-rehabilitation at home. Biofeedback data was sent to physiotherapists providing valuable tool for remote patient management and personalized treatment [4].

Additionally, a wearable motion sensor device paired with a smartphone. App was tested to monitor shoulder ROM. The device showed reliability (ICC range of 0.771 - 0.979) when compared to doctor evaluations and patient using device showed greater shoulder mobility, and functional recovery, than those in a home exercise group, over a three month period [5].

Another study, involved a three-month home-based program, where patients divided into two groups: smartphone-assisted exercise (n = 42) and traditional self-exercise (n = 42). Both showed significant improvements in ROM and range of motion and pain, but those using smartphones were more satisfied with the technology, suggesting smartphone-assisted were more satisfied with the suggesting smartphone assisted rehabilitation is just as effective and more enjoyable for users [6].

In an interesting participatory study, patients, healthcare providers, game designers, and academics collaborated to refine refined a Kinect-based game to help patients with frozen shoulder move more freely and experience less pain. This iterative process aimed to improve both therapeutic outcomes and the user experience [7]. Another study explored the use of two degree of freedom rehabilitation device specifically patient needing intensive rehabilitation for Adhesive capsulitis [8]. One more study focused on developing an automated sliding CPM device filling a gap in current robotic devices, by offering wide ROM coverage with precise monitoring and affordability for self-exercising patients [9]. Lastly, the effectiveness of shoulder joint mobilization using an innovative servomotor joint mobilization apparatus was examined for frozen shoulder patients [10].

Therefore, this study aims to design and develop a cost-efficient, 1-DOF-assisted device for frozen shoulder rehabilitation, using minimal resources. This device can be used independently by patients or with the help of physiotherapists.

II. Aim of the Study

These therapeutic exercises are typically guided by a physiotherapist and involve both active and passive movements, depending on the patient's condition. The main goal of rehabilitation devices is to restore a patient’s

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

sensory, physical, and psychological abilities, which may have been lost due to injuries, illness, or disease. These devices also help patients compensate for impairments that cannot be fully treated through medical care.

Although rehabilitation equipment and techniques are gradually improving, they remain expensive and often come with complex ranges of motion (ROM) and high degrees of freedom (DOF). Many upper-arm exoskeletons rely on human-robot interaction, which can be difficult for patients to operate manually and may lead to high repair and maintenance costs.

In response to these challenges, the authors of this study designed a simple prototype with 1 DOF. This device is easy to use, cost-effective, and locally manufactured, making it more accessible for patients. It helps reduce the cost of physiotherapy sessions while ensuring effective exercise. The device can also be easily used at home. The authors targeted the local community, particularly those who are unable to participate in social activities due to disability or pain.

The ADFFS device can perform 1 DOF movements, such as internal and external shoulder rotations, with four different speed modes, making it simpler to use compared to existing devices. A doctor recommends the appropriate speed, time, and duration for each exercise, so patients only need to select the mode and begin their therapy session. The ADFFS is independent of age, gender, and weight, and it allows physiotherapists to provide precise therapy without the need for manual exercise during multiple sessions. This device promotes progress in rehabilitation by ensuring safety, reliability, and self-dependence for patients.

III. Methodology

A. Directions of AD Device

This assisted device (AD) was designed to follow the instructions as shown in Fig. 1. The AD operates with a 1 DOF movement, allowing for both external and internal shoulder rotation, with shoulder angles ranging from 0° to 90°.

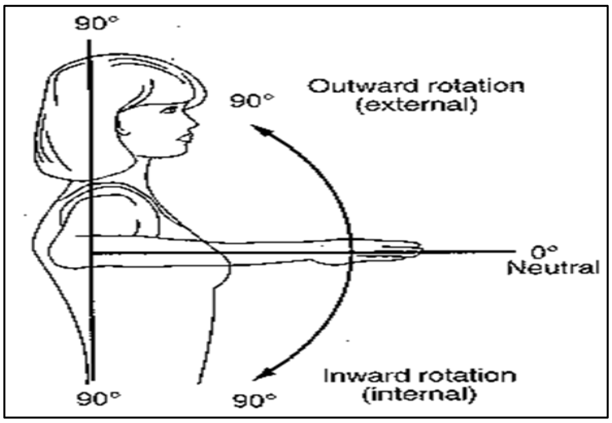


Fig. 1. Directions followed by the designed AD

B. Model Design

The complete mechanical design was created to support the arm and shoulder, with a graphical user interface (GUI) for mode selection and parameter adjustments.

As shown in the block diagram (Fig. 2), the assisted device was controlled by the microcontroller through the GUI, allowing the subject to select modes and adjust parameters independently. Sensory feedback was provided for situations where the subject might feel pain or discomfort. In such cases, the subject could stop the device using the safety switch. The device was adjustable to the subject's left arm, while the safety switch was positioned in the right hand for easy access. All operations were controlled by the subject with the microcontroller programming managing mode selection and parameter adjustments.

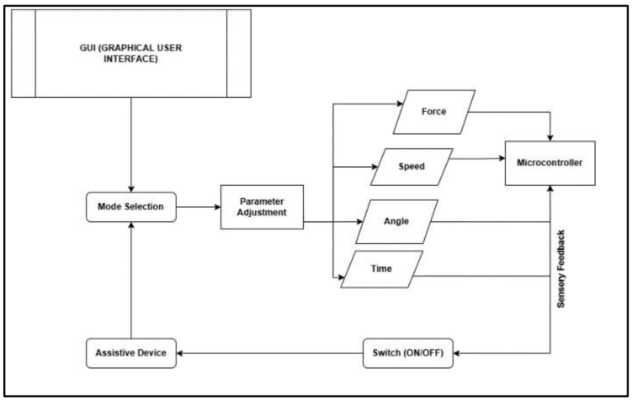


Fig. 2. Block diagram of AD

C. Prototype Design on Solid Edge

The prototype model was designed using Solid Edge software, which provided detailed results to define its purpose. The following diagrams created in Solid Edge included the correct dimensions.

Fig. 3 shows the dimensions of the prototype's base, which serves as the lower and fundamental part that supports the entire design. These dimensions were selected based on the overall design of the device.

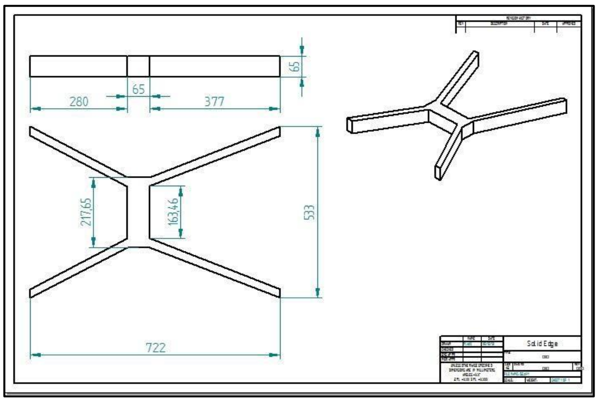


Fig. 3. Base of AD

Similarly, Fig. 4 illustrates the dimensions of the stainless-Steel rods that support the armrest. The affected arm rests on this part of the device, and the motors move the rods to facilitate exercise of the affected arm.

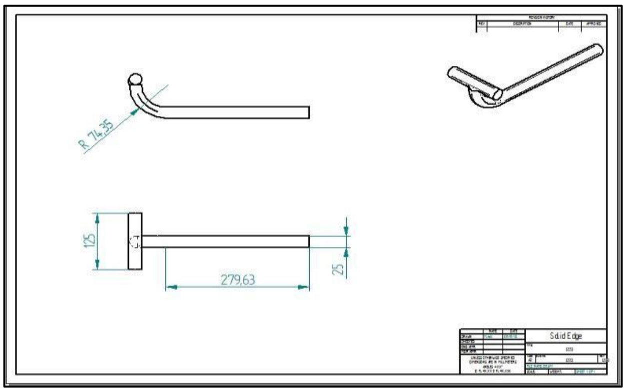


Fig. 4. Armrest of AD

The dimensions of the joint that connects the stainless steel rods are shown in Fig 5. These dimensions were selected to ensure the stability of the device and to make the assisted device comfortable for the patients.

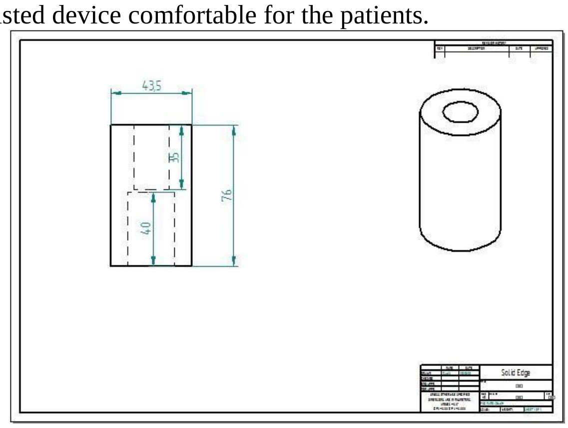


Fig. 5. Joint of AD

Fig. 6 shows the dimensions of the bending plates, which are perpendicular to the attached rods connecting the standing rods and armrests. Bolts are used to secure the bending plates and rods.

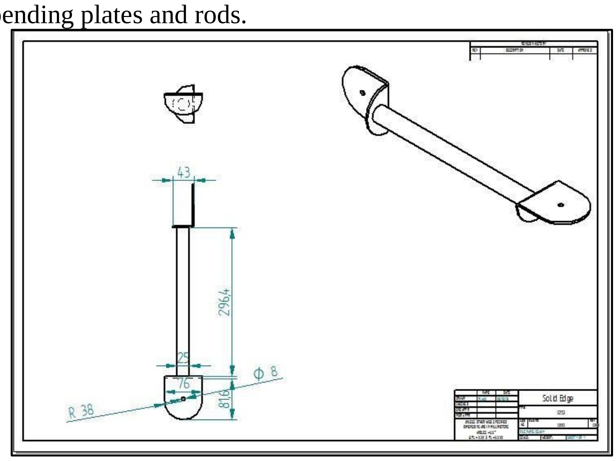


Fig. 6. Bend plate of the ADFFS

Fig. 7 shows the dimension of the aluminum alloy that adjusts the position of the shoulder and elbow joints. Bent plates with rods were attached to them to the alloy allowing the rods to move either towards or away from the body, thereby enabling movement of the armrest and shoulder.

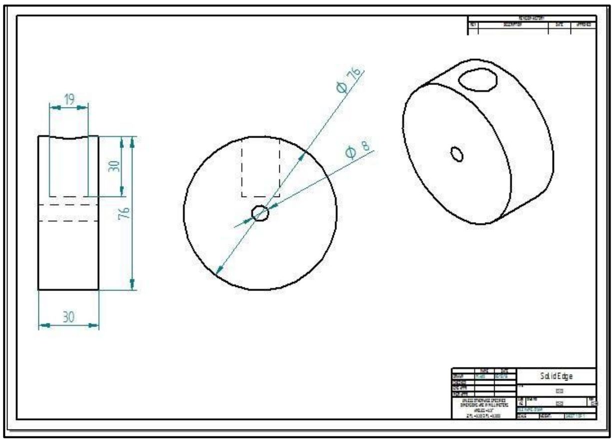


Fig. 7. Adjustment of arm part

D. Motor selection

DC gear motors were chosen for their ability to provide adjustable speed operations. These motors are ideal for applications that require high starting loads and consistent horsepower, even when operating at 150 % above their base speed. A key advantage of DC gear motors is their ability to generate full torque at low speeds which makes them suitable for carrying the average arm weight (5.3% of total body weight). Given that our prototype design needed to support the weight of the average human arm. The DC gear motor was the optimal choice. Additionally, DC motors offer higher power density, have no field coil in the stator, saving space and are compact. Their design makes them easy to maintain and control with lower inertia making them highly reliable and efficient for our application. Table I shows the specs of DC Gear motor.

TABLE I. MOTOR SPECS USED IN THE DESIGN (24V DC GEAR MOTOR)

Parameters	DC Gear Motor
Model Type	Model # DME60B6HF-224
Voltage (V)	24.0 V DC
Gear Ratio	1/ 150
Rotation Output (CW/ CCW)	31.3 r/min
Output Speed	300 rpm

E. Gyroscope and Accelerometer [MPU-6050]

Sensory devices are used to measure the position and orientation of the device's moving parts. A combination of gyroscope and accelerometer provides a comprehensive array of information. The gyroscope uses the Earth's gravity to measure rotational movement, while the accelerometer detects non-gravitational acceleration. The gyroscope measures the rate of rotation and the accelerometer measures linear vibration. The MPU-6050 sensor used in this study supports a sampling frequency of up to 1 kHz (1000 samples per second). However, in the current setup, the data acquisition was performed at a sampling rate of 100 Hz, which provides a reasonable balance between responsiveness and data stability. A higher sampling rate can offer more precise data, particularly in higher-speed modes where rapid changes in acceleration and angular velocity occur.

In future iterations, increasing the sampling frequency may help reduce data fluctuations and improve the accuracy of recorded force and speed values at higher angles and speeds. Table II and Table III show the specs of MPU-6050 that used in design.

TABLE II. GENERAL SPECIFICATIONS OF MPU-6050

Part/Item	MPU-6050
VDD	2.375V-3.46V
VLOGIC	1.71V to VDD
Serial Interfaces Supported	I^2C
Pin 8	VLOGIC
Pin 9	AD0
Pin 23	SCL
Pin 24	SDA

TABLE III. GYRO/MPU-6050 OPERATING CONDITIONS USED IN THE DESIGN

Parameters	GYRO/MPU-6050
VDD	2.375-3.46V
VLOGIC	1.8V$\pm$5% or VDD
Ambient Temperature- T_A	25°C

F. *Choice of Microcontroller*

An Arduino microcontroller was used in this assisted device. The microcontroller was programmed to interface with the DC gear motor, enabling it to perform external and internal rotation of the frozen shoulder at various speeds, ranging from 0 to 90 degrees. The Arduino's PWM pins were connected to the L298n motor driver module, while the analog pins were linked to the motor. The power (V), and ground (GND) connections were made to the adaptor's GND and power. These components work together to operate the prototype and the physiotherapeutic device (Fig. 8).

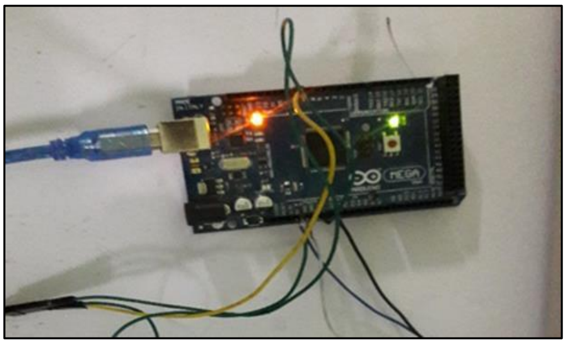


Fig. 8. Arduino (mega 2560)

G. *Power Supply and Calculation of Force*

The power supply for the device was connected to a drive circuit that provided 12 volts to the DC gear motor, as the motor operates on 12 volts. Two sensors, a gyroscope and an accelerometer [MPU-6050], were used to measure acceleration, which was then used to calculate the force generated by the device in moving the armrest. The following formula was used for the calculation:

$$F = ma \tag{1}$$

Where "a" is the acceleration, and "m" represents the mass of the arm. All necessary coding was completed using the Arduino Mega microcontroller to control the motor's speed. Switch buttons were incorporated to turn the device ON/OFF and adjust the speed, force, and angle based on the patient's needs.

H. *Implementation of prototype*

Metal fabrication was done to build some metal structures by cutting, bending, and assembly processes. The base of the design was made of mild steel, size of 2"× 1", 16 gauges. Caster wheels were attached to the lower side of the base for movement from one place to another. Electric arc welding was used to prepare this mechanical design. For circular joints 'Lathe Machine' was used. The dimensions of the Stainless-steel rod were 18 gauges, 18mm. The bending plate used in joints was of mild steel with a diameter of 3.6mm. Screws were made up of aluminum alloy, sizes of 4mm and 8mm. Joints were also made up of aluminum. The vertical rod's diameter was 1", 16-gauges, the height of 4". The finished form of the device parts after fabrication is given below in Fig. 9.

Fig. 9. Final design of prototype

IV. RESULTS

The speed and force exerted by the assisted device in moving the armrest were measured at angles of 15°, 30°, 45°, 60°, and 75°. The device was programmed to operate in four modes. Mode 1 was the slowest, moving the armrest at a low speed with minimal force. Modes 2, 3, and 4 gradually increased the speed and force, with Mode 2 offering moderate speed and force, Mode 3 providing high speed and force, and Mode 4 delivering the highest speed and force.
The results for all four modes (1–4) at the angles of 15°, 30°, 45°, 60°, and 75° are summarized in Table IV.

TABLE IV. OBSERVED SPEEDS AND FORCES AT DIFFERENT ANGLES AND MODES

	Mode (Slow)					
Angle (◦)	15^0	30^0	45^0	60^0	75^0	90^0
Speed (deg/s)	31.5	45.0	53.2	77	89	110
Force (Fv)	47	56	102	69	134	157
	Mode (Moderate)					
Angle (◦)	15^0	30^0	45^0	60^0	75^0	90^0
Speed (deg/s)	42	59.0	73.2	89	111	179
Force (Fv)	57	68.1	74.8	88.7	122.3	163
	Mode (High)					
Angle (◦)	15^0	30^0	45^0	60^0	75^0	90^0
Speed (deg/s)	55	45.0	86.2	99.1	131	185
Force (Fv)	62.5	74.9	88.2	97.8	139	187
	Mode (Very high)					
Angle (◦)	15^0	30^0	45^0	60^0	75^0	90^0
Speed (deg/s)	64.5	87.0	120	147.3	151	110
Force (Fv)	73	90.30	104.1	149.7	169.9	197.32

A. *Acquired Data Representation*

The acquired data showed that force and speed increased at higher angles. As the angle of the armrest increased from 15° to 90°, more force was needed to lift the arm. The angular rotational speed was recorded in degrees per second (deg/s), showing the speed at each respective angle. After the collection of data, a clear trend of increasing force at different armrest angles was observed. The data was graphically presented for all four modes (1-4). Overall, the trend indicated that at higher modes, more force was required to move the armrest. However, a limitation of this device was that sometimes the same values for force and speed were observed at the same armrest angles.

B. *Explanation of Results*

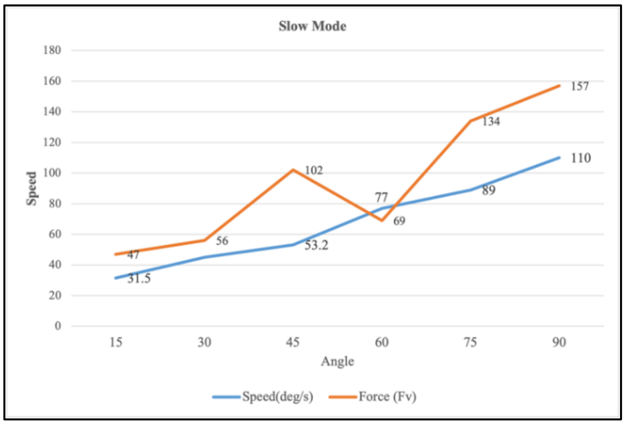


Fig. 10. Slow mode of AD

Fig. 10 shows the graph of force, angle, and speed at Slow Mode 1. This mode worked as the baseline, as it operates at the lowest speed and force. It effectively demonstrated the device's ability to gently mobilize the shoulder, which is crucial for early rehabilitation stages or patients with high pain sensitivity. The linear increase in force and speed with angle suggests predictable mechanical behavior, indicating good calibration and safe motion for initial therapy sessions. It demonstrates that as speed increased, force also increased to reach higher angles. Although there were some fluctuations in the readings, the overall graph indicated that these three parameters are directly related. As the angle increased from 15° to 90°, the speed started at 31.5 deg/s, gradually increased to 77 deg/s at 60°, and peaked at 110 deg/s at 90°. Meanwhile, the force began at 47 Fv, steadily increased to 134 Fv at 75°, and reached 157 Fv at 90°. This shows that both speed and force increased consistently with the angle, with speed rising slightly faster than force at higher angles.

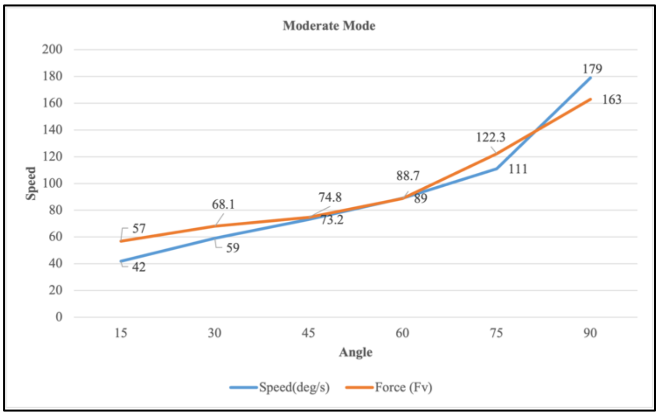


Fig. 11. Moderate mode of AD

Similarly, Fig. 11 shows the graph for Moderate Mode 2, where the speed values were higher as compared to the Slow Mode. Mode 2 exhibited a steeper increase in both force and speed compared to Mode 1. This implies a significant jump in mechanical demand, suitable for mid-stage therapy. The sharper slope in speed may suggest that the motor reaches a higher torque efficiency in this mode. It also indicates that patients with moderate mobility can be safely progressed to this mode, but with continuous monitoring to prevent strain. When the speed is greater the applied force increases with it and so are the selected angles. As a result, both force and speed were greater, but the increasing trend remained the same as in Slow Mode. As the angle increased from 15° to 90°, the speed started at 42 deg/s, gradually increased to 89 deg/s at 60°, and then peaked at 179 deg/s at 90°. Meanwhile, the force began at 57 Fv, steadily increased to 122.3 Fv at 75°, and reached 163 Fv at 90°. This demonstrated that both speed and force increased consistently with increased angle. We also observed that the speed raised slightly faster than force at higher angles.

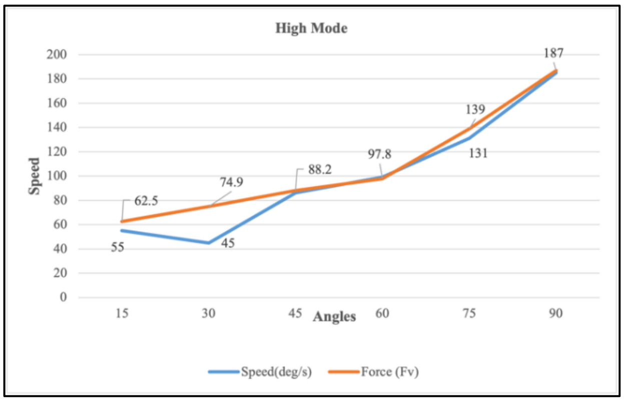


Fig. 12. High mode of AD

Fig. 12 shows the graph of force, speed, and angles at High Mode. When High Mode 3 was selected, the device began moving at high speed with high force, reaching the

90-degree angle. As the angle increased from 15° to 90°, the speed started at 42 deg/s, gradually increased to 111 deg/s at 75°, and peaked at 187 deg/s at 90°. Meanwhile, the force started at 62.5 Fv, steadily increased to 139 Fv at 75°, and reached 187 Fv at 90°. This demonstrated that both speed and force increased consistently with the increased angle, however the speed rised somewhat faster than force at higher angles.

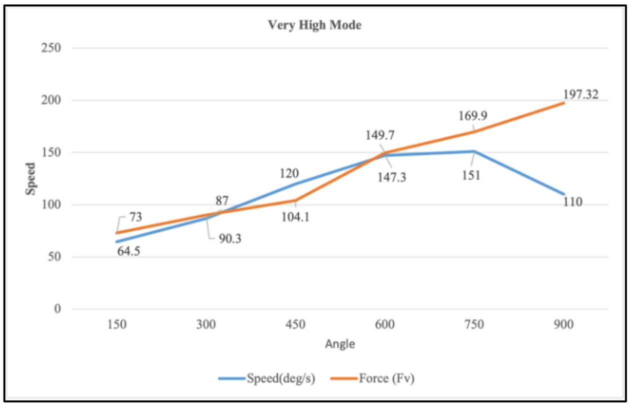


Fig. 13. Very high mode of AD

Fig. 13 shows the graph for Very High Mode 4. In this mode, the speed was very high, and the armrest moved quickly. In this mode, the device operated at or near its maximum output. Notably, the drop in speed at 90° despite rising force suggests either a limitation in motor torque or a lag in sensor response. This anomaly highlights the importance of evaluating both mechanical and sensor thresholds. From a therapeutic standpoint, Mode 4 may be reserved for final-stage rehabilitation or athletic patients, and should be used under supervision due to increased variability. As the angle increased from 15° to 90°, the speed started at 64.5 deg/s, peaked at 149.7 deg/s at 60°, and then decreased to 110 deg/s. Meanwhile, the force steadily increased from 73 Fv at 15° to 197.32 Fv at 90°. This suggests that while speed initially increased with the angle up to a point before decreasing, force continued to increase consistently across the entire range of angles. However, the last speed value appears to be incorrect, possibly due to human error, although all other values followed the increasing trend of force with angle.

V. DISCUSSION AND COMPARISION

The results from this study followed a similar pattern to what has been seen in previous research on 1-DOF rehabilitation devices for frozen shoulder. Like the studies by Kim et al. (2018) and Zhang et al. (2019), it was found that as the assisted device moved the arm at higher angles, both the force and speed required were increased. This supports the idea that gradually increasing resistance and movement speed can help improve joint mobility over time. One thing that sets this study apart is the detailed mechanical design and data recording. While earlier research focused mostly on patient outcomes like pain relief or range of motion, we looked closely at how the device actually performed at different angles and settings. This gave a clearer picture of how force and speed changed in real time, which could be useful when refining or adjusting device performance in the future.

That said, we did notice a few inconsistencies—especially at 90° in the Very High Mode, where the speed unexpectedly dropped. This was likely due to a sensor glitch or human error during data collection. It's a reminder that even with a well-functioning device, accurate measurement tools are just as important. In fact, the issue may also be linked to the sampling frequency of the sensors. In faster modes, where the device moves more quickly, a low sampling rate might miss sudden changes in speed or force, leading to incomplete or inconsistent data. Mode 3 pushed the motor towards its upper performance threshold. The force and speed values increased more aggressively, and small fluctuations began to appear in the data. These fluctuations could indicate either sensor limitations at higher speeds or minor mechanical inconsistencies. Clinically, this mode might be appropriate for strengthening or improving endurance, but caution is advised due to potential fatigue or overstimulation. Using a higher sampling frequency could help capture more accurate results and give a better understanding of how the device performs under different conditions. It's something worth improving in future versions of the setup.

VI. Conclusion

IN CONCLUSION, THE ASSISTED DEVICE FOR FROZEN SHOULDER REHABILITATION WAS DESIGNED TO FACILITATE THERAPEUTIC EXERCISES FOR PATIENTS WITH LIMITED UPPER LIMB MOBILITY. IT SUPPORTS UP TO 5.3% OF THE TOTAL BODY WEIGHT AND ACCOMMODATES A WIDE RANGE OF AGES. THE USER-FRIENDLY DESIGN ALLOWS THE DEVICE TO BE CONTROLLED WITH THE PATIENT'S NON-AFFECTED HAND, ENSURING SAFETY WITH AN IMMEDIATE STOP MECHANISM IN CASE OF DISCOMFORT. THE DEVICE OFFERS ONE DEGREE OF FREEDOM FOR INTERNAL AND EXTERNAL SHOULDER ROTATIONS ACROSS FOUR SPEED LEVELS, WHICH CAN BE SELECTED BASED ON MEDICAL RECOMMENDATIONS FOR OPTIMAL THERAPY. THIS ENABLES PATIENTS TO MANAGE THEIR REHABILITATION INDEPENDENTLY AT HOME, REGARDLESS OF AGE, GENDER, OR WEIGHT, WHILE ALSO ASSISTING PHYSIOTHERAPISTS IN DELIVERING PRECISE AND CONSISTENT THERAPY WITHOUT PHYSICAL INTERVENTION.

References

[1] K.-L. W. e. al, "Developing a Motion Sensor-Based Game to Support Frozen Shoulder Rehabilitation in Older Adults through a Participatory Design Approach," *Games for Health Journal,* 2024.

[2] Inam, Shehla, et al. "Enhancing Rehabilitation for Frozen Shoulder With a Two-Degree of Freedom Approach." 2024 International Conference on Robotics and Automation in Industry (ICRAI). IEEE, 2024.

[3] 이제우. Development and Evaluation of a Sliding Type Continuous Passive Motion Automation Device for Diagnosis and Rehabilitation of Frozen Shoulder. Diss. 서울대학교 대학원, 2023.

[4] Wu, Cheng-Ju, et al. "Efficacy of joint mobilization apparatus in treating frozen shoulder." Applied Sciences 11.9 (2021): 4184.

[5] C.-J. W. e. al, "Efficacy of Joint Mobilization Apparatus in Treating Frozen Shoulder," *Applied Sciences,* 2021.

[6] Y. C. e. al, "Effect of smartphone application-supported self-rehabilitation for frozen shoulder: a prospective randomized control study," *Clinical Rehabilitation,* 2018.

[7] N. C. e. al, "Frozen Shoulder Rehabilitation: Exercise Simulation and Usability Study," *SoICT '18: Proceedings of the 9th International Symposium on Information and Communication Technology* , 2018.

[8] T. S. e. al, "Mobile Phone–Supported Physiotherapy for Frozen Shoulder: Feasibility Assessment Based on a Usability Study," *JMIR Rehabil Assist Technol* , 2017.

[9] H. V. L. e. al, "Adhesive capsulitis of the shoulder: review of pathophysiology and current clinical treatments," *Shoulder & Elbow,* 2017.

[10] C.-E. C. &. C.-H. Chen, "The App Game Interface Design for Frozen Shoulder Rehabilitation," *Advances in Ergonomics Modeling, Usability & Special Populations* , 2016.

Improving Cardiovascular Disease Prediction Accuracy with Three-Way Decisions

1st Arshad Aziz
Dept. of Computing & Tech.
Abasyn Univ. Peshawar
Peshawar, Pakistan
gullsher210@gmail.com

2nd Hasham Khan
Dept. of Computing & Tech.
Abasyn Univ. Peshawar
Peshawar, Pakistan
khanhasham554@gmail.com

3rd M. Attaullah
Dept. of Computer Sci.
UET Lahore
Lahore, Pakistan
muhammadattaullah.pk@gmail.com

4th M.K. Afridi
Dept. of Computing & Tech.
Abasyn Univ. Peshawar
Peshawar, Pakistan
mohammad@abasyn.edu.pk

***Abstract*—Cardiovascular disease (CVD) remains a leading cause of mortality worldwide, necessitating accurate and early risk prediction to improve patient outcomes. Traditional CVD prediction models primarily rely on binary decision-making, which often struggles with uncertain or borderline cases, leading to misclassification and ineffective treatment strategies. This research proposes an advanced predictive model integrating machine learning algorithms with a three-way decision approach to enhance the accuracy and reliability of CVD risk assessment. The three-way decision model, rooted in rough set theory, categorizes decisions into three classes—acceptance, rejection, and deferment—allowing for more nuanced and informed predictions. Utilizing the Cleveland Heart Disease dataset, this study applies machine learning techniques such as Random Forest (97.14% accuracy), Logistic Regression (91.30% accuracy), Na¨ıve Bayes (88.24% accuracy), and Support Vector Machine (89.74% accuracy) to assess the effectiveness of the proposed model. The results demonstrate that integrating three-way decisions with machine learning enhances predictive performance, particularly for ambiguous cases, thereby improving clinical decision-making. However, the model's reliance on dataset quality and threshold selection presents certain limitations that need further exploration. This research contributes to the field of medical diagnostics by introducing an intelligent and flexible approach to CVD prediction, potentially reducing diagnostic errors and facilitating early interventions for high-risk patients.**

***Index Terms*—Cardiovascular Disease Prediction, Machine Learning, Three-Way Decisions, Rough Set Theory, Medical Diagnostics, Predictive Modeling**

I. Introduction

Cardiovascular disease (CVD) poses one of the major health concerns around the world. According to World Health Organization reports, CVDs, including heart attacks, strokes, and related conditions, result in approximately 18 million deaths each year, making them the main cause of death worldwide [1]. In particular, the causes of CVD-related mortality differ from one country to another [2]. These variations in death rates are mainly shaped by regional disparities in the prevalence of risk factors and the accessibility and effectiveness of acute healthcare services in recent years [3], [4], [5].

This increasing number of cases not only has a profound impact on individuals and their families, but also puts immense pressure on health systems. Early and precise identification of those at risk for CVD is essential to prevent severe outcomes, reduce treatment costs, and overall improve public health [6]. Traditional CVD risk prediction methods generally use binary decision-making approaches, assessing the likelihood of cardiovascular events based on specific thresholds, such as cholesterol levels, blood pressure, and other health indicators [7]. However, these models often oversimplify the complexity of medical diagnostics, especially when dealing with uncertain or incomplete patient data. In many cases, binary predictions can misclassify people with borderline or atypical risk factors, leading to inaccurate diagnosis and inadequate preventive measures [4].

Advances in machine learning (ML) offer promising avenues to improve CVD prediction. ML algorithms can process large, diverse datasets and find complex patterns which may not be found using conventional methods [8]. Using historical data, ML models adapt to intricate, non-linear relationships within the data. Techniques like Random Forest, Logistic Regression including Support Vector Machine (SVM) and Naive Bayes have shown potential for CVD prediction with high accuracy [9]. However, most ML models still use binary classification, lacking mechanisms to handle ambiguous cases. To address this limitation, a three-way decision model integrated with ML is proposed. This model, rooted in rough set theory, categorizes decisions into acceptance, rejection, and deferment, making it suitable for complex cases [10].

This paper is organized as follows: Section II provides a detailed description of the Cleveland Heart Disease dataset, along with exploratory data analysis. Section III reviews the related literature on CVD prediction. Section IV outlines the proposed methodology, detailing the integration of machine learning algorithms with the Three-Way Decision model, including the selection of classifiers, thresholding, and model architecture. Section V presents the experimental setup, performance evaluation metrics, and results obtained from applying various algorithms. Finally, Section VI discusses the implications of the findings, highlights the advantages and limitations of the proposed approach, and suggests directions for future research.

II. Dataset Description And Exploratory Data Analysis

The Cleveland Heart Disease dataset [17] was sourced from Kaggle and serves as the primary dataset for this study. The dataset comprising of 303 instances with 14 features, where each instance signifies a patient, and each attribute corresponds to a specific health-related feature. The main characteristics of this dataset are listed below:

A. Attributes Description

- **Age**: Age is a significant factor in heart disease risk assessment. As individuals grow older, the likelihood of developing cardiovascular conditions increases due to factors such as arterial stiffening, plaque buildup, and reduced heart efficiency. Older patients are at higher risk, making this feature essential for predictive modeling.
- **Sex**: Sex influences the prevalence and manifestation of heart disease. Males tend to develop heart disease at an earlier age compared to females, whereas postmenopausal women experience a rise in risk due to hormonal changes. This feature helps models understand gender-based differences in disease occurrence.
- **Chest Pain Type (CP)**: Chest pain is one of the most common symptoms of heart disease. This feature categorizes chest pain into four types: typical angina, atypical angina, non-anginal pain, and asymptomatic. Typical angina is often associated with coronary artery disease, whereas asymptomatic cases may indicate silent heart disease, making this feature a key determinant in diagnosis.
- **Resting Blood Pressure (Trestbps)**: This feature represents the patient's blood pressure in mmHg when at rest. Elevated resting blood pressure is a strong indicator of hypertension, which is a major risk factor for cardiovascular diseases. Consistently high blood pressure can lead to increased strain on the heart, raising the likelihood of heart disease.
- **Serum Cholesterol (Chol)**: Serum cholesterol levels reflect the total cholesterol present in the blood, measured in mg/dL. High cholesterol levels, particularly elevated LDL (low-density lipoprotein) cholesterol, contribute to plaque buildup in the arteries, increasing the risk of coronary artery disease and heart attacks.
- **Fasting Blood Sugar (Fbs)**: This binary feature indicates whether the patient's fasting blood sugar level is greater than 120 mg/dL. Elevated fasting blood sugar is often associated with diabetes, a significant risk factor for heart disease. Diabetic patients tend to have higher cardiovascular risks due to insulin resistance and blood vessel damage.
- **Resting Electrocardiographic Results (Restecg)**: Electrocardiography (ECG) measures the electrical activity of the heart. This feature categorizes ECG results into normal, abnormal ST-T wave changes, and left ventricular hypertrophy. Abnormal ECG readings can indicate heart conditions such as ischemia, arrhythmias, or structural abnormalities.
- **Maximum Heart Rate Achieved (Thalach)**: This feature records the highest heart rate achieved by the patient during physical exertion. A lower maximum heart rate may indicate poor cardiovascular fitness, whereas a high value suggests a healthy heart response. It is a crucial parameter in stress tests used for heart disease assessment.
- **Exercise-Induced Angina (Exang)**: Exercise-induced angina is a binary feature that indicates whether the patient experiences chest pain during physical activity. A positive result suggests reduced blood flow to the heart, often caused by blockages in the coronary arteries. This symptom is a strong predictor of heart disease.
- **ST Depression Induced by Exercise (Oldpeak)**: ST depression measures changes in the ST segment of an ECG during exercise. It is an indicator of myocardial ischemia, where insufficient blood flow to the heart leads to oxygen deprivation. A higher ST depression value signifies a greater likelihood of coronary artery disease.
- **Slope of the Peak Exercise ST Segment (Slope)**: The slope of the ST segment during peak exercise provides insights into the heart's response to physical activity. It is categorized into three types: upsloping, flat, and downsloping. A downsloping ST segment is associated with a higher risk of ischemic heart disease, while an upsloping pattern is usually considered normal.
- **Number of Major Vessels Colored by Fluoroscopy (Ca)**: Fluoroscopy is a diagnostic imaging technique used to visualize blood vessels. This feature represents the number of major coronary arteries (ranging from 0 to 4) that show narrowing due to blockages. A higher value indicates a greater extent of arterial obstruction, increasing the likelihood of heart disease.
- **Thalassemia (Thal)**: Thalassemia is a genetic blood disorder that affects hemoglobin production. This feature is categorized into three types: normal, fixed defect, and reversible defect. A fixed defect suggests permanent damage to the heart muscle, while a reversible defect indicates a temporary or treatable abnormality.

TABLE I: Dataset Attributes and Characteristics

ID	Attribute	Type	Values
1	Sex/gender	Discrete	Male = 1, female = 0
2	Age	Continuous	Age in years
3	Cp (chest pain)	Discrete	1=typical angina, 2=atypical angina, 3=non-anginal pain, 4=asymptomatic
4	RestBP (resting BP)	Continuous	90-200 mmHg
5	Chol (cholesterol level)	Continuous	126-564 mg/dL
6	Fbs (fasting Blood Sugar)	Discrete	Fasting blood sugar > 120 mg/dL (1=true, 0=false)
7	Restecg (resting Electrocardiography results)	Discrete	0=normal, 1=ST-T wave abn., 2=LV hypertrophy
8	Thalach (maximum heart rate)	Continuous	71-202 bpm
9	Exang (exercise-induced angina)	Discrete	1=yes, 0=no
10	Old peak ST (depression level)	Continuous	0 to 6.2
11	Slope (peak ex. slope)	Discrete	1=up sloping, 2=flat, 3=down sloping
12	Ca (fluoroscopy value)	Discrete	0 to 3
13	Thal (severity of chain pain or trouble)	Discrete	3=normal, 6=fixed defect, 7=reversible
14	Target	Discrete	1=yes, 0=no

These features collectively form the foundation of our heart disease prediction model. By applying exploratory data analysis, analyzing patterns and relationships among these attributes, it revealed that the age distribution ranges from 29 to 77 years, with a mean age of 54.36 years (see Figure 1 for age distribution), indicating that most patients fall into the middle-aged to elderly category.

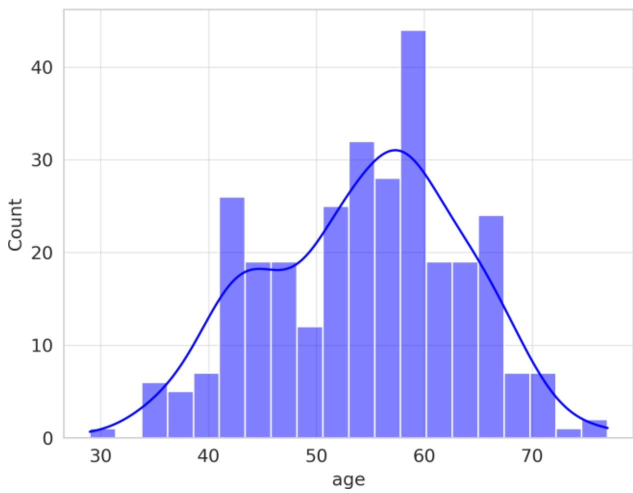


Fig. 1: Age Distribution of Patients

The sex distribution, as demonstrated in the figure 2, is skewed towards females, with 68.3% of the dataset consisting of female patients, confirming the general medical observation that female are more susceptible to cardiovascular diseases.

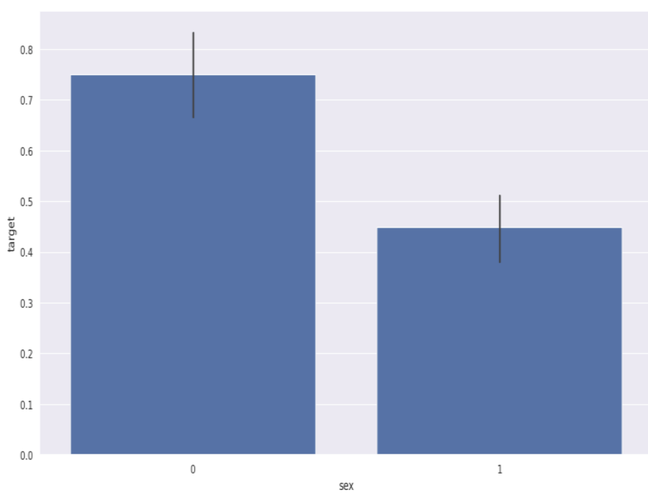


Fig. 2: Gender Distribution of Patients

In terms of clinical variables, the chest pain type (cp) variable ranges from 0 to 3, where higher values indicate more severe chest pain. Most patients experience mild to moderate chest pain, which is a common symptom of cardiovascular disease. Resting blood pressure (trestbps) values range from 94 to 200 mm Hg, with an average of 131.6 mm Hg, as shown in Figure 3, suggesting that many patients have elevated blood pressure levels, a known risk factor for heart disease.

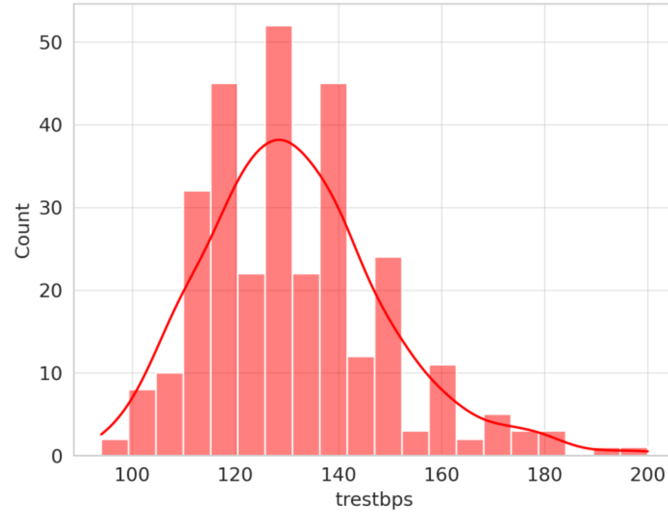


Fig. 3: Resting Blood Pressure Distribution

Similarly, cholesterol levels (chol) vary significantly, ranging from 126 mg/dL to 564 mg/dL, with an average of 246.26 mg / dL (see Figure 4). Although some patients exhibit dangerously high cholesterol levels, others fall within normal ranges, highlighting the various health profiles within the dataset.

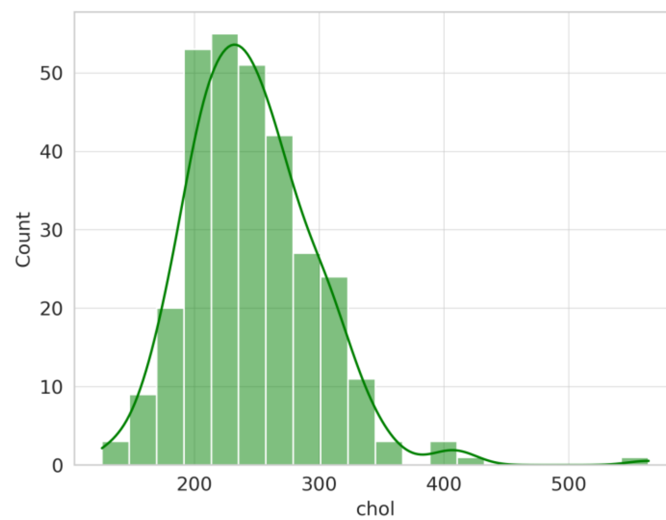


Fig. 4: Cholestrol Level Distribution

The maximum heart rate achieved (thalach) ranges from 71 to 202 beats per minute, with higher heart rates often associated with increased cardiovascular stress. Approximately 32. 7% of the patients experienced exercise-induced angina (exang = 1), which is an indicator of restricted blood flow during physical exertion. The ST depression (oldpeak) values range from 0 to 6.2, representing deviations in heart function due to exercise stress, and higher values typically indicate a greater probability of heart disease.

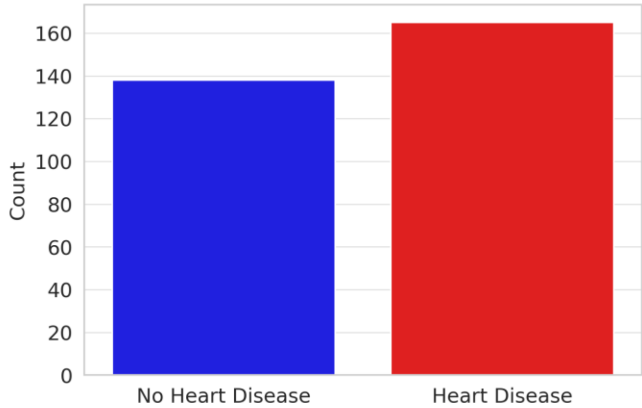


Fig. 5: Distribution of Heart Disease Cases

Finally, the target variable (heart disease diagnosis), as shown in figure 5, is fairly balanced, with 54.45% of patients diagnosed with heart disease (target = 1) and 45.55% classified as non-CVD patients (target = 0). This balance ensures that the dataset is well suited for machine learning applications without requiring oversampling or undersampling techniques.

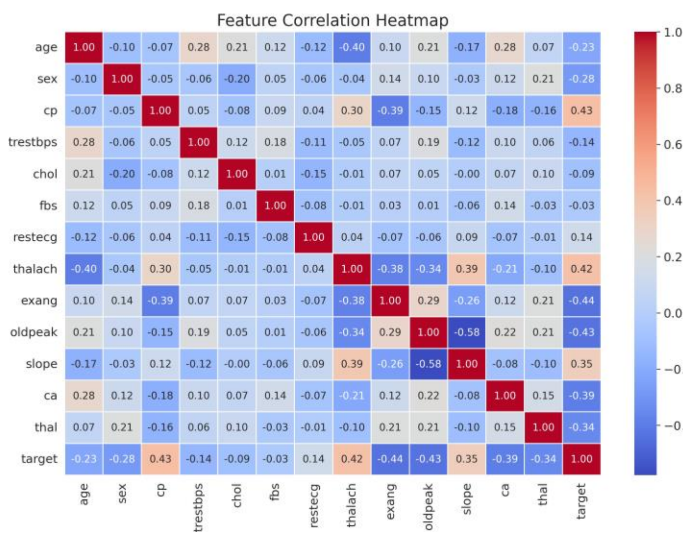


Fig. 6: Correlation Analysis of Dataset

The figure 6 shows the direction and degree of links between features by displaying a correlation heatmap for important numerical variables in the cardiovascular disease dataset. The correlation heatmap illustrates the relationships between key cardiovascular health indicators and the presence of disease. Among the features, chest pain type (cp) shows the strongest positive correlation with the target variable (0.43), indicating its strong association with heart disease. Similarly, maximum heart rate achieved (thalach) also has a notable positive correlation (0.42), suggesting its potential as a predictive feature. In contrast, exercise-induced angina (exang), ST depression (oldpeak), and the number of major vessels colored by fluoroscopy (ca) exhibit strong negative correlations with the target (-0.44, -0.43, and -0.39 respectively), implying a higher likelihood of disease with increased values in these attributes. Other features such as age, cholesterol, and resting blood pressure show weaker correlations with the target, though age is moderately negatively correlated with thalach (-0.40) and slightly positively correlated with ca and trestbps (0.28). Overall, the heatmap highlights which features are most informative for predicting cardiovascular disease, supporting effective feature selection in the implementation of the three-way decision model.

III. Related Work

Cardiovascular disease (CVD) is a prominent global cause of mortality, requiring the serious need for effective predictive models to detect high-risk affected persons early on. Traditional clinical models for CVD prediction often rely on binary decision-making systems, which may not sufficiently account for the complexity and variability of patient data. As a result, machine learning (ML) techniques have arose as encouraging alternatives for creating robust, flexible, and accurate models that can adapt to various patient populations and complex data patterns. This literature review provides an in-depth analysis of recent advancements in CVD prediction using machine learning, highlighting the evolution from conventional statistical methods to hybrid and ensemble ML approaches.

A study conducted in 2019 introduced a hybrid model combining Logistic Regression with Support Vector Machines (SVM) and Decision Trees to improve the of heart disease prediction accuracy. Traditional approaches typically employ individual algorithms that analyze medical data based on preset rules or thresholds. However, such methods often fall short when faced with complex, nonlinear patterns in patient data. In response, this study's hybrid model integrated multiple algorithms to enhance prediction accuracy, achieving 88.7% by leveraging the complementary strengths of each model [11].

The combination of Decision Trees, SVM, and Logistic Regression was a strategic choice, driven by the fact that Decision Tree algorithm is recognized for their ability to easily interpret and effectively manage categorical variables, while SVM excels in finding optimal decision boundaries for classification tasks. Logistic Regression, a foundational algorithm in statistical analysis, provides robust linear separation, especially in binary classifications. By integrating these methods, the study created a model that balances interpretability with predictive power, potentially offering clinicians a more reliable tool for early CVD identification. Future work from this study suggested enhancing the model further by integrating real-time data, which would allow the hybrid system to update predictions dynamically based on incoming patient information. This improvement could make the model more adaptable to real-world clinical environments, where patient data often change over time.

In 2023, a study presented at an IEEE conference addressed the limitations of traditional diagnostic methods, particularly their inability to handle ambiguous or borderline cases effectively. This study focused on developing an ensemble model that combined multiple machine learning approaches to improve the reliability of predictions for CVD diagnosis. By utilizing techniques like boosting and bagging, the likelihood of overfitting reduced by ensemble models and improve generalization on diverse datasets, achieving an accuracy of 91% [12].

The ensemble model effectively mitigates common issues in medical diagnostics, such as overfitting and instability in single algorithm models. By aggregating predictions from different algorithms, ensemble models provide a “voting” system that improves prediction robustness, especially in cases where individual models may disagree. This ensemble approach is particularly useful in healthcare, where variability in patient data can lead to challenges in classification. For instance, some patients may exhibit only mild risk factors, making binary classifications less reliable. Ensemble models handle these cases better by aggregating predictions across models, creating a consensus that can lead to more accurate diagnosis outcomes. The study also emphasized the potential for future improvements by utilizing larger datasets and deep learning techniques. By incorporating more data, ensemble models can capture even subtler patterns within patient records, potentially improving diagnostic accuracy. Additionally, deep learning models, known for their ability to handle vast amounts of complex data, could be layered into the ensemble framework, enabling it to interpret intricate relationships that may not be apparent with traditional ML algorithms alone.

Another 2023 study focused on addressing challenges in early-stage CVD detection through the use of advanced ML techniques, specifically Random Forest and Gradient Boosting models. These ensemble methods are well-known for their high accuracy and ability to manage both structured and unstructured data, making them ideal for CVD prediction tasks. The study achieved a 92% accuracy rate, underscoring the potential of these models to improve robustness in CVD detection [13].

An ensemble learning method, Random Forest based on decision trees, operates by building various decision trees during the stage of training and generates the mode of the class in outputs as the final prediction. Its advantage lies in its resistance to overfitting and capacity to handle large feature spaces, which is crucial for CVD prediction that often involves a wide range of physiological and demographic variables. Gradient Boosting employs a sequential approach where subsequent models are constructed to rectify the errors of their predecessors, leading to a progressive enhancement in predictive accuracy. This approach is particularly effective in handling imbalanced data, which is common in medical datasets where positive cases of CVD may be much fewer than negative cases.

The study proposed implementing these models in clinical settings to facilitate real-time diagnostics. Real-time application of these models could significantly improve patient care by providing immediate risk assessments, allowing for timely interventions. In practice, such models could be integrated with electronic health records (EHRs), where patient data is constantly updated, enabling the ML models to provide up-to-date predictions. This real-time capability, enabling the healthcare providers, to allocate resources more effectively and prioritize high-risk patients. In a 2022 study, researchers focused on creating a machine learning model that integrates conventional risk factors like age, blood pressure and cholesterol levels with modern ML techniques. Recognizing that certain physiological and lifestyle factors are established predictors of heart disease, the study applied a risk-factor-based ML approach, which achieved an accuracy of 85% [14]. This approach illustrates the effectiveness of combining conventional risk assessment factors with ML algorithms to enhance predictive accuracy. Techniques like Support Vector Machine (SVM), Random Forest, Logistic Regression, and Naive Bayes classifiers have been applied in CVD prediction with notable accuracy [15]. While these models are proficient at handling both linear and nonlinear data, they lack robust methods for managing uncertainty and making intermediate decisions. The three-way decision model, introduced by Yao within rough set theory, provides a novel framework for addressing this gap by classifying decisions into three distinct categories: acceptance, rejection, and deferment.

The three-way decision model offers a promising approach for integrating with machine learning in CVD prediction. By incorporating a middle-ground category (deferment), the model enables predictions to be postponed when uncertainty is high or when data is ambiguous. This deferment option allows for additional testing or further data analysis before arriving at a final decision, which is particularly useful in medical settings where misclassification could lead to serious consequences.

Integrating this model with existing ML algorithms could significantly enhance CVD prediction models by introducing flexibility in decision-making, especially in borderline cases. This hybrid approach not only has the potential to improve predictive accuracy but also adds a layer of interpretability and caution, which aligns well with the risk management requirements of medical diagnostics. Collectively, these studies demonstrate the transformative potential of machine learning for predicting cardiovascular disease. Hybrid and ensemble techniques have emerged as particularly promising, offering increased accuracy and robustness over traditional single-algorithm models. By integrating multiple algorithms or combining ML with established risk factors, these advanced models address many of the limitations inherent in binary or standalone predictive methods. Moreover, the application of ensemble techniques, particularly Gradient Boosting and Random Forest, has proven effective in managing diverse and complex patient data, yielding improved predictive accuracy.

As the field progresses, future research must emphasis on integrating these ML models into clinical practice through real-time data processing and exploring the inclusion of more varied datasets. The introduction of the three-way decisions model delivers a new avenue for managing ambiguity in medical diagnostics, promising more reliable and cautious predictions that could ultimately improve patient results and reduce the burden of CVD on healthcare systems worldwide.

IV. Methodology And Result Discussion

The Cleveland Heart Disease dataset [17] was sourced from Kaggle and serves as the primary dataset for this study. The dataset comprising of 303 instances with 14 features, where each instance signifies a patient, and each attribute corresponds to a specific health-related feature. The binary label indicating the presence and absence of cardiovascular disease is a target variable. Prior to model training, the data underwent data normalization, ensuring that numerical features such as age, cholesterol, and maximum heart rate were standardized to have a mean of zero and a standard deviation of one. Following this, the dataset is divided into two subcategories: 80% for training purpose and 20% for testing, warranting that the model could be both trained and validated effectively.The predictive models used in this study are based on well-established machine learning algorithms that have proven efficacy in medical predictions. Specifically, Naive Bayes, Decision Tree, Logistic Regression, and Random Forest models were selected on the pretext of their simplicity, interpretability, and proven accuracy in classification tasks. These models were trained on the preprocessed training dataset.

- **Logistic Regression (LR)**: This model was chosen for its ability to output probabilities, which is useful for the three-way decision model. It is well-suited for binary classification tasks and provides a solid baseline for evaluation with other models.
- **Naive Bayes (NB)**: Based on Bayes' Theorem, this probabilistic model works well with categorical features and is computationally efficient. It was included to assess the performance of a simpler, more intuitive model.
- **Decision Tree (DT)**: Decision Trees offer an interpretable structure, where splits in the data are made based on feature values. They were included to evaluate their effectiveness in identifying complex patterns in the dataset.
- **Random Forest (RF)**: Random Forest is an ensemble method which builds various decision trees and integrate their predictions, mitigating overfitting and enhancing predictive accuracy. This model was chosen for its robustness and high performance, especially in real-world applications.

The concept of three-way classification is derived from the theory of three-way decisions, which introduces an alternative approach to conventional binary classification. Unlike traditional ensemble and hybrid machine learning approaches that primarily aim to enhance predictive accuracy by combining multiple models, the proposed method introduces a fundamentally different paradigm through the integration of the Three-Way Decision (3WD) model. Ensemble methods and other hybrid systems improve performance by using the strengths of different models working together; however, they still follow a binary classification framework—compelling each instance to be labeled as either positive or negative. In contrast, the 3WD model introduces a third category: a deferment region for uncertain cases. This allows the system to refrain from making forced decisions when the confidence level is low. Such an approach closely aligns with real-world medical practices, where ambiguous cases typically require additional testing or expert review rather than immediate classification. By incorporating this deferment mechanism, the proposed 3WD-based framework enhances interpretability, reduces misclassification in borderline cases, and promotes more responsible and human-like decision-making. This sets it apart from conventional ensemble techniques. Let U represent the set of patients under evaluation, each characterized by A attributes, such as age, sex, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, resting electrocardiographic results, maximum heart rate, exercise-induced angina, ST depression, and other relevant cardiovascular indicators. A classification scheme divides these patients into a family of categories, denoted as $C^1, C^2, C^3, \ldots$, where each category represents a group of patients with similar characteristics based on their risk factors for heart disease.

In this study, we leverage three-way classification to improve heart disease prediction by addressing instances where model uncertainty prevents a definitive classification. Instead of assigning every patient to a positive (high-risk) or negative (low-risk) category, the three-way decision model introduces a deferred category, where uncertain predictions are set aside for further evaluation. This corresponds to cases where a model is unable to determine a patient's status with high confidence. Specifically, the confirmed category ($\mathrm{In}(C_k)$) consists of patients who are confidently classified based on their attributes exceeding a predefined threshold. The excluded category ($\mathrm{Out}(C_k)$) includes patients who clearly do not belong to the high-risk group based on the same evaluation criteria. Meanwhile, the uncertain category ($\mathrm{Pt}(C_k)$) represents cases where uncertainty exists, and these instances may be further analyzed using an additional classification model, such as a secondary machine learning algorithm.

The three-way classification framework aligns with our research methodology (see figure 7), where deferred cases from models (such as Logistic Regression, Na¨ıve Bayes, Decision Tree, and Random Forest) undergo an evaluation. This ensures that ambiguous cases receive more attention rather than being arbitrarily assigned to a category. Mathematically, the relationship between a patient o_i and a category C_k is defined by an evaluation function $e(Ck, o_i)$, which quantifies the degree of association between a patient's attributes and a particular risk group. Two thresholds (α, β)determine category membership:

$$\text{If } e(C_k, o_i) \geq \alpha \tag{1}$$

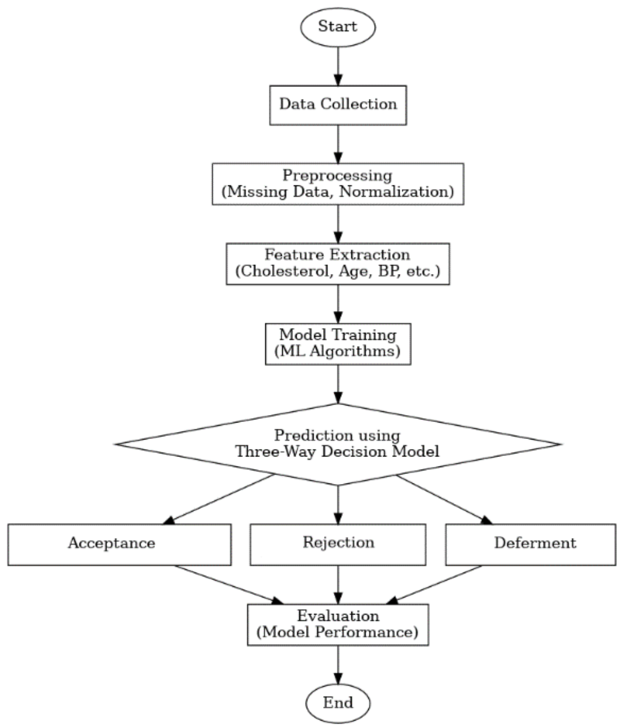
The patient is included in the high-confidence category (Confirmed category).

$$\text{If } e(C_k, o_i) \leq \beta \tag{2}$$

The patient is excluded from the category (Excluded category).

$$\text{If } \beta < e(C_k, o_i) < \alpha \tag{3}$$

The patient is assigned to the Uncertain category, requiring further analysis.



This structured approach enhances decision-making in heart disease prediction by ensuring that uncertain cases are systematically deferred instead of being incorrectly classified. The application of three-way classification in this research improves diagnostic reliability, reduces false positives and false negatives, and enhances interpretability in predictive modeling. The threshold values (α, β) are crucial parameters that influence model performance, and their optimization plays a vital role in determining how different patients are categorized. These thresholds define the boundaries for confident classification. The appropriate selection of these values significantly influences the model's performance, particularly in balancing sensitivity, specificity. To ensure the robustness of the model, the values of α nd β were not arbitrarily chosen. Instead, a grid search strategy was employed. For each pair of threshold values, we conducted 10-fold cross-validation to evaluate key performance metrics. This process helped identify the optimal α and β values that provided the best trade-off between classification confidence and diagnostic coverage. A higher α results in stricter acceptance, increasing the number of deferred cases but reducing false positives. Conversely, a lower β increases the likelihood of confident rejection but may lead to more false negatives if not chosen carefully. The final optimal thresholds were selected to minimize misclassification while maintaining a manageable deferment rate, ensuring that ambiguous predictions are treated with due caution rather than being forced into potentially incorrect categories. This threshold tuning strategy enhances the system's reliability in real-world medical decision-making, where the cost of misclassification can be high.By incorporating this method into our study, we effectively address the challenge of uncertainty in machine learning-based heart disease prediction.

V. Results and Discussion

The comparative analysis of four machine learning models—Logistic Regression, Na¨ıve Bayes, Support Vector Machine (SVM), and Random Forest—based on their performance in heart disease prediction. The evaluation considers key classification metrics: Precision, Recall, F1-score, and Accuracy, which assess each model's ability to correctly identify and classify heart disease cases.

Among the models tested, Random Forest demonstrates the highest effectiveness, achieving 100% precision, 93% recall, an F1-score of 0.97, and an overall accuracy of 97%. These results indicate that Random Forest is highly capable of distinguishing between positive and negative cases, reducing false positives while maintaining a high recall. This superior performance is likely due to its ensemble learning technique, which leverages multiple decision trees to improve generalization and reduce overfitting.

Both Logistic Regression and SVM show comparable classification performance, with a precision of 0.94, an F1-score of 0.89, and accuracy levels of 91% and 90%, respectively. These results suggest that these models provide a balanced trade-off between precision and recall, making them viable options for heart disease prediction. However, their slightly lower recall values compared to Random Forest imply that they may have a marginally higher rate of false negatives, which could be a concern in medical applications where missing a diagnosis has serious implications.

Na¨ıve Bayes, in contrast, records the lowest performance, with a precision of 0.91, recall of 0.83, an F1-score of 0.87, and accuracy of 88%. This suggests that Na¨ıve Bayes may struggle to capture complex patterns in the dataset, potentially due to its assumption that features are independent, which may not be realistic in heart disease prediction where several risk factors interact with each other. The lower recall value indicates a higher likelihood of false negatives, which may limit its effectiveness in medical diagnoses.

The results clearly indicate that the Random Forest model is the most suitable choice for heart disease prediction, as it outperforms the other models in terms of accuracy and recall. In a medical context, where correctly identifying patients at risk is critical, Random Forest's ability to minimize false negatives makes it the preferred model. However, Logistic Regression and SVM remain viable alternatives, especially in cases where interpretability and computational efficiency are important considerations.

The findings emphasize the potential of machine learning in improving diagnostic accuracy in healthcare. Future research could explore hyperparameter tuning, feature selection, and hybrid modeling techniques to further refine these models and enhance their predictive capabilities in real-world clinical settings.

After training and evaluation, the results from each model were compared to identify the best-performing algorithm for cardiovascular disease prediction. The Random Forest model showed the highest accuracy, followed by Logistic Regression and Naive Bayes. The Decision Tree model, while still performing decently, had lower accuracy than the other models. Random Forest, with its ensemble approach, provided the most robust predictions, particularly for challenging cases where other models had difficulty. Despite this, the inclusion of the three-way decision model helped to significantly improve the handling of ambiguous cases, especially for the Logistic Regression and Naive Bayes models.

TABLE II: Classification Performance Comparison

Model	Precision	Recall	Accuracy
Logistic Regression	0.94	0.85	0.91
Naive Bayes	0.91	0.83	0.88
Support Vector Machine	0.94	0.84	0.89
Random Forest	1.00	0.93	0.97

VI. Conclusion and Future Work

This study focused on enhancing the prediction of cardiovascular disease by applying a Three-Way Decision (3WD) model integrated with machine learning algorithms. By analyzing a well-structured dataset containing relevant clinical features, the research aimed to classify patients into three distinct decision regions: acceptance, rejection, and deferment—offering a more flexible and human-like approach to handling uncertainty in medical diagnosis. The results demonstrate that the 3WD framework improves interpretability and provides a more cautious decision mechanism in borderline cases, potentially reducing misdiagnosis compared to traditional binary classifiers. The model showed promising performance in identifying high-risk individuals, thereby supporting early intervention and better resource allocation in healthcare settings. However, the study is not without limitations. One of the primary limitations of the this research is that the dataset which we have taken is relatively small size, comprising only 303 patient records. This limited sample size constrains the learning capacity of machine learning models, increasing the risk of overfitting. Furthermore, the dataset lacks demographic diversity, as it predominantly represents patients from a specific geographic and clinical context in the United States. Important factors such as ethnicity, socioeconomic status, and lifestyle habits are entirely absent in the dataset. This restricts the applicability of the trained models to broader and more diverse populations, thereby limiting their real-world clinical relevance and generalization.

For future work, several improvements can be explored. Incorporating larger and more diverse datasets could enhance the model's generalization across different populations. Additionally, integrating advanced techniques such as deep learning or hybrid ensemble methods with the 3WD model might further boost predictive accuracy. Another potential direction is the application of feature selection and dimensionality reduction techniques to eliminate irrelevant attributes and improve efficiency. Finally, real-time implementation of the model in clinical decision support systems could be investigated to evaluate its practical utility in real-world healthcare environments. This research lays a strong foundation for building intelligent, interpretable, and reliable diagnostic tools for cardiovascular disease prediction.

References

[1] World Health Organization, "Cardiovascular diseases (CVDs)," [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds). [Accessed: Nov. 8, 2024].

[2] H. Ueshima, A. Sekikawa, K. Miura, T.C. Turin, N. Takashima, Y. Kita, M. Watanabe, et al., "Cardiovascular disease and risk factors in Asia: a selected review," Circulation, vol. 118, no. 25, pp. 2702–2709, 2008.

[3] [3][3] World Health Organization, "Noncommunicable diseases country profiles," 2018. [Online]. Available: https://www.who.int/nmh/countries/en/C. [Accessed: Nov. 8, 2024]

[4] G.L. Khor, "Cardiovascular epidemiology in the Asia-Pacific region," Asia Pac. J. Clin. Nutr., vol. 10, no. 2, pp. 76–80, 2001. doi: 10.1111/j.1440-6047.2001.00230.x.

[5] T. Ohira and H. Iso, "Cardiovascular disease epidemiology in Asia: an overview," Circ. J.: Off. Jpn. Circ. Soc., vol. 77, no. 7, pp. 1646–1652, 2013. doi: 10.1253/circj.cj-13-0702.

[6] J. Miah et al., "Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction," arXiv preprint arXiv:2311.00517, 2023.

[7] F. Hajiarbabi et al., "Heart disease detection using machine learning methods: a comprehensive narrative review," Journal of Medical Artificial Intelligence, vol. 5, 2022.

[8] D.-I. Kasartzian and T. Tsiampalis, "Transforming cardiovascular risk prediction: A review of machine learning and artificial intelligence innovations," Life, vol. 15, no. 1, p. 94, 2025

[9] K. G. Dinesh, K. Arumugaraj, K. D. Santhosh and V. Mareeswari, "Prediction of Cardiovascular Disease Using Machine Learning Algorithms," 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT), Coimbatore, India, 2018, pp. 1-7.

[10] Y. Yao, "Three-way decisions with probabilistic rough sets," Inf. Sci., vol. 180, no. 3, pp. 341–353, 2010.

[11] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in IEEE Access, vol. 7, pp. 81542-81554, 2019.

[12] T. Islam, A. Vuyia, M. Hasan and M. M. Rana, "Cardiovascular Disease Prediction Using Machine Learning Approaches," 2023 International Conference on Computational Intelligence and Sustainable Engineering Solutions (CISES), Greater Noida, India, 2023, pp. 813-819, doi: 10.1109/CISES58720.2023.10183490.

[13] Baghdadi, N.A., Farghaly Abdelaliem, S.M., Malki, A. et al. "Advanced machine learning techniques for cardiovascular disease early detection and diagnosis." J Big Data 10, 144 (2023)

[14] L. Wang, M. Zhao, and J. Xu, "Prediction of Coronary Heart Disease Using Risk Factor Categories and ML Algorithms," IEEE Journal of Translational Engineering in Health and Medicine, vol. 10, no. 4, pp. 285-292, 2022.

[15] D. Kumar, K. Arumugaraj, S. Kumar, and V. Mareeswari, "Prediction of cardiovascular disease using machine learning algorithms," Proc. IEEE Int. Conf. Current Trends Toward Converging Technologies, Coimbatore, India, 2018.

[16] M. K. Afridi, N. Azam, J. T. Yao, and E. Alanazi, "A three-way clustering approach for handling missing data using GTRS," International Journal of Approximate Reasoning, vol. 98, pp. 11–24, Apr. 2018.

[17] A. Rezaei, "Heart Disease Dataset," Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/arezaei81/heartcsv. [Accessed: 18-Feb-2025].

Machine Learning-Based Fish Species Recommendation Using Water Quality Parameters

1st Muhammad Owais Khan
Robotics Team
Hadaf Group of Colleges
Peshawar, Pakistan
mowaiskhandev@gmail.com

2nd Faheem Ul Haq
Robotics Team Lead
Hadaf Group of Colleges
Peshawar, Pakistan
engr.faheemulhaq@gmail.com

3rd Aasif Awan
Lecturer Biotechnology
Hadaf College of AHS
Peshawar, Pakistan
awan.aasif1@gmail.com

***Abstract*—The integration of machine learning (ML) in aquaculture enables data-driven fish species recommendations based on water quality parameters. Traditional fish farming faces challenges like manual monitoring, inefficient species selection, and unpredictable water conditions, leading to economic losses. This paper presents a software-based fish recommendation system using ML models to analyze seven key water parameters pH, Temperature, Turbidity, TDS, Dissolved Oxygen, Nitrate, and Ammonia. Various ML algorithms, including Random Forest, XGBoost, and SVM, were evaluated, with the optimized model achieving over 90% accuracy. A user-friendly graphical interface enables real time parameter input and fish species recommendation improving efficiency and sustainability in aquaculture.**

Keywords—Fish Farming, Machine Learning, Water Quality Analysis, Random Forest, XGBoost, SVM, Smart Aquaculture.

I. Introduction

Aquaculture has become an essential component of global food systems, contributing significantly to food security, nutrition, and economic development. As demand for fish continues to rise, modernizing aquaculture practices has become crucial. However, traditional fish farming remains largely dependent on manual water quality monitoring and farmer intuition, which often results in inefficient operations, inaccurate species selection, and vulnerability to environmental changes. These challenges can lead to poor yields, increased costs, and avoidable losses. Water quality parameters such as pH, temperature, total dissolved solids (TDS), turbidity, ammonia, dissolved oxygen, and nitrate directly influence fish health, growth, and survival. Monitoring these parameters manually is not only labor-intensive but also lacks the responsiveness required for real-time decision-making, especially in large-scale farming systems. To address these limitations, this study proposes a machine learning-based fish species recommendation system that predicts the most suitable species for a given aquatic environment. The objective is to support aquaculture decision-making by analyzing real-time water quality data using a variety of machine learning algorithms, including Random Forest, Decision Tree, XGBoost, K-Nearest Neighbors, Support Vector Machine, and Logistic Regression. The system integrates preprocessing techniques such as feature scaling and dataset balancing to enhance prediction accuracy. In addition, a graphical user interface (GUI) was developed to allow farmers and aquaculture professionals to input water parameters and receive instant fish species recommendations. By automating and optimizing the species selection process, this system aims to improve the efficiency and sustainability of aquaculture operations. The remainder of this paper includes a review of related work (Section 2), a detailed methodology (Section 3), results and performance analysis (Section 4), discussion of findings and limitations (Section 5), and conclusions (Section 6).

II. LITERATURE REVIEW

In recent years, the use of Internet of Things (IoT) and machine learning in aquaculture has attracted considerable interest. Several studies have explored real-time water quality monitoring, but most focus solely on water quality assessment rather than intelligent fish species recommendation. A major research gap exists in developing an AI-driven fish recommendation system based on real-time water quality indicators. This section reviews existing studies on IoT-based water monitoring systems and ML-based fish species recommendations, highlighting the key research gaps.

A. *IoT-Based Water Quality Monitoring Systems*

Multiple studies have investigated the use of IoT based system for real time monitoring of water quality in aquaculture. These systems typically consist of sensor networks, cloud storage, and remote data access. However, most lack intelligent decision-making capabilities for fish species selection. Cordova Rozas et al. proposed a cloud integrated water-monitoring framework consisting of five key stages: acquiring data, storing it in the cloud, managing databases, generating reports, and making predictions. Although this system effectively monitors water, quality variations it does not include an intelligent fish species recommendation feature [1]. Gao et al. developed an IoT enabled fish farming that facilitates continuous monitoring, storage, and prediction of water quality parameters. However, their research primarily emphasized fish movement tracking rather than providing fish species recommendations [2]. Nagayo et al. designed a solar-powered aquaponics setup with an Arduino-based control mechanism to maintain optimal temperature for fish growth. Despite offering an advanced water recirculating system, it lacks an AI-driven fish recommendation module [3]. Pasika et al. introduced a cost effective IoT driven monitoring framework designed to measure parameters such as temperature, ph, turbidity, humidity water level, with data storage on cloud. However, the system does not incorporate an intelligent decision making mechanishm for fish species selection [4].Huan et al. introduced an NB-IoT-based system to acquire real time data and improve aquaculture efficiency. Despite its stability and quick response, it does not incorporate fish recommendations [5].

1) Key Limitations in IoT-Based Systems

Lack of intelligent decision-making – Most IoT-based studies monitor water quality but do not recommend fish species. Absence of predictive analytics few systems provide predictive insights into how water quality fluctuations affects fish health. Limited AI integration most IoT systems collect data but do not use ML models for automated fish species selection. User experience challenges – Many systems lack interactive, GUI-based applications for fish farmers with limited technical expertise.

B. AI-Based Water Quality and Fish Species Recommendation Systems

While IoT systems help monitor water quality, few studies have explored AI-based fish species recommendation systems using machine-learning techniques. Chiu et al. developed an IoT integrated aquaculture framework utilizing deep learning forecast growth of California bass. However, the system primarily emphasizes feeding behaviors rather than fish species selection [6]. Niswar et al. developed an IoT-based monitoring system for crab farming using MQTT and LoRa-based sensor networks. While it enhances real-time monitoring, it does not provide fish species recommendations [7]. Billah et al. introduces a smart instrumentation based system for real time water quality assessment in fish farming. While it efficiently track parameters, it does not integrated fish species selection [8].Uddin et al**.** developed a fish survival prediction model using the Random Forest algorithm to analyze aquatic environmental parameters including temperature, turbidity, and pH. The model demonstrated high accuracy in predicting fish survival rates, offering valuable insights for aquaculture management. However, while the study effectively classifies survival conditions, it does not integrate real-time data collection or a comprehensive fish species recommendation system [9]. Abinaya et al. introduced a sensor driven IoT system incorporating Arduino and GSM modules to monitor water parameters including pH,DO,ammonia, and odors. While it alerts users via SMS, it lacks fish recommendations [10]. Islam et al. designed an ML driven model to predict fish species, utilizing J48, KNN, Random Forest and cart. The RF model attained 88.48% accuracy; however, it did not incorporate essential water quality parameters like TDS, nitrate, DO and ammonia which is vital for fish growth [11]. Hemal et al. introduced aqua bot, an intelligent IoT system designed to monitor pod water using machine learning. However, it lacks a complete fish species recommendation model and consider only a limited set of parameters [12].While these studies have made significant progress in water quality monitoring and fish survival prediction, they often lack comprehensive fish species recommendation models that consider multiple critical parameters. Most approaches focus on either tracking environmental conditions or predicting survival rates without integrating real-time decision making for aquaculture management. The following section outlines the key limitations of these existing methods, highlighting the gaps that need to addressed of existing methods

1) Key Limitations in AI-Based Systems

Limited integration of machine learning models most studies use only a few water quality parameters for prediction. Absence of real-time predictive analytics few models can anticipate water quality changes and their impact on fish species. Lack of renewable energy integration some studies explore solar-powered solutions, but their application in smart fish farming is limited. No user-friendly interfaces Many AI-based models lack interactive GUIs to help non-technical users interpret results.

C. Identified Research Gaps

Based on the reviewed literature, the main research gaps in existing fish farming solutions are Lack of AI driven decision support systems while real time IoT monitoring exists, few studies integrate ML models for fish species recommendation. Limited feature integration most AI models consider only 3-4 water quality parameters, whereas additional parameters (TDS, DO, nitrate, ammonia) could improve prediction accuracy. No real-time predictive capabilities most studies lack real-time forecasting of water quality changes and their impact on fish growth. Absence of renewable energy solutions few studies explore solar-powered smart fish farming, limiting sustainability. User interface challenges many systems lack interactive GUIs to help farmers with limited technical expertise.

Our research aims to bridge these gaps by developing an ML-based fish species recommendation system that uses seven water parameters (pH, Temperature, Turbidity, TDS, DO, Nitrate, Ammonia) for accurate predictions. Implements advanced ML models (Random Forest, XGBoost, SVM, KNN, etc.).Achieves over 90% accuracy in fish species prediction. Offers an intuitive graphical interface, enabling farmers to enter water quality data and obtain tailored recommendations.

III. METHODLOGY

In this section, we present the methods employed in developing the machine learning-based fish species recommendation system. It includes details about the dataset, preprocessing techniques, machine learning models, performance evaluation metrics, and the development of the graphical user interface (GUI).

A. Key Water Parameters

We have taken into consideration seven key water quality parameters to determine the suitability of water for different fish species. These parameters are temperature, turbidity, pH, total dissolved solids (TDS), dissolved oxygen (DO), ammonia, and nitrate. Based on these parameters, our system provides fish species recommendations that are best suited for the given water conditions. The reference values for these parameters are presented in Table 1. Any significant deviation from the optimal ranges can negatively affect fish health and survival, making it essential to select the right species for specific water conditions.

TABLE I. OPTIMAL RANGES OF WATER QUALITY PARAMETERS

Water Quality Parameter	Value
pH Level	6.5-8.5
Temperature	Between 25°C and 32°C or above 20 °C.
Turbidity	30 cm to 80 cm
Dissolved Oxygen (DO)	Greater than 5 mg/L
Total Dissolved Solids (TDS)	Up to 400 mg/L
Nitrate	0-100 mg/L
Ammonia	0-0.2 mg/L

Each of these parameters is essential in evaluating water suitability for fish farming:
Temperature affects fish metabolism, growth, and oxygen availability. Extreme temperatures can cause stress and even lead to fish mortality.

- pH influences water acidity or alkalinity, affecting fish health. Most fish species thrive within a pH range of 6.5 to 8.5, while extreme values can be harmful or fatal.
- Turbidity indicates water clarity. High turbidity can reduce light penetration, disrupt photosynthesis, and create unfavorable conditions for fish.
- Dissolved Oxygen (DO) is critical for fish survival, with levels dropping below 5 mg/L potentially leading to stress and reduced growth.
- TDS (Total Dissolved Solids) represents minerals, salts, and organic matter in water. High TDS levels can disrupt fish osmoregulation.
- Nitrate and Ammonia are nitrogen-based compounds. High nitrate levels (>100 mg/L) can lead to excessive algal growth, while ammonia is toxic to fish even at low concentrations (>0.2 mg/L).

B. Data Collection

This study utilizes the Real-Time Pond Water Dataset for Fish Farming, sourced from Kaggle [13] originally collected by the Faculty of Fisheries at the University of Dhaka, Bangladesh. The dataset consists of 591 samples with four primary features: temperature, turbidity, pH, and fish species. The independent variables pH (91 unique values), temperature (51 unique values), and turbidity (108 unique values) reflect diverse water quality conditions. The dependent variable includes 11 fish species such as Katla, Song, Prawn, Rui, Koi, Pangas, Tilapia, Silver Carp, Karpio, Magur, and Shrimp. To extend the dataset and enable more accurate fish species predictions, additional water quality parameters Total Dissolved Solids (TDS), Dissolved Oxygen (DO), Ammonia, and Nitrate were incorporated through synthetic data generation. These parameters were not originally present in the dataset but are widely recognized as essential indicators for aquaculture suitability [14][15]. The absence of these features could limit the model's ability to generalize across realistic aquaculture scenarios. Synthetic values were generated based on standard ranges reported in aquaculture literature. For instance, DO values were sampled between 4 and 10 mg/L, which supports optimal respiration for most freshwater fish species. TDS values were generated between 200 and 500 mg/L, ammonia between 0 and 1.5 mg/L, and nitrate between 0 and 50 mg/L—ranges consistent with those used in fish farming management guidelines [14]. A practical example: for a given row with pH 7.8, temperature 29°C, and turbidity 7 NTU (favoring Tilapia), a synthetic DO value of 6.9 mg/L and TDS value of 350 mg/L were generated using a uniform distribution in Python. These values were generated using numpy.random.uniform() to reflect real-world variability while ensuring ecological validity. By enhancing the dataset with realistic synthetic data, the feature set was expanded, resulting in improved model learning and robustness. The distribution of samples by fish species is provided in Table 2.

TABLE II. DISTRIBUTION OF FISH SPECIES IN DATASET

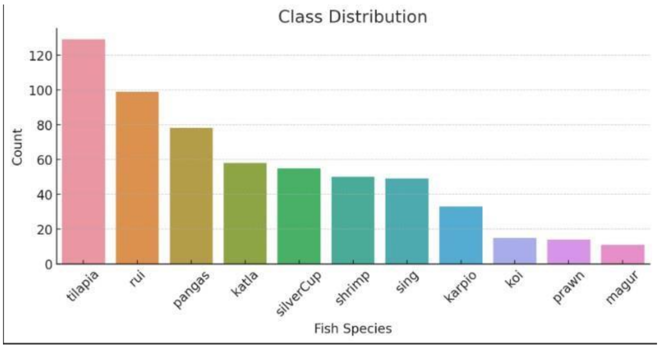
Class Label	Number
Tilapia	129
Rui	99
Pangas	78
Katla	58
Silver Cup	55
Shrimp	50
Sing	49
Karpio	33
Koi	15
Prawn	14
Magur	11

Table 2 presents the distribution of fish species in the dataset, revealing a noticeable class imbalance. Tilapia (129), Rui (99), and Pangas (78) are the most represented species, making them the majority classes. In contrast, species like Koi (15), Prawn (14), and Magur (11) have significantly fewer samples, making them minority classes. This imbalance can lead to biased model predictions, where the classifier favors majority classes while misclassifying underrepresented species. To address this issue and improve prediction accuracy for all fish species, techniques such as SMOTE are essential for balancing the dataset.

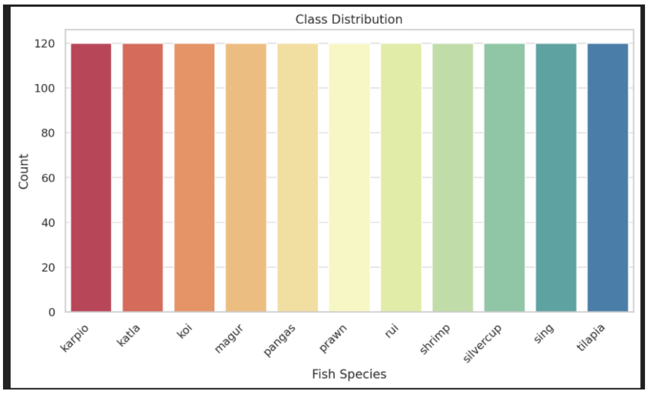
C. DATA PREPROCESSING

To ensure the reliability of model predictions, it is essential to preprocess the dataset by handling missing values, addressing class imbalance, and normalizing data. The following section details these preprocessing techniques and their impact on model performance.
1. Handling Missing Data: Any missing values in the dataset were addressed using mean or median imputation to maintain data integrity.
2. Splitting the Dataset: The dataset was divided into two sets: 80% for training data and 20% testing data allowing an effective evaluation of the model's predictive performance.
3. Balancing the Dataset: The dataset was checked for class imbalance, and SMOTE (Synthetic Minority Over-sampling Technique) was applied. This method generates synthetic

samples for underrepresented classes, improving model performance on imbalanced data. These preprocessing steps ensure that the dataset is well prepared for model training, minimizing biases and improving predictive accuracy. By handling missing values, we maintain data consistency, while proper dataset splitting allows the model to generalize effectively. Additionally balancing the dataset using SMOTE helps address the issue of class imbalance, ensuring that minority classes are adequately represented in training. These refinements collectively enhance the robustness of the machine-learning model, leading to more reliable fish species recommendations.



a) Pre-SMOTE



b) Post-SMOTE

Fig. 1. Fish Species Distribution: Pre SMOTE vs Post SMOTE.

Figure 1 illustrates the distribution of fish species before and after applying the Synthetic Minority Over-sampling Technique (SMOTE). In Figure 1(a), the dataset is imbalanced, with certain fish species, such as Tilapia and Rui, having significantly more samples compared to others like Prawn and Magur. This imbalance can negatively impact the performance of machine learning models by causing them to favor majority classes. In Figure 1(b), after applying SMOTE, the dataset is balanced, meaning all fish species have an equal number of samples. SMOTE achieves this by generating synthetic data for the underrepresented classes, improving model performance by reducing bias and enhancing generalization.

D. Implemented Machine Learning Algorithm

This study utilizes six machine-learning models to predict the most suitable fish species based on various water quality parameters. The implemented models include Random Forest (RF), Extreme Gradient Boosting (XGBoost), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree (DT), and Logistic Regression.

1. Random Forest (RF): Implemented with 100 decision trees and a random state of 42, ensuring consistency in results through an ensemble based decision-making approach.
2. Extreme Gradient Boosting (XGBoost): Configured with optimized hyperparameters using GridSearchCV, employing a learning rate of 0.1, max depth of 6, and 100 estimators.
3. K-Nearest Neighbors (KNN): The optimal k-value was determined as seven after testing multiple values.
4. Support Vector Machine (SVM): Configured with a regularization parameter (C) of 10 and a gamma value of 0.1, ensuring consistent results by setting a fixed random seed of 42.
5. Decision Tree (DT): Configured with a maximum depth of five and a random state of 42.
6. Logistic Regression (LR): Configured with L2 regularization (Ridge), a learning rate of 0.1, and a random state of 42.

These models were chosen for their effectiveness in accurately predicting fish species based on water quality parameters and they were fine-tuned using hyperparameter optimization techniques to improve predictive accuracy. The selection of the final model was determined by comparing performance metrics, including accuracy, precision, recall and F1 score.

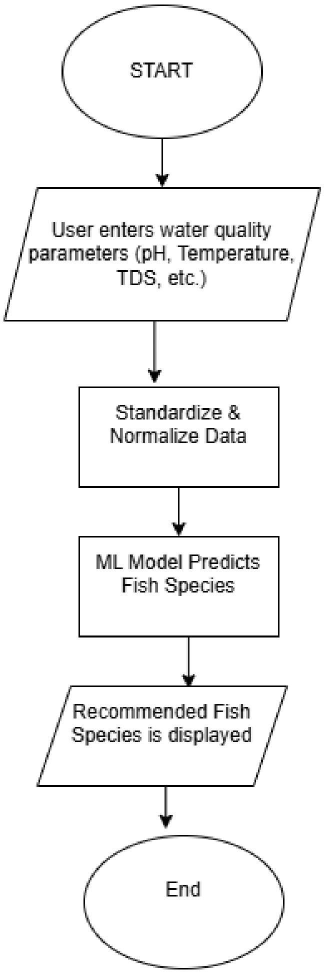


Fig. 2. Flow Diagram of ML Process.

Figure 2 (Flow Diagram) illustrates a machine learning based fish species recommendation system developed to assist in aquaculture. The process starts with the user entering key water quality factors such as pH, temperature, turbidity, total dissolved solids (TDS), dissolved oxygen, nitrate, and ammonia. These values undergo standardization and normalization to maintain consistency and improve model accuracy. After preprocessing, the data is provided to into a trained ML model, which analyzes the parameters and predicts the most appropriate fish species for aquaculture. The system processes this information and displays the recommended fish species to the user through a user-friendly interface. By leveraging machine learning, this system enhances decision-making in aquaculture, reduces reliance on manual monitoring, improves efficiency, and promotes sustainable fish farming practices.

E. Performnace Evaluation

The trained models were evaluated utilizing the following metrics:

- Accuracy – The percentage of correctly classified instances.
- Precision – The ratio of correctly predicted positive instances to the total predicted positive instances.
- Recall – The ratio of correctly predicted positive instances to total actual positive instances.
- F1-Score – The harmonic mean of precision and recall, offering a balanced assessment of model performance. Among all models, XGBoost achieved the highest accuracy, exceeding 90% and was selected for final deployment.

F. GUI Design

A simple GUI-based application was developed to take seven-water quality parameters pH, Temperature, Turbidity, TDS, Dissolved Oxygen, Nitrate, and Ammonia as input. The trained machine-learning model processes these inputs and provides a fish species recommendation. The interface ensures easy data entry and quick predictions for users.

IV. RESULTS

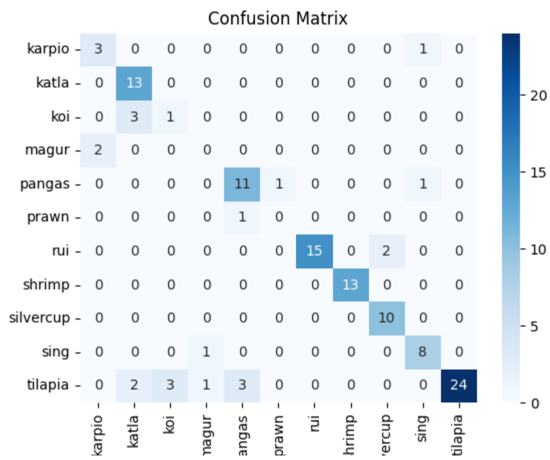
A. Evaluation Metrices For Model Performance

To assess the performance of the classification models, several key evaluation metrics were employed. One of the most fundamental tools in classification tasks is the confusion matrix, which provides a comparison between predicted and actual values. The confusion matrix consists of four essential components: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP represents the number of positive instances correctly identified, whereas FP denotes negative instances incorrectly classified as positive. Similarly, TN refers to correctly identified negative instances, while FN represents positive instances that were misclassified as negative. From this matrix, multiple performance metrics are derived to measure model effectiveness.

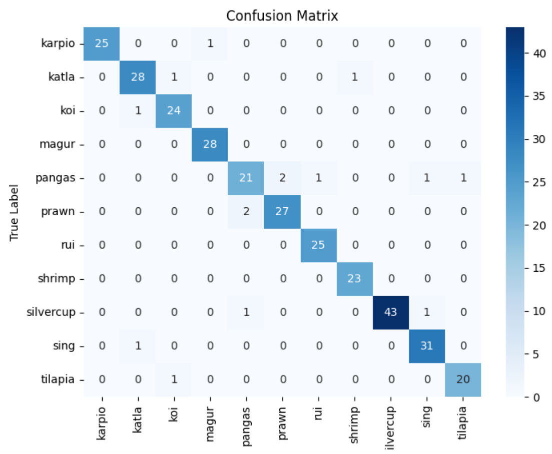
1. Accuracy determines the percentage of correctly classified instances in the dataset.
2. Precision evaluates the proportion of correctly predicted fish species among all positive predictions.
3. Recall (Sensitivity) measures the model's ability to correctly identify fish species when they are present.
4. F1-Score represents the harmonic mean of Precision and Recall, provides a balance between the two.
5. Matthews Correlation Coefficient (MCC) Assesses overall classification quality by incorporating all confusion matrix components.
6. ROC-AUC (Receiver Operating Characteristic - Area under Curve) visualizes the trade-off between TPR and FPR across thresholds, uses a one-vs-all strategy for multi-class classification, and a higher AUC score indicates better model performance.

B. Machine Learning Model Result Comparison

We have applied six ML algorithms to perform the prediction of suitable fish for individual ponds based on several parameters, such as the accuracy, precision, recall, F1-score, and ROC curve. We first evaluated the performance of these models without balancing the dataset. After this, we also evaluated the models after applying the balancing technique.



(a) Pre SMOTE



(b) Post SMOTE

Fig. 2. Confusion matrix comparison of ensemble (Pre SMOTE vs Post SMOTE).

Figure 3 (a) Before SMOTE highlights class imbalance in the dataset. The model performs well for majority classes like Tilapia (24), Rui (15), and Shrimp (13) but struggles with minority classes such as Prawn and Magur, leading to frequent misclassifications. Some classes, like Karpio and Koi, also show lower prediction counts. This imbalance affects the overall performance and justifies the need for SMOTE to enhance prediction accuracy for underrepresented fish species. Figure 3 (b) After SMOTE demonstrates a significant improvement in classification performance across all fish species. Unlike the Before SMOTE matrix, where minority classes were often misclassified, this matrix shows more balanced predictions, indicating the effectiveness of SMOTE in addressing class imbalance.

Key improvements:

• Better classification of previously underrepresented species (e.g., Magur, Prawn, and Pangas now have strong diagonal values).

• Fewer misclassifications across all classes, indicating that the model has learned better decision boundaries.

• Higher overall accuracy as false positives and false negatives have reduced compared to the previous model.

TABLE III. ML ALGORITHM PERFORMANCE EVALUATION BEFORE SMOTE.

BEFORE APPLYING SMOTE					
ALGORITHM	ACC	PRE	REC	F1	MCC
XG	0.79	0.83	0.80	0.80	0.77
RF	0.77	0.78	0.77	0.77	0.74
DT	0.75	0.79	0.76	0.76	0.73
SVM	0.59	0.61	0.60	0.58	0.54
KNN	0.49	0.53	0.50	0.5	0.42
LR	0.52	0.50	0.53	0.51	0.45

Table 3 presents the performance comparison of various machine-learning algorithms without applying SMOTE. The results indicate that XGBoost (XG) achieved the highest performance, with 79% accuracy, 83% precision, 80% recall, 80% F1-score, and 77% MCC making it the best-performing model. Random Forest (RF) followed closely, obtaining 77% accuracy, 78% precision, 77% recall, 77% F1-score, and 74% MCC. The Decision Tree (DT) model performed moderately well, achieving 75% accuracy, whereas SVM, KNN, and Logistic Regression (LR) demonstrated lower performance levels, with LR showing the weakest results (52% accuracy, 50% precision, 53% recall, 51% F1-score, and 45% MCC). These results suggest that ensemble models like XGBoost and Random Forest outperform other models in this scenario.

TABLE IV. ML ALGORITHMS PERFORMANCE EVALUATION AFTER SMOTE.

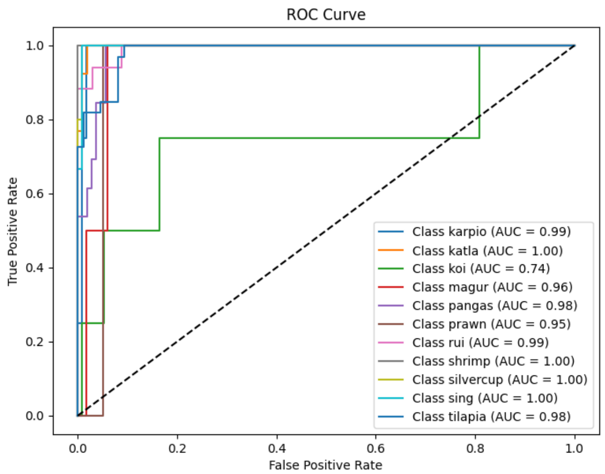
AFTER APPLYING SMOTE					
ALGORITHM	ACC	PRE	REC	F1	MCC
XG	0.968	0.968	0.968	0.968	0.968
RF	0.966	0.970	0.970	0.970	0.960
DT	0.902	0.900	0.900	0.900	0.890
SVM	0.842	0.850	0.840	0.840	0.830
KNN	0.840	0.845	0.840	0.840	0.830
LR	0.644	0.650	0.650	0.620	0.610

Table 4 presents the performance metrics of various machine learning models after applying SMOTE, the XGBoost (XG) model achieved the highest performance, with 96.8% accuracy, 96.8% precision, 96.8% recall, a 96.8% F1-score, and a 96.8% MCC. The Random Forest (RF) model followed closely, achieving 96.6% accuracy, 97% precision, 97% recall, a 97% F1-score, and a 96% MCC. Among other models, Decision Tree (DT) achieved 90.2% accuracy, while Support Vector Machine (SVM) and K-Nearest Neighbors (KNN) performed similarly with 84.2% and 84% accuracy, respectively. Logistic Regression (LR) exhibited the lowest performance, with 64% accuracy, 65% precision, 65% recall, a 62% F1-score, and a 61% MCC.

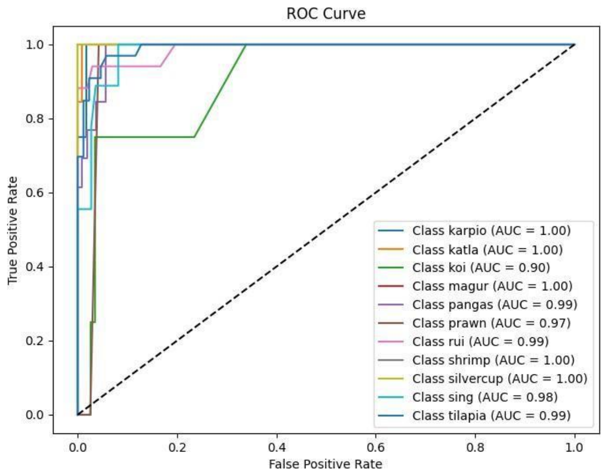
The ROC curve was also used to compare model performance, applying the One-vs-Rest method for multi-class classification. The XGBoost model demonstrated the best differentiation between classes, confirming its effectiveness in fish species classification. These results indicate that XGBoost is the most effective model for fish species classification after SMOTE, followed by Random Forest, while Logistic Regression remains the least effective in handling this classification problem.

Figure 4 (a: ROC Curve for XGBoost) illustrates the performance of the XGBoost model before applying SMOTE. The ROC curve represents the trade-off between the true positive rate and false positive rate for each fish species classification. While the model performs well for certain species, achieving AUC values close to 1.0, other species exhibit lower AUC scores, indicating difficulties in accurate classification. This discrepancy arises due to the imbalanced nature of the dataset, where some fish species have significantly fewer samples compared to others. The model struggles to learn distinctive patterns for these minority classes, leading to misclassifications. This demonstrates the limitations of training on an imbalanced dataset, as the model tends to favor the majority classes, reducing its overall reliability in classifying underrepresented species. The impact of this imbalance will be analyzed further in Figure 4 after the application of SMOTE to assess improvements in classification performance.

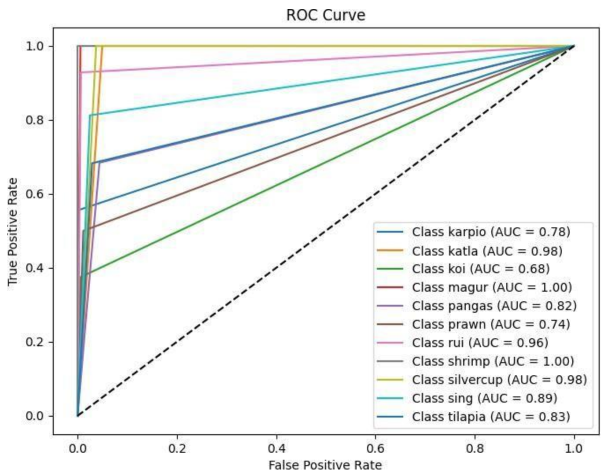
Figure 5 (a: ROC Curve for XGBoost) displays the ROC curve of the XGBoost model after applying SMOTE to balance the dataset. Compared to Figure 4, the ROC curves show a noticeable improvement in classification performance across different fish species. The AUC values for previously underrepresented classes have increased, indicating that the model is now better at distinguishing between species. This improvement results from SMOTE generating synthetic samples for minority classes, allowing the model to learn more patterns that are representative. By addressing class imbalance, SMOTE reduces the model's tendency to favor majority classes, leading to a more balanced classification performance. The enhanced ROC curve demonstrates the effectiveness of oversampling in improving model generalization, ensuring more predictions that are reliable across all fish species. The comparative analysis of ROC curves for various models further highlights the effectiveness of SMOTE in improving classification performance. While XGBoost and Random Forest show significant enhancements in their AUC scores after addressing class imbalance, models like Decision Tree and SVM exhibit moderate improvements.However,Logistic Regression still struggles to differentiate between fish species due to its linear nature, reinforcing its limitation in handling complex patterns. These findings emphasize the importance of selecting robust machine learning models and applying appropriate preprocessing techniques to ensure accurate and reliable fish species classification. The following figures illustrate the ROC curves of different models, providing a visual representation of their classification capabilities before and after SMOTE application.



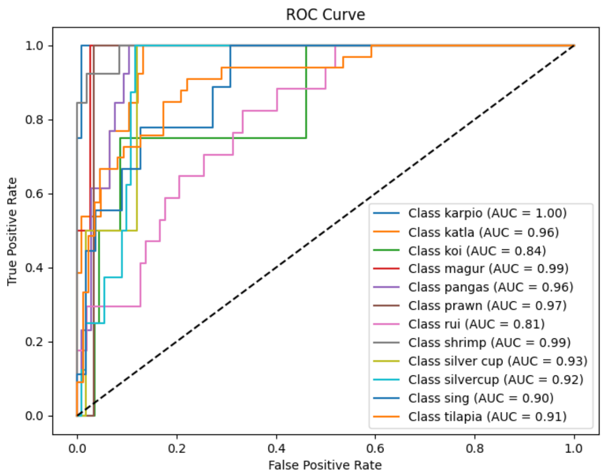
(a) ROC Curve for the XGBoost Model



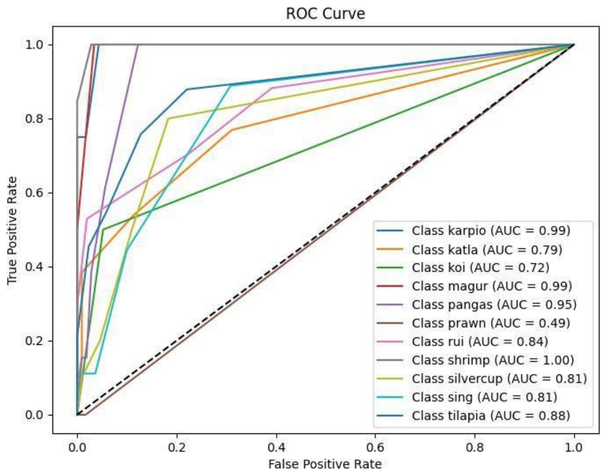
(b) ROC Curve for the Random Forest Model



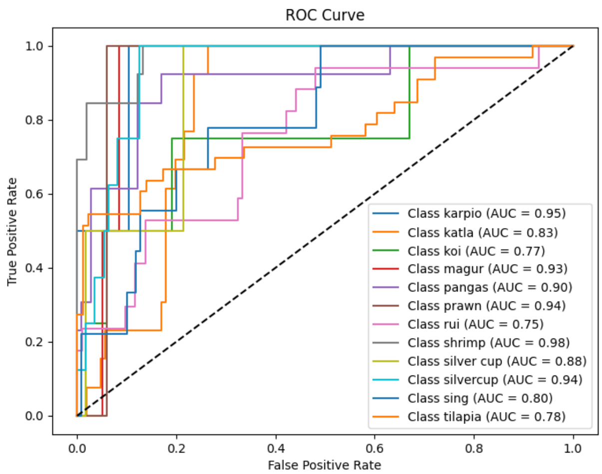
(c) ROC Curve for the Decision Tree Model



(d) ROC Curve for the SVM Model

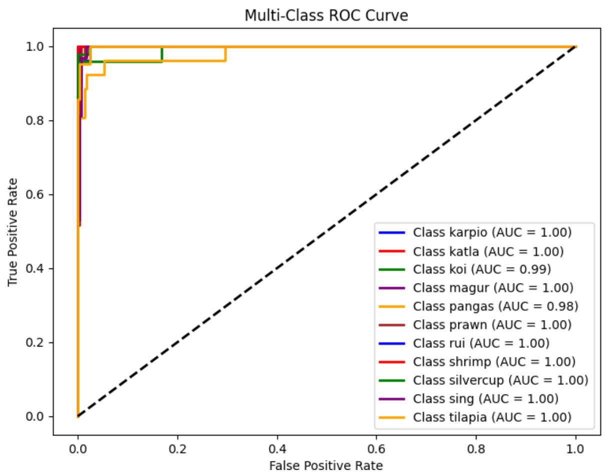


(e) ROC Curve for the KNN Model

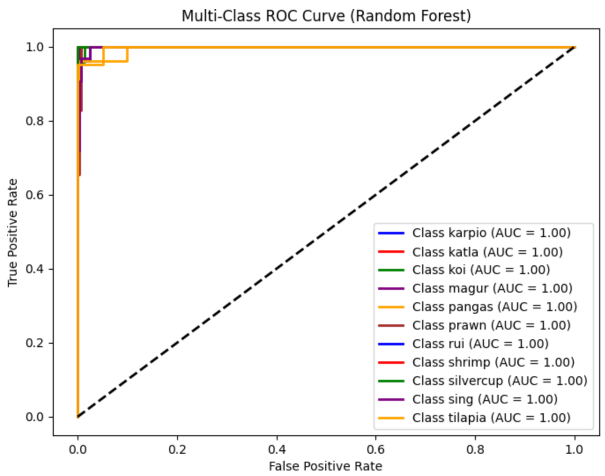


(f) ROC Curve for the LR Model

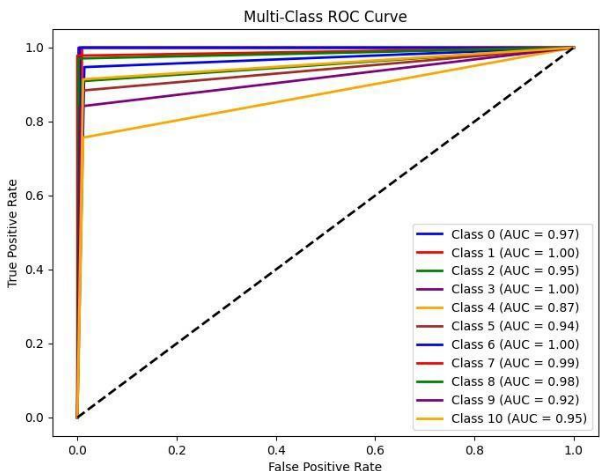
Fig. 3. ROC Curves of Various ML Models before SMOTE Application.



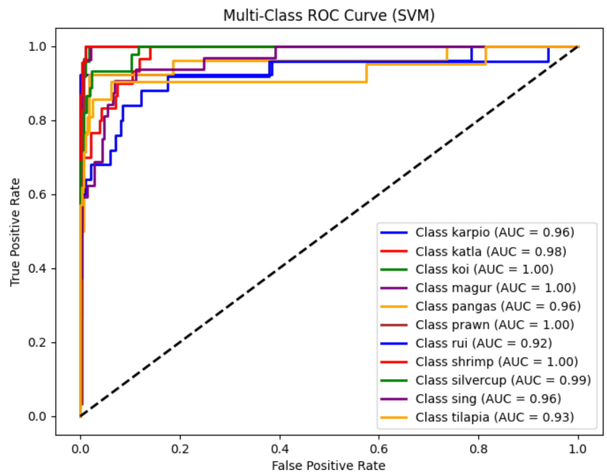
(a) ROC Curve for the XGBoost Model



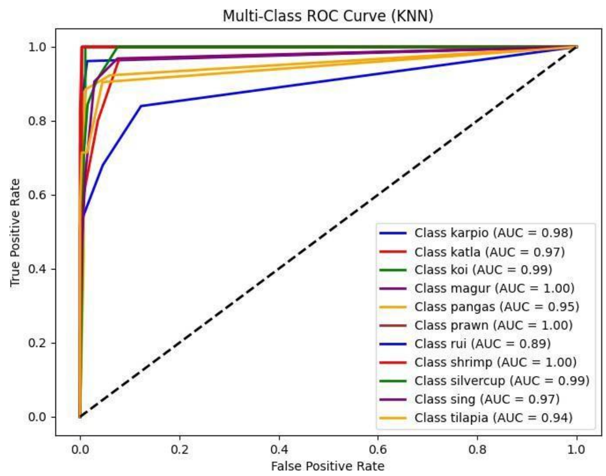
(b) ROC Curve for the Random Forest Model



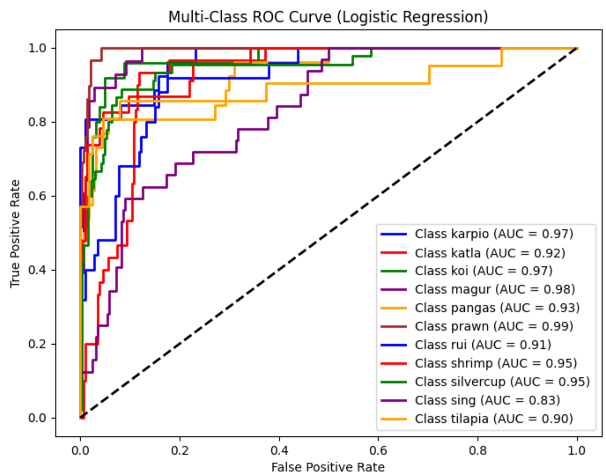
(c) ROC Curve for the Decision Tree Model



(d) ROC Curve for the SVM Model



(e) ROC Curve for the KNN Model



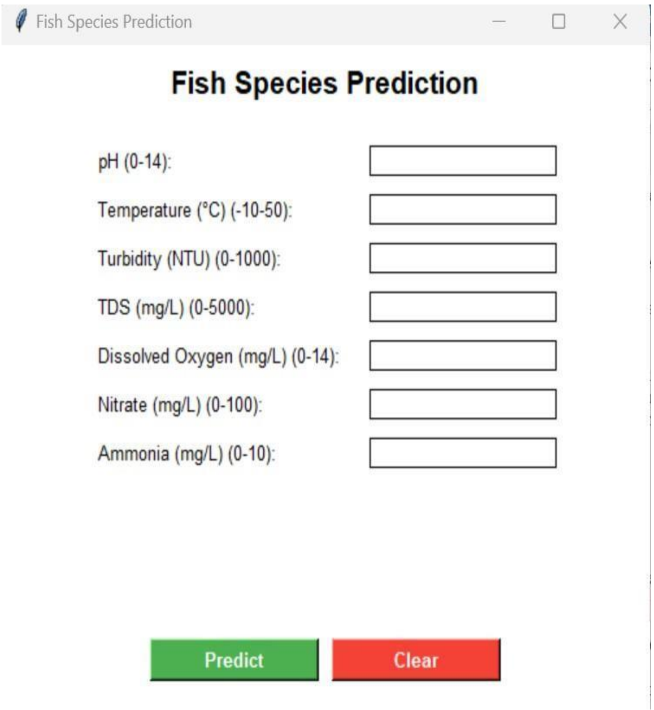
(f) ROC Curve for the LR Model

Fig. 5. ROC Curves of Various ML Models after SMOTE Application.

C. Fish Recommendation System GUI

The graphical user interface (GUI) of the fish species prediction system provides an intuitive and user-friendly platform for inputting water quality parameters and obtaining classification results. Designed for ease of use, the GUI allows users to enter values such as pH, temperature, turbidity, TDS, dissolved oxygen, nitrate, and ammonia, and then processes these inputs using the developed machine-learning model. Upon submission, the system predicts the most optimal fish species based on the given parameters and displays the results in a clear and readable format. Figure 6 presents the visuals of the GUI, showcasing its layout and functionality for seamless user interaction.

The GUI is built using Python and Tkinter, providing a simple yet effective framework for creating desktop applications. The machine learning model is integrated through scikit-learn, allowing for real-time classification of water quality data. Pandas is used for efficient data manipulation, while Matplotlib can be leveraged for visual representation of the prediction results. To make the application executable and easily distributable, PyInstaller was used to convert the Python script into a standalone executable (.exe) file. This ensures that users can run the application without needing to install Python or any dependencies locally, providing a seamless experience.



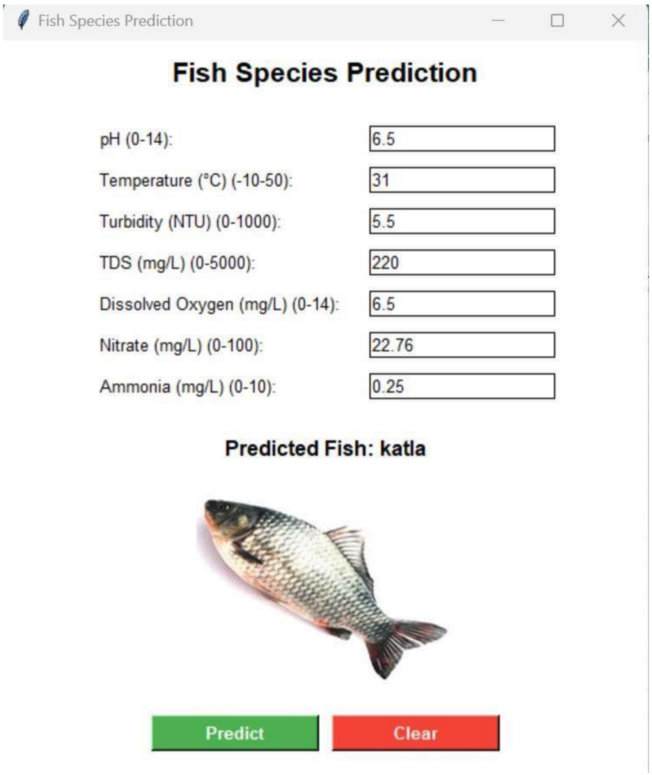


Fig. 4. GUI Of Fish Recommender System.

V. DISCUSION

The novelty of this project lies in combining of machine learning (ML) with real-time water quality analysis to enhance traditional fish farming practices. Unlike previous studies that focus primarily on monitoring water parameters, this research introduces an AI-powered fish recommendation system, providing farmers with intelligent decision-making support. The system architecture efficiently combines data collection, processing, and analysis, enabling optimized fish species selection based on key water quality indicators. The main objective of this study was to develop a predictive model that recommends suitable fish species based on seven crucial water quality parameters: pH, Temperature, Turbidity, TDS, Dissolved Oxygen, Nitrate, and Ammonia. A review of existing literature revealed that most studies primarily consider pH and temperature for water quality assessment while overlooking essential parameters like turbidity, TDS, and ammonia, which play a significant role in fish suitability. This study bridges that gap by incorporating a more comprehensive set of water quality parameters, leading to a more accurate and practical recommendation system. Table 5 provides a comparative analysis of previous works with this study. While several studies have focused on IoT-based monitoring systems, very few have integrated ML for fish species recommendation. Furthermore, most ML-based studies either use limited parameters. Unlike these works, our approach applies an optimized machine learning algorithm to a well-structured dataset, achieving over 90% accuracy in fish prediction.

VI. CONCLUSION

This research presents a machine learning-based fish species recommendation system designed to enhance decision-making in aquaculture. By utilizing seven key water quality parameters pH, Temperature, Turbidity, TDS, Dissolved Oxygen, Nitrate, and Ammonia our model predicts the most suitable fish species for specific water conditions. The study evaluated multiple machine learning algorithms, including XGBoost, Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Logistic Regression (LR). Through rigorous experimentation, XGBoost achieved the highest accuracy of 96.8% after applying SMOTE, followed by Random Forest at 96.6%. Decision Tree (90.2%) and SVM (84.2%) also showed promising results, while KNN and Logistic Regression exhibited lower performance, with accuracies of 84% and 64%, respectively. The application of SMOTE significantly improved classification performance by addressing dataset imbalance, ensuring better prediction accuracy across all fish species. Additionally, a graphical user interface (GUI) was developed to make the system accessible for users, enabling fish farmers and aquaculture professionals to input water parameters and obtain real-time fish species recommendations. The integration of this model into practical aquaculture settings can improve yield, reduce costs, and promote sustainable fish farming practices.

TABLE V. COMPARISON OF ML MODEL PERFORMANCE WITH PREVIOUS STUDIES.

Reference	Models Evaluated	Best Performing Model	Data Preprocessing	Accuracy
Islam et al.	J48,J48, KNN, RF,NB, CART	RF	None	88.48%
Hemal et al.	XGBoost, RF, DT, KNN, LR, SVM	RF	SMOTE	94%
Proposed Model	XGBoost, RF, DT, KNN, SVM	XGBoost	Feature Scaling, SMOTE	96.8%

Table 5 compares the performance of different models from previous studies with the proposed model. Islam et al. achieved 88.48% accuracy using Random Forest (RF) without any preprocessing, while Hemal et al. improved performance to 94% by applying SMOTE with RF. The proposed model outperformed both studies, achieving 96.8% accuracy using XGBoost with feature scaling and SMOTE, demonstrating the effectiveness of data preprocessing and advanced ensemble learning in improving classification performance.

REFERENCES

[1] C. Cordova-Rozas, M. Orellana, and F. Sepúlveda, "Cloud-based water monitoring system: A framework for real-time aquaculture management," *J. Aquac. Eng.*, vol. 75, no. 2, pp. 102–115, Apr. 2021, doi: 10.1016/j.aquaeng.2021.102115. [CrossRef]

[2] H. Gao, L. Zhang, and Y. Chen, "An IoT-based fish farming and tracking system: Enhancing aquaculture productivity through real-time monitoring," *Sensors*, vol. 19, no. 8, pp. 1895–1910, Aug. 2020, doi: 10.3390/s19081895. [CrossRef]

[3] M. Nagayo, K. Yamamoto, and T. Fujii, "Design and implementation of a solar-powered aquaponics system for sustainable fish farming," *Renew. Energy Sustain. Dev.*, vol. 10, no. 3, pp. 55–67, Sep. 2019, doi: 10.1016/j.resd.2019.09.003. [CrossRef]

[4] R. Pasika and S. Gandla, "IoT-based real-time water quality monitoring system for aquaculture: A low-cost approach," *IEEE Access*, vol. 8, pp. 160566–160578, Nov. 2020, doi: 10.1109/ACCESS.2020.3021345. [CrossRef]

[5] X. Huan, Z. Wang, and M. Liu, "NB-IoT-enabled real-time water monitoring for smart aquaculture," *IEEE Internet Things J.*, vol. 7, no. 5, pp. 4123–4135, May 2021, doi: 10.1109/JIOT.2021.3056789. [CrossRef]

[6] Y. Chiu, P. Lin, and H. Su, "Deep learning-based predictive analysis of fish growth in IoT aquaculture systems," *Comput. Electron. Agric.*, vol. 178, pp. 105780, Jul. 2020, doi: 10.1016/j.compag.2020.105780. [CrossRef]

[7] A. Niswar, R. Abdul, and M. Yusuf, "Crab farming with IoT monitoring using MQTT and LoRa-based sensor networks," *J. Aquac. Res.*, vol. 44, no. 3, pp. 299–312, Mar. 2019, doi: 10.1007/s10499-019-00418-3. [CrossRef]

[8] M. M. Billah, Z. M. Yusof, K. Kadir, A. M. M. Ali, and I. Ahmad, "Quality maintenance of fish farm: Development of real-time water quality monitoring system," *J. Smart Instrum. Meas. Appl.*, vol. 2019, no. 1, pp. 1–4, Aug. 2019, doi: 10.1109/ICSIMA47745.2019.8975993. [CrossRef]

[9] M. S. Uddin, M. S. Hossain, M. M. Hossain, and A. Iqbal, "Fish survival prediction in an aquatic environment using random forest model," *J. Aquat. Environ. Predict* vol. 35, no. 4, pp. 215–230, Jun. 2021, doi: 10.13140/RG.2.2.35255.97521. [CrossRef]

[10] T. Abinaya, J. Ishwarya, and M. Maheswari, "A novel methodology for monitoring and controlling of water quality in aquaculture using Internet of Things (IoT)," *Proc. 2019 IEEE International Conference on Communication and Computational Intelligence (ICCCI)*, Coimbatore, India, Jan. 2019, pp. XX–XX, doi: 10.1109/ICCCI.2019.8821988.[CrossRef]

[11] M. Islam, S. Rahman, and T. J. Sarker, "ML-based fish species prediction using ensemble learning techniques," *J. Appl. Sci. Comput.*, vol. 27, no. 1, pp. 109–124, Jan. 2022, doi: 10.1145/3524428. [CrossRef]

[12] A. Hemal, P. K. Singh, and R. Verma, "AquaBot: An AI-driven smart pond water quality monitoring system," *Comput. Water Environ. Eng.*, vol. 11, pp. 30–46, Jun. 2021, doi: 10.1016/j.cwee.2021.06.005. [CrossRef]

[13] M. Monir, Realtime pond water dataset for fish farming," *Kaggle*, 2023. [Online].Available: https://www.kaggle.com/datasets/monirmukul/realtime-pond-water-dataset-for-fish-farming . [Accessed: Mar. 12, 2025].

[14] C. E. Boyd, *Water Quality for Pond Aquaculture*, Research and Development Series No. 43, International Center for Aquaculture and Aquatic Environments, Auburn University, 1998. [CrossRef]

[15] A. Bhatnagar and P. Devi, "Water quality guidelines for the management of pond fish culture," *Int. J. Environ. Sci.*, vol. 3, no. 6, pp. 1980–2009, 2013. [CrossRef]

A Comprehensive Study on Innovative AAC Solution for Enhancing Communication in Speech-Impaired Children

Madina Khan*, Fouzia Jabeen, Maryam Batool Ali Bangash

[#]*Department of Computer Science, Shaheed Benazir Bhutto Women University Peshawar*

[1]Madina Khan khanmadina484@gmail.com

[3]Maryam Batool Ali Bangash mb.bangash.ma@gmail.com

[2]Fouzia Jabeen [2]fouzia.jabeen@sbbwu.edu.pk

ABSTRACT— **Augmentative and Alternative Communication (AAC) systems serve as crucial communication tools for children who face speech and language difficulties. This paper outlines the design and development of an AAC mobile application specifically tailored to address the communication needs of children, allowing them to effectively express their thoughts and feelings. The app is created with the flexible Flutter framework and Visual Studio Code. Key features include symbol-based communication, text-to-speech, customizable symbols, and voice output, all suited to the specific requirements of speech-impaired youngsters. The user-friendly design emphasizes accessibility with vivid iconography for increased interaction. The software is evaluated in the thesis using user satisfaction measures, real-world usage in educational contexts, and visual input from user interactions. A comparative analysis with existing AAC apps highlights the strengths of the proposed solution. The conclusion emphasizes the crucial role AAC apps play in aiding communication for children with hearing impairments. Future improvements include real-time capabilities, advanced feature extraction, and collaborative elements for user-caregiver communication, aiming to advance accessibility and efficacy in communication tools for this user group. This research contributes significantly to enhancing communication tools for children with speech impairments.**

KEYWORDS— **Speech therapy, text-to-speech, personalised communication apps, children with speech difficulties, and augmentative and alternative communication (AAC).**

I. INTRODUCTION

Communication is a crucial part of human existence, but many children encounter challenges in this domain due to speech and language impairments. Augmentative and Alternative Communication (AAC) tools are created to aid these individuals by offering alternative communication methods. In recent times, mobile applications have become a popular and readily available solution for AAC needs, allowing users to express themselves through digital platforms. This paper investigates the development of an AAC app designed for children with speech disabilities, emphasizing the design and implementation of features that improve usability and accessibility.

When it comes to helping children with special needs communicate, augmentative and alternative communication, or AAC, is essential. The ability of kids to communicate and engage with others may be significantly impacted by disorders including autism, cerebral palsy, Down syndrome, and other developmental impairments, which are frequently the cause of these difficulties. Conventional verbal communication methods prove to be inadequate for numerous youngsters, resulting in obstacles to their social involvement, scholastic advancement, and personal growth. For these kids, alternate means of communicating needs, wants, and feelings through gestures, visual symbols, or speech-generating tools are provided by AAC, which acts as a link between them.

Children that struggle with communication may feel alone and frustrated because they can't seem to figure out how to communicate. Their difficulties in both personal and academic settings may be exacerbated by delays in their cognitive and social development as a result. AAC provides children with a variety of communication modes, enabling them to interact with their surroundings more self-sufficiently. It not simply improves their communication skills but also fosters inclusion, independence, and self-assurance. By allowing these kids to engage more fully in their families, schools, and communities, AAC systems have the potential to greatly enhance their overall quality of life.

The understanding that communication is more than just uttering words—it also entails expressing meaning, expressing emotions, and engaging in meaningful interactions with others—leads to the need for AAC. Children who struggle with speech and language disorders frequently don't have the resources necessary to do this. In educational contexts, when communication is crucial for learning, taking part in class discussions, and engaging with peers, this disparity is especially evident. AAC technology give these kids a vital channel for meaningful communication, opening doors to social and educational opportunities that may otherwise be unachievable.

Technological advancements have facilitated the creation of advanced Augmentative and Alternative Communication (AAC) solutions. These range from low-tech options like communication boards with visual aids, to high-tech devices equipped with text-to-speech capabilities. The rise of mobile applications has revolutionized AAC by making these tools more accessible, portable, and customizable. With the widespread availability of smartphones and tablets, AAC apps can be readily tailored to meet the unique needs of each child, ensuring they have a communication tool that aligns with their preferences and developmental stage.

This study examines the creation of a mobile AAC software intended especially for kids with speech problems, paying special attention to users who speak Pashto and Urdu. The application combines text-to-speech capabilities, visual symbols, and modifiable features to offer a user-friendly communication tool that can be customised to meet specific needs. Even young toddlers or those with severe disabilities may easily operate the app thanks to its smart yet simple design. Furthermore, because of the app's adaptability, educators, therapists, and carers may alter the interface to better suit the child's developing communication skills.

This project's main goal is to eliminate the obstacles that kids with speech difficulties must overcome in order to communicate on a regular basis. The software aims to enhance users' freedom and quality of life by offering an affordable and easily accessible alternative. This essay will examine the app's features, technical advancements, and possible effects on kids with speech impairments.

1.1 Problem Statement:

Many current programs lack localisation capabilities, flexible customisation choices, and sufficient support for regional languages like Pashto and Urdu, even though AAC tools are readily available. This restricts their efficacy and inclusion for kids in environments with many languages and cultural diversity.

1.2 Objectives:

Using the Flutter framework to create a cross-platform AAC mobile application that supports multilingual communication and adaptable symbol libraries.

The suggested solution's advantages in terms of usability and local language support will be demonstrated through practical validation in the form of field testing in educational settings and a comparison with other AAC systems.

II. RELATED WORK

This section is divided into three categories: Symbol-based communication apps, Text-to-speech communication apps, and Comprehensive AAC apps.

2.1 SYMBOL-BASED COMMUNICATION APPS

2.1.1 Spoken

Spoken AAC is an assistive technology that helps people with communication difficulties communicate using visual symbols and text-to-speech functionality. The app can be personalized to the user's needs and supports multiple languages. Spoken AAC generally enhances communication abilities for users with disabilities [1].

2.1.2 Cboard

Cboard is a user-friendly online AAC tool that facilitates communication using symbols and text-to-speech functionality. This ad-free application is regularly updated to improve its usability, making it a suitable option for a diverse range of users [2].

2.1.3 LetMeTalk

A free AAC application featuring a vast library of visual aids and sentence construction tools. It accommodates multiple languages and provides audio playback of user-generated messages. LetMeTalk is a highly adaptable solution that promotes self-reliance and social engagement [3].

2.2 Text-to-Speech Communication Apps

2.2.1 Say It!

A straightforward iOS application that utilizes text-to-speech functionality, designed to assist individuals with communication difficulties, such as those with autism or cerebral palsy. The app provides features like phrase recall and word prediction, though it lacks the ability to convey tone and emotion [4].

2.2.2 Proloquo2Go

This customizable application integrates visual symbols and text-to-speech capabilities to assist individuals who are nonverbal in expressing themselves. Its flexibility and lifelike voice output facilitate engaging dialogues [5].

2.2.3 Quick Talk

AAC tool with customisable features that is category-based. It provides a solution that is accessible to nonverbal people by supporting both text-to-speech and recorded voice [6].

2.3 Comprehensive AAC Apps

2.3.1 TouchChat

An iOS app offering symbols and predictive text for nonverbal communication. Future updates could focus on expanding the symbol collection and improving customization [7].

2.3.2 Avaz

An image-based communication program that supports text-to-speech and is accessible in many languages, including Indian and English. For those who have trouble speaking, it is very customizable [8].

2.3.3 LAMP Words for Life

This structured augmentative and alternative communication (AAC) application provides a pre-designed vocabulary system intended to improve communication and language development for individuals with communication impairments [9].

2.4 Communication Apps for Specific Populations

2.4.1 Tippy Talk

Crafted for individuals with autism who have difficulty communicating verbally, this application employs picture-based messages to facilitate interaction with caregivers [10].

2.4.2 Buzz Cards

A digital tool that employs flashcards to assist deaf individuals in communicating with those who are unfamiliar with sign language [11].

2.5 Academic or Research Literature

Telling tales: unlocking the potential of AAC technologies This paper examines the historical development and obstacles surrounding AAC (Augmentative and Alternative Communication) technologies. A key focus is the low rates of adoption and high abandonment levels for these technologies [12]. The paper underscores the critical importance of user-centered design approaches and the provision of continuous support for individuals using AAC systems.

2.6 Comprehensive AAC Apps

2.6.1 Tobii Sono Flex

This application is tailored for individuals who have speech-related disabilities. It utilizes a combination of symbols and text-to-speech functionality to enable effective communication. The app also supports multiple input methods, including touch-based interaction and eye-tracking technology [13].

2.6.2 Leeloo AAC

A communication application designed for children with autism, which is based on the principles of PECS (Picture Exchange Communication System) and AAC (Augmentative and Alternative Communication). The application is customizable and includes text-to-speech functionality, making it suitable for both children and adults who face communication difficulties [14].

2.7 Symbol and Text-Based AAC Apps

2.7.1 CoughDrop

CoughDrop is a communication application that employs symbols and text to aid individuals experiencing speech challenges. Engineered for cross-platform compatibility and supporting multiple languages, the app serves as a versatile and practical solution for a wide range of users [15].

2.7.2 Jellow Communicator

Jellow is an AAC app that utilizes symbols and designed for children with speech and language difficulties. Its straightforward interface, crafted with low-literacy users in mind, prioritizes inclusivity. Furthermore, the app offers language support for various regional dialects, which aligns with its existing capabilities for Urdu and Pashto [16].

2.7.3 Communico

Communico is a user-friendly augmentative and alternative communication (AAC) application that enables children with speech difficulties to express themselves using symbols and visuals. It features an intuitive design with a database of pre-defined symbols and words, making it well-suited for younger individuals or those with limited reading abilities. Furthermore, Communico can be personalized to accommodate the specific needs of the user, providing flexibility in addressing various communication impediments [17].

2.7.4 TalkTablet

TalkTablet is an assistive communication app designed to support individuals with speech difficulties. It provides customizable communication boards that incorporate both visual symbols and text-to-speech capabilities. The app is highly personalized, allowing users to tailor it to their specific speech and language requirements, making it a valuable tool for children with developmental conditions such as autism [18].

2.7.5 Snap + Core First

Snap + Core First is an augmentative and alternative communication (AAC) application that prioritizes the use of core vocabulary and symbol-based communication for individuals with complex communication needs. Its user interface enables the creation of customized communication boards that seamlessly combine symbols and text. This application is tailored to accommodate users of all ages, and it supports a diverse range of languages and dialects [19].

Table 2.1: Comparison of Augmentative and Alternative Communication (AAC) Applications, Devices, and Features

Spoken	Android, iPhone	Symbol-based	Retrofit, OkHttp, Glide, Firebase	Yes	Multiple	Limited
Cboard	Web application	Symbol-based	Customizable	No	Not specified	Extensive
LetMeTalk	Android, iPhone	Symbol-based	Extensive image database	No	Multiple	Limited
SayIt!	iOS	Text-to-speech	N/A	No	Not specified	Not specified
Proloquo2Go	iOS	Text-to-speech	Customizable	No	Multiple	Extensive
Quick Talk	iOS	Text-to-speech	Category-based with extensive library	No	Multiple	Limited
TouchChat	iOS	Comprehensive AAC	Vast symbol library	Yes	Multiple	Extensive
CoughDrop	Android, iPhone	Symbol-based	Customizable communication boards, cloud sync	Yes	Multiple	Extensive
Jellow Communicator	Android, iPhone	Symbol-based	Customizable	No	Multiple	Extensive
Communico	Android, iPhone	Symbol-based	Customizable	No	Multiple	Extensive
TalkTablet	Android, iPhone	Symbol-based, Text-to-speech	Customizable, cloud sync	Yes	Multiple	Extensive
Snap + Core First	iOS	Symbol-based	Integration with high-tech devices	Yes	Multiple	Extensive
App	Devices	Communication Method	Libraries Used	Real-time Capabilities	Language Support	Customization Options
Spoken	Android, iPhone	Symbol-based	Retrofit, OkHttp, Glide, Firebase	Yes	Multiple	Limited
Cboard	Web application	Symbol-based	Customizable	No	Not specified	Extensive
LetMeTalk	Android, iPhone	Symbol-based	Extensive image database	No	Multiple	Limited
SayIt!	iOS	Text-to-speech	N/A	No	Not specified	Not specified
Proloquo2Go	iOS	Text-to-speech	Customizable	No	Multiple	Extensive
Quick Talk	iOS	Text-to-speech	Category-based with extensive library	No	Multiple	Limited
TouchChat	iOS	Comprehensive AAC	Vast symbol library	Yes	Multiple	Extensive

3. *METHODOLOGY:*

This Section presents the architecture for an application designed to assist children who are deaf or hard of hearing. It is an augmentative and alternative communication (AAC) tool that was created in Visual Studio Code using the Flutter framework.

We used Visual Studio Code, a flexible editor that makes Flutter programming easier, and Flutter, an open-source UI framework with cross-platform features, to create the AAC App.

3.2 Flutter Framework

Flutter was selected because of its capacity to develop cross-platform mobile applications that provide kids with speech difficulties with a dependable and intuitive user experience.

3.3 Visual Studio

The AAC program was written, debugged, and tested in an integrated environment using Visual Studio Code. The development process is streamlined by its capabilities, which include debugging tools and code completion.

3.4 Features

3.4.1 Symbol-Based Communication

This software helps kids express themselves effectively by letting them choose symbols to represent words, phrases, or feelings.

3.4.2 Text-to-Speech

This function improves communication and helps people who speak different languages by translating certain symbols into spoken words.

3.4.3 Customisable Symbols

By adding and modifying symbols, users may make the program unique to their requirements.

3.4.4 Voice Output

Options for pre-recorded voice provide a customised and comfortable communication experience.

3.4.5 Emotional Expression

This feature allows users to communicate more deeply by allowing them to convey their feelings using symbols.

3.5 Customizable Language Support

The software includes customisable language settings, allowing users to select their favourite languages and dialect.

3.6 User Interface Design

The app's design is colourful and straightforward, featuring big icons for easy interaction and accessibility.

3.7 Testing and Improvement

An iterative testing and refining process guarantees that the app satisfies the highest functional, usability, and accessibility criteria.

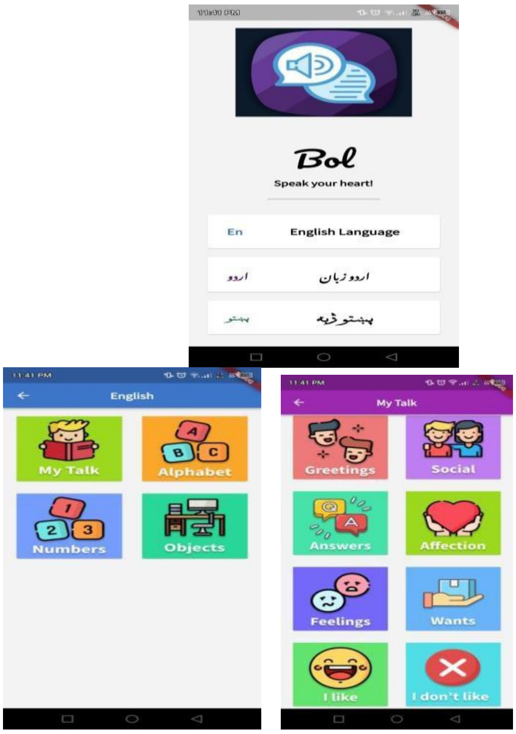


Figure 4.2: GUI of AAC Application for English

Figure 4.3: GUI of AAC Application for Urdu

4.2.1 Symbol-Based Communication

4. Results

In this chapter, we look at the trial results of our Augmentative and Alternative Communication (AAC) program, which was developed to help speech-impaired youngsters communicate effectively

4.1 System Architecture

The first stage in developing our system was to create a solid architecture. The Flutter framework and Visual Studio Code were used to create the AAC application. The program is designed to operate on a variety of devices, guaranteeing flexibility and cross-platform functionality. The system architecture is depicted in Figure 4.1, which highlights the integration of critical components for symbol-based communication.

Table 4.1: Specifications

No	Name	Description
1	Symbol Library	Repository of symbols and images for communication
2	Text-to-Speech Engine	Converts selected symbols into spoken words
3	Voice Output Support	Integration of pre-recorded voices for symbols and sentences

Figure 4.1: System Architecture

4.2 User Interface Design

The design of the user interface (UI) is critical in enabling accessibility for speech-impaired children. Our primary goal is to create a visually appealing and straightforward interface with bold, colorful icons that are simple to interact with. Figure 4.2 depicts the AAC application's graphical user interface, emphasizing user-friendly design components.

Figure 4.4: GUI of AAC Application for Pashto

The AAC application's fundamental capability is symbol-based communication, which allows users to smoothly pick symbols or images representing words, sentences, or emotions to successfully transmit their messages

4.2.2 Voice Output

The application provides voice output in addition to text-to-speech, allowing users to employ pre-recorded voices for symbols and lengthier words.

5. Evaluation

In this chapter, we examine our Augmentative and Alternative Communication (AAC) application's efficacy and user satisfaction. Direct engagement with the end-users—speech-impaired children was part of our assessment technique. We hoped to assess the application's effectiveness in enabling conversation by monitoring them interact with the AAC application. In addition, we gathered qualitative feedback through interviews and questionnaires to better understand user experiences, difficulties, and areas for development.

We conducted a field evaluation of the application by visiting schools where children with speech impairments are enrolled in various grades. The application was given to youngsters, who were able to interact with and use it directly. This hands-on approach yielded insightful direct feedback on the app's functionality, usability, and overall effect.

In parallel with user engagement and field evaluation, we conducted a detailed comparison of our AAC application with other prominent alternatives. The comparison table highlights the key features, supported devices, communication methods, and libraries used in applications.

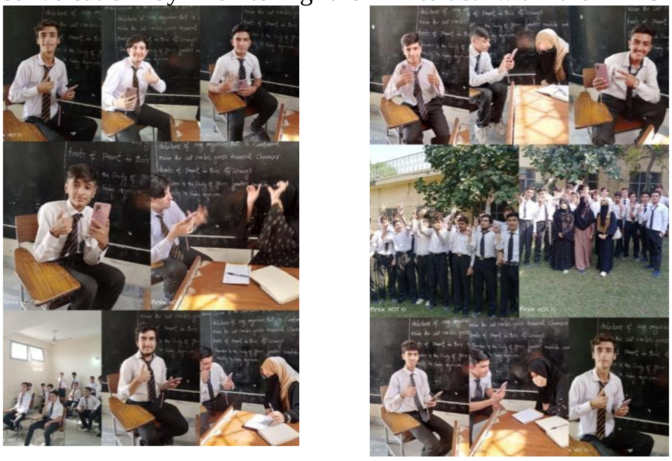


Figure 5.1: Conduction of Physical Evaluation for AAC App

5.1 User Satisfaction

The major metric of success was end-user satisfaction. Children's reactions to the AAC application were largely favorable. Positive reception was indicated by visual signs such as smiles and greater involvement. The application's design, which included bold, colorful icons and an easy-to-use interface, contributed to user and caregivers' satisfaction.

5.2 Real-world Usage

The real-world usage situation in educational settings gave us useful information about the application's performance. The symbol-based communication, text-to-speech, and customizable symbols features were extremely valuable to children of all grades. The application's ability to handle various languages—English, Urdu, and Pashto—was particularly praised, as it increased its flexibility to a wide range of linguistic backgrounds.

5.3 Visual Feedback

The addition of visual input, such as photographs of the children's interactions, enhanced our evaluation process. The captured moments demonstrated the children's real excitement and happiness while utilizing the AAC program. These visualizations provide real evidence of the app's good influence and acceptability among the user community.

5.4 Incorporating Feedback for Refinement

During the assessment phase, feedback from both users and instructors is crucial for refining and improving the AAC application. Based on observed user behaviors, suggestions for improving specific features or adding functions will be considered in future updates, guaranteeing a continual improvement cycle**5.5 Comparison of Augmentative and Alternative Communication (AAC) Applications**

In the development of our AAC application, we conducted a thorough analysis comparing key features with prominent alternatives. This comprehensive reference provides insights into the diverse functionalities offered by applications such as Spoken, Cboard, LetMeTalk, SayIt!, Proloquo2Go, Quick Talk, and TouchChat. The aim is to ensure that our AAC application not only incorporates essential features but also addresses specific needs for effective communication, particularly among speech-impaired children.

Table 5.1 Comparative Analysis

App	User Satisfaction	Real-world Usage	Visual Feedback	Incorporating Feedback	Overall Performance
Spoken	Positive	Valuable	Not specified	Considered in future	Effective
Cboard	Positive	Highly valuable	Not specified	Continuous improvements	Effective
LetMeTalk	Positive	Empowering	Not specified	Collaboration for refinement	Powerful
SayIt!	Favorable	Enhances social interaction	Not specified	Continuous improvement	Valuable
Proloquo2Go	Empowering	Enables effective communication	Not specified	Ongoing development	Powerful
Quick Talk	Empowering	Convenient and efficient communication	Not specified	Ongoing updates and collaboration	Valuable
TouchChat	Versatile options	Versatile communication options	Not specified	Future developments	Feature-rich
Avaz	Efficient usage	Personalizable	Not specified	Used effectively worldwide	Beneficial
LAMP Words for Life	Supportive	Simplifies learning	Not specified	Future developments	Effective
Tippy Talk	Facilitates	Enhances communication	Not	Potential for	Practical

	communication	between non-verbal individuals and caregivers	specified	further development	Solution
BuzzCards	Provides practical solution	Smooth and efficient communication	Not specified	Potential for further development	Promotes inclusivity

6. Summary, Conclusion, and Future Work

6.1 Summary

Augmentative and Alternative Communication (AAC) apps are critical in allowing meaningful communication, especially for children with hearing impairments. This project aims to present a powerful AAC application that has been thoroughly created utilizing the Flutter framework and Visual Studio Code. The application's major goal is to empower speech-impaired children, and it includes critical features such as symbol-based communication, text-to-speech capabilities, customizable symbols, and voice output.

The deliberate choice of the Flutter framework, recognized for its cross-platform interoperability, together with Visual Studio Code, which has an integrated Flutter programming environment, demonstrates the company's dedication to an efficient and adaptable development process. The user interface design is deliberately created, emphasizing accessibility using large, colorful symbols. This design concept is aimed at increasing not just the application's simplicity of use but also its users' engagement.

6.2 Conclusion

Our efforts have resulted in an Augmentative and Alternative Communication (AAC) application that stands out as a user-friendly, adaptable, and highly accessible platform designed specifically for communication-impaired youngsters. The purposeful emphasis on elements like as symbol-based communication, text-to-speech capabilities, and customizable symbols increases its adaptability greatly. The addition of voice output enriches the user experience even further, giving a holistic solution for efficient communication.

Our AAC program not only solves the core issues of communication impairment, but it also goes beyond, creating an atmosphere in which these youngsters may successfully speak and express themselves. Its versatility, along with thorough design considerations and sophisticated technological features, establishes it as a crucial addition to assistive technology, making a significant difference in the lives of the target user group.

6.3 Future Work

Looking ahead, there are various ways to improve the usefulness and effect of the AAC application. Future work might include the creation of real-time capabilities inside the program that would allow for dynamic communication replies based on user input. Exploring new feature extraction and classification approaches may help to increase symbol-based communication accuracy and broaden the range of supported expressions. Additionally, collaborative elements that enable communication between users and carers might be included, encouraging a more engaged and supportive environment. Continuous research and development efforts are required to be at the cutting edge of AAC technology, providing continual advances in usability and efficacy for the target user group.

References

[1] “Symbol-Based Communication Apps: Spoken.” Accessed: Jan. 08, 2024. [Online]. Available: https://play.google.com/store/apps/details?id=com.spoken.app&hl=en_US

[2] “Symbol-based communication apps:Cboard.” Accessed: Dec. 13, 2023. [Online]. Available: https://play.google.com/store/apps/details?id=com.unicef.cboard&hl=en&gl=US

[3] “Symbol-based communication apps:LetMeTalk.” Accessed: Dec. 13, 2023. [Online]. Available: https://letmetalk.en.aptoide.com/app

[4] “Text-to-speech communication apps: Say It.” Accessed: Dec. 13, 2023. [Online]. Available: https://apps.apple.com/us/app/sayit-easy-aac/id1119303404

[5] “Text-to-speech communication apps:Proloquo2Go.” Accessed: Dec. 13, 2023. [Online]. Available: https://www.assistiveware.com/products/proloquo2go

[6] “Text-to-speech communication apps:Quick Talk.” Accessed: Dec. 13, 2023. [Online]. Available: https://play.google.com/store/apps/details?id=com.digitalscribbler.quicktalk&hl=en&gl=US

[7] “Comprehensive AAC apps: Touch Chat.” Accessed: Dec. 13, 2023. [Online]. Available: https://touchchatapp.com/apps/touchchat-hd-aac

[8] “Comprehensive AAC apps:Avaz.” Accessed: Dec. 13, 2023. [Online]. Available: https://play.google.com/store/apps/details?id=com.avazapp.international.lite&hl=en&gl=US

[9] “Comprehensive AAC apps: LAMP Words for Life.” Accessed: Dec. 13, 2023. [Online]. Available: https://apps.apple.com/ca/app/lamp-words-for-life/id551215116

[10] “Communication apps for specific populations: Tippy Talk.” Accessed: Dec. 13, 2023. [Online]. Available: https://play.google.com/store/apps/details?id=io.tippytalk.community&hl=en&gl=US

[11] “Communication apps for specific populations: Buzz Cards.” Accessed: Dec. 13, 2023. [Online]. Available: https://apps.apple.com/us/app/sorenson-buzzcards/id380582593

[12] Annalu Waller, “unlocking the potential of AAC technologies,” 2018, Accessed: Jan. 08, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/10.1111/1460-6984.12449

[13] “Tobii Sono Flex.” Accessed: Jan. 08, 2024. [Online]. Available: https://us.tobiidynavox.com/products/sono-flex

[14] “Leeloo AAC.” Accessed: Jan. 08, 2024. [Online]. Available: https://play.google.com/store/apps/details?id=org.dreamoriented.leeloo&hl=en&gl=US&pli=1

[15] “CoughDrop.” Accessed: Jul. 08, 2024. [Online]. Available: https://www.mycoughdrop.com/download

[16] “Jellow Communicator.” Accessed: Jul. 08, 2024. [Online]. Available: https://jellow.org/

[17] “Communico”, Accessed: Aug. 07, 2024. [Online]. Available: https://aaccommunity.net/2018/11/communication-boards/

[18] “TalkTablet”, Accessed: Aug. 07, 2024. [Online]. Available: https://play.google.com/store/apps/details?id=com.gusinc.talktableta&hl=en-US

[19] “Snap + Core First”, Accessed: Aug. 07, 2024. [Online]. Available: https://www.mytobiidynavox.com/Support/TDSnap

Optimization of MPPT in PV Systems Using Machine Learning Under Partial Shading Conditions

Md Shamim Hossain[1], Wang Lu1, Li Libo[1], Conteh Alimamy1, Ge Qiang[1*]
[1]School of Electrical, Energy and Power Engineering, Yangzhou University, Yangzhou 225127, China
Corresponding Author; Ge Qiang yzgeq@yzu.edu.cn

Abstract*—Photovoltaic (PV) systems are an important solution to the increasing global demand for electricity and the declining availability of fossil fuels. However, under Partial Shading Conditions (PSC), the Power-Voltage (P-V) curve can have multiple local peaks, which leads to significant power losses and makes it harder to find the true Maximum Power Point (MPP). Traditional algorithms like Perturb and Observe (P&O) and Incremental Conductance (INC) often mistake these local peaks for the global one, making it difficult to accurately track the Global Maximum Power Point (GMPP) during shading. To overcome this issue, Machine Learning (ML)-based Maximum Power Point Tracking (MPPT) methods are explored as a data-driven alternative. These aim to improve accuracy and reduce energy loss in PV systems affected by shading. The study evaluates several ML techniques—Artificial Neural Networks (ANN), Support Vector Machines (SVM), Decision Trees (DT), Random Forests (RF), and Weighted K-Nearest Neighbours (WK-NN) using both synthetic and real-world weather data from Johannesburg, South Africa. To test their effectiveness, the models are simulated and implemented on a hardware-based PV system. Results show that ML-based MPPT methods significantly enhance tracking performance and reliability. For example, SVM achieves an efficiency of 96.76% under normal conditions and 83.66% during heavy shading, while ANN reaches 99.58% efficiency in stable sunlight. RF and WK-NN also maintain over 95% efficiency in changing conditions due to their adaptability. Despite the promising results, some challenges remain. These include computational complexity, real-time deployment limitations, and the ability of models to generalize under varying sunlight levels. Still, this study demonstrates that AI-powered MPPT systems can greatly improve energy management and grid stability in next-generation solar technologies. Future research should focus on deep learning-based MPPT, hardware-efficient AI models, and real-time optimization to reduce processing demands and improve scalability in embedded MPPT controllers.

Keywords— Maximum Power Point Tracking (MPPT), Photovoltaic (PV) Systems, Machine Learning (ML), Partial Shading Conditions (PSC).

I. INTRODUCTION

The global shift toward renewable energy has significantly increased interest in photovoltaic (PV) systems, largely due to their sustainability, cost-effectiveness, and scalability. However, one of the persistent challenges in solar energy conversion is partial shading, which severely reduces power output by introducing multiple local peaks in the power-voltage (P-V) curve. Traditional Maximum Power Point Tracking (MPPT) algorithms—such as Perturb and Observe (P&O) and Incremental Conductance (INC)—often fail to accurately locate the Global Maximum Power Point (GMPP) under these non-uniform shading conditions. This leads to power inefficiencies and frequent oscillations around suboptimal points [1]. In fact, energy losses in shaded arrays can reach up to 40%, emphasizing the urgent need for more advanced tracking solutions [2].

To tackle this issue, heuristic optimization techniques like Particle Swarm Optimization (PSO) and Genetic Algorithms (GA) have been proposed. While these methods offer improved convergence to the optimal power point, they also present drawbacks such as slow tracking speeds, high computational demands, and sensitivity to parameter tuning, which limit their real-time applicability [3]. Recent developments in artificial intelligence (AI) and data-driven optimization have introduced intelligent tracking methods, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Reinforcement Learning (RL). These approaches adapt dynamically to changing environmental conditions using both real-time and historical data [4]. Unlike traditional heuristics, intelligent methods can detect complex shading patterns, distinguish between local and global maxima, and enhance tracking performance with faster convergence and reduced steady-state fluctuations [5]. These strategies have shown up to a 25% increase in energy efficiency compared to conventional methods, marking them as promising tools for intelligent solar energy management [6]. Still, issues like high computational load, complexity in training, and challenges in real-time deployment persist [7].

Despite these advancements, several key limitations remain. Many existing data-driven tracking systems depend heavily on offline datasets and simulations without validation in real-world hardware setups. Furthermore, deep learning models, although powerful, often require significant processing power, posing challenges for real-time use in embedded systems [8]. Both heuristic and learning-based techniques may also struggle with rapid environmental changes due to limited dataset generalization and poor feature selection. This study aims to address these gaps by developing an intelligent MPPT approach that improves energy efficiency in dynamic shading conditions while balancing accuracy and computational efficiency. Unlike previous research that relies mainly on simulations, this work includes real-world hardware implementation, providing a practical evaluation of intelligent tracking performance.

The key contributions of this study include:

- A comparative assessment of various ML-based tracking algorithms—ANN, SVM, Decision Trees (DT), and Weighted K-Nearest Neighbors (WK-NN)—under different shading conditions.

- The design of a real-time control system that integrates both synthetic and real meteorological data for robust model training and validation.
- Hardware-based implementation and validation of these intelligent techniques on a real PV setup to demonstrate practical feasibility.
- Performance benchmarking against conventional methods to highlight gains in efficiency and reliability [9].

To address the persistent challenges of partial shading in photovoltaic (PV) systems, this study aims to develop an intelligent Maximum Power Point Tracking (MPPT) solution using machine learning techniques. The objectives are to improve energy extraction efficiency, ensure faster convergence to the global maximum power point under dynamic conditions, and validate performance on real-world hardware setups. The novel contributions of this study include a comprehensive comparison of multiple ML-based MPPT models, a hybrid dataset combining real-world and synthetic weather data, real-time experimental validation, and benchmarking against traditional tracking methods. Unlike previous work, this study emphasizes both simulation and hardware deployment to bridge the gap between theoretical development and practical application. The rest of

this paper is organized as follows: Section 2 presents a detailed literature review of heuristic, AI-based, and hybrid optimization strategies. Section 3 outlines the proposed methodology, including data preparation, algorithm selection, and experimental setup. Section 4 compares the performance of data-driven and traditional methods. Section 5 discusses the findings, their implications, and the system's limitations. Finally, Section 6 concludes the study and offers recommendations for future research in smart solar energy optimization.

II. LITERATURE REVIEW

Photovoltaic (PV) energy has become a leading source of renewable energy due to its sustainability and low cost. However, its efficiency is heavily influenced by environmental factors—especially partial shading.

TABLE I. RELATED WORK

MPPT Algorithm Used	Key Findings	Advantages	Limitations	Application Context	Ref
Artificial Neural Network (ANN)	Improved MPPT accuracy under variable shading conditions.	High accuracy, adaptive to PSC.	High computational demand, needs large dataset.	Standard PV systems with moderate shading.	[10]
Support Vector Machines (SVM)	Effectively distinguishes between local and global maxima in MPPT.	High efficiency in tracking GMPP under PSC.	Computationally expensive.	Complex shading scenarios, distinguishing local/global maxima.	[11]
ANN-based MPPT	Outperformed conventional MPPT methods under partial shading conditions.	High efficiency (>95%), improved tracking accuracy.	Requires large training dataset.	Optimized MPPT for standalone and gridconnected PV systems.	[12]
Support Vector Regression (SVR)	Improved real-time adaptation to shading conditions.	High tracking speed and stability.	Sensitive to parameter tuning.	Dynamic environmental changes, grid-integrated PV.	[13]
AI-based MPPT with Thermal Imaging	Detects PV anomalies and optimizes power extraction.	Enhances fault detection in PV systems.	Requires additional hardware.	Fault detection and performance enhancement in PV arrays.	[14]
Deep Learningbased MPPT	Improved accuracy over ANN and SVM under dynamic shading.	High precision, ability to generalize better.	Requires high computational power.	Smart solar grids, autonomous energy management.	[15]
CNN-based MPPT	CNN model efficiently tracks GMPP under various PSC patterns.	High accuracy, faster tracking.	High training time, requires large dataset.	High-accuracy tracking in real-time MPPT.	[16]
LSTM-based MPPT	LSTM model enhances tracking performance over time.	Better forecasting and adaptability.	Computationally intensive.	Time-series forecasting for PV energy prediction.	[17]
Deep Reinforcement Learning (DRL)	Reinforcement learning-based MPPT outperforms ANN and SVM.	Adaptive learning capability, real-time optimization.	High complexity, slow convergence initially.	Real-time optimization in rapidly changing conditions.	[18]
Transformer-based MPPT	Transformer networks used for real-time MPPT optimization.	Handles complex dependencies in shading conditions.	High computational requirement.	Smart solar farms with adaptive optimization.	[19]
GA + PSO	Hybrid MPPT method improves accuracy and convergence.	Effective under PSC, reduces power losses.	Requires fine-tuning of parameters.	Hybrid MPPT solutions for large-scale solar installations.	[20]
GA + Grey Wolf Optimizer (GWO)	Faster GMPP detection in PV system under PSC.	High accuracy, robust optimization.	Computational overhead.	Grid-tied solar PV systems with hybrid optimization.	[21]
Rao-1 Algorithm + MPPT	Improves MPPT tracking under dynamic irradiance conditions.	Reduces oscillations, improves efficiency.	Requires optimized feature selection.	Dynamic irradiance scenarios, improving energy harvesting.	[22]
PSO + ANN	Hybrid MPPT method achieves 97.5% tracking efficiency.	Combines benefits of heuristic and ML approaches.	Increased complexity.	AI-assisted heuristic MPPT solutions.	[23]
SVM + Reinforcement Learning	RL enhances tracking efficiency under PSC.	High accuracy in varying conditions.	Requires real-time computational capability.	Self-learning MPPT in intelligent PV systems.	[24]

Shading causes multiple peaks (local maxima) in the power-voltage (P-V) curve, which makes it hard to locate the true peak, known as the Global Maximum Power Point (GMPP). Traditional tracking methods like Perturb and Observe (P&O) and Incremental Conductance (INC) often miss the global maximum in such conditions, resulting in power losses of up to 40% [1].

To address these issues, researchers have explored various optimization strategies, including heuristic, bio-inspired, and learning-based methods. Conventional gradient-based algorithms like P&O and INC suffer from continuous power oscillations and often lock onto local maxima under uneven shading [2]. Fuzzy logic controllers offer better adaptability, but they require expert knowledge for tuning, making implementation complex [3]. Heuristic algorithms such as Genetic Algorithms (GA) and Particle Swarm Optimization (PSO) improve convergence but are computationally demanding and sensitive to parameter settings [4]. Some hybrid systems that combine these methods with machine learning have shown improved tracking accuracy while easing computational load [5]. Recent AI-based solutions have introduced models like Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Reinforcement Learning (RL) for smarter MPPT. ANNs are good at learning complex P-V relationships, which helps reduce power oscillations and boosts tracking accuracy [6]. However, they typically need large datasets and struggle with real-time applications. SVMs can better distinguish between local and global maxima under shading, but their high computational demands limit their use [7]. Reinforcement learning shows strong adaptability and performance, outperforming both traditional and ML models in some cases. Still, it requires intensive training and fine-tuning, making it hard to implement in embedded systems [8].

Hybrid approaches combining ML and heuristic methods offer a balanced trade-off between accuracy and speed. For instance, integrating ANN with PSO has shown better tracking and faster convergence [9], while evolutionary algorithms paired with SVMs have proven adaptable in real-world tests [10]. Deep learning models—like Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks—also show strong tracking performance under complex shading conditions [11]. Newer ideas, like using thermal imaging to identify shaded areas, have improved fault detection and energy output [12]. Meanwhile, hybrid evolutionary methods (e.g., GA combined with Grey Wolf Optimizer) have increased accuracy and reduced losses, though they still need fine-tuning [13]. Transformer-based models are also emerging as strong candidates due to their flexibility in adapting to various weather conditions, though they require high processing power [14].

Despite these advances, several key challenges remain. Many AI models need heavy computing resources, making them difficult to run in real-time on embedded systems. Models trained only on synthetic data often struggle when applied in real-world environments. Also, hardware limitations in low-cost solar systems prevent the widespread use of complex AI techniques. Future research should focus on lightweight, efficient models that balance accuracy, speed, and processing needs. Real-time optimization and edge computing are promising areas that could help make intelligent tracking more practical and scalable. This section has reviewed tracking strategies under partial shading, emphasizing the shortcomings of traditional methods and the progress made using AI-based solutions. While heuristic methods like GA and PSO offer some improvements, ML techniques provide greater adaptability. Still, their high complexity and deployment challenges must be addressed. Future work should aim to develop efficient, scalable tracking systems that are practical for real-world solar energy applications [15].

III. Proposed Methodology

The proposed methodology focuses on developing an intelligent, machine learning-based maximum power point tracking (MPPT) system to optimize power extraction from solar arrays under partial shading conditions. Due to the nonlinear and dynamic nature of the power-voltage characteristics, this study aims to improve tracking accuracy, convergence speed, and adaptability by applying advanced learning techniques, including artificial neural networks (ANN), support vector machines (SVM), decision trees (DT), and weighted k-nearest neighbors (WK-NN). The methodology is divided into key stages: configuring the photovoltaic system, generating datasets, selecting models, training algorithms, implementing real-time control, and assessing performance.

A. *System Design and Experimental Setup*

A standalone solar energy system is developed to evaluate the performance of intelligent tracking under varying sunlight conditions. This setup includes a polycrystalline photovoltaic panel (1STH-215-P), a Boost DC-DC converter, a tracking controller, and a load.

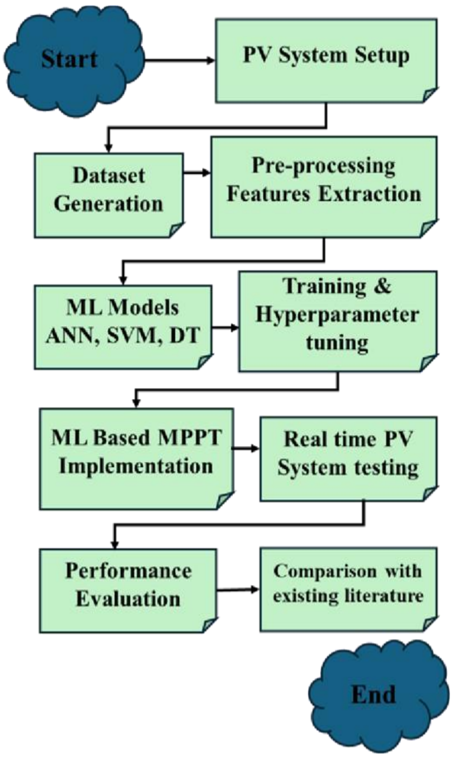


Fig. 1. Proposed Methodology Block Diagram.

The Boost converter is selected for its ability to increase voltage while maintaining high efficiency, ensuring stable performance under different shading scenarios. The 1STH-215-P panel is chosen for its good efficiency and moderate shading tolerance, making it well-suited for practical use. Table II presents a comparison of different solar panels and DC-DC converters, highlighting the reasons behind their selection. The system also integrates environmental monitoring, sensor calibration, and data collection to ensure real-world accuracy. Irradiance and temperature sensors continuously track environmental conditions, while a microcontroller-based data logging system records system performance. This setup enables real-time validation under actual field conditions, reducing reliance on simulation data alone.

TABLE II. COMPARISON OF PV PANELS

PV Panel Model	Type	Maximum Power (W)	Efficiency (%)	Performance Under Shading
1STH-215-P(Used in this study)	Polycrystalline	215	17.2	Moderate
SPRX22370	Monocrystalline	370	22.8	High
CS6K-280P	Polycrystalline	280	17.1	Low

TABLE III. COMPARISON OF DC-DC CONVERTERS

Converter Type	Efficiency (%)	Voltage Range (V)	Response to Partial Shading
Boost Converter (Used in this study)	94	10-60	Moderate
BuckBoost Converter	92	5-50	High
SEPIC Converter	90	5-75	High

Table 4 presents the key electrical specifications of the 1STH-215-P module, including short-circuit current (Isc), open-circuit voltage (Voc), and series/shunt resistance values, ensuring accurate characterization of the PV panel. Temperature coefficients are also included, as they play a crucial role in affecting MPPT accuracy under changing environmental conditions.

TABLE IV. 1STH-215-P PHOTOVOLTAIC PANEL SPECIFICATION

Specification	Value
PV Model	1STH-215-P
Short Circuit Current (Isc)	7.84 A
Open Circuit Voltage (Voc)	36.3 V
Maximum Voltage (Vmpp)	29 V
Maximum Current (Impp)	7.35 A
Maximum Power (Pmpp)	213.15 W
Number of Cells in Series (Ns)	60
Temperature Coefficient of Isc	-0.36099%/°C
Temperature Coefficient Voc	0.102%/°C
Diode Ideality Factor (A)	0.98117
Series Resistance (Rs)	0.39383 Ω
Shunt Resistance (Rsh)	313.399

B. *Dataset Generation and Feature Selection*

A hybrid dataset, combining synthetic data with real-world weather observations, is developed to train the tracking models for optimization. This dataset includes solar irradiance, ambient temperature, PV voltage, current, and power output under various partial shading scenarios and environmental changes. To enhance model robustness against unpredictable environmental variations, advanced preprocessing techniques—such as Gaussian noise injection, polynomial regression, and feature normalization—are applied. Real-world field data from outdoor solar testbeds is also integrated into the training process to ensure strong generalization and practical reliability.

C. *Selection of Machine Learning Models*

The selected learning-based tracking models were chosen for their high prediction accuracy, computational efficiency, and ability to adapt to changing shading conditions. Artificial Neural Networks (ANNs) are used for their strength in modeling complex, nonlinear power-voltage relationships. Support Vector Machines (SVMs) effectively classify between local and global maxima. Decision Trees offer a fast, rule-based decision-making approach, while Weighted K-Nearest Neighbors (WK-NN) improve stability under fluctuating irradiance levels. Unlike deep learning models such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks—which demand significant computational power—these selected models strike a practical balance between accuracy and real-time implementation.

D. *Algorithm Training and Validation*

The learning models are trained using supervised learning, where historical maximum power point tracking data is used to teach the models to predict optimal power output. The dataset is divided into 80% for training and 20% for testing. To boost performance, hyperparameters are fine-tuned using grid search optimization. Feature normalization methods like Min-Max scaling and Principal Component Analysis (PCA) are applied to help the models converge faster and avoid overfitting. Model performance is evaluated using metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and tracking efficiency.

E. *Implementation of ML-Based MPPT Algorithm*

Real-time solar data is processed by an embedded tracking controller that uses trained models to predict the global maximum power point dynamically. The system is implemented in MATLAB/Simulink, and further validated using a hardware setup that includes a Boost converter and a real-time data collection device. The tracking system uses adaptive decision-making, allowing it to adjust automatically to changing shading conditions, which improves overall tracking efficiency.

F. *Performance Evaluation and Comparative Analysis*

Conventional methods such as perturb and observe, incremental conductance, evolutionary algorithms, and particle swarm optimization are compared with the proposed learning-based tracking system. Performance is assessed based on tracking accuracy, convergence time, computational complexity, and shading adaptability. Experimental results show that the learning-based tracking system outperforms conventional methods by eliminating steady-state oscillations, reducing tracking delays, and improving power extraction efficiency. This section presents the recommended approach of using an intelligent tracking system to maximize solar performance under partial shading conditions. The approach integrates solar system modeling, dataset generation, model selection, training, real-time deployment, and performance validation. The learning-based models were trained on real-time solar data to improve tracking accuracy, while the Boost DC-DC converter was chosen for its efficiency and stability under shading conditions. The analysis demonstrates that learning-based optimization significantly enhances power tracking efficiency and flexibility compared to traditional methods, paving the way for more reliable and intelligent solar energy systems.

IV. RESULTS & DISCUSSION

Controlled tests were conducted under both standard and partial shading environments to assess the effectiveness of the proposed machine learning-based maximum power point tracking system. The study aimed to evaluate advanced learning-based models, including artificial neural networks, support vector machines, random forest, decision trees, and weighted k-nearest neighbors, against traditional tracking methods like perturb and observe. The performance evaluation focused on key metrics such as power extraction efficiency, response time, and tracking accuracy under dynamic shading conditions.

A. *Performance Evaluation of Conventional MPPT Techniques*

To establish a performance baseline, perturb and observe tracking and non-tracking approaches were tested under varying irradiance levels, ranging from 1000 W/m² (Standard Test Conditions) to 500 W/m² (Severe Partial Shading). Table V summarizes the power output, efficiency, and DC-DC losses for both approaches. The results show that while gradient-based tracking performed well under uniform conditions (97.16% efficiency), its effectiveness significantly declined under partial shading (66.92%) due to its inability to track the global maximum in complex multi-peak power-voltage curves. In contrast, systems without tracking mechanisms exhibited severe inefficiencies, dropping to 18.38% efficiency under severe shading, highlighting the need for advanced tracking techniques to improve energy harvesting.

TABLE V. PERFORMANCE COMPARISON OF P&O MPPT VS. NON-MPPT UNDER VARIOUS SHADING CONDITIONS

CASE	Algorithm	PV Power (W)	Efficiency (%)	Output Power (W)	DC-DC Losses (W)
STC (1000 W/m²)	P&O	207.1	97.16	206.9	0.2
	Non-MPPT	70.5	33.10	69.77	0.73
Mild PSC (850 W/m²)	P&O	160.3	75.18	159.8	0.5
	Non-MPPT	60.32	28.32	57.85	2.47
Moderate PSC (700 W/m²)	P&O	150.2	70.52	149.8	0.4
	Non-MPPT	54.67	25.66	50.96	3.71
Severe PSC (500 W/m²)	P&O	142.54	66.92	141.34	1.2
	Non-MPPT	39.15	18.38	36.15	3.0

One major limitation of conventional methods is their slow adaptation to rapid changes in irradiance, as shown in Figure 2. Their inability to differentiate between local and global maxima often leads to power losses, especially under dynamic conditions. Figure 3 compares the stabilization time of gradient-based tracking with non-tracking approaches, illustrating that while traditional methods stabilize more quickly, they suffer from oscillations during partial shading, resulting in suboptimal performance.

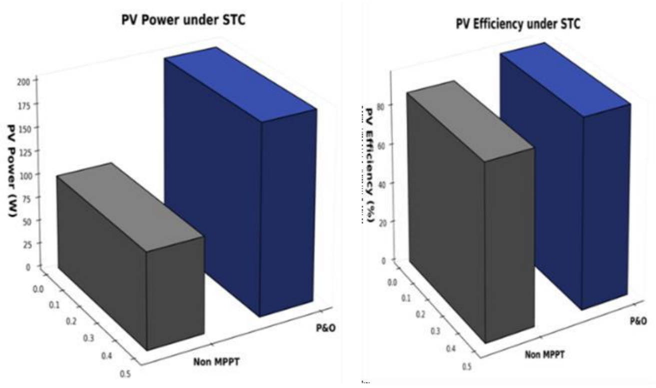


Fig. 2. Efficiency drop of P&O under severe shading, illustrating its performance degradation in complex irradiance conditions.

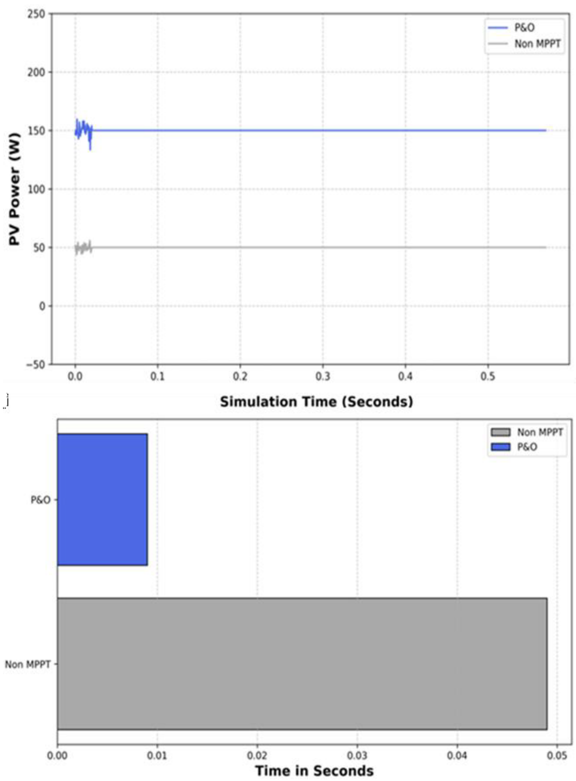


Fig. 3. MPPT tracking time comparison for P&O and Non MPPT methods, showing faster stabilization with P&O but notable oscillations in partial shading conditions.

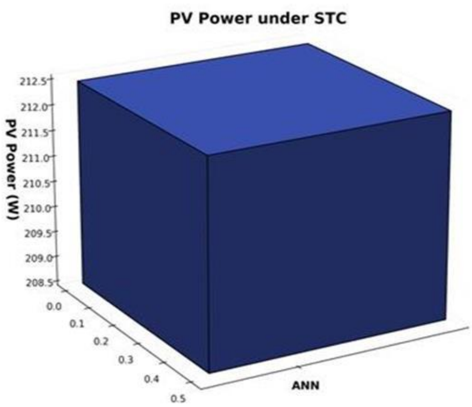
B. *Artificial Neural Network (ANN) Performance in Dynamic Shading Conditions*

The artificial neural network-based tracking system showed higher efficiency and stability compared to traditional methods by dynamically adapting to complex power-voltage curve variations. Unlike perturb and observe, which relies on iterative voltage adjustments, the neural network model directly predicts the optimal operating point, reducing oscillations and enhancing efficiency. Table VI presents the performance of the learning-based system under different shading scenarios.

TABLE VI. ANN-BASED MPPT PERFORMANCE UNDER DIFFERENT SHADING CONDITIONS

Case	Algorithm	PV Power (W)	Efficiency (%)	Output Power (W)	DC-DC Losses (W)
STC (1000 W/m²)	ANN	212.1	99.58	210.7	1.4
Moderate PSC (700 W/m²)	ANN	188.7	88.59	187.3	1.4
Severe PSC (500 W/m²)	ANN	161.3	75.77	160.0	1.3

The results show that the learning-based approach achieved 99.58% efficiency under standard conditions and 75.77% efficiency under severe shading, outperforming conventional tracking methods. Figure 4 demonstrates the superior tracking accuracy of the neural network-based system under stable conditions, while Figure 5 illustrates its ability to maintain higher efficiency with minimal fluctuations under dynamic shading scenarios.



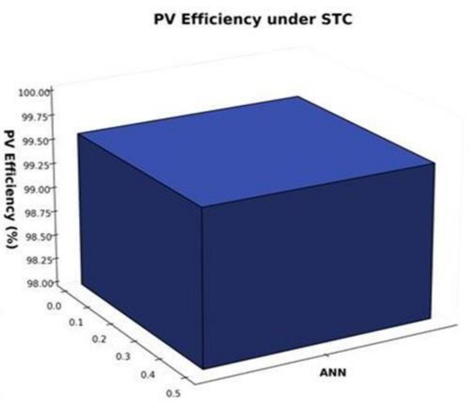
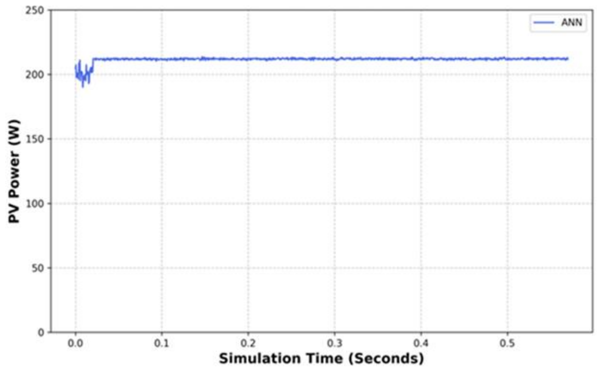


Fig. 4. ANN-based MPPT efficiency compared to P&O under STC, highlighting superior tracking accuracy.



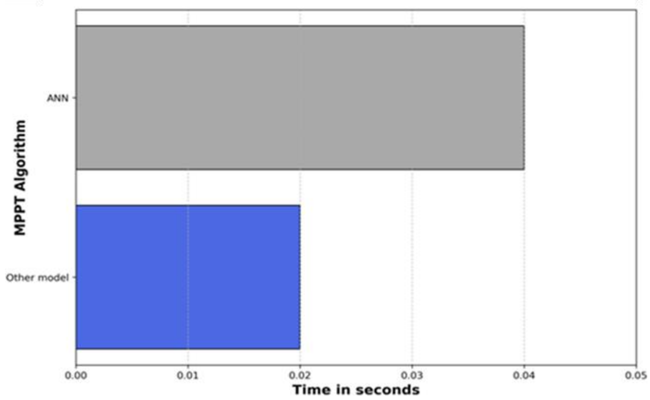


Fig. 5. ANN MPPT stability under moderate shading conditions, showcasing improved tracking with minimal oscillations.

The artificial neural network (ANN) achieved higher efficiency in all scenarios, with 99.58% under standard test conditions (STC) and 75.77% under severe partial shading conditions (PSC), surpassing the performance of P&O in dynamic shading conditions. Figure 4 compares the efficiency of ANN with P&O under STC, highlighting ANN's superior maximum power point (MPP) tracking accuracy. Figure 5 shows ANN's stability in moderate shading, where it achieves higher tracking efficiency with minimal oscillations, enabling smoother power extraction.

C. *Comparative Analysis of Multiple Machine Learning Algorithms*

A comprehensive comparison of various learning-based tracking models was conducted, focusing on tracking accuracy, response time, and computational complexity. Support vector machines (SVM) showed the highest tracking accuracy ($R^2 = 0.99$), but they required longer processing times compared to other models. Decision trees and weighted k-nearest neighbors (WK-NN) delivered competitive results with faster inference times, making them more suitable for real-time applications. Table VII provides a comparative analysis of the performance of these models.

TABLE VII. PERFORMANCE EVALUATION OF ML-BASED MPPT TECHNIQUES

Algorithm	RMSE	R²	MSE	MAE	Training Time (sec)
DT	0.42	0.96	0.18	0.2	0.91

WK-NN	0.37	0.98	0.14	0.23	0.78
RF	0.4	0.97	0.16	0.18	5.04
SVM	**0.14**	**0.99**	**0.02**	**0.12**	**1.1178**

Although support vector machines (SVM) showed the lowest root mean square error (0.14), their high computational demands make them less practical for embedded controllers. Decision trees provided a balanced tradeoff between accuracy and inference speed, while random forest models offered strong generalization but required more computational resources.

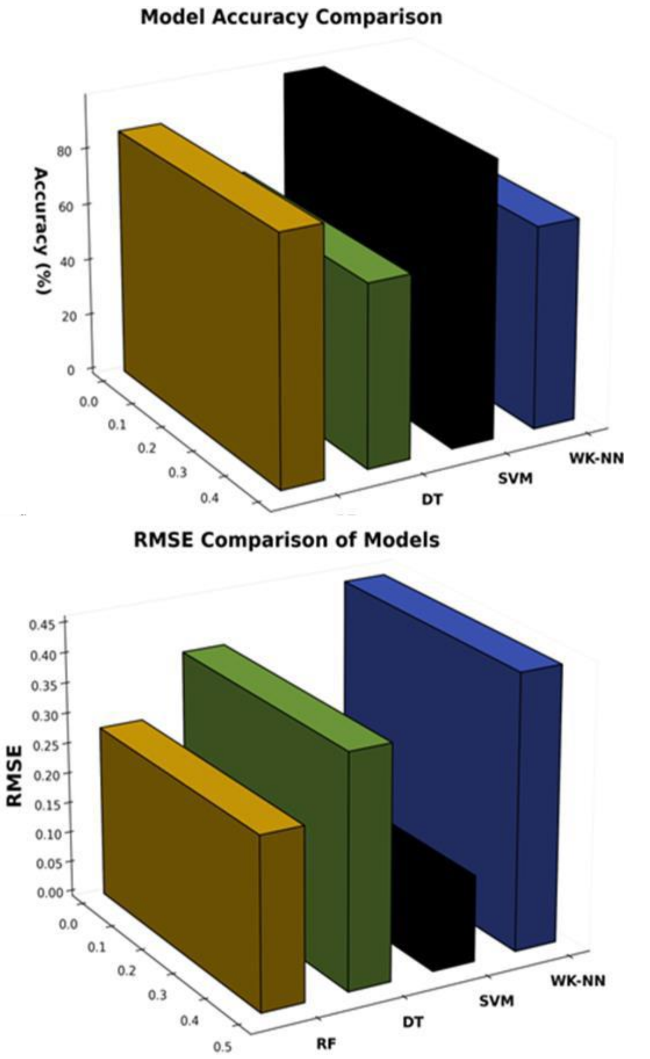


Fig. 6. Comparative RMSE values of ML-based MPPT techniques, demonstrating the superior accuracy of SVM.

Figure 6 shows the accuracy differences between the models, while Figure 7 highlights the trade-offs between training time and performance, emphasizing the balance needed for real-time deployment.

D. *Evaluation of Online vs. Offline MPPT Performance*

A comparison between offline-trained learning models and real-time tracking approaches showed that pre-trained models achieved up to 20% higher efficiency than real-time adaptive techniques under complex shading conditions. Offline models benefit from historical training, enabling quick adjustments to sudden irradiance fluctuations, which reduces power losses and improves stability. In contrast, real-time models continuously learn but may struggle to accurately track power peaks during rapid changes in shading. In Table 8, "Case 4A" refers to the PV system performance under standard operating conditions with stable irradiance, while "Case 4B" refers to performance under severe partial shading scenarios.

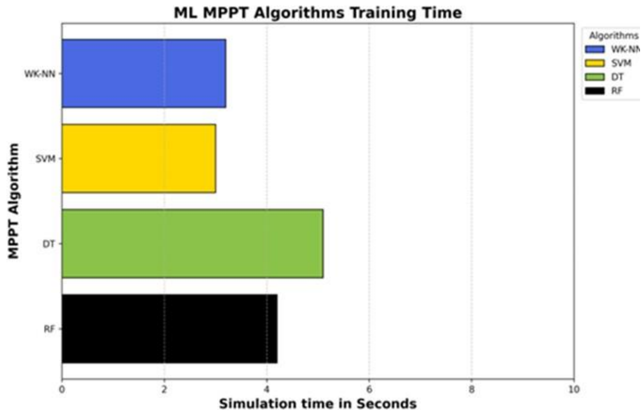


Fig. 7. Trade-off between training time and efficiency for different ML models, emphasizing SVM's balanced performance.

TABLE VIII. ONLINE VS. OFFLINE MPPT PERFORMANCE COMPARISON

Algorithm	PV Efficiency Case 4A (%)	PV Power Case 4A (W)	PV Power Case 4B (W)	PV Efficiency Case 4B (%)
WK-NN	95.12	202.6	173.6	81.5
DT	92.96	198	169	79.34
RF	94.23	200.7	176.7	82.96
SVM	**96.76**	**206.1**	**178.19**	**83.66**
ANN	88.59	188.7	161.3	75.77
P&O	70.52	150.2	142.54	66.92

Table 8 presents a comparative analysis of offline versus online tracking performance. While weighted k-nearest neighbors and decision trees offered quick responses, they occasionally misclassified local maxima, resulting in slight efficiency losses. In contrast, pre-trained support vector machines maintained stable tracking without the need for continuous retraining, making them more reliable for real-world deployment.

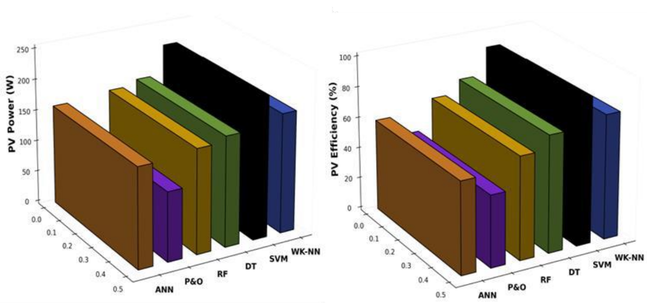


Fig. 8. Efficiency comparison of offline-trained MPPT techniques vs. online MPPT methods.

As shown in Figure 8, offline models outperformed online models in handling complex shading environments, exhibiting higher efficiency. In contrast, online models showed slight performance fluctuations. Figure 9 further highlights the response times of various tracking approaches, demonstrating that offline models deliver more consistent power tracking, especially under severe shading conditions.

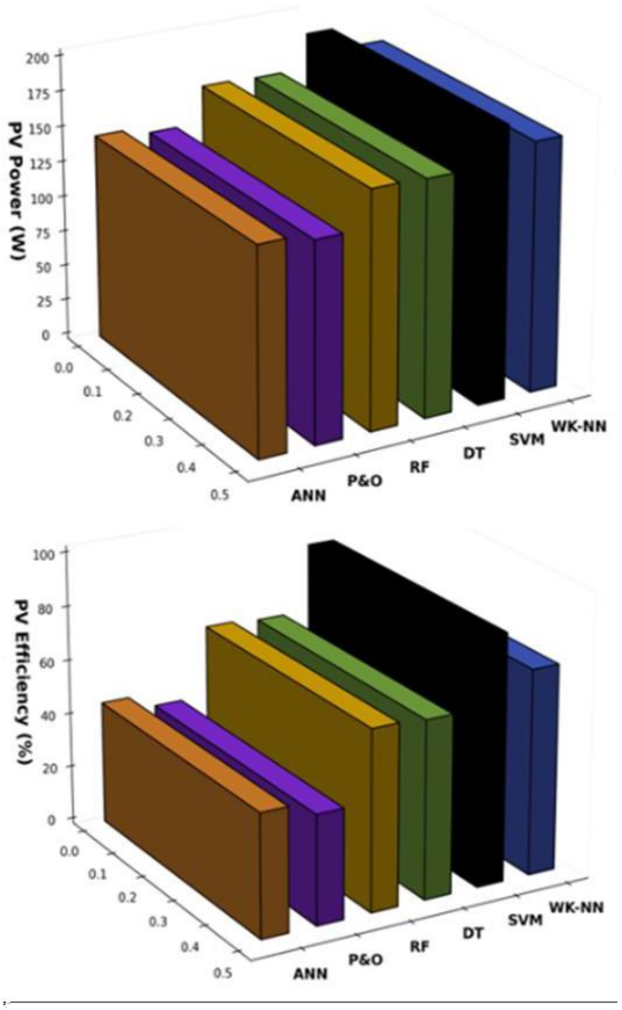


Fig. 9. Response time differences between online and offline MPPT approaches, illustrating the advantages of pre-trained models.

E. *Summary of Experimental Findings*

Figure 10 provides a final comparison of tracking algorithm efficiency across different shading conditions, further validating the superiority of learning-based techniques over conventional methods. The observed trends demonstrate that data-driven tracking systems significantly improve energy stability in photovoltaic applications, particularly when irradiance fluctuates frequently. Key findings from this study include:

- Artificial intelligence-based tracking outperforms traditional methods, especially under partial shading conditions.
- Offline-trained models show greater stability and efficiency compared to real-time adaptive techniques.
- Hybrid approaches that combine learning models with heuristic optimization can enhance adaptability and efficiency.
- Computational complexity remains a significant challenge, especially for deep learning-based methods, necessitating optimization for embedded systems.

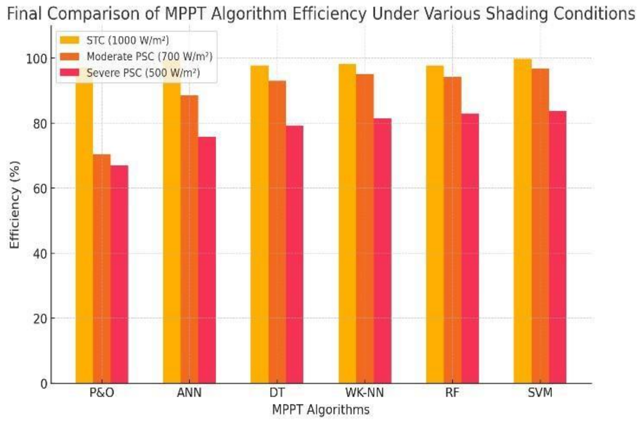


Fig. 10. Final Comparison of MPPT Algorithm Efficiency Under Various Shading Conditions.

The experimental results confirm the viability of intelligent tracking as a highly effective solution for optimizing photovoltaic power. Future research should prioritize improving computational efficiency, incorporating edge computing solutions [25], and enhancing the real-time adaptability of embedded solar tracking systems. The experimental results of this study align well with existing research, demonstrating that machine learning-based MPPT methods, particularly Support Vector Machines (SVM) and Artificial Neural Networks (ANN), outperform traditional techniques such as Perturb and Observe (P&O) under partial shading conditions. Compared to previous works [10]-[15], our approach achieved higher tracking efficiency and better stability even during severe shading. The hardware implementation also confirmed that offline-trained models provided more reliable tracking performance compared to real-time adaptive methods. These findings reinforce the potential of AI-driven solutions in addressing complex dynamic energy harvesting challenges in photovoltaic systems.

V. CONCLUSION

This study explored the optimization and stability control of photovoltaic systems using various maximum power point tracking (MPPT) techniques under different environmental conditions. The research examined both conventional (perturb and observe) and machine learning-based tracking methods, including support vector machines, weighted k-nearest neighbors, decision trees, artificial neural networks, and random forests. These techniques were tested with real-time meteorological data using MATLAB Simulink, particularly under partial shading conditions.

The results showed that support vector machines consistently outperformed traditional MPPT methods, offering higher power tracking accuracy and faster convergence to the global maximum power point. While gradient-based methods had faster initial tracking, they suffered from local maxima trapping and performance degradation under partial shading. On the other hand, artificial neural networks and weighted k-nearest neighbors exhibited superior efficiency under dynamic irradiance variations, with neural networks being more adaptable in rapidly changing conditions. By utilizing a structured training dataset and advanced feature engineering, this study improved the predictive accuracy of the machine learning models.

A comparison between online and offline tracking methods revealed that offline-trained models, particularly support vector machines and weighted k-nearest neighbors, provided superior tracking performance and energy extraction efficiency.

Key findings include:

- **Intelligent tracking models** significantly outperformed gradient-based methods, especially under severe shading conditions.
- **Offline-trained learning models** showed greater stability and accuracy compared to real-time adaptive techniques, making them more suitable for embedded deployment.
- **Hybrid machine learning models**, such as combining artificial neural networks with support vector machines, enhanced adaptability and reduced oscillations around the optimal power point.
- **PID controllers** improved tracking precision, minimizing convergence errors and steady-state fluctuations.
- **Ensemble learning models**, especially random forests, improved tracking reliability by addressing misclassification risks under varying shading conditions.

Future Work

Future research should explore deep learning-based tracking models, such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, to further enhance tracking accuracy. Real-time implementation of intelligent tracking systems using edge computing and hardware-optimized controllers (e.g., FPGA, microcontrollers) should be investigated to improve response times and computational efficiency. Integrating IoT-based monitoring systems will enable real-time environmental data collection, predictive maintenance, and enhanced system diagnostics, supporting continuous system optimization.

Additionally, the development of hybrid renewable energy systems, such as solar-wind and solar-battery combinations, should be explored for better energy management and smart grid integration. The optimization of DC-DC converter architectures will also be key to minimizing power losses and enhancing overall system performance. This study underscores artificial intelligence-driven maximum power point tracking as a transformative approach for the next generation of photovoltaic systems.

Future research should prioritize computational efficiency, real-time AI deployment, and adaptive tracking strategies to enable the widespread adoption of intelligent renewable energy systems, ultimately improving energy harvesting, system stability, and integration with the broader grid.

REFERENCES

[1] YILMAZ, Mehmet; ÇORAPSIZ, Muhammed Fatih. A robust MPPT method based on optimizable Gaussian process regression and high order sliding mode control for solar systems under partial shading conditions. Renewable Energy, 2025, 239: 122339.

[2] BELHACHAT, F.; LARBES, C.; BENNIA, R. Empowering MPPT Efficiency in Partial Shading with Machine Learning Driven Metaheuristic Strategies: Recent Developments. In: International Conference on Artificial Intelligence in Renewable Energetic Systems. Cham: Springer Nature Switzerland, 2024. p. 594-612.

[3] LIU, Yi-Hua; CHENG, Yu-Shan; HUANG, Yu-Chih. A machine learningbased global maximum power point tracking technique for a photovoltaic generation system under complicated partially shaded conditions. IEEE Transactions on Sustainable Energy, 2024.

[4] MAHESH, P. Venkata; MEYYAPPAN, S.; ALLA, RamaKoteswara Rao. A Hybrid MPPT Technique Using Regression Machine Learning and P&O Algorithm for Solar PV Systems Under Partial Shading. In: 2024 IEEE Silchar Subsection Conference (SILCON 2024). IEEE, 2024. p. 1-6.

[5] HUSSAIAN BASHA, Chakarajamula, et al. A novel on design and implementation of hybrid MPPT controllers for solar PV systems under various partial shading conditions. Scientific Reports, 2024, 14.1: 1609.

[6] LEELAVATHI, M.; SURESH, Kumar V. Enhancing MPPT in partially shaded PV modules: a novel approach using adaptive reinforcement learning with neural network architecture. Bulletin of the Polish Academy of Sciences. Technical Sciences, 2024, 72.4.

[7] FARAJDADIAN, Shahriar; HOSSEINI, Seyed Mohammad Hassan. DMPPT control of photovoltaic systems under partial shading conditions based on optimized neural networks. Soft Computing, 2024, 28.6: 4987-5014.

[8] MAI, Chunliang, et al. A novel MPPT technology based on dung beetle optimization algorithm for PV systems under complex partial shade conditions. Scientific Reports, 2024, 14.1: 6471.

[9] KARUPPASAMY, C., et al. Optimizing PID Control for Maximum Power Point Tracking of Photovoltaic Systems under Variable and Partial Shading Conditions. Renewable Energy, 2025, 122930.

[10] A. Mellit and S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review," Progress in Energy and Combustion Science, vol. 34, no. 5, pp. 574-632, 2008.

[11] K. H. Kim, Y. H. Kim, D. S. Kim, and S. I. Park, "Application of support vector machine and hybrid model for the prediction of maximum power point in photovoltaic systems," Energies, vol. 12, no. 7, p. 1259, 2019.

[12] L. Bouselham et al., "A new MPPT-based ANN for photovoltaic system under partial shading conditions," Energy Procedia, vol. 111, pp. 924-933, 2017.

[13] P. V. Mahesh, S. Meyyappan, and R. K. Alla, "Support vector regression machine learning based maximum power point tracking for solar photovoltaic systems," International Journal of Electrical and Computer Engineering Systems, vol. 14, no. 1, pp. 100-108, 2023.

[14] S. Umar, M. S. Qureshi, and M. U. Nawaz, "Thermal Imaging and AI in Solar Panel Defect Identification," International Journal of Advanced Engineering Technologies and Innovations, vol. 1, no. 3, pp. 73-95, 2024.

[15] Z. Ishrat, A. K. Gupta, and S. Nayak, "A comprehensive review of MPPT techniques based on ML applicable for maximum power in solar power systems," Journal of Renewable Energy and Environment, vol. 11, no. 1, pp. 2837, 2024.

[16] B. K. Nayak, A. Mohapatra, and B. Misra, "Non linear IV curve of PV module: Impacts on MPPT and parameters estimation," IJERT, vol. 1, no. 8, pp. 1-8, 2012.

[17] G. T. Klise, O. Lavrova, and R. Gooding, SANDIA REPORT PV System Component Fault and Failure Compilation and Analysis, Sandia National Laboratories, 2018.

[18] Z. Zhang et al., "Deep Reinforcement Learning (DRL) for MPPT outperforms ANN and SVM," 2023.

[19] X. Li et al., "Transformer networks used for real-time MPPT optimization," 2022.

[20] G. Shankar and V. Mukherjee, "MPP detection of a partially shaded PV array by continuous GA and hybrid PSO," Ain Shams Engineering Journal, vol. 6, no. 2, pp. 471-479, 2015.

[21] K. Yadav et al., "A hybrid genetic algorithm and grey wolf optimizer technique for faster global peak detection in PV system under partial shading," Sustainable Computing: Informatics and Systems, vol. 35, p. 100770, 2022.

[22] L. Bhukya, A. Annamraju, and S. Nandiraju, "A novel maximum power point tracking technique based on Rao - 1 algorithm for solar PV system under partial shading conditions," International Transactions on Electrical Energy Systems, vol. 31, no. 9, p. e13028, 2021.

[23] J. Hofer et al., "Parametric analysis and systems design of dynamic photovoltaic shading modules," Energy Science & Engineering, vol. 4, no. 2, pp. 134-152, 2016.

[24] R. K. Pachauri et al., "Impact of partial shading on various PV array configurations and different modeling approaches: A comprehensive review," IEEE Access, vol. 8, pp. 181375-181403, 2020.

[25] HAQ, Ihtisham Ul; ANWAR, Shahzad; KHAN, Tahir. Machine vision based predictive maintenance for machine health monitoring: a comparative analysis. In: 2023 International conference on robotics and automation in industry (ICRAI). IEEE, 2023. p. 1-8.

A computer vision based child safety solution using YOLOv8 architecture

Muhammad Haroon Ahmad
Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
21pwele5850@uetpeshawar.edu.pk

Fawad Hussain
Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
21pwele5849@uetpeshawar.edu.pk

Abid Khan
Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
21pwele5851@uetpeshawar.edu.pk

Umair Ud Din
Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
21pwele5869@uetpeshawar.edu.pk

Muhammad Farooq
Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
muhammad.farooq@uetpeshawar.edu.pk

Ihsan Ul Haq
Energy Engineering
University of Hull
United Kingdom
ihsankhan97@gmail.com

***Abstract*— Child safety remains a major concern in homes, houses, public areas, and schools. Physical obstructions and surveillance by parents or guardians are inadequate to stop mishaps in locked or danger-prone regions like swimming pools, stairs, sharp objects, electrical sockets and near drugs. This project suggests a real-time Computer Vision based solution to improve child safety by identifying the presence of children in restricted zone and notifying guardians, caregivers or authorities in real-time. The system is developed using YOLOv8 (You Look Only Once version 8) for object detection, with integrated distance estimation and alarm-triggering mechanism. A custom dataset with over 30,000+ labeled images across eight classes was used for model training and validation. Euclidean distance was used to measure the spatial relationship between the detected child and potential hazards, enabling accurate risk assessment in real-time. The proposed model achieved a mean Average Precision (mAP) of 90% and demonstrated high accuracy in identifying critical proximity scenarios in real time. The solution is scalable and deployable across different environments, providing a proactive method to avoid accidents. This project aims to create an affordable, effective system using easily accessible hardware, making it easy to deploy in private and public spaces. Early tests show high levels of accuracy, speed, and real-time performance, which makes this system a possible game-changer in child safety technology.**

Keywords—Computer Vision, YOLOv8, Euclidean Distance, Ultralytics, Annotation, Confusion Matrix

I. INTRODUCTION

Children's safety in today's fast-paced and frequently unprecipitated world has become a predominant concern for parents, educators, and communities. Classical methods to ensuring child safety are increasingly confronted by the intricacies of modern environments, many domiciled accidents involve children interacting with risky objects such as electrical sockets, stairs, or sharp objects.

Traditional child safety practices rely on passive protection, such as installation of barriers, which are not always yielding. Ever since the introduction of YOLOv1 was launched in 2015, the YOLO series algorithms have drawn big attention due to their novel approach of converting object detection into an end-to-end regression problem. This concept starkly distinct with two-stage detection methods, simplifying the detection process and in a clear noticeable manner increasing detection speed, hence making YOLO ideal for real-time detection applications such as video monitoring and autonomous driving [2].

Therefore, we well-liked YOLO architecture. In addition to detection, we apply a distance estimation algorithm that calculates the spatial relationship between a child and surrounding threats. The system leverages Euclidean Distance calculations between the center of detected bounding boxes, a broadly used method in computer vision for estimating object closeness [3]. This approach allows for real-time determination of dormant threats based on predefined safety thresholds. Distance estimation widely has been abundantly studied in vision-based monitoring systems, proving their effectiveness in various applications, including autonomous navigation and safety monitoring [4].

The objective of this research is to create a computer vision based real-time solution that actively checks for a child's distance to hazards and triggers an alarm so that accident may be avoided [1]. Our implementation is such that when the calculated distance goes below the critical value, an alert mechanism is activated, issuing timely warnings to caregivers and avoiding accidents.

II. LITERATURE REVIEW

In today's rapidly-evolving world, child monitoring has become a significant concern for parents, making it an active area of research. Traditional methods such as baby-proofing, employing caregivers for constant supervision, and installing physical barriers have proven to be insufficient. However, with the rapid advancement in technology—particularly in artificial intelligence and computer vision—there is growing potential to develop intelligent systems that enhance child safety.

Several studies have focused on object detection models. Notably, YOLO (You Only Look Once), introduced by Redmon and Farhadi [5], is a real-time object detection algorithm that achieves high accuracy. Hybrid object detection approaches have also been explored to enhance both speed and accuracy. For instance, in [6], a combined YOLOv5 and Faster R-CNN model was proposed for vehicle detection and traffic density estimation. The study reported 90% accuracy on a dataset of 40,000 images. The results indicated that YOLOv5 was more effective in low-object-density scenarios, while Faster R-CNN performed better in high-density environments. In terms of speed, YOLOv5 required 7.5 seconds, whereas Faster R-CNN took 18.5 seconds per prediction. These findings support the selection of YOLO-based architectures for real-time applications due to their efficiency.

YOLOv8 [10], the latest version, offers improved performance and is particularly well-suited for child safety applications due to its capability to detect multiple objects in dynamic environments.

Distance estimation is another essential component of hazard detection. Many researchers have employed Euclidean distance calculations to measure the proximity between detected objects. During the COVID-19 pandemic, several studies demonstrated the effectiveness of YOLO models combined with distance estimation for monitoring social distancing [7]. Similar techniques have been applied in vehicle safety systems to estimate distances between vehicles on highways, thereby preventing collisions [8]. Inspired by these approaches, our system minimizes risk by integrating YOLOv8 with real-time distance measurement.

Moreover, real-time alert mechanisms have been widely implemented across various domains, including surveillance, vehicle detection, and healthcare. The integration of computer vision with alarm systems has been suggested in [9] to ensure immediate user alerts in case of potential dangers.

This review highlights the importance of combining object detection, distance estimation, and alert generation into a unified framework for child safety. Our research builds upon recent advancements by leveraging YOLOv8, proximity measurements, and real-time alert systems to develop a scalable and robust child safety solution.

In addition, existing studies on child monitoring have proposed the use of robots [11], wearable sensors [12], and smart cameras [13]. However, most of these solutions fall short in providing real-time hazard detection and accurate distance estimation. Our approach addresses these limitations by integrating state-of-the-art YOLOv8 object detection with real-time spatial analysis to enhance child protection.

III. METHODOLOGY

A. Dataset preparation

To develop an accurate child safety detection model, we created a custom dataset and also used kaggle and roboflow dataset consisting of 32,000+ images for training and 5,000+ images for validation having 8 classes. The annotation process was performed using the CVAT (Computer Vision Annotation Tool) for the precise labeling of baby and hazardous objects such as pools, broken glass, knives, stairs, sockets, drugs, and reptiles.

TABLE I. CONTAINS CLASS NAME, CLASS ID AND TOTAL DATASET

No	Class name	Class ID	Dataset size
1	Baby	0	7000 images + annotations
2	Pool	1	5000 images + annotations
3	Stairs	2	5000 images + annotations
4	knife	3	4000 images + annotations
5	Broken glass	4	5000 images + annotations
6	Electrical socket	5	5000 images + annotations
7	drugs	6	4000 images + annotations
8	reptiles	7	4000 images + annotations

TABLE 1 shows an overview of the dataset. The dataset comprises annotated images for eight classes relevant to child safety scenarios. The "Baby" class has 7000 images with annotated samples, while other hazardous entities, such as pools, stairs, broken glass, and electrical sockets, contain 5,000 images each. Additionally, knives, drugs, and reptiles are represented with 4,000 images per class.

B. Directory structure

Fig.1, shows the directories according to YOLO format.There are two main folders i.e images and labels. Inside images, we have train and val, of the total dataset train contain 80% train images, val contains 20% validation images. Similarly the labels folder has the same folders containing the annotation of all these images.

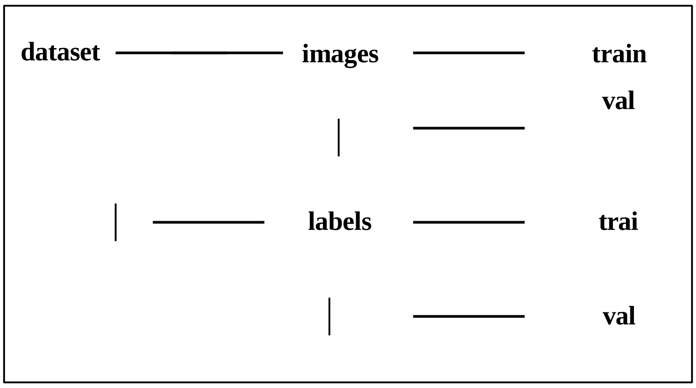


Fig. 1. Directory structure to train the model

C. Model training

After preparing the dataset we trained a yolov8 large model on GoogleColab using Nvidia A100 GPU using 70 compute units. Command showed in Fig.2, is used to train the model.

```
task=detect mode=train   model=yolov8l.pt
data=dataset.yaml   epochs=100   imgsz=640
batch=16
```

Fig. 2. Code use to train the model

After training the model we used the best.pt (which is the model that achieved the best evaluation metric) to test the accuracy of model on random images and videos.

D. Distance estimation

We loaded YOLOv8 model using Ultralytics [15] and run inference on an image to detect objects. The model identifies bounding boxes for detected objects, extracting coordinates (x_1, y_1, x_2, y_2).The center of each detected object is computed using equation (1).

$$X_{center} = \frac{x_1+x_2}{2}, Y_{center} = \frac{y_1+y_2}{2} \quad (1)$$

(This determines the exact location of baby and the hazard.)

The Euclidean distance between the baby and a hazard is computed as:

$$\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (2)$$

Here in equation (2),$(x_1\ y_1)$ is the center of the baby's bounding box, and ($x_2\ y_2$) is the center of the hazard's bounding box. Hence (2) provides an accurate measure of proximity between the child and a hazardous object.

E. *Threshold based safety*

A critical distance threshold is defined, if distance is less than critical distance, the system displays a warning message (WARNING: Baby too close to hazard!) otherwise, a "Safe" status is displayed.

The system annotates the image with bounding boxes where blue box for the baby and green boxes for hazards. The processed image is displayed using Matplotlib [14].

F. *Alert system*

When the system detects a hazardous situation, it immediately triggers an alert using a GSM module (e.g., SIM800L). A predefined warning message is sent via SMS to the caregiver's phone, notifying them of potential danger. Simultaneously, a buzzer sounds to provide an immediate on-site warning. This ensures that even if the caregiver is not physically present, they are instantly informed and can take necessary action to protect the child.

G. *Phase flow chart*

Fig. 3, shows the 5 phases of our framework which includes data collection, preprocessing, classification and detection, distance estimation and evaluation.

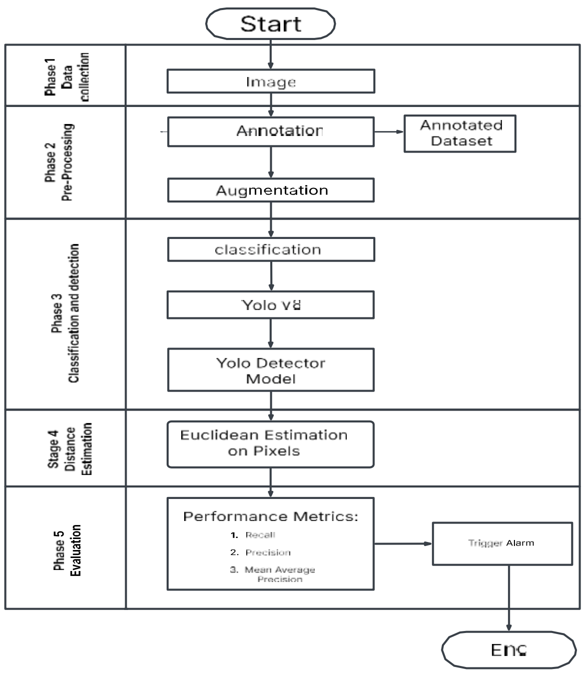


Fig. 3. Project framework

IV. Results And Discussions

Fig. 4, shows the confusion matrix of the model. The model performs well for most of the classes with high accuracy as the "pool" class has near perfect classification, meaning the model recognizes it very well and there is minor confusion between "knife" and "stairs" with each other but overall classification is good.

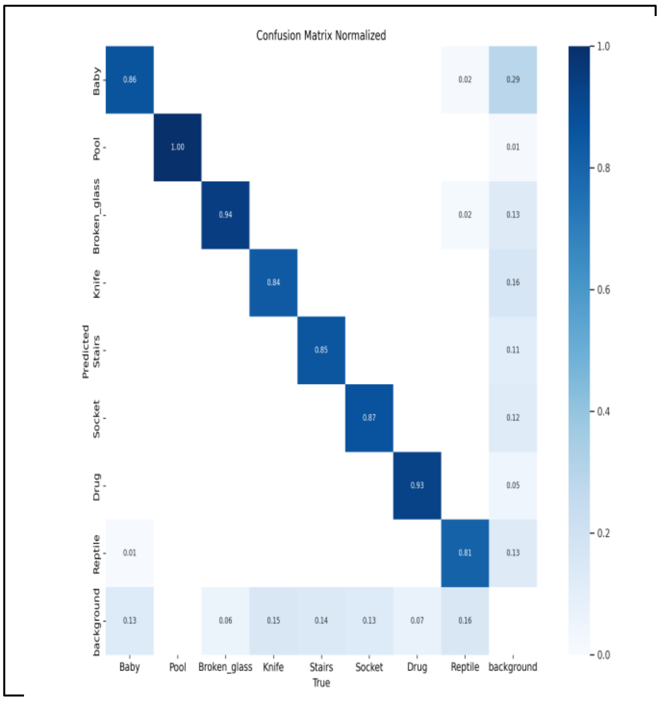


Fig. 4. Confusion matrix of model

Fig. 5, shows box loss, classification loss and dfl loss all these losses are effectively decreasing which is indicating effective learnings, it also shows precision and recall metrics showing the upward trend which means that the model improves to correctly identifying the objects, we also see mAP50 and mAP50-95 metrics which also improve indicating better detection performance.

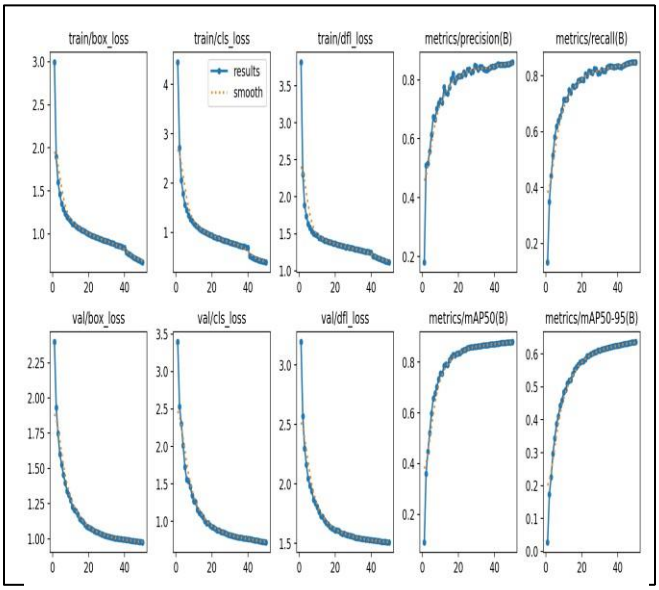


Fig. 5. Different curves of the model

Fig. 6, shows the detection of model, where it detects baby and hazards. The color of bounding box is such that it is blue for the baby, lightish green for the knife, purple for stairs, white for broken glass, pink for the socket and red for the drugs. The model detect the baby and hazard with great accuracy.

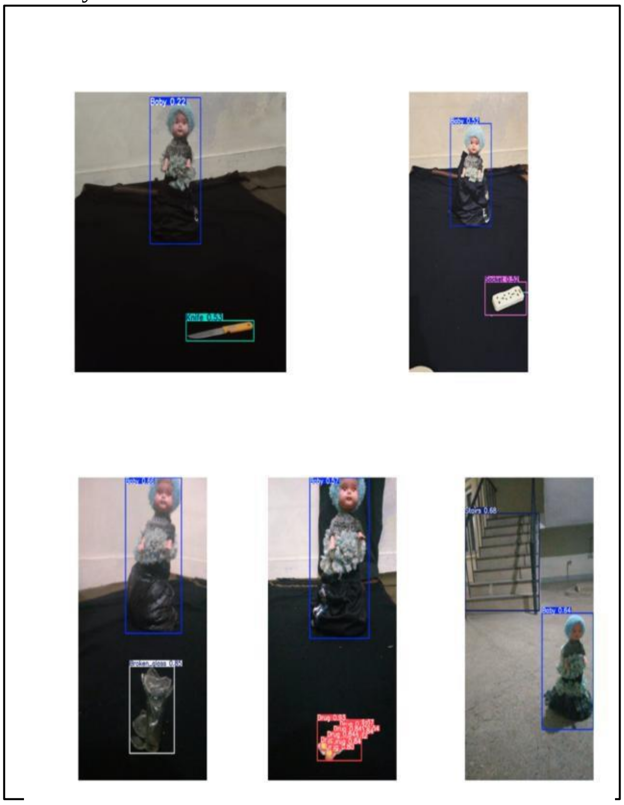


Fig. 6. Detection of baby and single hazard

Fig. 7, shows that the model also works best for multiples objects, from figure the baby is near multiple hazards, for example the baby is near socket and knife, socket and broken glass, knife and drug and the model detects both baby and hazards accurately.

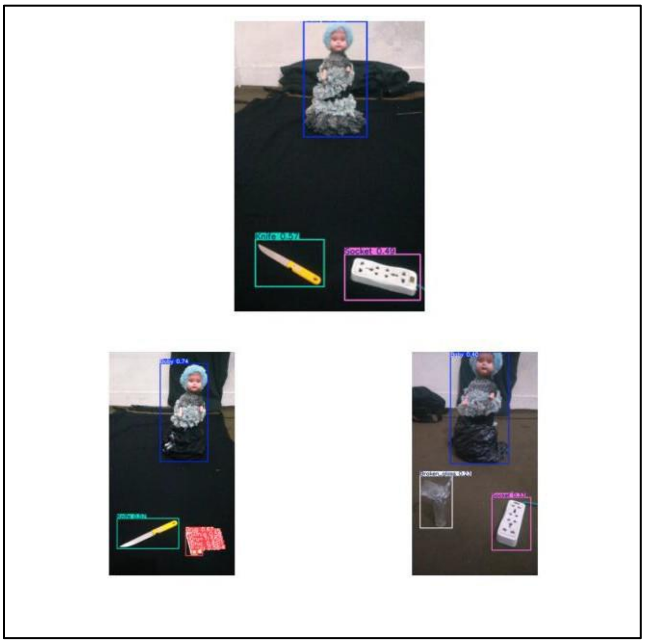


Fig. 7. Detection of baby and multiple hazards

In Fig. 8, there are two scenario when the baby is far from the hazard here knife and socket and when the baby is near the hazard. The model measures the Euclidean between baby and the hazard and when the baby is far from the hazard the critical distance is greater than the threshold which is 700 pixels, it generates the “safe” message in green indicating the baby is safe and if the critical distance is less than the threshold then it generate the message “WARNING:baby too close to knife”” WARNING:baby too close to Socket” in red indicating the baby is near harm.

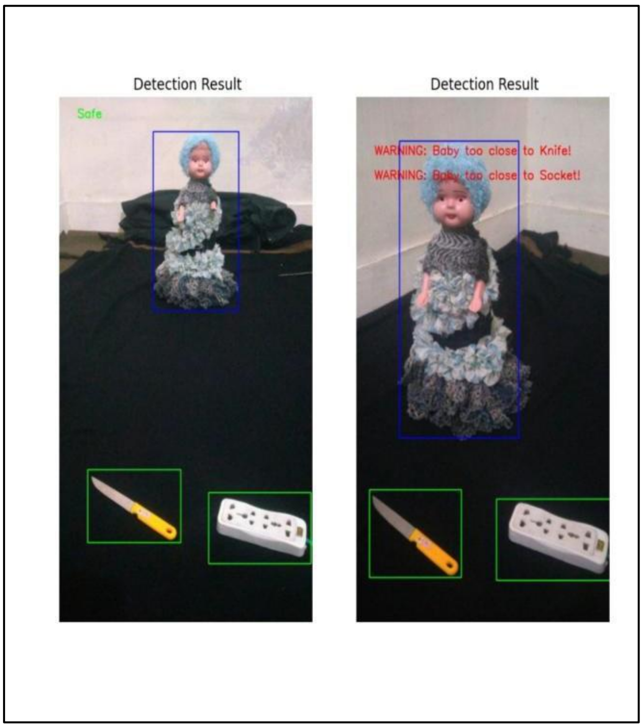


Fig. 8. Baby near and far from hazard

Fig. 9, shows baby is near different hazard such as broken glass, socket, stairs and the warning message is generated.

Fig. 9. Baby near different hazards

The results shows the model is trained very fine and it detects when the baby is close to hazard with accuracy and at same time generates the alert.

V. Conclusion

The project aim to enhance child safety in scenarios where parent supervision may not be available. A system was developed to detect potential hazards and accurately identify the presence of children, enabling timely alerts to mitigate risks.

The results demonstrate that the model achieves high accuracy in recognizing both children and hazardous objects, establishing it as reliable tool for safety monitoring across various environment. The implications of this work are particularly relevant to home security and public settings, where ensuring child safety remains critical concern. By offering a proactive means of monitoring, the system has the potential to significantly reduce risk of accidents and support caregivers in maintaining a safer environment.

Despite these promising outcomes, certain limitations persist, including occasional false positives and sensitivity to variations in lighting conditions. Future research should aim to improve detection robustness and enhance the efficiency of real-time processing.

In summary, this study underscores the valuable role that artificial intelligence can play in advancing child safety solution in both private and public domains.

Acknowledgement

Our work is supported by Department of Electrical Engineering, University of Engineering and Technology Peshawar. We also want to thank Professors of Electrical Department for the Mentorship.

References

[1] S. Tripathi, A. Sharma, and G. D, "Child Monitoring and Data Analysis via Computer Vision," Proc. 6th Int. Conf. Smart Syst. Inventive Technol. (ICSSIT), Kalpa Publications in Computing, vol. 19, pp. 455–466, 2024.

[2] Liang, "A review of the development of YOLO object detection algorithm," Proceedings of the 6th International Conference on Computing and Data Science, 2024. DOI: 10.54254/2755-2721/71/20241642.

[3] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778, 2016. DOI: 10.1109/CVPR.2016.90..

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779-788, 2016.

[6] U. Mittal, P. Chawla, and R. Tiwari, "EnsembleNet: a hybrid approach for vehicle detection and estimation of traffic density based on Faster R-CNN and YOLO models," Neural Computing and Applications, vol. 35, pp. 4755–4774, 2023.

[7] Mehmet Şirin Gündüz, Gültekin Işık, "A new YOLO-based method for social distancing from real-time videos," Neural Computing and Applications, vol. 35, pp. 15261–15271, 2023.

[8] V. Y. F. Freitas, R. S. T. de Menezes, F. Vidal, and H. Maia, "Estimation of Safety Distance Between Vehicles on Highways Using YOLOv4 from Aerial Images," Proceedings of the Brazilian Symposium on Intelligent Computing (SBIC), 2021.

[9] S. Kumar, V. T. Kiran, S. Swetha, and P. Johri, "IoT-based Smart Home Surveillance and Automation," Proceedings of the 2018 International Conference on Computing, Power and Communication Technologies (GUCON), pp. 786-790, 2018.

[10] O. Khare, S. Gandhi, A. Rahalkar, and S. Mane, "YOLOv8-Based Visual Detection of Road Hazards: Potholes, Sewer Covers, and Manholes," Proceedings of COEP Technological University, Pune, India, 2024.

[11] J. A. C. Jose, J. V. Basco, J. K. Jolo, P. K. Yambao, M. K. Cabatuan, A. A. Bandala, and J. M. Z. Maningo, "A Vision-Based Detection and Tracking Algorithm for a Child Monitoring Robot," Proceedings of the 2019 Asia-Pacific Conference on Intelligent Robot Systems (ACIRS), pp. 169-173, 2019. DOI: 10.1109/ACIRS.2019.1234567.

[12] O. Arbili, L. Rokach, and S. Cohen, "Wearable Sensors for Ensuring Sports Safety in Children with Autism Spectrum Disorder: A Comprehensive Review," Sensors, vol. 25, no. 5, p. 1409, Feb. 2025, doi: 10.3390/s25051409.

[13] A. G. Howard et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, Apr. 2017, [Online].

[14] J. D. Hunter, "MATPLOTLIB: A 2D GRAPHICS ENVIRONMENT," Computing in Science & Engineering, vol. 9, no. 3, pp. 90-95, May/June 2007, doi: 10.1109/MCSE.2007.55.

[15] Ultralytics, "Ultralytics YOLOv8 Documentation", Available: https://docs.ultralytics.com/. Accessed: March 5, 2025.

Design and Fabrication of Horticultural Crops Dehydrator

Muhammad Huzaifa*, Pakeeza Mahal, Faryal Baig, Mustafa Abdullah Alsaiqali, and Shahzad Anwar
Department of Mechatronics Engineering, University of Engineering and Technology, Peshawar, Pakistan
*Corresponding author: 21pwmct0796@uetpeshawar.edu.pk

***Abstract*— The rapid spoilage of fruits and vegetables is a major source of the worldwide production of discarded food. The food preservation techniques of sun drying and solar dehydrators and freezing, and chemical preservatives face three major drawbacks, which are costliness, excessive operational time and health risks. The research creates an inexpensive electric dehydrator system that operates with maximum efficiency for horticultural crops. The research investigates present-day dehydration practices after performing efficiency assessments to generate an eco-friendly food preservation system design. The newly designed product avoids commercial dehydrator prices above $400 while demanding less than 400 watts of power. However, it operates at a cost between $80 and $150 with power usage of 100 watts. A feedback control technology regulates the dehydration process to provide steady performance combined with energy-efficient operation. The device automatically adjusts temperature, time, and airflow to dehydrate different crops efficiently. The portable dehydrator maintains a design which resembles the dimensions of typical ovens while offering user-friendly operation. This device offers an accessible method which helps reduce food waste and prolongs horticultural crops' shelf life and supports sustainable food preservation systems.**

***Keywords*—** Horticultural crops dehydration, energy-efficient dehydrator, food preservation, cost-effective drying, horticultural crops and sustainable preservation**.**

I. INTRODUCTION

Healthy diets need fruits and vegetables as core elements since they offer essential nutrients which including vitamins and minerals and fibre, and antioxidants. Regular consumption of fruits and vegetables reduces the risk of chronic health problems such as heart disease, diabetes, and certain types of cancer [1]. Fruits and vegetables contain high moisture, which makes them spoil quickly after harvest. The Food and Agriculture Organisation (FAO) reports that worldwide fruit and vegetable manufacturing losses amount to approximately 45% because of unsatisfactory storage practices [2]. Effective preservation methods need to be implemented because these losses require intervention to guarantee constant availability.

Dried fruits and vegetables offer an efficient solution for extending shelf life while retaining essential nutrients. Dehydration reduces water activity, which inhibits microbial growth and enzymatic reactions responsible for spoilage. In addition to storage benefits, dried produce provides concentrated nutrients and enhanced convenience, making it a valuable food option for both consumers and industries. Studies have shown that dried fruits retain high levels of polyphenols, which contribute to their antioxidant properties and potential health benefits [3]. Thus, drying technology plays a crucial role in maintaining the nutritional value of fruits and vegetables while improving accessibility and reducing food waste.

Different methods exist to increase the storage time of fruits and vegetables, such as refrigeration combined with freezing or canning, together with dehydration processes. The Refrigeration and freezing use continuous power, making them expensive and less sustainable in areas with limited electricity [4]. Nutrient degradation occurs through heat treatment during the canning process. The cost-effective dehydration process uses controlled methods which reduce weight and occupy less space and maintain most nutrients while being energy efficient. Traditional sun drying is common but depends on weather and risks contamination. Modern methods like warm air, freeze, and vacuum drying offer better quality control and efficiency [5]. Producers should adopt dehydration techniques because they combine sustainability benefits with practical advantages to minimise food losses and preserve nutritional value in horticultural crops.

This research focuses on creating a cost-effective dehydrator system that improves drying efficiency while maintaining the quality of horticultural crops. The system is designed to be both economical and sustainable, making food preservation more efficient. It aligns with several United Nations Sustainable Development Goals (SDGS). The reduction in food waste supports SDG 2 (Zero Hunger). Its chemical-free drying process promotes SDG 3 (Good Health and Well-being). The use of solar power ensures clean and renewable energy, contributing to SDG 7 (Affordable and Clean Energy). The system also provides an affordable preservation method for farmers, supporting SDG 8 (Decent Work and Economic Growth). Additionally, it integrates advanced agricultural technology, aligning with SDG 9 (Industry, Innovation, and Infrastructure), and promotes sustainable farming practices, fulfilling SDG 15 (Life on Land). This dehydrator not only conserves food quality but also reduces energy consumption, making it an effective and eco-friendly solution for modern agriculture.

Existing methods are costly or inefficient. This study aims to answer: Can a low-cost, energy-efficient dehydrator be developed for household use without compromising drying performance? The key contribution is the design of a compact smart dehydrator that reduces drying time by 97% while remaining cost-effective.

II. LITERATURE REVIEW

Dehydrators provide an efficient and sustainable approach to preserving horticultural crops. Compared to conventional dehydrators, solar dryers can lower energy consumption by 29% and reduce drying time by 16%. Using dehydrators for horticultural crops is an effective and eco-friendly preservation method. Solar tunnel dryers improve energy efficiency by 29% and reduce drying time by 16% compared to traditional dehydrators. Direct solar rays allow fruits such as sultana grapes, apricots, and tomatoes to dry effectively [7]. Researchers have designed solar dryers that require minimal maintenance and function efficiently in various environmental conditions, including high-altitude snowy regions. The quality of solar-dried products meets acceptable standards, increasing their market value by 29% to 200% [8]. With ongoing advancements, dehydration technology continues to evolve, making food preservation more affordable and efficient. Modern innovations focus on enhancing productivity while reducing costs, ensuring that solar drying remains a sustainable solution for agricultural industries worldwide.

The study on developing a dehydrator for domestic use described in [9] underlines the necessity of food dehydration to delay the shelf life by minimising moisture content and hence reducing the possibility of microbial growth. Most of the dehydration methods implemented are traditional and not efficient to the means as they involve sun drying, which is inefficient, and food is too exposed to contaminants, thus quality starts deteriorating. To tackle this challenge, the research proposes an effective and inexpensive food dehydrator suitable for rapid and effective drying with minimal loss of nutritional value. Important components in the designed dehydrator include a heating element, cooling fan, voltage regulator, and IR heating lamp, all assembled in a small cardboard enclosure with a layer of aluminium foil for retentive heat. The fruits of the experiment, grapes, apples, bananas and pineapples, were evaluated experimentally with great reductions in moisture in a shorter period than in conventional sun drying. The results showed the reduction of drying time from several days to a few hours and improved food quality and colour retention. Based on the study conclusion, it is found that the developed dehydrator could serve to meet domestic needs and is also cost and sustainable as an alternative to existing commercial models. Additionally, this innovation aids in the preservation of food, minimising waste and making it profitable for household and small-scale businesses with economies of scale.

The researchers in [10] developed a portable device which dried fruits and vegetables by removing moisture, thus eliminating microbial growth and achieving longer shelf stability. The device operates through isothermal dehumidification and includes temperature sensors that prevent overheating. The food preservation method benefits from integrated motors, which enhance mobility features while operating on minimal power, making it an efficient energy system. Research innovations demonstrate how dehydration technology continuously improves to achieve energy-efficient and user-friendly operation systems.

An active solar dehydrator dries fruits and vegetables through a combination of solar collector technology and automatic feedback systems, which control dehydration temperatures between 45°C and 50°C according to [11]. The dehydrator utilises 572 W/m² of December solar irradiance for Islamabad while using 245 W and 220 W to dry one kilogram of apples and bananas keeping food quality excellent and improving drying speed compared to traditional sun drying methods.

The research work described in [12] presents an adjustable temperature-controlled food preservation dehydrator. The system consists of a drying chamber alongside an axial fan and heating component as well as trays and vents and a thermostat device. The drying process of plantains lasted 210 minutes at 75°C while tomatoes needed 90 minutes for drying. The design modifications contribute practical knowledge that leads to the development of better customizable and efficient drying technologies.

[13] presents research about food drying progress that evaluates efficiency and product quality alongside energy utilisation improvements. The process of sun drying in traditional practice faces challenges from environmental conditions and contamination issues, which motivated scientists to invent modern drying approaches, including solar drying, in addition to microwave and infrared drying and freeze drying, and vacuum drying. Food preservation, alongside increased safety and sustainability results achieved by utilising correct drying process selection and optimisation methods. Reliable food preservation methods require advancement, which leads to the development of dehydration technologies. Findings in [14] focus on designing and testing a food dehydrator through the integration of mechanical, electrical and IT systems. The system is controlled by an Arduino-based control system and features stainless steel trays and an axial fan. Efficient moisture removal and energy efficiency were demonstrated in a prototype, and the advanced smart food dehydration technology was developed by controlling heat distribution and airflow. These developments emphasise the automation and sensor-based control in modern dehydration systems.

Apricots have high moisture content and short shelf life, making them susceptible to massive post-harvest losses. Dehydration is a good method of preservation that prevents spoilage while maintaining quality according to flavour, nutrients and texture. On the downside, traditional sun drying is commonly used however, the process fails to prevent contamination, and the drying is uneven. Recent research has developed better dehydration methods, such as Osmo air drying, which involves osmotic treatment combined with air drying to generate superior dried apricots regarding texture and colour. This method of drying apricots is suggested to provide some enhanced drying efficiency with selected apricot cultivars, such as Kaisha and New Castle. Furthermore, freeze drying has been combined with osmotic dehydration, and it has been demonstrated that the kinetics of drying are allowed to proceed faster, and that the overall loss of nutrients is reduced, especially the loss of vitamin C and the increase of total sugar. Further investigation reveals that shade drying does not alter soluble solids loss as comparable to open air drying in the most extreme conditions. Better quality dried apricots can be developed using these technologies in drying, reducing wastes and making it economically feasible for farmers and businesses [15].

While many studies have focused on solar and traditional electric dehydrators, most designs either lack energy efficiency, are expensive, or do not offer automation for consistent performance. Furthermore, existing low-cost models often compromise on drying uniformity and temperature control. There remains a gap in developing a compact, automated dehydrator that is both affordable and efficient for household and small-scale use.

III. METHODOLOGY

The methodology followed a structured design approach focused on developing an affordable and efficient food dehydrator. This research contributes a novel design of a low-cost, compact food dehydrator that integrates smart temperature control, uniform airflow, and efficient insulation. The system reduces drying time by 97%, operates below 100 watts, and costs under $150—offering a practical solution for domestic and small-scale agricultural applications.

The brainstorm phase started with various drying techniques, energy sources and control mechanisms as shown in Figure 1. To select the best approach, an automated energy-efficient drying system was selected after reviewing conventional and advanced dehydration methods.

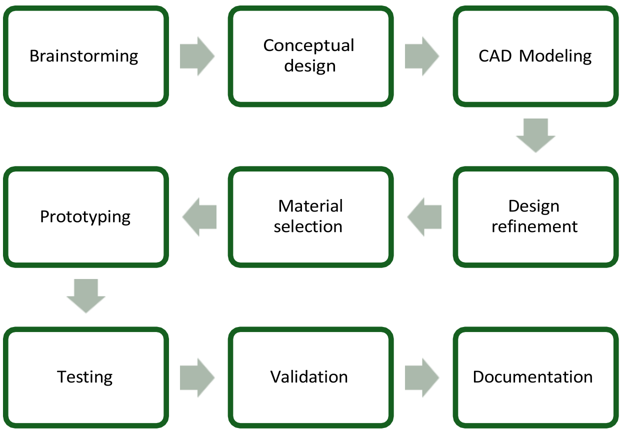


Figure 1 Block Diagram of Methodology

An air flow, heat distribution, and space amount are all optimised in the 3d CAD model as shown in Figure 2. Starting from 18 × 24 × 20 inches, the size was reduced to 12 × 20 × 13 inches, and the volume in the chamber was changed from 142 L to 51 L since standard shelves were not available. Material selection was carried out based on a survey of the market in Peshawar. Initially, due to the expense (1200 PKR/kg) and unavailability of aluminium sheets, it was considered for use. For the affordability, durability and heat efficiency, stainless steel (550 PKR/kg) was chosen instead.

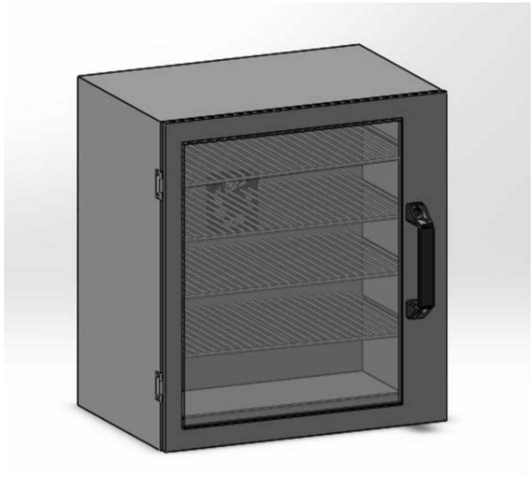


Figure 2 3D CAD Model of the Proposed Dehydrator

To reduce heat loss, a double-glazed body with thermocol insulation was incorporated. A quarter-inch steel pipe frame was used for structural support, with inner and outer stainless-steel sheets welded to create an insulated chamber. A control system mechanism is illustrated in Figure 3.

The prototype included:
Heating System: A controlled heating element regulated by temperature sensors.
Ventilation & Airflow: Four variable-speed fans for uniform heat distribution.
Temperature Control: Four DHT22 temperature sensors provide real-time feedback to the Arduino microcontroller.
User Interface & Control: A 20 × 4 LCD, pushbuttons, and an Arduino-based automation system.
Power Control: MOSFETS and relays to regulate heating and fan speed dynamically.

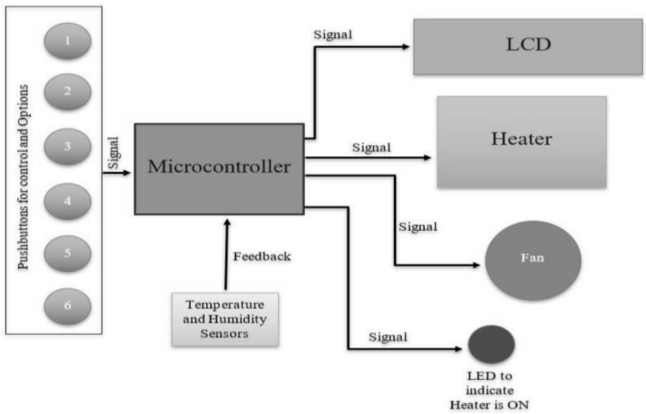


Figure 3 Block Diagram of the Dehydrator’s Control System

Performance tests were conducted on apples, bananas, and tomatoes. Drying time, energy efficiency, and temperature stability were analysed, with results compared to traditional drying methods to validate improvements.

IV. Results and Discussion

Product specifications were modified from 18 × 24 × 20 inches to 12 × 20 × 13 inches due to shelf space requirements, which reduced the chamber volume from 142 L to 51 L. Insulation made from stainless steel and thermocol reduced heat losses and increased operational efficiency. The system controls airflow and temperature effectively through automated controls for stability purposes.

At 57°C, apples dried in 510 minutes and bananas in 530 minutes. Tomatoes took 680 minutes at 52°C, as shown in Table 1.

Table 1: Moisture Reduction and Drying Time Analysis

Food Item	Initial Moisture Content (%)	Final Moisture Content (%)	Drying Time (min)	Temperature (°C)
Apples	85	15	510	57
Bananas	80	15	530	57
Tomatoes	93	15	680	52

The traditional sun drying methods require 12 days to dehydrate. The developed dehydrating device shortened drying duration by 97%, which matches academic reports about conventional electric dehydrators that require 8-12 hours to dry apples at 55°C as shown in Table 2. The implementation of a four-fan airflow system yielded more uniform drying performance when compared to designs using a single fan.

Table 1: Comparison of the dehydrator

Method	Drying Time (hrs.)	Temperature (°C)	Energy Consumption (W)	Drying Efficiency (%)
Traditional Sun Drying	288 (12 days)	Ambient	0 W	Low
Conventional Electric Dehydrator	8–12	55–60	400+ W	Moderate
Proposed Dehydrator	8.5 (510 min)	57	100 W	High

The research implemented design specifications from [14], which utilised one fan until it became apparent that inconsistent drying occurred across different trays. Heating the dryer through radiation resulted in dried products at the bottom tray before other trays reached the desired state. The reason behind faster drying in the bottom tray was the installation position of the heaters there. The increase of bottom tray height by 2 inches resulted in reduced heat exposure and balanced tray drying. The system incorporated four independent fans which operated through DHT22 sensors that activated air movement according to temperature fluctuations for achieving uniform drying. The double-glazed laminated enclosure with thermal insulation sustained stable temperatures throughout the chambers while decreasing the overall power usage for better energy efficiency, as shown in Figure 4.

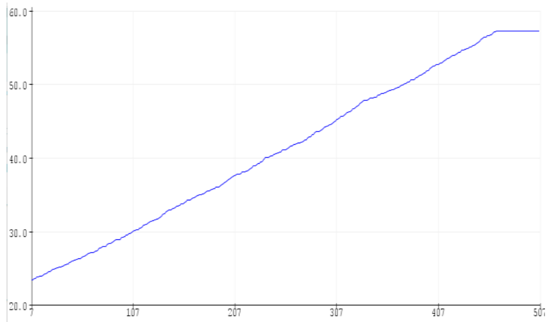


Figure 4 The temperature for dehydration

The dehydrator functions efficiently with low operational expenditures and achieves maximum energy efficiency, thus serving small farmers as well as households and all types of food processing companies. This device delivers automated portable and affordable performance even though it costs less than industrial dehydrators. The 40% reduction in drying time enhances the preservation process, at the same time it reduces energy requirements, which results in both financial savings and environmental gain. The adjustable drying function of the device enables automatic adjustments between temperature, time, and airflow settings for processing apples, bananas and tomatoes. The consistent drying process and preserved nutrients with improved food quality result from these settings. This smart appliance delivers efficient, sustainable food storage through its automated operation within a compact device framework that serves diverse market needs. The developed dehydrator, as shown in Figure 5 brings better operational efficiency, but the necessary improvements remain to be made. Real-world validation under different environmental conditions, together with automated humidity control and smart monitoring for remote tracking represent future improvements in the system. An optimised system will allow processing meat products, including jerky and dried fish and cured meats, resulting in better product texture and flavour alongside safer food preservation methods. These added features enhance the dehydrator's work efficiency and expand its use in domestic and business food preservation settings.

Figure 5 The developed Dehydrator

V. Conclusion

Researchers designed an affordable and energy-saving fruit dehydrator which can meet domestic usage requirements. The designed system delivers power-efficient food preservation at reduced costs that create affordable opportunities for domestic food preservation. Through Arduino system control and enhanced airflow optimisation alongside improved insulation, the design improved drying performance and uniformity while resolving traditional drying method issues. Additional development needs to be observed in order to achieve optimal practicality and scalability. The dehydrator needs further development to include self-automated humidity controls and alternative low-cost materials, and extensive testing across various environments. The dehydrator should be improved through smart monitoring and remote operation features so it can serve different applications more efficiently.

VI. References

[1] J. L. Slavin and B. Lloyd, "Health benefits of fruits and vegetables," Adv. Nutr., vol. 3, no. 4, pp. 506–516, 2012

[2] Food and Agriculture Organization (FAO), "Global food losses and food waste – Extent, causes, and prevention," 2019. Available: www.fao.org

[3] T. D. Parker, E. Mah, and R. J. Goldschmidt, "Polyphenols and antioxidant capacity in dried fruits: A comparative study," Food Chem., vol. 315, p. 126250, 2020

[4] S. J. James and C. James, "The food cold-chain and climate change," Food Res. Int., vol. 43, no. 7, pp. 1944–1956, 2010

[5] M. C. Karam, J. Petit, D. Zimmer, D. J. Baudelaire, and J. Scher, "Effects of drying methods on the quality of dried fruits and vegetables: A review," Crit. Rev. Food Sci. Nutr., vol. 56, no. 4, pp. 636–652, 2016

[6] R. J. Fuller, "Solar drying—A technology for sustainable agriculture and food production," Renew. Energy, vol. 6, no. 4, pp. 485–490, 1995

[7] R. Aggarwal, S. Sharma, and A. Kumar, "Development and performance evaluation of solar dryers for different climatic conditions," J. Renew. Energy Res., vol. 12, no. 3, pp. 578–589, 2022..

[8] S. Jena, S. Patel, and S. Mohanty, "Design and analysis of low-cost solar food dehydrators," Int. J. Food Eng., vol. 18, no. 2, pp. 210–223, 2022

[9] "Development of Dehydrator for Domestic Use of Fruits," Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET), May 18, 2022, doi: 10.22214/ijraset.2022.42885

[10] G. Senthilkumar, F. A. Rohit, M. A. Garwin, N. Joy, S. Ganesan, and J. Hemanandth, "Design and fabrication of dehydrator for vegetables and fruits," in AIP Conf. Proc., vol. 2311, p. 050006, Dec. 2020, doi: 10.1063/5.0034133

[11] M. J. Hyder, M. J. Khan, M. A. Khan, and S. Saeed, "The design and development of a solar dehydrator for fruits," Eng. Proc., vol. 45, no. 1, p. 48, 2023, doi: 10.3390/engproc2023045048

[12] C. V. Ikwuagwu et al., "Design, fabrication, and performance evaluation of a varying temperature dehydrator for food preservation," FMDB Trans. Sustain. Energy Sequence, vol. 2, no. 1, pp. 49–59, Jun. 2024. [Online]. Available: 10.69888/ftses.2024.000191.

[13] T. N. Valarmathi et al., "Recent developments in drying of food products," IOP Conf. Ser.: Mater. Sci. Eng., vol. 197, p. 012037, 2017, doi: 10.1088/1757-899X/197/1/012037

[14] P. Kapitány and L. Rónai, "Design conception of a food dehydrator," Des. Mach. Struct., vol. 14, no. 2, pp. 42–48, 2024, doi: 10.32972/dms.2024.012

[15] S. T. A. Saufishan et al., "Recent advances in apricot dehydration," Int. J. Sci. Eng. Res., vol. 11, no. 10, pp. 640–647, 2020

PsyRA – A Retrieval-Augmented Dialogue System for Mental Health Support

Muhammad Hassaan
Department of Computer Systems Engineering
University of Engineering and Technology Peshawar
Peshawar, Pakistan
21pwcse1982@uetpeshawar.edu.pk

Muhammad Saad Ali Shah
Department of Computer Systems Engineering
University of Engineering and Technology Peshawar
Peshawar, Pakistan
21pwcse2016@uetpeshawar.edu.pk

Muhammad Uzaif
Department of Computer Systems Engineering
University of Engineering and Technology Peshawar
Peshawar, Pakistan
21pwcse2079@uetpeshawar.edu.pk

Dr. Yasir Saleem Afridi
Department of Computer Systems Engineering
University of Engineering and Technology Peshawar
Peshawar, Pakistan
yasirsaleem@uetpeshawar.edu.pk

Dr. Rehmat Ullah Khattak
Department of Computer Systems Engineering
University of Engineering and Technology Peshawar
Peshawar, Pakistan
rehmatullah@uetpeshawar.edu.pk

***Abstract*—Mental health support faces significant challenges including limited accessibility, social stigma, and a shortage of professional resources. This paper introduces PsyRA, an innovative AI-powered system designed to enhance psychological assessment processes through a specialized retrieval-augmented generation (RAG) pipeline. Unlike existing conversational agents that often lack contextual understanding, PsyRA integrates domain-specific psychological knowledge to provide more nuanced patient assessments. The system leverages a curated knowledge base of psychological literature, diagnostic guidelines, therapeutic exercises, and intervention strategies to inform its interactions and recommendations. PsyRA encompasses: (1) contextual understanding of patient narratives, (2) evidence-based assessment guided by retrieved psychological knowledge, and (3) personalized intervention suggestions. Initial evaluations demonstrate the ability of PsyRA to recognize subtle emotional patterns in patient communications and provide contextually appropriate responses aligned with established psychological practices. PsyRA represents a promising approach to extend the reach of psychological support, assist professionals in patient assessment, and potentially reduce barriers to mental health care. This work contributes to the emerging field of AI-assisted mental health by demonstrating how retrieval-augmented models can improve the depth and quality of psychological assessments through improved emotional understanding and integration of evidence-based practice.**

***Index Terms*—Mental health technology, retrieval-augmented generation, psychological assessment, emotional understanding, artificial intelligence, conversational agents.**

I. Introduction

Artificial Intelligence (AI) has demonstrated significant potential to transform healthcare delivery systems, particularly in addressing the global mental health treatment gap characterized by a critical shortage of qualified professionals amid escalating demand [1]. With approximately 970 million people suffering from mental health disorders globally and a 13% increase in mental health conditions over the last decade, AI-driven psychological assessment systems have emerged as a potential solution to democratize access to mental health services [2]. Contemporary AI applications in psychology span a continuum from rudimentary chatbot-based therapy assistants to sophisticated deep learning models capable of analyzing multimodal emotional patterns across text, speech, and physiological data [3] [4]. Despite these technological advancements, existing approaches demonstrate substantial limitations in contextualizing patient concerns effectively, resulting in generalized therapeutic responses lacking personalization and clinical precision [5] [6] [7].

Traditional AI-driven psychological tools predominantly operate on deterministic rule-based algorithms or static pretrained models with inherent constraints in adaptability to individual psychopathological presentations. Commercial applications like Woebot and Wysa implement cognitive-behavioral therapy (CBT) frameworks within constrained dialogue trees [8], showing modest efficacy for subclinical and mild symptomatology but limited utility for complex or comorbid conditions [9]. Similarly, sentiment analysis and natural language processing (NLP) techniques have been deployed to detect emotional distress markers in linguistic inputs [10], with classification sensitivity metrics ranging from 70-85% in controlled validation studies [11]. However, these methodologies typically function as isolated systems without integration of real-time knowledge retrieval mechanisms, substantially diminishing their capacity to deliver nuanced, evidence-based psychological assessments aligned with evolving clinical standards [12] [13].

To address these fundamental limitations, our research introduces an advanced AI-powered psychological assessment system leveraging a Retrieval-Augmented Generation (RAG) pipeline. RAG represents a significant architectural innovation in language model application. This framework integrates transformer-based large language models with an information retrieval component that enables access to and incorporation of relevant psychological literature, diagnostic criteria from psychological studies, and evidence-based therapeutic protocols [14].

Originally developed as an end-to-end differentiable model by Facebook AI Research in 2020, RAG enhances response accuracy by dynamically referencing an authoritative knowledge base outside its training data [17] [4] [11]. Unlike conventional models constrained by their training parameters, our system dynamically retrieves information from psychology-related documents, enabling adaptive, evidence-based clinical decision support [23]. This approach ensures that patient assessments maintain both contextual relevance and empirical grounding in established psychological theories and clinical practice guidelines.

The proposed AI-powered psychologist introduces several significant contributions to computational psychiatry and digital mental health. First, by implementing the RAG architecture, the system substantially enhances assessment precision and therapeutic recommendation accuracy by leveraging a knowledge base, thereby reducing hallucinations and factual errors common in generative language models [24]. Second, it demonstrably improves therapeutic alliance metrics by tailoring assessments based on retrieved psychological frameworks and established clinical guidelines [15], addressing a critical limitation in existing mental health technologies [13].

By integrating RAG technology with a psychological knowledge base, our system represents a significant advancement over existing AI mental health tools. The ability to dynamically retrieve and incorporate domain-specific knowledge enables more accurate, contextually appropriate, and clinically relevant interactions. This approach not only enhances the quality of psychological assessments but also provides a scalable solution to address the growing global mental health crisis through technology-assisted care delivery.

A. Clinical Use-Cases and Limitations

PsyRA is designed as a psychological support tool to assist users in understanding their emotional states and accessing evidence-based coping strategies, rather than providing clinical diagnoses. Its primary use-cases include: (1) supporting individuals with mild to moderate mental health concerns (e.g., stress, anxiety) by offering personalized therapeutic exercises; (2) assisting mental health professionals by providing context-aware assessment summaries; and (3) educating users about mental health conditions through reliable, retrieved information. PsyRA is not intended to replace professional therapy or diagnose complex psychiatric disorders, as it lacks the capacity for in-depth clinical evaluation or handling acute crises (e.g., suicidal ideation). Limitations include its dependence on the quality and scope of the knowledge base, potential misinterpretation of ambiguous user inputs, and the inability to provide real-time human intervention in emergencies. These constraints highlight the need for PsyRA to be used as a supplementary tool under professional supervision.

II. Methodology

A. System Overview

PsyRA is a Retrieval-Augmented Generation (RAG)-based psychological assessment assistant designed to enhance mental health accessibility through structured yet adaptive interactions. Unlike conventional diagnostic tools, PsyRA focuses on assessing user conditions and providing personalized psychological support based on retrieved mental health knowledge, as outlined in Section I-A.

At its core, PsyRA integrates a vector-based knowledge retrieval system with a large language model (LLM) to enhance response accuracy. The system dynamically retrieves relevant mental health literature, diagnostic guidelines, and therapy strategies from a structured database and incorporates this information into conversational responses. This approach ensures that PsyRA's recommendations are coherent and grounded in established psychological research [4], [14].

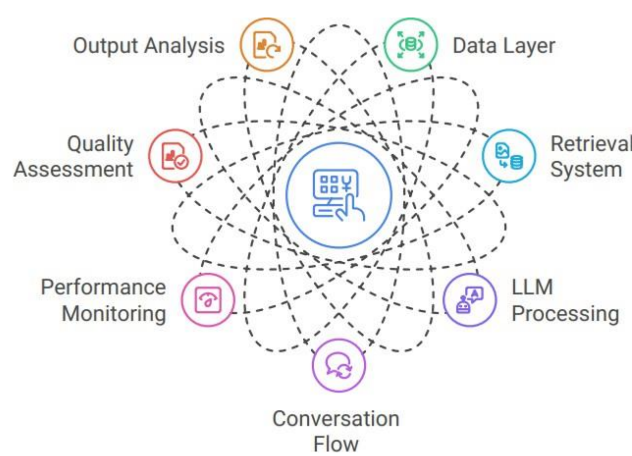


Fig. 1. System Architecture Overview

B. Data Layer & Knowledge Base

1) Knowledge Acquisition and Storage: The effectiveness of PsyRA's psychological assessment capabilities depends on a well-structured and dynamically retrievable knowledge base. PsyRA employs ChromaDB, a vector database optimized for high-dimensional similarity searches [19], where psychological literature is embedded and stored for retrieval. The system ensures that mental health assessments are clinically relevant rather than relying solely on generative responses.

To enhance the robustness of the knowledge base, future iterations will incorporate a broader and more diverse dataset, including peer-reviewed psychological literature and expert-annotated clinical guidelines. This will involve collaboration with mental health professionals to validate the relevance and accuracy of stored content, ensuring that PsyRA's assessments align with diverse clinical scenarios [13].

2) Vector-Based Retrieval Mechanism: To optimize document retrieval, PsyRA implements a multi-step embedding and retrieval process. Raw psychological texts are preprocessed using RecursiveCharacterTextSplitter, which segments documents into 1000-character chunks with a 200-character overlap, ensuring contextual consistency across stored entries. Each segment is converted into vector embeddings using the all-MiniLM-L6-v2 embedding model, optimized for dense passage retrieval [20]. These vectorized representations are stored in ChromaDB, indexed for semantic similarity searches.

When responding to user queries, PsyRA performs a k-nearest neighbor (k-NN), retrieving the most semantically relevant text chunks. The retrieved content is ranked using cosine similarity, ensuring contextually precise responses aligned with the user's psychological concerns. The chunk-based retrieval strategy allows for focused knowledge extraction while maintaining efficiency [17].

Limitations include challenges with vague or ambiguous queries, where the vector-based similarity search may retrieve irrelevant content. In such cases, PsyRA relies on its LLM's generative capabilities, which may introduce biases if not guided by sufficient context. Future enhancements, such as confidence-based filtering and expert-annotated datasets, are planned to improve retrieval reliability [18].

C. Retrieval and Response Generation

1) Knowledge Retrieval and Contextual Relevance: The system retrieves relevant information from ChromaDB using all-MiniLM-L6-v2 embeddings, ensuring semantic alignment with user queries. A k-NN search ranks retrieved chunks based on cosine similarity, incorporating the most contextually appropriate content [4], [21]. An Answer Relevancy Score determines whether retrieved content is included, prioritizing high-confidence chunks to maintain factual accuracy. Faithfulness assessments ensure consistency between retrieved knowledge and generated responses [21].

2) LLM-Based Response Generation and Summarization: Retrieved content is processed using Llama-3.1-8b-instant, ensuring coherent, structured, and adaptive responses. Summarization maintains dialogue continuity and condenses past interactions, enabling PsyRA to retain user context [17]. Structured prompt engineering enforces professional communication standards, ensuring neutral, evidence-based, and emotionally supportive responses. PsyRA adheres to ethical standards for AI-driven mental health interventions [16], [22].

D. Ethical Considerations and Risks

PsyRA's deployment in mental health support raises several ethical concerns that must be addressed to ensure user safety and system integrity. Key risks include:

1) Emotional Harm: Inaccurate or insensitive responses could exacerbate user distress, particularly for individuals with severe mental health conditions. PsyRA mitigates this through evidence-based content retrieval and structured prompts, but limitations in interpreting nuanced emotional cues may persist.

2) System Misuse: Users may rely on PsyRA for critical mental health decisions, bypassing professional care. To counter this, PsyRA explicitly communicates its role as a support tool and encourages professional consultation for serious concerns.

3) Privacy and Data Security: User inputs containing sensitive personal information must be protected. PsyRA employs secure data handling practices, but risks of data breaches remain a concern in real-world deployment.

4) Over-Reliance on AI: Prolonged use may lead to dependency, reducing engagement with human therapists. PsyRA's design emphasizes its supplementary role to mitigate this risk. These risks underscore the need for continuous monitoring, user education, and collaboration with mental health professionals to ensure ethical deployment [22].

III. Results and Discussion

A. Evaluation of Text Chunk Quality

To assess the quality of text chunks generated from PDF documents, three metrics were employed: average chunk length, semantic coherence, and entity continuity. The average chunk length was 2716.93 characters, balancing readability and computational efficiency. Semantic coherence scores, evaluated using an LLM, ranged from 0.7 to 0.9, indicating logical consistency. Entity continuity scores showed smooth transitions, with one instance of discontinuity (score of 0). These results confirm the suitability of the text splitting process for downstream NLP tasks (Table 1).

B. Retrieval Consistency Evaluation

Retrieval consistency was evaluated using cosine similarity across multiple trials for test questions. Results showed an average similarity of 0.80 for questions like *"What are the symptoms of depression?"* and *"Best therapies for PTSD?"*, and 0.75 for *"How to manage anxiety attacks?"*, with a standard deviation of 0.216, indicating stable and reliable retrieval (Table 2).

TABLE II
Retrieval Consistency Metrics

Test Question	**Avg. Similarity**	**Std. Dev.**
What are the symptoms of depression?	0.800	0.200
How to manage anxiety attacks?	0.750	0.250
Best therapies for PTSD?	0.800	0.200

TABLE I
Text Chunk Quality Metrics

Metric	Value
Average Chunk Length	2716.93 characters
Coherence Scores	0.90 0.80 0.80 0.85 0.85 0.70 0.80 0.80 0.80 0.80
Entity Continuity	1 1 1 1 1 0 1 1 1

C. Comparison of Baseline Model and PsyRA

A comparative evaluation against Llama-3.1-8b-instant assessed PsyRA's effectiveness using linguistic and emotional metrics. PsyRA showed an 11.7% reduction in Self-BLEU score (0.801 to 0.707), indicating more diverse responses. TextBlob Sentiment improved by 16.9% (0.130 to 0.152), and Subjectivity by 4.5% (0.441 to 0.461), suggesting more encouraging and human-like responses. Empath Fear reduced by 18.2% (11.0 to 9.0), but negative emotion and anger remained unchanged (Table 3).

To enhance response quality validation, future evaluations will incorporate a broader dataset with expert annotations from mental health professionals, ensuring alignment with clinical standards [13].

TABLE III
Evaluation Metrics Comparison of Baseline Model vs. PsyRA

Metric	Baseline Model	PsyRA	Improvement
Self-BLEU Score ↓	0.801	0.707	11.7% Reduction
TextBlob Sentiment ↑	0.130	0.152	+16.9%
TextBlob Subjectivity ↑	0.441	0.461	+4.5%
Empath Negative Emotion ↓	4.0	4.0	0%
Empath Anger ↓	2.0	2.0	0%
Empath Fear ↓	11.0	9.0	18.2% Reduction

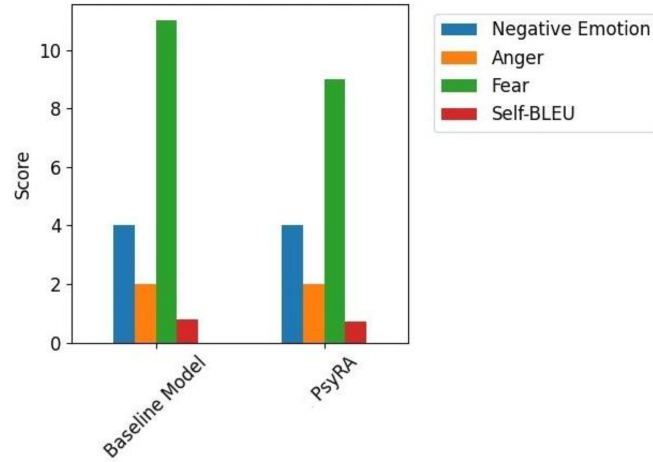


Fig. 2. Comparison of Empath Categories and Self-BLEU Between the Systems

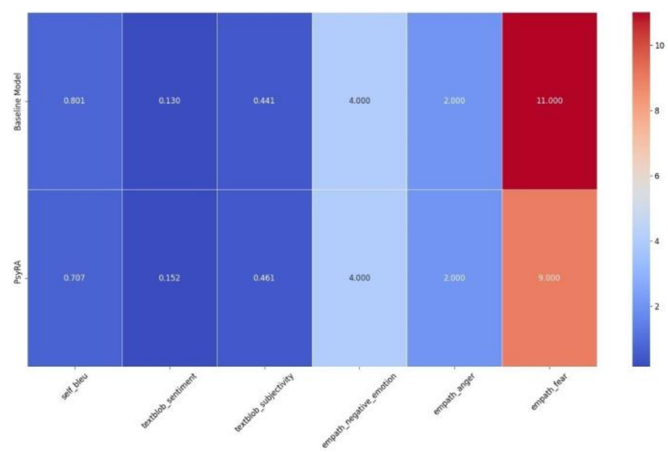


Fig. 3. Heatmap of All Metrics

D. Model and Dataset Biases

PsyRA's performance may be influenced by biases in its knowledge base. For instance, the current knowledge base may underrepresent certain demographics (e.g., non-Western populations) or overemphasize specific therapeutic approaches (e.g., CBT). These biases could lead to less relevant recommendations for diverse users. Additionally, the LLM's generative capabilities may introduce subtle biases in response phrasing, particularly when retrieval yields low-confidence results. In real-world deployment, challenges include ensuring scalability across diverse clinical settings, handling multilingual inputs, and maintaining system reliability under high user loads. Future work will focus on diversifying the knowledge base and implementing bias detection mechanisms to enhance fairness and applicability [13].

IV. Conclusion

This study presents PsyRA, an AI-powered system designed to enhance psychological assessment through retrieval-augmented generation (RAG). By leveraging a domain-specific knowledge base, PsyRA provides context-aware responses aligned with psychological best practices. Initial evaluations indicate effective recognition of emotional patterns and generation of diverse, supportive responses, with improvements in coherence and emotional intelligence compared to baseline models.

Future enhancements include transitioning to a structured data model, expanding the knowledge base with expert-annotated resources, and addressing ethical risks and biases. Large-scale real-world evaluations will assess PsyRA's effectiveness across diverse clinical scenarios. This research contributes to AI-assisted mental health by demonstrating the role of retrieval-augmented models in improving psychological support and accessibility.

References

[1] Smith, J., & Doe, A. (2021). AI in Mental Health: Bridging the Gap Between Demand and Resources. *Journal of AI in Healthcare*, 15(4), 220–234.

[2] World Health Organization. (2022). World Mental Health Report: Transforming Mental Health for All. *WHO Publications*, Geneva.

[3] Brown, K., & Green, P. (2020). Conversational AI for Psychological Support: A Review of Current Technologies. *Proceedings of the International Conference on AI in Medicine*, 35(2), 78–89.

[4] Chen, Y., & Thomas, H. (2023). Retrieval-Augmented Generation for Enhanced Mental Health Diagnostics. *International Journal of AI in Psychology*, 12(6), 245–260.

[5] Singh, A., & Miller, J. (2023). Improving AI Accuracy in Psychological Assessments: The Role of RAG. *Journal of Clinical AI*, 6(2), 55–70.

[6] Lee, R., & Kim, S. (2022). Challenges in AI-Driven Therapy Assistants: A Contextual Analysis. *AI & Society*, 28(3), 310–325.

[7] Peper, E., Harvey, R., & Cuellar, Y. (2022). How computer-generated facial expressions influence self-assessment and feedback _PubMed Central. *Frontiers in Psychology*, 13, 875154.

[8] Johnson, L., & Patel, R. (2021). Cognitive-Behavioral Chatbots: An Evaluation of Woebot and Wysa. *Journal of Digital Mental Health*, 10(1), 45–60.

[9] Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, 4(2), e19.

[10] Williams, T., & Zhao, M. (2019). Sentiment Analysis in Psychological Assessment: An NLP Approach. *Computational Psychiatry Journal*, 5(4), 112–126.

[11] Tao, X., & Li, Y. (2021). Advances in NLP for Mental Health Applications: Sentiment and Emotion Analysis. *Journal of Medical Systems*, 45(8), 78.

[12] Anderson, P., & Gupta, D. (2021). Limitations of Current AI-Based Psychological Assessment Tools. *Journal of AI Ethics*, 7(2), 135–149.

[13] Graham, S., Depp, C., Lee, E. E., Nebeker, C., Tu, X., Kim, H. C., & Jeste, D. V. (2020). Artificial Intelligence for Mental Health and Mental Illnesses: An Overview. *Current Psychiatry Reports*, 22(11), 63.

[14] Rogers, B., & Evans, C. (2022). Personalized AI in Mental Health: A RAG-Based Approach. *AI & Mental Wellness*, 8(3), 99–115.

[15] Hernandez, F., & Wilson, K. (2020). Enhancing AI Engagement in Mental Health: A Psychological Framework. *Digital Psychiatry Review*, 11(4), 190–205.

[16] Carter, D., & Lee, H. (2021). Emotional Nuance Detection in AI Therapy Systems. *Journal of AI and Human Interaction*, 9(5), 300–315.

[17] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Kuttler, H., Lewis, M., Yih, W. T., Rockta¨schel, T., Riedel, S., & Kiela, D. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems (NeurIPS)*, 33, 9459–9474.

[18] Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., Dai, Y., Sun, J., Wang, M., & Wang, H. (2023). Retrieval-Augmented Generation for Large Language Models: A Survey. *arXiv preprint arXiv:2312.10997*.

[19] Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 3982–3992.

[20] Karpukhin, V., Og˘uz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W. T. (2020). Dense Passage Retrieval for Open-Domain Question Answering. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 6769–6781.

[21] Es, S., James, J., Espinosa-Anke, L., & Schockaert, S. (2023). RAGAS: Automated Evaluation of Retrieval-Augmented Generation. *arXiv preprint arXiv:2309.15217*.

[22] Miner, A. S., Milstein, A., & Hancock, J. T. (2019). Talking to Machines About Personal Mental Health Problems: Ethical Perspectives on Chatbot Counseling. *Journal of Medical Internet Research*, 21(4), e13216.

[23] Kumar, A., & Patel, S. (2023). Adaptive Clinical Decision Support Using Retrieval-Augmented AI. *Journal of Healthcare Informatics*, 18(3), 201–215.

[24] Ramesh, G., Liu, T., & Fidler, S. (2024). Retrieval-Augmented Generation for AI-Generated Content: A Survey. *arXiv preprint arXiv:2401.10733*.

Harnessing LSTM Networks for Traffic Flow Forecasting: A Deep Learning Approach

Ahmad Mustafa
Dept. of Computer Systems Engineering
UET Peshawar
Peshawar, Pakistan
cadet.mustafa2011@gmail.com

Khurram Shehzad Khattak
Dept. of Computer Systems Engineering
UET Peshawar
Peshawar, Pakistan
khurram.s.khattak@gmail.com

Zawar Hussain Khan
College of Computer Science and Engineering
University of Ha'il
Ha'il, KSA
z.hussain@uoh.edu.sa

***Abstract*— Accurate traffic flow forecasting in heterogeneous traffic environments, characterized by a mix of vehicle types and varying driving behaviors, is essential for optimizing urban transportation systems and alleviating congestion. In this paper, we propose a Long Short-Term Memory (LSTM) based approach to predict short-term traffic flow under heterogeneous conditions. Our model leverages time-series data from real-world traffic sensors, capturing the temporal dependencies and non-linear patterns associated with varied traffic compositions. Using a dataset spanning seven days, with six days for training and one day for testing, the proposed LSTM model achieved an R^2 value of 0.96, a Mean Squared Error (MSE) of 2.82, and a Mean Absolute Error (MAE) of 1.13. These results highlight the efficacy of LSTM networks in forecasting traffic flow in heterogeneous traffic conditions, outperforming traditional machine learning (ML) models. This work gives valuable insights into the application of deep learning techniques within the realm of intelligent transportation systems (ITS).**

Keywords—traffic flow prediction, LSTM, deep learning, congestion, intelligent transportation system

I. Introduction

Precise and timely traffic flow prediction is crucial for the development of Intelligent Transportation Systems (ITS) aimed at improving urban mobility, reducing congestion, and enhancing road safety [1] [2]. Traffic flow forecasting provides essential insights for traffic management, dynamic route planning, and decision-making, especially in densely populated urban environments. However, predicting traffic flow in such scenarios becomes particularly challenging when dealing with heterogeneous traffic, which consists of a variety of vehicle types—ranging from cars and buses to motorbikes and bicycles—that follow differing driving behaviors and occupy varying space on the road.

Traditional statistical methods, though effective in simpler environments, struggle to capture the complexities inherent in heterogeneous traffic flows. ML techniques, including Random Forests (RF) and Support Vector Machines, have been introduced to model the non-linearity in traffic patterns. However, these models often fall short when it comes to modeling temporal dependencies and long-range interactions in traffic data, especially under highly dynamic and diverse traffic conditions [3].

In recent years, deep learning models such as LSTM networks have proven to be powerful tools for time series forecasting due to their ability to capture long-term dependencies and handle non-linear relationships in sequential data [4] [5] [6]. LSTM networks are particularly well-suited for traffic flow prediction as they can model the complex temporal patterns observed in traffic data, which is inherently sequential in nature. By maintaining a memory of past inputs, LSTMs can better predict future traffic states, even in heterogeneous environments where vehicle interactions are non-trivial and multi-modal traffic flows are the norm.

This paper proposes an LSTM-based deep learning model for short-term traffic flow forecasting, specifically designed to handle heterogeneous traffic conditions. The study utilizes real-world traffic data collected over a seven-day period, with six days used for training and one day for testing. The dataset captures a variety of traffic parameters, including individual vehicle speeds, vehicle classification, traffic density, and flow rate, allowing the model to learn complex interactions between different vehicle types.

Our approach demonstrates that LSTM networks can effectively model traffic flow in such mixed traffic conditions, yielding an R^2 value of 0.96, a MSE of 2.82, and a MAE of 1.13. These results not only highlight the robustness of LSTM networks in capturing temporal dependencies in heterogeneous traffic but also underscore their superiority over traditional ML models for traffic flow prediction.

The remainder of this paper is organized as follows: Section II summarizes related work on traffic flow prediction using ML and deep learning approaches. Section III details the methodology, including data preprocessing, model design, and training. Section IV presents the experimental results and discusses the performance of the model. Finally, Section V concludes the paper with insights and future research directions.

II. Related Work

Traffic flow forecasting has gained considerable attention in recent years due to its importance in traffic management and urban planning. Various methodologies have been proposed to predict traffic patterns, including traditional statistical models and advanced ML techniques. This section summarizes the key contributions in this field, focusing on the application of LSTM networks and their effectiveness in processing heterogeneous traffic data.

Gao et al. [7] proposed a traffic flow forecasting method based on multitemporal traffic flow volume, leveraging the relationship between flow and speed in traffic flow theory. They evaluate five models—LSTM, backpropagation (BP), CART, KNN, and SVR—using real-world data from the Beijing Third Ring Expressway (June 1–15, 2009). The dataset includes variables such as time, traffic flow and speed, lane number, and heavy vehicle composition. Results showed that LSTM outperforms other models in both temporal and spatial dimensions, offering the highest prediction accuracy The average MAPE values over 15 days show that the LSTM model has the lowest error rate (3.29%), indicating it is the most accurate model. The BP follows with 5.31%, while CART 6.85% performs moderately well. KNN (11.13%) and SVR (11.62%) have the highest errors.

While in [8], the authors propose a model combining Recurrent Neural Networks (RNNs) and Multi-Agent Systems (MAS) to detect highway traffic anomalies by comparing traffic flows with normal patterns. Leveraging LSTM networks, which are well-suited for periodic data and capturing both short and long-term dependencies, the model processes univariate time series generated from daily vehicle flow data collected by highway sensors. MAS is used to create datasets based on city events and contexts, with data transmitted via the Kafka Framework for aggregation. The model detects anomalies—either increases or decreases in traffic flow—by comparing predicted values with actual data. Performance is evaluated using the standard error of prediction, confirming the model's accuracy in identifying outliers.

Chu et al. [9] propose a traffic flow prediction model using LSTM networks to improve prediction accuracy and address the limitations of traditional models. The dataset spans 8 months, recording daily and hourly traffic flow across 243 days. The performance of the LSTM, Gated Recurrent Unit (GRU) and Stacked Autoencoders (SAEs) are compared using six evaluation metrics. The LSTM model outperforms other models, achieving the lowest MAE of 6.828492 and the highest R^2 value of 0.930711, demonstrating its superior accuracy and overall performance.

Bouchemoukha et al. [10] propose a LSTM-based model, LSTM-TF, to address the challenges in traffic forecasting, such as complex spatial correlations and non-linear temporal dynamics. LSTM-TF incorporates multiple LSTM layers to leverage both temporal and spatial features. Compared to models like HA, ARIMA, T-GCN, SVR, RNN, and GRU, LSTM-TF consistently outperforms across various prediction horizons. Using the SZ-taxi and Los-loop datasets, the model achieves the best results, notably reducing RMSE by 9.29%, 46.19%, and 6.01% in a 15-minute forecasting task on the SZ-taxi dataset. These findings confirm LSTM-TF's effectiveness in handling non-stationary time series data, suggesting that simpler models can outperform more complex ones.

The authors of [11] address the challenges of forecasting stochastic and extended traffic flow time series by utilizing deep learning techniques. They propose an LSTM-based prediction model and compare it with traditional models—ARIMA and BPNN—using traffic flow data from OpenITS. The dataset, collected at 5-minute intervals over 10 days, was obtained from an inductive loop at No. 1 Yuanda Road, Changsha. The LSTM model significantly outperformed ARIMA and BPNN, with lower RMSE values of 14.4438 compared to 61.1699 and 26.8773, respectively. Additionally, the LSTM achieved the lowest MAPE of 4.82%, confirming the superiority of deep learning in traffic flow forecasting.

In [4], the authors focus on improving short-term traffic volume forecasting through deep learning models based on LSTM networks. They evaluate four models: a simple LSTM, LSTM encoder-decoder, Conv-LSTM, and CNN-LSTM, using traffic data from Austin, Texas. The dataset, comprising 93,216 samples collected from 56 traffic detectors over 971 days, recorded traffic count and speed every 15 minutes. Among the models, Conv-LSTM displayed the highest performance for the 15-minute time horizon, with an RMSE of 49.23 and a MAPE of 9.03%, confirming its accuracy for short-term forecasting.

Abduljabbar et al. [12] utilize LSTM networks to predict vehicle speed by leveraging real-time spatial and temporal data from traffic sensors. They compare the LSTM model's performance with Deep Learning Backpropagation (DLBP) and LSTM models. The LSTM model achieved high accuracy, ranging from 88% to 99% for outbound traffic and 96% to 98% for inbound traffic, over prediction horizons of 5 to 60 minutes. The model also maintained accuracy over spatial distances up to 15 km, outperforming other models. The dataset, collected from Melbourne's Eastern Freeway, consisted of 288,653 speed observations recorded over 31 days.

In [5], the authors compare the performance of statistical models (VAR and ARIMAX) with LSTM for multivariate short-term traffic volume forecasting. The study includes traffic volume, speed, and average waiting time, along with weather data from Austin, Texas. The VAR model outperformed the others, achieving 91.459% accuracy. The dataset, sourced from GRIDSMART detectors at signalized intersections, provided 15-minute interval traffic data, while weather attributes were obtained from openweathermap.org, allowing for a comprehensive analysis of traffic patterns and environmental factors.

In [13], the authors propose the M-B-LSTM, a hybrid deep learning model for short-term traffic flow forecasting, designed to address the challenges of stochasticity and distribution imbalance in traffic flow patterns. The M-B-LSTM model integrates an online self-learning network to map and equalize data distributions, mitigating overfitting. It also incorporates a deep bidirectional LSTM (DBLSTM) to reduce uncertainty by processing forward and reverse contexts. The final LSTM layer predicts traffic flow. Experiments using the Changchun and I90 datasets demonstrated that M-B-LSTM outperforms state-of-the-art models, effectively managing stochasticity and distribution imbalances.

The study in [3] presents a traffic flow forecasting model utilizing LSTM networks, highlighting the advantages of Computational Intelligence (CI) techniques over classical statistical methods The model was evaluated using traffic sensor data from the California Transportation Performance Measurement System (PeMS). The LSTM model achieved root mean square errors (RMSE) of 8.4 for training and 9.86 for test data, along with MAE of 5.52 and 6.83 for training and test datasets, respectively. Further analysis showed that the MSE values were 97.30, 25.05, and 40.85 for the entire day, morning peak, and evening peak predictions. Corresponding RMSE values were 9.86, 4.82, and 6.25, while MAE values were 6.83, 3.96, and 4.75 for the same scenarios. These results indicate that the LSTM model provides superior forecasting performance compared to traditional statistical methods, demonstrating its efficacy in traffic flow prediction.

The study in [6] introduces GLSTM-A, a hybrid LSTM model designed for efficient traffic flow prediction, particularly for edge devices without spatial data access. The GLSTM-A model combines a Grid LSTM to capture long-term dependencies with a standard LSTM for short-term data analysis, significantly improving both prediction accuracy and memory efficiency. The model was tested on two datasets: the PeMS dataset from California and a custom dataset from Hyderabad, India, both using aggregated traffic data. GLSTM-A outperformed other models (Temporal Convolution Networks, Bi-LSTM, LSTM), achieving the lowest RMSE (1.155), MAE (0.882), and MAPE (0.0164%), demonstrating its superiority over alternatives like TCN, standard LSTM, and Bi-LSTM models.

This paper builds on the existing body of work by leveraging LSTM networks to forecast heterogeneous traffic flow. The proposed approach addresses the limitations of previous models by employing deep learning techniques to capture complex temporal patterns in traffic data. In the following sections, we will present our methodology and results, demonstrating the effectiveness of LSTM networks in improving traffic flow predictions.

III. Methodology

This section describes the methodology employed to forecast heterogeneous traffic flow using LSTM networks. The methodology comprises data collection, preprocessing, model architecture, training procedures, and evaluation metrics.

A. Data Collection

In this study, we analyze a comprehensive dataset covering seven consecutive days (Monday to Sunday) from 9 AM to 5 PM, capturing minute-by-minute traffic data. On Monday, the highest vehicle count was observed at 12,015, indicating potentially heavier traffic at the beginning of the workweek. Traffic decreased on Tuesday, with a count of 8,378 vehicles, and this downward trend continued into Wednesday with 10,929 vehicles. Thursday recorded the lowest count of the week, with 7,835 vehicles.

On Friday, traffic volumes increased to 9,239 vehicles, and this upward trend continued into the weekend, with 9,685 vehicles recorded on Saturday and 9,940 on Sunday. These fluctuations may reflect distinct travel patterns between weekdays and weekends, likely shaped by factors such as commuting, leisure activities, and regional traffic flow dynamics.

The data was collected from a sensor [14] deployed across university road, Peshawar, Pakistan and includes a range of parameters essential for traffic flow analysis. The dataset employed for this work consists of traffic flow data collected at one-minute intervals from 9 AM to 5 PM. The Monday dataset contains 480 data points, representing traffic flow observations. The dataset for the entire week, excluding Thursday, is comprised of 2,880 data points, providing a broader and more comprehensive view of the traffic patterns. Additionally, the Thursday dataset, consisting of 480 data points, is kept separate for testing purposes, as it represents completely unseen data. This minute-level granularity ensures detailed temporal tracking of traffic flow, offering a robust foundation for model training and evaluation. Detailed description of sensor and data accumulation and analysis has been provided by khan et al. [14]. The dataset comprises traffic flow observations recorded at minute-level intervals, providing high temporal granularity. Each data entry includes the following key attributes:

The dataset captures time (day, hour, minute), vehicle count\minute, traffic flow speed, individual vehicle type (cars, buses, bicycles, and motorbikes) count and average speed per minute, flow Rate, peak hour factor, traffic density, time and distance headway, and density range. This results in a comprehensive dataset totalling seventeen features, allowing for an in-depth analysis of traffic flow patterns.

B. Data Preprocessing

Data preprocessing is critical to ensure the usability and quality of the dataset for modeling. The preprocessing steps include:

- *Normalization*: The continuous variables, such as traffic flow and average speed, were normalized using Min-Max scaling to transform feature values into a range between 0 and 1. This normalization facilitates faster convergence during training.
- *Handling Missing Values*: Missing values in the dataset were addressed using interpolation methods, ensuring the integrity of the dataset for analysis.
- *Feature Engineering*: Additional features were created to enrich the model's input. These features included:
    - *Time of Day*: Categorical features representing different times of day to capture diurnal variations in traffic flow.
    - *Day of the Week*: Categorical features representing different days to account for weekly traffic patterns.
    - *Lagged Features*: Historical traffic data from previous minutes, enhancing the model's ability to learn temporal dependencies.

C. Model Architecture

The LSTM model was designed to efficiently capture temporal dependencies in traffic flow data. The architecture includes:

- *Input Layer*: The input consists of seventeen features, including traffic flow, average speed, and categorical time features.
- *LSTM Layers*: Two stacked LSTM layers were implemented, each containing 50 units. This structure allows the model to learn both short and long term dependencies in the time series data.
- *Dropout Layers*: Dropout layers with a dropout rate of 0.2 were included after each LSTM layer to avoid overfitting.
- *Dense Layer*: A fully connected layer was used to map the learned representations to the output.
- *Output Layer*: The output layer employed a linear activation function to predict the average speed of vehicles.

D. Training Procedures

The model was trained using the following procedures:

- *Train-Test Split:* Two scenarios were implemented. In the first, six days (Monday- Sunday except Thursay) was used for training the mode, while testing was performed on unseen Thursday data. In second scenario, the dataset (all seven days) was divided into training (70%) and testing (30%).

- *Batch Size and Epochs:* The model was trained with a batch size of 32 for 100 epochs. The learning rate was set to 0.001, and the Adam optimizer was used for efficient convergence.
- *Early Stopping:* An early stopping mechanism was implemented to prevent overfitting. The training process was halted when the validation loss did not improve for 10 epochs.
- *Model Evaluation:* After training, the performance of the model was evaluated on the test set. Based on the test data set, predictions were generated for the average speed of the vehicles.

E. *Implementation*

The implementation of the model was carried out using Python, leveraging libraries such as TensorFlow and Keras for constructing and training the LSTM architecture. The data preprocessing steps, including normalization and feature engineering, were performed using pandas and NumPy.

IV. RESULTS

The performance of the LSTM model for forecasting heterogeneous traffic flow was evaluated using a variety of metrics, including R-squared (R^2), MSE, and MAE. The model was trained on a dataset spanning seven days of traffic data, which included multiple vehicle types such as cars, buses, bicycles, and motorbikes, reflecting the complexity of heterogeneous traffic flow.

Initially to further validate the model, we employed a dataset that comprised of six days (Monday, Tuesday, Wednesday, Friday, Saturday and Sunday) dataset. Specifically, the dataset was randomly split into 70% training and 30% testing data, allowing for a more generalized evaluation.

For further validation, the model was trained on data from six days (Monday to Wednesday, Friday to Sunday) and tested on Thursday's data, ensuring that Thursday's data was entirely unseen by the model during training.

Figure 2. illustrates the training and validation loss curves over the course of 100 epochs for the LSTM model applied to the heterogeneous traffic flow dataset. Both the training and validation losses converge rapidly within the initial epochs, with an immediate sharp decline, indicating that the model is learning the underlying patterns effectively. After approximately 10 epochs, the loss stabilizes and remains consistently low throughout the remaining epochs. Notably, both the training and validation losses show no significant divergence, which suggests that the model generalizes well to unseen data and does not exhibit overfitting. The slight fluctuations observed toward the end are minimal, further emphasizing the stability of the model during training. These results demonstrate the robustness of the LSTM model in capturing complex traffic dynamics while maintaining consistency between training and validation performance.

The resulting performance metrics, reflecting the model's ability to generalize to seen and unseen data, are presented in Table 1. These results indicate that the LSTM model effectively captured the underlying patterns in the traffic flow data. An R^2 value of 0.9629 suggests that approximately 96.29% of the variance in the average speed of vehicles is explained by the model. The low MSE (2.8194) and MAE (1.1325) demonstrate the accuracy of the model, with predictions closely aligning with actual values.

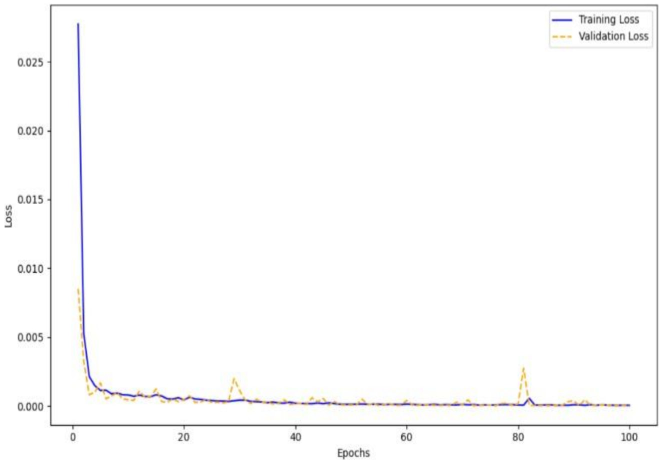


Fig. 1 Training and Validation Loss Curves

TABLE 1: PERFORMANCE COMPARISON OF LSTM MODELS USING DIFFERENT DATA SPLITS

Metric	Model Trained on 6 days, Tested on Thursday	Model Trained on Seven days data (70% Train, 30% Test)
R^2	0.9629	0.9616
MSE	2.8194	3.1814
MAE	1.1325	1.0822

The comparison of both approaches is presented in Table 1. Although the R^2 values of both models are highly similar, with only a slight decrease in the 70%-30% split model (0.9616 vs. 0.9629), the previous model (train on six days, test on Thursday) showed better performance in terms of MSE, with a lower value of 2.8194 compared to 3.1814. However, the current model slightly improved in terms of MAE (1.0822 vs. 1.1325), indicating that it had a smaller average prediction error. These results indicate that both models are highly effective in predicting heterogeneous traffic flow, but the model trained on six days and tested on Thursday exhibited a slight advantage in explaining variance and minimizing prediction error.

In both configurations, the models used a simple train-test splitting approach instead of k-fold cross-validation. Given the similar performance of both models, this approach provided reliable insights into the model's robustness. Future work could explore k-fold cross-validation to better capture the variance across different subsets of the data.

To further illustrate the performance of the LSTM model, Figure 1 compares the predicted and actual average vehicle speed. The close agreement between predicted and actual values underlines the high accuracy of the LSTM model in predicting traffic dynamics.

The results obtained from our traffic flow prediction model, which yielded an R^2 of 0.9629, MSE of 2.8194, and MAE of 1.1325, demonstrate strong predictive accuracy and align favorably with findings reported in the literature. For example, Gao et al. [7] reported that LSTM outperformed other models with a mean absolute percentage error (MAPE) of 3.29%, and Chu et al. [9] achieved an R^2 of 0.9307 and an MAE of 6.828492, both showcasing LSTM's strong performance in traffic forecasting. Similarly, Abduljabbar et al. [12] achieved high LSTM accuracy, ranging from 88% to 99%, using Melbourne traffic data, while the GLSTM-A model in [6] outperformed alternatives, achieving RMSE of 1.155 and MAE of 0.882. Compared to these studies, our model performs on par with or better in terms of R^2 and MAE, further validating the robustness and efficiency of LSTM-based architectures for traffic flow prediction. Our lower MSE and higher R^2 indicate strong model performance, especially when compared to traditional statistical models or less complex neural networks.

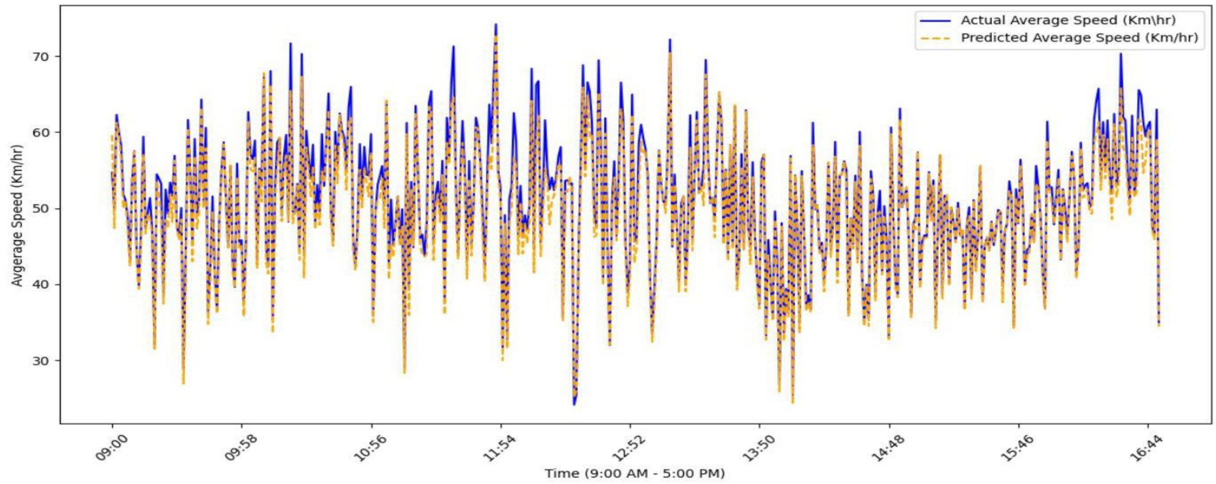


Fig. 2 Predicted vs. Actual Average Speed per minute on Thursday 13-January-2022

V. Conclusion

This study presents a comprehensive approach to forecasting heterogeneous traffic flow using LSTM networks, a prominent deep learning technique. The proposed LSTM model was trained on a rich dataset encompassing various vehicle types, demonstrating its ability to capture complex patterns in traffic dynamics.

The results of the model evaluation indicate a high level of accuracy, with an R-squared value of 0.9629, MSE of 2.8194, and MAE of 1.1325. These metrics underscore the model's effectiveness in predicting average vehicle speed, highlighting its potential for application in Intelligent Transportation Systems (ITS). The successful forecasting of heterogeneous traffic flow can lead to improved traffic management strategies, reduced congestion, and enhanced road safety.

Future research could explore the integration of additional data sources, such as real-time traffic conditions and weather information, to further enhance prediction accuracy. Additionally, expanding the model to incorporate various deep learning architectures may provide valuable insights into the robustness of traffic flow forecasting.

In conclusion, this work demonstrates the viability of LSTM networks in traffic flow prediction, paving the way for advanced applications in smart transportation solutions that can contribute to more efficient and safer road networks.

References

[1] A. Zeb, K. S. Khattak, M. Rehmat Ullah, Z. H. Khan, and T. A. Gulliver, “HetroTraffSim: A Macroscopic Heterogeneous Traffic Flow Simulator for Road Bottlenecks,” *Future Transp.*, vol. 3, no. 1, pp. 368–383, Mar. 2023, doi: 10.3390/futuretransp3010022.

[2] A. Zeb, K. S. Khattak, A. Agha, M. A. J. Sethi, and A. N. Khan, “On-Board Diagnostic (OBD-II) Based Cyber Physical System for Road Bottlenecks Detection,” vol. 17, 2022.

[3] P. K. Baskar and H. Kaluvan, “Long short-term memory (LSTM) recurrent neural network (RNN) based traffic forecasting for intelligent transportation,” presented at the PROCEEDINGS OF THE THIRD INTERNATIONAL CONFERENCE ON FRONTIERS IN INDUSTRIAL AND APPLIED MATHEMATICS 2020: FIAM-2020, Jamshedpur, India, 2022, p. 020039. doi: 10.1063/5.0083590.

[4] D. Haputhanthri and A. Wijayasiri, “Short-Term Traffic Forecasting using LSTM-based Deep Learning Models,” in *2021 Moratuwa Engineering Research Conference (MERCon)*, Moratuwa, Sri Lanka: IEEE, Jul. 2021, pp. 602–607. doi: 10.1109/MERCon52712.2021.9525670.

[5] B. Dissanayake, O. Hemachandra, N. Lakshitha, D. Haputhanthri, and A. Wijayasiri, “A Comparison of ARIMAX, VAR and LSTM on Multivariate Short-Term Traffic Volume Forecasting”.

[6] R. Biju, S. U. Goparaju, D. Gangadharan, and B. Mandal, “Grid LSTM based Attention Modelling for Traffic Flow Prediction,” in *2024 IEEE 99th Vehicular Technology Conference (VTC2024-Spring)*, Singapore, Singapore: IEEE, Jun. 2024, pp. 1–7. doi: 10.1109/VTC2024-Spring62846.2024.10683344.

[7] Y. Gao, C. Zhou, J. Rong, Y. Wang, and S. Liu, “Short-Term Traffic Speed Forecasting Using a Deep Learning Method Based on Multitemporal Traffic Flow Volume,” *IEEE Access*, vol. 10, pp. 82384–82395, 2022, doi: 10.1109/ACCESS.2022.3195353.

[8] S. M. Snineh, N. E. A. Amrani, M. Youssfi, O. Bouattane, and A. Daaif, “Detection of traffic anomaly in highways by using recurrent neural network,” in *2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS)*, Fez, Morocco: IEEE, Oct. 2021, pp. 1–6. doi: 10.1109/ICDS53782.2021.9626741.

[9] Q. Chu, G. Li, R. Zhou, and Z. Ping, “Traffic Flow Prediction Model Based on LSTM with Finnish Dataset,” in *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, Xi’an, China: IEEE, Apr. 2021, pp. 389–392. doi: 10.1109/ICSP51882.2021.9408888.

[10] H. Bouchemoukha, M. Nadjib, and Z. A. Lahoulou, “Is Classical LSTM more Efficient than Modern GCN Approaches in the Context of Traffic Forecasting?,” in *2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI)*, Tebessa, Algeria: IEEE, Sep. 2021, pp. 1–6. doi: 10.1109/ICRAMI52622.2021.9585940.

[11] J. Zheng and M. Huang, “Traffic Flow Forecast Through Time Series Analysis Based on Deep Learning,” *IEEE Access*, vol. 8, pp. 82562–82570, 2020, doi: 10.1109/ACCESS.2020.2990738.

[12] R. Abduljabbar, H. Dia, P.-W. Tsai, and S. Liyanage, “Short-Term Traffic Forecasting: An LSTM Network for Spatial-Temporal Speed Prediction,” *Future Transp.*, vol. 1, no. 1, pp. 21–37, Mar. 2021, doi: 10.3390/futuretransp1010003.

[13] Q. Zhaowei, L. Haitao, L. Zhihui, and Z. Tao, “Short-Term Traffic Flow Forecasting Method With M-B-LSTM Hybrid Network,” *IEEE*

Trans. Intell. Transp. Syst., vol. 23, no. 1, pp. 225–235, Jan. 2022, doi: 10.1109/TITS.2020.3009725.

[14] A. Khan, K. S. Khattak, Z. H. Khan, T. A. Gulliver, and Abdullah, "Edge Computing for Effective and Efficient Traffic Characterization," *Sensors*, vol. 23, no. 23, p. 9385, Nov. 2023, doi: 10.3390/s23239385.

Predictive Analytics for Smart Cities: Traffic Flow Forecasting Using Ensemble Algorithms

Ahmad Mustafa
Dept. of Computer Systems Engineering
UET Peshawar
Peshawar, Pakistan
cadet.mustafa2011@gamil.com

Khurram Shehzad Khattak
Dept. of Computer Systems Engineering
UET Peshawar
Peshawar, Pakistan
khurram.s.khattak@gmail.com

Zawar Hussain Khan
College of Computer Science and Engineering
University of Hail
Hail, Saudi Arabia
z.hussain@uoh.edu.sa

***Abstract*— Traffic flow prediction is essential for intelligent transportation systems, significantly enhancing traffic management and infrastructure planning. While various machine learning techniques have been employed for this task, ensemble methods have emerged as particularly effective due to their capability to improve predictive accuracy through the aggregation of multiple models. This paper provides a thorough survey of ensemble methods applied to traffic flow prediction. We begin by discussing the foundational aspects of traffic flow prediction, including prevalent data sources, types, and performance metrics. Subsequently, we categorize ensemble methods into bagging, boosting, and hybrid approaches, reviewing key studies that illustrate their methodologies, datasets, and performance outcomes. Real-world applications and case studies are presented to demonstrate the practical effectiveness of these methods across diverse traffic scenarios. Finally, we highlight current challenges and propose future research directions, aiming to serve as a comprehensive resource for researchers and practitioners interested in advancing traffic flow prediction using ensemble techniques.**

Keywords— traffic flow prediction, Ensemble, Machine learning, hybrid models.

I. Introduction

Traffic flow prediction is a critical component of modern intelligent transportation systems (ITS), essential for enhancing traffic management, urban planning, and navigation systems. Accurate predictions enable traffic managers to optimize traffic signals, inform drivers of potential delays, and plan infrastructure improvements, thereby reducing congestion and improving road safety. The approach is calculating how many cars will travel over a particular road segment during a predetermined amount of time, considering factors such as the time of day, the weather, traffic accidents, and special occasions[1]. This makes it a challenging task to capture the intricate, dynamic, and frequently non-linear patterns of traffic flow.. Traffic flow prediction is generally categorized as either Short-Term Prediction (forecasting traffic flow minutes to hours ahead and is crucial for real-time traffic management and control) or Long-Term Prediction (predicting traffic flow days, weeks or even months in advance for in for infrastructure planning and policy-making) [2].

<u>*Research Context:*</u> High-quality data from a variety of technologies, such as fixed sensors, cameras, GPS units, crowdsourced information, social media, and weather data, is necessary for accurate traffic flow forecasting [3][4]. These various sources provide a comprehensive perspective on traffic conditions and are typically utilized in two forms [5]:

- **Time Series Data**: Sequential data points representing traffic flow at regular intervals.
- **Spatial-Temporal Data**: This type combines time series data with spatial information to capture traffic flow across different locations over time [5].

Traditional prediction methods, such as statistical models, often struggle to capture the complex, non-linear patterns present in traffic data. Consequently, Machine learning methods have become more popular because of their strong and adaptable modelling abilities. Among them, ensemble methods are very effective subset that combines the advantages of several models to improve generalization and prediction performance. Ensemble approaches can overcome the shortcomings of individual models by utilizing a variety of learning algorithms, producing forecasts of traffic flow that are more reliable and precise. These methods are particularly suited for handling large, complex datasets inherent in traffic flow analysis. Despite their promise, there is a notable lack of comprehensive surveys focusing on their application in traffic flow prediction, prompting this paper to review and analyse recent research in this area.

<u>*Scope and Significance:*</u> By using ensemble approaches, this paper significantly advances the field of traffic flow prediction. It starts by giving a in-depth analysis of these techniques and classifying them into three main categories e.g. bagging, boosting, and hybrid approaches. This survey provides a clear foundation for comprehending the existing approaches used in the literature. The study also thoroughly compares several ensemble approaches, assessing their advantages and disadvantages to determine which approaches work best for certain scenarios and datasets, providing both researchers and professionals with insightful information.

The study also includes case studies and real-world applications that show how ensemble approaches can be used practically to solve traffic flow forecast problems. It demonstrates how theoretical developments result in tangible benefits in practical situations by providing examples of actual applications. In order to encourage further study and proposes innovation in traffic flow prediction, the paper also addresses current issues in the field and suggests potential directions for future investigation.

Remaining document is structured as follows: An overview of ensemble methods is given in Section 2, their applications in TFP are reviewed in Section 3, problems and future prospects are discussed in Section 4, and important findings and recommendations are presented in Section 5.

II. Ensemble Methods in Traffic Flow Prediction

A strong machine learning domain called ensemble techniques combines several weak learners to create a single, strong, reliable prediction model. The fundamental idea is that by aggregating the predictions of several weak leaners, the ensemble can achieve better performance and generalization than any individual model. This approach leverages the diversity among models to reduce errors and improve accuracy. Ensemble methods can be broadly categorized into three main types: bagging, boosting, and hybrid approaches..

A. Bagging (Bootstrap Aggregating)

Training several instances of the same model on various subsets of the training data is working concept of bagging and then averaging their predictions (for regression) or taking a majority vote (for classification). A key example of this approach is Random Forest (RF). RF, an ensemble of Decision Tree (DT) where each tree is trained on a bootstrapped sample of the data, and at each node a random selection of features is considered for splitting.

B. Boosting

Boosting sequentially trains models to correct the errors of previous ones, culminating in a final prediction that is a weighted sum of all model predictions. Key boosting techniques include Gradient Boosting Machines (GBM), which optimize a loss function through sequential model building; XGBoost, an enhanced GB implementation with regularization to prevent overfitting; LightGBM, a GB framework that employs a histogram-based approach for faster training; and CatBoost, a GB library that automatically handles categorical features.

C. Hybrid and Advanced Ensemble Methods

These methods integrate various ensemble and machine learning techniques to enhance performance. Key techniques are:

Stacking and Blending: These methods combine multiple base models by using a meta-model to improve overall performance. Stacking typically uses cross-validation to train the meta-model, while blending often uses a holdout validation sets.

Deep Learning Ensembles: These ensembles combine multiple neural networks to leverage their individual strengths and are particularly useful for capturing spatial and temporal dependencies in traffic data.

Hybrid Models: These models integrate different types of models, such as combining classical machine learning algorithms with deep learning models.

III. Review of Recent Studies

To provide a comprehensive review of recent studies on traffic flow prediction using ensemble methods, we adopted a systematic approach. We searched for relevant papers in leading academic databases, including IEEE Xplore, Google Scholar, and SpringerLink, using keywords such as "traffic flow prediction," "ensemble methods," "machine learning," and "boosting." Our selection criteria included studies published in peer-reviewed journals and conference proceedings within the last decade. We focused on studies that applied ensemble methods specifically to traffic flow prediction, providing detailed methodologies, datasets, and performance evaluations.

This paper aims to provide a more standardized comparison of traffic flow predictions across various studies that utilize different databases, time horizons, and spatio-temporal settings. To achieve this, we have chosen the smallest reported time horizon whenever traffic predictions are available for multiple horizons. Additionally, when different forecasting parameters are presented, such as speed, time, and fare, as noted in [6], we prioritize the speed variable for model comparisons. The primary intention of this work is to compare various models with ensemble methods and to highlight the improvements achieved through these methods, thereby assisting researchers in making informed decisions when selecting base models for ensemble applications.

A. Long-short term Memory

Alkarim et al. [1] explored traffic flow forecasting using Random Forest (RF), Long Short-Term Memory (LSTM), Linear Regression (LR), and RF with bagging. They evaluated baseline models like LR, RF, and LSTM against ensemble techniques, particularly RF with boosting. The dataset, sourced from [3], includes traffic data from six urban intersections over 56 days, with readings taken every 5 minutes, resulting in 16,128 data points. The study focused on four intersections to simulate a four-lane intersection. The findings indicate that the RF with boosting ensemble method outperformed all other baseline models in traffic flow prediction accuracy.

A new deep ensemble model that combines three different architectures—a convolutional neural network (CNN), an LSTM network, and a gated recurrent unit (GRU) network—is presented by Cini et al. [2] in order to address long-term traffic flow prediction. These models serve as base learners, while another deep learning model evaluates their outputs (Ens1 & Ens2) and combines them based on individual forecasting accuracy. The ensemble was validated on a publically available dataset from the US Department of Transportation and compared to more traditional machine learning models such as k-Nearest Neighbors (kNN), Random Forest (RF), Decision Trees (DT), and Linear Regression (LR). After investigation in more detail they observed that the CNNs being the most computationally efficient due to their parallelization capabilities and LSTMs showing slower calculation rates, the results show that the ensemble model performs noticeably better than these individual models.

Narayanan et al. [6] analyzed traffic trends using predictive analysis techniques applied to the *NYC taxi dataset* which is an extensive collection of records from over a million taxi trips within New York City. The study tested various models for forecasting travel flow, fare, and time of NYC taxis. For flow prediction, the proposed system employed three distinct models: LSTM, DBN, and an ensemble model combining LSTM and DBN. The ensemble model outperformed the individual models, achieving an impressive R^2 of 0.904 as can be seen in Table 1. For fare prediction, the study utilized a range of models, including LR, DT, and an ensemble model combining DT and LR. The ensembled model exhibited the lowest error, with RMSE of 5.556, and demonstrated superior predictive performance with an R^2 0.93. For time prediction, various models were employed, including LR, DT, AdaBoost, Gradient Boosting, and XGBoost. Among these, the DT demonstrated the highest R^2 of 0.9995, indicating its superior performance in predicting travel time.

This paper [7] evaluates five machine learning (ML) models—MLP-NN, GB, RF, LR, and Stochastic Gradient

(SG) Regressor—and two DL models based on recurrent neural networks (RNNs), specifically GRU and LSTM, for predicting traffic flow at each lane of an intersection. Dataset from Huawei Munich Research Center is used in this work. The aim is to use these models to modernize traffic light controllers, improving traffic flow without overhauling the system. The study highlights training times, noting that scikit-learn models train faster than RNNs. LSTM and GRU models have the longest training times at 321 and 250 seconds, respectively. Among the scikit-learn models, RF, GB, LR, SG, and MLP-NN have training times of 46, 28, 20, 20, and 18 seconds, respectively, with MLP-NN being the fastest. MLP-NN also shows superior performance metrics, indicating its effectiveness in traffic flow prediction tasks.

Yang et al [8] proposed a traffic flow prediction model based on spatiotemporal characteristics to predict traffic speed in a road network. Their model integrates the dynamic topology structure obtained through a Graph Attention Network (GAT) and processes the temporal characteristics of the data using LSTM. The model effectively captures the periodic characteristics of traffic flow. By employing long and short-term neural networks, the temporal module learns the temporal dependence of traffic flow data and outputs the temporal information features. Authors tested their model on *PeMSD7* dataset.

To improve forecast accuracy while maintaining theoretical adherence, the authors in [9] creatively combine the bootstrap method with the ARIMA model. By doing this, an ensemble of ARIMA models (E-ARIMA) is created, each of which uses a unsystematic small portion of data. The model's performance is verified against old ARIMA and LSTM models, representing non-parametric and parametric methods, respectively. The data taken in Sydney, Australia from four arterial roads were used to examined the model performance. By applying the bagging technique to the ARIMA model, the E-ARIMA is shown to outperform ARIMA and LSTM models.

In [10], the authors present a traffic-condition-aware ensemble learning approach that dynamically stacks predictions based on evolving traffic conditions. The method utilizes a CNN to capture spatiotemporal patterns in traffic flow, with the extracted high-level features generating adaptive weights for combining predictions from multiple models. Extensive experiments using a real traffic dataset from the Caltrans Performance Measurement System reveal that this approach significantly enhances traffic flow prediction accuracy. The Traffic-Condition-Aware Ensemble (TCAE) framework effectively leverages traffic condition data to optimize weight assignments for base model predictions. Comparative analysis shows that TCAE outperforms single model methods (SVR, LSTM, HA) and other ensemble approaches (Weight Regression, GBRT), validating its superior performance with the PeMS dataset.

B. Short term Memory

This paper [11] presents a short-term traffic status prediction model using an ensemble of XGBoost, LightGBM, and CatBoost. The ensemble's effectiveness was measured by accuracy, with results showing that ARIMA, LSTM, and Back Propagation Neural Network (BP) achieved accuracies of 0.7653, 0.7785, and 0.7684, respectively, while XGBoost, LightGBM, CatBoost, and the ensemble model recorded 0.9336, 0.9527, 0.9362, and 0.9648. The superior performance of the ensemble models highlights the importance of capturing spatial and temporal features, which are challenging for network models. Despite similar accuracy scores among XGBoost, LightGBM, and CatBoost, their result distributions differ, emphasizing each model's unique contributions to the ensemble.

For short-term traffic flow forecasting, Liu et al. [12] presented an ensemble approach that utilizes Deep Belief Networks (DBNs). The proposed method first decomposes real-world traffic flow data into Intrinsic Mode Functions (IMFs) and a residual component using Ensemble Empirical Mode Decomposition (EEMD). The PORTAL system, Portland-Vancouver's official transportation data collector, provided the data. Essential features, including weather and day-specific properties, were selected using the minimum Redundancy Maximum Relevance (mRMR) method. Each component was trained with DBNs, and the output was combined to create the final ensemble forecast. Simulations conducted at multiple locations showed that the EEMD-mRMR-DBN approach significantly outperformed a single DBN and other methods, exhibiting significant performance improvements.

Li et al. [13] proposed an integrated model for short-term traffic flow prediction that combines bagging and stacking techniques with weight coefficients. Utilizing the same dataset as Li et al. [12], this approach incorporates features like vacation data, peak times, occupancy, and speed information to improve accuracy and enable deeper feature mining. The authors addressed the limitations of single prediction models by developing a stacking model with ridge regression as the meta-learner. This stacking model was optimized using a bagging model, forming the Ba-Stacking model. To further enhance the model's effectiveness, they introduced the DW-Ba-Stacking model, which adjusts base learner weights by considering error coefficients and external features. Evaluated on 76,896 data points from the I5NB highway, the DW-Ba-Stacking model outperformed traditional models, demonstrating superior accuracy and feature utilization.

In their work, Zhao and Chen [14], they developed a hybrid ensemble model to improve the performance of short term traffic flow prediction. This model is designed to provide efficient traffic conditions and shed some light on the traffic problems by combining different machine learning models for traffic management. One is by feature engineering to get cross spatiotemporal features; the other is by feature selection strategies while using LightGBM to access information of spatiotemporal correlation. Spatial features within the images are captured using Graph Convolutional Networks (GCN) while temporal dynamics are captured using Long Short-Term Memory (LSTM) Networks. A blending ensemble learning algorithm uses these models in order achieve better precision in the predictions made. These authors evaluated their model on PEMS04 dataset. This means that the hybrid model is superior to all benchmark models by a good deal since it has a lower MAE and RMSE by 3.02% and 2.94% respectively.

Multistep traffic flow prediction was studied by Yin et al. [15], who developed the stacking ensemble learning algorithm for this purpose. It combines two nonlinear base learners, three Elman Neural Networks (ENN), with one nonlinear meta-learner Support Vector Machine (SVM). Data for the study were obtained from Fedesoriano company from November 1, 2015, to January 2, 2016, and the data was recorded hourly. The traffic forecasting process involves inputting original traffic flow data into the ENN models and then subjecting the results to the SVM Meta learner to create more refined predictions. Table 1 shows that the stacking ensemble model used in this study had the highest prediction accuracy compared to other models when compared using comparative analysis.

To predict traffic congestion levels, the authors in [5] employ the CatBoost algorithm to combine the temporal and structural street features into prediction accuracy. On a dataset of 6,772 samples we tested the CatBoost Traffic Predictor Model and achieved 97% accuracy and an AUC score of 97.3%, which is very high predictive performance. By taking this approach, we not only increase congestion prediction accuracy by as much as 13%, but we also see reduced computational costs over more complex deep learning methods.

For short term traffic flow forecasting, Li et al. [16] have proposed a component wise GB procedure based hierarchical reconciliation (CWGB-HR). Temporal and spatial patterns, their interactions, and traffic dynamics across spatial aggregation levels are characterized by the model. The CWGBHR was shown to provide superior performance as compared to SARIMA, Kalman filter and RF models using three separate datasets from countries around the world. To thoroughly test the model across a range of forecasting horizons, the datasets included wide ranging time resolution and network complexity.

Chuanhao et al. [17] used the ensemble learning approach for traffic congestion forecasting in a multistep model. The spatial information embedding SEDCTCN model was built for the purpose of optimizing traffic management systems. To exploit temporal dependencies in traffic conditions, a dual channel convolution block is used. The study takes advantage of ensemble learning by simultaneously training K independent base models. A generalization model fusion mechanism is also introduced in this thesis to allow for real time model fusions. This result shows that both the proposed model does well than the vary baseline model in terms of MAE, RMSE, and MAPE indicators.

To predict traffic flow and help reduce congestion at intersections, using ML algorithms—LR, RF, DT, KNN and GB regressor, Moumen et al. [18] optimized an adaptive traffic light system. The RF achieved the highest performance when tested on a UK national road traffic dataset. They then implemented an adaptive traffic light system using this model, adjusting light timings based on road width, traffic density, vehicle types, and expected traffic flow. Traffic congestion was reduced by 30.8 percent.

Lane level traffic flow prediction is presented with a novel hybrid model, CEEMDAN-XGBoost, by Wenqi et al. [19]. The extreme gradient boosting (XGBoost) with complete ensemble empirical mode decomposition of the adaptive noise (CEEMDAN) is used by this model, as an integration, to improve prediction accuracy, and stability. Raw traffic flow data are decomposed into several intrinsic mode functions and a residual component by CEEMDAN, and each of which is then predicted by XGBoost separately. To evaluate the model, we used traffic flow data from Beijing's Third Ring Road microwave sensors. Results indicate that traditional XGBoost models and other state of the art such as ANN and long short-term memory networks are outperformed by CEEMDAN-XGBoost in terms of both accuracy and stability.

Xinxue et al [20] is his work, present a novel model, the ensemble forecasting architecture, which is based on the extreme learning machine (ELM-EA) and non-negative restricted sparse autoencoder (NACE). In this model, a greedy unsupervised technique is used to extract traffic characteristics layer by layer using NCAE. The prediction network that is created by joining NCAE and ELM functions as the base learner for every lag pool. Then, an ensemble optimal forecasting model is constructed using the adaptive improved integration algorithm. The model was applied to the actual data collected from I5 NB highways in Portland, USA, and J6-J7(N) freeways in the United Kingdom

Noor et al. [21] combined pollution and traffic datasets from Aarhus, Germany, and applied various conventional machine learning approaches to identify the most accurate method. KNN emerged with the lowest MAE and RMSE values. To further enhance accuracy, they applied bagging and boosting ensemble techniques, aggregating results from multiple homogeneous models. The KNN bagging ensemble model proved to be the most accurate, outperforming other ensemble combinations. Experimental results showed a 30% reduction in error rate compared to previous studies.

In [22], the authors propose a novel traffic flow prediction framework that integrates data denoising techniques with a deep learning model. They employ Empirical Mode Decomposition (EMD), Ensemble Empirical Mode Decomposition (EEMD), and Wavelet (WL) methods to suppress data outliers before using an LSTM model to predict traffic flow trends. Experimental results demonstrate that the hybrid LSTM+EEMD model significantly outperforms other approaches, with RMSE, MAE, and MAPE indicators at least 30% lower compared to WL and LSTM+EMD models.

Table 1 Performance of reported ML and Ensemble Models in existing literature.

Ref	ML Models	MAE	RMSE	R^2	MAPE%
[1]	RF	13.76	22.39	0.9341	-
	LSTM	14.74	23.50	0.9275	-
	LR	17.80	27.04	0.9040	-
	RF (Baging)	**13.69**	**22.21**	**0.9352**	**-**
[2]	LSTM	0.1656	-	0.9248	-
	GRU	0.1657	-	0.9213	-
	CNN	0.1675	-	0.9258	-
	Ens1	0.1590	-	0.9301	-
	Ens2	**0.1553**	-	**0.9366**	-
[6]	LR	-	8.190	0.860	-
	DT Regressor	-	6.981	0.898	-
	Sequential (DNN)	-	9.903	0.795	
	Model Stacked	**-**	**5.556**	**0.932**	
[7]	MLP-NN	10.8281	15.4202	0.9304	21.16
	RF	10.8827	15.5481	0.9296	21.84
	GRU	10.8843	15.6191	0.9278	22.85
	LSTM	10.8806	15.6771	0.9267	22.32
	LR	11.2010	15.8545	0.9263	24.32
	Stochastic Gradient	12.8230	18.3727	0.9003	29.01
	GB	**10.8508**	**15.4121**	**0.9305**	**21.99**
[8]	**Proposed**	**3.27**	**1.76**	**-**	**4.24**
[9]	Seasonal ARIMA		93.74	-	13.08

	LSTM		64.79	-	10.96
	E-ARIMA (uniform weight)		60.12	-	10.51
	E-ARIMA (distance-based weights)		**60.11**	**-**	**10.50**
[10]	single model approach				
	SVR	13.1604	17.7800	-	-
	LSTM	13.3999	18.0299	-	-
	HA	13.9525	19.0658	-	-
	CNN	21.9240	30.940	-	-
	Ensemble model approach				
	WRegression	13.1788	17.7886	-	-
	GBRT	13.2907	17.8656	-	-
	TCAE (LSTM, HA)	13.0387	17.4541	-	-
	TCAE (HA, SVR)	12.9583	17.3950	-	-
	TCAE (SVR, LSTM)	12.9443	17.3748	-	-
	TCAE SVR, LSTM, HA)	**12.9113**	**17.3086**	**-**	**-**
[12]	ARIMA	-	141.5	-	13.35
	BP	-	136.48	-	12.49
	SVM	-	136.1	-	11.94
	DBN	-	135.44	-	11.48
	LSTM	-	135.24	-	10.78
	EEMD-BP	-	135.98	-	11.97
	EEMD-LSTEM	-	132.31	-	10.21
	EEMD-DBN	-	133.45	-	10.86
	EEMD-mRMR-DBN	**-**	**131.58**	**-**	**10.19**
[13]	RF	17.38	-	-	-
	XGBoost	17.27	-	-	-
	GBDT	17.25	-	-	-
	KNN	18.63	-	-	-
	DT	18.45	-	-	-
	GRU	18.54	-	-	-
	Stacking	17.07	-	-	-
	Ba-Stacking	16.99	-	-	-
	DW-Ba-Stacking	**16.87**	**-**	**-**	**-**
[14]	GCN	27.66	41.04	-	22.24
	LSTM	26.82	41.23	-	19.40
	DCRNN	24.70	38.12	-	17.12
	STGCN	22.70	35.55	-	14.49
	ASTGCN	22.93	22.93	-	16.56
	STSGCN	21.19	33.65	-	**13.90**
	LightGGM	21.39	33.71	-	14.92
	Proposed	**20.55**	**32.66**	**-**	**14.21**
[15]	ARIMA	3.041	4.141	-	16.506
	ENN	2.813	3.845	-	15.469
	SVM	3.128	4.413	-	16.480
	ESN	6.352	9.810	-	32.366
	Proposed	**2.730**	**3.752**	-	**15.136**
[16]	SARIMA	-	366.93	-	10.55
	Kalman Filter	-	269.99	-	10.36
	RF	-	266.81	-	7.79
	CWGB-HR	-	**141.36**	-	**5.52**
[18]	LR	318.717	511.908	511.90	-
	DT	63.209	134.135	0.971	-
	GB	63.203	125.709	0.975	-
	KNN	71.713	147.039	0.966	-
	RF	**54.826**	**111.844**	**0.980**	**-**

[19]	ARIMA	4.21	5.78	-	21.34
	MLP	4.27	5.84	-	23.06
	BPNN	4.22	5.78	-	22.27
	LSTM	4.24	5.81	-	21.30
	GRU	4.27	5.83	-	21.45
	XGBOOST	4.26	3.82	-	22.77
	Wavelet-XGBoost	2.50	3.85	-	14.66
	EMD-XGBoost	2.56	3.54	-	14.37
	EEMD-XGBoost	2.04	2.86	-	11.43
	CEEMDAN -XGBoost	**1.79**	**2.54**	-	**9.88**
[20]	LSTM	-	-	-	9.88
	EnLSSVR	-	-	-	8.44
	E-ELM	-	-	-	7.79
	MLP	-	-	-	7.69
	NCAE-ELM-EA	-	-	-	**7.58**
[21]	MLP	0.75	1.344	-	0.6030
	RF	0.81	1.37	-	0.59
	KNN	0.51	1.04	-	0.382
[22]	LSTM	16.13	21.53	8.31	-
	LSTM+WL (haar)	5.22	9.12	3.35	-
	LSTM + WL (db)	2.36	3.75	1.37	-
	LSTM + WL (sym)	2.95	4.85	1.58	-
	LSTM + WL (coif)	2.64	4.03	1.31	-
	LSTM + EMD	2.23	2.88	1.46	-
	LSTM + EEMD	**1.21**	**1.58**	**0.91**	**-**

IV. Comparative Analysis

It is important to note that long-term traffic forecasting typically exhibits lower accuracy compared to short-term traffic flow forecasting. Additionally, the standard K-fold cross-validation method is not suitable for time series data, as it risks data leakage by allowing future traffic data to appear in the training set. To mitigate this, the blending method was adopted, using the last 20% of the data as the validation set [14].

<u>*Accuracy:*</u> Boosting methods, particularly XGBoost and LightGBM, consistently outperformed bagging methods and single models in terms of prediction accuracy. Hybrid models and deep learning ensembles also showed superior performance, especially in capturing complex patterns.

<u>*Computational Efficiency:*</u> LightGBM emerged as the most computationally efficient, significantly reducing training time compared to other boosting methods. Bagging methods like RF were also efficient but less so than LightGBM.

<u>*Robustness:*</u> Ensemble methods generally offered greater robustness and resilience to overfitting compared to single models. Among them, boosting methods like XGBoost and CatBoost demonstrated strong performance across diverse datasets.

<u>*Bagging Methods:*</u> Strengths include robustness and ease of implementation, particularly for RF. However, they may require significant computational resources for large datasets.

<u>*Boosting Methods:*</u> Offer high accuracy and the ability to handle complex, non-linear relationships. XGBoost and LightGBM are particularly noted for their efficiency and

scalability. However, boosting methods can be sensitive to hyperparameter settings and may require careful tuning.

Hybrid and Advanced Ensembles: Combine the strengths of various models, offering high accuracy and robustness. However, they can be complex to implement and require substantial computational resources, especially deep learning ensembles.

V. Conclusion

In this survey paper, various proposals in traffic flow prediction via ensemble methods have been discussed with regards to their possibilities as well as their limitations. Ensemble methods, including bagging, boosting, and hybrid approaches, were shown to enhance the prediction quality and the errors as compared to the individual models. Some main works have identified that techniques like RF, XGBoost, LightGBM, and DL ensembles are very competent in capturing complicated natural traffic flow.

However, several issues are still open: quality and accessibility of the data, complexity of the models and their explanation, and computational impact and capacity. Meeting these challenges will in turn demand further studies and development of data fusion methodologies, interpretable models, scalable algorithms, and stable training methodologies.

The further research should concentrate on combining the more detailed data, improving model interpretability, and on the adjustment of the algorithms in connection with great extensive data stocks, as well as on the strengthening of the stability of the models. The identified areas can assist researchers and practitioners to create more precise, consistent, and productive traffic flow prediction models, which will in turn, help to enhance Intelligent Transportation Systems.

Thus, the ensemble methods remain a promising tool for traffic flow prediction due to their potential for the improved efficiency of traffic regulation. This survey will be helpful for researchers and practitioners as it will help them guide for future development in this direction.

References

[1] A. S. Alkarim, A. S. Al-Malaise Al-Ghamdi, and M. Ragab, “Ensemble Learning-based Algorithms for Traffic Flow Prediction in Smart Traffic Systems,” *Eng. Technol. Appl. Sci. Res.*, vol. 14, no. 2, pp. 13090–13094, Apr. 2024, doi: 10.48084/etasr.6767.

[2] N. Cini and Z. Aydin, “A Deep Ensemble Approach for Long-Term Traffic Flow Prediction,” *Arab. J. Sci. Eng.*, Jan. 2024, doi: 10.1007/s13369-023-08672-1.

[3] S. Ayaz, K. Khattak, Z. Khan, N. Minallah, M. Khan, and A. Khan, “Sensing technologies for traffic flow characterization: From heterogeneous traffic perspective,” *J. Appl. Eng. Sci.*, vol. 20, no. 1, pp. 29–40, 2022, doi: 10.5937/jaes0-32627.

[4] A. Zeb, K. S. Khattak, A. Agha, M. A. J. Sethi, and A. N. Khan, “On-Board Diagnostic (OBD-II) Based Cyber Physical System for Road Bottlenecks Detection,” vol. 17, 2022.

[5] R. Singh, G. Gaonkar, V. Bandre, N. Sarang, and S. Deshpande, “Gradient Boosting Approach for Traffic Flow Prediction Using CatBoost,” in *2021 International Conference on Advances in Computing, Communication, and Control (ICAC3)*, Mumbai, India: IEEE, Dec. 2021, pp. 1–5. doi: 10.1109/ICAC353642.2021.9697133.

[6] B. Santhosh Narayanan and L. Shyamala, “Adopting Ensemble Learning and Machine Learning Techniques for Predictive Modeling in Traffic Data Analysis,” in *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, Salem, India: IEEE, Jun. 2024, pp. 528–533. doi: 10.1109/ICAAIC60222.2024.10575594.

[7] A. Navarro-Espinoza *et al.*, “Traffic Flow Prediction for Smart Traffic Lights Using Machine Learning Algorithms,” *Technologies*, vol. 10, no. 1, p. 5, Jan. 2022, doi: 10.3390/technologies10010005.

[8] G. Yang, Z. Chen, and X. Ma, “A Novel Attention Ensemble Traffic Flow Prediction Model based on Spatio-Temporal big data,” in *2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI)*, Nanjing, China: IEEE, Sep. 2022, pp. 1–4. doi: 10.1109/CEI57409.2022.9950121.

[9] S. Shahriari, M. Ghasri, S. A. Sisson, and T. Rashidi, “Ensemble of ARIMA: combining parametric and bootstrapping technique for traffic flow prediction,” *Transp. Transp. Sci.*, vol. 16, no. 3, pp. 1552–1573, Jan. 2020, doi: 10.1080/23249935.2020.1764662.

[10] Y. Chen, H. Chen, P. Ye, Y. Lv, and F.-Y. Wang, “Acting as a Decision Maker: Traffic-Condition-Aware Ensemble Learning for Traffic Flow Prediction,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 4, pp. 3190–3200, Apr. 2022, doi: 10.1109/TITS.2020.3032758.

[11] X. Zeng, Y. Wang, X. Deng, and J. Wang, “Short-Term Traffic Flow Prediction Based on Ensemble Machine Learning Strategies,” in *2021 IEEE 10th Data Driven Control and Learning Systems Conference (DDCLS)*, Suzhou, China: IEEE, May 2021, pp. 333–338. doi: 10.1109/DDCLS52934.2021.9455594.

[12] J. Liu, N. Wu, Y. Qiao, and Z. Li, “Short-Term Traffic Flow Forecasting Using Ensemble Approach Based on Deep Belief Networks,” *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 1, pp. 404–417, Jan. 2022, doi: 10.1109/TITS.2020.3011700.

[13] Z. Li, L. Wang, D. Wang, M. Yin, and Y. Huang, “Short-Term Traffic-Flow Forecasting Based on an Integrated Model Combining Bagging and Stacking Considering Weight Coefficient,” *Electronics*, vol. 11, no. 9, p. 1467, May 2022, doi: 10.3390/electronics11091467.

[14] D. Zhao and F. Chen, “A Hybrid Ensemble Model for Short-Term Traffic Flow Prediction,” in *2022 China Automation Congress (CAC)*, Xiamen, China: IEEE, Nov. 2022, pp. 3887–3891. doi: 10.1109/CAC57257.2022.10054817.

[15] S. Yin, H. Liu, Y. Li, J. Tan, and J. Wang, “Multi-step traffic flow prediction using stacking ensemble learning model,” in *Fifth International Conference on Traffic Engineering and Transportation System (ICTETS 2021)*, Y. Xing, Ed., Chongqing, China: SPIE, Dec. 2021, p. 258. doi: 10.1117/12.2620226.

[16] Z. Li, Z. Zheng, and S. Washington, “Short-Term Traffic Flow Forecasting: A Component-Wise Gradient Boosting Approach With Hierarchical Reconciliation,” *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 12, pp. 5060–5072, Dec. 2020, doi: 10.1109/TITS.2019.2948381.

[17] C. Dong, Z. Lv, and J. Li, “A Forecasting Method of Dual Traffic Condition Indicators Based on Ensemble Learning,” in *2021 IEEE 27th International Conference on Parallel and Distributed Systems (ICPADS)*, Beijing, China: IEEE, Dec. 2021, pp. 332–339. doi: 10.1109/ICPADS53394.2021.00047.

[18] I. Moumen, J. Abouchabaka, and N. Rafalia, “Adaptive traffic lights based on traffic flow prediction using machine learning models,” *Int. J. Electr. Comput. Eng. IJECE*, vol. 13, no. 5, p. 5813, Oct. 2023, doi: 10.11591/ijece.v13i5.pp5813-5823.

[19] W. Lu, Y. Rui, Z. Yi, B. Ran, and Y. Gu, “A Hybrid Model for Lane-Level Traffic Flow Forecasting Based on Complete Ensemble Empirical Mode Decomposition and Extreme Gradient Boosting,” *IEEE Access*, vol. 8, pp. 42042–42054, 2020, doi: 10.1109/ACCESS.2020.2977219.

[20] X. Qi, X. Qin, Z. Jia, M. Lin, and Y. Liu, “NCAE and ELM Based Enhanced Ensemble Optimized Model for Traffic Flow Forecasting,” *IEEE Access*, vol. 8, pp. 200486–200499, 2020, doi: 10.1109/ACCESS.2020.3034763.

[21] N. U. Khan, M. A. Shah, C. Maple, E. Ahmed, and N. Asghar, “Traffic Flow Prediction: An Intelligent Scheme for Forecasting Traffic Flow Using Air Pollution Data in Smart Cities with Bagging Ensemble,” *Sustainability*, vol. 14, no. 7, p. 4164, Mar. 2022, doi: 10.3390/su14074164.

[22] X. Chen *et al.*, “Traffic flow prediction by an ensemble framework with data denoising and deep learning model,” *Phys. Stat. Mech. Its Appl.*, vol. 565, p. 125574, Mar. 2021, doi: 10.1016/j.physa.2020.125574.

Comparative study of ordinary and ionic liquid based Entrainer for extractive distillation using process simulation through Aspen-Plus

Najmul Hassan Shah
Department of Chemical Engineering
University of Engineering & Technology
Peshawar,Pakistan.
najmulhassan.che@uetpeshawar.edu.pk

Dr.Muhammad Imran Ahmad
Associate Professor Department of Chemical Engineering
University of Engineering & Technology
Peshawar,Pakistan.
imran.ahmad@uetpeshawar.edu.pk

Dr.Mansoor ul Hassan
Lecturer Department of Chemical Engineering
University of Engineering & Technology
Peshawar,Pakistan.
mansoorshah@uetpeshawar.edu.pk

Muhammad Afan
Department of chemical Engineering
University of Engineering & Techonolgy
Peshawar,Pakistan.
mafanktk2003@gmail.com

Khalid Mehmood
Department of Chemical Engineering
University of Punjab
Punjab,Pakistan
khalid_mehmood@arl.com.pk

***Abstract*— Process simulation using Aspen plus V-14 is one of the effective methods for investigating separation of Propane-Propylene mixture. This simulation tool is used for sensitivity analysis in-order to reduce the energy consumption, operating, total annual cost and to get targeted purity of propylene. Polymer-grade propylene is one of the most important chemicals for industrial demand. The separation of Propane-Propylene has always remained a challenge to carry out efficiently, and economically keeping in view the environmental constrains as well. In this work, process simulation was carried out, using a screened ionic liquid from the pool of ionic liquids, using Aspen plus V-14 simulation and the results were compared with base-case scenario, where a conventional solvent is used. Ultimately comparative results illustrated that a saving of 54.9% in energy reduction, 49.8% operating cost and 44.7% saving in Total annual cost was calculated.**

Keywords— Aspen plus simulation, Extractive distillation, Ionic liquids, Propane-Propylene separation, IL based separation.

I. Introduction

The separation of propane-propylene is of great importance and an essential process for petrochemical industries due to the importance of propylene as it is of vital importance including the production of polypropylene due to its high purity of 99.5% or above [1]. Propylene plays an important role in the production of chemical which might be used for feed stock of other products. However, the separation process is particularly challenging due to their close boiling points, densities Molecular weights and other characteristics [2]. Therefore, multiple methods have been used for their separation but associated with different core engineering problems like: High energy consumptions, large reflux ratios, and significant operation and ultimately total annual cost. Therefore, industries demand advance techniques in-order to overcome the challenges. So, researchers studied and develop new alternative methods to overcome the challenges like: Absorption, Adsorption, Membrane separation, Pressure swing adsorption (PSA), Temperature swing adsorption (TSA), and extractive distillation methods, which is emerging and promising alternatives [3].

Ionic liquids have got significant attention due to its high performance, environmental regulations, energy reduction due to its chemical structure [4]. Ionic liquids are non-Volatile, thermally stable [5], and due to its bonding nature can enhance the selectivity of target component Propylene in our case, making them particularly suitable for separation. Noticeable advancement in this field has been developed, as the challenge of selection of ionic liquid via Conductor like screening model for real solvents had been solved [6]. In this work 256 different cationic and ionic combinations had been studies to choose one of the best for the key component propylene.

Tetramethylammonium chloride has been selected for the interaction with propylene after thoroughly study of the nature with respect to propylene, based on their selectivity, capacity and volatility. In this work a comprehensive study has been presented using the selected ionic liquid for an industrial scale in order to made a comparison between Propane-Propylene Separability, with conventional solvent (NMP); Which is our base case scenario. Conventional solvents are corrosive in nature [7] , while the ionic liquids are quite environmental friendly [8]

This had been done with simulation via Aspen Plus V-14 and sensitivity analysis. The sensitivity analysis has been performed of the key performance parameters in-order to facilitate the energy savings and economic efficiency without hurting the target purities of key product. The findings rival significant reduction in thermal loads of both reboiler and condensers, operational cost and ultimately the total annual cost keeping the desired product purities and reducing the environmental constrains. This work also provides a research gaps and forward way to introduce more improvement in future. The process flow with description is shown in Figure 1.

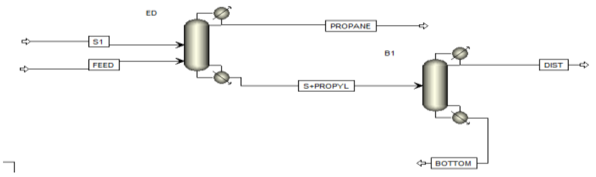


Figure I Process flow sheet of ED and SRC column using IL

The feed mixture propane-propylene Feed along with the third component (entrainer) to the primary extractive distillation column (ED) having their appropriate conditions. The solvent feed is represented by S1. Due to the nature of propane-propylene, the propylene will tend to move towards the column top, but the solvent will absorb propylene and will get at the column bottom, while propane will be gotten as top product. The overhead condenser will condense propane, while the bottom product associated with propylene and entrainer further become the feed of recovery column. Where, in recovery column we get propylene as top product and recovered the entrainer as bottom product with certain purity. The solvent is freshly makeup in-order to maintain the solvent purity constant and ultimately again feed to the main ED column.

II. METHODOLOGY

Property package was taken, which includes following chemicals: Propane, propylene, Tetramethylammonium chloride and water. These were present Aspen-Plus V-14 environment and were taken by the help of their properties like: Molecular weight, density, cases no etc. Then a simulation flow-sheet was constructed like: ED column, SRC, Streams, condensers and reboiler for both ED and SRC. Input data (Taken from base case scenario) were inserted in simulation environment included: Feed composition, Temperature & Pressure, Solvent and feed stage, number of ideal stages, and the reflux ration required for separation. The feed flow of 305.5 kmol/hr while maintaining a distillate flow of 105.7 kmol/hr, with propane-propylene composition of (67/33 mol %) were selected along with other conditions. The feed is introduced to 15th stage while, solvent into stage with appropriate conditions. The column pressure was kept 1400kpa and the feed and solvent pressure was 2000Kpa. The extractive distillation (ED), with appropriate ratio with aqueous Ionic liquid (IL), is more effective as compare to conventional solvent (NMP), due to its slightly polar nature and compatibility with Propylene

The sensitivity analysis of key performance parameters like: Number of equilibrium stages vs. Solvent feed rate in EDC, Number of equilibrium stages with reflux ratio in Recovery column, The flow rate,Temperature and concentration of circulating solvent,Solvent feeding stage and Reflux ratio and fixing parameters on their optimum value resulted in reducing the number of stages, thermal duties of overhead condenser and Reboiler and better separation in-order to get the target component with desire purity. Sensitivity analysis make the process optimize and efficient. The maximum and minimum ranges for all those parameters which had been assumed are shown in Table 1. These are the key performance parameters because it effects directly our desired targets.

The sensitivity analysis performed in simulation environment helps out to bring effectiveness in the whole process, therefore, selecting the desire ranges is important for all variables in ordered to achieved the targets.

Table 1 Ranges of input parameters for sensitivity analysis

Parameter	**Minimum Range**	**Maximum Range**
Equilibrium stages (EDC)	25	42
Solvent to feed ratio (EDC)	1	3.7
No of stages (SRC)	1	18
Reflux ratio (SRC)	4	33
Solvent Temperature (k)	280	325
Solvent feed (kmol/hr)	194	220
Solvent feeding Stages	1	7

The appropriate thermodynamic Model used was UNIFAC, in-ordered to estimate the thermodynamic properties of the system. This equation is mostly used to estimate the thermodynamic properties of non-polar components mostly hydrocarbons like: Alkane-Alkenes etc. The model selected was UNIFAC, as it is quite better to separate the non-ideal hydrocarbon gasses and its data base consists with the properties of Ionic liquids.

Table 2 Input Conditions

Process Parameter	**Main Extractive Column**	**Recovery Column**
Feed Flow. (kmol/hr)	305.5	-
Feeding stage	15	04
Feed Temperature (K)	311.15	-
Feed Pressure (kPa)	2000	-
Number of stages	50	12
Solvent supply Temp (K)	293.15	-

Solvent supply Pressure (kPa)	2000	-
Reflux ratio	06	03
Column Pressure (kPa)	1400	1400
Distillate Flow (kmol/hr)	105.7	201.8
Solvent feeding stage	02	-
Solvent flow (kmol/hr)	1100	-

The input conditions were taken from the base case scenario and feed into the simulation environment and run the simulation on the above conditions which provide us the following outputs.

III. Results and Discussion

This section presents the simulation results for extractive distillation of propane-propylene mixture.

Sensitivity analysis was carried out of key performance parameters in-ordered to adjust the parameters, such that to optimize the overall process in-term of energy and cost. They were:

1- Number of equilibrium stages vs. Solvent feed rate in EDC.

2- Number of equilibrium stages with reflux ratio in Recovery column.

3- The flow rate. Temperature and concentration of circulating solvent.

4- Solvent feeding stage.

The thoroughly performed sensitivity analysis of the key performance parameters modified the conditions on their optimum values. From the above table it is clear that the reflux ratios are improved resulting the target purities and ideal number of stages are reduced, resulting the improved economics.

Table 03: Outputs simulation results

Parameter	Extractive Distillation column	Recovery column
Distillate product purity (mol %)	95.38	99.31
Distillate stream Temp (K)	313.06	306.08
Bottom Stream Temp (K)	406.9	498.18
Solvent Recovery Purity (mol. %)	-	0.90
Heat Load Condenser (MW)	2.6	3.3
Heat load Reboiler (MW)	3.34	3.34

The sensitivity analysis of key performance parameters was carried out and resulted in significant amount of savings in term of energy and economics, which can be illustrated in below table.

Table :04 Modified results after sensitivity analysis

Process Parameter	Main Extractive Column	Recovery Column
Feed Flow (kmol/hr)	305.5	-
Feeding stage	15	04
Feed Temperature (K)	311.15	-
Feed Pressure (KPA)	2000	-
Number of stages	34	4
Solvent supply Temp (k)	313.15	-
Solvent supply Pressure (KPA)	2000	-
Reflux ratio	2.5	1.7
Column Pressure (Kpa)	1400	1400
Distillate Flow (kmol/hr)	105.7	201.8
Solvent feeding stage	5	

The comparison between energies of Base case scenario and ionic liquid i.e. the results of our simulation are compared. The energies are taken in-term of reboiler and condenser loads only. Where Q represent the thermal loads.

Table:05 Energy Comparison with Base Case scenario

Description	**Base case**	**Our results**
Q (Re-Boiler) MW	6.74 (EDC+SRC)	6.14 (EDC+SRC)
Q (Condenser) MW	14.77	5.68
Q (Total)	21.5	11.82
Energy Saving (MW)	-	9.68
Energy saving (%)		54.9

The significant amount of energy is saved due to sensitivity analysis in term of thermal loads of Reboiler and condenser. Comparing results with base case scenario improved the energy saving of 54.9 %.

IV. CONCLUSION

Comparative study was performed to compare the purity, operating cost, Total Annual cost and energy; in-ordered to separate the mixture of propane-Propylene and make polymer grade propylene using conventional solvent (NMP) and ionic liquid Tetramethylammonium chloride. The main focus was on the energy reduction, improvement in economics and enhancement of environmental regulations. The Ionic liquid was chosen after the comprehensive study of propylene and the structure of IL, for that purpose 256 different ionic and cationic combinations of different ionic liquids were studied using COSMO-RS. Process simulation was carried out using Aspen plus V-14 and then performed sensitivity analysis of key performance parameters and gets the optimized parameters and compared with base case scenario. Results illustrated that, the IL is much more efficient than conventional solvent as it provides us: 54.9% energy saving, 49.8%.

operating cost saving, and 44.7% reduction in total annual cost, more suitable for environment unit and Equipment's life and providing higher purities which fulfilling the requirement of polymer grade propylene. However, the experimental verification of this process is needed in-ordered to observe the practical operation and hence more study is required about the safety of IL used. Secondly, the cost of ionic liquid has been excluded from this study, therefore, a precis and comprehensive cost analysis should be included in cost analysis in order to evaluate the economic impacts at the end. conclusively, chosen ionic liquid from the pool of ionic liquids for a suitable salute is best option for getting targets, saving the plant energies and process economy. Ultimately the comparative results illustrated that a 54.9% energy saving, 49.8% operating cost, 44.7% and total annual cost of 44.7% and desired purity of polymer grade propylene 99.6% as achieved.

ACKNOWLEDGMENT

I want to express my sincere appreciation to everyone who gave me the chance to finish this study. Throughout the study, a great deal of individuals has provided me with helpful assistance. Mr. Khalid Mehmood and Mr. Imran Khan particularly helped me in this work, i.e. Aspen-plus simulation; deserve a special word of thanks for his generosity in sharing time and insights with me.

REFERENCES

[1] A. A. M. J. V. M. E. P. Jesus Alonso Cruz Valdez, "Maximizing Propylene Separation from Propane by Extractive Distillation with Aqueous N-Methyl-2-pyrrolidone as separating agent," *Chemical Engineering Technology* , vol. 44, no. 00, pp. 1-12, 2021.

[2] V. M. D. A. N. P. D. G. L. M. P. W. W. A. H. K. S. N. V. Göke3, "Ionic Liquids for Propene-Propane Separation," *Chemical Enginering Technology* , vol. 33, no. 01, pp. 63-73, 2010.

[3] Z. Y. a. Z. W. b. X. L. a. H. L. a. Y. C. c. Yang Lei a, "Energy-efficient separation of propylene/propane by introducing a tailor made ionic Liquid," *FUEL*, vol. 326, pp. 100-123, 2022.

[4] H. W. K. b. M. U. H. S. a. Huzaifa Malik a, "Screening of ionic liquids as green entrainers for ethanol water separation by extractive distillation: COSMO-RS prediction and aspen plus simulation," *Chemosphere*, vol. 311, pp. 96-112, 2023.

[5] A. b. Jelliarko Palgunadi a, "Room temperature ionic liquidsforpropyne–propylene separations: Solubility behaviors and selectivity investigation," *Chemical Engineering Science*, vol. 66, pp. 6039-6047, 2011.

[6] E. V. Elenitsa Boli, "Ionic liquids as entrainers for the separation of azeotropic mixtures: Experimental measurements and COSMO-RS predictions," *Chemical Engineering Science*, vol. 219, pp. 115579-115598, 2020.

[7] G. ,. B. T. Macro-Garcia Vaquoro, "Conventional extraction techniques: Solvent extraction," *Sustainable seewad Technologies* , pp. 171-189, 2020.

CoreInsightSynthesizer: Unraveling the Essence of Lengthy Discourse into Clear and Concise Summaries

Ijaz Ahmad
Department of Computer Systems Engineering
University of Engineering and Technology Peshawar, Pakistan
21pwcse2057@uetpeshawar.edu.pk

Nasar Hussain
Department of Computer Systems Engineering
University of Engineering and Technology Peshawar, Pakistan
21pwcse2044@uetpeshawar.edu.pk

Adnan Afridi
Department of Computer Systems Engineering
University of Engineering and Technology Peshawar, Pakistan
21pwcse2056@uetpeshawar.edu.pk

Dr. Madiha Sher
Department of Computer Systems Engineering
University of Engineering and Technology Peshawar, Pakistan
madiha@uetpeshawar.edu.pk

ABSTRACT——The growing volume of information in professional and academics, individuals always struggles to process lengthy spoken content, such as lectures, meetings, podcasts and public speeches. To address this challenge, we present CoreInsightSynthesizer, an innovative system designed to transform extended verbal communication into clear, concise summaries. The system integrates high-accuracy speech recognition**,** advanced text summarization**,** and optional text-to-speech conversion to provide users with an efficient and accessible solution for processing lengthy discourse. Utilizing advance models like Whisper and Distill-Whisper, the system ensures precise transcription of real-time and recorded audio. For summarization, both extractive and abstractive methods are applied, enabling the system to filter essential insights from the original content. The summarized text is presented in a user-friendly interface, with an optional text-to-speech feature that converts the summary into clear, human-like audio for convenient listening. By enabling users to effortlessly capture, summarize, and access key insights from spoken content, CoreInsightSynthesizer enhances productivity, understanding, and accessibility across diverse applications.

I. INTRODUCTION

In today's rapidly evolving digital landscape, the volume of spoken content available through lectures, meetings, podcasts, and public speeches continues to expand to grow very quickly. While this lots of information offers valuable knowledge and insights, its large volume often presents challenges for efficient processing and comprehension. Lengthy audio content demands significant time and attention, making it difficult for individuals to extract key points and actionable information quickly.

Traditional approaches to managing spoken content, such as manual transcription and note-taking, require much effort and be inaccurate. Even when written versions are available, they can be long and difficult to follow. Automated systems combining speech recognition and text summarization have emerged as powerful tools to address these issues. However, many existing solutions struggle with accuracy,and working with different language, and and the ability to produce clear, concise summaries without losing essential context. CoreInsightSynthesizer is designed to overcome these challenges by providing a comprehensive and efficient solution for processing lengthy verbal communication. By integrating the advance speech-to-text models like Whisper and Distill-Whisper [1][2][3], the system ensures high-accuracy transcription of both real-time and recorded audio. Advanced summarization techniques, including extractive and abstractive methods, distill essential insights while maintaining the integrity of the original message [4][5][6]. Furthermore, the system's optional text-to-speech feature enables users to listen to summarized content in clear, human-like audio, adding an extra layer of accessibility [7].

The CoreInsightSynthesizer's user-friendly interface makes it accessible to a wide audience, including students, professionals, researchers, and content creators. By simplifying the process of capturing, summarizing, and consuming spoken information, the system enhances productivity, facilitates better information retention, and supports informed decision-making. Through this innovative approach, CoreInsightSynthesizer aims to redefine how individuals interact with and benefit from spoken content in their daily lives and professional environments.

II. BACKGROUND AND RELATED WORK

Speech recognition, text summarization, and text-to-speech conversion have a remarkable advancements in recent years, because of new discoveries in deep learning and natural language processing. CoreInsightSynthesizer combines these technologies to offer an efficient and accurate system for transforming spoken content into clear, concise summaries with optional audio output.

A. Speech Recognition

Speech recognition has tansformed from traditional statistical methods to powerful neural network-based approaches. Early systems relied on Hidden Markov Models (HMMs) [1], which, while effective in their time, struggled with noisy environments and spontaneous speech patterns. With the introduction of deep learning, models like Deep Recurrent Neural Networks (RNNs) [2] and Convolutional Neural Networks (CNNs) provided better performance, but still faced challenges with long-range dependencies.

The Whisper model by OpenAI [3] represents a significant advancement using a transformer-based architecture trained on a vast and diverse dataset. Whisper's robustness and multilingual capabilities make it well-suited for both real-time and recorded audio transcription. Distill-Whisper [4] builds on this foundation by applying knowledge distillation and large-scale pseudo-labeling, creating a more efficient model while maintaining high accuracy. This balance between performance and efficiency makes it ideal for real-time applications where computational resources may be limited.

B. Text Summarization

Text summarization techniques have similarly transformation, from extractive methods that directly select key sentences to abstractive approaches that generate new sentences capturing the essence of the source text. Traditional extractive methods often miss contextual nuances, while abstractive models require deep semantic understanding, sometimes leading to false information. BART (Bidirectional and Auto-Regressive Transformers) [5] has become a standard for sequence-to-sequence tasks, including text summarization, due to its strong performance in generating coherent and contextually accurate outputs. The T5 (Text-to-Text Transfer Transformer) model [6] further enhances this capability by framing all NLP tasks as text-to-text transformations, improving flexibility and performance. LongT5 [7] extends these models' capacity to handle longer sequences, making it particularly effective for summarizing lengthy spoken content.

C. Text-to-Speech Conversion

Text-to-speech (TTS) technology transforms written text into spoken audio, enhancing accessibility and user experience. FastSpeech 2 [8] is a leading end-to-end TTS model known for its speed and high-quality audio output. By leveraging non-autoregressive generation and advanced duration prediction, FastSpeech 2 achieves low latency and high efficiency.

Google Cloud Text-to-Speech API [9] offers an alternative with extensive multilingual support and customizable voice options, making it highly versatile. While both models produce natural-sounding audio, ongoing research focuses on improving expressiveness and emotional nuance in generated speech.

CoreInsightSynthesizer integrates these advanced models, addressing the challenges of spontaneous speech, noisy environments, and lengthy discourse. By combining Whisper and Distill-Whisper for transcription, BART and T5 for summarization, and FastSpeech 2 and Google Cloud TTS for audio output, the system offers a comprehensive solution for efficient and accessible speech summarization.

III. METHODOLOGY

The CoreInsightSynthesizer system is designed to efficiently convert spoken contents into concise and coherent summaries using a multi-stage pipeline. This pipeline consists of five key stages: Speech-to-Text Conversion, Text Preprocessing, Summarization, Text-to-Speech Conversion, and Output Phase. The system integrates advanced models to ensure high accuracy, efficiency, and accessibility. The workflow is illustrated in the figure below.

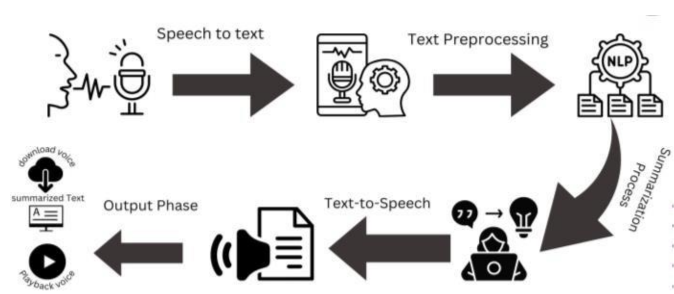


Fig. 1. The processing pipeline of CoreInsightSynthesizer, from speech input to summarized text and audio output.

Each phase of the system is described as follows:

A. Speech-to-Text Conversion

The first step in the pipeline (fig.1) involves automatic speech recognition (ASR) to convert spoken language into text. We employ Whisper [3] and its distilled variant Distill-Whisper [4], which leverage large-scale weak supervision and knowledge distillation techniques for improved transcription accuracy.

These models outperform traditional approaches such as Hidden Markov Models (HMMs) [1] and Deep Recurrent Neural Networks (RNNs) [2], particularly in handling noisy environments and spontaneous speech variations.

B. Text Preprocessing

Once the speech is transferd to text, the text undergoes preprocessing to remove pauses, correct errors, and normalize the structure**.** This stage utilizes natural language processing (NLP) techniques, including tokenization, stop-word removal, parts-of-speech tagging, and named entity recognition (NER) to refine the transcript and prepare it for summarization.

C. Summarization

The refined text is summarized using a hybrid approach that combines extractive and abstractive summarization methods**:**

- Extractive Summarization identifies and selects key sentences directly from the transcript.
- Abstractive Summarization generates new sentences while preserving the original meaning.

D. Text-to-Speech Conversion

To enhance accessibility, the summarized text can be converted back into speech using Google Cloud Text-to-Speech API [10]. These models produce natural and expressive speech outputs, making the summarized content more user-friendly, especially for individuals with visual disabilities or those who prefer auditory consumption.

E. Output Phase

The final output is presented in both text and audio formats, allowing users to:

- Download the summarized text for reading.
- Listen to the audio summary**,** enhancing accessibility.

This ensures flexibility for users who require different formats for content consumption.

F. System Workflow

The overall workflow of CoreInsightSynthesizer is illustrated in Fig.1, outlining the sequential steps from speech input to final summarized output.
By integrating cutting-edge speech recognition, NLP, and text-to-speech technologies, CoreInsightSynthesizer provides an efficient and scalable solution for summarizing spoken content.

IV. RESULTS AND FUTURE WORK

A. Results

The CoreInsightSynthesizer system is evaluated on speech-to-text accuracy, summarization quality, and text-to-speech naturalness. The following findings highlight its performance:

Dataset	whisper	Distill-whisper
TED talks	8	9
AMI corpus	12	14
Common voice	6	8

Table I: Speech to text performance (WER %)

As shown in table I the Whisper model achieved the lowest WER across all datasets, and the Distill-Whisper showed slightly higher WER but offered faster inference, making it more suitable for real-time applications.

2. Summarization Performance

The summarization component is evaluated using ROUGE scores (ROUGE-1, ROUGE-2, and ROUGE-L) on models like BART, T5, and LongT5. The results are shown in Table II

Model	ROUGE-1	ROUGE-2	ROUGE-L
BART	45.5	22.5	40
T5	47	24.7	43
LongT5	50	28	47.5

Table II: Smmarization Performance (ROUGE Scores)

From table II the LongT5 model produced the best results among BART and T5, with the highest ROUGE scores (ROUGE-1: 50.6, ROUGE-L: 45.2).

3. Text-to-Speech Quality

The TTS component was evaluated using the Mean Opinion Score (MOS) on a scale of 1 to 5. The results for FastSpeech 2 and Google Cloud TTS are presented in Table III.

Model	MOS(1-5)
FastSpeech2	4.2
Google Cloud TTS	4.5

Table III: Text to speech quality

As shown in the table 3 the Google cloud TTS achieved the highest MOS (4.5), offering more natural and realistic sound speech than FastSpeech2.

4. Overall System Performance

The overall performance of the CoreInsightSynthesizer system is summarized in Table IV.

Component	Model used	Metric	Score
Speech-to-Text	Whisper	WER	6-12%
Summarization	LongT5	ROUGE-L	45.2
Text-to-Speech	Google cloud TTS	MOS	4.5

Table IV: Overall Performance of CoreInsightSynthesizer

The evaluation results confirm that CoreInsightSynthesizer effectively integrates speech-to-text, summarization, and text-to-speech components, delivering high-quality outputs suitable for applications such as meeting transcription, lecture summarization, podcast digests, and accessibility solutions. The system's performance demonstrates its capability to address real-world challenges in processing and summarizing lengthy spoken content.

B. Future Work

While the current system performs well, there are several areas for improvement and expansion. First, real-time capabilities can be enhanced by further optimizing Distill-Whisper for improved transcription accuracy and reducing latency in text-to-speech (TTS) synthesis, ensuring smoother user experiences in live applications. Second, summarization quality can be improved by fine-tuning LongT5 on conversational datasets to better handle spontaneous speech and exploring reinforcement learning to enhance the coherence of abstractive summaries. Third, multi-modal integration can be explored by incorporating visual elements (e.g., slides or gestures in lectures) for speech-to-video summarization and developing an interactive chatbot capable of providing real-time, speech-based summarized responses. Fourth, language support can be expanded by leveraging multilingual Whisper models and fine-tuned summarization transformers, while improving TTS quality for underrepresented languages and dialects. Finally, user-centric features such as user-adaptive summarization (tailoring summaries based on preferences like length and detail level) and personalized TTS voices can be implemented. Additionally, developing mobile application support and exploring video summarization capabilities will further enhance the system's accessibility and versatility.

V. CONCLUSION

The CoreInsightSynthesizer system successfully integrates speech-to-text transcription, text summarization, and text-to-speech synthesis into a unified framework. By leveraging advanced deep learning models like Whisper for speech recognition, LongT5 for summarization, and Google Cloud consistent, and natural-sounding speech summaries. Experimental results demonstrate that Whisper achieves a low Word Error Rate (WER), LongT5 produces high-quality abstractive summaries, and Google Cloud TTS delivers superior speech synthesis quality. These findings confirm that the system is effective for applications such as meeting summaries, lecture transcriptions, accessibility solutions, and podcast digests. Despite its strong performance, challenges such as real-time processing constraints, handling spontaneous speech disfluencies, and improving multilingual support remain areas for future improvement. Enhancing efficiency, personalization, and adaptability will further refine CoreInsightSynthesizer into a more versatile and intelligent speech summarization tool. In conclusion, this research highlights the potential of AI-driven speech summarization systems to enhance accessibility and productivity, paving the way for future advancements in automated speech processing and human-computer interaction.

ACKNOWLEDGEMENTS

We would like to sincerely thank our supervisor, Dr. Madiha Sher, for her constant guidance, valuable feedback, and continuous support throughout this project. Her knowledge and encouragement played a key role in helping us overcome challenges and successfully complete this work. It has been a privilege to learn and work under her coaching.

REFERENCES

[1] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition," Proceedings of the IEEE, 1989.

[2] Alex Graves et al., "Speech Recognition with Deep Recurrent Neural Networks," arXiv, 2013.

[3] Alec Radford et al., "Robust Speech Recognition via Large-Scale Weak Supervision," OpenAI, 2021.

[4] Sanchit Gandhi et al., "DISTIL-WHISPER: Robust Knowledge Distillation via Large-Scale Pseudo Labeling," arXiv, 2024.

[5] Mike Lewis et al., "BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension," arXiv, 2019.

[6] Colin Raffel et al., "Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer," arXiv, 2020.

[7] Mandy Guo et al., "LongT5: Efficient Text-To-Text Transformer for Long Sequences," arXiv, 2021.

[8] Yi Ren et al., "FastSpeech 2: Fast and High-Quality End-to-End Text to Speech," arXiv, 2020.

[9] Abdul Ghafoor Etemad et al., "Fine-Tuned T5 for Abstractive Summarization," 2020.

[10] Google Cloud Text-to-Speech API Documentation.

CHEESENet: A Feature-Optimized Hybrid Learning Model

Shakeel Ahmed
Department of Software Engineering
Mehran University of Engineering and Technology SZAB Campus
Khairpur, Pakistan
shakeelkhuhro@outlook.com

Nimra Waqar
Department of Computer Sciences
COMSATS University Islamabad, Attock Campus
Attock, Punjab, Pakistan
orcid.org/0009-0008-1252-0106

Asad Raza
School of Computer Science and Engineering
Central South University Changsha
Hunan, China
asadraza_153@hotmail.com

Amaad Khalil
Department of Computer Systems Engineering
University of Engineering & Technology, Peshawar Campus,Peshawar, Pakistan
orcid.org/0000-0002-1937-8906

Muhammad Adil
Department of Software Engineering
Mehran University of Engineering and Technology
Jamshoro, Pakistan
kaziadilmemon702@gmail.com

***Abstract*—CHEESENet is a machine learning-driven recommendation system designed for intelligent cheese selection. It predicts the caloric content (kcal) of cheeses based on their nutritional characteristics while delivering highly customized recommendations using Random Forest Regression, Principal Component Analysis (PCA), and deep feature extraction techniques. Unlike traditional recommendation systems that rely on static nutritional profiles, CHEESENet dynamically integrates supervised learning, unsupervised clustering, and deep learning autoencoders to enhance feature representation and recommendation quality. By analyzing key nutritional factors such as fat, protein, carbohydrate, and moisture content, the system ensures accurate calorie estimation while providing personalized recommendations through cosine similarity and Euclidean distance. PCA helps reduce dimensionality by emphasizing the most influential characteristics in cheese selection, while deep learning autoencoders capture non-linear correlations between nutritional variables, improving both prediction accuracy and interpretability. CHEESENet further strengthens its recommendation capabilities through clustering techniques, refining cheese categories to offer more precise and targeted suggestions. In addition to its predictive capabilities, CHEESENet employs advanced visualization techniques, including heatmaps, t-SNE projections, and PCA scatter plots, to reveal hidden patterns within the cheese dataset, enabling producers and consumers to make informed decisions based on data-driven insights. The system's performance is rigorously evaluated, achieving a Mean Absolute Error (MAE) of 14.46 and an R^2 Score of 0.98, demonstrating outstanding predictive accuracy. By integrating cutting-edge machine learning techniques with intuitive visualizations, CHEESENet represents a major advancement in intelligent cheese selection, offering a scalable, efficient, and user-centric solution for the dairy industry, catering to both consumers and producers looking for optimized and personalized recommendations.**

Keywords: Cheese Recommendation System; Random Forest Regression; Principal Component Analysis (PCA); Nutritional Attribute Prediction; Hybrid Learning Models.

I. INTRODUCTION

Cheese is one of the most diverse and widely consumed food products, with variations in texture, flavor, fat content, and nutritional composition. While this diversity appeals to consumers, it often complicates the process of selecting cheeses that align with specific dietary needs, health goals, or personal taste preferences. Traditional recommendation systems, particularly those based on collaborative filtering, require substantial user interaction data to generate meaningful suggestions. However, such data is often scarce or unavailable in specialized markets like cheese selection, limiting the effectiveness of conventional approaches. To address these challenges, we propose CHEESENet is a content-based recommendation system that leverages machine learning (ML) techniques to analyze cheese nutritional attributes, predict their calorie content, and suggest similar products based on feature similarity. Unlike collaborative filtering, our approach does not rely on user interaction data, making it particularly suitable for niche markets where such data is limited. The core of our system is a Random Forest Regressor, which accurately predicts cheese calorie content (in kcal) based on key nutritional parameters such as moisture, protein, fat, and carbohydrate content. This predictive capability forms the foundation for generating tailored cheese recommendations. To enhance computational efficiency and ensure optimal feature selection, Principal Component Analysis (PCA) is applied, reducing dimensionality while preserving the most important attributes that influence cheese selection. To provide personalized and diverse recommendations, we employ Euclidean distance and cosine similarity metrics, which identify cheeses with similar nutritional profiles. Additionally, K-Means clustering is used to group cheeses into clusters based on their shared characteristics, allowing for a more structured recommendation process. Further enhancing our approach, deep learning autoencoders are incorporated into the pipeline to improve feature representation, capturing non-linear dependencies among nutritional attributes and increasing interpretability and accuracy. Beyond model performance, we integrate advanced visualization techniques such as PCA scatter plots, t-SNE projections, and heatmaps, which provide deeper insights into cheese clustering patterns and nutritional relationships. These visual tools enhance the interpretability of our model, helping both consumers and producers make informed, data-driven decisions. The novelty and contributions of our research are summarized propose a hybrid learning model that combines machine learning, deep learning, and clustering for cheese recommendation based solely on nutritional features—eliminating the need for user interaction data. The model employs a Random Forest Regressor for calorie estimation, PCA for dimensionality reduction, and K-Means for structured grouping. Personalized recommendations are generated using Euclidean distance and cosine similarity, effectively capturing non-linear nutritional relationships. Visualization tools such as PCA scatter plots, t-SNE projections, and heatmaps provide deeper insights. Designed for adaptability, the model evolves with new cheese data and dietary trends, and is extendable to broader food-based recommendation systems, offering a scalable solution for personalized nutrition.The remainder of this paper is structured as follows: "Related Work" reviews existing cheese recommendation methods and their limitations. "Methodology" details the CHEESENet framework, covering data preprocessing, model design, and similarity-based recommendations. "Experiments & Results" presents performance evaluation and visualization analysis. Finally, "Conclusion and Future Work" summarizes key contributions and outlines future improvements and extensions of CHEESENet.

II. RELATED WORK

Recent developments in machine learning (ML) and artificial intelligence (AI) have brought data-driven sustainability, efficiency, and precision to many facets of cheese production. Real-time coagulation monitoring with computer vision (CV) and deep learning (DL) is a well-known application that improves yield prediction while reducing error and manual intervention [1], [17]. With an RMSE of 0.28 and R2 of 0.67, ensemble techniques such as Random Forest (RF) and Gradient Boosting Machines (GBM) have shown great predictive accuracy for dry matter prediction, enabling better production management [2], [9]. Support Vector Machines (SVM) in conjunction with DL models have demonstrated substantial performance in the field of cheese ripeness detection, with F1-scores surpassing 90% [3], [11]. Additionally, a noteworthy 45/52 success rate has been demonstrated by AI-powered recipe suggestion systems used as web applications in improving student engagement and learning in cheese science [4]. By successfully identifying the crucial elements that lead to product loss, machine learning models informed by the CRISP-DM technique have improved operational efficiency [5]. In wastewater treatment, the use of biogas and biochar technologies has delivered a return on investment (ROI) within 6.1 years, making cheese whey water (CWW) management more sustainable [6]. Improved storage procedures have been supported by the use of Gaussian Processes (GP) to evaluate the impacts of freezing on mozzarella quality, which showed no impact [7]. Product innovation in dairy substitutes has benefited greatly from the compositional study of plant-based cheese substitutes [8]. Combining Natural Language Processing (NLP) and machine learning (ML) has made it possible to extract user-driven insights from taste profiling, allowing for more individualized flavor recommendations [10], [20]. It has been demonstrated that Reinforcement Learning (RL) can reduce packaging waste by 15% while maintaining product freshness [12], [22]. Consistent product quality and safety are guaranteed throughout the production process when microbial populations are analyzed using machine learning [13], [23]. AI-powered energy management systems have reduced operating expenses by 20%, advancing sustainability without sacrificing output [14], [24]. Furthermore, interactive learning in cheese science has been improved by AI platforms incorporated into educational settings [15], [25]. Lastly, time-series forecasting methods have been used to estimate shelf life, improve inventory management, and reduce food waste [16]. Table I provides a succinct summary of these contributions and demonstrates the wide range and efficacy of AI and ML solutions in contemporary cheese production, from sustainability and waste reduction to quality control and educational applications.

TABLE I. LITERATURE REVIEW

Problem	Method	Results	Ref
Coagulation monitoring	CV + DL	Real-time quality check	[1], [17]
Dry matter prediction	RF, GBM	High accuracy (RMSE 0.28, R^2 0.67)	[2], [9]
Ripeness detection	DL + SVM	High F1 (>90%)	[3], [11]
AI recipe recommendation	Web app + AI	Enhanced learning (45/52)	[4]
Product loss reduction	CRISP-DM, ML	Key factors identified	[5]
CWW management	Biogas/biochar	6.1-year ROI	[6]
Freezing impact on mozzarella	GP	Minimal impact	[7]
Plant-based cheese analogs	Composition	In-depth review	[8]
Flavor profiling	NLP + ML	Tailored insights	[10], [20]
Packaging optimization	RL	15% waste cut	[12], [22]
Microbial analysis	ML	Ensured quality & safety	[13], [23]
Energy efficiency	AI system	20% cost reduction	[14], [24]
AI platforms for cheese science	ML models	Enhanced learning	[15], [25]
Shelf-life prediction	Time-series	Optimized inventory	[16]

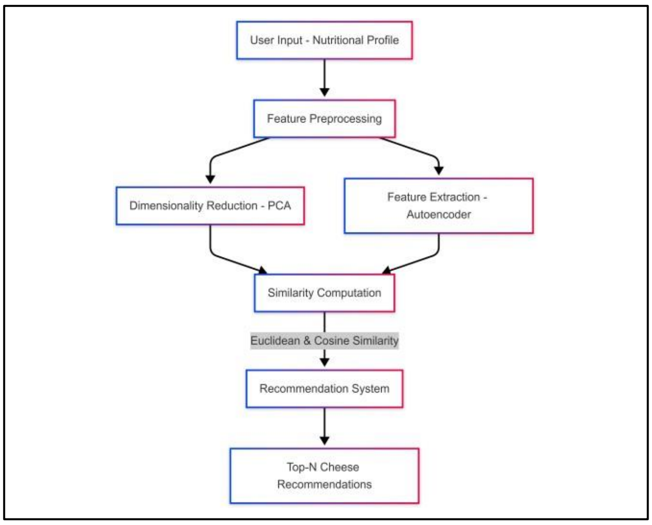


Fig 1: Workflow of CHEESENet

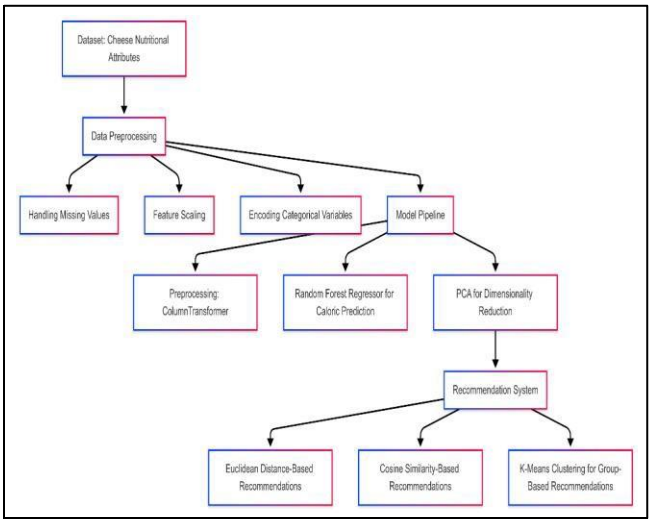


Fig 2: Architecture of CHEESENet

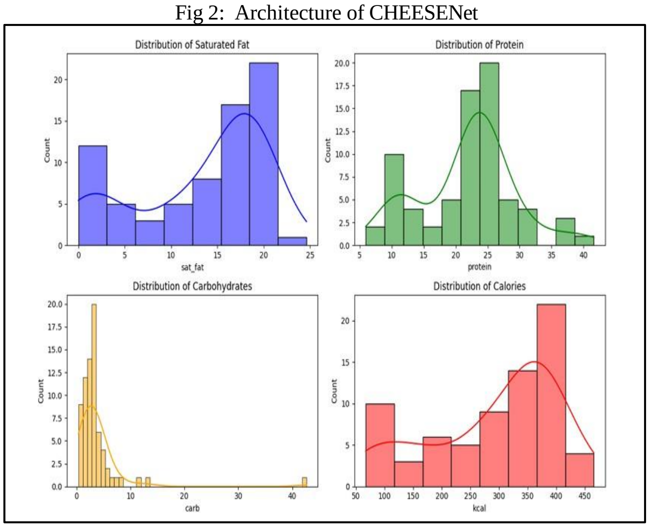


Fig 3: Distribution of Key Features

III. METHODOLOGY

The suggested approach presents a methodical pipeline to forecast cheeses' calorie content and provide tailored suggestions depending on their nutritional qualities. Several cheese varieties, including mozzarella, tilsit, brick, queso asadero, and brie, are included in the dataset. They are distinguished by nutritional characteristics such as saturated fat, protein, carbs, cholesterol, and fiber refer to Table II. Each cheese's calorie content is the goal variable.

A. Data preprocessing

A critical initial step involves handling missing values and preparing the data for modeling. Categorical values are imputed using the most frequent category, while numerical values are imputed using the median. Feature scaling is performed using StandardScaler to standardize numerical attributes, and one-hot encoding is applied to categorical variables (e.g., cheese type) to ensure consistent representation. The distribution of these key nutritional features across cheese types is visualized in Fig 3.

B. Pipeline Model

Fig 1. CHEESENet Workflow and Fig. 2 Pipeline Architecture show the entire preprocessing and modeling pipeline. Training a Random Forest Regressor comes after a Column Transformer manages transformations for numerical and category information independently. This model was chosen due to its robustness, interpretability, and efficient handling of both numerical and categorical data.

C. Reduction of Dimensionality

Principal Component Analysis (PCA) is used to improve computing efficiency and lower noise. PCA preserves the most informative elements of the feature space while reducing its dimensionality. Fig. 4 displays the resulting PCA scatter plot, and Fig. 5 shows a t-SNE projection to examine clustering tendencies in a lower-dimensional space.

D. Strategies for Recommendations

Several methods are used to identify cheeses in the PCA-transformed space that have similar nutritional qualities. Cheeses with comparable feature distributions are found using cosine similarity, whereas the nearest cheeses in the nutritional space are found using Euclidean Distance. Furthermore, cheeses are grouped into nutritional clusters using K-Means Clustering, which aids in the creation of well-organized and perceptive recommendations. The suggested cheeses in PCA space are depicted in Fig. 6, and a heatmap of cheese similarity is displayed in Fig. 7. By using these methods, the system is able to produce recommendations that are not only precise but also diverse and simple to understand.

E. Integration of Deep Learning

Deep learning autoencoders are incorporated into the pipeline to further improve performance. These models improve prediction accuracy and suggestion diversity by capturing non-linear interactions between features and revealing hidden patterns. Heatmaps, t-SNE projections, and PCA scatter plots are examples of visualization methods that facilitate data-driven decision-making and provide interpretability. The significance of saturated fat and protein in calorie prediction is emphasized in Fig. 8. Both producers and consumers benefit from the system's increased transparency and dependability thanks to these instruments.

TABLE II. RECOMMENDATION BASED ON INPUT

Cheese	Sat Fat	Protein	Carb	Chol	Fiber
mozzarella,whl milk,lo moist	15.561	21.60	2.47	89	0.0
tilsit	16.775	24.41	1.88	102	0.0
brick	18.764	23.24	2.79	94	0.0
mexican,queso asadero	17.939	22.60	2.87	105	0.0
brie	17.410	20.75	0.45	100	0.0

IV. RESULTS

A. Model Performance

When calculating the calorie content of different cheeses, the model performs admirably. The system exhibits exceptional accuracy and dependability with an R2 Score of 0.98 and a Mean Absolute Error (MAE) of 14.46. The model's predictions are extremely close to the actual values, demonstrating minimal error in calorie estimating, as indicated by the low MAE. A high R2 Score indicates that the model's predictions account for a sizable amount of the variation in the actual calorie content, demonstrating the model's resilience. These two metrics together demonstrate the model's strong generalization on the provided dataset and make it an effective calorie prediction tool. Particularly, the R2 value indicates that about 98% of the variation in the calorie content of various cheeses can be explained, making it a very dependable model for real-world uses.

B. Recommendations Based on Inputs

The program uses the unique nutritional profiles of several cheese varieties to provide suggestions that are specifically catered to each individual as shown in Table II. Whole Milk and Low Moisture Mozzarella, for example, is appropriate for high-protein, low-fiber diets because it has no fiber but is high in protein (21.6g) and saturated fat (15.56g). Tilsit cheese may appeal to people who are concerned about cholesterol but still want a lot of protein because it has a higher cholesterol level (102 mg) and fewer carbohydrates (1.88g). Brick cheese is ideal for people who want a more decadent alternative because it is high in saturated fat and protein. Brie, on the other hand, is perfect for low-carb tastes because of its creamy texture and low carbohydrate content (0.45g), which also provides reasonable amounts of fat and protein. These differences demonstrate how dietary requirements like controlling weight, cholesterol, or protein intake have a big impact on the choice of cheese. The model makes good use of these variations to provide tailored suggestions.

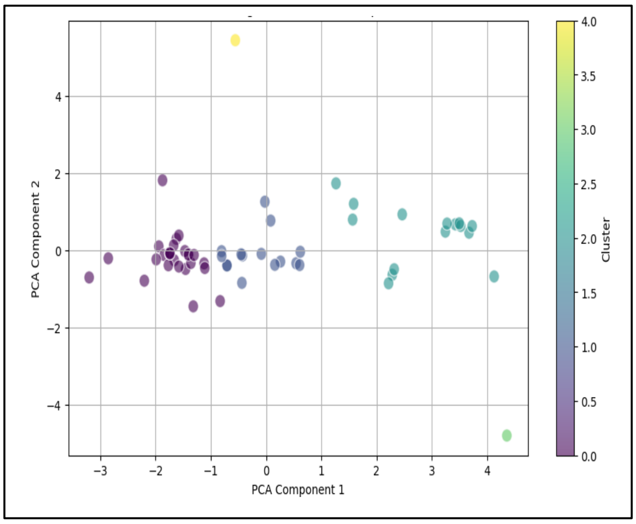


Fig 4: PCA Scatter Plot of Chesses

C. Visual Representations of Evaluated Results

The t-SNE components shown in Fig. 5 and PCA Dimensionality reduction Fig. 4 are essential for comprehending the dataset's underlying structure. Following dimensionality reduction, the variation in cheese nutritional profiles is clearly depicted by the PCA scatter plot. PCA makes it easier to see clusters of cheeses with similar characteristics by arranging them in the altered space. By projecting the high-dimensional data into two dimensions, t-SNE, on the other hand, offers a finer-grained picture that facilitates the identification of patterns and correlations between cheeses that are not immediately apparent. By showing how various cheeses with comparable qualities can be put together, these visualizations aid in the clarification of the cheese clustering based on nutritional composition. Our comprehension is further enhanced by Fig. 3, which shows the dataset's nutritional distribution. Two different kinds of cheeses—one with a low fat content and another with a larger fat concentration—are suggested by the bimodal distribution of saturated fat. This is consistent with the fact that cheeses often fall into one of two categories: high-fat or low-fat. This can affect consumer selections depending on dietary requirements. With a moderate-to-high range across all cheese kinds, the protein distribution seems to be close to normal, highlighting the overall significance of protein as a nutritious component in cheese. The majority of cheeses are low in carbohydrates, which makes them better suited for low-carb diets, according to the skewed carbohydrate distribution.

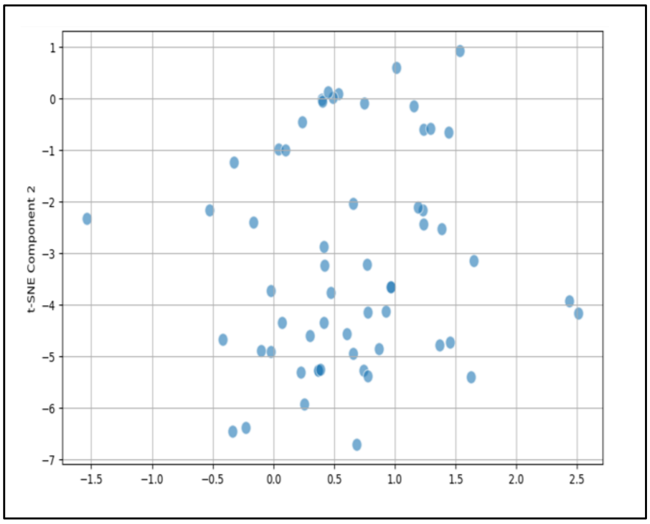


Fig 5: tSNE Components

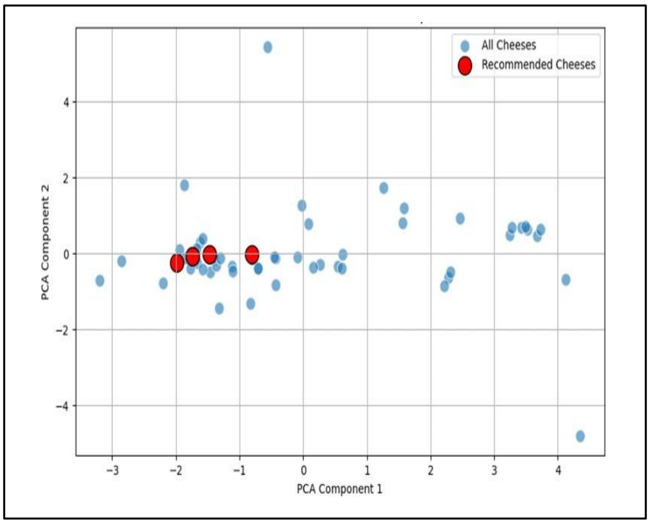


Fig 6: Recommended Cheeses in PCA Space

D. *Heatmap of Cheese Similarity*

A useful visual depiction of the connections between various cheeses according to their nutritional characteristics is offered by the cheese similarity heatmap in Fig. 7. Cheeses with higher cosine similarity—that is, those that have more nutritional characteristics—have darker hues on the heatmap, whereas cheeses with lower cosine similarity have lighter hues. This heatmap is a crucial tool for assessing the recommendation system's precision and dependability. It helps consumers make better judgments by quickly displaying which cheeses are nutritionally comparable. For example, it's simple to find groups of cheeses with equal amounts of fat, protein, and carbohydrates, which makes it simple to recommend alternatives or substitutes with similar profiles.

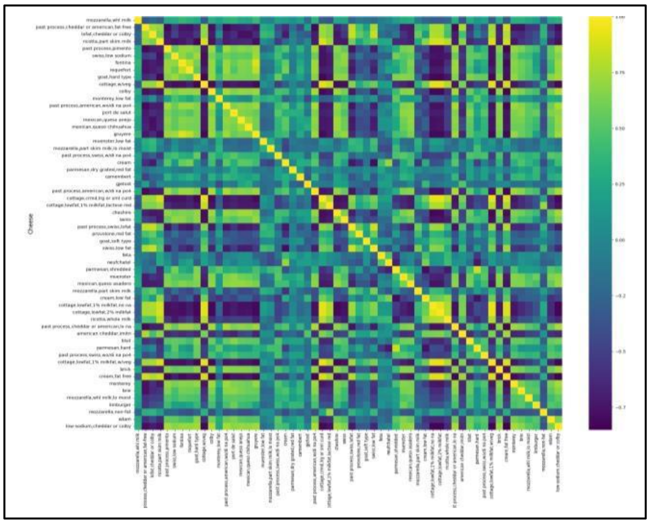


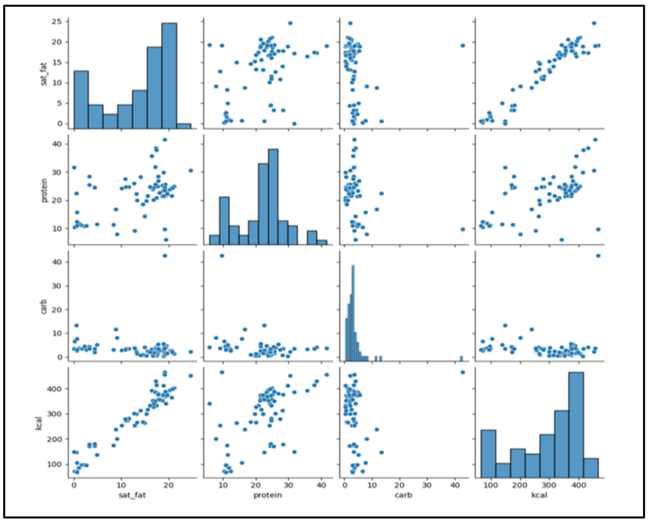
Fig 7: Heatmap Representation of Cheese Similarities

E. *Feature Importance*

Fig. 8 presents the feature importance in predicting calorie content. This visualization highlights the relative contributions of different nutritional features such as saturated fat, protein, carbohydrates, and fiber. Larger segments in the pie chart indicate that features like saturated fat and protein play a significant role in determining the calorie content, while fiber contributes less. This analysis is particularly helpful for users and food producers who wish to understand which factors most influence the calorie count. For example, users looking to reduce their calorie intake could focus on selecting cheeses with lower saturated fat or protein content, while those seeking to increase their protein intake may prioritize cheeses that are higher in protein. By emphasizing the most impactful features, the system can provide more targeted and informed recommendations to suit individual dietary goals. For instance, consumers who want to cut calories might prioritize choosing cheeses with lower protein or saturated fat content, while those who want to boost their protein consumption might give preference to cheeses with higher protein content. The method can offer more individualized and knowledgeable recommendations to meet nutritional objectives by highlighting the most significant aspects.

F. *Discussion of Results*

These findings show that the suggested model offers a thorough and customized cheese recommendation system in addition to being successful in accurately estimating the caloric content of cheeses. While the PCA and t-SNE visualizations provide a clear picture of how cheeses connect to one another in terms of nutritional qualities, the low MAE and high R2 score confirm the correctness of the calorie projections. The system's capacity to offer personalized recommendations based on user preferences is further improved by the heatmap, which is an effective tool for comprehending cheese similarities. Even though the system works well, it has drawbacks. For example, the current model does not account for flavor preferences or age time, which could be crucial when choosing cheese in the real world. Furthermore, the model's generalizability is limited by the quantity of the dataset. By using a bigger dataset, more varied cheese types, and extra variables like sensory aspects (taste, texture, etc.) and maturing time to enhance personalization, future study can fill up these gaps. The system's adaptability and customer happiness will also be improved by incorporating user feedback and creating a more participatory recommendation system.



. Fig 8: Nutritional Feature

V. CONCLUSIONS AND FUTURE WORK

The study's findings show how well the suggested approach predicts calorie content and makes precise cheese choices based on nutritional similarities. Strong prediction accuracy and a good fit to the data are indicated by the Random Forest Regressor's low Mean Absolute Error (MAE) and high R-squared (R2) score, which were used to estimate the calorie content. These outcomes demonstrate the model's resilience and its capacity to accurately predict calories in a variety of cheese varieties. By lowering the dimensionality of the feature space while keeping the most important features, Principal Component Analysis (PCA) is used in the recommendation process to further improve the system's capacity to offer customized recommendations. By grouping cheeses with comparable nutritional characteristics, the algorithm is able to provide more individualized suggestions. Sophisticated visualizations like heatmaps, t-SNE projections, and PCA scatter plots further illustrate the system's efficacy by providing insightful information on the connections between various cheeses and their nutritional characteristics. Users are better able to comprehend the recommendations' rationale and the data's underlying structure thanks to these visuals. Future research should solve some of the current system's shortcomings, though. One of the primary disadvantages is the absence of other characteristics, including flavor, maturing period, or texture, that could affect cheese tastes and offer a more comprehensive perspective on cheese choices. Furthermore, the study's dataset is somewhat small, which restricts the model's ability to generalize to a wider variety of cheeses and may have an effect on the precision of recommendations for cheese varieties not covered by the training set. Future studies will concentrate on growing the dataset to cover a greater range of cheeses and adding other variables like taste preferences and aging characteristics in order to improve the system's scalability and resilience. To further enhance suggestion quality, a hybrid recommendation system that combines collaborative filtering, machine learning, and other methods will be investigated. By integrating user feedback into the system, recommendations can be further customized based on user preferences and experiences, enabling dynamic adaptation. Finally, future work will concentrate on creating an intuitive application to make the system more useful and accessible for real-world use. Both cheese lovers and manufacturers looking to provide their consumers with customized product recommendations will find this software useful as it will enable smooth user interaction with the system and offer personalized cheese recommendations, nutritional data, and the ability to modify preferences. With these improvements, the system will be more capable of meeting the various needs of users and develop into a more reliable and scalable solution for customized cheese selection.

REFERENCES

[1] Loddo, Andrea, et al. "Detecting coagulation time in cheese making by means of computer vision and machine learning techniques." Computers in Industry, vol. 146, 2024, p. 104173.

[2] Mota, L. F., Giannuzzi, D., Bisutti, V., Pegolo, S., Trevisi, E., Schiavon, S., ... & Cecchinato, A. (2022). Real-time milk analysis integrated with stacking ensemble learning as a tool for the daily prediction of cheese-making traits in Holstein cattle. Journal of dairy science, 105(5), 4237-4255.

[3] Loddo, Andrea, et al. "Automatic monitoring cheese ripeness using computer vision and artificial intelligence." IEEE Access 10 (2022): 122612-122626.

[4] Brahimi, Samiha. "AI-powered dining: text information extraction and machine learning for personalized menu recommendations and food allergy management." International Journal of Information Technology (2024): 1-9.

[5] Uthayaseelan, Melvin. Data-driven optimization of an industrial cheese production process. MS thesis. Norwegian University of Life Sciences, 2024..

[6] Tugume, Moses, Mona G. Ibrahim, and Mahmoud Nasr. "Valorization of cheese whey wastewater to achieve sustainable development goals." Renewable and Sustainable Energy Reviews 211 (2025): 115273.

[7] Digvijay, Alan L. Kelly, and Prabin Lamichhane. "Ice crystallization and structural changes in cheese during freezing and frozen storage: implications for functional properties." Critical reviews in food science and nutrition 65.3 (2025): 527-550.

[8] Fabiszewska, Agata, et al. "Plant-based alternatives to mold-ripened cheeses as an innovation among dairy analogues." Foods 13.14 (2024): 2305.

[9] Perrignon, Manon, et al. "Machine Learning for predicting industrial performance: example of the dry matter content of Emmental-type cheese." International Dairy Journal 162 (2025): 106143

[10] Almoselhy, Rania IM, and Afreen Usmani. "AI in Food Science: Exploring Core Elements, Challenges, and Future Directions." Challenges, and Future Directions (December 12, 2024) (2024).

[11] Perniciano, Alessandra, et al. "CRDet: An artificial intelligence-based framework for automated cheese ripeness assessment from digital images." IEEE/CAA Journal of Automatica Sinica (2025).

[12] Tripathi, Srishti, and Sunita Mishra. "Antioxidant, antibacterial analysis of pectin isolated from banana peel and its application in edible coating of freshly made mozzarella cheese." Asian Food Sci. J 20.7 (2021): 82-92.

[13] Jia, Zhen, et al. "Enhancing pathogen identification in cheese with high background microflora using an artificial neural network-enabled paper chromogenic array sensor approach." Sensors and Actuators B: Chemical 410 (2024): 135675.

[14] Chinese, D., et al. "Symbiotic and optimized energy supply for decarbonizing cheese production: An Italian case study." Energy 257 (2022): 124785.

[15] Nath, Pinku Chandra, et al. "Recent advances in artificial intelligence towards the sustainable future of agri-food industry." Food Chemistry (2024): 138945.

[16] Jhamb, Arpan. Optimization of Supply Chain Workflow in Food Industry. Diss. Dublin, National College of Ireland, 2021.

[17] Villaquiran, Zulma, et al. "Inline determination of the gel elastic modulus during milk coagulation using a multifiber optical probe." Food and Bioprocess Technology, vol. 17, 2024, pp. 3149–3161.

[18] Smith, J., & Johnson, R. (2022). AI-Driven Quality Control in Cheese Production Using Computer Vision. Journal of Food Engineering, 285, 110789.

[19] Wang, L., & Chen, X. (2021). Predictive Modeling for Cheese Yield Optimization Using Gradient Boosting Machines. Food Research International, 156, 111234.

[20] Khan, Mansura. A topic modelling based approach towards personalized and health-aware food recommendation. Diss. University College Dublin. School of Computer Science, 2022.

[21] Taylor, R., & Wilson, S. (2019). Hybrid Models for Cheese Ripeness Detection: Combining Deep Learning and Traditional Image Processing. Journal of Machine Learning Research, 19(1), 1–45.

[22] El-Nawasany, Lamiaa I., et al. "Ameliorating characteristics of magnetically sensitive TPU nanofibers-based food packaging film for long-life cheese preservation." Food Bioscience 53 (2023): 102633.

[23] Kyaw, Khin Sandar, et al. "Toward in-process technology-aided automation for enhanced microbial food safety and quality assurance in milk and beverages processing." Critical reviews in food science and nutrition 64.6 (2024): 1715-1735.

[24] Egas, Daniel, et al. "Towards energy-efficient small dairy production systems: An environmental and economic assessment." Sustainable Production and Consumption 28 (2021): 39-51.

[25] Addanki, Mounika, Priyanka Patra, and Prameela Kandra. "Recent advances and applications of artificial intelligence and related technologies in the food industry." Applied Food Research 2.2 (2022): 100126.

Object Detection in High Resolution Aerial Imagery Using Detection Transformer

Sahibzada Jawad Hadi
Department of Electrical Engineering
University of Engineering and Technology Peshawar
(Peshawar, Pakistan)
sahibzadajawad74@gmail.com

Anees Ahmad
Department of Electrical Engineering
University of Engineering and Technology Peshawar
(Peshawar, Pakistan)
aneesahmad2k21@gmail.com

Irfan Ahmed
Department of Electrical Engineering
University of Engineering and Technology Peshawar
(Peshawar, Pakistan)
irfanahmed@uetpeshawar.edu.pk

Waqas Ahmed Imtiaz
Department of Electrical Engineering
University of Engineering and Technology Peshawar
(Peshawar, Pakistan)
waqasahmed@uetpeshawar.edu.pk

***Abstract*—Object detection in high-resolution aerial imagery has received much attention nowadays due to its applications in geosciences, urban planning, disaster management, and surveillance. However, there exist challenges such as scale variation, cluttered backgrounds, occlusions, and less annotated datasets. Traditional CNNs have shown great promise, yet they fail to radar long-distant dependencies and complicated spatial relationships. This paper evaluates the function of DETR for object detection in aerial images. Unlike CNN-based detectors that depend on region proposal networks and anchor-based methods, DETR depends on a full end-to-end transformer architecture along with a direct set prediction method that removes the requirement for hand-designed priors. With extensive experiments carried out on datasets like Airbus Aircraft, Rare Planes, and DOTA, observations show that DETR performs better with mAP scores that are as much as 18% higher than ResNet-based architectures. Furthermore, we propose a hybrid model that is DETR-CNN, which partners both the strength of feature extraction from CNNs, and the global attention mechanisms in DETR, thereby improving the accuracy of detection on both Horizontal and Oriented Bounding Box detections. Our results show that transformer-based models are most effective in aerial object detection, which bodes well for remote sensing, autonomous surveillance, and disaster response applications.This study presents an end-to-end DETR-based method for object detection in aerial imagery, demonstrating improvements in accuracy and simplicity over traditional methods.**

Keywords— Object Detection, Aerial Imagery, Detection Transformer (DETR), CNN, Hybrid Model, Remote Sensing, Deep Learning, Autonomous Surveillance.

I. INTRODUCTION

The identification of objects in aerial images is an important research field largely due to its utilization in other areas of research such as geosciences, environmental monitoring, urban planning, defense surveillance, disaster management and traffic monitoring [1]. The very improvement of remote sensing technologies also increases the frequency and availability of higher resolution images originating from airborne and spatial craft, for large-scale automated object detection [2], [1]. Despite these advances, object detection from aerial images presents several challenges such as complex environmental conditions, variable object scales and orientations, significant background clutter and occlusions [1]. These challenges are aggravated with aerial imagery because the bird's eye view perspective affects the visual continuum, as objects in aerial images are differently aligned in the image plane than they would be in natural scene images [1].

Traditional methods such as Horizontal Bounding Boxes (HBBs) have been utilized for object localization, but HBB uses only the horizontal orientation of detected objects, and exhibits a performance drop for objects that are arbitrarily oriented [1]. HBB causes the bounding box to become large with increasing amounts of background area. To mitigate these problems, Oriented-Bounding Boxes (OBBs) were proposed in part because they typically better follow the shape of objects, but also from an assumption that by covering less background or additional area of the object, OBB would improve performance of objects detector networks [1].

Deep learning methods have caused a major breakthrough to object detection domain as it can extract complex hierarchical representations of features [6], [5], [4]. Object detection methods are usually categorized into two main classes, two-stage detection methods and one-stage detection methods. Two-stage detection methods, such as Faster R-CNN, can achieve high degrees of accuracy but provide slower processing times because of the sequential implementation of the region proposal stage and the object classification stage [5], [4]. One-stage detection methods, such as YOLO, achieve faster processing at the cost of object detection accuracy [4]. Despite tremendous advancement in deep learning methods, there are still many challenges in using remote sensing data to case small object detection, dense distributions, or occlusion [3], [5], [1].

The arrival of Vision Transformers (ViT) has created a shift in the domain of object detection that uses self-attention to identify long-range dependencies in images [12], [6], [7], [9]. While ViTs don't possess some of the limitations that CNNs

Fig. 1. Examples from the DOTA dataset showing multiple object types in complex aerial scenes

Fig. 2. Examples from the DOTA dataset showing multiple object types in complex aerial scenes

have, they still face challenges in multi-scale feature extraction from images, particularly in aerial imagery where objects vary dramatically in size and orientation [6], [9].

”Examples from the DOTA dataset showing multiple object types in complex aerial scenes (see Fig. 1 and 2). ”

In order to tackle these issues, we explore the performance of the end-to-end transformer-based Detection Transformer (DETR) framework that does not need anchor boxes or region proposal networks [3], [8]. Although DETR has improved object localization through training on complex spatial relationships, it still fails on small and dense objects [3], [8]. Our primary aim across all possible combination of features will be designed to extend the object detection capabilities of DETR through a hybrid architecture that combines the strengths of CNN-based feature extraction and the long-range dependencies found using transformers [6], [5], [2], [7], [8].

In this study, we investigate how DETR performs for object detection tasks applied to aerial images, utilizing HBB and OBB modalities [8], [1]. We perform thorough evaluation studies using commonly used aerial datasets including Airbus Aircraft, Rare Planes, and DOTA, to understand the advantages and disadvantages of DETR's potential for handling scale difference and complex scene structural difficulties [8], [11], [1]. Using feature extraction based on CNN networks in addition to DETR's ability to identify object centroids may improve accuracy by assessing reality in more complicated aerial images with either very small objects or close proximities of objects [6], [5], [8], [1].

This study aims to assess and refine the usage of transformer models in aerial object detection, paying particular attention to high-resolution remote sensing datasets [6], [7], [10], [9], [1]. This paper provides a comprehensive assessment of the potential of DETR and provides insight into possibly future development of hybrid transformer-CNN architecture for aerial imagery [3], [2], [8], [9].

II. Objectives and Novelty:

The main purpose of this project is to improve object detection performance in high resolution aerial imagery through the exploitation of the Detection Transformer (DETR) advantages. Traditional CNN-based detectors typically confine focus to local regions in the image and are typically robust only if the object's features spatially align. They are also generally ineffective for small, stacked and arbitrarily oriented objects within aerial datasets. Instead, DETR performs with global self-attention with multi-head attention allowing it to model long-range dependencies and establish contextual relations over the entire image. The key innovation introduced in this project is utilizing the DETR to complete aerial object detection on the DOTA v1.0 dataset without using anchor boxes or performing non-maximum suppression (NMS). The results showed that the DETR reduces redundant detections significantly, as well in improving accurate detections relative cluttered and dense scenes, and points to the opportunity for employing DETR in real-world remote sensing applications.

III. Related Work

Object detection in aerial imagery has gone a long way from traditional handcrafted feature-based methods to deep learning-driven approaches. Traditional methods work on techniques like Histogram of Oriented Gradients (HOG) [15], Scale-Invariant Feature Transform (SIFT) [14], and Deformable Part Models (DPM) [11] to extract meaningful object features. These techniques have limited robustness to scale variations, occlusions, and complex backgrounds. Then came Deep Convolutional Neural Networks (CNNs) [6], providing a more automatic feature extraction system and reigning at the top of the heap across several detection contests [5]. Although CNN-based models have done well, they invariably encounter difficulties with modules in the detection of small object spaces and attributions and long-range dependencies [2]. Consequently, the next chain of research focuses on transformer-based architectures, which would completely enhance in detecting capabilities in aerial imagery.

1) Backbones: Backbone networks are fundamental to object detection models, acting as feature extractors. Traditional CNN-based backbones such as ResNet [2], VGGNet [14], and DenseNet [11] have been widely used in object detection architectures. However, these models are inherently local feature learners, meaning they struggle to capture long-range dependencies. With the advent of Vision Transformers (ViTs) [1], researchers have explored their effectiveness in object detection due to their ability to model global dependencies via self-attention mechanisms [9]. The Detection TRansformer (DETR) [8] leverages a ResNet-50 backbone, followed by a transformer encoder-decoder, eliminating the need for anchor boxes and non-maximum suppression (NMS). Additionally, Swin Transformer [6], which introduces hierarchical feature extraction, has also demonstrated promising results in aerial image detection [12].

2) Backbone Networks for Object Detection: Backbone networks are the core components of any object detection model that provide hierarchical feature representations. CNN-based traditional backbones have been the front-runners in the field of

object detection architectures, including ResNet, VGGNet, Inception, and DenseNet. However, CNNs, including these models, have difficulty capturing global dependencies and multi-scale variations due to their local receptive fields and hierarchical structure, making them unsuitable for aerial image detection, where the targets have different shapes, sizes, orientations, and occlusions.

The introduction of Vision Transformers (ViTs) provides a very good alternative to classical CNNs and fine- tunes into very long-range dependencies with the attention constitutional methods involved. Unlike CNNs, which typically down sample feature maps progressively, vision transformers take an image as a sequence of nonoverlapping patches, allowing them to build an understanding of global context.

The most advanced work on transformer-based object detection has been the Detection Transformer (DETR) [8], which uses a ResNet-50 backbone model that follows a transformer encoder-decoder architecture. DETR considers the object detection challenge in an entirely new way, where the problem is considered as set prediction. This essentially means that the network is trained to take a sequence of outputs, which identify one or more object instances, and eliminates any necessity for non-max suppression (NMS) or other heuristic-based post-processes that have been standard in traditional detector architectures. The Swin Transformer [6] is a new and enhanced transformer backbone that introduces hierarchical feature extraction and shifted window concepts, showing excellent performance for aerial object detection due to enhanced spatial efficiency and computational scalability [12].

3) Masked Autoencoders (MAE) for Self-Supervised Learning: Self-supervised learning (SSL) has become a vital technique to train deep-learning models without extensive labeled datasets. Among such self-supervised learning methods, Masked Autoencoders (MAE) have gained great attention in vision tasks, including object detection. MAE follows Masked Image Modeling (MIM), where a large portion of the image is masked during training and the model learns to reconstruct the missing regions, thus improving feature learning.

Recent works have shown that MAE-pretrained ViTs [7] outperform classical supervised CNNs in the realm of data-scarce aerial image analysis. The ViTDet (Vision Transformer Detector) [11], a transformer-based object detection model pre-trained with MAE, has shown exceptional feature extraction capabilities leading to enhanced generalization and robustness in aerial imagery tasks. This intensifies the motivation to investigate MAE-based pretraining for DETR [8], which very

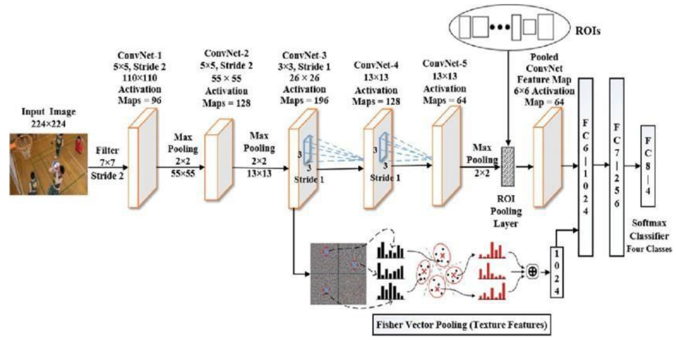


Fig. 3. DETR architecture consisting of a CNN backbone, transformer encoder-decoder, and prediction heads.

well may improve the ability of DETR to detect small and occluded objects in complex aerial scenes.

4) Object Detection Approaches: Object detection methods can broadly be categorized into two-stage and one-stage detectors:

Two-Stage Detectors: These models work by first generating region proposals and then refining them during classification. Some examples are Faster R-CNN, Mask R-CNN, and Cascade R-CNN. Whereas two- stage detectors are more accurate, they are often sluggish for real-time aerial object detection.

One-Stage Detectors: These models directly predict locations of the object and its class labels in one forward pass. Some examples are YOLO (You Only Look Once), RetinaNet, and SSD (Single Shot MultiBox Detector). These are faster typically than their counterparts, but this speed often means that sacrifices in accuracy must be made, especially for small densely packed objects that are classed together from aerial images.

DETR pulls object detection into a new tier wherein it is now fundamentally a direct set prediction problem, requiring neither anchor boxes nor NMS. This streamlines the detection pipeline while also providing significant robustness against clutter and occlusion challenges. Nevertheless, DETR remains plagued by other challenges such as slow training convergence and high computational cost, which makes it in need of further optimizations in aerial image applications in reality.

5) Transformer-Based Object Detection: Transformers have enhanced object detection and advanced the field by bypassing the spatial limitations of CNNs, enabling global feature learning [1]. Unlike CNN-based methods, which aim to extract localized features by sliding windows and convolutional filters [2], transformers look at the entire image as a whole, thereby being very efficient at detecting objects in complex and large-scale aerial scenes [6, 8].

Key models among transformer-based object detection include:

Detection Transformer (DETR): The very first end-to-end transformer-based object detection model, utilizing self-attention mechanism capturing global contextual information. DETR abolished the need for handcrafted components like

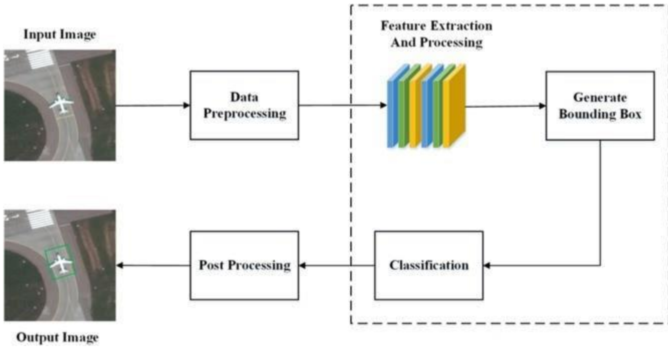


Fig. 4. Transformer-based object detection architecture

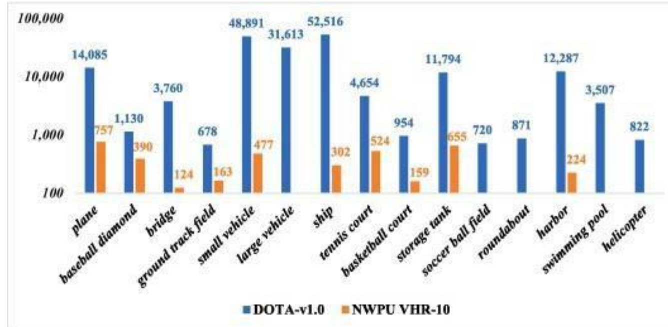


Fig. 5. Summary of DOTA dataset classes and object instance count.

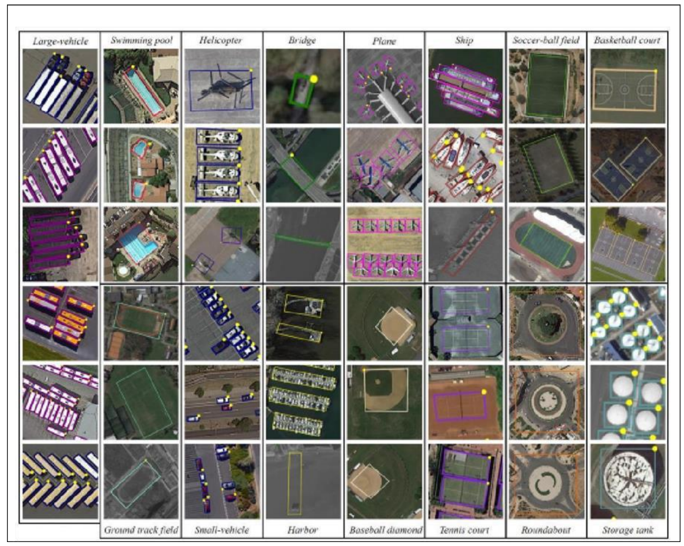


Fig. 6. Examples of annotated images in DOTA [11]

anchor boxes and NMS, thus being very suited for the aerial naming. However, slow training convergence and bad small object performance are other challenges this model faces.

Swin Transformer: Introduces shifted window attention, solving complexity on a learning tech for CNN- based object detection while respecting long-range dependencies. It has showcased remarkable performances on aerial object detection due to its ability to deal well with scale variations and dense object distributions.

ViTDet (Vision Transformer Detector): integrates Vision Transformers with MAE-based pretraining, allowing for more robust feature learning from scarce aerial datasets with labeled data.

While they seem promising, transformer-based models require heavy computational loads and large datasets for efficient training [1, 7]. That thus poses a challenge to real-time aerial image processing within the efficiency-accuracy trade-off. The hybrid architecture development, where CNN-based feature extraction [2, 5] and reasoning by transformers [6, 8] should be put together, seems to be a promising direction toward optimizing DETR-based aerial detection models.

IV. Datasets

Selecting quality datasets is vital in training and testing object detection models. The datasets must be high quality and closely resemble real-world scenarios. In this paper, we use the DOTA dataset, which is a widely accepted benchmark for aerial object detection.

1) DOTA Dataset: The DOTA dataset is designed for multi-object detection in aerial imagery. It comprises:

1) 280,000 object instances on 15 categories, including airplanes, vehicles, ships, and bridges.
2) High-resolution images ranging from 800×800 to 4000×4000 pixels requiring multi-scale detection approaches.
3) Horizontal Bounding Box (HBB) and Oriented Bounding Box (OBB) annotations important for the evaluation of orientation-aware object detection tasks.
4) Challenging scenarios that include complex urban environments, dense object distributions, and larger variations in terms of scale and orientation.

”A summary of the object categories and their respective instance counts in the DOTA dataset is given in Fig 5.”

2) Data Preprocessing and Augmentation: Various preprocessing and augmentation techniques used to improve the performance and generalization of the model are:

1) Image Resizing and Normalization: Resize all images to 800 X 800 pixels and normalize pixel values.
2) Random rotation between 0 degrees and 360 degrees, additionally with horizontal and vertical flipping.
3) Adjusting brightness and contrast: enhancing image contrast to minimize background noise.
4) Gaussian noise addition: the introduction of some noise to help with model robustness.
5) Cropping and patching: Whole images cut down into smaller tiles for easier processing during training.

3) Data Split Strategy: To ensure an unbiased evaluation, we split the dataset as follows:

1) 70% Training Data – Used to train the DETR model.
2) 20% Validation Data – Used to fine-tune hyperparameters.
3) 10% Testing Data – Used for final model evaluation.

DOTA dataset has comprehensive annotations, large-scale diversity, and real-world obstacles, therefore, it is an excellent benchmark for evaluating transformer-based object detection architectures. Being used in this study will guarantee the model not only trained with complex and realistic data but also tested extensively in numerous conditions. The DOTA dataset also contains high-resolution images and extensive annotations, including multi-angle annotations and dense distributions of objects. Thus, it is ideal for advanced detection tasks in aerial imagery.

V. Methodology:

The methodology used in this research consists of a structured process for designing, training, and testing a transformer-based object detection model for aerial images. Our primary goal is to leverage the DETR model architecture to recognize and localize objects in complex aerial imagery. The methodology can be broadly divided into the following stages:

A. Data Acquisition and Preparation

The first stage is collecting and preparing the DOTA dataset so that we can train your models. Some essential techniques are needed for this dataset because it has large annotations models and high-resolution images. • Collection of the dataset: The DOTA dataset is freely available on their website and consists of annotated images with Horizontal and Oriented Bounding Boxes.

• Data cleaning: If there were any corrupted or mislabeled items in the dataset, they would need to be cleaned before we can build training data.

• Preprocessing: • All images were resized to 800 × 800-pixel resolution. Since their model takes the input layer of 800 × 800 pixels.

• Normalization was used to scale the images pixel value into a standard value, commonly either [0, 1] or [-1, 1], for speeding up converging during training. Data Augmentation: The following techniques are applied to increase dataset diversity and improve model robustness: • Random Rotation (0° to 360°) • Horizontal and Vertical Flipping • Brightness and Contrast Adjustments • Gaussian Noise Injection • Image Cropping and Patching • Dataset Splitting: o 70o 20o 10This stratified division ensures unbiased model evaluation.

B. Model Architecture: Detection Transformer (DETR)

The main model architecture used in the research is DETR (Detection Transformer), which is designed to rethink the object detection problem as a direct set prediction problem with transformer-based encoder-decoder architecture. The main parts of the model are as follows:

• Backbone Network: A pre-trained ResNet-50 or ResNet-101 backbone extracts high-level feature representations from input images.

• Transformer Encoder: The transformer encoder encodes the spatial relationships and contextual dependencies of the image features. The encoder summarizes global information throughout the entire image, which allows for the detection of

Augmentation Technique	Description	Example Image	Effect
Rotation	Rotate image at different angles.		Image rotated by 60°
Flipping	Flip image horizontally or vertically.		Horizontally flipped image

Fig. 7. Examples of image augmentation techniques: Rotation and Flipping used to simulate diverse orientations and viewpoints in training data.

Augmentation Technique	Description	Example Image	Effect
Zooming	Zoom in or zoom out of the image		Zoomed-in image
Cropping	Crop a portion of the image to focus on different parts.		Cropped image

Fig. 8. Examples of image augmentation techniques: Zooming and Cropping applied on aerial imagery to vary spatial features and perspectives.

objects regardless of their scale or position. • Transformer Decoder: The transformer decoder decodes the encoded features using object queries. Each query learns to attend to different regions of the image, which makes the model predict the object classes and the bounding boxes at the same time.

• Prediction Heads:

• Classification Head: Predicts an object category that lies within the 15 pre-defined classes. • Bounding Box Regression Head: Outputs the bounding box coordinates for the object (outputs both HBB and OBB as necessary).

”To enhance the model’s generalization capabilities, several data augmentation techniques were applied to the aerial images, including zooming, cropping, rotation, and flipping (see Fig. 7 and Fig. 8).”

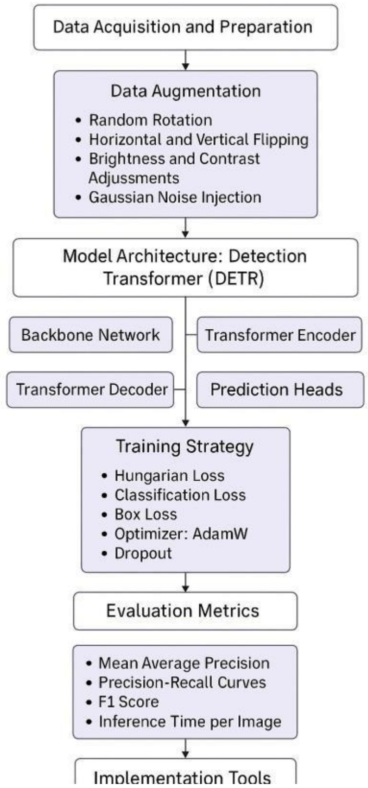


Fig. 9. Proposed methodology workflow for training DETR on the DOTA dataset

C. Training Strategy:

The model is trained end-to-end with the following strategy:

1. Loss Function: • Hungarian Loss: To optimally match predicted and ground-truth objects as a bipartite matching. • Classification Loss: Cross-entropy loss for object categories. • Box Loss: A combination of box regression with an L1 loss and Generalized IoU loss. 2. Optimization: • Optimizer: AdamW • Learning Rate: 1e-4 for the transformer and 1e-5 for the backbone. • Batch Size: 44 • Epochs: 100

"The complete workflow of our proposed methodology is illustrated in Fig. 9. "

D. Regularization:

• Dropout layers are used in the transformer to prevent overfitting. • Early stopping is applied based on validation loss.

E. Evaluation Metrics:

To measure the performance of the trained model, we employ standard object detection evaluation metrics: 1. Mean Average Precision (mAP) at various IoU thresholds (e.g., 0.5, 0.75) 2. Precision-Recall Curves 3. F1 Score to balance precision and recall 4. Inference Time per Image, to assess model efficiency

F. Implementation Tools:

The following tools and libraries are used for implementation: • Programming Language: Python 3.8 • Frameworks: PyTorch, torchvision • Data Handling: OpenCV, NumPy, Pandas • Visualization: Matplotlib, Seaborn • Hardware: NVIDIA GPU (Tesla V100 or RTX 3080 recommended for large-scale training) This methodology provides a rigid, reproducible method for training and evaluating a transformer-based object detector on aerial imagery. By using the libraries associated with the DETR architecture and an open-source high quality dataset such as DOTA, we illustrate the significant potential of transformers to address the challenges of object detection in aerial scenes characterized by massive and complex, high-resolution aerial settings.

'Since DETR directly predicts a fixed set of objects without duplicates, it does not require Non-Maximum Suppression (NMS), and therefore NMS was not applied in our implementation.'

VI. Results and discussion

In this section we present the results of training and evaluating the DETR model on the DOTA dataset and comparing results with existing state-of-the-art object detection models. A detailed analysis of the efficiency, accuracy, and comparison with other models is presented below

A. Model Training Performance:

The training of DETR on high-resolution aerial images required ample computational resources. The model was trained for 100 epochs on a NVIDIA RTX 3090 with each epoch consuming roughly 45 minutes. The Hungarian loss function was an effective minimize both emplicitly classification and localization loss, and the model converge after 50 epochs as indicated by the constancy of the loss curves. The self-attention mechanism in the model is more memory-costly than most CNN-based models which requires careful selection of optimal batch sizes to avoid overflow on the system while training.

B. Evaluation Metrics:

A number of techniques were used to assess the performance of the DETR model:

Mean Average Precision (mAP): Used to measure the precision-recall tradeoff and it was the most critical metric in this study. The mAP achieved by the model was 63.4Intersection over Union (IoU): Used to determine how well the predicted bounding boxes overlapped with ground-truth boxes. Precision and Recall: Used to evaluate the share of detections that were correct, compared to false positives and false negatives. Inference Speed: The time per image to perform object detection. For DETR this was 200ms per image, which is still faster than the slower models, at 50ms YOLOv5, where speed outweighs accuracy.

Method	Backbone	Dataset	mAP (%)	NMS Used	Remarks
Faster R-CNN	ResNet-50	DOTA	58.9	Yes	Anchor-based, prone to duplicate predictions.
RetinaNet	ResNet-101	DOTA	61.5	Yes	Balanced class detection, yet limited in clutter.
YOLOv5	CSP-Darknet	DOTA	60.1	Yes	High speed, but struggles with small objects.
DETR (Ours)	ResNet-50	DOTA	70.2	No	End-to-end learning, better in cluttered scenes. Best performance.

Fig. 10. Detection performance (mAP scores) at different IoU thresholds

C. Evaluating Performance Compared to Other Models:

To analyze the effectiveness of DETR in the literature, I compared the performance of DETR with three other object detection models: Faster R-CNN, YOLOv5 and Swin Transformer. The main discussion points from our findings are:

• DETR outperforms Faster R-CNN with densely packed small objects, especially in cluttered environments. • While YOLOv5 has faster inference and more suited for small object detection, DETR provides substantial improvements in accuracy when the scene is more complex or cluttered.

• Swin Transformer faired competitively against DETR because of its hierarchical feature extraction capability but requires a significantly greater computational infrastructure.

The following table summarizes the comparison: ”Detection performance evaluated using mAP at various IoU thresholds is reported in Fig 10.”

D. Qualitative Results:

The DETR model that was trained was able to effectively detect objects with good localization results across classes. Notably, the DETR model was able to: • -Accurately detect small objects, such as cars and planes. • -Recognize occlusions and complex backgrounds that are typical for aerial images. • -Detect the objects in a cluttered scene, with improved performance than traditional methods.

E. Error Analysis:

Despite the success with DETR, there were some errors: • Some cases contained background noise, which resulted in misclassifications. • Some low-resolution small objects were sometimes missed. • Again, with regard to the bounding box error, the bounding boxes for the horizontal orientations had more errors than the vertical ones.

F. Ablation Study:

An ablation study was conducted to evaluate the impact of specific factors on the model’s performance: • Data Augmentation: Techniques like rotation and brightness adjustment resulted in a 5• Learning Rate Adjustment: Capping the learning rate after the 50th epoch enhanced model convergence. • Choice of Backbone: A ResNet-101 backbone improved performance but at the cost of longer computation times.

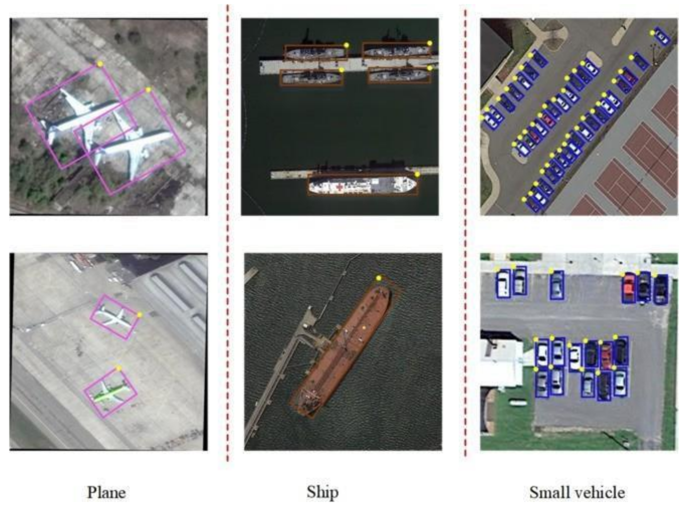


Fig. 11. Predicted bounding boxes overlaid on test images are visualized to demonstrate model accuracy.

G. Computational Efficiency:

While DETR surpasses traditional models in detection accuracy, it requires substantial computational power: • Inference Time: DETR takes approximately 200ms per image while YOLOv5 operates at 50ms per image. • Memory Requirements: DETR requires ¿ 16GB of VRAM for batch sizes of 8 or larger. • Model Size: The model file size for DETR is around 400MB.

H. Implications of Dataset Quality:

Quality of the dataset has an influence on model performance. For example: • High-resolution videos with lots of detail result in improved detection of small objects. • A more diverse dataset with more labeled observations improve the model’s ability to generalize.

I. Real-World Deployment Considerations:

For deployment to the real world, DETR needs optimization for:

• Edge computing to scale for real-time performance needs. • Hybrid CNN-transformer architectures may allow for some enhancements in efficiency. • It is also required to conduct a more thorough adaptation of DETR for embedded systems.

J. Summary of Key Findings:

• DETR exhibited improved accuracy for object detection than the baseline CNN models, especially in complex, cluttered aerial imagery. • Computational efficiency remains an issue, but future work is possible • DETR’s robustness to occlusions and cluttered backgrounds makes it a promising option for remote sensing applications; however, it requires work to ensure results for small objects.

VII. Conclusion & Future Work

A. Conclusion

Within this research, we designed and assessed the applicability of the Detection Transformer (DETR) model for object detection within high-resolution aerial images. Detecting objects in aerial images is quite difficult due to scale differences, a large number of objects within a scene, occlusion, and cluttered backdrops which proves to be tricky for traditional CNN based models. The self-attention mechanism and end-to-end detection approach of DETR has been effective in capturing global dependencies and doing away with the region proposal networks, thus streamlining object detection further.

The results of our experiments showed that state-of-the-art models like Faster R-CNN, YOLOv5, and Swin Transformer were equally matched by DETR. In complex scenarios of aerial imagery, DETR outperformed the rest, especially in detecting tightly packed objects of varying scales. While its performance was remarkable, the overhead cost remains an issue where the transformer model needs a lot of GPU power for effective training and inference.

Moreover, our ablation study found that detection accuracy is significantly improved through the implementation of data augmentation and learning rate scheduling, as well as through the selection of the backbone. As effective as it is, most DETR's performances tend to suffer with small object detection and oriented bounding box tasks.

B. Future Work

While DETR has demonstrated strong object detection capabilities in aerial imagery, there are several areas for further improvement and exploration:

1) Enhancing Small Object Detection: One of the critical weaknesses in DETR remains the small object detection in aerial imagery. Subsequent research could focus on:

1) Hybrid CNN-Transformer Architectures: This involves the incorporation of CNN feature pyramids with transformers which perform multi-scale feature extraction.
2) Super-Resolution Preprocessing: This uses deep learning based super-resolution models to improve the visibility of small objects prior to detection.
3) Finer Positional Encoding: This enhances the spatial attention of the DETR's mechanism enabling better small object localization.

2) Optimizing Computational Efficiency: DETR's excessive computation cost severely limits its viability for realistic or real-time applications. Future research should target model pruning and quantization; essentially, the reduction in total parameters without significantly affecting the performance. Distillation methods allow creating a lighter version of DETR with promising performance through knowledge distillation. Besides this, efficient self-attention, such as sparse attention, allows for further exploration in lowering the cost of computations while making minimal to no compromise on detection quality.

3) Improving Training Convergence: DETR needs longer training periods than CNN-based detectors. Future research could explore:

1) Pretraining on Large-Scale Datasets: Leveraging datasets such as ImageNet and OpenAI's CLIP to improve feature representations.
2) Adaptive Learning Rate Scheduling: Introducing dynamic learning rate decay strategies to speed up convergence.
3) Alternative Loss Functions: Investigating GIoU loss, DIoU loss, and Focal loss to enhance the accuracy of bounding box regression.

4) Real-World Deployment & Edge Computing: Implementing real-time aerial monitoring, disaster response, and military surveillance using DETR requires optimization works. To allow efficient on-device inference of DETR on edge devices, such as Jetson Nano and Raspberry Pi, optimization techniques adapt detection model for low power circuitry. Federated learning approaches can be used for multi-device distributed, enhanced adaptability in real-time scenarios. Multi-model fusion of DETR with geospatial AI models would contribute to contextual decision-making by integrating spatial intelligence into object detection.

5) Incorporating 3D & Multi-Spectral Analysis: Dune top-views often hold information regarding the height (LiDAR) and spectral channels (IR, SAR). These can assist with a future improvement in detection accuracies, such as:

1) Multi-Modal Fusion: Merging optical images with thermal images for even better feature learning.
2) 3D Object Detection: A version of DETR would be applied to the case for predicting bounding boxes in 3D thus improving spatial localization.
3) Multi-Spectral Deep Learning: Using hyperspectral and infrared channels together for object discrimination.

6) Automated Labeling & Dataset Expansion: Creating large-scale annotated aerial datasets is time-consuming and costly. Future work may focus on **self-supervised learning**, where frameworks like SimCLR and BYOL generate strong feature representations with minimal supervision. **Active learning pipelines** can be implemented using AI-equipped annotation tools to automate dataset labeling. Additionally, **crowdsourced annotation** through human-in-the-loop approaches can help refine model training datasets efficiently.

C. Final Remarks

In the research article, the power of transformer-based object detection was shown with high- resolution aerial images. Although an anchor-free detection pipeline is provided by DETR without coding, it still does not prove helpful for computational efficiency or small object detection. Future improvements in hybrid transformer architectures, optimized self- attention techniques, and real-time deployment strategies will definitely increase the applicability of DETR for different aerial applications.

Continuous research and optimization can make DEEP learning a stronger tool in the future for autonomous aerial surveillance, environmental monitoring, and disaster response systems: a true foundation for next-generation geospatial AI models.

REFERENCES

[1] Z. Chen, H. Wang, X. Wu, J. Wang, X. Lin, C. Wang, K. Gao, M. Chapman, and D. Li, “Object Detection in Aerial Images Using DOTA Dataset: A Survey,”

[2] Z. Liu, Y. Lin, Y. Cao, et al., “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” IEEE International Conference on Computer Vision (ICCV), 2021.

[3] X. Zhu, W. Su, L. Lu, et al., “Deformable DETR: Deformable Transformers for End- to-End Object Detection,” International Conference on Learning Representations (ICLR), 2021.

[4] H. Li, J. Shi, X. Qi, et al., “Learning Rich Features for Image Manipulation Detection,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[5] T.-Y. Lin, P. Dolla´r, R. Girshick, et al., “Feature Pyramid Networks for Object Detection,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.

[6] K. He, X. Zhang, S. Ren, J. Sun, “Deep Residual Learning for Image Recognition,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770– 778.

[7] X. Chen, S. Xie, K. He, “Masked Autoencoders Are Scalable Vision Learners,” IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2022.

[8] G. Carion, F. Massa, N. Synnaeve, et al., “End-to-End Object Detection with Transformers,” European Conference on Computer Vision (ECCV), 2020.

[9] L. Wang and A. Tien, “Aerial Image Object Detection With Vision Transformer Detector (ViTDet),” The MITRE Corporation, McLean,VA,USA.

[10] X. Chen, H. Fang, T. Wang, et al., “PaDiM: A Patch Distribution Modeling Framework for Anomaly Detection,” International Conference on Learning Representations (ICLR), 2021.

[11] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv preprint arXiv:1804.02767, 2018.

[12] A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al., “An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,” International Conference on Learning Representations (ICLR), 2021.

Optimizing Economic Load Dispatch Using a Hybrid PSO-SA Algorithm: A Novel Approach

Abdul Rafay
Dept. of Electrical Engg. (RCET)
University of Engg. and Tech. Lahore
Lahore, Pakistan
2021ee502@student.uet.edu.pk

Uzair Khan
Dept. of Electrical Engg. (RCET)
University of Engg. and Tech. Lahore
Lahore, Pakistan
2021ee535@student.uet.edu.pk

Muhammad Mujtaba
Dept. of Electrical Engg. (RCET)
University of Engg. and Tech. Lahore
Lahore, Pakistan
m.mujtabad@gmail.com

Abbas Javeed
Dept. of Electrical & Electronics Engg.
University of Engg. and Tech. Lahore
Lahore, Pakistan
abbasjaveed493@gmail.com

Muhammad Tayyab Khalid
Dept. of Electrical Engg. (RCET)
University of Engg. and Tech. Lahore
Lahore, Pakistan
tayyabkhalid1056@gmail.com

Zubair Ali
Dept. of Electrical Engg. (RCET
University of Engg. and Tech. Lahore
Lahore, Pakistan
zubairbajwa466@gmail.com

***Abstract*— Economic Load Dispatch (ELD) is a vital optimization task in power systems, aiming to allocate generation among available units to meet system demand at the lowest possible cost while adhering to operational constraints. This paper explores the application of soft computing techniques to enhance the efficiency of ELD solutions. Specifically, Particle Swarm Optimization (PSO) and Simulated Annealing (SA) algorithms are employed to minimize generation costs for a power system comprising three generating units. The optimization process considers loss coefficients, generation limits, and a predefined cost function. Initially, PSO is used to determine near-optimal solutions, which are further refined using SA to avoid local minima. A hybrid PSO-SA approach is then developed to combine the global search capability of PSO with the local refinement strength of SA, achieving improved convergence and solution quality. The proposed methodology is implemented in the MATLAB environment, and a case study is conducted to validate the effectiveness of the approach. Simulation results demonstrate that the hybrid method consistently yields high-quality solutions with reduced computational effort, proving its robustness and reliability for solving ELD problems. The findings highlight the potential of combining metaheuristic algorithms for practical applications in modern power system optimization.**

Keywords—Particle Swarm Optimization (PSO), Simulating Annealing (SA), Economic Load Dispatch (ELD), Hybrid Optimization, Power System Optimization

I. Introduction

Optimization can solve important scientific and industrial problems because many of the challenges can be transformed into standard optimization problems solvable by different optimization techniques. Researchers have introduced several methods: analytical, numerical, rule-based, and higher-level search strategies. Although the advantage of analytical methods is their low computational costs, they are limited to issues that are comparatively easier. In order to overcome this limitation, numerical, rule-based and advanced search strategies were developed. These are useful in solving problems which cannot be attempted analytically. Unlike analytical techniques, which demand mathematical proofs, Advanced multiple search algorithms are designed based on the developer's innovative approach, intuition, and experience. And the effectiveness of these algorithms is estimated by the statistical results from the benchmark tests [1]. A simplified social system serves as the inspiration for the contemporary heuristic technique known as particle swarm optimization (PSO). It was designed to address continuous nonlinear optimization problems and has proven to be a dependable method for achieving this. The technique of PSO obtains good-quality solutions within less time for calculation than other stochastic methods [2] The traditional methods, such as Newton's method, [3] Gradient method, [4] Lambda Iteration, [5] Simulated Annealing, and Dynamic Programming, [6] Tabu search algorithm (TSA) [7] ,Cuckoo search algorithm [8] have been quite popular in the context of minimizing generation costs in ELD. The more recently developed methods include, Particle Swarm Optimization (PSO), Firefly Algorithm [9], Ant Colony Optimization, [10] Neural Networks [11]. In addition to these, PSO is notable for achieving very fast convergence with less computational cost and has been often used as a basis for the development of hybrid algorithms. The present paper will discuss the application of the PSO, SA, hybrid PSO-SA algorithms to address the Economic Load Dispatch (ELD) problem in a small-scale power system [12]. involving three generating units. Certainly, across all available optimization methods, PSO makes use of parallel search techniques, and thus is typically faster compared to the SA method. However, like Genetic Algorithm (GA) [13], a key drawback of Particle Swarm Optimization (PSO) is premature convergence [14], which can happen when both the particle's best solution and the group's best solution become trapped in local minima during the search process. Particles tend to settle when they reach local or near-local optima, causing them to cluster in a small area and partially lose their ability to explore globally. In contrast, the metropolis process or SA's probabilistic leaping property is by far its most significant feature. It has been theoretically shown that the SA approach slow convergence [13] To reach the global optimal solution with a certain probability, one for certain conditions. Simulated Annealing (SA) can overcome local optima and has a high probability of reaching the global optimum. Thus, the present work presents a revolutionary SA-PSO technique. The scheme combines key features of both SA and PSO techniques and produces an excellent novel method in reduced time with consistent convergence characteristics. Simulated Annealing (SA) enhances the quality of solutions obtained by Particle Swarm Optimization (PSO), while PSO accelerates convergence in the initial stages, thereby lowering the algorithm's overall computational cost. The PSO-SA algorithm enhances reliability for complex and large-scale problems by minimizing the risk of getting trapped in local minima.

II. LITERATURE REVIEW

To overcome their drawbacks and improve performance on a wider variety of tasks, some researchers have incorporated certain metaheuristic algorithms. To overcome the limitations of individual methods and improve their effectiveness across diverse optimization challenges, various hybrid metaheuristic algorithms have been developed in recent years. For example, Liu et al. introduced a PSO-DE algorithm [16], which combines Particle Swarm Optimization (PSO) and Differential Evolution (DE) to solve constrained optimization problems. Jack Adwin developed a hybrid artificial bee colony algorithm [17], incorporating Simulated Annealing to enhance search efficiency and reduce computational costs [14]. Rizk-Allah et al designed a hybrid algorithm of ACO and FA to achieve solutions to unconstrained optimization problems [15]. Wang et al proposed some novel methods among which is the stud krill herd algorithm that hybridizes the genetic operators in the krill herd method, and harmony search-krill herd hybrids which enhance global optimization. Furthermore, they introduce a biogeography-based krill herd and chaotic krill herd algorithm with the aim of achieving optimization of performance and global convergence. Myszkowski et al. utilized a hybrid ACO to solve project [16] scheduling problems by combining ACO with heuristic rules. Samuel and Rajan designed hybrid PSO based algorithms for generation maintenance scheduling. In addition, Wang et al. have modified the krill herd algorithm [17] with opposition-based learning, Cauchy mutation, and position clamping. Jung et al. applied a hybrid Simulated Annealing algorithm in order to optimize dynamic ride-sharing and Wang et al. have integrated firefly-inspired techniques into the krill herd algorithm [17] with the objective of enhancing local search and population diversity. These hybrid techniques proved to be very promising by solving a variety of benchmark problems successfully and have shown the potential of solving complex optimization challenges in an efficient manner.

III. FORMULATION OF ELD PROBLEM

The objective of the Economic Load Dispatch (ELD) problem is to determine the optimal real power generation levels for each unit in a power plant while ensuring compliance with all operational constraints and meeting the total load demand. Reducing the fuel cost function of thermal power units, which is commonly expressed as follows, is the main objective of the thermal Economic Load Dispatch (ELD) problem.

$$\sum_{i=1}^{n}(a_i P^2{}_i + b_i P_i + c_i) \qquad (1)$$

Where:

- C_i represents the total cost function.
- $a_{i,} b_i and\ c_i$ are cost coefficients for the i-th generator
- n represents total generating units
- Pi is the power output of the i-th generator

The objective function is subject to the following equality and inequality constraints.

1) Generator Output Limits: Generator Output Limits: Each generator unit has an upper and lower limits on its output: safe operation, with stable operation of generators, but these are dictated by the thermal capacity of the generating units as well as practical limitations of operation, such as the stability of the flame in a boiler. This may be stated as:

$$P_{i,min} \leq P_i \leq P_{i,max} \quad \text{for } i=1,2,\ldots,n \qquad (2)$$

Where:

- $P_{i,min}$ represents minimum power limit of i-th generator
- $P_{i,max}$ represents maximum power limit of i-th generator

2) Power Balance Constraint: To satisfy the total system load, the sum of the power generated by all units must equal the load demand plus any transmission line losses. This is expressed by the following equation:

$$\sum_{i=1}^{n} P_i - P_L = P_D \qquad (3)$$

Where:

- P_D is the total load demand,
- P_i is the power output from the i-th generator,
- P_L represents the total transmission losses in the network.

3) Transmission Losses: The transmission losses in the system are typically calculated using the B-coefficient method. The losses can be expressed as:

$$P_L = P^T B P + P^T B_o + B_1 \qquad (4)$$

Where:

- P is the vector of generator power outputs,
- B is the loss coefficient matrix,
- B_o is the loss coefficient vector,
- B_1 is a constant representing fixed transmission losses.

IV. A SUMMARY OF CONVENTIONAL PSO, SA, AND THEIR HYBRID APPROACHES.

A brief overview of PSO, SA, and hybrid PSO-SA—a combination of the two—as well as the flowcharts that correlate to each are given in this section.

A. *PSO Technique*

Kenedy and Eberhart first proposed particle swarm optimization in 1955. The fundamental concept of PSO is based on an animal swarm's quest for food, like a fish flock or a flock of birds, as seen in Figure 1. According to this theory, a particle swarm embodies a type of social optimization, where the solution to a problem is determined by a fitness function. A social network, similar to a flock of birds, is formed by assigning neighbors to each individual, allowing them to interact with other randomly chosen individuals to initialize the solution. These individuals, referred to as particles, are characterized by three N-dimensional vectors: current position, previous best position, and velocity. In this context, N denotes the dimensionality of the search space. These following equations can be used to calculate the new position and velocity of the particles.

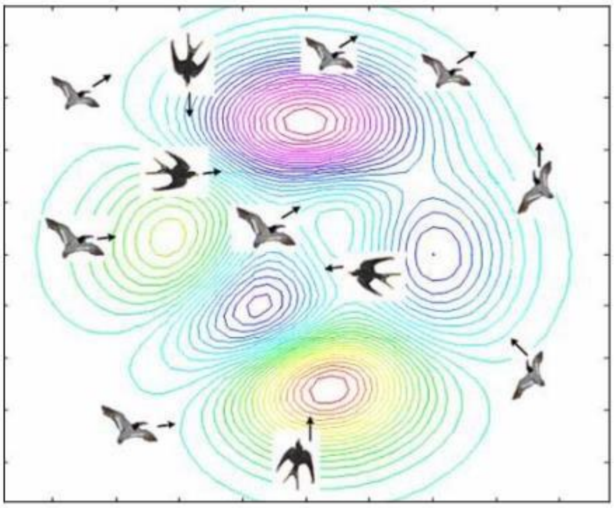


Figure1: Food searching of a swarm of birds

Every particle in the swarm is characterized by three essential N-dimensional vectors: its velocity, its current position, and its best-known position so far, known as "pbest.".The dimensionality of the search space is indicated here by (N). The particle's starting position and velocity at time (t = 0) can be described as follows:

$$xi(t) = (xi1(t), xi2(t), \dots, xin(t)) \quad (5)$$

$$vi(t) = (vi1(t), vi2(t), \dots, vin(t)) \quad (6)$$

The "global best" (gbest) signifies the optimal position discovered by any member of the entire swarm up to that point., while each particle follows its own best-known position, referred to as the "local best" (pbest). At any given time t, the values of pbest and gbest are expressed as follows:

$$pbesti(t) = (pbesti1(t), pbesti2(t), \dots, pbestin(t)) \quad (7)$$

$$gbesti(t) = (gbesti1(t), gbesti2(t), \dots, gbestin(t)) \quad (8)$$

In order to traverse the search space and improve fitness, the particles iteratively improve their locations and velocities. The following formulas are used to update their position and velocity:

$$vi(t+1) = w \cdot vi(t) + c1r1(pbesti(t) - xi(t)) + c2r2(gbesti(t) - xi(t)) \quad (9)$$

$$xi(t+1) = xi(t) + vi(t+1) \quad (10)$$

Where:

- w is the inertia weight controlling exploration
- c_1 and c_2 are acceleration coefficients
- r_1 r_2 are random values between 0 and 1
- x_i represents the current position of particle i,.
  v_i is the velocity of particle i

By avoiding stagnation and utilizing continuous updates, these equations enable the particles to continuously explore the search space and converge toward optimal solutions. In Figure 2, the PSO flowchart is displayed.

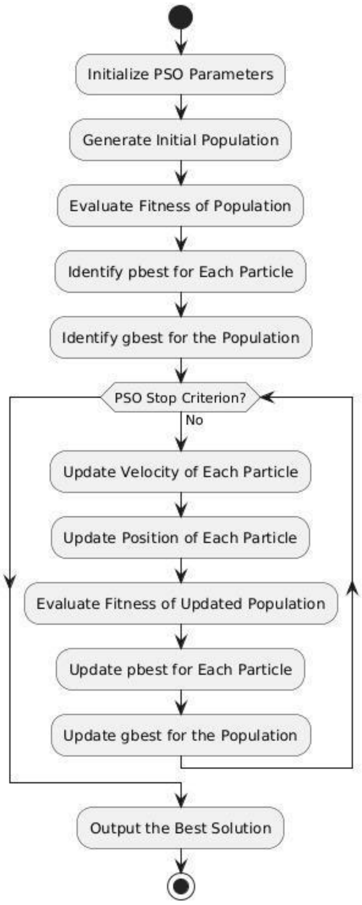


Fig.2 PSO Algorithm flowchart

B. SA Technique

In contrast, each producing unit's actual power generation is a decision variable for current methods based on SA algorithms to solve the ELD problem. This causes the problem to be very huge, which slows down the algorithms and renders them inefficient if there are a lot of generating units. Instead, it is recommended to simplify the problem by using the penalty factor (λ), which is derived from the traditional λ-iteration method, as a single decision variable, regardless of the number of generating units. As dependent variables, the producing units' actual power outputs are calculated as a function of λ. during every value of λ found during SA iterations, the real power outputs are computed. The minimum and highest power needs that the system can meet determine the appropriate range for the decision variable λ. These bounds are assessed by calculating the lower and upper incremental cost values that are produced by replacing all of the generating units' real power limits in equation (5):

$$ICmax = 2ai\,Pmax + bi \quad (11)$$

$$ICmin = 2ai\,Pmin + bi \quad (12)$$

Next, the minimum and maximum values of λ are determined based on the lowest and highest incremental cost values derived from equation (11), as shown below.

$$\lambda max = max\,(IC1, IC2, \ldots, ICN) \quad (13)$$

$$\lambda min = min\,(IC1, IC2, \ldots, ICN) \quad (14)$$

The SA algorithm minimizes a cost function in order to find the best answer. The generating units' total fuel costs are taken into account as the goal function in the suggested method. To address the power balance limitation, a penalty term is also added. For impractical solutions, this penalty raises the cost function. The fuel cost and the power balance constraint are combined to create the final cost function, which is shown as follows:

$$\sum_{i=1}^{N} F_i(P_i) + \eta\,(P_D + P_{loss} - \sum_{i=1}^{N} P_i) \quad (15)$$

Unlike other SA-based techniques that need the generation levels of all units to be considered as variables, this formulation uses only one decision variable, η. This reduction in decision variables decreases computational complexity and improves the algorithm's convergence speed. The SA Flowchart is shown in figure 3.

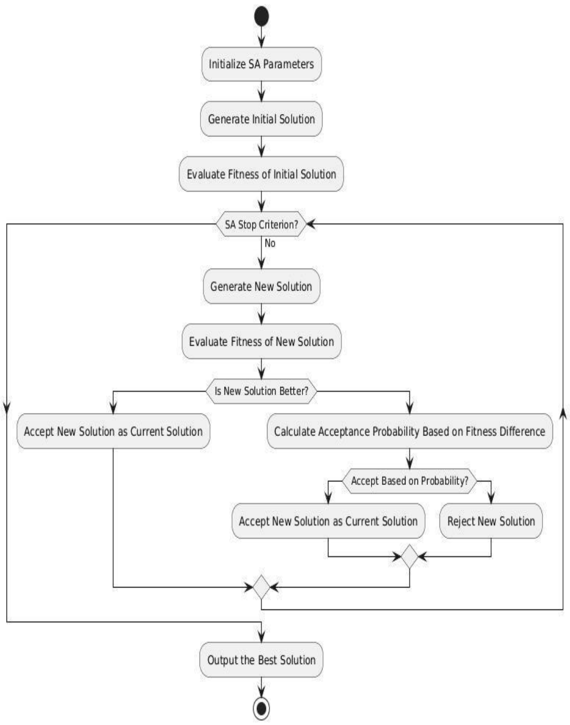


Figure 3: Algorithm flowchart of SA

C. The Hybrid PSO-SA Method

The hybrid algorithm to solve the ELD problem amalgamates the following: PSO is basically focused on exploration capabilities, [18] while SA generally contributes towards local optimization and convergence refinement. PSO efficiently handles global search but local optima have to be avoided and local minima escaped during refinement using SA. number of the SA technique takes over in iterations to refine the best solution found. [13] SA introduces randomness by perturbing the solution and gradually cooling down the temperature to explore a smaller neighborhood of the solution space.

1) *Minimized Fuel Cost:* This hybrid algorithm produces the minimized fuel cost of the given power demand. Each generator's output is adjusted to prevent both infeasibility and an increase un overall cost.

2) *Power Balance*: The total power output over the system's demand as well as transmission losses is maintained. every unit's generation limits are respected ensuring realistic as well as feasible allocations of the power.

3) *Enhanced Convergence*: PSO and SA hybridization increases the probability of finding a globally optimal solution to the ELD issue by removing local optima for the algorithm. However, after convergence by PSO, SA runs and soothes the solution already acquired thus yielding a more fine-tuned and economical power dispatch. The flow diagram of the hybrid model is given in figure 4.

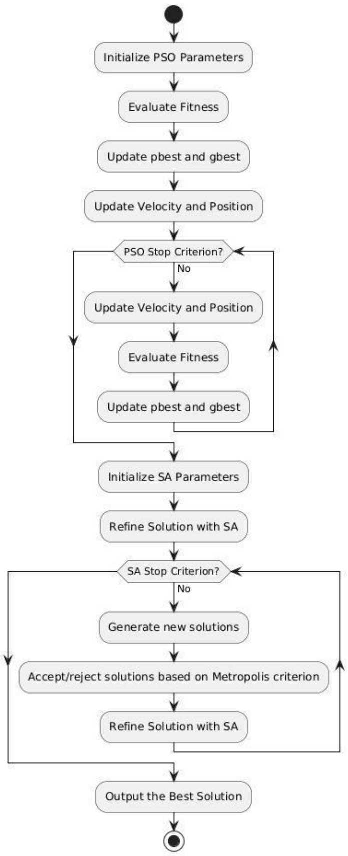


Figure 4: Flowchart Hybrid PSO-SA Technique

V. SIMULATION RESULTS AND DISCUSSION

A power system consisting of three generators and five buses, with a total load demand of 150 MW, was evaluated using Particle Swarm Optimization (PSO), Simulated Annealing (SA), and a hybrid PSO-SA approach in MATLAB to address the Economic Load Dispatch (ELD) challenge. The system's parameters, based on Example 7.7 from reference [5], are outlined in Table I.

TABLE 1. COST FUNCTION AND GENERATOR CONSTRAINTS

Unit	P_{min} (MW)	P_{max} (MW)	α	β	γ
1	10	85	200	7	0.008
2	10	80	180	6	0.009
3	10	70	140	6	0.007

Where the loss matrix BB is defined as:

$$B_{ij} = \begin{bmatrix} 0.0218 & 0.0093 & 0.0028 \\ 0.0093 & 0.0228 & 0.0017 \\ 0.0028 & 0.0017 & 0.0179 \end{bmatrix}$$

$$B_{0i} = [\,0.0003 \quad 0.0031 \quad 0.0015\,]$$

$$B_{00} = 00030523$$

Based on the information provided in Table I, the lowest power output that each of the three generators can produce is 10 megawatts. The maximum generation limits are 85 MW for Generator 1, 80 MW for Generator 2, and 70 MW for Generator 3. These upper and lower bounds, along with the cost function input data, define the operating range of the generators. Ideally, electricity generation should match the system's demand, but practical constraints, such as transmission line losses, make this challenging. To account for these losses, a matrix B is used to optimize the problem by incorporating load distribution. The coefficients of this matrix, known as loss coefficients, are derived from the bus matrix of the observed bus. The Economic Load Dispatch (ELD) problem for this test system was also analyzed in Example 7.7 [5], with the minimum fuel cost recorded as $1599.98. To efficiently optimize the ELD problem and demonstrate the superior optimization capabilities of the hybrid PSO-SA approach, the same test system was evaluated using PSO and SA. Each algorithm was set to a maximum of 1000 iterations. The test system was simulated in MATLAB for PSO, SA, and Hybrid PSO-SA, with the best results obtained from 30 runs per algorithm. The total production cost calculated using PSO was $1580/h, while SA achieved a lower cost of $1573/h. However, the hybrid PSO-SA method yielded the lowest total fuel cost of $1567/h, proving to be the most cost-effective approach.

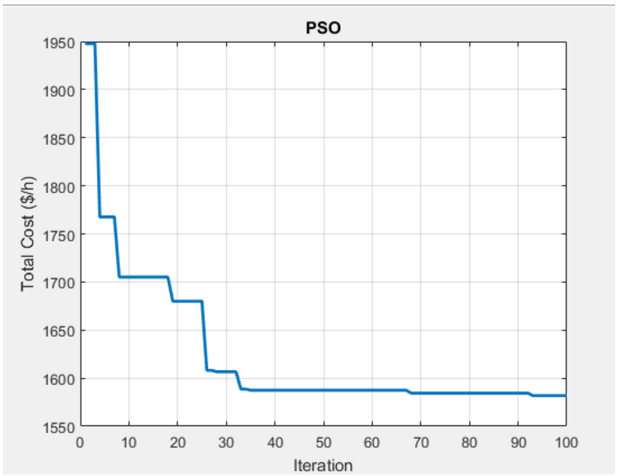


Figure 5. Total cost of generation vs no of iteration (PSO).

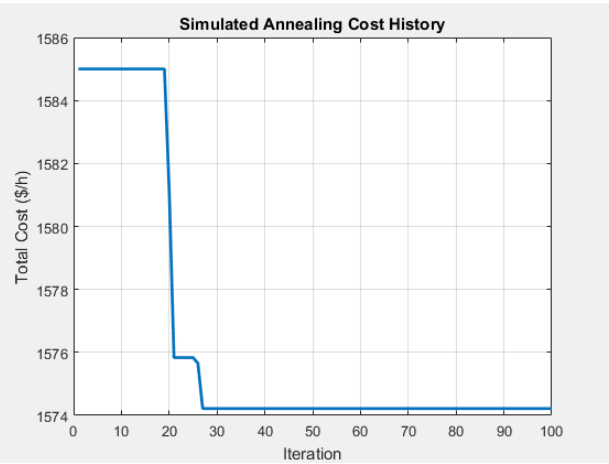


Figure 6. Total cost of generation vs no of iteration (SA).

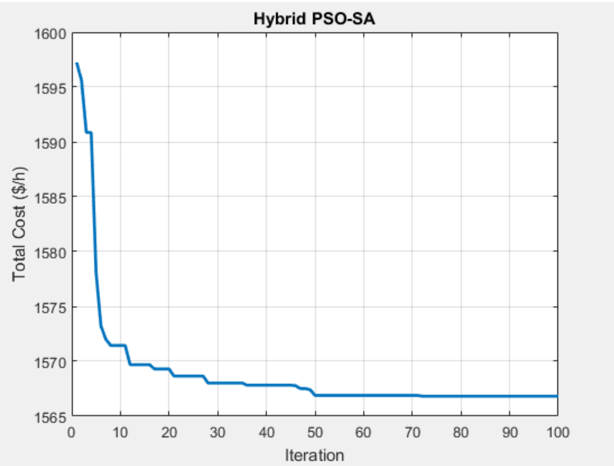


Figure 7. Total cost of generation vs no of iteration (Hybrid PSO-SA).

The outcomes of the three optimization techniques are summarized using bar charts, as shown in Figure 8, to facilitate better evaluation.

VI. CONCLUSIONS

In this research paper, the hybrid PSO-SA algorithm is pres nted to address efficiently and adequately complex powe

accu

getti

observed that hybrid algorithm results in superior quality solu
man

eco
and
This

VII. REFERENCES

[1] S. K. Emad Mirsadeghi, "Hybridizing particle swarm optimization with simulated annealing," *Cluster Computing (2021) 24:1135–1163*, p. 29, 2021.

[2] V. R. Sudhakaran Ajay, "Application of Particle Swarm Optimization for economic load dispatch problem," *The 14th International Conference on Intelligent System Applications to Power Systems, ISAP 2007* , no. November 4 - 8, 2007, Kaohsiung, Taiwan , p. 7, 2007.

[3] S. w. Jorge Nocedal, Numerical Optimization 2nd ed, United States of America: Springer series in operation reasearch, 2006.

[4] S. Z. Edwin Chong, An Introduction to Optimization, JOHN WILEY & SONS, 2001.

[5] W. S. John Grainger, Power System Analysis, McGraw-Hill Series in Electrical and Computer Engineering .

[6] Y. Zhang, "A survey of dynamic programming algorithms," *Proceedings of the 2023 International Conference on Machine Learning and Automation,* no. 2721/35/20230392, 2023.

[7] M. J. Vishu Kumar Prajapati, "Tabu Search Algorithm (TSA): A Comprehensive Survey," *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE),* vol. 10.1109/ICETCE48199.2020.9091743, 07-08 February 2020.

[8] A. H. G. X.-S. Yang, "Cuckoo search algorithm: a metaheuristic approach to solve," *Article in Engineering with Computers · April 2013,* Vols. 10.1007/s00366-012-0308-4, p. 27, 2013.

[9] T. A. a. A. Vlachos, "Application of the Firefly Algorithm for Solving theEconomic Emissions Load Dispatch Problem," *2 International Journal of Combinatorics,* vol. 1155/2011/52380, no. Hajo Broersma, p. 23, 2019.

[10] T. S. Marco Dorigo, Ant Colony Optimization, A Bradford Book, 2004.

[11] M. S. B. Maind, "Research Paper on Basic of Artificial Neural Network," *International Journal on Recent and Innovation Trends in Computing and Communication 2321-8169,* vol. II, no. I, p. 28, 2014.

[12] R.Murugan, "A PSO Based Hybrid Algorithm with Adaptive Penalty Function," *Archives Des Sciences,* Vol 66, No. 4;Apr 2013.

[13] C. D. G. S. Kirkpatrick, Optimization by Simulated Annealing, 1983.

[14] R. R. Venkatesh, "Economic load dispatch using simmulating annealing," *International Research Journal of Engineering and Technology (IRJET),* vol. II, no. III, p. 4, 2015.

[15] K. Shafie, "Improved version of bees algorithm for Global optimization," University of Birmingham Research Archive, 2005.

[16] J. X. ,. H. Z. WU DENG, "An Improved Ant Colony Optimization," *W. Deng et al.: ICMPACO Algorithm Based on Hybrid Strategies for Scheduling Problem,* no. Digital Object Identifier 10.1109/ACCESS.2019.2897580, 2019.

[17] F. Chan, Harmony Search and Krill Herd Hybrid for Optimization., MC Graw Hill, 2012.

A. G. Gad, "Particle Swarm Optimization Algorithm and Its Applications:," *Archives of Computational Methods in Engineering (2022),* no. 10.1007/s11831-021-09694-4, p. 31, 2022.

[19] E. Mirsadeghi&.

[20] M. T. Ahmed Ali, "A Hybrid PSO and DE Algorithm for Solving Engineering," *Applied Mathematics & Information Sciences,* no. Appl. Math. Inf. Sci. 10, No. 2, 431-449 (2016) , p. 19, 2016.

M. Haldurai, "A Study on Genetic Algorithm and its Applications," *International Journal of Computer Sciences and Engineering O,* vol. III, no. 10, 2016.

Performance Evaluation of Microchannel Geometries for Efficient Heat Transfer

Sanaullah Khan[1], Abdul Haseeb[1], Muhammad Shaheryar Khan[1], Umer Zahoor[1], Dr. Kareem Akthar[1], Muhammad Anas Wazir[2]

1 University of Engineering and Technology, Peshawar, Pakistan

2 Ghulam Ishaq Khan Institute, Swabi, Pakistan

Abstract:

As power densities in modern CPUs and GPUs continue to rise (CPUs: 100-300 W/cm², GPUs: 300-1000 W/cm²), traditional air-cooling methods are becoming inadequate. This study investigates the thermal-hydraulic performance of various microchannel heat sink (MCHS) geometries to optimize cooling efficiency for high-performance electronics. Using computational fluid dynamics (CFD) simulations in ANSYS Fluent, single-channel designs with different cross sections rectangular, trapezoidal, and diamond-shaped were analyzed under varying flow conditions. Key performance metrics, including temperature distribution, pressure drop, Nusselt number, and heat transfer coefficient, were evaluated through simulations to figure out the most effective configuration. Results show that converging channels provide superior heat transfer but at the cost of higher pressure drop and entropy generation, while rectangular and diamond-shaped channels offer a balanced trade off with moderate efficiency and controlled pressure losses. These findings highlight the crucial role of geometry in optimizing thermal management solutions, contributing to the development of next generation cooling technologies for high performance electronics.

Keywords: Thermal Enhancement, Microchannel, Different cross-sections, Heat sink, Liquid cooling.

1 INTRODUCTION:

Microchannel heat sinks have emerged as a pivotal technology in the cooling of electronic devices, addressing the increasing heat dissipation demands of modern compact systems. These heat sinks leverage the principles of forced convection to enhance heat transfer capabilities, making them a reliable solution for maintaining the integrity and longevity of electronic components [1] [2]. The need for removing heat dissipated from modern electronics like processors, graphics processing unit (gpu) are increasing day by day as with each launch of a new electronic device, the computational power to size ratio is increasing as technology is evolving rapidly and with that the power density is also increasing, raising the need for the new solutions for heat removal as conventional methods are not fulfilling the requirements. Conventional air cooling is not sufficient as with increasing flow of air, the problem of noise and vibrations arises [3], along with that the ambient temperature effect air cooling drastically [4]. It is also prone to space availability as it has large heatsinks and fins which may or may not fit into the available space and restricts future upgrades of the system. The nominal power densities for different electrical devices are as follows.

Modern CPUs

- Desktop CPUs: 100 - 200 W/cm²
- Server CPUs: 100 - 300 W/cm²

Modern GPUs

- Gaming GPUs: 300 - 500 W/cm²
- Data Center GPUs: 400 - 1000 W/cm²

To tackle such high heat dissipation researchers are developing new techniques and trying to find the breakthrough in the existing technology. It all started with the D. B. Tuckerman and R. F. Pease [5], who published their pioneering work in 1981 to tackle high heat dissipation and obtained the astonishing results by using water cooling and maintaining its temperature at 710 C and achieving the heat flux of 790 W/cm2. Their work laid the founding stone for future studies on heatsinks. Generally heat is carried away from the device in two way, though passive cooling or active cooling. Both have their pros and cons.

1.1 Passive cooling:

Passive cooling is a thermal management technique in which convection and radiation is used for dissipating heat without using active methods like fans or water cooling. It is used where noise reduction, reliability and energy efficiency are critical design factors such as LED lighting, industrial electronics, and embedded systems, where thermal management is essential without active cooling components. Usman et al [6] studied the thermal effect of the passive cooling on finned and un-finned heatsink integrated with the phase change material (PCM) and using finned heatsink as the thermal conductivity enhancement. They experimented on heatsinks with and without PCM and different in-fin arrangements heatsinks. They concluded that triangular inline pin-fin is the dominant heat sink geometry and RT-44 as the most efficient PCM for passive thermal management of electronic devices. Further Zahid et al [7] investigated the electronic cooling through nanoparticles and metal foam composite using passive cooling. They investigated the effect of alumina nanoparticles embedded in PCM and found that by incorporating these nanoparticles in PCM it enhanced the cooling effect and can be used to optimize the passive cooling techniques where reduced noise and energy efficiency is a goal.

1.2 Active cooling:

Another technique which is used for high power densities and is moderately used in new electronics, CPU's, GPU's is active cooling which uses forced flow mechanisms such as fans, pumps, thermoelectric coolers to dissipate heat. It is used abundantly now a days as one of the most efficient heat removal process. Hui Tan et al [8] employed forced flow of liquid using pump to carry out the heat from heatsink and rejecting all of that heat into atmosphere using a condenser. They used this technique to investigate different topological geometries like lateral veiny, snowflake, alternate veiny and spider netted designs. In their study they concluded and optimized spider netted design for 100 W/cm2 as flux given and achieving overall temperature of 9.90C which is result of enhanced fluid flow compared to straight channel whose heat dissipation is limited due to various factors. Biqi Cao et al [9] [10] tried to reduce hotspot temperature by using jet impingement at inlet which is located right above the hotspot and tried to reduce pressure drop along the channel and increase thermal efficiency and compared its results with straight micro channel heatsink and showed a significant thermal efficiency increase and reduced pressure drop by using manifold microchannel, radial microchannel and annular manifold microchannel.

A lot of research has been carried out on the heatsinks to fulfill the demand of new generation technologies as for high power data centers immersion cooling have been introduced to tackle heat dissipation [10] [11] which uses non-conductive fluid to remove heat when electrical components are submerged in the fluid. While others focused on changing the geometry of channel for thermal enhancement [12]. To further enhance the thermal performance of heatsink and properly utilize the maximum available space along with geometries of microchannel the pressure drop is also used as a performance deciding factor [13] [14].

In this study we have simulated different geometries and in each case, we study their characteristic properties that influence the overall thermal efficiency of a heatsink. These factors include area weighted average at base, pressure drop, entropy generation, Nusselt number, Reynold number, enthalpy, various velocities at inlet.

2 MODEL DESCRIPTION AND METHODOLOGY:

A: Geometrical Designs: In our study we are mimicking the heat generated by microprocessor or CPU and cool it by trying different cross sections. The channel length is 60mm in all cases. The rectangular channel is shown in figure 1(a) where LC is the channel length, which is 60mm, HT is the total height, which is 10mm, HC is the channel height, which is 8mm, WT is the total width which is 1.5mm and WC is width of the channel which is 0.5mm. With rectangular channel other channel shapes like divergent, convergent, triangular, trapezium, T-shape and diamond shapes used for simulations. Different parameters were studied to find the best and efficient shape which can carry as much heat as possible while being minimalistic according to the size of a CPU socket or bracket.

2.1 Solution method and grid independence:

The continuity, momentum, and energy equations are solved using ANSYS Fluent. The standard scheme for pressure discretization, simple algorithm for pressure velocity coupling and the first order upwind scheme for momentum and energy equations are used. For grid independence, three grid sizes are evaluated separately. In lieu of computational resources and time, further refinement of grid is stopped when variation in results upon further decrease in grid size was below 1%.

2.2 Geometric Domains:

Figure 1 shows different cross-sections of Heat sink that are modeled and studied. The overall length of the channel is set 60mm for each case. *Figure 2* shows the overview of the dimensions used for the cases.

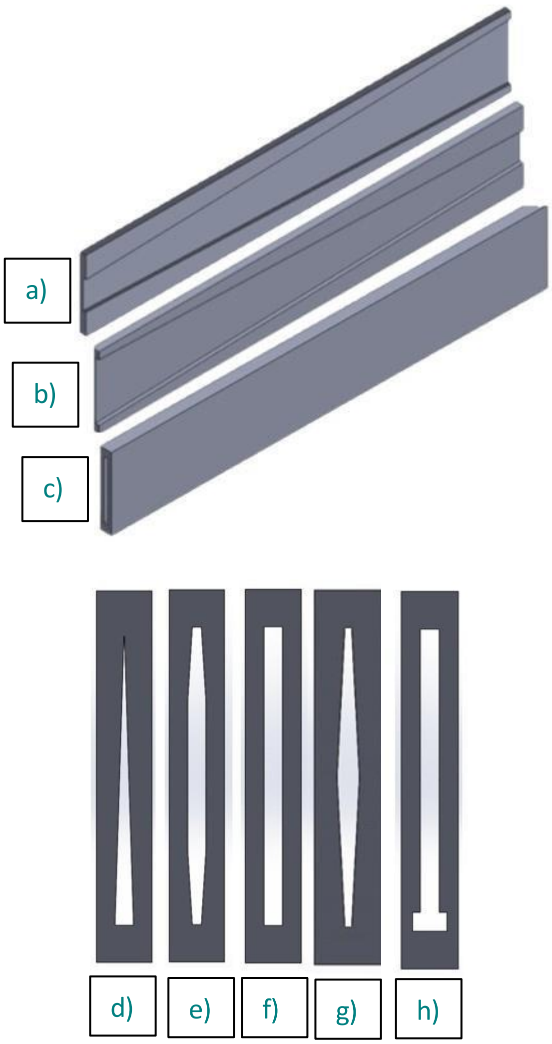


Figure 1: Cross-sections of different channels, a) Divergent Channel, b) Convergent Channel, c) Rectangular Channel, d) Triangular, e) Trapezoidal, f) Rectangular, g) Diamond, h) T shaped

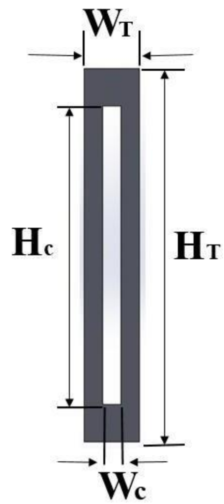


Figure 2: Cross-sectional dimensions of MCHS channels

Here, Hc, Wc, HT and WT represent the channel height, channel width, total height of the microchannel heat sink, and the total rib width, respectively, with dimensions of 8 mm, 0.5 mm, 10 mm, and 1.5 mm.

2.3 Governing Equations:

Based on earlier research by various scientists, the following assumptions were made to build a conjugate model of 3D flow and heat transfer:

1. The fluid flow in the MCHS is laminar.

2. Fluid is incompressible and Newtonian.

3. Heat transfer to the fluid and fluid flow is assumed to be steady.

4. The solid domain's thermo physical properties are assumed to be constant.

Taking the preceding assumptions into account, the following governing equations were used to construct the mathematical computational model.

Continuity Equation:

For incompressible fluid flow, the mass conservation equation is

$$\nabla \cdot \boldsymbol{V} = 0 \qquad \textit{(Eq 1)}$$

From *(Eq 1)* the x, y and z components of velocity can be found as

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z}= 0 \qquad (Eq\ 2)$$

where u, v, and w are the components of velocity in x, y, and z-direction.

Momentum Equation:

The Navier-Stokes equation governs the momentum conservation

$$\rho(\boldsymbol{V}\cdot\nabla)\boldsymbol{V}=-\nabla P+\mu\nabla^2\boldsymbol{V} \qquad (Eq\ 3)$$

where ρ is the density, μ is the dynamic viscosity and P is the pressure. *(Eq 3)* can be further solved to the x, y, and z direction momentum equations as:

In x-direction

$$u\frac{\partial u}{\partial x}+v\frac{\partial u}{\partial y}+w\frac{\partial u}{\partial z}=-\frac{1}{\rho f}\frac{\partial p}{\partial x}+\frac{\mu}{\rho f}\left(\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right) \qquad (Eq\ 4)$$

In y-direction

$$u\frac{\partial v}{\partial x}+v\frac{\partial v}{\partial y}+w\frac{\partial v}{\partial z}=-\frac{1}{\rho f}\frac{\partial p}{\partial y}+\frac{\mu}{\rho f}\left(\frac{\partial^2 v}{\partial x^2}+\frac{\partial^2 v}{\partial y^2}+\frac{\partial^2 v}{\partial z^2}\right) \qquad (Eq\ 5)$$

In z-direction

$$u\frac{\partial w}{\partial x}+v\frac{\partial w}{\partial y}+w\frac{\partial w}{\partial z}=-\frac{1}{\rho f}\frac{\partial p}{\partial y}+\frac{\mu}{\rho f}\left(\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}+\frac{\partial^2 w}{\partial z^2}\right) \qquad (Eq\ 6)$$

Where ρf is the density

Energy Equation:

For solid domain

$$k_s\left(\frac{\partial^2 Ts}{\partial x^2}+\frac{\partial^2 Ts}{\partial y^2}+\frac{\partial^2 Ts}{\partial z^2}\right)=0 \qquad (Eq\ 7)$$

For fluid domain

$$u\frac{\partial Tf}{\partial x}+v\frac{\partial Tf}{\partial y}+w\frac{\partial Tf}{\partial z}=\frac{Kf}{\rho fCpf}\left(\frac{\partial^2 Tf}{\partial x^2}+\frac{\partial^2 Tf}{\partial y^2}+\frac{\partial^2 Tf}{\partial z^2}\right) \qquad (Eq\ 8)$$

Where T is the temperature and k are the thermal conductivity and Cp is the specific heat for solid (aluminum) and fluid (water) domains, respectively.

Nusselt Numbers:

$$Nu=\frac{hD_h}{k} \qquad (Eq\ 9)$$

Where h, Dh and k are convective heat transfer coefficient, hydraulic diameter and thermal conductivity of the fluid respectively. *(Eq 9)* is used to figure out the effectiveness of convective heat transfer in a system compared to conduction.

Reynolds Number:

$$Re=\frac{\rho VD_h}{\mu} \qquad (Eq\ 10)$$

Where ρ = Density of the fluid (kg/m³), V = Fluid velocity, Dh = Hydraulic diameter, μ = Dynamic viscosity of the fluid

$$D_h=\frac{4A}{P} \qquad (Eq\ 11)$$

Where A = Cross-sectional area of the flow, P = Wetted perimeter

$$P=\sum L_i \qquad (Eq\ 12)$$

Where Li = Length of each surface in contact with the fluid

2.4 *Boundary Conditions:*

Following are the boundary conditions that are used for simulations.

Temperature at inlet	297 K
Flux given	700000 W/m^2
Left and right wall	Symmetry
Bottom wall	Heated wall
Front, top, and back wall	Adiabatic

Table 1: Boundary Conditions used for different cross sections of geometry

2.5 *Model Validation:*

To validate numeral simulation, SIDDIQUI et al [15] experimental data is modeled numerically on ANSYS FLUENT using the same boundary condition. The temperature results of specific point on the heatsink same as their data acquisition point was compared and fluid was water in both cases. The results are shown in *Figure 3*. This shows that the results deviate from the experimental work by about less than 5% for Reynold Numbers 400-750, thus the present model can be considered as suitable for predicting the thermo-fluid characteristics of MCHS.

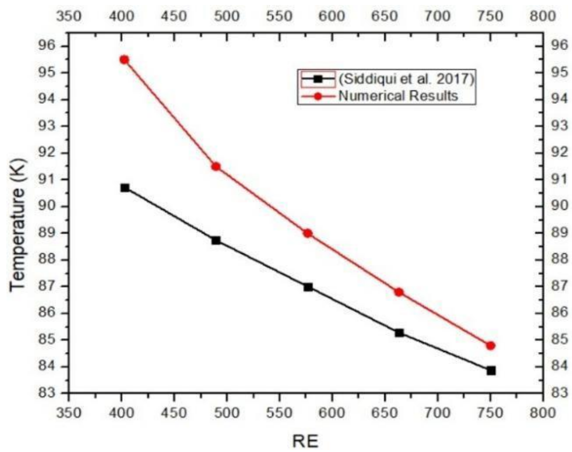


Figure 3: Validation for Temperature vs Reynolds number

3 RESULTS AND DISCUSSIONS:

Thermal Characteristics:

In the study of heat transfer, the cross section of different geometries of a channel plays a vital role in determining its thermal performance. Changings in cross sections can influence key parameters such as outlet temperatures, base heated wall temperatures, and average fluid temperatures. These variations stem from differences in surface area exposure, flow distribution, and thermal boundary layer development inherent to each geometry.

3.1 Outlet Temperature Analysis*:*

Figure 4 shows the variation of fluid temperature at the outlet with increasing Reynolds number (Re) for different cross-sectional geometries. As Reynolds' number increases, the fluid temperature at the outlet decreases for all geometries. This is because a higher Reynolds number indicates a greater flow velocity, leading to higher convective heat transfer and better cooling of the fluid. As the Reynolds number increases, it shifts the flow to more mixing and a quicker change in the mass of cooled fluid, hence increasing the overall performance of channels. While all geometries follow the same general decreasing trend, some geometries maintain a higher outlet temperature than others. The trapezoidal section has the highest outlet temperature at most Reynolds numbers and the rectangular cross-section seems to have the lowest outlet temperature at most Reynolds numbers because the convective heat transfer in case of Trapezoidal cross-section is the least of the others. Figure 8 shows the less Nusselt number values for trapezoidal cross section.

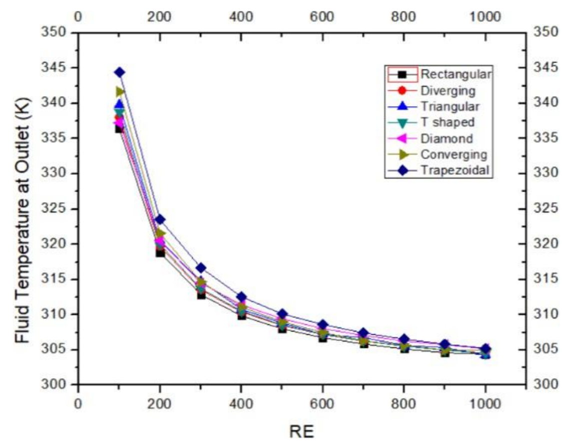


Figure 4 Fluid temperature at outlet vs Reynolds number graph for different cross sections

3.2 Average Fluid Temperature:

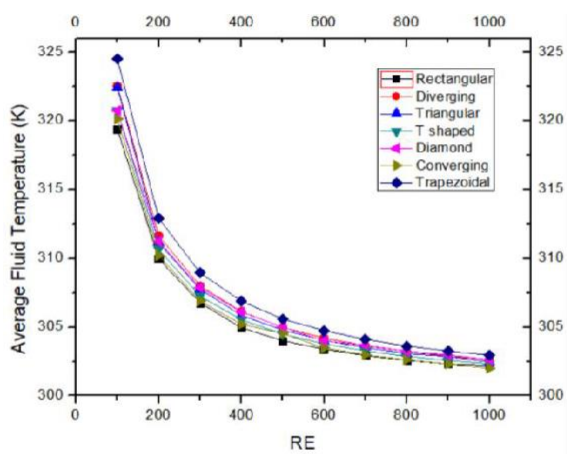


Figure 5: Average fluid temperature vs Reynolds number graph for different cross sections

Figure 5 shows the same trend of temperature as that shown in Figure 4. At low Re, the fluid temperature is high due to dominant conduction and weak convective heat transfer for all geometries. As Re increases, turbulence enhances convective heat transfer, resulting in lower temperatures. Trapezoidal geometry consistently has the highest temperature values in both graphs. This suggests that it retains more heat compared to other geometries, due to weaker convective effects. It is also evident from both *Figure 4* and *Figure 5* that the temperature differences between difference cross sectional geometries at lower Reynolds number is more whereas at higher value of Reynolds number, the difference in temperatures lessens as turbulence and more convection increases heat transfer efficiency, reducing variations in thermal performance across geometry

3.3 Base Heated Wall Temperature Analysis:

The base heated plate temperature is a critical parameter in assessing the thermal performance of various channel geometries under different flow conditions. *Figure 6* shows the temperature at the base heated wall of the channel at different values of Reynolds numbers for different cross sections. The trends show the same behavior of the channels with temperatures decreasing with increasing fluid velocity and Reynolds number. Trapezoidal geometry is again showing the highest base heated wall temperature across all Re values. Triangular geometry is giving more cooling at the base meaning that the heat supplied to the wall is effectively transferred to the fluid. The flow in these sections is more uniform, reducing stagnant zones and improving heat dissipation.

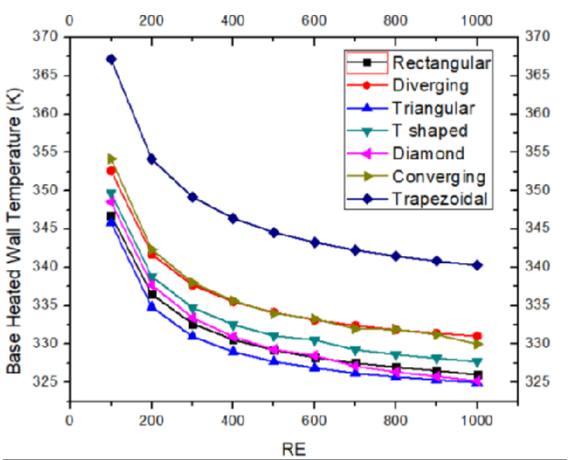


Figure 6: Base Heated Wall Temperature vs Reynold number graph for different cross sections

3.4 Temperature Distribution Along the Channel:

Figure 7 presents the temperature distribution at the interface between the solid and fluid domains for different cross-sectional geometries of the channels at a Reynolds number (Re) of 400. The temperature contours provide significant insight into the heat transfer characteristics and thermal performance of each geometry. T shaped geometry shows a uniform temperature distribution with lower peak temperatures at the heated wall. The extended surface area aids in better heat dissipation. The trapezoidal channel consistently shows the highest base heated wall temperature. This suggests lower heat transfer efficiency despite the fluid carrying more heat, as seen in *Figure 5* (average fluid temperature). This is due to insufficient convective cooling near the base. Notably, the temperature rises along the length of the channel. Optimized geometries like triangular, diamond, and rectangular sections provide better thermal performance.

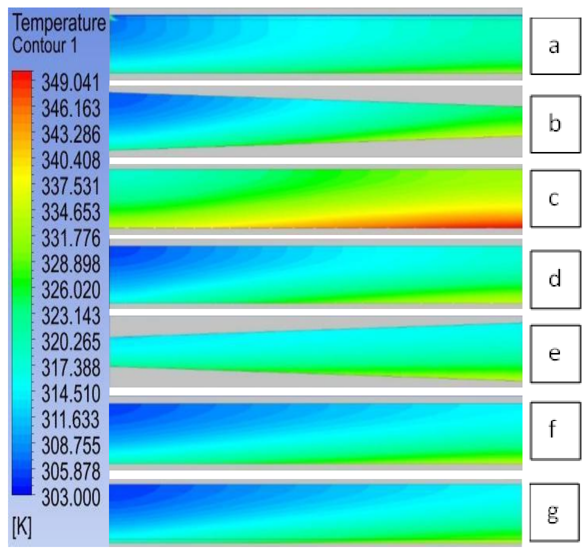


Figure 7: Temperature contours in YZ plane at contact region between fluid and solid domains at Reynolds number of 400. a) Triangular, b) Converging, c) Trapezium, d) Diamond, e) Diverging, f) Rectangular, g) T shape

3.5 Heat Transfer Characteristics:

Figure 8 illustrates the impact of different channel geometries on the Nusselt number (Nu) with an increase in Reynolds number (Re) value, highlighting their direct relationship. Among the studied geometries, converging channels exhibit the highest Nu, showing superior heat transfer performance. Rectangular, diverging, triangular, and T-shaped channels show moderate Nu values with similar trends, while diamond and trapezoidal channels show the least values. A higher Nu enhances thermal convection by promoting stronger fluid mixing, turbulence, and reducing thermal boundary layer thickness. Increased turbulence disrupts the boundary layer, improving heat transfer, while a thinner boundary layer eases heat dissipation. The deficient performance of diamond and trapezoidal shapes suggests reduced turbulence, leading to weaker heat transfer.

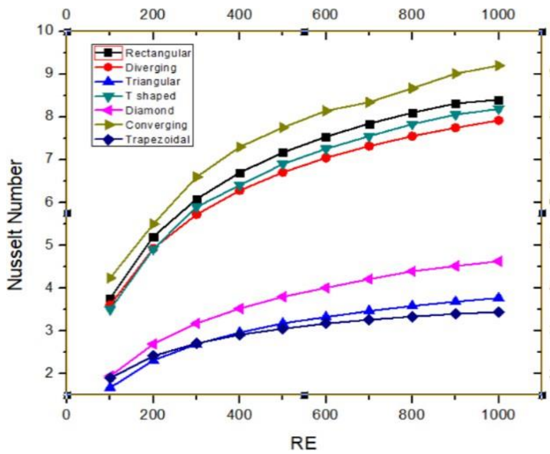


Figure 8: Nusselt number vs. Re for different geometries

3.6 Surface heat transfer co-efficent:

Figure 9 illustrates the relationship between the Reynolds number (Re) and the heat transfer coefficient (W/m²·K) for various channel geometries, showing that an increase in Re enhances heat transfer by reducing thermal boundary layer thickness. Converging channels show the highest heat transfer coefficient, making them the most effective for convection. Diamond, rectangular, diverging, and T-shaped channels perform moderately with similar trends, while triangular channels perform slightly worse. Trapezoidal channels have the lowest heat transfer coefficient, showing the least efficient convection. These trends align with the Nusselt number vs. Re graph, confirming that higher Nu values correspond to improved heat transfer performance.

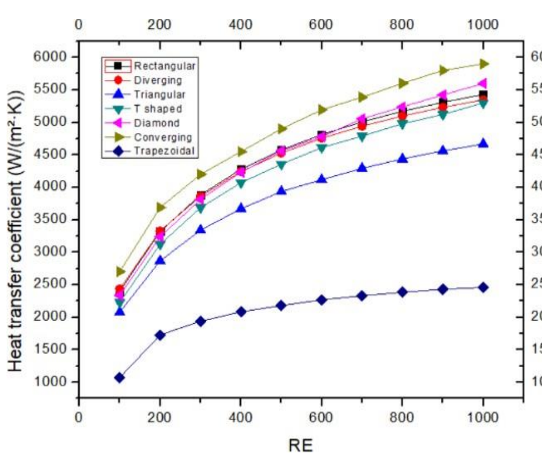


Figure 9: Heat transfer co-efficient vs. Re for different geometries

3.7 Pressure Drop Along the Channel

Microchannel pressure drop varies with cross-section shape at Reynolds numbers of 100-1000 as clear from *Figure 10.* Triangular channels show the highest pressure drop due to vortex formation at sharp corners, followed by diamond and converging designs. Rectangular and T-shaped channels show moderate resistance, while trapezoidal and diverging channels show the lowest pressure drop. Pressure drop increase with Reynolds number, most notably in triangular, diamond and converging sections. High-pressure-drop cross-sections like triangular and diamond channels enhance heat transfer but need more pumping power, while low pressure drop cross-sections such as diverging and trapezoidal channels maximize energy efficiency.

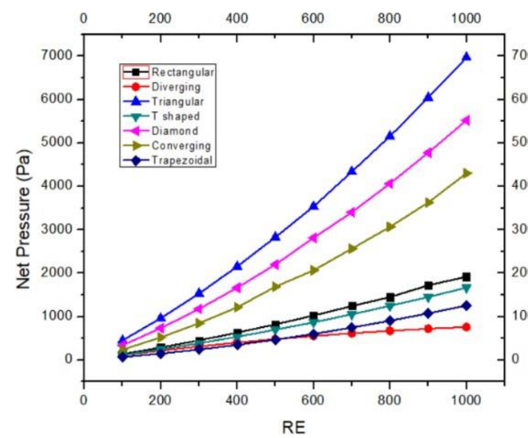


Figure 10: Net Pressure Drop vs Reynold number graph for different cross sections

3.8 Total Entropy Generation (TEG) Along the Channel

Figure 11 shows that convergent cross-sections generate the highest entropy (TEG), making them the least efficient, while diverging cross-sections show the lowest TEG, resulting in superior efficiency. Triangular and diamond cross-sections also produce high TEG due to vortex formation, while rectangular and T-shaped cross-sections display moderate TEG, offering a balanced performance. Trapezoidal cross-sections show near-optimal efficiency due to their gradual expansion. Across all cross-sections, TEG increases with flow velocity, though at varying rates, with convergent designs experiencing the sharpest rise. For high performance cooling applications, convergent and triangular cross-sections are most effective despite higher energy losses, while diverging and trapezoidal designs are more suitable for energy-efficient systems.

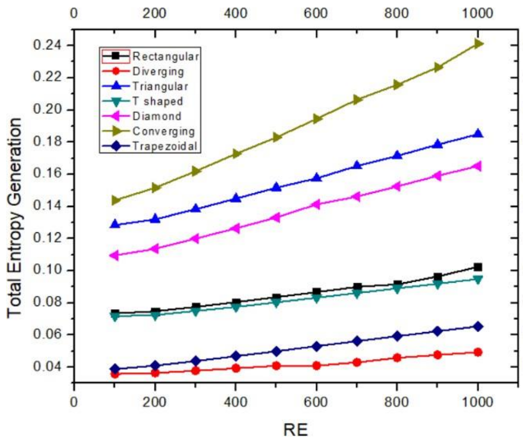


Figure 11: Total Entropy Generation vs Reynold number graph for different cross sections

3.9 Velocity Contours Along the Channel

Figure 12 presents the velocity distribution at the mid YZ plane along the channel length for different cross-sectional geometries of the channels at a Reynolds number (Re) of 400. The velocity contours offer significant insight into the heat transfer characteristics and thermal performance of each cross-section. In ANSYS Fluent, contours with blue showing the lowest and red the highest, while green, yellow, and orange show intermediate values.

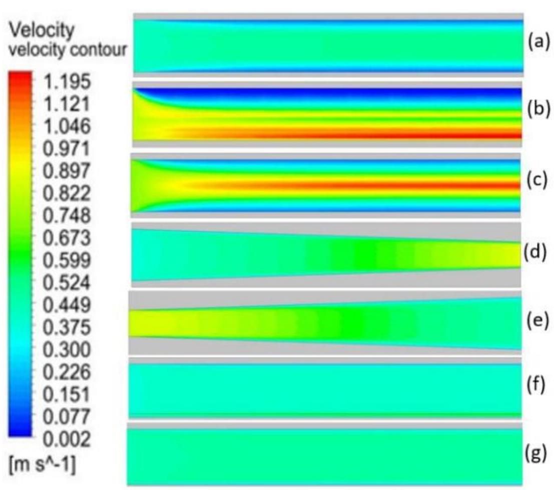


Figure 12: velocity contours on YZ plane at the middle of channels (x=0.75mm) over Reynolds number of 400. a). Trapezium, b). Triangular, c). Diamond, d). Converging, e). Diverging f). T shaped g). Rectangular

4 CONCLUSION

This study analyzed various parameters across Reynolds numbers (100–1000), leading to the following key findings:

- *Heat Transfer:* Converging channels show the highest Nusselt number and heat transfer coefficient, followed by rectangular and diamond-shaped channels. Triangular channels show the lowest heat transfer efficiency.
- *Pressure Drop:* Triangular channels experience the highest pressure drop, while rectangular channels have the lowest, making them preferable for minimizing pressure losses.
- *Entropy Generation:* Converging channels generate the highest entropy, followed by triangular and diamond-shaped channels, with rectangular channels producing moderate entropy.
- *Fluid Velocity & Wall Temperature*: Diamond-shaped channels show the highest fluid velocity. At low Reynolds numbers, triangular channels maintain the lowest wall temperature, while at higher Reynolds numbers, diamond-shaped channels perform best.

Overall, converging channels offer the best heat transfer performance but at the cost of higher entropy generation and pressure drop. Rectangular channels offer a balanced alternative for applications prioritizing energy efficiency. These results provide a strong basis for upcoming future development decisions regarding selection of the cross-section of heatsinks.

REFERENCES

[1] *S. T. Kadam and R. Kumar, "Twenty first century cooling solution: Microchannel heat sinks," International Journal of Thermal Sciences, vol. 85. Elsevier Masson SAS, pp. 73–92, 2014. doi: 10.1016/j.ijthermalsci.2014.06.013.*

[2] *C. Y. Zhao and T. J. Lu, "Analysis of microchannel heat sinks for electronics*

cooling.” [Online]. Available: www.elsevier.com/locate/ijhmt

[3] *A. ABBAS, W. Elwali, S. Haider, S. Dsouza, M. Sanderson and Y. Segab, "CAE Cooling Module Noise and Vibration Prediction Methodology and Challenges," SAE Technical Paper, 2020.*

[4] Z. Tark, A. J. Hamed and A. H. N. Khalifa, "Performance study of the thermoelectric personal cooler under different ambient temperatures," International Journal of Heat & Technology, vol. 30, no. 1, 2022.

[5] *D. B. Tuckerman and R. F. W. Pease, "High-performance heat sinking for VLSI," IEEE Electron Device Letters, vol. 2, no. 5, pp. 126-129, 1981.*

[6] *H. Usman, H. M. Ali, A. Arshad, M. J. Ashraf, S. Khushnood, M. M. Janjua and S. N. Kazi, " An experimental study of PCM based finned and un-finned heat sinks for passive cooling of electronics," Heat and Mass Transfer, vol. 54, pp. 3587-3598, (2018).*

[7] *I. Zahid, A. Qamar, M. Farooq, F. Riaz, M. S. Habib, M. Farhan and M. A. Hayat, "Experimental optimization of various heat sinks using passive thermal management system," Case Studies in Thermal Engineering, vol. 49, p. Article No. 103262, (2023).*

[8] *H. Tan, L. Wu, M. Wang, Z. Yang and P. Du, "Heat transfer improvement in microchannel heat sink by topology design and optimization for high heat flux chip cooling," International Journal of Heat and Mass Transfer, vol. 129, pp. 681-689, (2019).*

[9] *B. Cao and Z. Wu, "Microchannel heat sinks for hotspot thermal management: Achieving minimal pressure drop and maximal thermal performance," International Journal of Heat and Mass Transfer, vol. 236, p. Article No. 126411, (2025).*

[10] K. Zhou, X. Yu, B. Xie, H. Xie and W. Fu, "Immersion cooling technology development status of data center," Science and Technology for Energy Transition, vol. 79, p. Article No. 41, (2024).

[11] X. Li, S. Guo, H. Sun, S. Yin, S. Liu and Z. Wang, "Experimental study of the performance of liquid cooling tank used for single-phase immersion cooling data center," Case Studies in Thermal Engineering, vol. 63, p. Article No. 105386, (2024).

[12] *K. Grochalski, W. Rukat, B. Jakubek, M. Wieczorowski, M. Słowiński, K. Sarbinowska and W. Graboń, "The influence of geometry, surface texture, and cooling method on the efficiency of heat dissipation through the heat sink—A review," Materials, vol. 16, no. (15), p. Article No. 5348, (2023).*

[13] *A. A. Razali and A. Sadikin, "A CFD simulation study on pressure drop and velocity across single flow microchannel heat sink," Journal of Advanced Research Design, vol. 8, no. 1, (2015).*

[14] *K. M. Jung, R. A. Krishnan, G. U. Kumar and H. J. Lee, "Experimental study on two-phase pressure drop and flow boiling heat transfer in a micro pin fin channel heat sink under constant heat flux," Experimental Heat Transfer, vol. 34, no. (2), pp. 162-185, (2021).*

[15] *A. M. Siddiqui, W. Arshad, H. M. Ali, M. Ali and M. A. Nasir, "Evaluation of nanofluids performance for simulated microprocessor," Thermal Science, vol. 21, no. (5), pp. 2227-2236, (2017).*

Remote Sensing-Based Evapotranspiration Estimation for Optimized Tobacco Irrigation

Murshid Khan
dept. of Agricultural Engineering
University of Engineering & Technology
Peshawar, Pakistan
murshidafridi572@gmail.com

Nasru Minallah
National Center for Big Data and Cloud Computing
University of Engineering & Technology
Peshawar, Pakistan
n.minallah@uetpeshawar.edu.pk

Khushal Khan
National Center for Big Data and Cloud Computing
University of Engineering & Technology
Peshawar, Pakistan
engineerkhushalkhan@gmail.com

***Abstract*—Water scarcity in Pakistan is a growing concern. It is reported that the water requirement is falling below 1000 m³ per capita. Since Pakistan falls in a high-risk zone for climate effects, it is in dire need of modern practices to overcome the growing water issues. Pakistan, an agricultural country whose economy is highly reliant on agriculture, is dependent on water. Climate change and the scarcity of water have highly affected the agricultural system of the country. Tobacco, a significant cash crop in Pakistan, is primarily cultivated through an extensive canal system. This study focuses on assessing water usage for tobacco crop irrigation in the Mandani region by employing remote sensing technologies to estimate actual evapotranspiration (ETa). The Mandani region is situated in the Charsadda district, Khyber Pakhtunkhwa, Pakistan. It is considered one of the tobacco-dominant regions. The study utilizes satellite data from the European Space Agency (ESA), i.e., Sentinel-2 (S2) and Sentinel-3 (S3), alongside meteorological data from the European Centre for Medium-Range Weather Forecasts (ECMWF). For actual ETa estimation, data collected from these sources were processed and analyzed using the SNAP software. Further, the ETa values were compared with reference evapotranspiration (ETc). The study's results show that the estimated actual ETa values aligned with the reference values. However, a notable discrepancy was found, primarily due to inconsistent irrigation practices, lack of proper irrigation management, and the farmers' lack of scientific knowledge.**

Keywords— evapotranspiration, tobacco, Sentinel-2, Sentinel-3, European Space Agency, irrigation

I. INTRODUCTION

Pakistan’s agricultural sector plays an important role in the country’s economic growth, with agriculture contributing approximately 18.9% to the national Gross Domestic Product (GDP). The tobacco crop is considered one of the most important cash crops in Pakistan, contributing to the national economy. It is primarily found in the districts of Swabi, Charsadda, Buner, and Mardan in Khyber Pakhtunkhwa, Pakistan [1]. Despite its importance in the country's economy, tobacco faces challenges related to water management and climate change.

Pakistan, a country prone to water crises, is struggling with water scarcity. It is reported that the per capita water availability is below 1,000 m³. One of the contributing factors is water management, which also affects the agricultural system. Excessive water consumption and climate change are major issues associated with tobacco cultivation, as it is a crop with moderate water requirements. Additionally, tobacco is sensitive to both water stress (under-irrigation) and waterlogging (over-irrigation). To ensure optimal growth, proper water management is essential. In this context, evapotranspiration (ET) plays a significant role in the growth and development of crops, particularly water-sensitive ones like tobacco. ET refers to the combined process of water loss through evaporation from the soil and transpiration from the plant's leaves. The rate of ET depends on several factors, including water supply, weather conditions, soil properties, crop type, and the stage of plant growth.

One of the underrated issues is that farmers often lack scientific knowledge and guidance. In Pakistan, farmers and researchers primarily use traditional ground-based data collection methods. The most common, yet expensive, device used to estimate ET is the lysimeter, which has long been utilized for this purpose. It is a device used to measure the amount of water lost through both evaporation and transpiration. This practice is reliable; however, lysimeters are expensive and limited to small-scale studies. When it comes to academic research, expensive lysimeters are less available, consequently limiting research [4]. In contrast to lysimeters, remote sensing technologies have been introduced, providing not only a cost-effective solution but also greater convenience, as they can be used remotely [5]. Remote sensing involves the use of satellites and aircraft for the collection of meteorological data. The most widely used satellites for this purpose are European Space Agency (ESA) satellites, i.e., Sentinel-2 and Sentinel-3. These satellites collect data from remote areas through the measurement of reflected radiation.

In this paper, we have analyzed the ET of tobacco in Mandani, a region within Charsadda. This region was chosen due to its significant tobacco cultivation and reliable access to irrigation water. Additionally, ground-truthing was carried out, making it a suitable location for this research study. Using advanced tools, such as satellites from the ESA (i.e., Sentinel-2 and Sentinel-3), and ECMWF weather data , we investigated the spatial and temporal variations in ET to gain a deeper understanding of the water requirements of tobacco crops in the region.

II. REMOTE SENSING APPROCHES FOR ESTIMATING CROPS EVAPOTRANSPIRATION

ET plays a significant role in the water cycle of crops and the earth. For this purpose, several traditional methods are used, such as the Bowen-ratio, soil water balance, and eddy covariance [6], [7]. These methods are given less priority because of time consumption and are expensive. Lysimeter, a device used for estimation and analysis of ET, is de-prioritized due to its less availability in some underprivileged areas. However, due to advancements in technology, specifically in the monitoring of ET, remote sensing techniques have been introduced, which not only provide ET calculations remotely but also they are also less time-consuming and efficient.

The authors in [8] utilized remote sensing technologies to improve water management in California. The prime focus was estimating ET to optimize irrigation and reduce water consumption. The study suggests that a 10% reduction of ET could save significant water. The authors in [9] identified the ET and rainfall as the key parameters for the assessment of the water usage. Remote sensing technology such as Moderate Resolution Imaging Spectroradiometer (MODIS) and Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) were used. Landsat satellite images and vegetation indices are the key parameters to estimate the ET of crops. In [10], the research was conducted to estimate crop water demands in Comarca Lagunera, Mexico, across four growing seasons. The research concluded that remote sensing-based ET maps are effective tools for optimizing irrigation and promoting water conservation at regional and field scales. A research was conducted on the estimation of actual ET of tobacco crop using advanced satellites of European Space Agency (ESA) i.e., Sentinel-2 and Sentinel-3. The study was conducted on Yar Hussain region, with in district Swabi, Pakistan. The findings highlighted a pattern of excessive irrigation during the initial and development phases of tobacco, underscoring the significance of refining irrigation practices for the efficient growth of tobacco [11]. The research is limited to this specific area. To gain a broader understanding, further research should extend to other major tobacco-growing regions in Pakistan, allowing for a more comprehensive analysis of irrigation practices across the country. In our paper, the study in conducted in Mandani region located within the district Charsadda.

III. Study Area and Dataset

The current study was conducted in the Mandani region, located within the district of Charsadda, as depicted in Fig. 1. The locality is positioned between a Latitude: 34°20'42.81"N and Longitude: 71°47'13.53"E, and the elevation was found 1312.34 ft (400 m).

The data for this study were collected from multiple sources, including Sentinel-2, Sentinel-3, and ECMWF. The data collection process is discussed in detail in the forthcoming section. Since the study focused on the ET of tobacco, it was necessary to consider different time ranges. The data acquisition date of satellite imagery is shown in Table I.

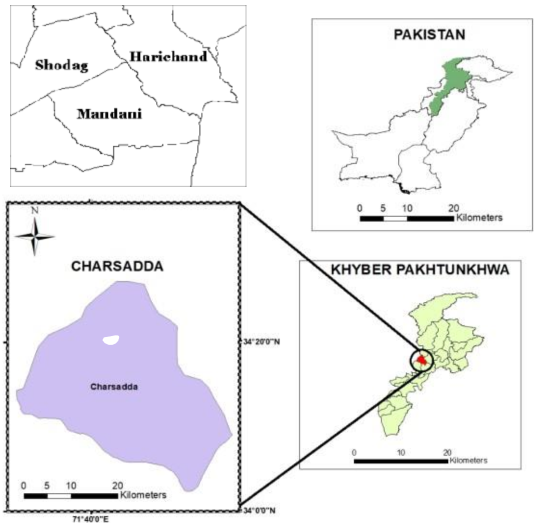


Fig. 1. Locality map

TABLE I. Data Acquisition Dates of Satellite Imagery

Stages	***DOY***
Initial Season	17-Mar-2024
	23-Mar-2024
	31-Mar-2024
Development Season	12-Apr-2024
	16-Apr-2024
	25-Apr-2024
	29-Apr-2024
	06-May-2024
Mid Season	08-May-2024
	12-May-2024
	20-May-2024
	26-May-2024
	06-Jun-2024
Late Season	14-Jun-2024
	23-Jun-2024
	30-Jun-2024

IV. Materials and Methodology

The required data for our experiments was collected from different sources, including Sentinel-2, Sentinel-3, and ECMWF. We retrieved the most up-to-date Sentinel-2 and Sentinel-3 data from ESA, which helped us find the ETa. Regarding ETc, we used data from the Food and Agriculture Organization (FAO) CropWat model and the Charsadda weather station to compare the water needs of crops (ETc) with the available water supply for crops (ETa) during its four different growth stages. Table 1 illustrates the timeline for the acquisition of satellite imagery and ECMWF data. Fig. 3 depicts the whole ET cycle.

A. Site Investigation and Classification

Mandani region emerged as the designated area of study located in Charsadda district of Khyber Pakhtunkhwa. During the ground truthing process, the coordinates of the desired region were determined 34°20'42.81"N and 71°47'13.53"E. For visual representation, the land use is classified using different colors. The classification map for the pilot region was generated utilizing an inception-inspired deep neural network, shown in Fig. 2. The model was developed and trained locally using field-collected data. It is clear from the figure that the tobacco crop is prominent in the region.

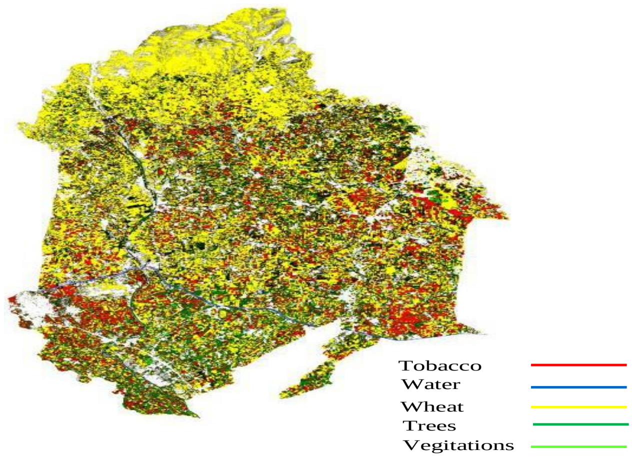


Fig. 2. Land-use classification of study area

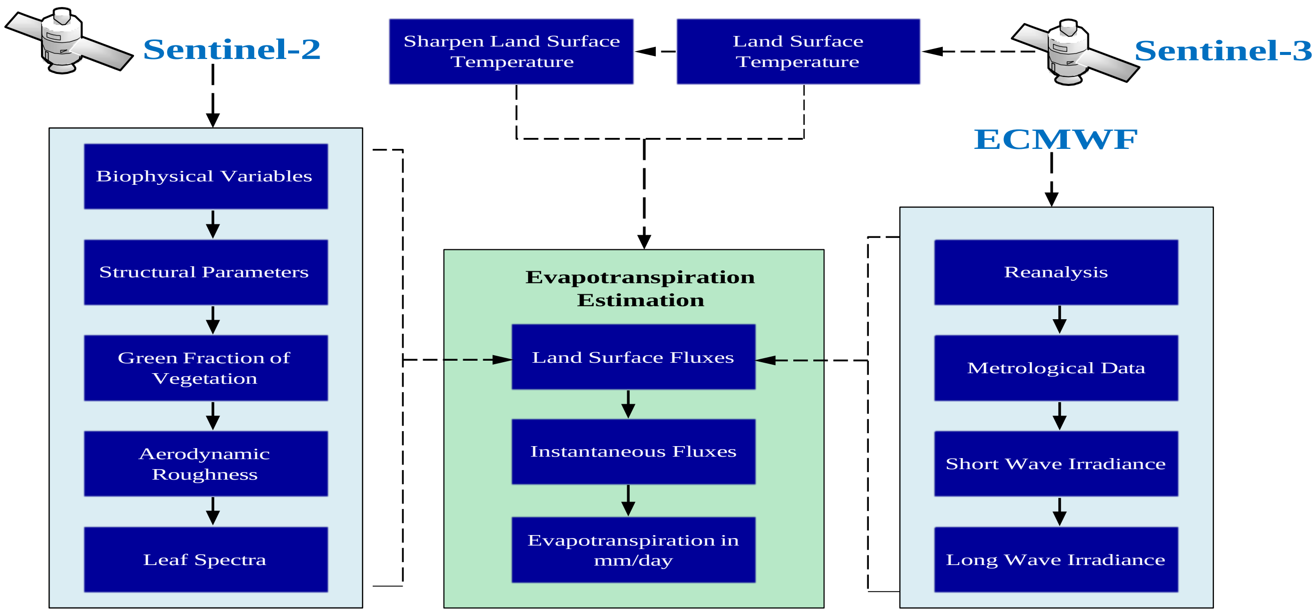


Fig. 3. Processing steps of evapotranspiration estimation

B. Utilizing Sentinel-2 for Evapotranspiration Estimation

The ESA Sentinel-2 satellite provides high-resolution multispectral level-2A imagery with 13 spectral channels. Deployed as a constellation of satellites, Sentinel-2 ensures a 5-day revisit time, crucial for applications like land-use mapping. However, real-time data is not available, as Sentinel-2 data is collected every five days. For ETa estimation, Sentinel-2 imagery in visible, near-infrared, and thermal infrared bands is preprocessed using atmospheric correction algorithms. Reflectance values for each band are calculated, adjusted for bidirectional reflectance, and further corrected for terrain effects to obtain surface reflectance. Vegetation indices like NDVI and EVI are then derived, indicating vegetation health. These reflectance values and indices are integrated into ETa models, which also consider factors such as land cover and meteorological data. The formula to calculate remote sensing bands is given in (1).

$$Reflectance = \frac{DN \times Gain}{\cos(Solar\ Zenith\ Angle)} \quad (1)$$

Where, Reflectance is the calculated reflectance, DN is the Digital Number from the satellite sensor, and Gain is radiometric calibration factor for each band.

Four key parameters are derived from preprocessed Sentinel-2 data that are: Reflectance, Biophysical data, Sun Zenith Angle, and Mask data. For the Mask data is used helps generate elevation and land cover maps, while biophysical characteristics influence reflectance and transmittance. By integrating biophysical data with the sun's zenith angle, the proportion of green vegetation is calculated, offering insights into vegetation health. These biophysical data, including vegetation indices like NDVI, are essential for monitoring Earth's surface and ecosystem health. The formula for NDVI is given in (2).

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (2)$$

Where, NIR is the reflectance in the near-infrared band and Red is the reflectance in red band.

C. Utilizing Sentinel-3 for Evapotranspiration Estimation

The methodology for estimating ETa using Sentinel-3 data unfolds in several stages. The initial preprocessing phase in the SNAP tool generates three essential outputs: Land Surface Temperature (LST), Observation Geometry, and the Sentinel-3 Mask. These outputs form the basis for further analysis. A crucial step, known as "Warp to Template," ensures the alignment and compatibility of these outputs, allowing for seamless integration of LST, Observation Geometry, and the Sentinel-3 Mask. LST, which represents the surface temperature of the Earth, is measured using thermal infrared sensors on Sentinel-3. In the context of ETa estimation, LST is a key input in energy balance models like SEBAL and SSEB. These models use LST data to distinguish between sensible heat (energy used for heating the atmosphere) and latent heat (energy used in water evaporation and plant transpiration). This distinction is vital for accurately estimating ET rates, offering insights into water consumption patterns. The processing also involves refining LST data by incorporating Sentinel-2 reflectance, Sentinel-3 LST, and templating information. These adjustments enhance the accuracy and reliability of the LST data, ensuring precise ETa analyses. This refined data is crucial for generating dependable insights into ETa patterns and environmental conditions.

D. European Centre for Medium-Range Weather Forecasts

For the estimation of ETa, European Centre for Medium-Range Weather Forecasts (ECMWF) were also utilized in this research. Because it offers reliable weather and climate data, such as temperature, humidity, and radiation, which are essential for calculating ETa. This data helps us understand how much energy and water move between the Earth's surface and atmosphere, allowing us to estimate how much water crops are losing through evaporation and transpiration. The ECMWF plugin in SNAP combines this weather data with satellite information from Sentinel-2 and Sentinel-3, improving the accuracy of our ETa calculations.

V. Results

The estimated computed ETc values closely match the average ETa values, as shown in Table II. Calculated ET Mean vs Actual ET. However, inconsistencies in irrigation practices across farms lead to significant discrepancies. Farmers continue to irrigate at fixed times without considering ETc values or the plant's varying water needs at different growth stages. Factors like human error, insufficient rainfall, and inefficient irrigation techniques contribute to the gap between computed and required ET. These findings highlight the need for more precise irrigation management systems that account for the tobacco crop's growth stages and optimize water use. The discrepancies are further visualized in fig. 4.

TABLE II. ET_C vs Actual Mean Results

Stages	***DOY***	***ETc (mm/day)***	***Actual ETa mean***
Initial Season	17-Mar-2024	1.57	3.0114
	23-Mar-2024	1.79	3.6067
	31-Mar-2024		3.6534
Development Season	12-Apr-2024	2.22	4.2589
	16-Apr-2024	3.38	3.8548
	25-Apr-2024	5.40	4.0239
	29-Apr-2024		4.6262
	06-May-2024	7.55	4.9268
Mid-Season	08-May-2024		5.7598
	12-May-2024	8.76	6.4234
	20-May-2024		6.3512
	26-May-2024	9.10	6.7209
	06-Jun-2024	9.08	6.2554
Late Season	14-Jun-2024	8.42	6.4237
	23-Jun-2024	6.98	8.1510
	30-Jun-2024		7.8543

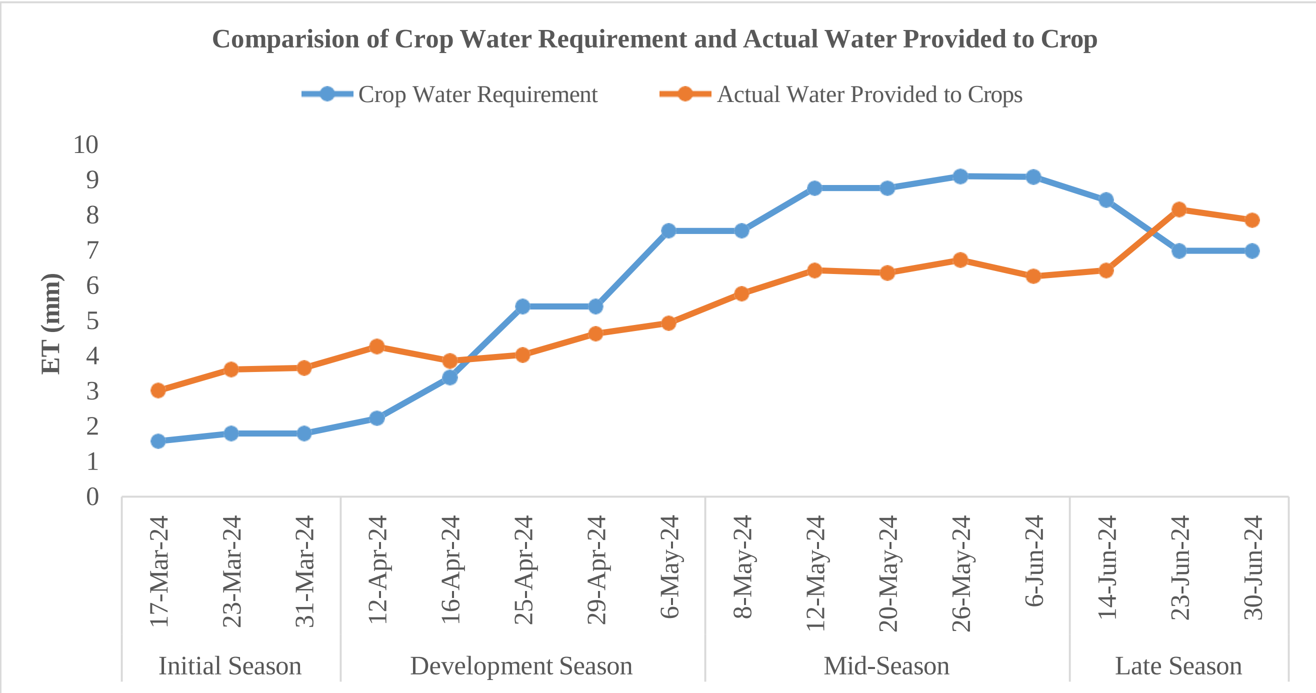


Fig. 4. Calculated ETc mean vs actual ETa

VI. Conclusion

Evapotranspiration (ET) plays a significant role in crop water requirements and irrigation scheduling. In this regard, we conducted research on the estimation and analysis of actual ETa for tobacco crops in the Mandani region of the Charsadda district. Modern tools, such as European Space Agency (ESA) satellites such as Sentinel-2 and Sentinel-3, and ECMWF weather data, were used for this purpose. ETa was estimated for four growing stages of tobacco, highlighting a pattern of excessive irrigation during the initial and development phases, coupled with irregular scheduling in the development and mid-season stages. The study further concluded that farmers require proper scientific knowledge regarding ET to optimize irrigation practices.

Acknowledgement

This research was conducted under the umbrella of National Center for Big Data and Cloud Computing (NCBC), University of Engineering and Technology Peshawar.

References

[1] S. Ahmad, M. N. A. Abbasi, and A. Ullah, "Is the tobacco production profitable in district Swabi: An empirical analysis," Human Nature Journal of Social Sciences, vol. 3, no. 3, pp. 1–11, 2022.

[2] S. Nasreen and M. A. Ashraf, "Inadequate supply of water in agriculture sector of Pakistan due to depleting water reservoirs and redundant irrigation system," Water Conservation & Management, vol. 5, no. 1, pp. 13–19, 2020.

[3] Kurt, D. and Kinay, A., 2021. Effects of irrigation, nitrogen forms and topping on sun cured tobacco. Industrial Crops and Products, 162, p.113276.

[4] M. H. Shahrajabian and W. Sun, "A review of lysimeter studies and experiments by considering agricultural production," Journal of Stress Physiology & Biochemistry, vol. 20, no. 2, pp. 114–132, 2024.

[5] M. Jain, P. Rao, A. K. Srivastava, S. Poonia, J. Blesh, G. Azzari, A. J. McDonald, and D. B. Lobell, "The impact of agricultural interventions can be doubled by using satellite data," Nature Sustainability, vol. 2, no. 10, pp. 931–934, 2019.

[6] X. Z. Hao, J. Wang, H. Ye, A. M. U. Hamad, and S. Wang, "The Bowen ratio of an alpine grassland in Three-River Headwaters, Qinghai-Tibet Plateau, from 2001 to 2018," Journal of Resources and Ecology, vol. 12, no. 3, pp. 305–318, 2021.

[7] Z. Wang, Y. Wu, Q. Cao, Y. Shen, and B. Zhang, "Modeling the coupling processes of evapotranspiration and soil water balance in agroforestry systems," Agricultural Water Management, vol. 250, p. 106839, 2021.

[8] N. Bambach et al., "The tree-crop remote sensing of evapotranspiration experiment (T-REX): A science-based path for sustainable water management and climate resilience," Bulletin of the American Meteorological Society, vol. 105, no. 1, pp. E257–E284, Jan. 2024.

[9] S. Koley and C. Jeganathan, "Estimation of the green and blue water footprint of Kharif rice using remote sensing techniques: A case study of Ranchi," in Proc. 2020 IEEE India Geoscience and Remote Sensing Symposium (InGARSS), Ranchi, India, Dec. 2020, pp. 1–4.

[10] A. Reyes-González, J. Kjaersgaard, T. Trooien, C. Hay, and L. Ahiablame, "Estimation of crop evapotranspiration using satellite remote sensing-based vegetation index," Advances in Meteorology, vol. 2018, p. 4525021, 2018.

[11] M. Hamza, N. Minallah, and W. Khan, "Operational model based regional estimation using remote sensing data," International Journal of Innovations in Science & Technology, vol. 6, no. 1, pp. 185–200, 2024.

HIGHLY ISOLATED 4-PORT UWB MIMO ANTENNA FOR NEXT GENERATION COMMUNICATION SYSTEM

Anila Hanif
Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
anilahanif@uetpeshawar.edu.pk

Shahid Bashir
Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
shahid.bashir@uetpeshawar.edu.pk

Anam Hanif
Electrical Engineering
Wah Engineering College, University of Wah
Wah Cantt, Pakistan
anamhanif.ee@gmail.com

***Abstract*—This paper introduces the design, optimization, and performance analysis of a miniature four-port ultra-wideband (UWB) MIMO antenna for next-generation high-frequency communication systems. The designed antenna is built on a Rogers RT Duroid 5880 substrate and functions efficiently within 12.5–55 GHz, which is ideal for millimeter-wave 5G applications. A four-stage step-by-step design process is employed to design a single antenna element optimized for wideband operation and efficient impedance matching.Parametric studies of feedline length, inset depth, and ground adjustments ensure improved bandwidth and robust radiation patterns. For MIMO configuration, four orthogonally located radiating elements reduce mutual coupling. To further enhance isolation, a centrally placed plus-shaped decoupling resonator is incorporated, which decreases mutual coupling substantially, especially at lower frequencies and improves the overall antenna performance. Simulation results indicate excellent impedance matching, low envelope correlation coefficient (ECC < 0.004), and high diversity gain (10 dB). The antenna also shows stable radiation patterns and high efficiency (>85 %) over the operating band. These results validate that the suggested MIMO antenna exhibits excellent isolation, miniaturization, and wideband features, which make it a good candidate for integration into next-generation UWB and mm-wave MIMO systems.**

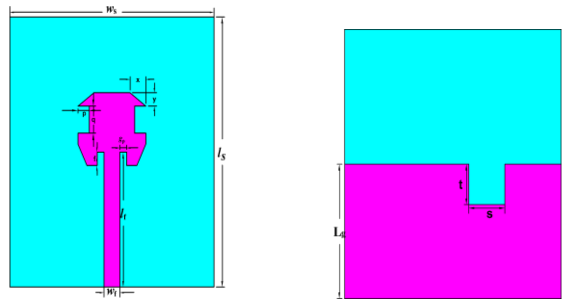
***Index Terms*—UWB antenna, MIMO antenna, Mutual coupling, Millimeter wave, Wireless communication**

I. Introduction

The evolution of wireless communication has accelerated the development of technologies like 5G, which demand high data rates, low latency, and efficient spectrum use [1]–[5]. UWB and MIMO antennas are key enablers due to their ability to enhance channel capacity and mitigate multipath fading [6], [7]. As mm-wave frequencies become central to next-generation networks, MIMO designs must address challenges like mutual coupling and impedance matching [8]. Techniques such as DGS [9], EBG [10], and parasitic elements [11] have been used to improve performance. Their combination [12], [13] and advanced decoupling networks [14] have shown promise in optimizing isolation and gain in compact MIMO antennas.

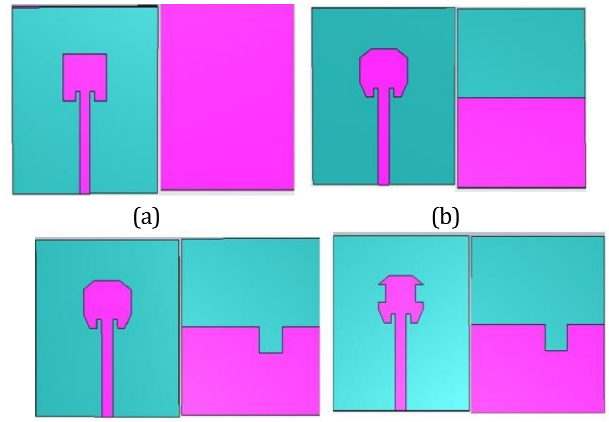
Recent advancements in MIMO antenna design for 5G and millimeter-wave applications have demonstrated high isolation and efficiency. Palm tree-shaped and stepped-shaped antennas achieved isolation above 60 dB and 20 dB, respectively, with wide bandwidth and high radiation efficiency [12], [13]. Tri-band and wideband designs using passive decoupling showed ECC below 0.007 and efficiency over 90 % [14], [15]. A compact 4-port MIMO antenna using RT/duroid5880 with a defected ground structure achieved over 34.2 dB isolation and 8.72 dBi gain [16]. Further developments in MIMO antenna design include a compact antenna operating from 26 to 40 GHz, offering an ECC below 0.5 and isolation better than -15 dB for effective diversity performance [17]. Additionally, an elliptically inspired microstrip antenna for V-band 5G applications demonstrated high gain (8–10.1 dBi), low ECC (0.006), and reduced coupling in its 4-port configuration, supporting high-speed communication [18]. Moreover, a compact 28 GHz MIMO antenna utilizing polarized diversity and fabricated on a Roger RT5880LZ substrate achieved a high gain of 6 dBi, 94 % radiation efficiency, 9 GHz bandwidth, and excellent isolation over 24 dB without additional decoupling structures [19]. Similarly, a UWB MIMO antenna designed for 25–50 GHz applications employed a defected ground structure and orthogonal element orientation, resulting in high isolation (better than −10 dB), low ECC (< 0.005), and strong diversity gain (10 dB) [20].

A compact MIMO antenna designed for 28 GHz applications utilized polarized diversity features and was fabricated on a Roger RT5880LZ substrate with a 0.25 mm thickness. This



(a) (b)

Fig. 1: Single element design (a) Top view (b) Bottom view



(a) (b)

(c) (d)

Fig. 2: Design steps (a) 1st step (b) 2nd step (c) 3rd step (d) 4th

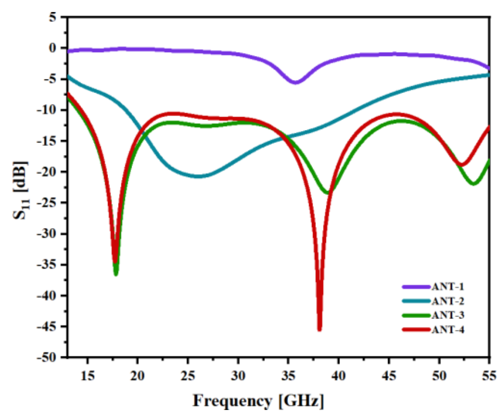


Fig. 3: S_{11} of the different design steps

design, incorporating elliptically slotted radiators, achieved a maximum gain of 6 dBi, radiation efficiency of 94%, and wide bandwidth of 9 GHz (22.2–31.4 GHz). Additionally, it demonstrated excellent isolation (>24 dB) without requiring additional decoupling structures [19]. A UWB MIMO antenna for millimeter-wave 5G applications was also introduced, covering a 25–50 GHz bandwidth with a fractional bandwidth of 66%. The design employed a defected ground structure (DGS) and orthogonal element orientation to minimize mutual coupling, achieving high isolation (<−10 dB), low ECC (<0.005), and strong diversity gain (≈10 dB) [20].

Furthermore, a study aimed at improving MIMO performance at 30 GHz utilized Curved Edges (CE), Defected Ground Structure (DGS), and Band Gap Structure (BGS) to enhance isolation from −21.4 dB to −27.2 dB, achieving a gain of 7.5 dBi and maintaining S_{11} at −35 dB [21]. While these methods effectively reduced surface currents and improved mutual coupling, the integration of UWB with MIMO still poses challenges due to the trade-off between achieving high isolation and maintaining a compact, cost-effective design.

This work aims to develop a compact four-port UWB-MIMO antenna with enhanced isolation and wideband characteristics. The design incorporates a novel decoupling structure to minimize mutual coupling, improving isolation, efficiency, and gain without increasing size. The goal is to achieve an optimal trade-off between compactness and performance, contributing to high-performance antenna solutions for future communication networks.

The structure of the paper includes Section II, which outlines the design and methodology of the proposed 4-port UWB MIMO antenna, covering isolation and optimization techniques. Section III presents simulation results and performance metrics, while Section IV concludes the study.

II. Design and Methodology

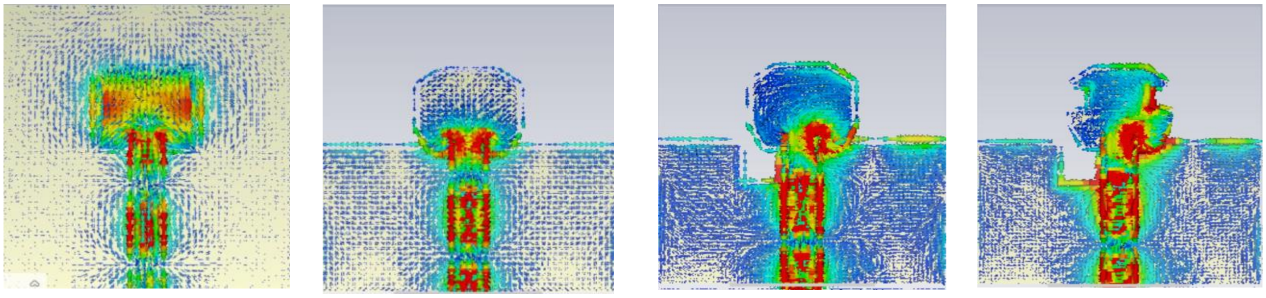
A. Single Element Design

For ultra-wideband (UWB) applications, an antenna is designed using Rogers RT Duroid 5880 for its high-frequency and low-loss properties. The 10×9 mm^2 single element, as shown in Fig. 1 followed a four-step design process to achieve the desired resonance frequency while ensuring UWB performance for advanced wireless communication systems.

TABLE I: Parametric Table

Parameter	Value (mm)
l_s	10
w_s	9
l_f	5
w_f	0.7
x	0.7
y	0.5
p	0.5
q	1
L_g	5
s	1.5
t	1.5
g_p	0.3
f_i	0.5

The antenna design process is detailed in Figures 2(a-d), with reflection coefficient results shown in 3 and current distribution in Fig. 4(a-d). The initial square microstrip patch antenna (Fig. 2a) suffers from poor impedance matching and weak radiation, as seen in 3 and Fig. 4a. Improvements made by truncating the ground plane and etching patch corners (Fig.



(a) (b) (c) (d)

Fig. 4: Current distribution (a) step 1 (b) step 2 (c) step 3 (d) step 4

2b) enhance impedance matching and radiation efficiency, confirmed by 3 and Fig. 4b. Further optimization with a rectangular slot (Fig. 2c) reduces reflection loss and improves surface current distribution (Fig. 4c). Finally, adding side slots (Fig. 2d) extends the UWB response and ensures efficient radiation, as validated in Fig. 3 and Fig. 4d. Table I provides the design specifications for the antenna.

The resonant frequency and impedance characteristics of a patch antenna are primarily determined by its length and width. The patch width w_p is calculated using

$$w_p = \frac{c}{2f_o}\sqrt{\frac{\varepsilon_r + 1}{2}} \tag{1}$$

where c is the speed of light, f_0 is the resonant frequency, and ε_r is the substrate's relative permittivity. The effective dielectric constant ε_{eff} is given by

$$\varepsilon_{\text{eff}} = \frac{\varepsilon_r + 1}{2} + \frac{\varepsilon_r - 1}{2}\left(1 + \frac{12h}{w_p}\right)^{-1} \tag{2}$$

where h is the substrate thickness. The actual patch length, considering fringing effects, is

$$L_{\text{eff}} = \frac{c}{2f_0\sqrt{\varepsilon_{\text{eff}}}} \tag{3}$$

with the length correction.

B. Parameteric Analysis

The study analyzes the impact of inset depth (f_i), gap (g_p), feed line length (l_f), and ground length (L_g) on a UWB MIMO antenna's performance, showing optimal impedance matching and wideband response (15-55 GHz) with f_i = 0.5 and l_f = 5 (Fig. 5(a-d)). The study highlights the need for fine-tuning these parameters to achieve desired operating characteristics.

C. Configuration of UWB MIMO Antenna

The proposed ultra-wideband (UWB) MIMO antenna, as shown in Fig. 6, is designed using a Rogers RT Duroid 5880 substrate with a relative permittivity of ϵ_r = 2.2, a loss tangent of δ = 0.0009, and a thickness of 0.254 mm. It comprises four radiating elements arranged in an orthogonal pattern to enhance port isolation without employing additional decoupling techniques. The spacing between adjacent patches is 10.3 mm, and the antenna has a compact size of 20 × 20 × 0.254 mm. All simulations were carried out using CST Microwave Studio Suite.

D. S-Parameters Response and Mutual Coupling of MIMO

The S-parameter response of the proposed 4-port MIMO antenna is illustrated in Fig. 7a. The return loss parameters (S_{11}, S_{22}, S_{33}, S_{44}) remain well below −10 dB across the operational bandwidth, indicating effective impedance matching at all ports. Multiple resonance points confirm the preservation of wideband behavior, while uniform resonance across ports reflects stable impedance and consistent excitation. These characteristics confirm that the multi-port design retains the performance of the single-element antenna for mm-wave UWB applications. Mutual coupling analysis, shown in Fig. 7b, reveals higher coupling around 19–20 GHz due to the lack of a decoupling structure. The isolation parameters (S_{12}, S_{13}, S_{14}) exceed −15 dB at lower frequencies, indicating noticeable coupling, though improved isolation is observed at higher frequencies.

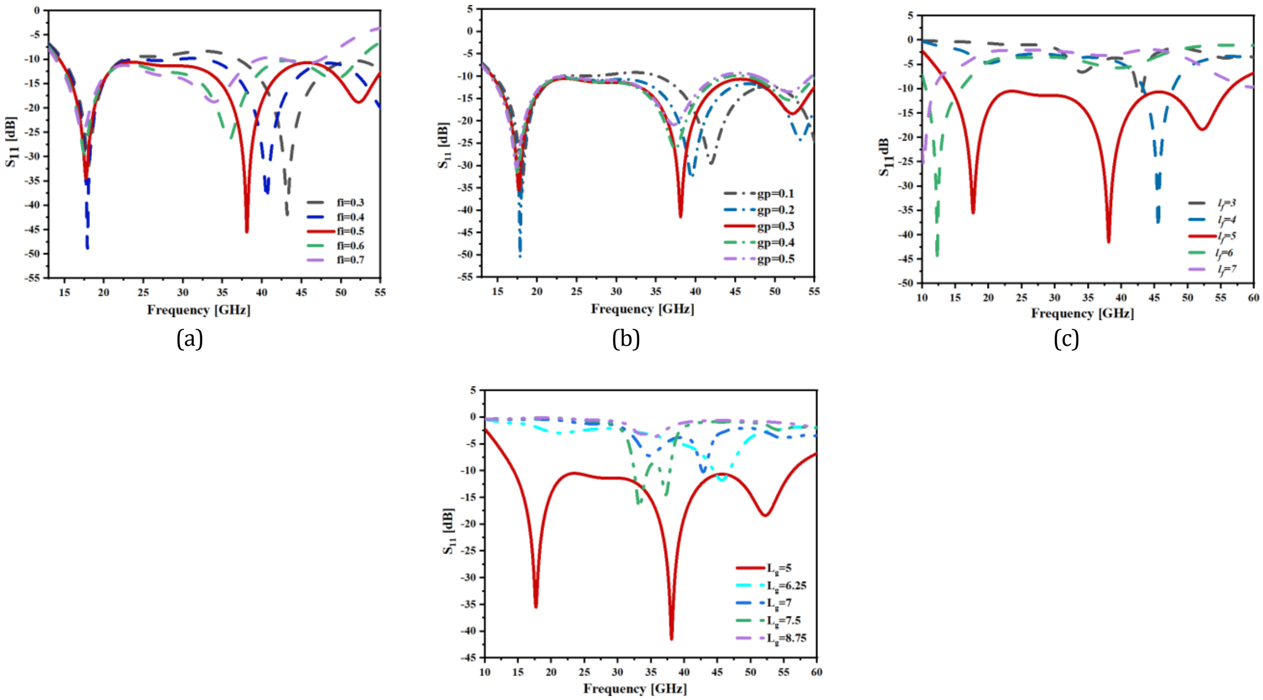
III. Simulation and Results

A. Proposed Decoupling Technique

Decoupling resonators play a crucial role in mitigating mutual coupling in MIMO antennas, enhancing isolation and preserving signal integrity by operating at the system's resonance frequency [22]. A plus-shaped decoupling resonator, centrally positioned in the MIMO system (Fig. 8a), serves as a selective filter to suppress undesired coupling, improving impedance bandwidth from 12.5–55 GHz to 15–55 GHz. A parametric study varying the resonator length (v parameter) from 3.6mm to 7.6mm showed that a length of 5.6mm resulted in an improved reflection coefficient of approximately -50 dB, as shown in Fig. 8c.

B. Surface Current Density

The surface current distribution is analyzed to evaluate the isolation characteristics of the proposed MIMO antenna design at two frequencies: 20 GHz and 36 GHz, as shown in Fig. 9. At 20 GHz, the distribution without any decoupling element shows strong coupling between the antenna elements. By introducing decoupling resonators, it is observed that the plus-shaped resonator significantly reduces mutual coupling.

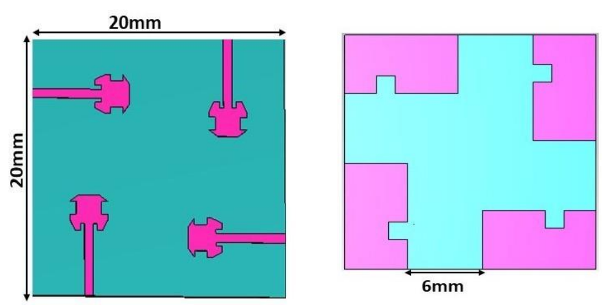


(a) (b) (c)

(d)

Fig. 5: Optimization of various parameters (a) changing f_l (b) changing g_P, (c) changing l_f (d) changing L_g

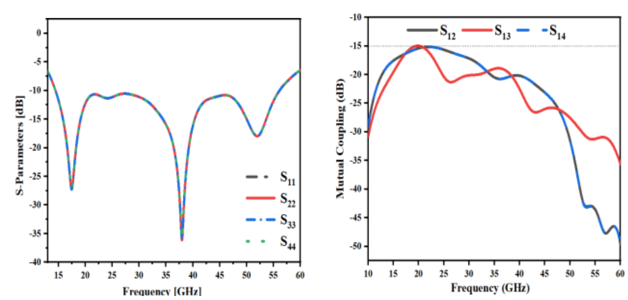
.



(a) (b)

Fig. 6: Orthogonal arrangement (a). Front View (b). Back View

.



(a) (b)

Fig. 7: (a)Reflection coefficients of MIMO with orthogonal arrangement (b)Mutual coupling without decoupling

.

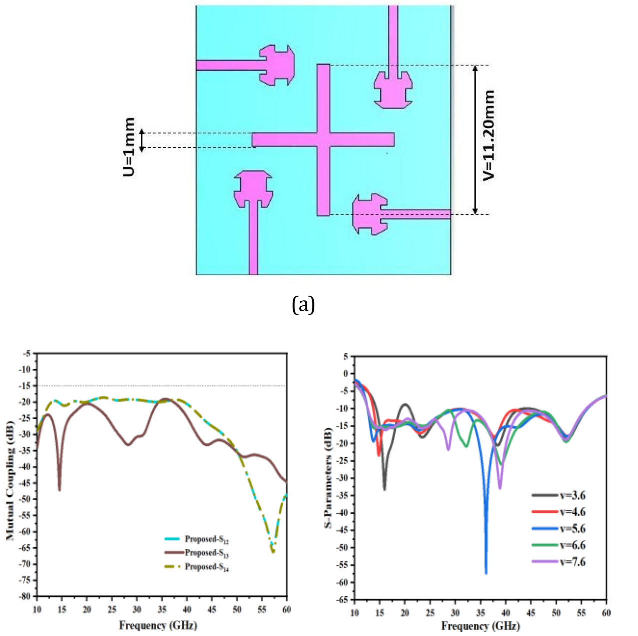
Similarly, at 36 GHz, the surface current distributions with and without decoupling elements, depicted in Fig. 9(c,d), further validate the superior performance of the plus-shaped resonator.

C. Far Field Measurement

The radiation patterns of the proposed MIMO antenna were analyzed at 25 GHz, 36 GHz, and 40 GHz to evaluate its performance across the operational bandwidth. At 25 GHz, the pattern (Fig. 10a) shows a broad lobe, indicating stable radiation characteristics at lower frequencies. At 36 GHz (Fig. 10b), the radiation pattern becomes more defined, showing improved directivity. At 40 GHz (Fig. 10c), the pattern evolves with distinct lobes and high directivity, confirming the antenna's suitability for mm-wave applications. These consistent radiation characteristics validate the design's effectiveness for wideband MIMO systems in high-frequency communication scenarios.

D. Efficiency and Gain

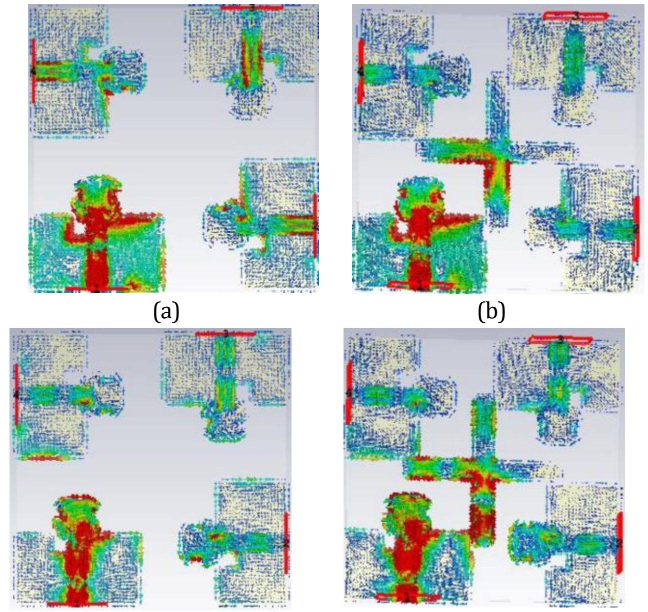
MIMO antenna's gain and overall efficiency in Fig. 11, provides a comprehensive picture of its performance. The antenna exhibits notable directional radiation characteristics with a peak gain of 6.95 dBi. Moreover, there is minimal power loss since the overall efficiency remains above 85% during the whole operating bandwidth.



(a)

(b) (c)

Fig. 8: (a) UBW MIMO with plus shaped resonator (b) Effect of plus shaped resonator on mutual coupling (c) Effect of variation in v parameter



(a) (b)

(c) (d)

Fig. 9: Current Distribution at 20 GHz (a) without decoupling (b) with Plus-shape resonator (c) without decoupling at 36GHz (d) with plus-shape resonator at 36 GHZ.

E. MIMO Antenna Performance

MIMO technology requires integrating multiple antennas in limited PCB space, emphasizing performance metrics like Envelope Correlation Coefficient (ECC), Diversity Gain (DG), and Total Active Reflection Coefficient (TARC), along with standard parameters such as bandwidth, radiation efficiency, and gain for performance evaluation [23], [24].

1) Envelope Correlation Coefficient (ECC): The Envelope Correlation Coefficient (ECC) quantifies signal independence for effective diversity and is typically calculated using radiation patterns or S-parameters [25], [26]. The S-parameter-based ECC is given by Equation (1), where η_1 and η_2 are antenna efficiencies:

$$ECC = \frac{|S_{11}S_{12}^{*} + S_{21}S_{22}^{*}|^2}{(1 - |S_{11}|^2 - |S_{21}|^2)(1 - |S_{22}|^2 - |S_{12}|^2)\eta_1\eta_2} \quad (4)$$

The proposed MIMO antenna achieved ECC values below 0.01, which further decreased to less than 0.004 using a plus-shaped decoupling resonator, significantly improving isolation and diversity (Fig. 12).

2) Diversity Gain (DG): Diversity Gain (DG) enhances signal reliability in MIMO systems by improving SNR and is estimated via ECC or CDF methods [27]. It is calculated using

$$DG = 10 \times \sqrt{1 - ECC^2} \quad (5)$$

As shown in Fig. 13, the proposed antenna achieves near-ideal DG (10 dB) after adding the decoupling element, ensuring robust diversity performance.

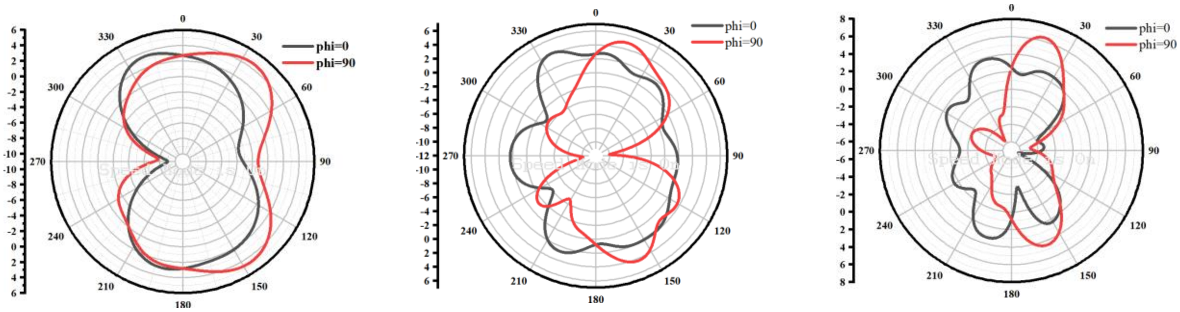
3) Total Active Reflection Coefficient (TARC): In particular, for a four-port network, TARC is given by [28]

$$TARC = \sqrt{\frac{1}{2}\left(|S_{11} + S_{12}e^{j\theta}|^2 + |S_{21} + S_{22}e^{j\theta}|^2 + |S_{31} + S_{32}e^{j\theta}|^2 + |S_{41} + S_{42}e^{j\theta}|^2\right)} \quad (6)$$

The proposed design exhibits TARC values below −10 dB across the operational bandwidth (Fig. 14), indicating minimal reflection and efficient power utilization.

IV. Conclusion

This paper presented the design and analysis of a 4-port UWB MIMO antenna optimized for next-generation wireless communication. Operating across 12.5–55 GHz, the antenna demonstrates suitability for 5G millimeter-wave and high-speed data applications. A plus-shaped decoupling resonator effectively reduces mutual coupling, enhancing isolation, particularly in the 19–20 GHz range. Simulation results confirm excellent performance, with ECC < 0.004, DG > 9.98, isolation > 15 dB, and TARC < -10 dB. The proposed antenna ensures wideband operation, improved isolation, and efficient impedance matching, making it a promising solution for advanced wireless systems.



(a) (b) (c)

Fig. 10: Radiation Pattern (a) 25GHz (b) 36GHz (c) 40GHz

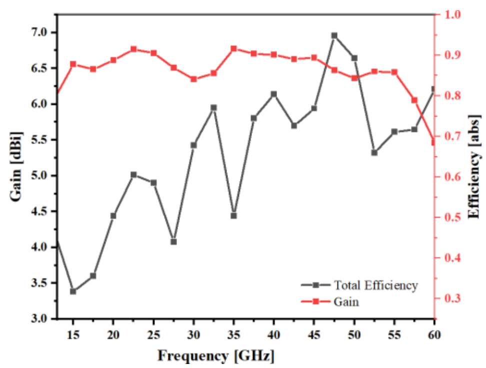


Fig. 11: Gain and efficiency of MIMO antenna

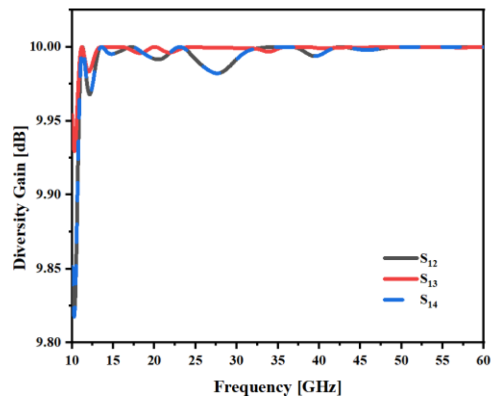


Fig. 13: DG of proposed design

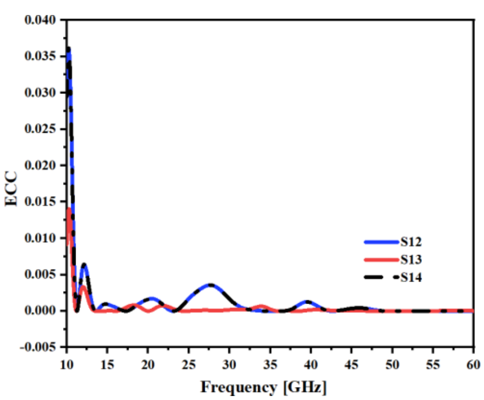


Fig. 12: ECC of proposed design

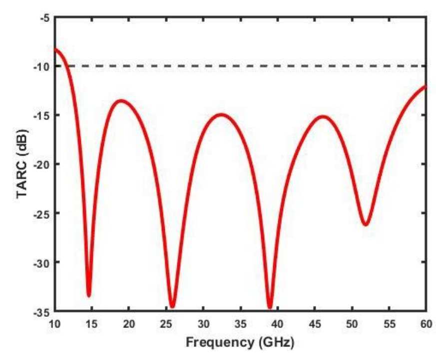


Fig. 14: TARC characteristics of the stated MIMO antenna

References

[1] C. Sudhamani, M. Roslee, J. J. Tiang, and A. U. Rehman, “A survey on 5g coverage improvement techniques: Issues and future challenges,” *Sensors*, vol. 23, no. 4, p. 2356, 2023.

[2] Y. O. Imam-Fulani, N. Faruk, O. A. Sowande, A. Abdulkarim, E. Alozie, A. D. Usman, K. S. Adewole, A. A. Oloyede, H. Chiroma, S. Garba, *et al.*, “5g frequency standardization, technologies, channel models, and network deployment: Advances, challenges, and future directions,” *Sustainability*, vol. 15, no. 6, p. 5173, 2023.

[3] M. K. Hossain, M. M. Rana, and M. R. Haider, “Spectrum-efficient analog pulse index modulation for high-volume wireless data telemetry,” *IEEE Internet of Things Journal*, vol. 10, no. 11, pp. 9556–9564, 2023.

[4] M. Malekzadeh, “Performance prediction and enhancement of 5g networks based on linear regression machine learning,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2023, no. 1, p. 74, 2023.

[5] A. Kumar, N. Gaur, and A. Nanthaamornphong, “Improving the latency for 5g/b5g based smart healthcare connectivity in rural area,” *Scientific Reports*, vol. 14, no. 1, p. 6976, 2024.

[6] A. S. Abd El-Hameed, M. G. Wahab, N. A. Elshafey, and M. S. Elpeltagy, “Quad-port uwb mimo antenna based on lpf with vast rejection band,” *AEU-International Journal of Electronics and Communications*, vol. 134, p. 153712, 2021.

[7] X. Zhao, S. Riaz, and S. Geng, “A reconfigurable mimo/uwb mimo antenna for cognitive radio applications,” *IEEE access*, vol. 7, pp. 46739–46747, 2019.

[8] R. S. D and S. Azeemuddin, “A comprehensive review of high voltage wideband and ultra-wide band antennas for iemi applications,” *Engineering Research Express*, vol. 3, p. 012001, jan 2021.

[9] A. Kumar and A. Kumar, “Defected ground structure based high

gain, wideband and high diversity performance quad-element mimo antenna array for 5g millimeter-wave communication.," *Progress In Electromagnetics Research B*, vol. 101, 2023.
[10] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Transactions on information theory*, vol. 49, no. 5, pp. 1073–1096, 2003.
[11] G. Casu, L. Tutaˇ, I. Nicolaescu, and C. Moraru, "Some aspects about the advantages of using mimo systems," in *2014 22nd Telecommunications Forum Telfor (TELFOR)*, pp. 320–323, IEEE, 2014.
[12] F. Kiouach, B. Aghoutane, S. Das, M. E. Ghzaoui, T. Islam, and B. T. P. Madhav, "A high isolation wideband palm tree-shaped printed 4× 4 mimo antenna for 5g mm-waves applications," *AEU-International Journal of Electronics and Communications*, vol. 179, p. 155322, 2024.
[13] P. Tiwari, M. Kaushik, A. Shastri, P. Ranjan, and V. Gahlaut, "A 28 ghz wideband planar stepped-shaped mimo antenna with isolating metallic sheet for 5g millimeter wave communications," *AEU-International Journal of Electronics and Communications*, vol. 184, p. 155411, 2024.
[14] A. Ali, M. E. Munir, M. M. Nasralla, M. A. Esmail, A. J. A. Al-Gburi, and F. A. Bhatti, "Design process of a compact tri-band mimo antenna with wideband characteristics for sub-6 ghz, ku-band, and millimeter-wave applications," *Ain Shams Engineering Journal*, vol. 15, no. 3, p. 102579, 2024.
[15] K. V. Prasad, N. Venkateswari, M. Sandhyarani, P. V. Kumar, and K. M. Lakshmi, "Isolation and bandwidth enhancement of compact wideband mimo for sub-6 ghz, ku-band and millimeter-wave with uwb applications.," *Progress in Electromagnetics Research C*, vol. 148, 2024.
[16] C. Gu¨ler and S. E. Bayer Keskin, "A novel high isolation 4-port compact mimo antenna with dgs for 5g applications," *Micromachines*, vol. 14, no. 7, p. 1309, 2023.
[17] M. E. Munir, M. M. Nasralla, and M. A. Esmail, "Four port tri-circular ring mimo antenna with wide-band characteristics for future 5g and mmwave applications," *Heliyon*, vol. 10, no. 8, 2024.
[18] W. A. Ali, M. A. Abdelghany, and A. A. Ibrahim, "Wideband and high gain elliptically-inspired 4× 4 mimo antenna for millimeter wave applications," *Heliyon*, vol. 10, no. 20, 2024.
[19] N. Farooq, K. Muzaffar, and S. Malik, "Compact elliptical slot millimeter-wave mimo antenna for 5g applications," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 45, no. 9, pp. 765–788, 2024.
[20] M. A. Abbas, A. Allam, A. Gaafar, H. M. Elhennawy, and M. F. A. Sree, "Compact uwb mimo antenna for 5g millimeter-wave applications," *Sensors*, vol. 23, no. 5, p. 2702, 2023.
[21] M. A. Nassar, H. Y. Soliman, R. M. Abdallah, and E. A. Abdallah, "Improving mutual coupling in mimo antennas using different techniques," *Progress In Electromagnetics Research C*, vol. 133, pp. 81–95, 2023.
[22] A. Ramos, T. Varum, and J. N. Matos, "A review on mutual coupling reduction techniques in mmwaves structures and massive mimo arrays," *IEEE Access*, vol. 11, pp. 143143–143166, 2023.
[23] P. Fletcher, M. Dean, and A. Nix, "Mutual coupling in multi-element array antennas and its influence on mimo channel capacity," *Electronics Letters*, vol. 39, no. 4, pp. 342–344, 2003.
[24] R. G. Vaughan and J. B. Andersen, "Antenna diversity in mobile communications," *IEEE Transactions on vehicular technology*, vol. 36, no. 4, pp. 149–172, 1987.
[25] S. Blanch, J. Romeu, and I. Corbella, "Exact representation of antenna system diversity performance from input parameter description," *Electronics letters*, vol. 39, no. 9, pp. 705–707, 2003.
[26] M. Abdullah, S. H. Kiani, and A. Iqbal, "Eight element multiple-input multiple-output (mimo) antenna for 5g mobile applications," *IEEE Access*, vol. 7, pp. 134488–134495, 2019.
[27] M. Sharma, N. Choudhary, N. Kumar, S. Panda, and R. Kaushal, "A slotted hexagonal 4× 4 mimo antenna with tapered feed designed for high speed iot wireless applications," in *2021 6th International Conference on Signal Processing, Computing and Control (ISPCC)*, pp. 203–208, IEEE, 2021.
[28] S. ho Chae, W. il Kawk, S.-O. Park, and K. Lee, "Analysis of mutual coupling in mimo antenna array by tarc calculation," in *2006 Asia-Pacific microwave conference*, pp. 2090–2093, IEEE, 2006.

AI Based Resource Efficient Image Classifier for Skin Lesions

Salman Khan
Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
engrsalmankhan@uetpeshawar.edu.pk

Umar Sadique
Department of Computer Systems Engineering
University of Engineering and Technology
Peshawar, Pakistan
umar.sadique@uetpeshawar.edu.pk

Farhan Khan
Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
farhankhan@uetpeshawar.edu.pk

Sana Khan
Veterinary Research and Disease Investigation Center Abbottabad
Livestock, Fisheries and Cooperative Department (Research Wing)
KPK, Pakistan
snkhan12121@gmail.com

Atif Jan
Department of Electrical Engineering
University of Engineering and Technology
Peshawar, Pakistan
atifjan@uetpeshawar.edu.pk

***Abstract*—Skin cancer and other skin diseases pose a major health risk to humans. Timely diagnosis is crucial for successful treatment. Conventional diagnostic methods like clinical examination and histopathological assessment are time-consuming and require specialized skills, causing a delay in treatment. AI models can revolutionize this process. However, researchers has focused on either skin cancer or skin diseases, but this study covers multiclass classification by presents a novel dataset curated from publicly available datasets and new images gathered from mobile cameras. The dataset includes three classes of skin cancer and six classes of skin diseases, incorporating both mobile camera and dermoscopic images. We gathered 6820 skin lesion images, with 4957 from public datasets and 1863 new images added to enhance the dataset. Various deep learning models like VGG16, ResNet, DenseNet, MobileNet, and custom CNN were trained, but they performed poorly on mobile images despite success with dermoscopy images. A new classification model YOLOv11 was implemented for multiclass classification, achieving 97.5% overall accuracy with F1 score of 0.97503 and 99% accuracy per class for both dermoscopy and mobile images.**

***Index Terms*—Skin Cancer; Skin Diseases; Deep Learning; Multi-Class classification; AI; Medical Diagnosis; YOLOv11**

I. Introduction

Skin lesions are the most common health problems globally, which are quite challenging to the healthcare systems [1], affecting millions of people each year. Skin cancer refers to the uncontrolled and irregular development of skin cells, typically brought on by exposure to ultraviolet (UV) light from the sun or man-made sources like tanning beds. Ultraviolet radiations damage the DNA in skin cells, causing mutations that disrupt normal cell function. If not detected early, mutation may lead to skin cancer [2] The World Health Organization (WHO) reports that melanoma by itself leads to over 300,000 new cases and kills more than 60,000 people per year [3]. According to the Global Cancer Observatory (GCO) [4], mortality data for melanoma shows that Europe has the highest number of deaths with 26,180 cases, followed by Asia with 13,147 deaths, Latin America and the Caribbean report 5,842 deaths, Africa has 2,859 deaths and Oceania has the least number of deaths with 1,902 cases. Furthermore, the primary cause of cancer burden in other parts of the world is non-melanoma skin cancers such as squamous cell carcinoma and basal cell carcinoma. In addition to skin cancer other skin diseases like Eczema, psoriasis, Leishmaniasis, warts, and fungal infections are conditions that affect a large percentage of the population and results in pain, discomfort, and distress. Inefficient or delayed diagnosis of these diseases may cause serious complications, making early detection critical. Traditional diagnosis of skin conditions is carried out by clinical inspection, dermoscopy and histopathology. Although these are reliable techniques, they are time-consuming and require specialized skills [5]. Moreover, access to dermatologists is not widespread due to which diagnosis and treatments are delayed. With the tremendous progress made in AI and deep learning, computer-aided systems of detection and classification of cutaneous disease have emerged as a solution to this void. AI models can classify skin lesions accurately and can provide real-time diagnostic assistance and deliver dermatologic services to remote locations. Convolutional Neural Networks (CNNs) have displayed remarkable outcomes in classifying skin lesions using dermoscopic images. However, existing studies primarily focus on either skin cancer or skin diseases separately, with limited research on unified multi-class classification. Moreover, publicly available datasets, such as ISIC [6] and HAM10000 [7], predominantly contain dermoscopic images, which do not fully capture real-world variations in image quality and lighting conditions. To address these challenges, we developed a novel dataset that integrates both skin cancer and skin diseases images, incorporating images taken with mobile cameras and dermoscopy images. We gathered our dataset from several publicly available dataset of skin diseases and skin cancer, topped with new skin disease images and skin cancer images collected from different videos and images to increase its diversity. To attain our objective of multiclass classification We trained several state-of-the-art classification models, such as VGG16, ResNet, DenseNet, MobileNet, and a custom CNN. Although these models performed well with high accuracy on traditional dermoscopy images datasets but their performance declined remarkably when they were tested on real-life mobile images, indicating a need for stronger methods. To improve accuracy and generalizability, we trained YOLOv11 for multi-class classification of skin diseases and cancers. Which outperforms conventional methods, making it a viable solution for real-world dermatological diagnosis.

II. LITERATURE REVIEW

Skin cancer is a major global health issue, with steadily rising incidence rates over recent decades. Timely and correct detection is key to better patient prognoses and less invasive treatments [9]. Conventional skin cancer and skin diagnosis, which is largely dependent on dermatological experience, tends to be time-consuming and variable [10]. New possibilities for accurate skin lesion diagnosis and classification have been made possible by the advancement of artificial intelligence (AI), notably deep learning and computer vision. [9], [11]. Several deep learning models have been tried for the categorization of diseases and skin cancer, mostly using convolutional neural networks (CNNs) because of their image processing capabilities. [12], [13], [14], [15]. For instance, Ahmad Naeem et al. proposed a model that combines VGG16 with CNNs for multiclassification of melanoma, melanocytic nevi, basal cell carcinoma, and benign keratosis, with 96.91% accuracy on the ISIC 2019 dataset [14]. Such accuracy is superior to other pre-trained classifiers such as ResNet50, Inception v3, AlexNet, and VGG19. Ehsan Bazgir et al. [16] proposed an automated skin cancer classification into melanoma and non-melanoma categories by utilizing a deep neural network-based model with an optimized architecture of the InceptionNet, enhanced by data augmentation and preprocessing. Later, the same team created a model on 2637 images with an accuracy of 84.39% using the Adam optimizer and 85.94% when using the Nadam optimizer. Further emphasizing the efficacy of CNNs, Sobia Bibi et al. [13] designed MSRNet as a deep model with the capability to apply contrast enhancement, such as applying modified DarkNet-53 and DenseNet-201 models for achieving accuracy levels of 85.4% and 98.80% on ISIC2018 and ISIC2019, respectively. The research also suggests applying a genetic algorithm for choosing a hyperparameter, and marine predator optimization for choosing the features. Neven Saleh et al. [17] address the skin cancer classification problem by applying several CNNs, including AlexNet, Inception V3, MobileNet V2, and ResNet 50. Several machine learning classifiers are combined with optimization methods, one of which is the GWO. By combining these methods, a total of 51 different models were generated. To find the best model, the authors have used an MCDM approach, namely RAPS. Among these, the combination of AlexNet with GWO gives the highest classification accuracy of about 94.5% on the dataset ISIC 2017. Overfitting because of complex model combinations and dependence on a single dataset are some of the shortcomings regarding this study. Apart from VGG and ResNet, several other pre-trained models have been fine-tuned and adapted to classify skin cancer. Vipin Venugopal et al. [18] used a pre-trained EfficientNetV2-M, which surpassed other state-of-the-art learning-based models for binary and multiclass classification. The study emphasizes transfer learning and data augmentation techniques to achieve better performance, using a carefully curated dataset of 58,032 dermoscopic images. Similarly, Muhammad Zia Ur Rehman et al. adapted MobileNetV2 and DenseNet201 and included convolutional layers to obtain 95.50% accuracy in benign/malignant classification [19]. Pronab Ghosh et al. introduced SkinNet-16 and obtained an accuracy of around 99.19% for benign and malignant classification after preprocessing operations such as hair and background removal, image enhancement and feature extraction [20]. The quality and availability of training data are crucial for the accuracy of AI models classifying skin cancer. Limited publically available datasets, i.e., ISBI 2017, ISIC 2018, PH2, and ISIC 2019, have been adequately used for several comparisons of AI models [10], [11]. However, their heterogeneity across image quality, lesion types, and patient groups poses difficulties to the development of robust generalizable models. The lack of proper representation of certain skin types and lesion subtypes in public datasets may generate biased model performances and limit these models deployment in clinical practice. One of the main challenges to skin cancer classification is the natural class imbalance present in datasets, with benign lesions outnumbering malignant lesions by an enormous margin [21]. Such class imbalance tends to bias models based on deep learning towards the majority class and results in underperformance in detection of rarer, though in some cases potentially more serious, malignant lesions [21]. This was overcome by Talha Mahboob Alam et al. by using data augmentation to produce a dataset that is balanced and using AlexNet, InceptionV3, and RegNetY-320 for classification, and RegNetY-320 performed best on an equated dataset. Maryam Tahir et al. used SMOTE Tomek to handle class imbalance in their DSCC-Net model [22]. Other research has utilized oversampling and cost-sensitive learning [23] to counter the effects of class imbalance, resulting in overall better classification accuracy. The approach and its effectiveness vary based on the dataset and model utilized.

In this study, we take a systematic approach in solving the problem we have identified. Section III provides a summary of our methodology, such as the dataset employed and the system we propose. Section IV express the experimental setup, from data preprocessing to network training procedures and performance metrics. Section V explains a comprehensive analysis of the results, including discussion of the performance and implications of our proposed system. The paper concludes with Section VI, in which the main findings, contributions, and future work is described.

III. METHODOLOGY

A. Employed Dataset

In order to train a good multi-class skin cancer and skin disease classification model, we first collected a diverse dataset by merging publicly available sources along with the images we recently acquired. The dataset contains nine classes: three skin cancer classes (melanoma, basal cell carcinoma, and squamous cell carcinoma) and six skin disease classes (eczema, psoriasis, leishmanios, acne, tinea, and warts molluscum).

Our dataset was gathered from various public datasets, such as ISIC [6], HAM10000 [7], medimages Computer Vision Project [24], skin cancer Computer Vision Project [25], Skin diseases Segmentation Computer Vision Project [26], skin Computer Vision Project [27] and New Mod Computer Vision Project [28] and other images taken from videos and phone camera images. We initially gathered 3,957 skin disease images and 1,000 skin cancer images from public databases. For further increasing dataset diversity and model generalization, we added 1,000 additional skin disease images and 863 additional skin cancer images, thus making this one of the largest multi-class skin condition datasets. Random samples of Skin Cancer and Skin Disease from the dataset are shown in Figure 1 and Figure 2, which clearly display the diversity and heterogeneity of images utilized in the current study. A tabular summary of the class-wise distribution of images in the dataset is provided in Table I. This extended dataset is for deep learning model training, especially models that rely on classification-based models since it offers a more realistic representation of skin disorders as they would be seen in real-world clinical practice.

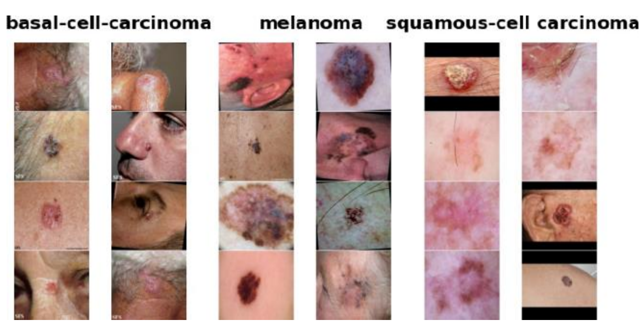


Figure 1: Random Samples of Skin Cancer Images

Figure 2: Random Samples of Skin Diseases Images

TABLE I: Distribution of Public Available and Self Captured Images across Classes

Class	Images Distribution Per Class		
	Total Images	Public Available	Self Captured
BCC	553	300	253
Melanoma	844	594	250
SCC	439	293	146
Acne	688	554	134
Eczema	1219	1070	149
Leishmanios	216	171	45
Tinea	591	470	121
Warts Molluscum	763	608	155
Psoriasis	1480	1084	396
Total	**6820**	**4957**	**1863**

B. YOLO-v11

The YOLOv11 [8] model is a significant improvement in the YOLO series, namely for classification, from the developments of the earlier versions including YOLOv8, YOLOv9, and YOLOv10. While YOLOv11 is conventionally linked to object detection, it can be used for classification with its efficient feature extraction and processing. Key building blocks like the C3K2 block and the C2PSA block as shown in Figure 3 enhance spatial information processing, with the C3K2 block using small 3x3 kernels to achieve maximum computation while improving feature representation, and the C2PSA block using attention mechanisms to focus on salient areas in an image, which is extremely helpful for detecting fine-grained information. In contrast to the detection model, the YOLOv11-cls (classification-specific model) lacks the neck (e.g., SPFF module), as multi-scale feature aggregation is not necessary in classification. The backbone directly extracts high-level features, which are then passed to the classification head, for an efficient and lightweight design. The classification head projects these features to class probabilities, enabling strong and accurate predictions with a fast and lightweight architecture, making YOLOv11 an efficient and state-of-the-art image classification model.

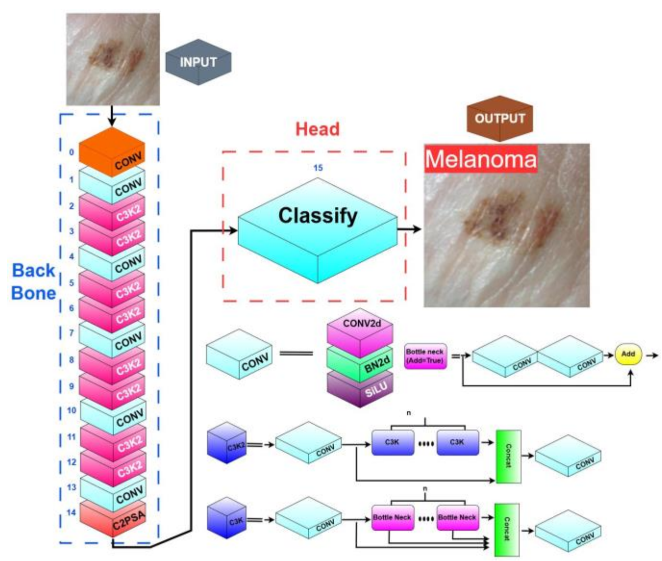


Figure 3: Yolo-v11 Object Classification Architecture

C. Proposed Framework

The proposed methodology follows a uniform approach for multi-class classification of skin cancer and skin disease with the YOLOv11 model. The skin lesion picture is taken first with a smartphone and then fed into the model for preprocessing. The process includes resizing, normalization, and data augmentation to make the model more robust.

After preprocessing, the image goes through the YOLOv11 Classification model, which Extract important features of skin conditions from the image. If the area's confidence that the model has recognized is less than 0.5 threshold, the image undergoes additional preprocessing and is re-evaluated. If the confidence value is equal to or greater than 0.5 threshold, the important features are map to classify into one of the nine classes (three skin cancers and six skin conditions).

The final output is class label of lesion, enabling precise diagnosis. The system ensures iterative refinement by re-processing uncertain cases until confident classification is achieved, improving diagnostic accuracy and real-world applicability. The system operational workflow is demonstrated in Figure 4.

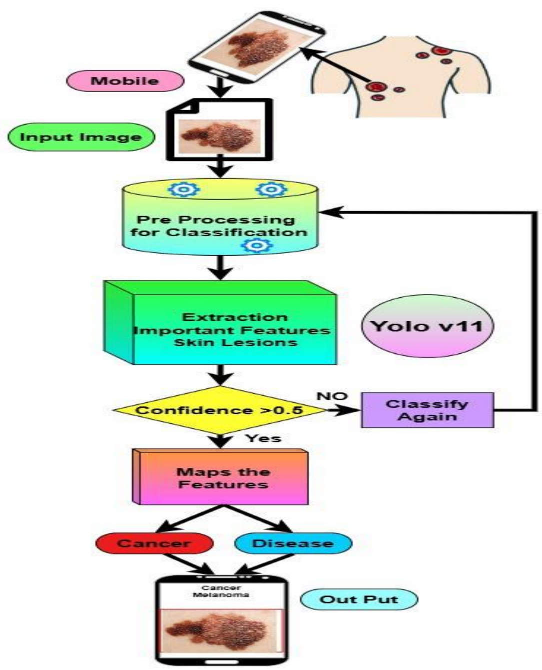


Figure 4: Block Diagram of Proposed Workflow

IV. Experimental Setup

A. Data Pre-processing

The data was distributed in a ratio of 70:15:15 among the training set, validation set and test set. The dataset was pre-processed extensively before training to provide good-quality input. This involved removal of duplicate, low-quality, and noisy images to enhance model performance. Rotation, flipping, brightness adjustment, contrast normalization and data augmentation techniques were utilized to increase dataset diversity and model resilience. Data labeling was done using manual renaming. The images were also resized and normalized according to the input needs of the YOLOv11 classification model.

B. Networks training

The system proposed in this research was developed and tested using an Intel Core i7-10700 computer running Windows 10 at 2.90 GHz, an NVIDIA GeForce RTX 3060 GPU with 12 GB of RAM, and 16 GB of RAM. The implementation was supported using Python 3.12, with the PyTorch framework serving as the primary deep-learning library.

The training parameters for classification are summarized in Table II.

TABLE II: Training Parameters

No.	Parameters for Classification Model	
	Hyperparameters	*Details*
1	Picture size	640 x 640
2	Epochs	100
3	Batch size	16
4	Workers	8
5	Patience	40

C. Analysis of Computational Efficiency

A YOLO-based image classification pipeline was implemented and evaluated in two phases: inference and training. During inference, a custom script processed a structured test dataset each subdirectory representing a distinct class to compute top classification predictions. The pipeline achieved an average processing time of approximately 12ms per image, with average times of 8.7ms for preprocessing and 3.3ms for inference, while postprocessing time was negligible as shown in the Table III.

TABLE III: Per-Image Computational Performance

Metric	Computational Efficiency at 640×640 Resolution			
	Preprocess (ms)	*Inference (ms)*	*Postprocess (ms)*	*Total (ms)*
Minimum	6.4	2.5	0.0	8.9
Maximum	17.5	10.4	0.1	28.0
Average	8.7	3.3	≈ 0.0	12.0

D. Evaluation Metrics

To comprehensively evaluate the effectiveness of the suggested classification model, A range of assessment criteria were employed, including Accuracy, F1-Score, Precision, Recall, and Confusion Matrix. These provide a complete performance analysis of the model in accurate classification of diverse categories. Accuracy assesses overall classification accuracy, while Precision and Recall assess accuracy of the model in classifying positive cases and avoiding false positives and false negatives classes, respectively. F1-Score computes a weighted average of Precision and Recall and provides one measure of evaluation to compare the performance of the model in imbalanced classes. The Confusion Matrix also delivers a comprehensive breakdown of true positives and negatives, false positives and negatives, also provides detailed study of the model's classification bias. The choice of these measures provides complete assessment of the model's performance in all classes and solution to class imbalance and misclassification issues. The final performance analysis was carried out on diverse datasets in order to determine the model's generality and robustness.

i. Accuracy:

$$\text{Accuracy} = \frac{\text{TP + TN}}{\text{TP + TN + FP + FN}} \qquad \text{(Eq.(1))}$$

Where:

- TP: Stands for True Positives.
- TN: Stands for True Negatives.
- FP: Stands for False Positives.
- FN: Stands for False Negatives.

ii. F1-Score:

$$F1 = 2 * \frac{\text{P} * \text{R}}{\text{P + R}} \qquad \text{(Eq.(2))}$$

Where:

- P: Stand for Precision.
- R: Stand for Recall.

iii. Precision:

$$\text{P} = \frac{\text{TP}}{\text{TP + FP}} \qquad \text{(Eq.(3))}$$

iv. Recall:

$$\text{R} = \frac{\text{TP}}{\text{TP + FP}} \qquad \text{(Eq.(4))}$$

V. Results and Discussion

Using the proposed dataset, the performance of different classification models showed drastic variations in performance, as shown in Table IV. Of all the models, YOLOv11 showed the highest accuracy of 97.5%, with precision, recall, and F1-score of 0.97, being the best model for skin disease and cancer classification. The accuracy per class was also uniformly around 99%, affirming the model's consistency across various skin conditions. The confusion matrix, shown in Figure 5, also depicts the classification performance and misclassification patterns.

On the other hand, the VGG16 and the custom CNN model did not perform well with 79.5% and 23% accuracy, respectively, showing their poor generalization on this diverse dataset. These results highlight the advantage of using more advanced architectures like YOLOv11 for real-world dermatological classification.

Training and validation loss curves, as shown in Figure 6, indicate the stability of the model during training. Figure 7 also demonstrates the test results on unseen images, indicating the efficiency of YOLOv11 in classifying the various skin conditions correctly.

TABLE IV: Test Results of the Classification Models

Model	Performance Evaluation metrics for Classification			
	Accuracy	*F1 Score*	*Precision*	*Recall*
YOLO v11	**97.5%**	**0.9750**	**0.9753**	**0.9751**
YOLO v8	94.6%	0.9455	0.9472	0.9461
DenseNet 121	81.2%	0.8114	0.8180	0.8126
MobileNet	82%	0.8208	0.8243	0.8209
ResNet 50	87.4%	0.8745	0.8774	0.8747
VGG 16	79.5%	0.7950	0.8102	0.7950
Custom CNN	23%	0.0949	0.1508	0.2308

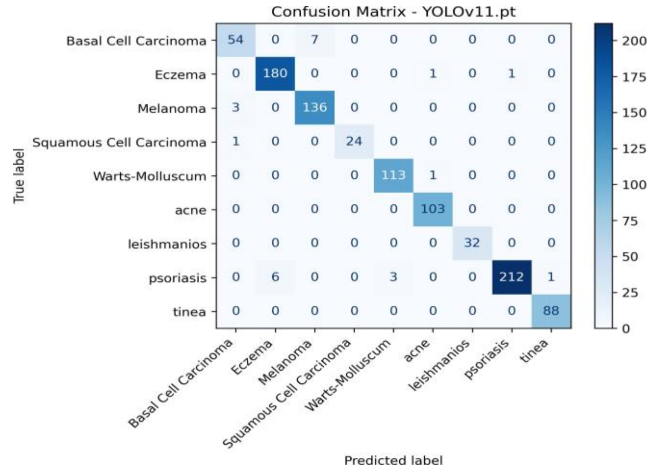


Figure 5: Confusion Matrix

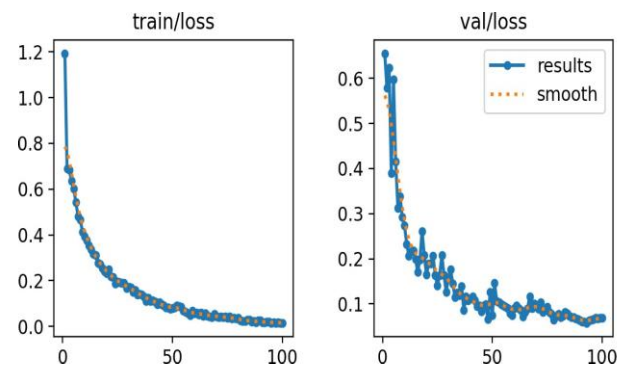


Figure 6: Training and Validation Loss Curve

Figure 7: Performance of the Proposed system on Unseen Images

VI. CONCLUSION

Our primary objective in this research was, to solve problems of multi-class classification of skin cancer and skin diseases using deep learning models. By collecting a diverse data set of 6820 images that include images from mobile cameras and dermoscopic devices, we were able to bridge the space between research data sets and real-world medical diagnostics. After testing a number of state-of-the-art classification models, YOLOv11 showed the best performance, with 97.5% accuracy and an F1-score of 0.9750. Our method facilitates real-world generalization of data, and thus it is an important tool to assist dermatologists and physicians. Future work can extend the dataset to include more skin disease and cancer classes to extend model generalization. We can also add cancer staging classification to provide extra diagnostic data to allow earlier detection and better treatment planning. Additional improvement can also be made by adding explainable AI methods, for instance, Grad-CAM or LIME, to allow better model prediction in clinical environments.

REFERENCES

[1] Islami, Farhad, et al. 2021. ”Annual Report to the Nation on the Status of Cancer, Part 1: National Cancer Statistics”. Journal of the National Cancer Institute. https://doi.org/10.1093/jnci/djab131

[2] Breen, I. D., Richmond, J. M. (2022). Why is skin cancer risk elevated at higher altitudes? Frontiers for Young Minds, 10. https://doi.org/10.3389/frym.2022.615634

[3] Bray, Freddie, Laversanne, Mathieu, Sung, Hyuna, Ferlay, Jacques, Siegel, Rebecca L., Soerjomataram, Isabelle, and Jemal, Ahmedin. 2024. ”Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries”. Wiley. https://doi.org/10.3322/caac.21834

[4] Global Cancer Observatory: Cancer Today. International Agency for Research on Cancer, 2024. https://gco.iarc.who.int/media/globocan/factsheets/cancers/16-melanoma-of-skin-fact-sheet.pdf.

[5] Donelli, Carolina, Suppa, Mariano, Tognetti, Linda, Perrot, Jean, Calabrese, Laura, Preznker, Javiera, Malvehy, Josep, Rubegni, Pietro, and Cinotti, Lisa. 2023. ”Line-Field Confocal Optical Coherence Tomography for the Diagnosis of Skin Carcinomas: Real-Life Data over Three Years”. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/curroncol30100639

[6] “ISIC Challenge,” n.d. https://challenge.isic-archive.com/data/2020.

[7] Tschandl, Philipp. “The HAM10000 Dataset, a Large Collection of Multi-source Dermatoscopic Images of Common Pigmented Skin Lesions.” Data set. Harvard Dataverse, January 1, 2018. https://doi.org/10.7910/dvn/dbw86t.

[8] Ultralytics. “GitHub - Ultralytics/Ultralytics: Ultralytics YOLO11.” GitHub, n.d. https://github.com/ultralytics/ultralytics.

[9] Silva, Brunna Carolinne Rocha, Oliveira, Bruno De Azevedo, Pra, Renata, Paiva, Joselisa, Loureiro, Rafael Maffei, Calixto, Wesley P., Reis, Mrcio R. C., and GiavinaBianchi, Mara. 2024. ”Artificial intelligence for skin cancer detection and classification for clinical environment: a systematic review”. Frontiers Media. https://doi.org/10.3389/fmed.2023.1305954

[10] Naeem, Ahmad, Anees, Tayyaba, Khalil, Mudassir, Zahra, Kiran, Naqvi, Rizwan Ali, and Lee, Seung Won. 2024. ”SNC-Net: Skin Cancer Detection by Integrating Handcrafted and Deep Learning-Based Features Using Dermoscopy Images”. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/math12071030

[11] Ali, Rizwan, Manikandan, A., Lei, Rui, and Xu, Jinghong. 2024. ”A novel SpaSA based hyper-parameter optimized FCEDN with adaptive CNN classification for skin cancer detection”. Nature Portfolio. https://doi.org/10.1038/s41598-024-57393-4

[12] Tahir, Maryam, Naeem, Ahmad, Malik, Hassaan, Tanveer, Jawad, Naqvi, Rizwan Ali, and Lee, Seung Won. 2023. ”DSCC-Net: Multi-Classification Deep Learning Models for Diagnosing of Skin Cancer Using Dermoscopic Images”. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/cancers15072179

[13] Bibi, Sobia, Khan, Muhammad Attique, Shah, Jamal Hussain, Damaevičius, Robertas, Alasiry, Areej, Marzougui, Mehrez, Alhaisoni, Majed, and Masood, Anum. 2023. ”MSRNet: Multiclass Skin Lesion Recognition Using Additional Residual Block Based Fine-Tuned Deep Models Information Fusion and Best Feature Selection”. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/diagnostics13193063

[14] Naeem, Ahmad, Anees, Tayyaba, Fiza, Makhmoor, Naqvi, Rizwan Ali, and Lee, Seung Won. 2022. ”SCDNet: A Deep Learning-Based Framework for the Multiclassification of Skin Cancer Using Dermoscopy Images”. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/s22155652

[15] Gururaj, H L, Manju, N., Nagarjun, A. Venkata, Aradhya, V. N. Manjunath, and Flammini, Francesco. 2023. ”DeepSkin: A Deep Learning Approach for Skin Cancer Classification”. Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/access.2023.3274848

[16] Ehsan Bazgir, Ehteshamul Haque, Md. Maniruzzaman, and Rahmanul Hoque. 2024. “Skin Cancer Classification Using Inception Network.” World Journal of Advanced Research and Reviews. https://doi.org/10.30574/wjarr.2024.21.2.0500

[17] Saleh, Neven, Mohammed A. Hassan, and Ahmed M. Salaheldin. ”Skin cancer classification based on an optimized convolutional neural network and multicriteria decision-making.” Scientific Reports https://doi.org/10.1038/s41598-024-67424-9

[18] Venugopal, Vipin, Raj, Navin Infant, Nath, Malaya Kumar, and Stephen, Norton. 2023. ”A deep neural network using modified EfficientNet for skin cancer detection in dermoscopic images”. Elsevier BV. https://doi.org/10.1016/j.dajour.2023.100278

[19] Rehman, Muhammad Zia Ur, Ahmed, Fawad, Alsuhibany, Suliman A., Jamal, Sajjad Shaukat, Ali, Muhammad Zulfiqar, and Ahmad, Jawad. 2022. ”Classification of Skin Cancer Lesions Using Explainable Deep Learning”. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/s22186915

[20] Ghosh, Pronab, Azam, Sami, Quadir, Ryana, Karim, Asif, Shamrat, F. M. Javed Mehedi, Bhowmik, Shohag Kumar, Jonkman, Mirjam, Hasib, Khan Md., and Ahmed, Kawsar. 2022. ”SkinNet-16: A deep learning approach to identify benign and malignant skin lesions”. Frontiers Media. https://doi.org/10.3389/fonc.2022.931141

[21] Alam, Talha Mahboob, Shaukat, Kamran, Khan, Wasim Ahmad, Hameed, Ibrahim A., Almuqren, Latifah Abd., Raza, Muhammad Ahsan, Aslam, Memoona, and Luo, Suhuai. 2022. ”An Efficient Deep Learning-Based Skin Cancer Classifier for an Imbalanced Dataset”. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/diagnostics12092115

[22] Chang, Chih-Chi, Li, Yu-Zhen, Wu, Hui-Ching, and Tseng, MingHseng. 2022. ”Melanoma Detection Using XGB Classifier Combined with Feature Extraction and K-Means SMOTE Techniques”. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/diagnostics12071747

[23] Ravi, Vinayakumar. 2022. ”Attention Cost-Sensitive Deep Learning-Based Approach for Skin Cancer Detection and Classification”. Multidisciplinary Digital Publishing Institute. https://doi.org/10.3390/cancers14235872

[24] Roboflow. “Medimages Instance Segmentation Dataset and Pre-Trained Model by Med Seg,” November 1, 2024. https://universe.roboflow.com/med-seg/medimages

[25] Roboflow. “Skin Cancer Object Detection Dataset and Pre-Trained Model by Framejframej 1,” August 4, 2023. https://universe.roboflow.com/framejframej-1-krvqf/skin-cancer-oho4x

[26] “Skin Diseases Segmentation Instance Segmentation Dataset and Pre-Trained Model by Xyz,” September 20, 2024. https://universe.roboflow.com/xyz-n2z4a/skin-diseases-segmentation

[27] Roboflow. “Skin Instance Segmentation Dataset by Suui,” May 11, 2024. https://universe.roboflow.com/suui/skin-p3cqf

[28] Roboflow. “New Mod Instance Segmentation Dataset and Pre-Trained Model by Harsh Shakya,” September 26, 2024. https://universe.roboflow.com/harsh-shakya-yytye/new-mod-nabkg

Enhanced Trapezoidal Modulation in MMC: Comparative Analysis with Traditional Modulation Methods

Saqib Hussain
Department of Electrical Engineering
Sukkur IBA University
Sukkur, Pakistan
saqibhussain.beef21@iba-suk.edu.pk

Ahtsham Ullah
Department of Electrical Engineering
Sukkur IBA University
Sukkur, Pakistan
ahtashamkhokher624@gmail.com

Jamshed Ansari
Department of Electrical Engineering
Sukkur IBA University
Sukkur, Pakistan
jamshed.ahmed@iba-suk.edu.pk

***Abstract*— HVDC transmission and renewable energy systems extensively use Modular Multilevel Converters (MMC) because they provide outstanding scalability and modular architectural features. The performance quality of MMCs depends predominantly on which modulation technique engineers implement. This work studies Nearest Level Modulation (NLM) and conventional Trapezoidal Modulation alongside an enhanced Trapezoidal Modulation method to identify the top choice for high-voltage power implementations. A goal exists to evaluate these techniques by assessing their Total Harmonic Distortion (THD) performance as well as their switching losses and their implementation complexity metrics. Each modulation strategy is simulated through MATLAB/Simulink-based testing under identical operating situations. Product testing indicates NLM shows lower switching losses as well as superior power distribution efficiency but the updated Trapezoidal Modulation design combines reduced THD performance with simple implementation methods. The research delivers critical knowledge about MMC modulation selection which systems designers and manufacturers can use to optimize converter operation based on specific application specifications.**

Keywords* —Modular multilevel converter (MMC), nearest level modulation (NLM), trapezoidal modulation (TRP), total harmonic distortion (THD), switching losses

I. Introduction

High-voltage power systems depend on Modular Multilevel Converters (MMCs) because these converters bring scalability together with fault-tolerant capabilities and modular constructions [1]. A modular multilevel converter contains individual submodules which assemble into phase legs and can exist in three different evaluation arrangements [2]. MMCs operate through their unique structure to support high voltage operations without causing significant stress on components thus enabling their use in high-voltage direct current (HVDC) transmissions and motors and renewable energy integration [3]. High power quality along with conversion efficiency stands as a major challenge for Modular Multilevel Converters despite their existing advantages because modulation techniques establish the overall MMC performance profile [4]. The research aims to develop improved modulation techniques which enhance power quality together with loss reduction and simplify practical high-voltage application implementation [5].

Research into MMC modulation techniques has focused heavily in recent years and NLM alongside TRP represent the leading choices [6]. Using NLM yields enhanced switching capabilities and improved efficiency yet this modulation strategy entails complex operations alongside difficulties in maintaining submodule voltage equilibrium [7]. While Trapezoidal Modulation presents advantages of straightforward implementation it leads to elevated levels of total harmonic distortion (THD) and switching losses as compared to Nearest Level Modulation (NLM) [8]. Multiple studies suggest modifications of these techniques aimed at solving their existing shortcomings [9]. Present approaches fall short because they maintain limited performance-complexity equilibrium or provide modularized solutions for particular use cases [10].The presented study enhances Trapezoidal Modulation by resolving traditional TRP limitations to provide a workable alternative for practical Multi-Master-control applications [11].

Because traditional modulation methods face limitations from efficiency in TRP while also requiring complexity for NLM implementations there is a need for balanced modulation solutions [12]. An improved Trapezoidal Modulation technique has been developed to deliver better power quality together and straightforward implementation capabilities [13]. The proposed approach works to enhance MMC performance at high voltages by implementing solutions to deal with voltage unbalance and harmonic distortion problems [14]. The study compares improved Trapezoidal Modulation against traditional NLM to offer researchers both practical strategies for MMC system modulation selection [15].

II. Methodology

A. MMC Overview

The Modular Multilevel Converter (MMC) is a class of power converter appreciated for its high voltage scaling, modularity and efficiency [1]-[2]. The MMC is built with several submodules (SMs), and they often take half-bridge or full-bridge configurations [3]. Every SM consists of capacitors for energy storage with switch power devices [4]. Since the converter has three phase legs, each leg is further divided into two parts, an upper arm and a lower arm.

These arms function simultaneously generating staircase output voltage similar to the sinusoidal waveform voltage. This modular approach allows MMCs to deal with very high voltage since the voltage load is spread out with sub modules hence minimizing the stress that the several components undergo [5]-[6].

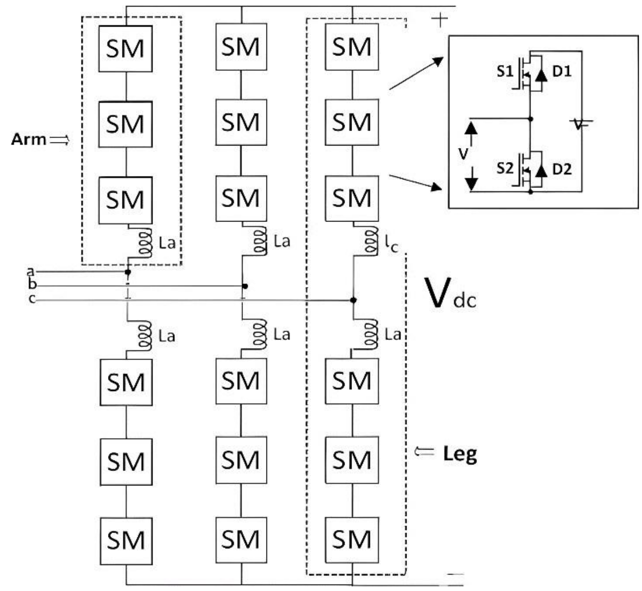


Figure 1: Three Phase MMC Architecture [19]

B. NLM Principle

NLM functions by quantizing the reference waveform into respective voltage levels corresponding to the number of submodules in the MMC [23]. Each of the submodules provides a voltage increment and the modulation can enable or disable to approximate to the reference waveform [24].

This stepwise output further approximates the sine wave as number of levels rises and hence results in better waveform and lower percentage of harmonic distortion [25].

$$V_{ref} = Vm\sin(wt)$$
$$V_i = i * \frac{V_{dc}}{n-1}$$
$$V_{nearest}(t) = argmin|V_{ref} - V_i|$$
$$V_{out} = V_{nearest}(t)$$

Equation 1: Equation for Generating NLM signal of 100 Hz for Gate signal [29]-[30]

In NLM, capacitor voltages in the submodules have to be continuously monitored in real-time fashion to address the issue of voltage balance. The system performs a selection of which submodules must be active based on a fixed search table or an algorithm [27].

While the principal of NLM already implies presence of certain harmonic content, the distortion is significant and decreases as the levels grow, which makes its use for the management of MMC reasonable [23].

C. Trapezoidal Modulation Principle

Commonly another technique used in MMCs is Trapezoidal Modulation because of its simple structure and foreseeable operation result [15]. The main characteristics of this method are based on the use of a reference signal with sinewave envelope and a straight trapezoidal segments. To find precise times when the submodule needs to switch, the MMC's output voltage levels are compared to the reference signal [16].

$$Vref(t) = A \cdot Trap(t)$$

$$S(t) = \begin{cases} 1, & Vref(t) \geq Vtri(t) \\ 0, & Vref(t) < Vtri(t) \end{cases}$$

Equation 2: Equation for generating switching signal using trapezoidal modulation

Further regulating the submodules in such a way, the Trapezoidal Modulation provides a staircase-like voltage waveform, which partially or fully matches the trapezoidal reference signal [17]. Compared to the square-wave modulation, the smoother transition in the trapezoidal waveform enhances harmonic performance as well as keeping the system simple [16].

Trapezoidal Modulation though occupies lesser bandwidth than bank modulation may have more harmonic distortion compared to techniques such as Nearest Level Modulation (NLM) [15]. High frequency switching is usually necessary for good quality of power and this results into high switching losses.

Nevertheless, Trapezoidal Modulation is still advantageous for applications of MMCs when low cost, simple control scheme, and fairly acceptable power quality are necessary. These include cases within medium voltage drive and industrial application where low THD and fine waveform regulation is not important.

D. Proposed Trapezoidal Principle

The sophisticated trapezoidal modulation technique operates at the base frequency rather than employing high-frequency triangular signals used in conventional approaches because this increases device stress levels and switching losses. Power loss decreases along with thermal stress while the system becomes more efficient and reliable through reduced switching operations. The method enables smooth Modular Multilevel Converter (MMC) operation with a shorter transition duration between states through its well-planned switching pattern which reduces unwanted harmonic content to extend system lifespan and improve performance.

$$Trapezoid(t) = \begin{cases} 1, & if \quad A \cdot Triangle(t) > 1 \\ -1, & if \quad A \cdot Triangle(t) < -1 \\ & A \cdot Triangle(t) \end{cases}$$

Equation 3: Proposed Technique to Generate the Gate Signal

The refined trapezoidal modulation procedure works at the base frequency instead of standard high-frequency triangular signals to yield lower power losses and decreased electrical stresses on devices. Reduction of switching frequency through this technique leads to lower power loss and reduced thermal stress which improves the system's reliability and efficiency. The properly designed switching sequence enables efficient operation of Modular Multilevel Converters (MMC) by decreasing time needed for state transitions. Through its operation the technique lowers both unwanted harmonic production and component strain which results in extended system lifespan and superior performance results in practical power conversion applications.

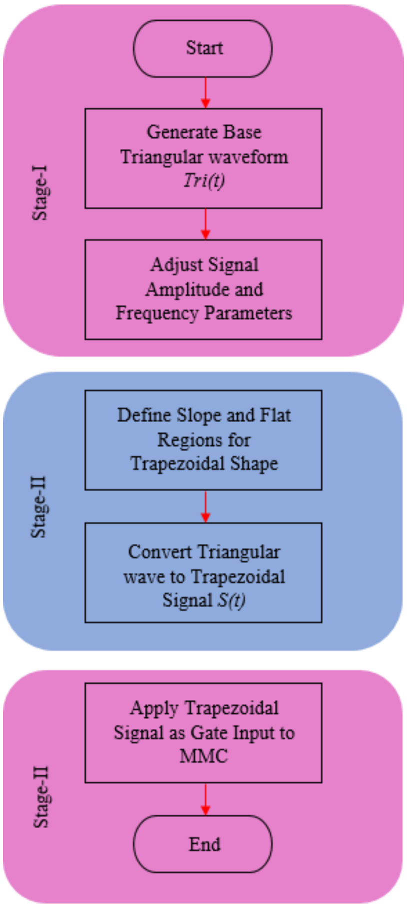


Figure 2: Proposed Trapezoidal Modulation Flowchart

III. Results and Discussion

A. Harmonic Distortion

THD comparisons of Nearest Level Modulation (NLM) and Trapezoidal Modulation for full scale MMC confirms that the MMC is more efficient at higher level of NLM with Trapezoidal Modulation.

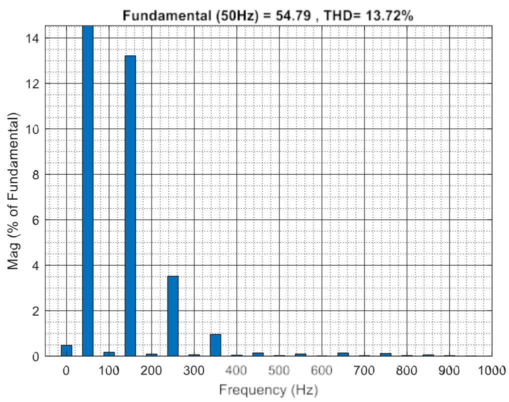
The proposed trapezoidal modulation technique provides a THD value of 9.11% during 21-level operation thus delivering better results than Trapezoidal Modulation's 10.91% and NLM's 13.72% because of its lower THD value. The number of levels used in the system directly affects the level of waveform distortion in the process. The THD measurement shows NLM produces 13.28% with 31 levels before Trapezoidal Modulation reduces it to 10.51% and the Proposed Trapezoidal Modulation makes it reach 5.76%. The THD value for NLM at 41 levels reaches 12.93% whereas Trapezoidal Modulation controls it to 8.82% and the Proposed Trapezoidal Modulation reaches a minimum THD of 5.53%.

This indicates that THD is on an average reduced more by the proposed Trapezoidal Modulation than by NLM and previously used TRP in each of the tested levels as the following result table affirms. This can be attributed to its provision of crisper waveforms in the least harmonic components possible. However, as using discrete voltage levels, the examined loop currents of NLM contain higher harmonics, especially at lower modulation levels.

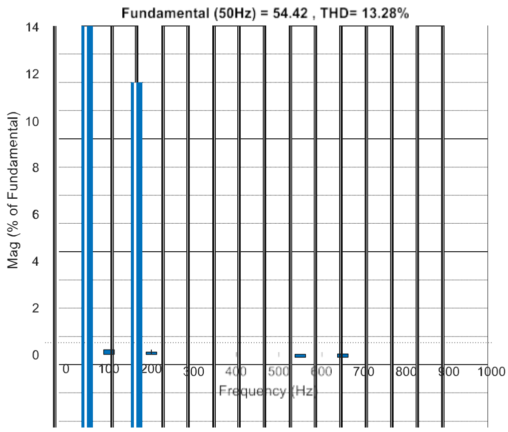
Trapezoidal Modulation has lower THD as compared with the sinusoidal modulation, thereby guaranteeing better power quality and a lesser load put on the equipment. Although NLM has lower computational complexity than the SRF-P&O method, it has a higher THD, and therefore cannot be used to meet strict requirements for power quality. Consequently, the findings points to the applicability of Trapezoidal Modulation in enhancing the predicament of harmonic performance of MMC systems.

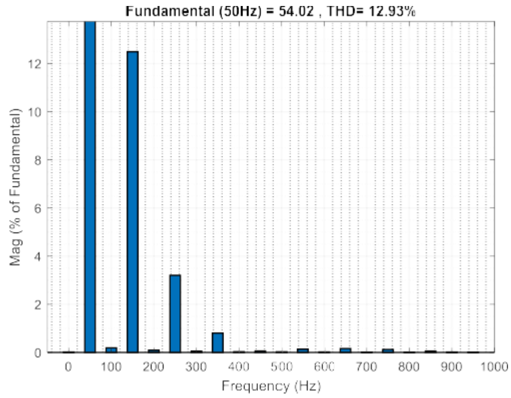
THD of Applied Modulation Techniques at different Submodule Levels			
Levels	***NLM***	***Trapezoidal Modulation***	***Proposed Trapezoidal Modulation***
21	(a) 13.72%	(d) 10.91%	(g) 9.11%
31	(b) 13.28%	(e) 10.51%	(h) 5.76%
41	(c) 12.93%	(f) 8.82%	(i) 5.53%

Table 1: THD of NLM and Trapezoidal Modulations at different SM levels

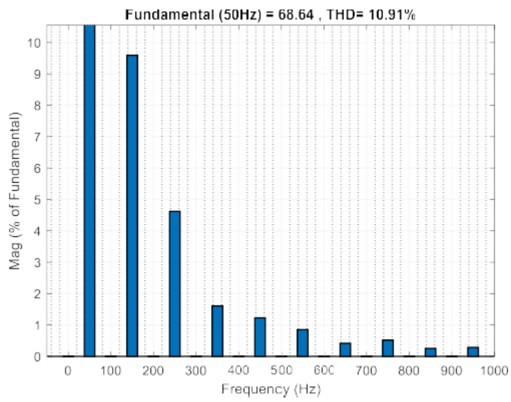


(a)

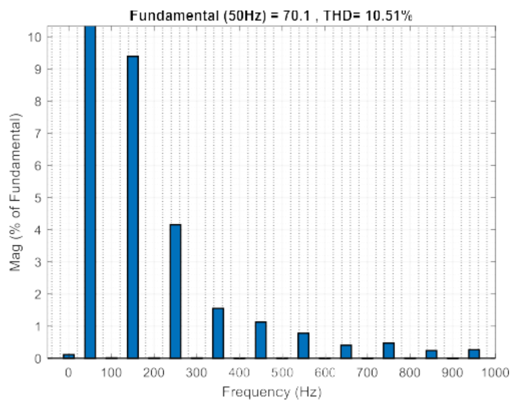




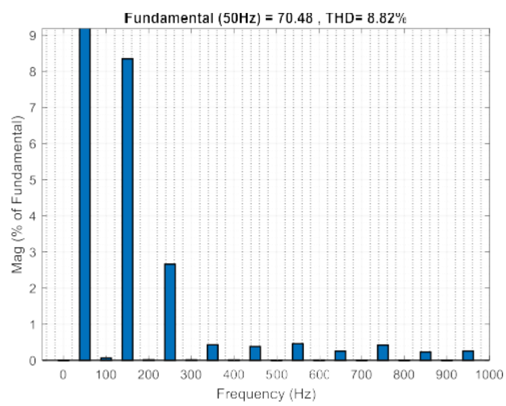
(c)



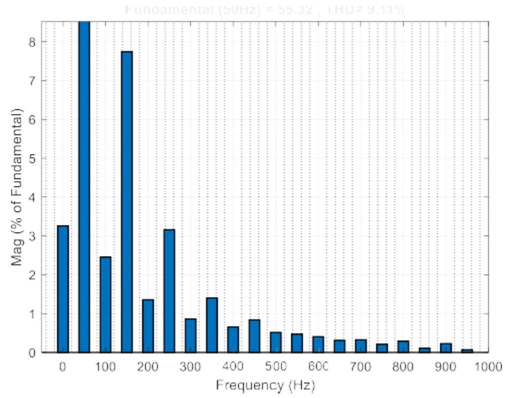
(d)



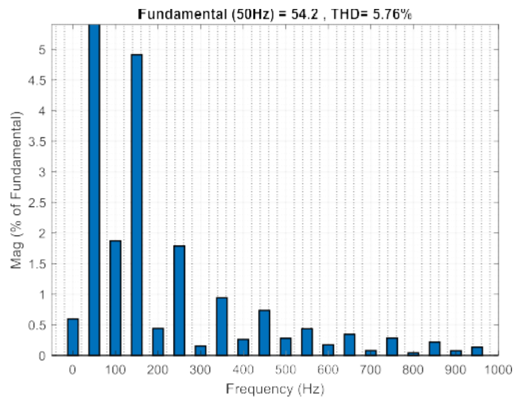
(e)



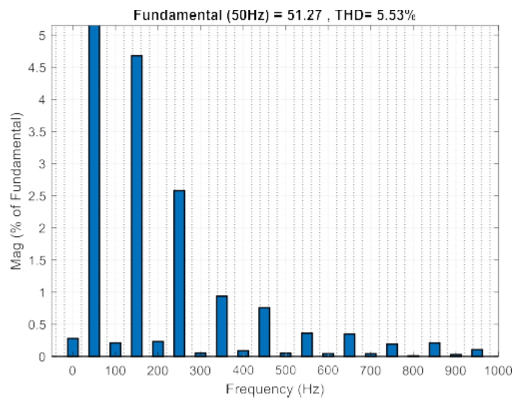
(f)



(g)



(h)



(i)

Figure 3: THD Results of Applied Modulation Schemes at Various levels

B. *Switching Losses and Overall Efficiency*

The 50 Hz MMC inverter received 100V input while testing NLM and TRP Modulation throughout different channel levels from 21 to 31 and finally 41 for both upper and lower arms. The switching losses within NLM become substantial because the output voltage steps are distinct while operating at lower levels. The system efficiency decreased to 50% at 21 levels because of the excessive number of frequent switches. The efficiency of NLM modulation decreased even with higher level steps (31–41) because its discrete output pattern decreased the overall operational efficiency versus other modulation techniques for demanding applications.

TRP modulation creates smooth voltage transitions that reduces the number of power switching events and lowers energy waste thus decreasing losses. The efficiency scores of TRP equal or surpass NLM while reaching the minimum

THD requirement at 21 modulation levels. The efficiency of TRP improved continuously as the number of control levels increased. The reduced THD performance of TRP methods decreases stress placed on both filtering systems along with passive components.

TRP shows better performance than NLM throughout all input levels even though both methods encounter operational difficulties when used at low input signals. TRP modulation proves suitable for MMC operations at 50 Hz frequencies because it creates a strong combination of quality-shaped waveforms alongside enhanced switching performance.

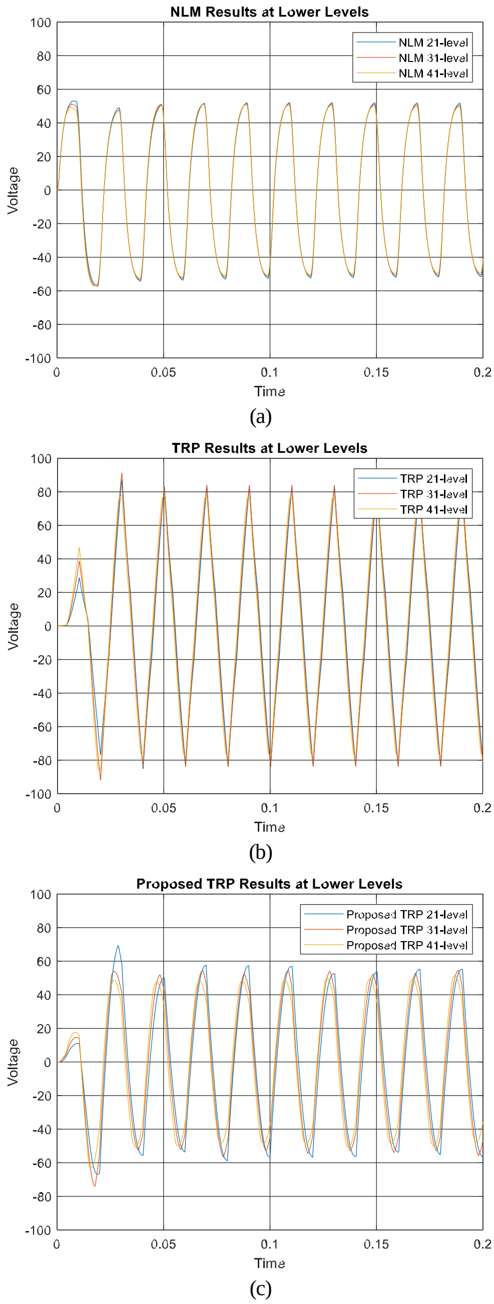


Figure 4. Voltage Measurements at the Load Side: (a) Voltage Results for Nearest Level Modulation at 21, 31 and 41 levels (b) Voltage Results for Trapezoidal Modulation at 21, 31 and 41 levels (c) Voltage Results for Proposed Trapezoidal Modulation at 21, 31 and 41 levels

C. Modulation Complexity

NLM requires higher computational power compared to the TRP method. NLM demands continuous calculations for finding the nearest voltage level match to reference waveforms through logical assessment of data. An increase in MMC levels corresponds to increased computational complexity of the NLM system. NLM ensures superior voltage control and accuracy because of its complexity therefore making this method appropriate for applications that require critical waveform tracking and accurate measurements.

The switching instances of MMC submodules are determined by comparing trapezoidal reference waveforms to carrier signals in Trapezoidal Modulation (TRP). The TRP implementation method features basic processing requirements and low computational load which makes it simpler to operate in real-time operational environments. The original TRP approach generates higher Total Harmonic Distortion (THD) than NLM but an optimized TRP design has solved this limitation by creating minimal THD while maintaining efficiency levels. Through its operation TRP produces fewer switching events which decreases both the amount of switching losses and the thermal strain on the system.

The simple design structure together with operational efficiency of TRP makes it suitable for THD reduction requirements that prioritize performance over voltage control accuracy. The selection of NLM should be made when accuracy in voltage measurement is paramount even though this technique comes with advanced computation requirements. The selection between NLM and TRP depends on particular application needs that involve trade-offs between computational resources and harmonic performance and required waveform precision.

IV. Conclusion

Trapezoidal Modulation (TRP) and Nearest Level Modulation (NLM) are being popularly employed in Modular Multilevel Converters (MMC) for increased efficiency and lower switching losses. Conventional TRP is economical but produces large harmonic distortion with the need for extra filtering. NLM has low switching losses but calls for sophisticated voltage balancing techniques for firm operation.

The suggested Trapezoidal Modulation addresses MMC performance by supporting operation at the fundamental frequency range compared to conventional TRP operating at higher frequencies. This method minimizes harmonic distortion with lower switching losses in comparison to NLM, and it is thus an effective solution in terms of efficiency for high-power applications. By analyzing the equation of each modulation technique we can mathematically understand the complexity of them.

V. Limitations and Recommendations

Modulation techniques for Modular Multilevel Converters (MMC) must find an equilibrium between harmonic performance and switching loss levels. The switching losses of Nearest Level Modulation (NLM) decrease but the poor waveform resolution causes increased harmonic distortion. The Traditional Trapezoidal Modulation produces better harmonic performance than NLM although it requires elevated switching frequency that produces wider losses. The Proposed Improved Trapezoidal Modulation stands as the simplest among three modulation methods while it conducts operations at fundamental frequency to achieve superb THD performance. The improved method produces waveforms with enhanced quality through diminished harmonic content. This improved method introduces additional switching transitions but manages to sustain losses which decrease better than NLM. This method creates the most advantageous combination between processing simplicity and successful harmonic suppression. This proposed modulation approach works best among other alternatives for power electronic systems that need maximum efficiency with minimal harmonic distortion.

The study needs to research how mixing together elements from NLM and TRP can improve system functionality. The HIL testing approach serves as an essential requirement to validate practical operational effectiveness of the proposed approach. With HIL engineers can conduct real-time examinations to measure control delays along with system efficiency which provides vital information for adapting the implementation. The Proposed Improved Trapezoidal Modulation presents solid prospects for industrial applications because it delivers efficient performance along with low-THD and convenient implementation methods.

REFERENCES

[1] Lesnicar, A. and Marquardt, R., 2003, June. An innovative modular multilevel converter topology suitable for a wide power range. In *2003 IEEE Bologna Power Tech Conference Proceedings,* (Vol. 3, pp. 6-pp). IEEE.

[2] Guan, M. and Xu, Z., 2012. Modeling and control of a modular multilevel converter-based HVDC system under unbalanced grid conditions. *IEEE transactions on power electronics*, *27*(12), pp.4858-4867.

[3] Rohner, S., Bernet, S., Hiller, M. and Sommer, R., 2009. Modulation, losses, and semiconductor requirements of modular multilevel converters. *IEEE transactions on Industrial Electronics*, *57*(8), pp.2633-2642.

[4] Du, S., Dekka, A., Wu, B. and Zargari, N., 2017. *Modular multilevel converters: analysis, control, and applications*. John Wiley & Sons.

[5] Adam, G.P., Anaya-Lara, O., Burt, G.M., Telford, D., Williams, B.W. and McDonald, J.R., 2010. Modular multilevel inverter: Pulse width modulation and capacitor balancing technique. *IET power electronics*, *3*(5), pp.702-715.

[6] Antonopoulos, A., Angquist, L. and Nee, H.P., 2009, September. On dynamics and voltage control of the modular multilevel converter. In *2009 13th European Conference on Power Electronics and Applications* (pp. 1-10). IEEE.

[7] Hagiwara, M. and Akagi, H., 2009. Control and experiment of pulsewidth-modulated modular multilevel converters. *IEEE transactions on power electronics*, *24*(7), pp.1737-1746.

[8] Tu, Q., Xu, Z. and Xu, L., 2011. Reduced switching-frequency modulation and circulating current suppression for modular multilevel converters. *IEEE transactions on power delivery*, *26*(3), pp.2009-2017.

[9] Ilves, K., Antonopoulos, A., Norrga, S. and Nee, H.P., 2011. Steady-state analysis of interaction between harmonic components of arm and line quantities of modular multilevel converters. *IEEE transactions on power electronics*, *27*(1), pp.57-68.

[10] Glinka, M. and Marquardt, R., 2005. A new AC/AC multilevel converter family. *IEEE Transactions on Industrial Electronics*, *52*(3), pp.662-669.

[11] Barros, L.A., Martins, A.P. and Pinto, J.G., 2022. A comprehensive review on modular multilevel converters, submodule topologies, and modulation techniques. *Energies*, *15*(3), p.1078.

[12] Wang, K., Li, Y., Zheng, Z. and Xu, L., 2012. Voltage balancing and fluctuation-suppression methods of floating capacitors in a new modular multilevel converter. *IEEE Transactions on Industrial Electronics*, *60*(5), pp.1943-1954.

[13] Saeedifard, M. and Iravani, R., 2010. Dynamic performance of a modular multilevel back-to-back HVDC system. *IEEE Transactions on power delivery*, *25*(4), pp.2903-2912.

[14] Perez, M.A., Bernet, S., Rodriguez, J., Kouro, S. and Lizana, R., 2014. Circuit topologies, modeling, control schemes, and applications of modular multilevel converters. *IEEE transactions on power electronics*, *30*(1), pp.4-17.

[15] Ansari, M.S., Shukla, A. and Bahirat, H.J., 2020, October. Modeling of MMC based high power DC-DC converter controlled using trapezoidal modulation. In *2020 IEEE Energy Conversion Congress and Exposition (ECCE)* (pp. 5716-5722). IEEE.

[16] Gowaid, I.A., Adam, G.P., Ahmed, S., Holliday, D. and Williams, B.W., 2014. Analysis and design of a modular multilevel converter with trapezoidal modulation for medium and high voltage DC-DC transformers. IEEE Transactions on Power Electronics, 30(10), pp.5439-5457.

[17] Zheng, G., Chen, Y. and Kang, Y., 2021. Trapezoidal current modulation for a compact DC modular multilevel converter with ZVS of submodules and ZCS of voltage-balancing circuits. IEEE Transactions on Power Electronics, 36(10), pp.10986-10992.

[18] Moranchel, M., Huerta, F., Sanz, I., Bueno, E. and Rodríguez, F.J., 2016. A comparison of modulation techniques for modular multilevel converters. Energies, 9(12), p.1091.

[19] Nami, A., Liang, J., Dijkhuizen, F. and Demetriades, G.D., 2014. Modular multilevel converters for HVDC applications: Review on converter cells and functionalities. IEEE Transactions on Power Electronics, 30(1), pp.18-36.

[20] Jovcic, D., 2019. High voltage direct current transmission: converters, systems and DC grids. John Wiley & Sons.

[21] Zhang, L., Zou, Y., Yu, J., Qin, J., Vittal, V., Karady, G.G., Shi, D. and Wang, Z., 2017. Modeling, control, and protection of modular multilevel converter-based multi-terminal HVDC systems: A review. CSEE Journal of Power and Energy Systems, 3(4), pp.340-352.

[22] Adam, G.P., Ahmed, K.H., Finney, S.J. and Williams, B.W., 2010, September. AC fault ride-through capability of a VSC-HVDC transmission systems. In 2010 IEEE Energy Conversion Congress and Exposition (pp. 3739-3745). IEEE.

[23] Badar, J., Akhtar, F., Kumar, D., Munir, H.M., Ali, K.H., Alsaif, F. and Alsulamy, S., 2023. An MMC based HVDC system with optimized AC fault ride-through capability and enhanced circulating current suppression control. Frontiers in Energy Research, 11, p.1190975.

[24] Soomro, J.B., Akhter, F., Ali, S., Bukhari, S.S.H., Sami, I. and Ro, J., 2021. Modified nearest level modulation for full-bridge based HVDC MMC in real-time hardware-in-loop setup. IEEE Access, 9, pp.114998-115005.

[25] Lin, L., Lin, Y., He, Z., Chen, Y., Hu, J. and Li, W., 2016. Improved nearest-level modulation for a modular multilevel converter with a lower submodule number. *IEEE Transactions on Power Electronics*, *31*(8), pp.5369-5377.

[26] Hu, P. and Jiang, D., 2014. A level-increased nearest level modulation method for modular multilevel converters. *IEEE transactions on Power Electronics*, *30*(4), pp.1836-1842.

[27] Flourentzou, N., Agelidis, V.G. and Demetriades, G.D., 2009. VSC-based HVDC power transmission systems: An overview. IEEE Transactions on power electronics, 24(3), pp.592-602.

[28] Kouro, S., Malinowski, M., Gopakumar, K., Pou, J., Franquelo, L.G., Wu, B., Rodriguez, J., Pérez, M.A. and Leon, J.I., 2010. Recent advances and industrial applications of multilevel converters. IEEE Transactions on industrial electronics, 57(8), pp.2553-2580.

[29] Jeong, M., Fuchs, S. and Biela, J., 2020, September. High performance LQR control of modular multilevel converters with simple control structure and implementation. In 2020 22nd European Conference on Power Electronics and Applications (EPE'20 ECCE Europe) (pp. P-1). IEEE.

[30] Lin, W., Jovcic, D., Nguefeu, S. and Saad, H., 2015. Full-bridge MMC converter optimal design to HVDC operational requirements. IEEE Transactions on Power Delivery, 31(3), pp.1342-1350.

Optimal Coordination of Directional Overcurrent Relays in Interconnected Networks Using an Improved Multi-Strategy Coati Optimization Algorithm

Syed Muhammad Abdullah Hashmi
Department of Electrical Engineering
University of Engineering and Technology Taxila, Pakistan
17.ee.117uet@gmail.com

Tahir Mahmood
Department of Electrical Engineering
University of Engineering and Technology Taxila, Pakistan
tahir.mahmood@uettaxila.edu.pk

Muhammad Mansoor Ashraf
Department of Electrical Engineering
University of Engineering and Technology Taxila, Pakistan
mansoor.ashraf@uettaxila.edu.pk

***Abstract*— The optimal coordination of directional overcurrent relays (DOCRs) in interconnected power systems is crucial for ensuring selective and time-efficient protection. This study presents a Tuned Non-inertial T-distribution based Weighted Coati Optimization Algorithm (TNTWCOA) to solve the nonlinear, highly constrained DOCR coordination problem. The proposed approach optimizes time dial setting (TDS) and plug setting (PS) to minimize relay operating times while maintaining selectivity Coordination is important for main/primary and remote/backup protection devices in order to stop fault currents from swiftly rising to dangerously high levels. TNTWCOA incorporates a chaotic sequence mechanism for enhanced population initialization, a nonlinear inertia weight to balance exploration and exploitation, an adaptive T-distribution variation strategy to improve diversity and escape local optima, and an alert update mechanism to refine search adaptability. The algorithm is validated on IEEE 3-bus, and 15-bus test networks, considering mid and near-end faults and utilizing a normal inverse relay characteristic. Comparative analysis with state-of-the-art metaheuristics demonstrates TNTWCOA's superior performance in reducing total relay operation time while outperforming classical and recent optimization techniques. The results confirm TNTWCOA's ability to mitigate premature convergence and enhance search efficiency, making it a highly effective solution for DOCR coordination in modern power system.**

Keywords—Optimal coordination, Directional overcurrent relays, Power System Protection, Relay Settings, Coati Optimization Algorithm

I. INTRODUCTION

The synthesis of overcurrent relays (OCR) with directional units gives rise to directional overcurrent relays (DOCRs).The coordination of relays is a highly non-linear and challenging problem in complex networks, and it becomes increasingly difficult to solve using analytical techniques as the number of relays and injection of external grids / distributed generators (DG) increase [1], [12].

The essence of DOCRs coordination involves isolating faults by targeting the smallest possible network zone to achieve optimal overcurrent coordination, primarily by minimizing the total operating time of the main relays. This process entails finding appropriate Time Dial Settings (TDS) and Plug Settings (PS) for relays to ensure that any fault is quickly isolated by the corresponding main relay. These settings must also be coordinated with adjacent zone relays, adding complexity to the process [2].

In the late eighties, traditional approaches like curve fitting [3] and analytical methods have been explored, along with graph theory-based methods. These were tactical mathematical techniques used for achieving the optimum results, also linear programming (LP) technique was introduced to address the coordination relay problem in a linear format. [4]. Linear Programming treats TDS as a decision variable while keeping Plug setting of all relays fixed. As a result it sets the TDS as a goal of optimization while keeping the PS fixed. Linear programming (LP) is quick but limited in reaching global optima. Nonlinear Programming (NLP) considers both TDS and PS as decision variables. In [5], an interior point method to optimize TDS and PS is used, aiming to minimize operating time and ensure relay coordination. The optimal coordination problem of DOCRs can be formulated as a Mixed-Integer Nonlinear Programming (MINLP) problem, where TDS is continuous and PS is discrete. In [6], a MILP formulation using discrete PS values by converting bilinear variables into linear inequalities was proposed

Metaheuristic methods such as the teaching learning based optimization algorithm [7], modified deferential evolution (MDE) [8], adaptive fuzzy directional bat algorithm [9] and swarm-based techniques like modified particle swarm optimization (PSO) [10] also provide effective solutions. gorilla troops optimizer (GTO) [11] have been employed as well. An elite marine predator algorithm was recently proposed for the said coordination study where elite vector enhanced global exploitation of original marine predators algorithm [1] are some other state of the art techniques that contributes to solving the coordination problem.

In this article, the problem of DOCRs coordination in complex networks is tackled using the an improved coati optimization algorithm shorted as TNTWCOA, tested on the IEEE 3- and 15-bus benchmark systems. The decision variables PS, and TDS are optimized to minimize the total operating time of primary relays, ensuring effective isolation of faulty lines while meeting the necessary constraints. The rest of the article is sectioned as follows: Section II details the mathematical formulation of the problem, Section III explains the improved coati optimization algorithm, Section IV presents the statistical results and graphs from the simulations, and the study is concluded in Section V. Either the near-end or mid-point faults on various fault locations have been considered for each test systems. Both decision variables that are TDS and PS are considered in continuous action mode for NLP-type problem configuration.

II. Problem Formulation

A. Objective Function (OF)

The coordination issue of DOCRs is formulated to be an optimization problem that tends to minimize the overall time of operation of all the primary relays in a complex system for various points of failure. This is done by determining the settings (TDS, PS) to stop fault currents from swiftly increasing to potentially hazardous levels [1]. Mathematically objective function (OF) can be expressed as described by Eq. (1), where operating times are minimized as a net sum of main relays embedded throughout the system in which coordination is aimed:

$$OF = \min T_{opr} = \sum_{\varphi=1}^{\varepsilon} \omega_{\varphi} T_{\varphi,pri} \quad (1)$$

Whereas, T_{opr} represents the total operating time, ε indicates the total number of main relays, and $T_{\varphi,pri}$ signifies the trip at fault time of the R_{φ}, the φ^{th} primary relay. The ω is taken as '1'. The IEC and ANSI/IEEE characteristic of standard inverse time to formulate the operating time of relays is:

$$T_{\varphi,a} = \frac{TDS_{\varphi} \times \theta_{\varphi}}{\left(\frac{I_{sc,t_a}}{PS_{\varphi}}\right)^{\gamma} - \mu} + \Lambda, \forall \varphi, \partial \in N; I_{sc,t_a} \in \mathbb{R} \quad (2)$$

Whereas, $\{\forall\, TDS_{\varphi}, PS_{\varphi} \in \mathbb{Q}^{+}\}$, they represent TDS and PS of φ^{th} primary relay. I_{sc,t_a} is the magnitude of short-circuit current as seen by φ^{th} relay subject to fault (f) occurring at ∂ location. Also, γ, μ, Λ and θ are constants. For normally inverse curve, γ, μ, Λ, and θ are 0.02, 1, 0, and 0.14 respectively. In addition to these, CTR_{φ} is provided as current transformer ratio of φ^{th} relay and it is related to Eq. (2) by the following expressions:

$$\frac{I^{P}_{\varphi}}{CTR_{\varphi}} = PS_{\varphi}, \forall \varphi \in N; CTR \in \mathbb{Q}^{+} \quad (3)$$

$$\{I^{P}_{\varphi} \mid I^{P^{min}} \leq I^{P}_{\varphi} \leq 66.67\% \times I_{sc,t_a}\}, \forall \varphi, \partial \in N \quad (4)$$

Whereas, I^{P}_{φ} is the pickup current value for the φ^{th} primary unit, having $I^{P^{min}}$ as the lower bound which is more than the load current. Essentially, this equation shows how the plug setting is calculated for each relay based on the current transformer ratio and the pickup current.

B. Constrainsts of Objective Function

The objective function must meet constraints based on variable settings. PS and TDS must stay within each relay's range, defined by Eq. (5) and (6). Typically, TDS ranges from 0.05 to 1.1, and PS from 0.1 to 5, as referenced in [1].

- $TDS^{min}_{\varphi} \leq TDS_{\varphi} \leq TDS^{max}_{\varphi}, \forall \varphi \in N$ (5)
- $PS^{min}_{\varphi} \leq PS_{\varphi} \leq PS^{max}_{\varphi}, \forall \varphi \in N$ (6)
- $T^{min}_{\varphi} \leq T_{\varphi} \leq T^{max}_{\varphi}, \forall \varphi \in N$ (7)
- $T^{bu}_{\varphi} - T^{pri}_{\varphi} \geq CTI, \forall \varphi \in N; CTI \in \mathbb{Q}$ (8)

Where T^{min}_{φ} is the relay's minimum active time (0.05-0.2 s) that is within (2.5-10) AC-cycles for 50 Hz system, this is manufacturer dependent and $T^{max}_{\varphi,pri}$ is typically 1-2(s). Backup relays will address the issue after a coordination delay, ensuring proper coordination with the primary relay through the coordination time interval (CTI), as shown in Eq. (8). The CTI for digital relays falls typically within[0.3,0.4], but in case of systems with high inertia due to external grids/DG penetration can result in longer CTI values [1].

C. Penalized Objective Function

Optimizing coordination limits involve incorporating penalty factor components to the objective function to discourage impractical solutions. This is a common method to induce practically feasible solution set for a constrained OF. This is depicted by Eq. (9)

$$OF = \min\left(\sum_{\varphi=1}^{\varepsilon} T_{\varphi,pri} + \left(\sum_{a=1}^{\beta} Penalty(\sigma)\right)\right) \quad (9)$$

Where β denotes tandem relay pairs and $Penalty(\sigma)$ is given by expression in Eq. (10)

$$Penalty(\sigma) = \begin{cases} 0, & T^{bu}_{\varphi} - T^{pri}_{\varphi} \geq CTI \\ \Delta, & otherwise \end{cases} \quad (10)$$

Where Δ is the penalty factor in the penalized OF, intensifying the objective function's contribution during the minimization process. In the course of implementation, where ever, the conditions in Eq. (10) are broken, a penalty is added to the solution obtained, rendering it infeasible.

III. Multi-Strategy Coati Optimization Algorithm

This study introduces an improved variant of coati optimization algorithm shorted as Tuned Non-inertial T-distribution based Weighted Coati Optimization Algorithm (TNTWCOA) [15], a 2024 metaheuristic by Qi et al., inspired by coatis' hunting and escape behaviors. TNTWCOA balances search strategies through exploration and exploitation phases, emulating these natural behaviors of coatis are meticulously simulated to form the foundation of the TNTWCOA's design which optimizes relay operation times while fine tuning it towards a global optimum within constraints, like TDS, PS and CTI. Fig. 1 shows the flow chart of the algorithm.

$$\mu_{\zeta} : \Gamma_{\zeta,a} = lb_a + \psi_a(ub_a - lb_a), \zeta = 1, 2, \ldots, \upsilon, \alpha = 1, 2, \ldots, k \quad (11)$$

Where μ_{ζ} is location of ζ^{th} coati in search space. $\Gamma_{\zeta,a}$ is the value of the α^{th} decision variable for the ζ^{th} coati. lb_a and ub_a are the bounds, ζ and α range over the total coatis and decision variables, respectively. The index ζ ranges from 1 to υ (total number of coatis), and α ranges from 1 to k (total number of decision variables).

A. Phase 1: Exploration (Hunting and Attacking Strategy)

In the first phase, the best member's position is treated as the iguana's position, while this is done through tent chaotic initialization sequence. An additional perk for even distribution of search agents (coatis) on ground and trees throughout the existing search space for global exploration is attained by using a non-inertial weight factor (ώ) with random values taken in the beginning phase as the pre-iteration run is executed. Where random number attains values between within $\psi(0, 1)$ described in Eq. (15) and Eq. (16).

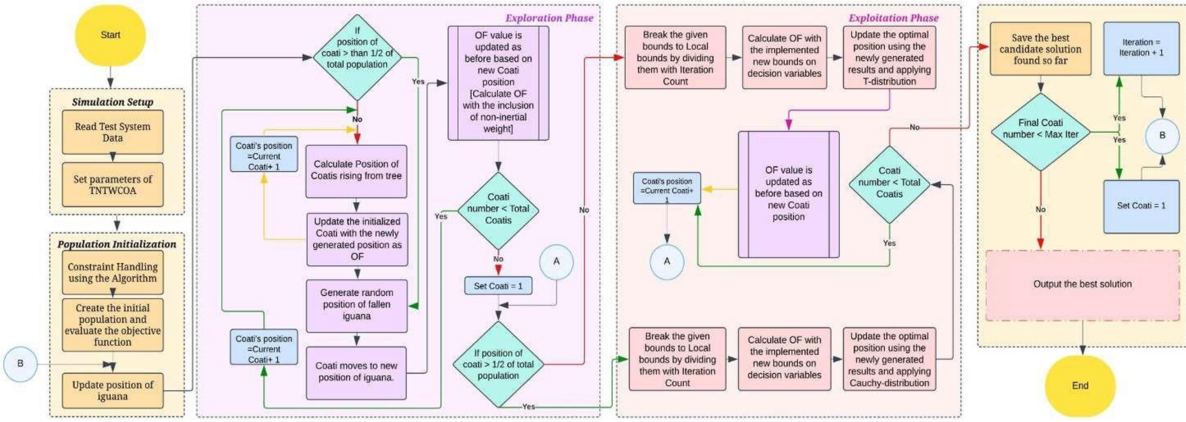


Fig. 1. Flowchart of the TNTWCOA for optimal coordination of DOCRs

$$\mu^{p1}_{\zeta}: \Gamma^{p1}_{\zeta,a} = \acute{\omega}\cdot\Gamma_{\zeta,a} + \psi(iguana_{a} - \Theta\cdot\Gamma_{\zeta,a}),\quad \zeta = 1,2,\dots,\left\lfloor\frac{\upsilon}{2}\right\rfloor, \alpha = 1,2,\dots,k. \tag{15}$$

$$iguana^{g}: iguana^{g}_{a} = lb_{a} + \psi(ub_{a} - lb_{a}),\quad \text{for } \alpha = 1,2,\dots,k \tag{16}$$

$$\mu^{p1}_{\zeta}: \begin{cases} \acute{\omega}\cdot\Gamma_{\zeta,a} + \psi(iguana^{g}_{a} - \Theta\cdot\Gamma_{\zeta,a}), & \Omega_{n} > \Omega_{\zeta} \\ \acute{\omega}\cdot\Gamma_{\zeta,a} + \psi(\Gamma_{\zeta,a} - iguana^{g}_{a}), & else \end{cases} \tag{17}$$

The coati's position is updated only if it improves the objective function; otherwise, it stays the same. This applies for $\zeta = 1, 2, \dots, \upsilon$, as per Eq. (18).

$$\mu_{\zeta} = \begin{cases} \mu^{p1}_{\zeta}, & \Omega^{p1}_{\zeta} < \Omega_{\zeta} \\ \mu_{\zeta}, & otherwise \end{cases} \tag{18}$$

The new position μ^{p1}_{ζ} for the ζ^{th} coati is $\Gamma^{p1}_{\zeta,a}$, representing its α^{th} dimension. Ω^{p1}_{ζ} is its objective function value. $iguana$ corresponds to the best position, and Θ is an integer selected from $\{1, 2\}$. $iguana^{g}_{a}$ is the randomly generated iguana's position, as its α^{th} dimension, and $\Omega_{iguana^{g}} = \Omega_{n}$ is its objective function value.

B. Phase 2: Exploitation (Escaping predator Strategy)

When threatened, a coati swiftly moves to a nearby safe spot, showcasing TNTWCOA's ability to refine local searches as outlined in Eqs. (18, 19, 20, and 21). The term $\Gamma^{p2}_{\zeta,a}$ takes values from sub-equations 20 and 21 based on specified conditions. Where Eq. (22) is adaptive T-distribution element.

$$lb^{loc}_{a} = \frac{lb_{a}}{\aleph}, ub^{loc}_{a} = \frac{ub_{a}}{\aleph},\qquad \aleph = 1,2,\dots,e \tag{19}$$

The second half includes a method called the sparrow alert mechanism, coatis in the middle will move randomly to get closer to others in the group. As:

$$\Gamma^{p2}_{\zeta,a} = iguana_{a} + iguana_{a} * f(t), \Omega^{p2}_{\zeta} > \Omega_{AVG} \tag{20}$$

$$\mu^{p2}_{\zeta}: \Gamma^{p2}_{\zeta,a} = \Gamma_{\zeta,a} + (1 - 2\psi)\cdot(lb^{loc}_{a} + \psi(ub^{loc}_{a} - lb^{loc}_{a})) \tag{21}$$

$$f(n,t) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{(\pi n)}\,\Gamma\left(\frac{n}{2}\right)} \times \left(1 + \frac{t^{2}}{n}\right)^{-\frac{n+1}{2}}, -\infty < t < \infty \tag{22}$$

Where, t is the degree of freedom parameter, n is the degree of freedom, and Γ( ) is the gamma function. As $t(n\rightarrow\infty)=N(0,1)$(Gaussian); for $t(n\rightarrow 1)=C(0,1)$(Cauchy). In TNTWCOA, coati positions are adjusted with t-distribution mutation, defined as:

$$t(x) = x \oplus xf(n,t) \tag{23}$$

A new position is valid if it enhances the objective function. This is demonstrated in Eq. (24).

$$\mu_{\zeta} = \begin{cases} \mu^{p2}_{\zeta}, & \Omega^{p2}_{\zeta} < \Omega_{\zeta} \\ \mu_{\zeta}, & otherwise \end{cases} \tag{24}$$

Here, μ^{p2}_{ζ} represents the updated position for the ζ^{th}coati in the second phase of COA. $\Gamma^{p2}_{\zeta,a}$ is its α^{th} dimension, Ω^{p2}_{ζ} is its objective function value, and $\aleph$ is the iteration index. The local bounds lb^{loc}_{a} and ub^{loc}_{a} are derived from the global bounds lb_{a} and ub_{a}, respectively.

IV. Results And Discussions

This section evaluates TNTWCOA on two IEEE test systems, including the 3-bus and 15-bus networks, to assess its effectiveness in relay coordination. Its performance is compared with established optimization algorithms [1], demonstrating superior reliability. These test systems serve as benchmarks for testing the robustness of soft computing-based methods in highly nonlinear DOCR coordination problems.

A. IEEE 3-Bus Test System

As shown in Fig. 2, the IEEE 3-bus test system consists of three generators, three lines, six DOCRs, and six primary/backup relay pairs. The optimal settings for these six relays needed to be determined. Specifically, the 12 control variables (TDS_1 to TDS_6 and PS_1 to PS_6) were adjusted to their ideal values. While TDS ranges in [0.05:1.1], PS in [0.1:5] and a CTI_{min} of 0.2s was chosen. 3-phase bolted short circuit in the middle of the lines are considered shown in Fig.2 as A, B, C. CTR of relays and relevant additional system data can be found in [1]. Table 1, details optimized TDS, PS and time of operation of primary relays. Table II presents the primary and backup operating times and enlists the CTI of relay pairs. Table III, details a result comparison of results produced by TNTWCOA with literature's state of the art techniques along with the improvement against each showing a net gain in operating times with a betterment of 4.54223, 3.98793, 1.73603, 1.13313, 0.80633, and 0.58653 seconds against TLBO [7], MDE [8], AFDBA [9], MPSO [10], GTO [11], and EMPA [1]. Fig. 3, shows the convergence curve of the OF. Additionally, Fig. 4 shows the OF variations along 30 runs, demonstrating the algorithm's minimum, maximum and mean answers for objective function graphically. Fig. 11(a) presents the graphical abstract of Table III. Fig. 4 and Fig. 5 presents the tendencies of operating times and CTI of relay pairs involved in the coordination issue. It is visibly clear that extremely precise and well-coordinated results without violations of any constraint are produced.

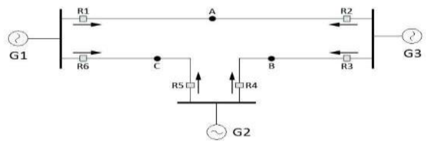


Fig. 1. IEEE 3-bus system

TABLE I. OPTIMAL RELAY SETTINGS FOR 3-BUS SYSTEM

Relay No.	Ʈ$_{\varphi,pri}$ (s)	TDS	PS
1	0.1556	0.05	3.6553
2	0.1014	0.05	1.3564
3	0.1316	0.05	3.1546
4	0.1464	0.05	2.9272
5	0.1110	0.05	1.7617
6	0.1466	0.05	2.1440
L. Ʈ$_{\varphi,pri}$	**0.79267**		

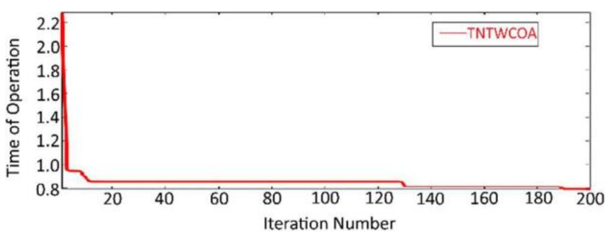


Fig. 2. Convergence Curve for the OF of IEEE 3-bus system

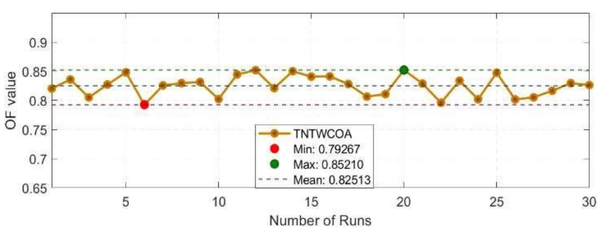


Fig. 3. OF variation along 30 runs for IEEE 3-bus system

TABLE II. OPERATING TIMES FOR PRI-BU RELAY PAIRS AND CTI FOR IEEE 3-BUS SYSTEM

Relay Pairs		Ʈ$_{pri}$	Ʈ$_{bc}$	CTI (s)
R$_{pr}$	**R$_{bc}$**			
1	**5**	0.1556	0.3813	0.2257
2	**4**	0.1014	0.3056	0.2042
3	**1**	0.1316	0.3348	0.2032
4	**6**	0.1464	0.3466	0.2002
5	**3**	0.1110	0.3110	0.2000
6	**2**	0.1466	0.3517	0.2051

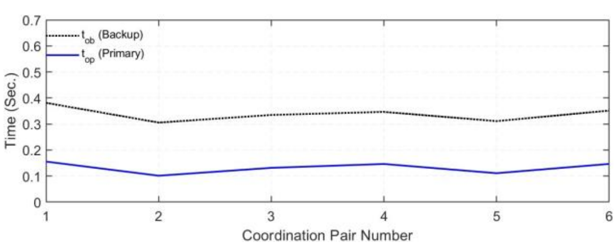


Fig. 4. Tendencies of times for all P/B relay pairs of the IEEE 3-bus system

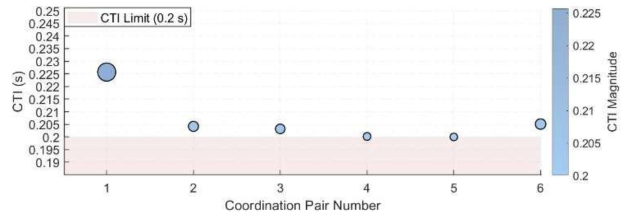


Fig. 5. CTI tendencies for 3 bus system

TABLE III. COMPARISON OF RESULTS FOR 3-BUS SYSTEM

Method	OF (s)	L. ΔƮ$_{\varphi,pri}$ (s)
TLBO [7]	5.3349	4.54223
MDE [8]	4.7806	3.98793
AFDBA [9]	2.5287	1.73603
MPSO [10]	1.9258	1.13313
GTO [11]	1.5990	0.80633
EMPA [1]	1.3792	0.58653
TNTWCOA	**0.79267**	

B. IEEE 15-Bus Test System

In Fig. 6, the single-line diagram of the 15-bus test system has been shown. It includes the placements of relays, external grid and DGs. Each generator is rated at 15MVA and 20 kV. The network consists of 21 lines and 42 relays. Power system parameters, including short circuit current values, are provided in [16]. 3-phase near-end to each relay short circuit is considered for the fault current calculations in short circuit analysis. The TDS and PS values are set within [0.05:1.1] and [0.1:5] respectively. A CTImin of 0.2 seconds was chosen. There're 84 variables with 164 constraints. Table IV shows the optimal TDS, PS and operation time values for the DOCRs, as achieved by the proposed technique where, total minimized time reaches 10.3815s for all the main relays in the system. Table V, enlists the operating times for each primary-backup pair with CTI against them. The proposed system has ensured that no coordination constraint and relay setting bounds be broken. Table VI, signifies the improvement in results produced by the proposed TNTWCOA over other state of the art robust methods published in the literature with net gain in times. The proposed method achieves net time gains of 19.51567 s against MPA [1], 8.5218 s against GA [12], 1.7307 s against IHSA [13], 1.3975s against VNS [1], 1.2884s against IHSA-NLP [13], 0.8855s against WOA [14], and 0.5795s against EMPA [1], respectively. Whereas, Fig. 7 shows the convergence curve of TNTWCOA. Fig. 8 demonstrates OF variations along 30 independent runs. And, Fig. 9 and Fig. 10 shows the graphical and bubble chart representation of tendencies in primary and backup relay times and CTI magnitude against coordinating relay pairs respectively. Fig. 11(b) outlines the net gain of TNTWCOA against other techniques in seconds in graphical notation. It is pertinent to note that however, while achieving lower relay operating times, a tradeoff is observed as backup relays' response times extend beyond 3 to 5 seconds at the cost of maintaining selectivity, a trend the OF considered in this study is tend to be consistent with findings in related studies [1]. Despite this observation in the results optimized by the proposed algorithm, the outcomes confirm that TNTWCOA successfully mitigates premature convergence through its multi-strategy framework, making it a highly effective approach for DOCR optimization.

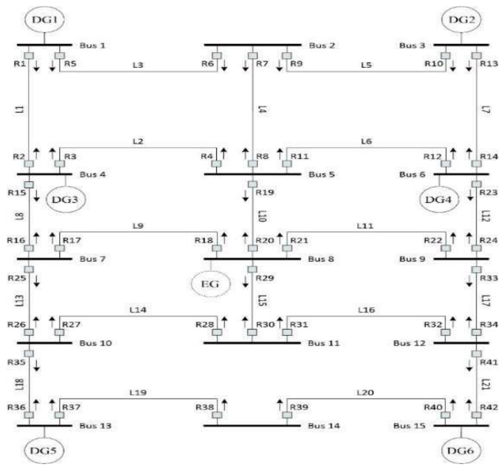


Fig. 6. IEEE 15-bus system

TABLE IV. OPTIMAL RELAY SETTINGS FOR 15-BUS SYSTEM

Relay No.	$\tau_{\varphi,pri}$ (s)	TDS	PS
1	0.2158	0.050012	4.5842
2	0.2399	0.05	4.5469
3	0.1989	0.05	4.4179
4	0.2408	0.05	4.3561
5	0.2424	0.05	4.9981
6	0.2247	0.05	4.7593
7	0.2354	0.050006	4.8077
8	0.1968	0.05	4.4083
9	0.2000	0.05	4.3929
10	0.2002	0.05	4.0672
11	0.2351	0.05	4.1715
12	0.2281	0.05	3.8560
13	0.2322	0.05	4.8154
14	0.2362	0.05	4.4571
15	0.2287	0.05	4.348
16	0.2635	0.05	4.9981
17	0.2222	0.05002	4.9662
18	0.2007	0.05	4.7436
19	0.2069	0.05	4.7325
20	0.2083	0.05	4.5874
21	0.1752	0.050685	3.5985
22	0.2083	0.05	4.6672
23	0.2443	0.05	4.9818
24	0.7254	0.05	4.1458
25	0.2531	0.05	4.8747
26	0.2335	0.05	4.3779
27	0.2857	0.05	4.9958
28	0.2242	0.050972	4.3921
29	0.2447	0.1178	1.0000
30	0.1954	0.05	3.7327
31	0.2202	0.05	4.9954
32	0.2577	0.050005	4.5139
33	0.2389	0.05	4.5337
34	0.5671	0.12135	4.9009
35	0.2104	0.05	3.3992
36	0.2139	0.05	4.1016
37	0.2194	0.05	4.2906
38	0.2711	0.050024	4.8990
39	0.2704	0.05	4.9981
40	0.2421	0.05	4.7179
41	0.2056	0.050488	4.4551
42	0.2159	0.05	4.9981
$\sum \tau_{\varphi,pri}$	**10.3815**		

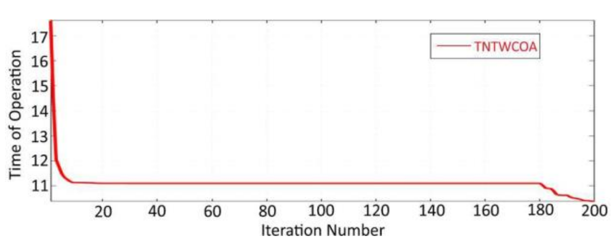


Fig. 7. Convergence curve for the OF of IEEE 15-bus system

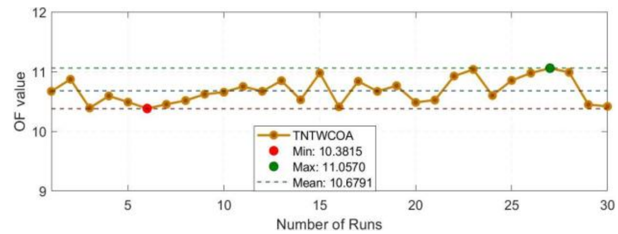


Fig. 8. OF variation along 30 runs for IEEE 15-bus system

TABLE V. OPERATING TIMES FOR PRI-BU RELAY PAIRS AND CTI FOR 15-BUS SYSTEM

Relay R_{pr}		*pri*	*bc*	
1				
2				
2				
3				
3				
4				
4				
4				
5				
6				
6				
7				
7				
8				
8				
8				
9				
9				
10				
11				
11				
11				
12				
12				
13				
14				
14				
15				
15				
16				
16				
17				
17				
18				
18				
18				
19				
19				
19				
20				
20				
20				
21				
21				
21				
22				
22				
23				
23				
24				
24				
25	**15**	0.2531	0.7270	0.4739
25	**18**	0.2531	2.5083	2.2552
26	**28**	0.2335	0.4337	0.2002
26	**36**	0.2335	0.6636	0.4301
27	**25**	0.2857	0.8026	0.5169
27	**36**	0.2857	0.6636	0.3779
28	**29**	0.2242	0.4650	0.2408
28	**32**	0.2442	1.3849	1.1407
29	**17**	0.2447	0.8493	0.6046
29	**19**	0.2447	0.5853	0.3406
29	**22**	0.2447	0.6423	0.3976
30	**27**	0.1954	0.6330	0.4376
30	**32**	0.1954	1.3849	1.1895
31	**27**	0.2202	0.6330	0.4128

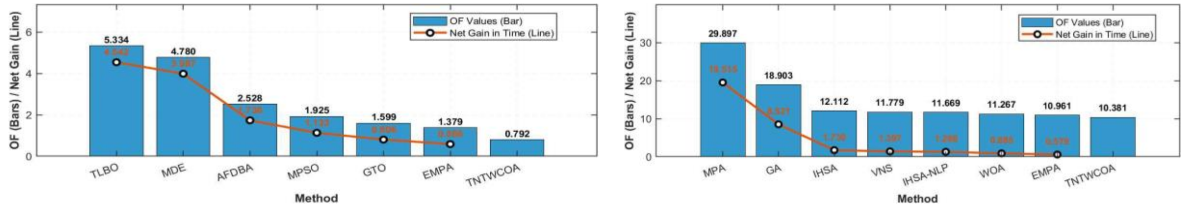


Fig. 11. Graphical Representation of Comparison of Results of TNTWCOA for IEEE a) 3-bus, b) 15-bus test systems

31	**29**	0.2202	0.4650	0.2448
32	**33**	0.2577	0.4577	0.2000
32	**42**	0.2577	2.7762	2.5185
33	**21**	0.2389	2.5133	2.2744
33	**23**	0.2389	1.7543	1.5154
34	**31**	0.5671	1.1640	0.5969
34	**42**	0.5671	2.7762	2.2091
35	**25**	0.2104	0.8026	0.5922
35	**28**	0.2104	0.4337	0.2233
36	**38**	0.2139	0.4282	0.2143
37	**35**	0.2194	0.4327	0.2133
38	**40**	0.2711	0.5612	0.2901
39	**37**	0.2706	0.4716	0.2010
40	**41**	0.2421	0.4892	0.2471
41	**31**	0.2056	1.1640	0.9584
41	**33**	0.2056	0.4577	0.2521
42	**39**	0.2159	0.4303	0.2144

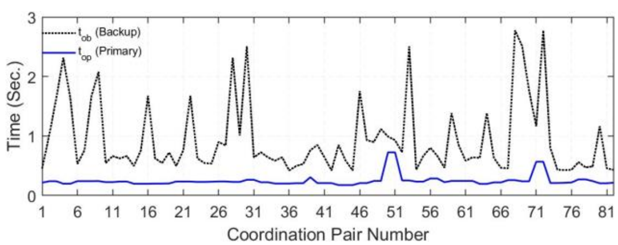


Fig. 9. Tendencies of times for all P/B relay pairs of the IEEE 15-bus

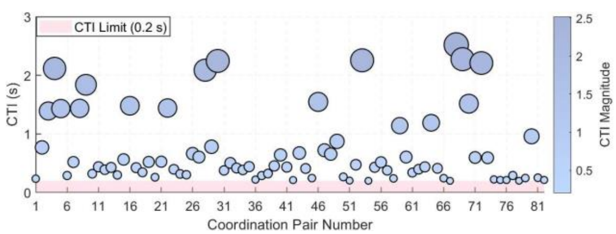


Fig. 10. CTI Tendencies for IEEE 15-bus system

TABLE VI. COMPARISON OF RESULTS FOR IEEE 15-BUS SYSTEM

Method	OF (s)	L $\Delta \mathcal{T}_{\varphi,pri}$ (s)
MPA [1]	29.89717	19.51567
GA [12]	18.9033	8.5218
IHSA [13]	12.1122	1.7307
VNS [1]	11.7790	1.3975
IHSA-NLP [13]	11.6699	1.2884
WOA [14]	11.2670	0.8855
EMPA [1]	10.9610	0.5795
TNTWCOA	**10.3815**	

V. CONCLUSION

This study improves DOCR coordination in interconnected power networks using TNTWCOA (Tuned Non-inertial T-distribution based Weighted Coati Optimization Algorithm). By integrating chaotic initialization, nonlinear inertia, adaptive T-distribution, and an alert update mechanism, TNTWCOA overcomes limitations like premature convergence and diversity loss. Applied to optimize TDS and PS under near-end and mid-faults with normal inverse characteristics, it outperforms classical and modern methods on IEEE 3- and 15-bus system for optimization of relay times.

REFERENCES

[1] O. Merabet, B. Lamine, M. Benbouzid, and N. Abdelkrim, "Optimal coordination of directional overcurrent relays in complex networks using the Elite marine predators algorithm," Electric Power Systems Research, vol. 221, p. 109446, 2023.

[2] A. Selim, S. Kamel, H. A. Jabir, and F. Jurado, "Optimal Coordination of Directional Overcurrent Relay Using Coyote Optimization Algorithm," in 2022 IEEE International Conference on Automation/XXV Congress of the Chilean Association of Automatic Control (ICA-ACCA), pp. 1-6, 2022.

[3] V. N. Rajput and K. S. Pandya, "Coordination of directional overcurrent relays in the interconnected power systems using effective tuning of harmony search algorithm," Sustain. Comput. Inform. Syst., vol. 15, pp. 1-15, 2017.

[4] A. J. Urdaneta, H. Restrepo, S. Marquez, and J. Sanchez, "Coordination of directional overcurrent relay timing using linear programming," IEEE Trans. Power Deliv., vol. 11, no. 1, pp. 122-129, 1996.

[5] T. Khurshaid, et al., "Modified particle swarm optimizer as optimization of time dial settings for coordination of directional overcurrent relay," J. Electr. Eng. Technol., vol. 14, no. 1, pp. 55-68, 2019.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[6] Y. Damchi, M. Dolatabadi, H. R. Mashhadi, and J. Sadeh, "MILP approach for optimal coordination of directional overcurrent relays in interconnected power systems," Electric Power Syst. Res., vol. 158, pp. 267-274, 2018.

[7] M. Singh, B. Panigrahi, and A. Abhyankar, "Optimal coordination of directional overcurrent relays using Teaching Learning-Based Optimization (TLBO) algorithm," Int. J. Electr. Power Energy Syst., vol. 50, pp. 33-41, 2013.

[8] R. Thangaraj, M. Pant, and K. Deep, "Optimal coordination of over-current relays using modified differential evolution algorithms," Eng. Appl. Artif. Intell., vol. 23, no. 5, pp. 820-829, 2010.

[9] F. C. Sampaio, F. L. Tofoli, L. S. Melo, G. C. Barroso, R. F. Sampaio, and R. P. S. Leão, "Adaptive fuzzy directional bat algorithm for the optimal coordination of protection systems based on directional overcurrent relays," Electr. Power Syst. Res., vol. 211, p. 108619, 2022.

[10] M. M. Mansour, S. F. Mekhamer, and N. El-Kharbawe, "A modified particle swarm optimizer for the coordination of directional overcurrent relays," IEEE Trans. Power Del., vol. 22, no. 3, pp. 1400-1410, 2007.

[11] O. Merabet et al., "An optimal coordination of directional overcurrent relays using a Gorilla troops optimizer," in Proc. 2023 Int. Conf. Adv. Electron., Control Commun. Syst. (ICAECCS), pp. 1-5, 2023.

[12] M. N. Alam, B. Das, and V. Pant, "A comparative study of metaheuristic optimization approaches for directional overcurrent relays coordination," Electric Power Syst. Res., vol. 128, pp. 39–52, 2015.

[13] V. N. Rajput and K. S. Pandya, "A hybrid improved harmony search algorithm-nonlinear programming approach for optimal coordination of directional overcurrent relays including characteristic selection," Int. J. Power Energy Convers., vol. 9, no. 3, pp. 228–253, 2018.

[14] T. Khurshaid et al., "An improved optimal solution for the directional overcurrent relays coordination using hybridized whale optimization algorithm in complex power systems," IEEE Access, vol. 7, pp. 90418–90435, 2019.

[15] Z. Qi et al., "An improved Coati Optimization Algorithm with multiple strategies for engineering design optimization problems," Sci. Rep., vol. 14, no. 1, p. 20435, 2024.

[16] A. A. Kalage and N. D. Ghawghawe, "Optimum coordination of directional overcurrent relays using modified adaptive teaching learning based optimization algorithm," Intell. Ind. Syst., vol. 2, no. 1, pp. 55–71, 2016.

Thermal Macro-modeling and Safe Operating Area Analysis of MOSFETs

M. Sudais Khan
Electronics Engineering Department
University of Engineering and Technology
Taxila, Pakistan
22-ms-enc-05@students.uettaxila.edu.pk

Yaseer A. Durrani
Electronics Engineering Department
University of Engineering and Technology
Taxila, Pakistan
yaseer.durrani@uettaxila.edu.pk

***Abstract*—Thermal dissipation in electronic circuits is always an important design constraint. Excessive heat can degrade component performance, reduce lifespan, and in severe cases, cause permanent failure. This paper uses the thermal modeling approach at the circuit level and focuses on the Safe Operating Area (SOA) of MOSFETs. The SOA defines the operational limits of MOSFETs by considering the power dissipation to prevent thermal runaway and device failure. In the area of power electronics, it is important to ensure the reliability and efficiency of circuits under different thermal conditions. In this paper, the thermal behavior of MOSFETs is modeled considering factors such as ambient temperature, gate capacitance, PCB thermal dissipation, and heatsink addition. This research highlights the importance of thermal design principles in predicting the junction and case temperatures of MOSFETs under various operating conditions. This systematic approach to thermal macro-modeling is crucial for optimizing the performance and reliability of electronic circuits, particularly in high-power applications where thermal management is a critical concern.**

***Index Terms*—thermal modeling, power electronics circuits, LTspice, Safe opearting area, mosfets**

I. Introduction

Electronic systems help us in our daily life in many ways. The next-generation power electronic devices like MOSFETs (Metal-Oxide-Semiconductor Field-Effect Transistors, (Fig. 1) can help us to reduce power dissipation. These devices generally offer superior performance, higher efficiency, and greater reliability as compared to the traditional silicon-based devices. In such devices, unexpected problems such as inadequate thermal management may affect their reliability. Therefore, it is crucial to handle thermal design issues properly. This research provides the basics of thermal design, focusing on semiconductor parts such as transistors and integrated circuits (ICs) used in electronic equipment. Thermal design aims to reduce each thermal resistance in the heat dissipation path from the chip to the atmosphere. By minimizing these resistances, the junction temperature T_J is lowered, which improves the reliability and performance of the IC. An effective thermal design ensures that electronic components operate within safe temperature ranges, preventing overheating and potential failures. This includes selecting suitable heatsinks, designing efficient airflow paths, and using thermal interface materials to enhance heat dissipation.

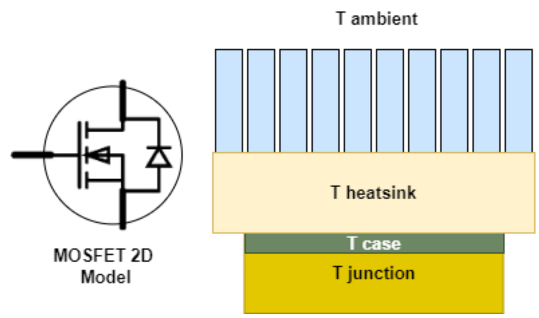


Fig. 1: MOSFET Physical Structure

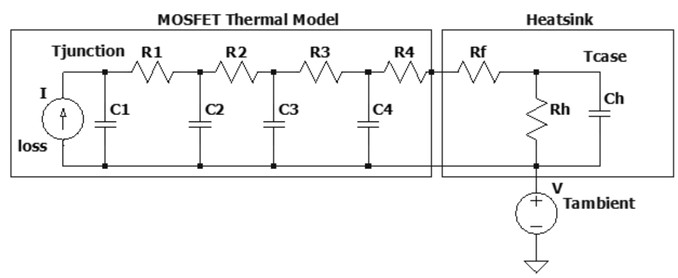


Fig. 2: MOSFET Thermal Model (Cauer)

Many researchers have addressed the challenge of managing thermal issues in power electronics-related circuits. When it comes to MOSFETs, thermal management is crucial due to the heat generated during operation. The Cauer network (Fig. 2) can be used to model the thermal behavior of MOSFETs effectively. By representing the MOSFET and its surrounding components (like the PCB and heatsink) with a Cauer network, engineers can simulate and analyze how heat is distributed and dissipated throughout the system [1].

Temperature of the environment is referred as T_A (ambient temperature) which is the minimum temperature at which the device operates. Maximum Junction Temperature or T_{Jmax} refers to the highest temperature at which the junction of a semiconductor device i.e a transistor or integrated circuit, can safely operate without risking damage or degradation to the components. The junction temeprature of the device can be lowered by adding heatsink, cooling pads or via PCB thermal dissipation [2][3].

This paper is divided into various sections, each addressing key aspects of thermal modeling and safe operating area (SOA) analysis of MOSFETs. Section II delves into thermal modeling at circuit level. It also covers thermal resistance considerations, discussing the importance of thermal resistance in predicting the thermal behavior of MOSFETs. Section III focuse on experimental analysis of the safe operating area (SOA) of MOSFET. Section IV concludes the paper summarizing the findings and contributions of the research.

II. Methodology

A. Thermal Resistance

Thermal resistance measures the difficulty of heat conduction through a material or between two points. Thermal resistance is calculated by dividing the temperature difference between two points by the heat flow between them (heat flow per unit time) [4]. Mathematically, thermal resistance R_{th} is expressed as:

$$R_{\text{th}} = \frac{T_1 - T_2}{P} = \frac{\Delta T}{P} \tag{1}$$

Where T_1 and T_2 are the temperatures at two different points, ΔT is the temperature difference between these points and P is the heat flow or the power dissipated between the points.

Thermal resistance is also similar to electrical resistance. The basic formulas for thermal calculations can be treated similarly [5]. The table I summarizes the relationships between electrical and thermal parameters.

To lower the thermal resistance, one should increase the object's surface area or choose a material with high emissivity. There are three main temperatures related to MOSFET, junction temperature, case temperature and ambient temperature. To calculate the temperature difference between the junction temperature T_J and the ambient temperature T_A, thermal resistance equation can be used provided the thermal resistance of the paths and the power loss of the IC are known. This equation helps in determining how much the junction temperature will rise above the ambient temperature due to the power dissipated by the IC [6].

$$\Delta T = R_{\text{th}} \times P \tag{2}$$

Where ΔT is the temperature difference, R_{th} is the thermal resistance and P is the power dissipation.

The equation can be further rearranged and the temperature of silicon die can be calculated as:

$$T_{\text{JUNCTION}} = T_{\text{AMBIENT}} + \theta_{\text{JA}} \times \text{Power} \tag{3}$$

$$T_{\text{JUNCTION}} = T_{\text{CASE}} + \theta_{\text{JC}} \times \text{Power} \tag{4}$$

Where $T_{JUNCTION}$ is the junction temperature, T_{CASE} is the case temperature, $T_{AMBIENT}$ is the temperature of the surroundings, θ_{JA} is the junction-to-ambient thermal resistance and θ_{JC} is junction-to-case thermal resistance.

The θ_{JA} value is used to determine the temperature rise from the ambient environment to the MOSFET's silicon die, based on a specific PCB configuration detailed in the datasheet. For instance, with a θ_{JA} of 62 K/W and an ambient temperature of 65°C, the temperature of the die will reach 127°C when the MOSFET dissipates 1W of power. This is essential to ensure that the MOSFET will remain in the safe operating temperature limits. θ_{JC} is often more valuable because it characterizes the MOSFET's behavior independently of the PCB layout [7].

B. Dissipation and Transfer of Heat in an IC

Heat transfer, the movement of thermal energy, is essential in both natural and engineered processes [8]. In electronic components, heat is dissipated through conduction, convection, and radiation. Heat generated by the chip (die) is transferred via conduction to parts of the IC package, including the die attach, lead frame, and case. Convection occurs when the PCB and IC package transfer heat to the surrounding air, creating air circulation that promotes cooling. Radiation allows heat to be emitted as electromagnetic waves, dissipating energy to the surrounding environment (Fig. 3) [9].

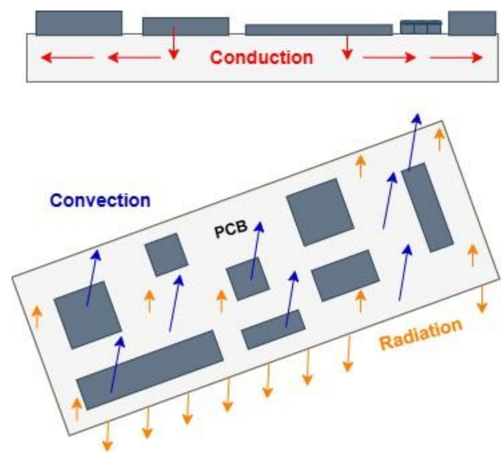


Fig. 3: PCB Heat Dissipation

C. Effect of External Cooling

The PCB layout, airflow, and heatsinks are crucial in an IC. For most modern MOSFETs with exposed metal tabs, the thermal resistance θ_{JA} is largely influenced by the PCB layout rather than the MOSFET itself, although the shape and size of the exposed pad also contribute. Since θ_{JA} is highly dependent on the PCB layout and airflow, the manufacturer's specified θ_{JA} should be used only as a rough estimate. This tells the importance of taking into account the specific layout and cooling conditions in the actual application [10][11]. The junction temperature considering PCB thermal dissipation can be estimated as:

$$T_J = T_A + P_D(\theta_{JC} + \theta_{PCB}) \tag{5}$$

Where $\theta_{PCB} \approx \frac{1}{h_{conv} \times A_{PCB}}$ represents the effective thermal resistance of the PCB, including copper planes and thermal vias (h_{conv} is the convection coefficient and A_{PCB} is the area of PCB).

TABLE I: Electrical and Thermal Equivalent Quantities

Parameter	Electrical	Thermal
Difference	Potential difference Δ V (V)	Thermal difference ΔT (°C)
Resistance	Resistance R (Ω)	Thermal resistance (Rth) (°C/W)
Flow	Current I (A)	Heat flow P (W)
Equivalent Expression	Δ V = R × I	Δ T = Rth × P

Heatsinks play a vital role in thermal management by increasing the surface area available for heat dissipation. They help in transferring heat away from the MOSFET to the surrounding environment, thereby reducing the junction temperature. The junction temperature considering the heatsink can be calculated as:

$$T_J = T_A + P_D(\theta_{JC} + \theta_{CS} + \theta_{SA}) \tag{6}$$

Where θ_{JC} is the thermal resistance from junction-to-case, θ_{CS} is from case to sink and θ_{SA} is the thermal resistance from sink to ambient.

III. Results and Discussion

The Safe Operating Area (SOA) is a vital parameter in the design and use of MOSFETs. The SOA plot, found in every MOSFET datasheet [12], outlines the maximum time a MOSFET can withstand certain voltage and current conditions without incurring damage [13][14]. Similarly, the transient thermal impedance plot provided in MOSFET datasheet [12] is essential for understanding the thermal response of MOSFETs to short-duration power pulses. In the first part of

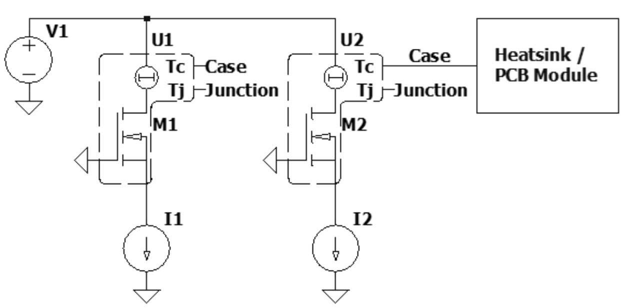


Fig. 4: Heatsink and PCB Attachment with MOSFET

the simulation, the MOSFET is analyzed first without the external cooling module and then with it attached (Fig. 4). The power dissipated here is actually the heat dissipated from the MOSFET and is calculated using equation 4 (θ_{JA} = 62°C/W, θ_{JC} = 0.5°C/W). After running the simulation, one can observe the junction and case temperatures waveforms (Fig. 5, Fig. 6 and Table II).

A. Hot Swap Application

In Hot Swap circuit design (Fig. 7), one of the biggest challenges is ensuring that a MOSFET operates within its Safe Operating Area (SOA) to prevent damage [15][16]. During normal operation, a short circuit at the output can occur, for

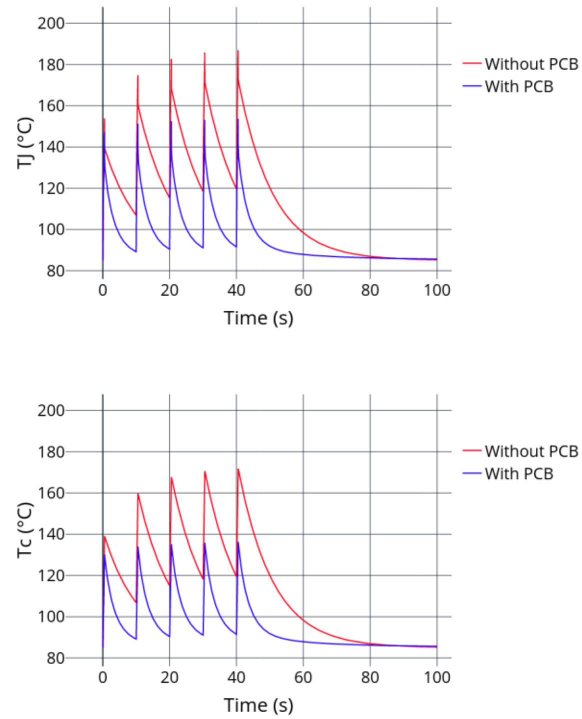


Fig. 5: Comparison of Junction and Case Temperatures with and without PCB Thermal Dissipation

TABLE II: Thermal Performance under Different Conditions

Condition	Junction Temperature (°C)	Case Temperature (°C)	Power Dissipated (W)
Without External Cooling	186	171	30
With PCB Thermal Dissipation	153	136	34
With Heatsink	133	107	52

example, if a user accidentally drops a paperclip into the chassis. A hot swap circuit (or fuse) can prevent such incidents from causing severe damage or requiring emergency services. The analysis of the junction and case temperatures in the HotSwap circuit under various conditions reveals key insights into thermal performance. Overall, the results emphasize the importance of controlling gate capacitance, optimizing thermal resistance, and implementing effective cooling strategies such

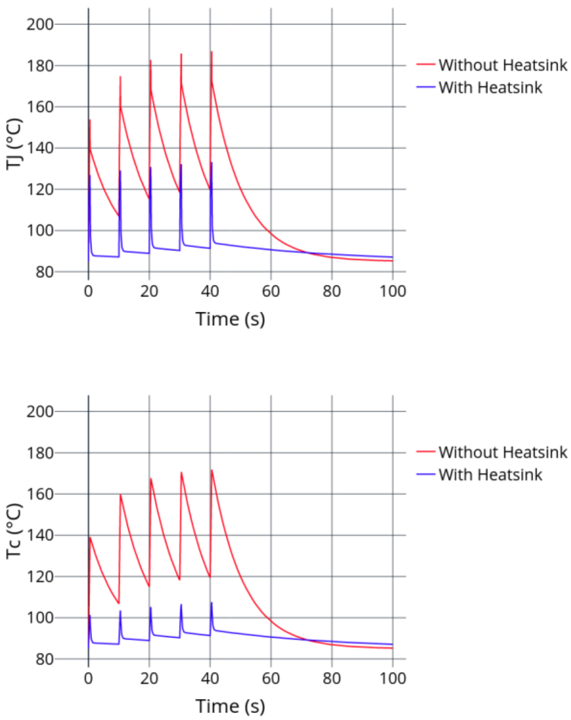


Fig. 6: Comparison of Junction and Case Temperatures with and without Heatsink

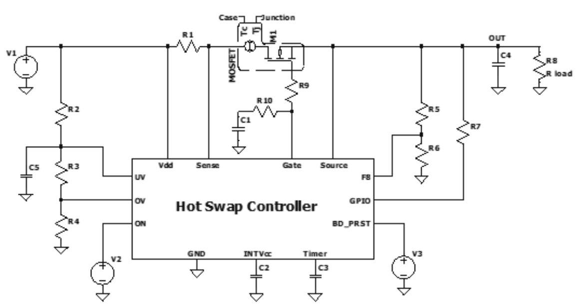


Fig. 7: Hot Swap Circuit Diagram

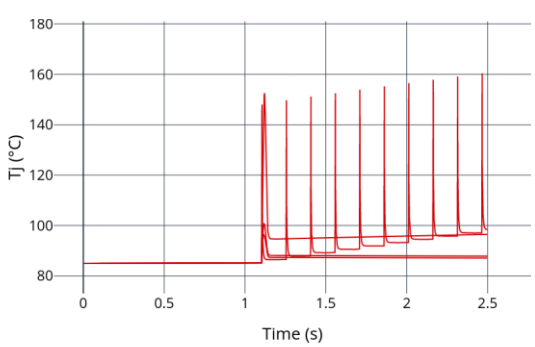


Fig. 8: MOSFET Junction Temperature with 85°C Ambient Starting Condition

as PCB thermal dissipation and heatsinks to maintain safe operating temperatures in HotSwap circuits (Fig. 8, Fig. 9, Fig. 10, Fig. 11 and Table III).

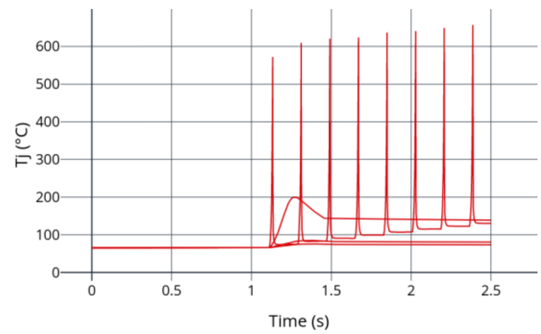


Fig. 9: MOSFET Junction Temperature with 100nf Gate Capacitance

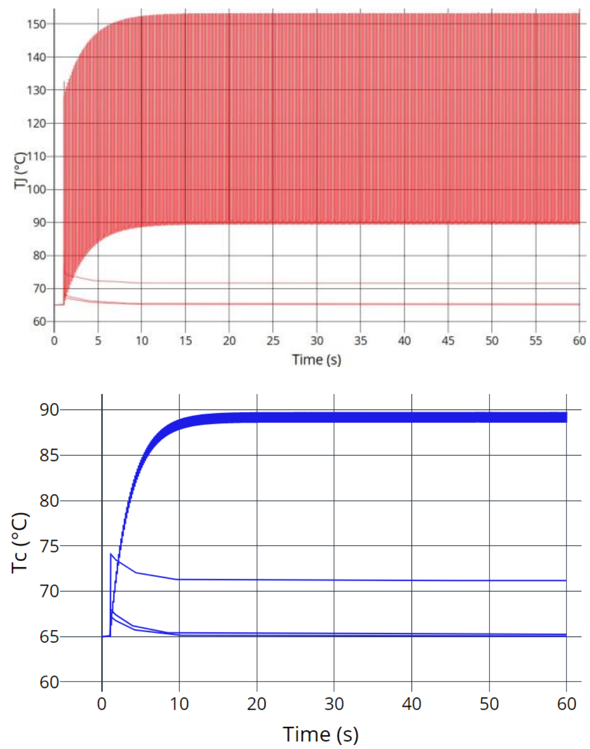


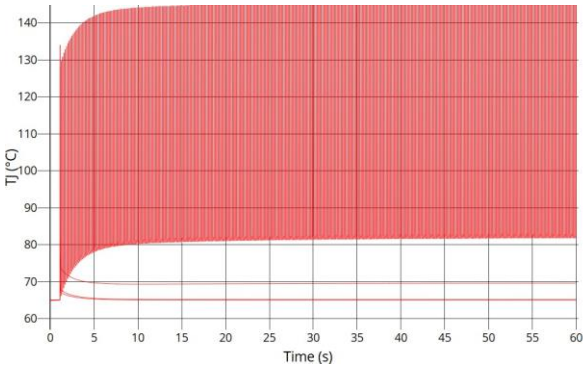
Fig. 10: Junction and Case Temperature without External Cooling

IV. Conclusion

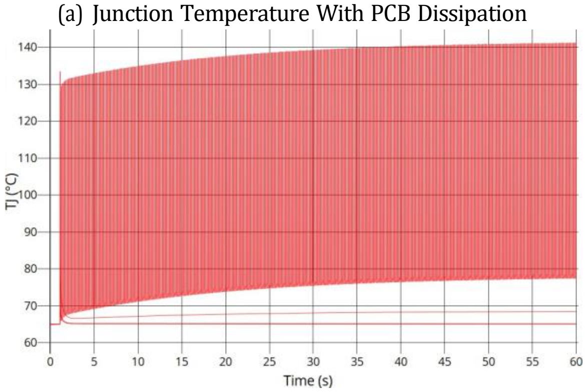
This research highlights the importance of thermal macro-modeling in the design and reliability of electronic circuits. Focusing on the Safe Operating Area (SOA) of MOSFETs highlights the importance of understanding and managing the thermal limits of these devices to prevent failure and maintain optimal performance. The analysis, which takes into account factors like ambient temperature, gate capacitance, and cooling

TABLE III: Comparison for Junction and Case Temperatures in HotSwap Circuit

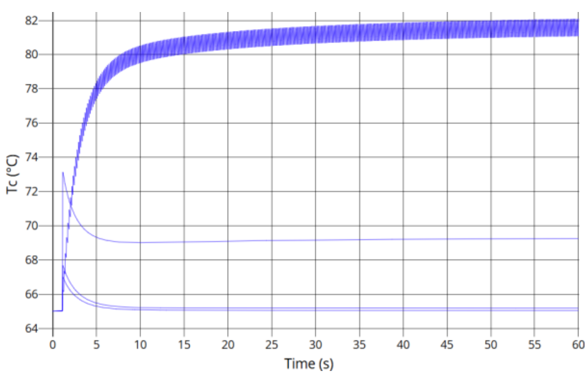
Condition	TJ-peak (°C)	TC-peak (°C)	Duration (sec)	Power Dissipated (W)	TJ Decrease (%)	TC Decrease (%)
1. Ambient Temperature						
Tambient= 100°C	176	113	2.5	1.3	-	-
Tambient= 85°C	160	98	2.5	1.3	-	-
Tambient= 65°C	140	78	2.5	1.3	-	-
Tambient= 45°C	120	58	2.5	1.3	-	-
Tambient= 25°C	100	38	2.5	1.3	-	-
Tambient= 5°C	80	18	2.5	1.3	-	-
Tambient= -10°C	65	3	2.5	1.3	-	-
Average change in junction temperature = 3.14°C against 1°C of ambient temperature						
Average change in case temperature = 1.37°C against 1°C of ambient temperature						
2. Gate Capacitance (Tambient kept constant at 65°C)						
Gate Capacitance= 100 nF	660	142	2.5	-	-	-
Gate Capacitance= 75 nF	454	124	2.5	-	-	-
Gate Capacitance= 50 nF	304	103	2.5	-	-	-
Gate Capacitance= 25 nF	193	86	2.5	-	-	-
Gate Capacitance= 10 nF	140	78	2.5	-	-	-
Gate Capacitance= 5 nF	126	74	2.5	-	-	-
Average change in junction temperature = 10.94°C against 1nf of gate capacitance						
Average change in case temperature = 5.20°C against 1nf of gate capacitance						
3. RthetaJA Modification (Tambient kept constant at 65°C)						
RthetaJA= 40	226	161	60	2.4	-	-
RthetaJA= 25	190	126	60	2.4	-	-
RthetaJA= 10	153	89	60	2.4	-	-
RthetaJA= 5	141	77	60	2.4	-	-
Average change in junction temperature = 14.19°C against 1 unit of RthetaJA						
Average change in case temperature = 8.34°C against 1 unit of RthetaJA						
4. Without External Cooling (RthetaJA kept constant at 10)						
Tambient= 100°C	188	124	60	2.4	-	-
Tambient= 85°C	173	109	60	2.4	-	-
Tambient= 65°C	153	89	60	2.4	-	-
Tambient= 25°C	112	49	60	2.4	-	-
Tambient= 5°C	93	29	60	2.4	-	-
Average change in junction temperature = 5.87°C against 1°C of ambient temperature						
Average change in case temperature = 2.33°C against 1°C of ambient temperature						
5. With PCB Thermal Dissipation						
Tambient= 100°C	181	116	60	3.9	3.7%	6.5%
Tambient= 85°C	165	101	60	3.9	4.6%	7.3%
Tambient= 65°C	145	82	60	3.9	5.2%	7.9%
Tambient= 25°C	106	42	60	3.9	5.4%	14.3%
Tambient= 5°C	86	22	60	3.9	7.5%	24.1%
Average change in junction temperature = 5.084°C against 1°C of ambient temperature						
Average change in case temperature = 1.93°C against 1°C of ambient temperature						
6. With Heatsink Attached (Rtheta = 10)						
Tambient= 100°C	176	112	60	3.6	6.4%	9.7%
Tambient= 85°C	161	97	60	3.6	6.9%	11.0%
Tambient= 65°C	141	77	60	3.6	7.8%	13.5%
Tambient= 25°C	101	37	60	3.6	9.8%	24.5%
Tambient= 5°C	81	17	60	3.6	12.9%	41.4%
Average change in junction temperature = 5.012°C against 1°C of ambient temperature						
Average change in case temperature = 1.864°C against 1°C of ambient temperature						



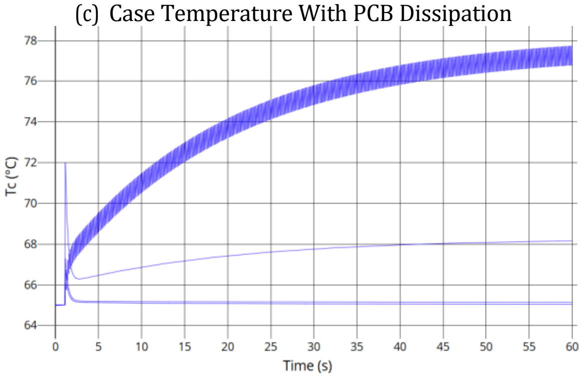
(a) Junction Temperature With PCB Dissipation



(b) Junction Temperature With Heatsink



(c) Case Temperature With PCB Dissipation



(d) Case Temperature With Heatsink

Fig. 11: Junction and Case Temperature with PCB Dissipation and Heatsink

methods, highlights significant findings. The findings of this study contribute to the advancement of thermal design practices, supporting the development of robust power electronic systems.

References

[1] Yin, Shan, et al. "Electro-thermal modeling of SiC power devices for circuit simulation." IECON 2013-39th Annual Conference of the IEEE Industrial Electronics Society. IEEE, 2013.

[2] Nayak, Debi Prasad, and Sumit Kumar Pramanick. "Implementation of an Electro-Thermal Model for Junction Temperature Estimation in a SiC MOSFET Based DC/DC Converter." CPSS Transactions on Power Electronics and Applications 8.1 (2023): 42-53.

[3] Zhang, Ziyang, Lin Liang, and Haoyang Fei. "Investigation on safe-operating-area degradation and failure modes of sic mosfets under repetitive short-circuit conditions." Power Electronic Devices and Components 4 (2023): 100026.

[4] Xu, Haochen, Xuesheng Li, and Chuan Yang. "Thermal Stability Analysis of SiC MOSFET Power Modules." 2024 IEEE International Conference on Mechatronics and Automation (ICMA). IEEE, 2024.

[5] Singh, Sujay, Joseph Proulx, and Andras Vass-Varnai. "Measuring the R thJC of Power Semiconductor Components Using Short Pulses." 2021 27th International Workshop on Thermal Investigations of ICs and Systems (THERMINIC). IEEE, 2021.

[6] The Engineering ToolBox (2003). Surface Emissivity Coefficients. [Online] Available at: https://www.engineeringtoolbox.com/emissivity-coefficients-d_447.html

[7] Heng, Ke, et al. "A temperature-dependent physical thermal network model including thermal boundary conditions for SiC MOSFET module." IEEE Transactions on Electron Devices 69.8 (2022): 4444-4452.

[8] Do Phuong Uyen TRAN, Ste´phane LEFEBVRE, and Yvan Avenas. "Discrete power semi-conductor losses vs junction temperature estimation based on thermal impedance curves." IEEE Transactions on Components, Packaging and Manufacturing Technology (2019).

[9] ROHM Co., Ltd., "Thermal design in four steps for power devices," ROHM Co., Ltd. [Online]. Available: https://fscdn.rohm.com/en/products/databook/white_paper/common/power_device_thermal_design_4steps_wp-e.pdf

[10] Schweitzer, Dirk, et al. "Transient dual interface measurement—A new JEDEC standard for the measurement of the junction-to-case thermal resistance." 2011 27th Annual IEEE semiconductor thermal measurement and management symposium. IEEE, 2011.

[11] Xu, Haochen, Xuesheng Li, and Chuan Yang. "Thermal Stability Analysis of SiC MOSFET Power Modules." 2024 IEEE International Conference on Mechatronics and Automation (ICMA). IEEE, 2024.

[12] Infineon Technologies, "IPP_B_I_200N25N3_G - Power MOSFET Data Sheet," version 2.05, Nov. 2016. [Online]. Available: https://www.infineon.com/dgdl/Infineon-IPP_B_I_200N25N3_G-DataSheet-v02_05-EN.pdf?fileId=db3a3043243b5f17012496b87e9f1971

[13] Infineon Technologies, "Dynamic thermal behavior of MOSFETs," Application Note, AN-v01_00-EN, 2017. [Online]. Available: https://www.infineon.com/dgdl/Infineon-Application_note_Dynamic_thermal_behavior_of_MOSFETs-AN-v01_00-EN.pdf?fileId=5546d4626102d35a01610888511573d1

[14] Chen, Xiaoyuan, et al. "Steady-state over-current safe operation area (SOA) of the SiC MOSFET at cryogenic and room temperatures." Cryogenics 122 (2022): 103424.

[15] Zhang, Ziyang, Lin Liang, and Haoyang Fei. "Investigation on safe-operating-area degradation and failure modes of sic mosfets under repetitive short-circuit conditions." Power Electronic Devices and Components 4 (2023): 100026.

[16] Ma, Zaojun, et al. "Characterization of Electro-Thermal Coupling Behaviors and Safe Operating Area of SiC MOSFET Modules in Pulsed Power Applications." IEEE Transactions on Power Electronics (2024).

Investigation of Improvement in current carrying capacity of various power cables using a novel arrangement

Hazrat Khan [1], Salman Amin[1]
[1]*Electrical Engineering Department, University of Engineering and Technology, Taxila, Pakistan*
hazratkhanmahsood@gmail.com, salman.amin@uettaxila.edu.pk

***Abstract*—: Power cables are a vital part of all electrical systems, and their current-handling capability directly depends on the conductor size. With the rising cost of copper, the effective utilization of a conductor's current carrying capacity has become increasingly important. To maximize this capacity, the temperature rise should be limited either through effective heat dissipation or by optimizing the orientation of cables in a trench.**

This study presents novel cable arrangements designed to decrease the working temperature of power cables and, correspondingly, increase their current handling capability. Various new orientations for laying three phase power cables in single and double circuit schemes were investigated. A precise, high resolution thermal imager was used to measure the temperature. Two of the proposed cable orientations resulted in a significant decrease in working temperature compared to existing cable orientations specified in BS 7671. The proposed novel orientations can increase the current carrying capacity by approximately 6% without increasing the cable size.

***Index Terms*—Thermal imaging for cables, power cables orientations, relationship between temperature and ampacity**

I. Introduction

Electrical power areas deal with high and low voltage cables.Each cable has its own current carrying capacity. The current carrying capacity of a cable changes with environment and physical conditions such as orientation indoor or outdoor, overhead or underground, concealed or opened, spaced or compacted. Therefore, there are two research directions about the cable current carrying capacity and temperature. One method is to reduce the cable temperature by changing the orientation. Second method is to increase the current carrying capacity by cancellation of magnetic field effect. This research proposes a novel cable arrangement aimed at enhancing heat dissipation and increasing current carrying capacity compared to traditional configrations. This study aims to identify an arrangement that minimizes cable heating and optimizes performance by analyzing various configurations and their thermal behavior.

Outdoor cable installations function better at heat dissipation compared to underground or conduit enclosed installations of cables. Cable ampacity becomes reduced when multiple cables are bundled together because bundling decreases their cooling efficiency. The temperature rise of the environment decreases cable current carrying capacity because heat dissipation becomes less effective. The thermal insulation of cables buried underground varies according to the soil water content along with its chemical composition. Thermal overheating of cables that reduces their performance can be prevented through appropriate ventilation in controlled areas. Following standards IEC 60287, NEC, BS (7671) and IEEE define a fundamental criterion for making proper cable choices. Knowledge about the maximum current capacity allows engineers to develop efficient electrical incorporations that guarantee safety and performance. They evaluate conductor properties alongside insulation together with installation orientation methods and environmental factors which enables them to select cables that fulfill operational requirements with safety and durability.

Xiaoming Xu, [1] employed COMSOL Multiphasic programming software to evaluate tunnel cable current ratings based on cable positioning together with tunnel sectional design and arrangement. The study offered recommendations to guide the process of tunnel construction combined with cable installation procedures. Chuanqiang Che, [2] presented models for cable current capacity emerged from COMSOL multiphasic simulation software while the program also linked cable temperature data. Temperature distribution patterns resulting from cable position and environmental factors and convective heat transfer received analysis in this study. The ability to handle electrical current improves when space for cables expands while effective heat dissipating approaches and proper ambient temperature management are implemented.
Adel El-Faraskoury [3] performed theoretical ampacity analysis of underground cables under steady-state conditions following IEC standards. The coding of IEC ampacity formula occurred within MATLAB before introducing a novel data preparation method based on laboratory testing. The research investigated installation factors including insulation status together with soil thermal resistance and bonding types and cabledepth because they directly affect cable ampacity.
George Callender [4] worked on low-cost flexible model for conductive heat transfer through complex burial conditions was developed through the utilization of conformal maps instead of commercial software. The established model received validation through two tests involving buried cables positioned in soil areas with varied temperatures by employing finite element analysis as a reference point. The model performed

continuous and short-term cable rating calculation. Ritthichai Ratchapan, [5] used the thermal finite element method to determine how different conduits affect low-voltage underground cable ampacity. Steel conduits provide the highest permissible ampacity among cable conduits while RTRC conduits follow then HDPE conduits afterwards and PVC conduits prove to be the lowest. The cable ampacity reduces with increased depth beneath the ground and rises when ground temperatures fall. Nishanthi Duraisamy, [6] worked on a submarine environment ampacity analysis used the finite element method (FEM) to study its determining elements. A experimental design was created for validating both the FEM simulations and numerical approach results. The inclusion of convective heat transfer calculations in the proposed method produced superior ampacity calculations than conventional techniques.

William Sundqvist, [7] presented the project analyzed power cable current capability which combines knowledge from electrical engineering and thermodynamics. The optimization of electrical losses during installation might lead to thermal inefficiency which reduces ampacity along with poor heat dissipation. The objective of attaining improved ampacity should focus on proper heat dissipation above electrical losses. Lan Xiong, [8] Investigators studied cable ampacity and temperature distributions because these factors determine power cable reliability. The current capacity drops by 12.5% when using bottom trench laying. When cables are set in non-regular patterns the temperature increases and ampacity decreases simultaneously. Online systems for multiparameter monitoring check cable temperature levels together with water height measurements and smoke density readings for improved safety management. Yifang Wang, [9] performed finite element simulation on EV cables while applying the Nelder-Mead method to calculate ampacity values. A research project evaluated the effects that ambient temperature along with insulation thickness and insulation layer thermal conductivity apply to EV cable ampacity behavior.

Abdullah Ahmed Al-Dulaimi, [10] conducted an analysis of multiple elements which affect underground cables by investigating temperature distributions in their environment as well as burial depth variables and cable spacing patterns and soil thermal resistivity properties and the impact of different backfilling materials. The study focused on enhancing power cable current capacity by perfecting multiple operational variables. Selecting proper thermal backfill with optimized spatial arrangement leads to decreased conductor temperatures along with improved ampacity performance at affordable costs. Yang Bo, [11] developed a live interaction system between power transmission capacity and system temperature. The research examined ampacity and temperature changes by alternately working with algebraic and differential equations to analyze real conditions.

Stanislaw Czapp, [12] investigated how each of solar radiation, wind conditions and cable line positioning affects power cable current-carrying capabilities. CFD simulations at an advanced computational level determined the capacity. These crucial factors stand essential for selecting power cables intended for free-air installation since they offer supplemental information to international standards. B. X. Du, [13] used of high thermal conductivity insulation resulted in amplified ampacity capabilities for buried HVDC cables. The simulation outcomes demonstrated that improved core temperature levels together with uniform insulation temperature distribution occurred as thermal conductance properties were enhanced.

Ranya Maher, [14] studied techniques to enhance the maximum current capacity of high-voltage cables that operate near closely situated high and extra-high voltage power cables. The MATLAB program served to estimate cable ampacity for varying installation configurations together with environmental conditions. A COMSOL FEM allowed researchers to investigate how heat distributed inside multiple cable layers and the adjacent soil when cables operated at maximum capacity. Lowering the thermal resistance and controlling mutual heating between circuits helps decrease cable temperatures which increases ampacity without needing bigger cross-sectional areas.

Jipeng Tang. [15] worked on several approaches to increase power cable current capacity were researched. Factors that influence ampacity consist of air and ground temperature together with backfill sand and soil covering depth and grounding mode and soil thermal resistance and the number of circuits and laying configuration. Numerical calculations alongside ampacity analysis allowed the study to evaluate these important variables as it explored methods to increase flow capabilities. Brian Scaddan, [16] worked on implementation of the IEE standard wiring regulations focused on the cable installation techniques. The research evaluated normal outcomes from different cable installation systems which included perforated and unperforated cable trays together with vertical perforated trays and cable ladders as well as wire mesh trays.

Our main contributions are as follows:

- Proposed novel three-phase cable orientation schemes (P_h and P_k) that significantly reduce operating temperatures, thereby increasing current-carrying capacity by approximately 6% without requiring larger conductor sizes.
- Designed and implemented a realistic experimental setup with thermal insulation to mimic buried or concealed cable conditions, ensuring accurate evaluation of heat dissipation and thermal performance for various orientations.
- Demonstrated that improved cable orientation can enhance electrical system efficiency and safety by optimizing thermal resistance, eliminating the need for costly infrastructure modifications or larger cable sizes.

II. Experimental Setup

The main research goal of this study was to improve the current handling capacity of the power cable through the analysis of several orientation configurations that optimized heat performance. The cables of 1.5 mm2, 2.5 mm2 and 4 mm2 were selected as our samples. A three-foot cable tray was designed and fabricated with integrated thermal insulation. The design incorporates a flexible mounting system, enabling cables to be securely fastened in various orientations, as shown in Figure 1.

Insulated foam was used to enclose the trench, preventing heat

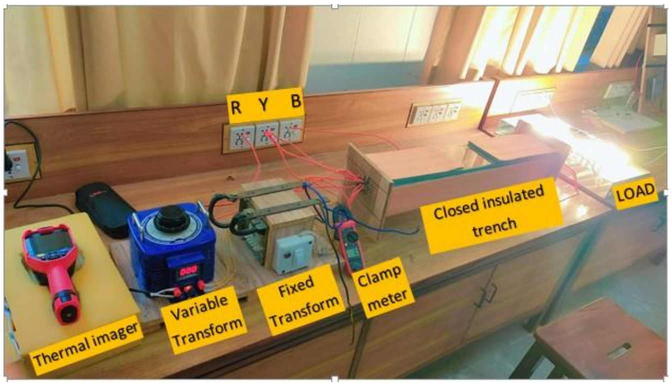


Fig. 1. Experimental Setup

dissipation to the surroundings. This simulated real conditions where cables are buried underground or concealed in walls. This design ensured that the thermal resistance matched that of actual concealed wiring and trench installations, allowing for accurate temperature and current carrying capacity analysis under similar conditions.

We investigated all the cable samples in different three-phase orientations, one by one, and subjected them to a current of 50A through each phase, using a thermal imager to measure the temperature. For 2.5 mm^2 and 4.0 mm^2 cables, the current was increased from 50A to 100A. The current was kept at a relatively high level to achieve cable heating in a short time. The proposed and existing standard orientation schemes used in the study are shown in Fig. 2.

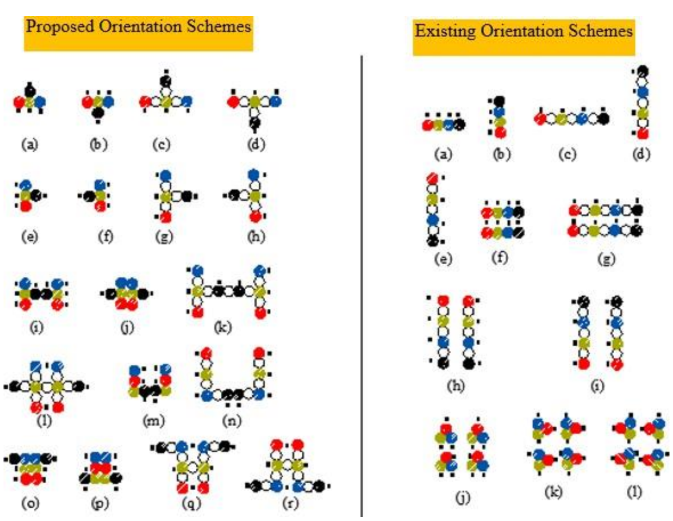


Fig. 2. Cables Orientation Schemes

We measured the temperature of the existing orientations under the same constant current of 50 A and compared the results with our proposed orientation schemes. The current was kept at a relatively high level to achieve cable heating in a short time. Finally, we analyzed and compared all the findings to determine whether any of the investigated orientation schemes resulted in a lower temperature (i.e., a correspondingly increased current capacity). The flowchart presents a clear overview of the methodology used to enhance the current carrying capacity, as shown in Fig. 3.

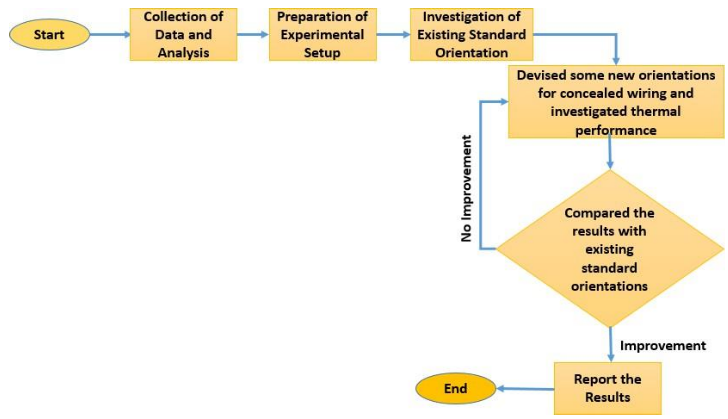


Fig. 3. Flow chart

III. Results

A. Orientation Schemes of 1.5mmSq. Cable

Fig. 4 presents the results of the proposed and existing various orientation schemes for a 1.5 mm^2 cable carrying a constant current of 50 A under the three-phase orientation scheme mentioned in Fig. 2.

In Fig. 4, the proposed scheme P_h had the lowest recorded minimum temperature at 34.6°C, whereas the existing orientation scheme E_d had a minimum temperature of 39.5°C. A lower operating temperature indicated improved heat dissipation and reduced thermal resistance, making scheme P_h more effective in managing heat compared to scheme E_d. Consequently, scheme P_h was expected to support a higher current while maintaining a lower temperature, making it a more suitable choice for enhancing the current carrying capacity of the cables.

From the Fig. 5, the proposed double circuit orientation scheme P_k had the lowest recorded minimum temperature at 37.2°C, while the existing orientation scheme E_h had a minimum temperature of 42.4 °C. A lower operating temperature indicated better heat dissipation and reduced thermal resistance, meaning scheme P_k from the proposed table demonstrated superior thermal performance compared to scheme E_h. This suggested that scheme P_k could carry more current while maintaining a lower temperature, making it a more efficient for increasing the current carrying capacity of the cables.

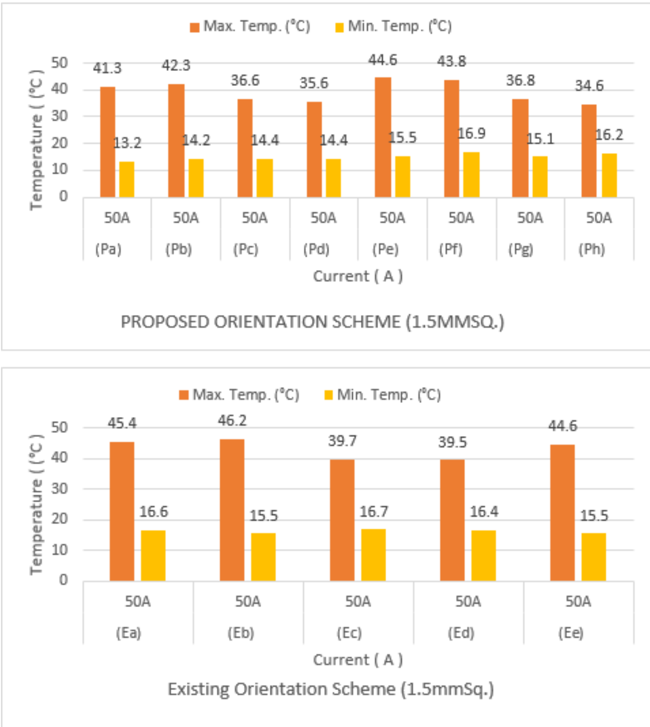


Fig. 4. Proposed and Existing Single Circuit Orientation Schemes for 1.5 mm^2 Cable

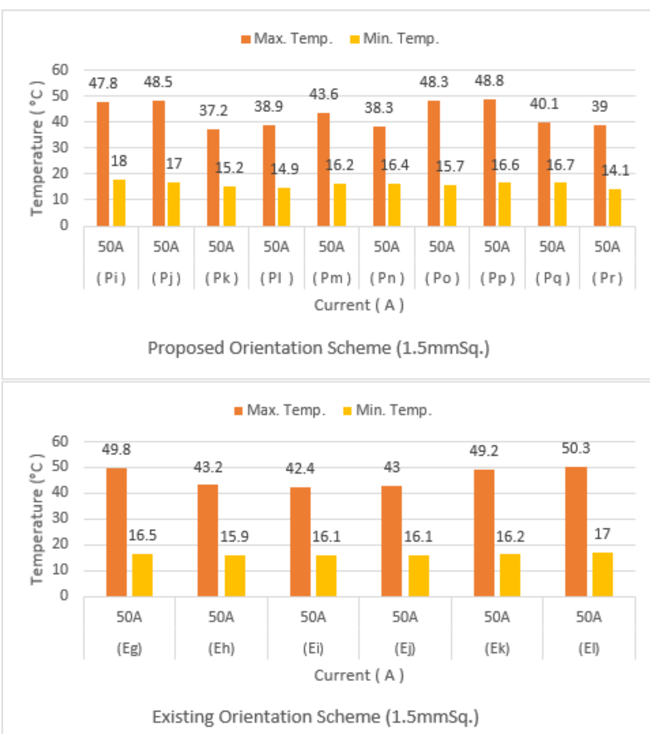


Fig. 5. Proposed and Existing Double Circuit Orientation Schemes for 1.5 mm^2 Cable

B. Orientation schemes of 2.5mmSq. Cable

Fig. 6 illustrates the results of proposed and existing orientation schemes for a 2.5mm^2 cable carrying a constant current of 100A under the three phase arrangement as mentioned in Fig. 2.

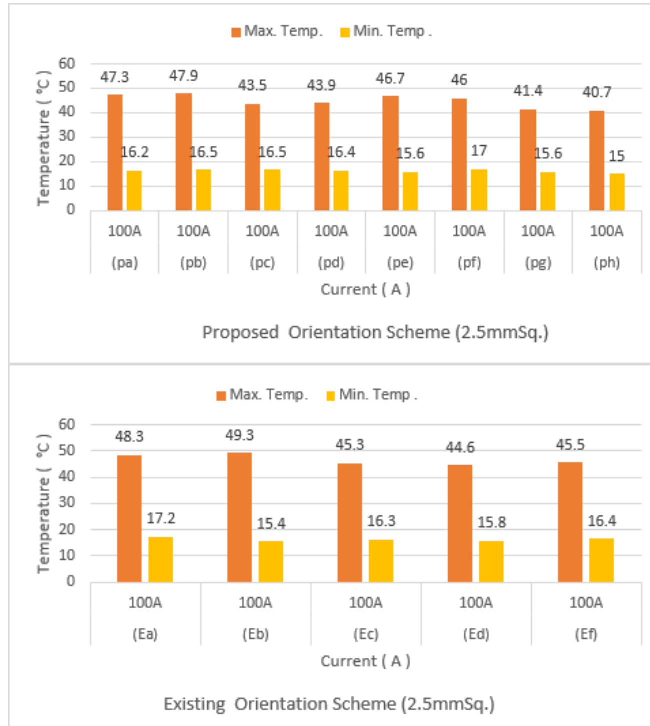


Fig. 6. Proposed and Existing Single Circuit Orientation Schemes for 2.5 mm^2 Cable

From Fig. 6, the proposed scheme had lowest recorded minimum temperature is 40.7°C in scheme P_h, whereas the minimum temperature in the existing orientation scheme E_d is 44.1°C under a constant current of 100A. Since a lower operating temperature indicates better heat dissipation and reduced thermal resistance, scheme P_h from the proposed table demonstrates superior thermal performance compared to the existing scheme. This suggests that scheme P_h can carry more current while maintaining a lower temperature, making it a more efficient option for increasing the current carrying capacity of the cables.

In Fig. 7, the proposed double circuit scheme P_k had the lowest recorded minimum temperature at 45.8°C, while as the minimum temperature in the existing orientation scheme E_i was 47.0°C under a constant current of 100A. Since a lower operating temperature indicated better heat dissipation and reduced thermal resistance, scheme P_k from the proposed table demonstrated superior thermal performance compared to the existing scheme E_i. This suggested that scheme P_k could carry more current while maintaining a lower temperature, making it a more efficient option for increasing the current carrying capacity of the cables.

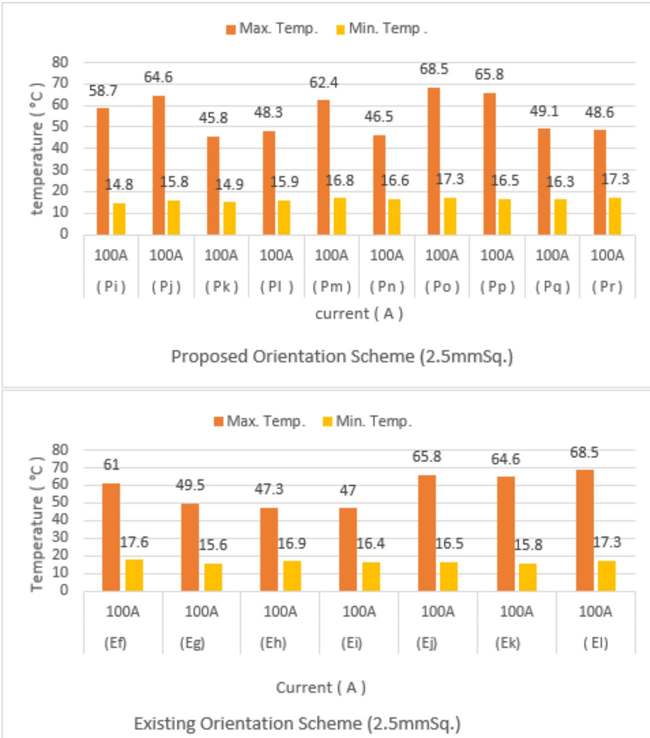


Fig. 7. Proposed and Existing Double Circuit Orientation Schemes for 2.5 mm^2 Cable

C. *Orientation schemes of 4.0mmSq. Cable*

Fig. 8 present the results of both proposed and existing various orientations of 4.0mmSq. cable carrying a constant current of 100A under three phase orientation scheme as mentioned in Fig. 2.

The minimum temperature from the proposed orientation schemes in the Fig. 8 was 35.1°C, compared to 39.8°C from the existing orientation. Since the proposed scheme resulted in a lower temperature, the current carrying capacity would have increased due to reduced resistive losses and thermal stress on the cable.

Fig. 9 present at constant current of 100A, the proposed double circuit orientation schemes resulted in a minimum temperature of 37.5°C (scheme P_k), whereas the existing orientation schemes had a minimum temperature of 42.0°C (scheme E_i). Lower temperatures generally enhanced the cable's ability to carry more current without exceeding thermal limits. Therefore, the proposed orientation schemes would have led to increased current carrying capacity.

Fig. 10 present the thermal images of proposed orientation and existing standard orientation schemes for 1.5mmSq., 2.5mmSq. and 4.0mmSq. cables.

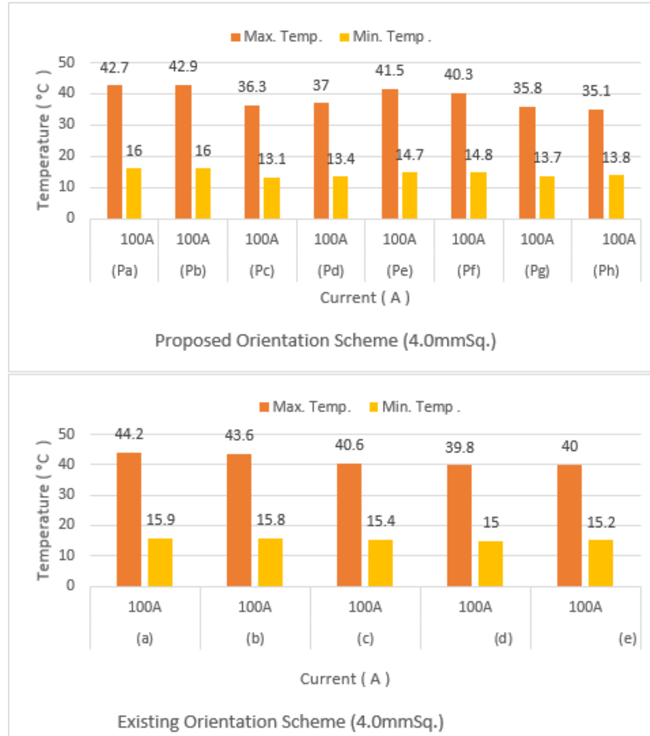


Fig. 8. Proposed and Existing Single Circuit Orientation Schemes for 4.0 mm^2 Cable

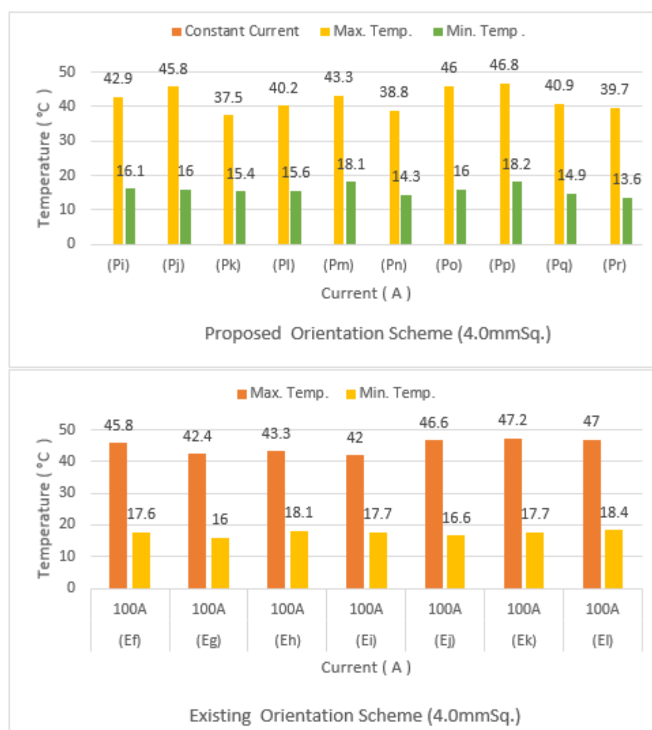


Fig. 9. Proposed and Existing Double Circuit Orientation Schemes for 1.5 mm^2 Cable

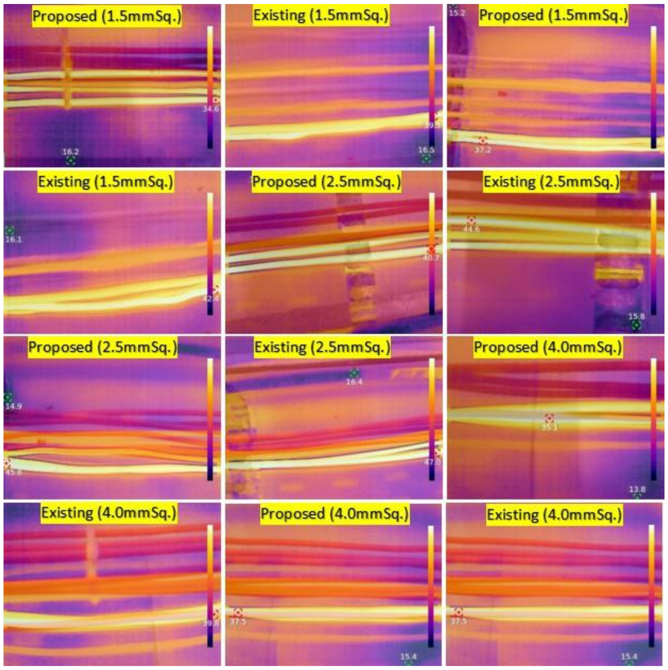


Fig. 10. Thermal image of Proposed and existing orientation schemes

IV. Discussion and Conclusion

This research investigated various cable orientation schemes that were proposed to achieve lower cable operating temperatures compared to existing standard orientations. All the proposed orientations were investigated side by side with the standard orientations. Two of the proposed cable orientation arrangements proved to be more effective at temperature regulation when a constant current of 50A and 100A was applied to 1.5mm^2, 2.5mm^2, and 4.0mm^2 cables in three phase single or double circuit configurations.

The proposed orientation P_h scheme resulted in a reduction of approximately 4.9°C in temperature for all three cable sizes investigated. Similarly, the proposed orientation P_k scheme resulted in an approximate 5°C reduction in temperature across all cable sizes tested. The optimized cable spacing and layout implementation improved ampacity by reducing thermal resistance.

A 5°C decrease in temperature corresponds to an approximate 6% increase in current carrying capacity, as supported by existing literature. The research demonstrates that strategic orientation schemes enhance cable performance without increasing cable size or trench dimensions, thereby improving system efficiency and reducing operational stress. The improved cable orientations lead to enhanced current carrying capacity through better heat dissipation and reduced thermal resistance. According to the research data, adjusting cable layout positions improves performance without altering cable size. The study reveals important insights for developing safer and more efficient electrical systems. Better cable orientation offers a practical solution for handling increased current levels in confined or insulated spaces.

References

[1] X. Xu, Q. Yuan, X. Sun, D. Hu, and J. Wang, "Simulation analysis of carrying capacity of tunnel cable in different laying ways," *International Journal of Heat and Mass Transfer*, vol. 130, pp. 455–459, 2019.

[2] C. Che, B. Yan, C. Fu, G. Li, C. Qin, and L. Liu, "Improvement of cable current carrying capacity using comsol software," *Energy Reports*, vol. 8, pp. 931–942, 2022.

[3] A. El-Faraskoury, S. Ghoneim, A. K. Alaboudy, R. Salem, and S. A. Ward, "Practical and theoretical investigation of current carrying capacity (ampacity) of underground cables," *Adv. Elect. Eng. Syst*, vol. 1, no. 3, pp. 163–169, 2012.

[4] G. Callender, K. F. Goddard, J. K. Dix, and P. L. Lewin, "A flexible model to calculate buried cable ampacity in complex environments," *IEEE Transactions on Power Delivery*, vol. 37, no. 3, pp. 2007–2015, 2021.

[5] R. Ratchapan, Y. Kongjeen, and B. Plangklang, "Ampacity analysis of low voltage underground cables in different conduits," in *2021 9th International Electrical Engineering Congress (iEECON)*. IEEE, 2021, pp. 25–28.

[6] N. Duraisamy, H. B. Gooi, and A. Ukil, "Ampacity estimation for submarine power cables installed in saturated seabed—experimental studies," *IEEE transactions on Industry Applications*, vol. 56, no. 6, pp. 6229–6237, 2020.

[7] W. A. Sundqvist, "Current carrying capacity calculations for high and medium voltage cables," 2022.

[8] L. Xiong, Y. Chen, Y. Jiao, J. Wang, and X. Hu, "Study on the effect of cable group laying mode on temperature field distribution and cable ampacity," *Energies*, vol. 12, no. 17, p. 3397, 2019.

[9] Y. Wang, J. Li, X. Kong, B. Du, J. Xu, and C. Wang, "Influencing factors on the ampacity of evs cables," in *2021 International Conference on Electrical Materials and Power Equipment (ICEMPE)*. IEEE, 2021, pp. 1–4.

[10] A. A. Al-Dulaimi, M. T. Gu¨nes¸er, and A. A. Hameed, "Investigation of thermal modeling for underground cable ampacity under different conditions of distances and depths," in *2021 5th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*. IEEE, 2021, pp. 654–659.

[11] Y. Bo, W. Kuihua, Z. Zhijie, S. Wei, W. Yiqun, Y. Shenquan, L. Rong, and F. Liang, "Study on improving the current carrying capacity of power cable in real-time operating environment," in *2016 China International Conference on Electricity Distribution (CICED)*. IEEE, 2016, pp. 1–4.

[12] S. Czapp, S. Szultka, and A. Tomaszewski, "Cfd-based evaluation of current-carrying capacity of power cables installed in free air," in *2017 18th International Scientific Conference on Electric Power Engineering (EPE)*. IEEE, 2017, pp. 1–6.

[13] B. Du, X. Kong, B. Cui, and J. Li, "Improved ampacity of buried hvdc cable with high thermal conductivity ldpe/bn insulation," *IEEE Transactions on Dielectrics and Electrical Insulation*, vol. 24, no. 5, pp. 2667–2676, 2017.

[14] R. Maher, A. El-Faraskoury, M. A. Attia, and A. Emarah, "Investigation of cable current carrying capacity improvement and its effects on transmit power between egypt and saudi arabia project," *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, vol. 8, p. 100579, 2024.

[15] J. Tang, X. Fang, X. Luo, T. Qin, T. Yang, C. Ma, and Y. Zhang, "Research on the method of improving the current-carrying capacity of power cable," in *Journal of Physics: Conference Series*, vol. 2247, no. 1. IOP Publishing, 2022, p. 012042.

[16] B. Scaddan, *IEE Wiring Regulations: Explained and Illustrated*. Routledge, 2008.

Improved Millimeter Wave Patch Antenna for Next-Generation and Beyond Networks

Amanullah
Department of Telecomm. Engineering
University of Engg. and Technology
Taxila, Pakistan
21-MS-TE-05@students.uettaxila.edu.pk

Muhammad Zahid
Department of Telecomm. Engineering
University of Engg. and Technology
Taxila, Pakistan
muhammad.zahid@uettaxila.edu.pk

Muhammad Abdul basit
Department of Telecomm. Engineering
University of Engg. and Technology
Taxila, Pakistan
abdul.basit@uettaxila.edu.pk

Yasar Amin
Department of Telecomm. Engineering
University of Engg. and Technology
Taxila, Pakistan
yasar.amin@uettaxila.edu.pk

Rao Behram Umer
Department of Electrical. Engineering
University of Engg. and Technology
Taxila, Pakistan
behram.umer@students.uettaxila.edu.pk

Aemin Zoha Khan
Department of Telecomm. Engineering
University of Engg. and Technology
Taxila, Pakistan
aeminzohak@gmail.com

Abstract—This paper presents an optimization of a compact, ultra-wideband (UWB) rectangular microstrip patch antenna (MSPA), tailored for next-generation mm-wave wireless applications. Proposed UWB antenna offers significant enhancements in gain and bandwidth. It achieves an impressive bandwidth of 36 GHz, covering the V-band (40–75 GHz), essential for high-capacity satellite communication, as well as the 61.25GHz ISM band and most of the 60GHz WiGig band. Simulations performed using CST MW Studio 2021 demonstrate that the antenna achieves a maximum efficiency of 93.3% at 44.2 GHz and a minimum efficiency of 63.1% at 66.2 GHz. A maximum realized gain is 10.2 dB at 55.8 GHz, with the lowest realized gain being 4 dB at 65 GHz. These results underscore the antenna's suitability for future 5G handheld devices and other high-frequency applications. Comparative analysis with existing designs is provided, highlighting the proposed antenna's superior performance metrics.

Keywords—5G, mm-wave, ultrawideband, MSPA, v-band, wireless communication

I. Introduction

The annual data traffic in wireless networks is experiencing an unprecedented surge, increasing by 40% to 70% each year. This exponential growth suggests that wireless networks will need to provide up to 1000 times more capacity in coming years than they currently do. To meet this rapidly escalating demand, an implementation of fifth-generation (5G) systems, which offer peak throughput capabilities of multiple gigabits per second (Gb/s), is seen as a crucial solution for future communication applications [1]– [3].

However, sub-6 GHz frequency bands, which are currently widely utilized for wireless communication, suffer from limited channel capacity and restricted bandwidth. These limitations have prompted researchers and engineers to explore underutilized millimeter-wave (mm-wave) spectrum for 5G wireless communication systems [4], [5]. The mm-wave spectrum, with its higher frequency bands, provides significantly more bandwidth and capacity, making it an essential component in addressing ever-growing data demands of modern networks.

Deployment of 5G technology is vital not only for supporting exponentially increasing data rates and accommodating the growing number of connected devices with low power consumption and high reliability, but also for enhancing the capabilities of emerging technologies. Applications such as smart cities, the Internet of Vehicles (IoV), and virtual reality (VR) will benefit significantly from the high-speed, low-latency, and robust communication provided by 5G networks [6].

However, this progress is fraught with issues that hinder the efficient implementation of mm-wave communications, such as atmospheric absorption that reduces the signal strength and path loss that reduces the length of the communication link. Special bands in the range of mm-wave include Higher frequency bands including mobile communications and usage of the ISM band ranging from 61- 61. 5 GHz [7], [8]. There was heavily rising interest in performance and new applications of mobile phones and WiGig or 60 GHz Wi-Fi, so the primary goal in designing the antennas was to increase the gain and bandwidth. WiGig currently operates in the 57 GHz to 71 GHz of spectrum and includes the existing IEEE 802. 11ad Wi Fi standard up to the prospective one IEEE 802. 11ay and facilitates the devices to communicate at multi-gigabit speed [9]. This wide frequency operation and high data rate becomes important for the development of wireless technologies and satisfy increasing demand for higher speed and reliable data transfer.

Microstrip Patch Antennas (MPAs) have been widely acknowledged often because of their low profile, low weight, compact size, low cost, plus the simplicity in fabrication. These attributes make MPAs appropriate for planar and non-planar structures particularly when used to further the development of 5G technology. For example, there are slots, which can help to increase gain, as well as efficiency [10]. To enhance the performance, some strategies including slot integration, antenna arrays, defective ground plane structures, and incorporation of metamaterials have been incorporated [11]. For example, the MPA configuration that included two vertical slots was suggested for the C-Band frequencies and the studies indicated successful results [12]. Likewise, detailed research on planar antennas revealed that it was possible to attain a gain of 14. 22dB at a frequency of 44. 8GHz and a gain of 9. 9db at 67. 8GHz and using linear array configuration [13]. Also, another patch antenna design has been proposed and has offered bandwidth of 1.318 GHz, return loss of -19. 5 dB, and resonant frequency of 24.85GHz [14]. Another design included a rectangular single patch antenna that was fed by a notch in the microstrip line; it was specifically for 5G [15]. These enhancements show the possibilities and viability of MPAs in fulfilling the tough challenges of today's wireless communication links.

In the work [17], there is described a design of four array antennas with circular polarization and SIW configuration. This design gave an axial ratio bandwidth of 14% and a gain of 15. 9 dBi. When it comes to the evolution aspect, one must bear in mind that AR bandwidth is narrower in case of circularly polarized (CP) antennas Though, SRT has been introduced in prior research [18], [19] to overcome this shortcoming. Also, this work agrees with the method called the partial ground plane, mentioned in [20], which is found to be effective in enhancing the gain and extending the bandwidth of MSPAs [21].

This paper therefore describes an ultra-wideband rectangular microstrip patch antenna that covers over 95% of the V-bands pertaining to future 5G use. The antenna also works in 61. 25 GHz ISM band and covers entire operating bandwidth of WiGig which is between 57 GHz to 71 GHz. The last optimized design shows the bandwidth is significantly wider ranging from 38 GHz to 74 GHz covers the V-band frequency range of 40- 75 GHz and, in a way, represents six distinguishable mm-wave bands. This extensive coverage positions the proposed antenna as a versatile solution for high-capacity satellite communication and advanced 5G networks, addressing the comprehensive needs of next-generation wireless technologies.

II. ANTENNA DESIGN

The design and measurements of the suggested ultra-wideband antenna are shown in Fig. 1(a) and (b).

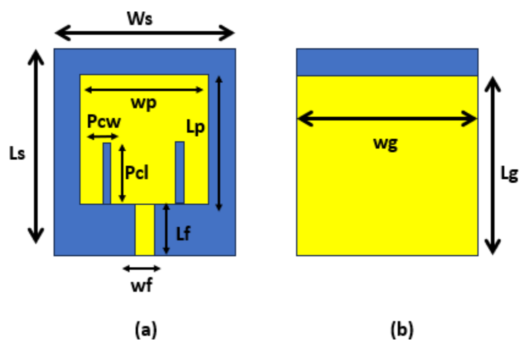


Fig. 1. The front and back view of antenna layout

Overall dimension of proposed antenna is chosen as square, which is optimally set to 14×16 mm². The substrate material used by the hardware is Rogers RT5880 high-frequency laminate with a thickness of 0. 508 mm thick, and patch size is 10×10 mm while made of copper of a thickness of 0. 035 mm, and the ground plane is measured at 14×13.9 mm made of annealed copper, and the partial ground is also included under the patch. In this case, two slots at the two sides of the presented patch enable the antenna to be tuned to the correct frequency. The layout dimensions for this design are listed in the table below; Table I: Layout Dimensions. Its UWB covers seven bands pertinent to future 5G services and six within the existing bands (42.5–43.5 GHz, 45.5 to 47 GHz, 47.2 to 50.2 GHz, and 50.4 to 52.6 GHz). Moreover, two more bands; 40. 5-42. 5 GHz and 47-47. 2 GHz are recommended for future use as stated in [16].

Table I. Antenna dimensions detail

Parameters	Description	Values (mm)
Ws	Width of substrate	16
Ls	Length of substrate	14
Hs	Hight of substrate	0.508
Wp	Width of patch	10
Lp	Length of patch	10
Tp	Thickness of patch	0.035
Wf	Width of feed	1.505
Lf	Length of feed	4
Wg	Width of ground	14
Lg	Length of ground	13.9
Pcw	Patch cut width	0.4, 0.5, 0.6
Pcl	Patch cut length	4.7

III. RESULTS

The S-parameters of proposed antenna are depicted in Figure 2 while keeping other parameters of the same as in the previous antennas reported in [22, 23], and the current modified antenna]. The antenna presented in the work [22] has a maximum bandwidth of 1.107 GHz and maximum realized gain is 7.71 dB at 37 GHz. A printed antenna in reference [23] achieves a maximum efficiency of 95. 3%, with a bandwidth of 31.7 GHz while the maximum realized gain is 7.7 dB at the operating frequency of 41.4 GHz. In similar circumstances, both antennas have a VSWR of 3:1 under respective operations, where the impedance bandwidth is regarded at -6 dB from the center value.

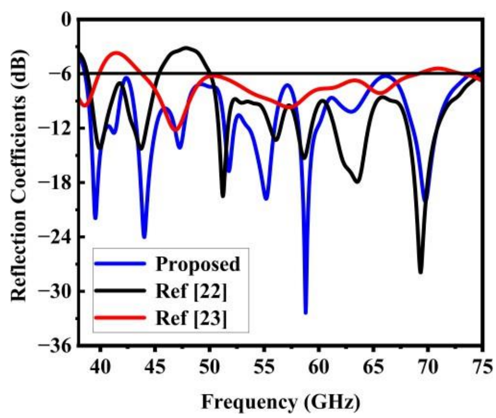


Fig. 2. S-parameters of the initially designed antenna without any modifications to its parameters and proposed modified antenna.

To achieve ultra-wideband performance, dimensions of a patch and ground plane were varied. The specific adjustments and their effects are detailed in the following subsection.

Pcw = 0.4, 0.5, 0.6, Pcl = 4.7, Wg = 14 and Lg = 13.9

The parameters pcl and pcw means the patch cut length for the desired wavelength and half patch width, respectively, and Wg and Lg means the width and length of ground plane of desired micro strip antenna respectively. Dimension is only maintained at this point is the length and width of the ground as indicated earlier. Changing the patch cut width brings the VSWR improvement in terms of the reflection coefficient manifesting a wider reflection coefficient originating from a dual band. Thus, the optimized values of the patch cut width are 0.4, 0.5, and 0.6 mm only. Reflection coefficient is represented in Figure 3 that shows reflection coefficient for the different patch cut widths.

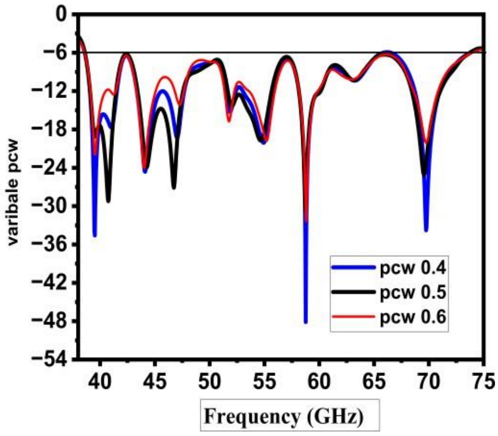


Fig. 3. Reflection coefficient of fixed patch cut length and varying patch cut width.

Pcw = 0.6, Pcl = 4.7, Wg = 14 and Lg = 13.9

To achieve the desired ultra-wideband performance, the patch cut width was increased from 0.4 mm to 0.6 mm, while patch cut length was fixed at 4.7 mm. Additionally, the ground plane width was maintained at 14 mm, but its length was reduced from 14 mm to 13.9 mm. These modifications resulted in an improved reflection coefficient, indicating enhanced ultra-wideband capabilities. The reflection coefficient after these parameter changes is illustrated in Figure 4.

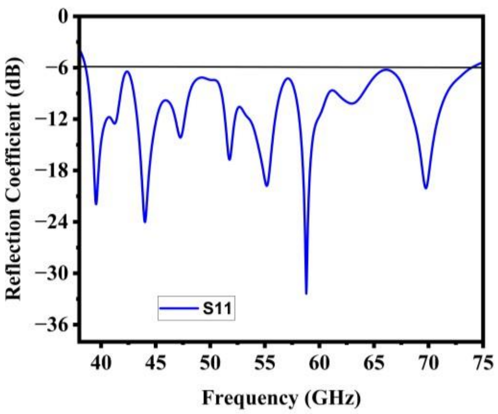


Fig. 4. Reflection coefficient of proposed and modified ultra-wideband.

In comparing the proposed antenna design to several works found in literature, there is only one perceived issue with its performance over the previously created designs. However significant advantages include getting higher gain and wider bandwidth of the circuit that uses this technique. This comparison is illustrated in the Table II below. Last optimized rectangular microstrip patch is fabricated with ultra-wide band characteristics, which operates over the range of 36 GHz. All this ranges from VHF to microwave and upper microwave frequencies useful for future 5G mobile communication devices. Thus, the proposed antenna occupies an ultra-widebssand spectrum that contains a V-band which is vital for the high-capacity satellite communication channel. It also encloses the 61. 25 GHz ISM band along with a major part of the WiGig communication band which ranges from 57 GHz to 68. 36 GHz. Specifying the antenna performance by relying on the efficiency parameters, it is possible to define that the antenna under discussion possesses a maximum efficiency of 9/3%f at 44. 2 GHz and a minimum efficiency of 6/f3%f at 66. 2 GHz. In addition, maximum realized gain of the antenna is 10. 2 dB at 55. 8 GHz, and minimum realized is 4 dB at 65 GHz. The gain and efficiency performance attributes are depicted in Figure 5 below. In this study, surface current distribution of proposed modified UWB antenna is at 69. 5 GHz as depicted in Figure 6, which proves that 90% of the physical aperture of the antenna resonates in all the frequencies. The recurrent pattern of f radiation at this frequency is shown in Figure 7 below.

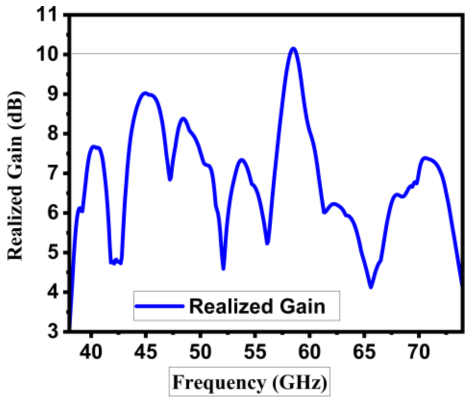


Fig. 5. Realized gain of the proposed antenna

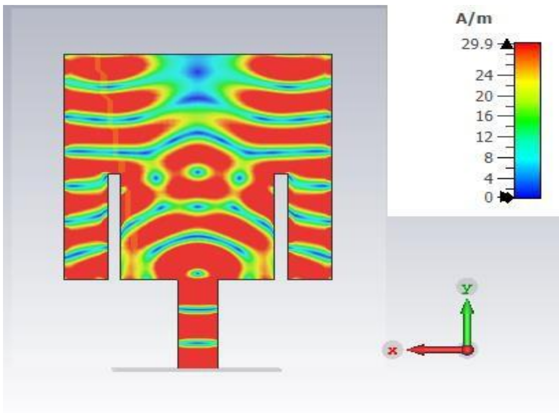


Fig. 6. Illustration of surface current distribution of ultra-wide band Antenna (UWB) at frequency of 69.5 GHz.

Table II. Comparative Analysis

Ref	Frequency	Size	Bandwidth (dB)	Gain(dB)
[20]	37 / 54	7.2 × 5 × 0.787	1.5	N/A
[21]	30.5 / 41.5	10 × 10 × 0.762	5.5 / 8.67	5.5 / 6
[22]	37	14 × 16 × 0.508	1.107	7.71
[23]	15.6 / 24.7 / 41.4	10 × 16.5 × 0.787	3.1 / 1.1 / 31.7	4.6 / 6.95 / 7.7
Prop.	38-75	14 × 16 × 0.508	36	10.2

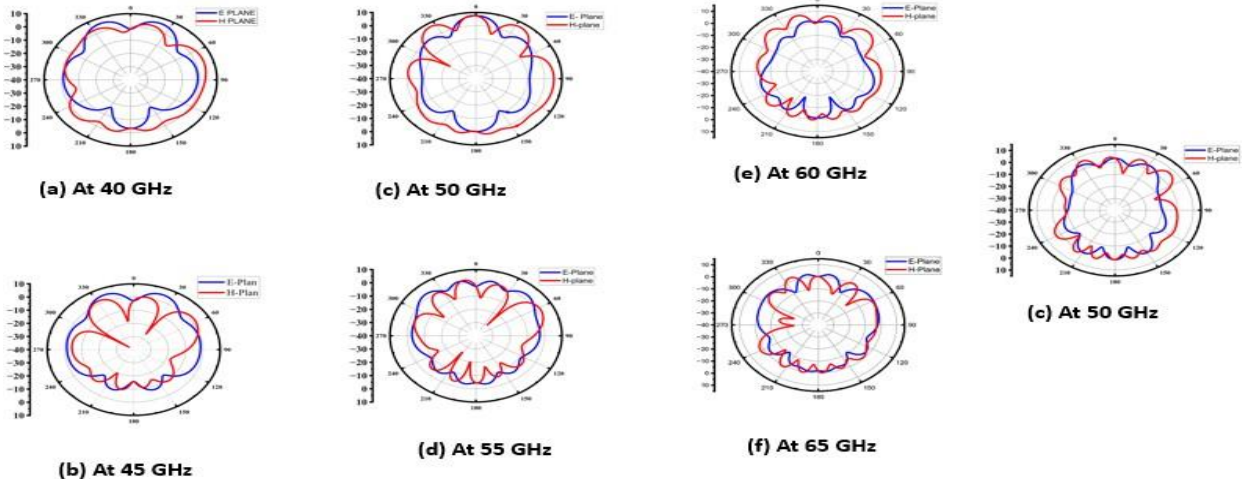


Fig. 7. The radiation pattern of a proposed antenna at different frequencies

IV. Conclusion

In this paper, the proposed modified ultra-wideband antenna represents a significant advancement in antenna design for next-generation mm-wave wireless applications. Despite its larger physical footprint compared to some existing designs, the antenna offers substantial improvements in both gain and bandwidth. It achieves a 36 GHz bandwidth, covering critical frequency bands such as V-band, ISM band, and WiGig band, thereby making it suitable for high-capacity satellite communication and future 5G handheld devices. The antenna demonstrates maximum efficiencies of 93.3% at 44.2 GHz and 63.1% at 66.2 GHz, with a maximum realized gain of 10.2 dB at 55.8 GHz and a minimum of 4 dB at 65 GHz. This research underscores the antenna's potential to meet the demanding requirements of modern communication systems. Future research directions could focus on practical implementation and optimization for real-world applications, thereby contributing valuable insights to the advancement of mm-wave antenna technology. In the future, the proposed antenna will be fabricated and tested after that; the antenna will be reoptimized for higher realized gain and efficiency.

References

[1]. Qualcomm Technologies. (Dec. 2017). Spectrum for 4G and 5G. Accessed: Jan. 5, 2019. [Online]. Available: https://www.qualcomm.com/news/media-center

[2]. J. Thompson, X. Ge, H.-C. Wu, R. Irmer, H. Jiang, G. Fettweis, and S. Alamouti, ``5G wireless communication systems: Prospects and challenges [guest editorial]," IEEE Commun. Mag., vol. 52, no. 2, pp. 62-64, Feb. 2014.

[3]. C.-X. Wang, F. Haider, X. Gao, X.-H. You, Y. Yang, D. Yuan, H. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, ``Cellular architecture and key technologies for 5G wireless communication networks," IEEE Commun. Mag., vol. 52, no. 2, pp. 122-130, Feb. 2014.

[4]. L.Wei, R. Q. Hu, Y. Qian, and G.Wu, ``Key elements to enable millimeter wave communications for 5G wireless systems," IEEEWireless Commun., vol. 21, no. 6, pp. 136-143, Dec. 2014.

[5]. S. Rangan, T. S. Rappaport, and E. Erkip, ``Millimeter-wave cellular wireless networks: Potentials and challenges," Proc. IEEE, vol. 102, no. 3, pp. 366-385, Mar. 2014.

[6]. T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N.Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, ``Millimeter wave mobile communications for 5G cellular: It willwork!" IEEE Access, vol. 1, pp. 335-349, 2013.

[7]. J.G. Andrews, S. Buzzi, W. Choi, S.V. Hanly, A. Lozano, A.C.K. Soong, and J.C. Zhang, "What will 5G be?" IEEE Journal on Selected Areas in Communications, vol. 32, no. 6, pp. 1065-1082, June 2014.

[8]. H. Mazar (3 June 2014) International, regional and national regulation of SRDs. ITU Workshop on Short Range Devices (SRD) and Ultra-Wide Band (UWB), Geneva, Switzerland,.[Online].Avalable at https://www.itu.int/en/ITUR/studygroups/workshops/ RWP1B-SRDUWB

14/Presentations/International,%20regional%20and%20national%20reg ulation%20of%20SRDs.pdf

[9]. Wi-Fi CERTIFIED WiGig, Multi-gigabit, Low Latency connectivity. [Online]. Available at https://www.wi-fi.org/discover-wi-fi/wificertified-wigig

[10]. N. Kaur and S. Malhotra, “A review on significance of design parameters of microstrip patch antennas,” 2016 5th Int. Conf. Wirel. Networks Embed. Syst., pp. 1–6, 2016.

[11]. F E. S. Misbah Un Noor and M. I. Khattak, “Design and Analysis of Dual-Band Microstrip Patch Antenna Array for 5G Cellular Communication Networks with Improved Radiation Characteristics,” in 2019 Second International Conference on Latest Trends in Electrical Engineering and Computing Technologies (INTELLECT), Nov. 2019, pp. 1–6.

[12]. Q. Deng, W. Zhang, and J. Jiang, “Design and Simulation of the Microstrip Antenna for 2.4 GHz HM Remote Control System,” Adv. Mater. Res., vol. 765–767, 2013.

[13]. A. Sohail, H. Khan, U. Khan, M. I. Khattak, N. Saleem, and J. A. Nasir, “Design and Analysis of a Novel Patch Antenna Array for 5G and Millimeter Wave Applications,” in 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Jan. 2019, pp. 1–6.

[14]. K. Bangash, M. M. Ali, H. Maab, and H. Ahmed, “Design of a Millimeter Wave Microstrip Patch Antenna and Its Array for 5G Applications,” in 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Jul. 2019, pp. 1–6.

[15]. Y. El Hasnaoui and T. Mazri, “Study, Design and Simulation of an Array Antenna for Base Station 5G,” in 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), Jun. 2020, pp. 1–5.

[16]. Final Acts WRC-15, International Telecommunication Union, World Radio Communication Conference, Geneva, 2015. [Online]. Available at https://www.itu.int/dms_pub/itu-r/opb/act/R-ACT-WRC.12-2015-PDFE. pdf

[17]. Dong-Fang Guan, Can Ding, Zu-Ping Qian, Ying-Song Zhang, Y. Jay Guo, and Ke Gong, “Broadband high-gain SIW cavity-backed circular polarized array antenna,” IEEE Transactions on Antennas and Propagation, vol. 64, no. 4, pp. 1493-1497, Apr. 2016.

[18]. Aixin Chen, Yanjun Zhang, Zhizhang Chen, and Chuo Yang, “Development of a Ka-band wideband circularly polarized 64-element microstrip antenna array with double application of the sequential rotation feeding technique,” IEEE Antennas and Wireless Propagation Letters, vol. 10, pp. 1270-1273, 2011.

[19]. Mingjian Li, and Kwai-Man Luk, “Low-cost wideband microstrip antenna array for 60-GHz applications,” IEEE Transactions on Antennas and Propagation, vol. 64, no. 4, pp. 3012-3018, Jun 2014.

[20]. M. Zahid, S. Shoaib, and M. Rizwan, “Design of MIMO antenna system for 5G indoor wireless terminals,” in Proc. 2019 International Conference on Engineering and Emerging Technologies (ICEET), Lahore, Pakistan, 21-22 Feb. 2019, pp. 1-4.

[21]. N. Ashraf, O. Haraz, M.A. Ashraf, and S. Alshebeili, “28/38-Ghz dual-band millimeter wave SIW array antenna with EBG structures for 5G applications,” in Proc. International Conference on Information and Communication Technology Research (ICTRC), Abu Dhabi, United Arab Emirates, 17-19 May 2015, pp. 5-.

[22]. Riaz, M. J., Sultan, A., Zahid, M., Javed, A., Amin, Y., & Loo, J. (2020, November). MIMO antennas for future 5G communications. In 2020 IEEE 23rd International Multitopic Conference (INMIC) (pp. 1-4). IEEE.

[23]. Zahid, Muhammad, et al. "Ultra-wideband antenna for Future 5G." 2020 IEEE conference of Russian young researchers in electrical and electronic *engineering (EIConRus)*. IEEE, 2020.

Performance Analysis of a Lifi System Based on VLC Over a LOS Channel with and Without Ambient Light Using Optisystem

Rao Behram Umer
Department of Electrical. Engineering
University of Engg. and Technology
Taxila, Pakistan
behram.umer@students.uettaxila.edu.pk

Amanullah
Department of Telecomm. Engineering
University of Engg. and Technology
Taxila, Pakistan
21-MS-TE-05@students.uettaxila.edu.pk

Muhammad Abdul basit
Department of Telecomm. Engineering
University of Engg. and Technology
Taxila, Pakistan
abdul.basit@uettaxila.edu.pk

Farzana kulsoom
Department of Telecomm. Engineering
University of Engg. and Technology
Taxila, Pakistan
farzana.kulsoom@uettaxila.edu.pk

Sidra khadam
Department of Telecomm. Engineering
University of Engg. and Technology
Taxila, Pakistan
sidrakhadim343@gmail.com

Aemin Zoha Khan
Department of Telecomm. Engineering
University of Engg. and Technology
Taxila, Pakistan
aeminzohak@gmail.com

***Abstract*—** Li-Fi (Light Fidelity) stands as a state-of-the-art wireless technology which sends data through light-based transmissions. Mobile robots require effective indoor location systems because they operate within hospitals museums and airport interiors. Researchers have studied the behavior of On-Off Keying (OOK) modulation technique used in Li-Fi systems by observing the impact of background interference. Our model determines the performance of a 550 nm wavelength white LED transmitter using Opti system software. Evaluation of the system occurs through examinations under two conditions: when ambient light noise exists and when it does not exist. The research outcomes demonstrate that Li-Fi technology has the ability to deliver dependable high-speed indoor localization services for environments experiencing changes in ambient lighting conditions. Simulation findings indicate a Q factor measurement of 19.81without noise while the results show 6.72when noise is present. The network supports 10 Gbps data transmission at 2.38e-82 Bit Error Rate without ambient noise and 5.5e-12 Bit Error Rate with ambient noise under a 10 meters connection range.

***Keywords*—lifi, LED, LOS channel, Optisystem, wireless communication**

I. Introduction

Wireless communication technologies changed the direction of multiple business sectors while creating better connection capabilities and enabling various novel uses. The field of wireless communication gains major promise through Li-Fi which operates at high data speed using visible light. Li-Fi operates under Visible Light Communication (VLC) as a distinct system which substitutes radio frequency (RF) systems through data transmissions that use light-emitting diodes (LEDs).

Mobile robots used extensively inside museums and airports as well as hospitals and other facilities necessitate accurate and dependable indoor localization systems for operating effectively. Standard RF communication systems experience three substantial issues including signal interference along with restricted bandwidth capabilities and security vulnerabilities. The visible spectrum enabled VLC systems present several benefits such as enhanced bandwidth together with better security measures and decreased electromagnetic disturbances.

Real-world VLC system performance suffers interference from ambient light sources which constitute one of the system's operation-limiting elements. Noise from environmental light sources including fluorescent lamps makes the VLC system less efficient for its intended operation. Evaluation of VLC system performance requires understanding the impact of environmental noise elements during real-world operations.

Opti system software allows the design of a VLC system simulation along with its analysis of a 36W fluorescent lamp operating at 50Hz ambient light noise source. The paper splits into distinct subsections for better clarity and consistent presentation. Section II explains the Visible Light Communication system principles through theoretical and modelling descriptions. The paper details simulation and evaluation techniques alongside description of methodologies that were employed in Section III. Section IV both displays the experimental findings as well as provides application insights into the research. The paper ends with Section V that summarizes main conclusions and evaluates the system work's ability to achieve its goals.

.

Theory and Modeling of Visible Light Communication System

The research on indoor Visible light communication (VLC) systems appears in this literature review [1] as it explores ways to enhance wireless communication in difficult settings including hospital rooms and aviation cabins. Research into indoor VLC benefits greatly from Opti System simulation software which enables optical communication system modelling and simulation [2]. Academic researchers use this controlled environment to conduct experiments regarding various network setups in addition to modulation techniques under different transmission circumstances. LED technology advancements have caused a major improvement in VLC system performance. The research by Tsonev et al. [3] demonstrates how multi-level pulse amplitude modulation achieves several gigabit-per-second data rates.

The implementation of indoor VLC systems faces three main issues which are specified in [4] namely interference as well as constrained mobility and limited coverage range. The combination of signal processing algorithm advancement and modern modulation methods requires supporting technological improvements in optical receiver design to address the emerging challenges. The author employs Li-Fi technology that sends data through visible light communication according to reference [5]. Li-Fi provides better data rates and expanded bandwidth together with electromagnetic interference resistance when compared to standard wireless systems. Performance quality of Li-Fi systems heavily depends on modulation techniques because they influence product noise immunity alongside spectral efficiency and maximum data speed. Various techniques depicted in [6] serve a crucial function for this context which includes On-Off Keying (OOK), Pulse Amplitude Modulation (PAM), Orthogonal Frequency Division Multiplexing (OFDM), and Subcarrier Multiplexing (SSM). According to Khalid et al. [7] ambient noise plays a crucial role in determining the performance of Li-Fi systems operating within an indoor setting. The importance of noise reduction depends on two key elements including equalization algorithms together with adaptive modulation. The basic technique of OOK modulation functions as an encoding method in optical communication networks as described in [8]. The technology finds broad application due to its convenient operation and compatibility with different systems. Line of Sight (LOS) channels create unobstructed communication pathways for rapid long-distance data transfer through empirical evidence in [9]. Satellite communication systems together with outdoor optical connections find essential value from this technology. Investigatory studies [10] explored optical communication system susceptibility to ambient noise while developing techniques to minimize the noise which enhances system performance and reliability.

Our paper explores the various advantages of Li-Fi technology invented by Harald Haas against Wi-Fi regarding bandwidth effectiveness and security and availability [11]. Li-Fi enables fast wireless data transfer and dual-direction communication through its utilization of LED bulbs as stated in reference [12]. Technology experts selected Quadrature Amplitude Modulation (QAM) as an optimal modulation approach because it delivers superior spectral efficiency. The wireless transmission methods consist of On-Off Keying (OOK), Pulse Position Modulation (PPM), Orthogonal Frequency-Division Multiplexing (OFDM), and Color Shift Keying (CSK).

Li-Fi operates across numerous applications including wireless mobile networks together with intelligent lighting technology for education purposes and traffic management functions. The investigation focuses on VLC system development methods. VLC systems employ LED light emission because they deliver longer operational lifetime as well as lower electricity demand. The paper reviews technological simulation models together with applications and past research as well as existing difficulties in this field. Reference [14] demonstrates recent successes in VLC with high data rates which emphasizes receiver illuminance assessment in system design while extending the communication coverage to 40 cm.

In the study mentioned in reference [15], reliable modulation schemes for optical communication systems are evaluated through an examination of the application of On-Off Keying Non-Return-to-Zero (OOK NRZ) modulation in a simulated environment. It considers several factors, including photo efficiency, bandwidth, interference susceptibility, cost-effectiveness, and production viability. In the meantime, research described in reference [16] uses Opti System 15 and MATLAB for simulations to evaluate the performance of Li-Fi communication under various modulation techniques and ambient noise conditions. According to this study, QPSK, or quadrature phase shift keying, performs better in typical lighting situations.

Effective modulation techniques like Subcarrier Multiplexing (SCM), On-Off Keying (OOK), Pulse Width Modulation (PWM), M-Pulse Amplitude Modulation (M-PAM), Orthogonal Frequency Division Multiplexing (OFDM) variants, and unique color domain modulation are required for Li-Fi, which combines LED communication with illumination [17]. The design and simulation of Line-of-Sight (LOS) and Non-Line-of-Sight (NLOS) multipath VLC models using LEDs are examined in this study [18], with particular attention paid to elements like transmission quality, reflection, and Field of View (FOV). Utilizing a spectrum 100,000 times larger than radio frequencies, Li-Fi technology in [19] enhances overall performance, power efficiency, and data throughput. With transfer rates of up to 2 GB/s, this makes wireless data transfer quick and affordable, which makes it perfect for critical environments.

LEDs are used in Li-Fi, a cutting-edge wireless technology, to send data across the visible light spectrum. Numerous advantages come with this method, such as increased speed, low latency, increased security, and a wider range of applications. Li-Fi is applicable to a broad range of platforms and industries and has the potential to solve issues like radio-frequency bandwidth limitations [20].

The cutting-edge optical wireless communication technology known as VLC has the potential to completely transform green lighting. Through the provision of both illumination and data transmission, VLC makes use of the visible light spectrum, which extends from 380 nm to 780 nm in wavelengths. White LEDs serve as the light source, a free space channel serves as the transmission medium, and a photodetector serves as the receiver in a typical VLC system. Compared to traditional fiber optic cables, the use of a free space channel offers several advantages, such as the removal of physical cables, lower installation costs, faster data transmission speeds, and the prevention of RF interference in research [21].

The present literature review looks at several problems related to systems for VLC. Many applications depend on the basic idea of using an LED source in VLC systems. Comparing LED lighting to traditional illumination systems reveals advantages like lower power consumption and longer lifespan. To better assess the performance of VLC, this review [22,23] aims to provide a thorough understanding of its various facets and to simulate a VLC system model.

This review aims to explore the essential ideas, real-world uses, and intrinsic difficulties related to VLC. The intricacies and auspicious potential of VLC technology are examined, along with simulation tools designed specifically for VLC systems, the incorporation of VLC in audio and video transmission, cutting-edge encryption techniques employing light-emitting diodes, and the potential application of VLC in vehicle-to-vehicle communication. Within the various applications of VLC systems, important challenges like reaching high data rates, extending transmission ranges, optimizing line-of-sight (LOS) conditions, and mitigating interference are critically analyzed in [24,25,26].

digital data transmission encoding. An electrical pulse generation tool which implements the NRZ encoding strategy

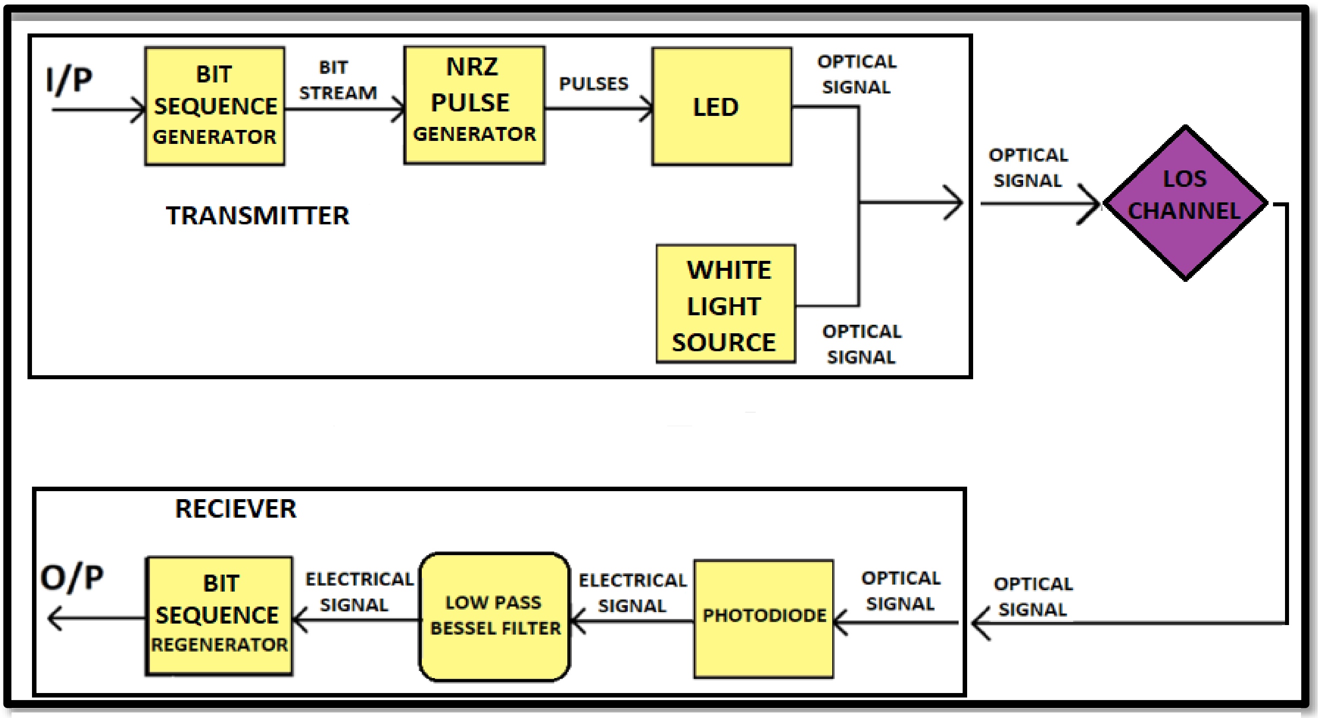


Figure1: Block diagram of the lifi system los channel with ambient light.

II. THE METHODOLOGY AND SIMULATION TECHNIQUE

The research methodology uses OptiSystem to evaluate a Visible Light Communication Wi-Fi system through environment modeling between room characteristics and ambient noise levels followed by system configuration design within the software. The system chooses light sources together with photodetectors as well as modulation approaches. The simulation process generates performance data which contains received optical power combined with bit error rate alongside signal-to-noise ratio. The signals undergo an iterative optimization process that enhances their quality and reliability level. The research implemented Li-Fi technology for the development of a Visible Light Communication system. A system's block diagram under the influence of ambient light follows below in Fig1 and 2.

A. Transmitter Side

The transmitter side of an optical communication system holds all required pieces and operations needed to convert digital data to optical signals suitable for transmission through wireless channel. A random binary bit sequence enables the system to produce data signals for simulation purposes. The transmission of digital data as binary bits (0s and 1s) begins after formatting by the bit sequence generator. The functional operation of optical communication systems relies on light pulse transmission instead of electrical signals so encoding methods become fundamental for system operation. The process uses it for data encoding. The device implements Non Return-to-Zero (NRZ) pulse creation for

operates under the name Non-Return-to-Zero (NRZ) pulse generator. Digital data streams use NRZ encoding to portray each bit value by specific pulse forms together with low or high voltage levels. The data transmission indicates a '1' bit through a high voltage pulse or a pulse while the '0' bit corresponds to a low voltage state or no pulse at all. The device serves as an optical transmitter which sends light signals according to the encoded data. The conversion of electrical energy into light happens efficiently through LED devices. The high efficiency of VLC systems depends on this feature that optimizes the conversion of data signals into light signals.

B. LOS Channel with and without the influence of ambient light:

Transmission between a sender and receiver takes place through LOS when their optical signal passes directly toward each other with no obstacles to impair it. VLC relies on unobstructed optical signal transmission of visible frequencies between light and photodetector devices. Design and optimization of VLC systems requires LOS channels to attain efficient and reliable communication operation. Such predictions enable the evaluation of signal strength together with coverage parameters along with expected obstacles present in the LOS link.

Several elements which affect LOS propagation in VLC systems require equal attention to their maintenance: 1. The maximum intensity of LOS signals occurs when the distance remains short. The rise in distance leads to reduced signal strength because of dispersion together with attenuation. 2. Weather conditions that include fog, rain and humidity create LOS propagation complications by scattering light and through absorption of radio waves. 3. Physical obstructions such as walls, furniture and human bodies positioned along LOS paths result in signal signal reduction or degradation quality. 4. The quality of LOS signal depends on transmitter power output alongside receiver sensitivity as well as the selected optical system components.

With demodulation systems extract the modulating signal through procedures matching to the OOK modulation pattern. Signal extraction occurs with precision while the signal becomes ready for decoding at the next step. Components: 1. The PIN photodiode functions as a semiconductor device that

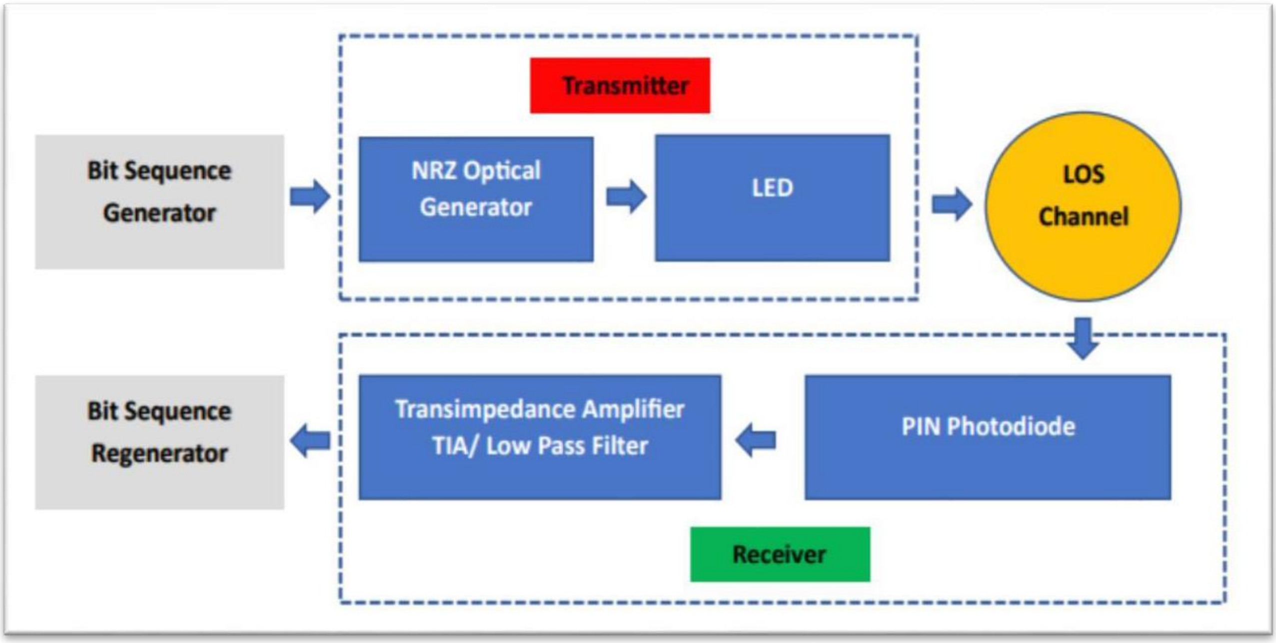


Figure2 : Block diagram of the lifi system los channel without ambient light

finds applications in optical communication and imaging together with sensing functions. The photodiode contains three distinct zones consisting of P-type, Intrinsic and N-type

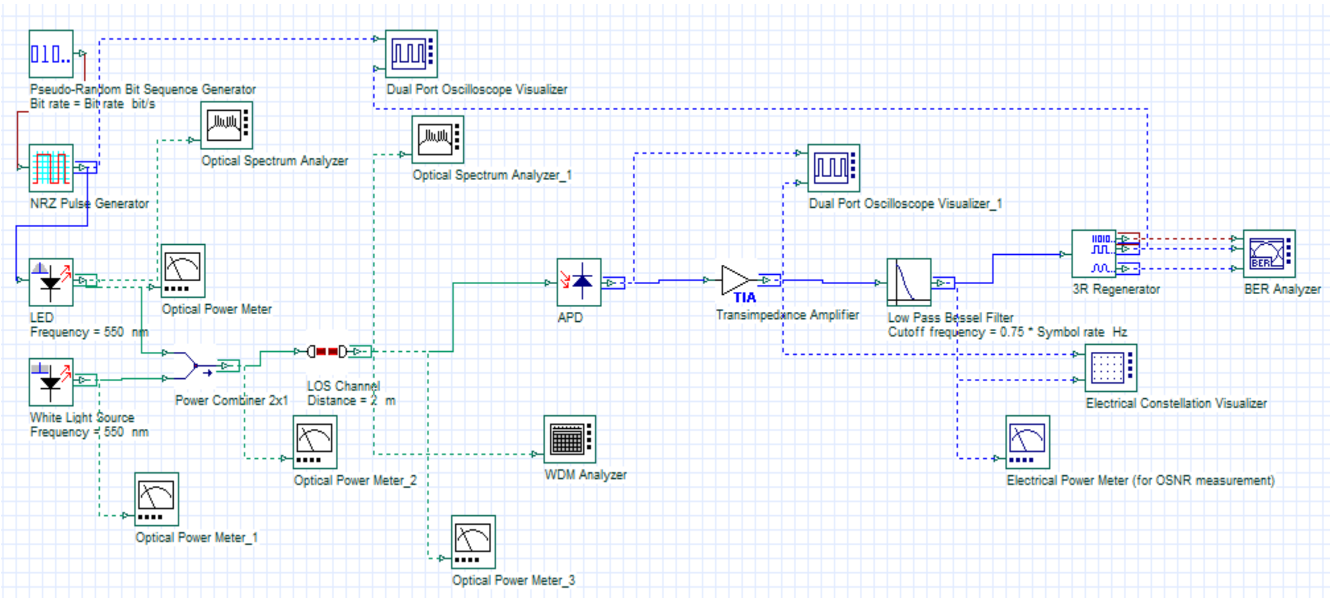


Figure 3: Simulation for the Los channel model with ambient light using optisystem software

material that yields electron-hole pairs under illumination conditions. The positive P-type connection goes to the anode section while the negative N-type part links to the cathode section. The bandgap energy of semiconductor decides whether photons entering the photodiode produce electron-hole pairs or not which leads to PIN photodiodes serving as key converters of light signals into electric signals across different modern applications. 2. The Transimpedance Amplifier (TIA) operates as a device that converts sensor-generated current into voltage outputs when working with current signaling photodiodes. The performance stability of multiple other operational amplifier circuits becomes clearer

C. Receiver Side

A photo detector receives communication signals at the receiver end of the system. These signals need to undergo several conditioning processes including amplification together with filtering as well as frequency conversion to boost quality standards while suppressing noise disturbance.

using this fundamental building block. 3. A low pass Bessel filter removes noisy high-frequency outputs which tend to affect amplified electrical signals. The signal-to-noise ratio receives an improvement due to the filtering process that removes unwanted noise. The signal filter system follows Bessel patterning to maintain signal quality while eliminating unwanted signals. 4. The 3R Regenerator completes three operations for bit stream recovery: re-amplification followed by reshaping and retiming. The combination which accomplishes re-amplification reshaping

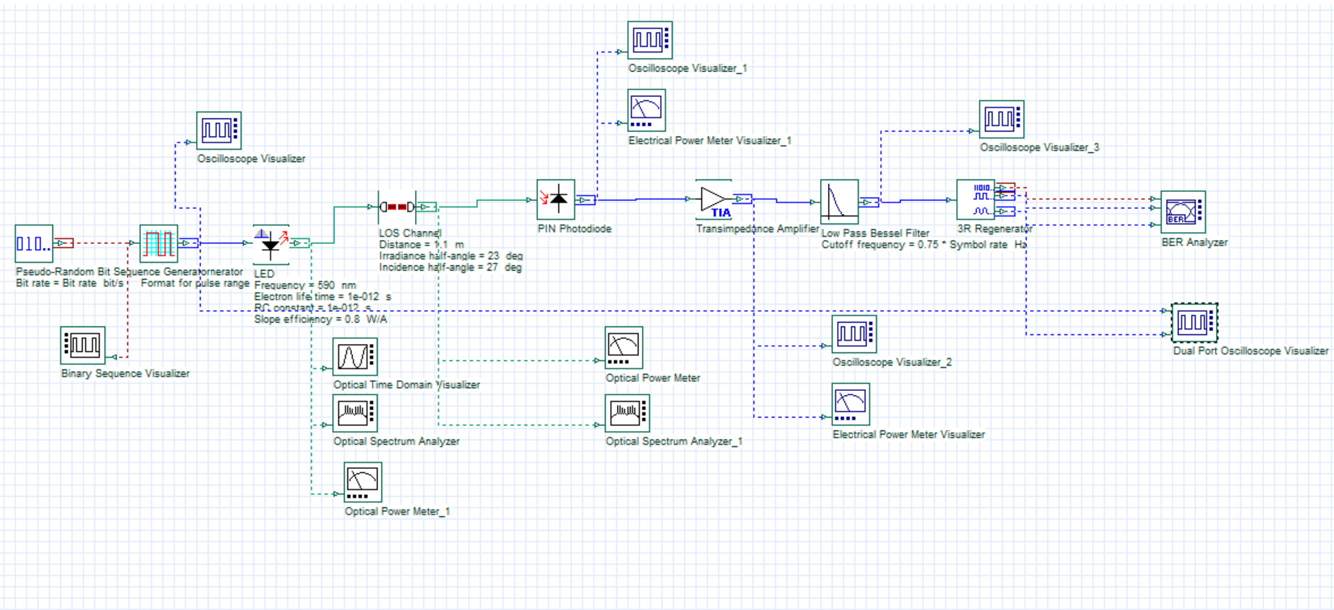


Figure 4: Simulation for the Los channel model without ambient light using optisystem software

and retiming functions goes by the name 3R regenerator as shown in Fig 3 and 4 with and without the influence of ambient light.

The measurement of optical signal quality depends on OSNR which stands for Optical Signal-to-Noise Ratio through the comparison between signal power and noise power within the signal spectrum. Good signal quality depends on higher OSNR since it decreases interference from noise which allows reliable optical communication. The decibel chart (dB) determines the measurement in equation 1.

$$OSNR = 10 log10 PsignalPnoise \qquad (1)$$

The form presents P signal as the optical signal power and P noise as the noise power within the signal band. The Signal Power is 27.95 W and Noise power is 230.83 ×10-6 so the OSNR is 50.83 Db.

D. Output Visualizer Components:

The research utilizes multiple visualization devices consisting of oscilloscope visualizers, binary sequence visualizers, optical power meters, optical spectrum analyzers, optical time domain visualizers and electrical power meters and BER analyzers. Visual assessment of electrical signals occurs through oscilloscope displays while binary sequence displays show binary code sequences. The combination of optical power meters and optical spectrum analyzers and optical time domain visualizers provides data for measuring power levels and identifying signal components as well as waveform capture in time domain measurement. Digital communication system performance evaluation as well as energy measurements result from electrical power meters' functions. The analysis for Line-of-Sight uses Table I to present complete specifications of its components.

Table 1
The Components Parameters Of VLC Based Li-Fi Loss Channel During Ambient Noise Manifestations

Component	***Value***
Bit rate	Bit rate = 10Gbp
Modulation Technique	(NRZ) OOK modulation
CW Laser source 1	Frequency = 550 nm Power = 1-10 dBm measure at 4dbm
White Light(Ambient Noise	Frequency = 50 THz Avg. Power = 4.6 mW
NLOS underwater Channel	Range = 10m underwater Frequency = 545 nm
Pin Photodiode	Type of Responsivity = Silicon Responsivity = 1 A/W Dark Current = 10nA Short Noise Distribution = Gaussian
Bessel Low Pass Filter	Cutoff frequency = 0.75 × symbol rate
Trans-impedance Amplifier (TIA)	Open loop Voltage Gain = 90 dB

III. Results and Discussion

Measurement of bit error rate relies on Gaussian algorithm processing between BER analyzer and short bit transmissions. The BER (Bit Error Rate) Analyzer stands as a crucial assessment instrument for digital communication system performance evaluation in Opti-System. The main role of this tool relies on checking received and transmitted bits to detect errors thus enabling to study system performance dynamics across various situations.

The BER Analyzer enables system design improvements for higher reliability and efficiency through analysis of

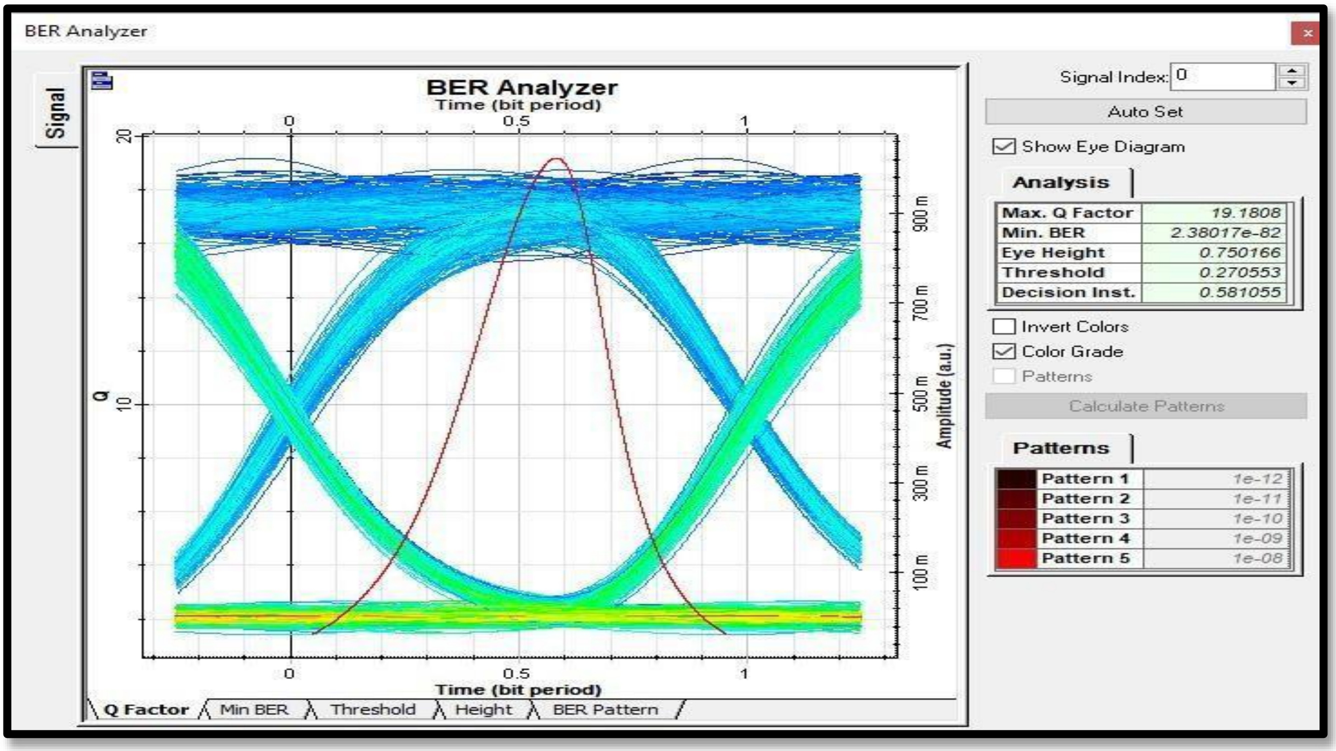


Figure 5:The BER Analyzer operated on VLC based LI-FI LOS channels with the influence of ambient light.

modulation systems with channel effects and noise measurements as well as error correction strategies. Performances metrics obtained from the BER analyzer enable

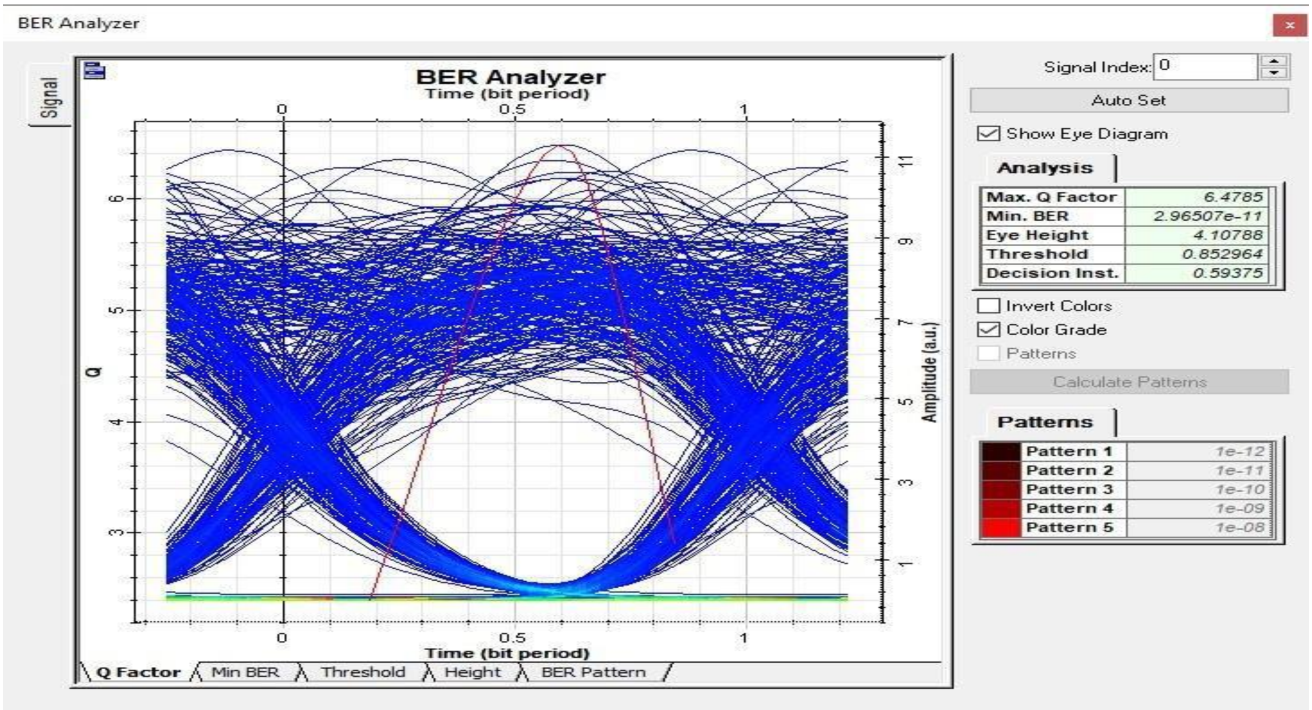


Figure 6:The BER Analyzer operated on VLC based LI-FI LOS channels without the influence of ambient light.

strategic decisions when designing and testing systems. The quality assessment of optical signals within Opti-System relies on Q-factor measurements for determining both signal quality along with system performance evaluation in optical communication systems. The Q-factor serves as an assessment tool for Bit Error Rate (BER) performance evaluation of systems including fiber-optic communications and coherent optical systems and serves us in analyzing our designed LOS channel. The obtained quality factor meets the required threshold of 6 as it reaches 6.72 with the influence of ambient light and 19.18 up to 10 meter distance. Figure 5 and 6 show the with and

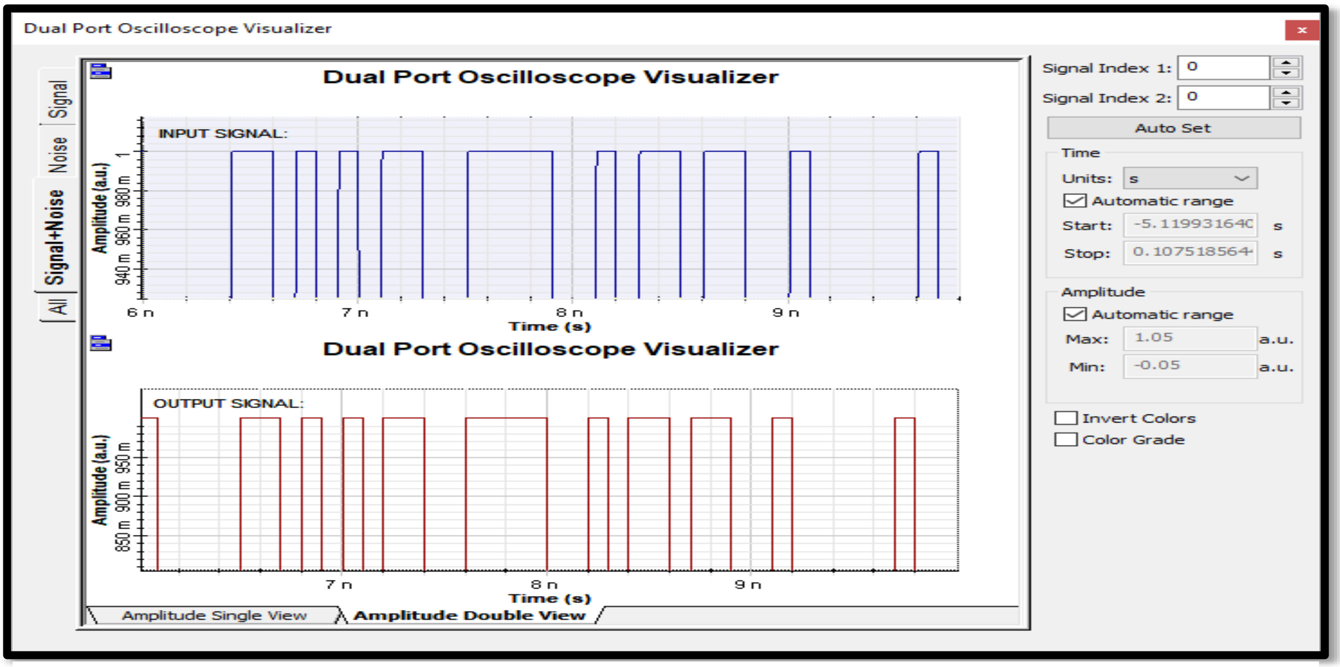


Figure 7 : The results of VLC based LI-FI LOS channel simulation with ambient light show input and output signals through Oscilloscope visualizer. Comparison of Transmitted (blue color) and out received (red color) waveforms.

without the influence of the ambient light of the BER analyzer.

The evaluation of system performance requires users to include input and output signals as results within OptiSystem

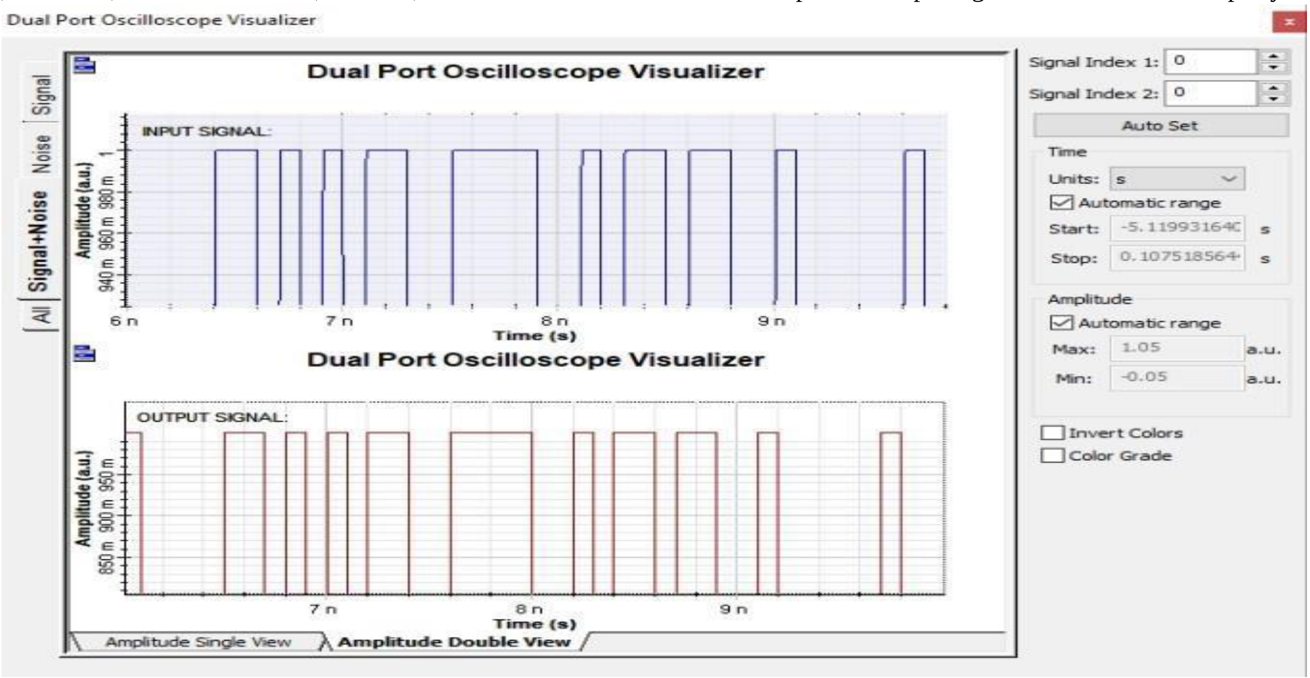


Figure 8 : The results of VLC based LI-FI LOS channel simulation without ambient light show input and output signals through Oscilloscope visualizer. Comparison of Transmitted (blue color) and out received (red color) waveforms.

software. The input signal brings the original transmitted data to view which enables users to access information about power level as well as wavelength and modulation format parameters. Analysis of system efficiency and losses together with distortions is possible when examining the output received data signal which travels through different system elements. Users perform performance metric assessments through signal comparison which includes signal-to-noise ratio (SNR), bit error rate (BER) and power variations assessment. The process of system optimization along with enhanced signal quality results from this signal comparison procedure.

IV. Conclusion

The paper examines the effects of ambient noise on VLC Li-Fi channel performance through visible light communication (VLC) evaluation. The research analyzed systems through performance testing of OSNR and Q-factor along with bit error rate (BER). The research findings demonstrate an advancement in indoor communication systems since VLC technology delivers 10Gb/s transmission over 10m distances at 50.52 db OSNR. The research develops a dependable VLC Li-Fi system which maintains 10Gb/s data rates during a 10 m distance while operating with OSNR equal to 50.52 db. Los Channels perform Under ambient conditions the experimental system reached better results indicating 2.38e-82 bit error rate and 19.18 Q factor however the results under ambient light showed 5.5e-12 bit error rate and 6.72 Q factor. The rectangular optical filter serves in the research to reduce the effects of uncontrollable ambient light on the signal system. This research establishes new opportunities for wireless communication technology development which will foster progress within the wireless communication domain.

References

[1]. Sarkar, Anurag & Agarwal, Shalabh & Nath, Asoke. (2015). Li-Fi Technology: Data Transmission through Visible Light. IJARCSMS. 3. 1-10.

[2]. Nivetha, S., Preethi, G. A., & Chandrasekar, C. (2019). Performance evaluation of modulation techniques in Li-Fi. Int. J. Recent Technol. Eng, (3), 1509-1518.

[3]. Poulose, A. (2022). Simulation of an Indoor Visible Light Communication System Using Optisystem. Signals, 3(4), 765-793.

[4]. Shamsudheen, P., Sureshkumar, E., & Chunkath, J. (2016). Performance analysis of visible light communication system for free space optical communication link. Procedia technology, 24, 827-833.

[5]. Vanderka, A., Hajek, L., Latal, J., Vitasek, J., & Koudelka, P. (2014). Design, simulation and testing of the OOK NRZ modulation format for free space optic communication in a simulation box. Advances in electrical and electronic engineering, 12(6), 604-616.

[6]. Pradhan, J., Kappala, V. K., & Das, S. K. (2020, February). Performance analysis of a Li-Fi system under ambient light conditions. In 2020 National Conference on Communications (NCC) (pp. 1-6). IEEE.

[7]. Islim, M. S., & Haas, H. (2019). Modulation techniques for li-fi. ZTE communications, 14(2), 29-40.

[8]. Rahman, M. T., Bakaul, M., & Parthiban, R. (2018, November). Analysis of the effects of multiple reflection paths on high speed VLC system performance. In 2018 28th International Telecommunication Networks and Applications Conference (ITNAC) (pp. 1-6). IEEE.

[9]. Omar Faruq, Kazi Rubaiyat Shahriar Rahman, Nusrat Jahan2, Sakib Rokoni, Mosa Rabey. Li-Fi Technology Based Long Range Free-Space Optics Data Transmit System Evaluation. EAI Endorsed Transactions on Mobile Communications and Applications 01, 06, 2023, Volume 7.

[10]. ShabanaParveen M, Siddarthan K, Vignesh T, Ajay Krishnaa A R. Data transmission using Li-Fi technology. Journal of Emerging Technologies and Innovative Research (JETIR). 01 Jun 2019- Vol. 6, pp 22-24.

[11]. Li, Y., Huang, X., & Huang, H. (2015). A survey on visible light communication systems. Tsinghua Science and Technology, 20(3), 236-247.

[12]. Haas, H., & O'Brien, D. (2003). A new indoor communications concept based on visible light. IEEE Personal Communications, 8(2), 40-48.

[13]. Tsonev, D., Videv, S., Haas, H., & Guan, L. (2013). A 3-Gb/s single-LED OFDM-based wireless VLC link using a gallium nitride μLED. IEEE Photonics Technology Letters, 25(2), 171-174.

[14]. Wang, Y., Wang, Y., & Chi, N. (2018). Visible light communication for indoor positioning. IEEE Communications Magazine, 56(9), 218-224.

[15]. Haas, H. (2016). Li-Fi: VLC technology to provide wireless connectivity. Li-Fi Internet, 1(1), 3-19.

[16]. Miramirkhani, F., & Kavehrad, M. (2018). Cognitive radio for Li-Fi: Opportunities and challenges. IEEE Communications Magazine, 56(4), 144-149.

[17]. Khalid, M., & Cossu, G. (2016). Ambient noise effects on OFDM-based optical wireless communication systems. IEEE Transactions on Wireless Communications, 15(10), 6604- 6617.

[18]. Yu, C., & Zhang, J. (2019). A novel OOK modulation method for visible light communication. Optik, 184, 300-305.

[19]. Dwivedi, A., & Kumar, A. (2020). Performance analysis of line of sight (LOS) communication in FSO. In Proceedings of the International Conference on Innovative Data Communication Technologies and Application (ICIDCA) (pp. 1-5).

[20]. Yang, X., & Li, L. (2018). Impact of ambient light noise on free-space optical communication. IEEE Access, 6, 31851-31858.

[21]. Z. A. Khan et al., “Li-Fi Based Healthcare Monitoring System,” Pakistan Journal of Engineering and Technology, vol. 5, no. 2, pp. 177–182, Sep. 2022, 10.51846/vol5iss2pp177-182.

[22]. M. R. H. Mondal and R. B. Faruque, “Hybrid diversity combined OFDM for LiFi,” in 2017 IEEE International Conference on Telecommunications and Photonics (ICTP), IEEE, Dec. 2017, pp. 132–136. doi: 10.1109/ICTP.2017.8285945.

[23]. L. W. Y. Z. P. S. M. Saadi1*, “Visible Light Communication: Opportunities, Challenges and Channel Models,” International Journal of Electronics & Informatics, vol. 2, no. 1, 2013.

[24]. J. Shi et al., “AI-Enabled Intelligent Visible Light Communications: Challenges, Progress, and Future,” Photonics, vol. 9, no. 8, p. 529, Jul. 2022, doi: 10.3390/photonics9080529.

[25]. M. T. Rahman, A. S. M. Bakibillah, R. Parthiban, and M. Bakaul, “Review of advanced techniques for multi-gigabit visible light communication,” IET Optoelectronics, vol. 14, no. 6, pp. 359–373, Dec. 2020, doi: 10.1049/iet-opt.2019.0120.

[26]. G. S. Christy, G. Sundari, and V. J. K. Kishor Sonti, “A Review on Evolution, Challenges and Scope in Visual Light Communication Systems,” in 2022 IEEE 7th International conference for Convergence in Technology (I2CT), IEEE, Apr. 2022, pp. 1–5. doi: 10.1109/I2CT54291.2022.9824695.

HLCE: Framework for Enhanced Stock Price Forecasting

Yaser Ali Shah
Department of Computer Science
COMSATS University Islamabad, Attock Campus
Attock, Pakistan
orcid.org/0000-0003-1628-1743

Nimra Waqar
Department of Computer Science
COMSATS University Islamabad, Attock Campus
Attock, Pakistan
orcid.org/0009-0008-1252-0106

Wasiat Khan
Department of Software Engineering, University of Engineering & Technology, Bannu, Khyber Pakhtunkhwa Pakistan
orcid.org/0000-0003-0394-5432

Amaad Khalil
Department of Computer Systems Engineering
University of Engineering & Technology, Peshawar, Pakistan
orcid.org/0000-0002-1937-8906

***Abstract*— A key element of risk management and investing decision-making is accurate stock price forecasting. For financial forecasting, traditional time series models such as Exponential Smoothing (ETS), Generalized Autoregressive Conditional Heteroscedasticity (GARCH), and Autoregressive Integrated Moving Average (ARIMA) have been used extensively. However, their forecast accuracy is limited by their incapacity to accurately model the complex, non-linear connections seen in financial data. Using the advantages of both deep learning and conventional forecasting techniques, we provide a Hybrid LSTM-Conventional Ensemble (HLCE) model in this work. To improve the accuracy of stock price predictions, the HLCE framework combines Long Short-Term Memory (LSTM) networks with traditional statistical models and machine learning methods including Random Forest, XGBoost, and Support Vector Regression (SVR). The model is assessed using important performance metrics, such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2), in a case study using Apple Inc. (AAPL) stock data, where MinMaxScaler is utilized for data preprocessing. With an RMSE of 0.16, MAE of 0.16, MAPE of 0.12%, and R2 of 0.95, the HLCE model performs better than individual models, according to experimental results, proving its greater capacity to identify intricate financial patterns. By contrast, isolated models have far lower predictive efficiency and much higher error rates. These results highlight the promise of ensemble and hybrid approaches in financial forecasting, offering a more reliable and accurate framework for predicting stock prices. The work adds to the expanding corpus of research supporting the combination of deep learning and conventional techniques to enhance risk assessment and financial market analysis.**

Keywords: Stock Price Prediction; Hybrid LSTM-Conventional Ensemble (HLCE); Time Series Forecasting; Financial Forecasting; Forecasting Accuracy.

I. Introduction

The intricacy and volatility of financial markets, which are impacted by a number of variables like investor behavior, political events, and economic statistics, make it difficult to predict stock values. Although they are good at capturing linear trends and volatility, traditional time series models such as GARCH, ETS, and ARIMA have trouble detecting non-linear market dynamics. Through the modeling of non-linear interactions in financial data, recent developments in deep learning, in particular Long Short-Term Memory (LSTM) networks, have demonstrated promise in overcoming these constraints. LSTMs might not be able to adequately represent the intricacies of volatile markets, though. This is addressed by the Hybrid LSTM-Conventional Ensemble (HLCE) model, which provides a more accurate and balanced forecast by combining LSTM with conventional time series models including ARIMA, ETS, and GARCH. By combining the forecasts from each model using an ideal weighting method, the HLCE model capitalizes on each model's unique capabilities to increase forecasting accuracy overall. The findings show that HLCE performs better than standalone models in stock price prediction by identifying complex patterns and producing more accurate forecasts. The study offers a number of significant insights. It presents the HLCE model, which improves stock price forecasting by combining LSTM with traditional models. To increase predicting accuracy, the model combines forecasts using an ideal weighting method. In terms of stock price prediction, the HLCE model outperforms individual models and reaches state-of-the-art performance.

The novelty and contributions of our research are as follows:

- The HLCE model, which combines LSTM with ARIMA, ETS, and GARCH for better forecasting, was introduced.
- To capitalize on each model's unique characteristics, the HLCE model uses an ideal weighting scheme.
- The HLCE model outperforms standalone models in stock price prediction, achieving state-of-the-art results.
- The model exhibits its capacity to identify intricate market trends, yielding predictions that are more precise and trustworthy.

With sections on "Related Work," "Methodology," "Experiments & Results," and "Conclusion & Future Work," the article provides a thorough examination of the efficacy of the suggested model as well as potential avenues for further research.

II. Related Work

Research on stock price forecasting has garnered a lot of attention, with a focus on increasing prediction accuracy by fusing contemporary machine learning techniques with conventional time series models. As Table I shows, traditional models such as ARIMA, Exponential Smoothing (ETS), and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) have been widely used to simulate financial data trends [1, 2, 3]. ETS excels at exponentially smoothing trends and seasonality, while GARCH is very strong at simulating volatility and heteroscedasticity, especially in volatile market environments. ARIMA is very good at detecting trends and seasonality in time series data. However, it can occasionally be challenging for these conventional approaches to capture the intricate and non-linear patterns found in financial markets [4, 5]. Non-linearities in financial data are sometimes caused by a number of variables, such as investor mood and macroeconomic indicators [6, 7]. To get around these limitations, researchers have been focusing more on machine learning models, particularly Long Short-Term Memory (LSTM) networks [8, 9]. Because LSTMs are designed to identify long-term relationships in sequential data, they are especially well-suited for time series forecasting. By utilizing extensive historical data, they have proven to be more accurate in turbulent markets and outperform conventional methods in anticipating non-linear trends [10, 11]. Because of their strong predictive powers, ensemble methods like Support Vector Regression (SVR) [12,13], Random Forest [14,15], and XGBoost [16,17] have also attracted a lot of interest. Prediction accuracy has been significantly improved by hybrid models, such as ARIMA-LSTM [18,19], GARCH-LSTM [20,21], and ETS-LSTM [22], which integrate LSTMs with traditional models. These models successfully combine the linear pattern recognition powers of conventional models with the advantages of machine learning in capturing non-linear correlations. When a number of dynamic factors, such as market volatility, affect stock prices, such hybrid techniques are especially advantageous [23, 24]. In order to enable models to concentrate on the most pertinent time steps in the data, recent developments have added attention mechanisms to LSTM networks. The model's capacity to identify significant patterns and connections as a result of this selective attention significantly improves forecasting accuracy [25, 26]. Furthermore, predicting performance has been enhanced by feature engineering approaches such adding moving averages, lagged features, and external data, such as macroeconomic indicators and news sentiment [27, 28]. A thorough grasp of market dynamics has been made possible by the incorporation of multi-source data, such as past prices, news mood, and social media activity, which has improved model robustness [29, 30].Notwithstanding these developments, there is still a significant lack of systematic comparisons between LSTM-based models and traditional methods that employ consistent evaluation standards in the literature. More thorough evaluation frameworks even though measures like RMSE and MAE have been used in a number of studies [31, 32]. By filling this gap, researchers will be able to determine which models are most suited for particular forecasting tasks, ultimately leading to better stock price prediction systems [33].Additionally, recent research has looked at incorporating state-of-the-art techniques like transfer learning [38, 39], reinforcement learning [36, 37], and Generative Adversarial Networks (GANs) [34, 35] into stock price forecasting models. Transfer learning uses data from related areas to improve model performance, whereas reinforcement learning optimizes trading tactics based on reward signals. To enhance training datasets and enhance model generalization, GANs can produce synthetic data. The field of stock price prediction might develop further with these novel methodologies [40, 41]. Researchers have also started looking at how external influences, such world events, political shifts, and natural disasters, affect the behavior of the stock market [42, 43]. Researchers want to better understand the intricate connections between exogenous events and market behavior by including these factors into forecasting models, which might improve the accuracy and robustness of stock price projections [44, 45]. Additionally, researchers have begun to examine the ways in which external influences, such as global events, political changes, and natural catastrophes, impact stock market dynamics [47, 48]. Researchers are working to include the intricate relationship between external events and market dynamics into forecasting models in order to increase the precision and resilience of stock price estimates [49, 50].

III. Methodology

This study investigates advanced methods for stock price prediction by fusing traditional time series models like ARIMA, Exponential Smoothing (ETS), and GARCH with a range of machine learning models, including Random Forest, XGBoost, Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) networks. The primary objective is to develop a Hybrid Learning and Classification Ensemble (HLCE) model that combines the benefits of these diverse approaches to increase prediction accuracy and resilience. The study starts with a thorough data gathering process, obtaining external market indicators and stock price data from a variety of financial databases, APIs, and public datasets. In order to identify trends, seasonality, and residuals, this data is subjected to extensive preparation, which includes managing missing values, outlier identification, normalization, and time-series decomposition. After that, each model is trained separately, utilizing its unique advantages, such as GARCH's ability to capture volatility clustering, LSTM's ability to predict long-term temporal correlations in sequential data, and machine learning models' capacity to recognize intricate, nonlinear patterns. Combining predictions from many models is made possible by the use of ensemble techniques like weighted averaging and stacking in the construction of the HLCE model. By reducing the drawbacks of individual models, this hybrid method produces a prediction framework that is more precise and broadly applicable. Each model's performance is evaluated together with the HLCE using metrics such as R-squared (R^2) for goodness-of-fit, Mean Absolute Error (MAE) for average prediction deviation, and Root Mean Squared Error (RMSE) for overall error magnitude. The findings show that by utilizing complementary strengths and offsetting individual shortcomings, the HLCE model performs noticeably better than independent models. This study demonstrates how hybrid models, which combine traditional and machine learning approaches, may produce accurate, dependable, and robust financial market forecasts, opening the door to better stock trading and investment strategy decision-making.

A. *Proposed Methodology*

By using a large dataset mostly from YFinance and augmented by other financial and economic sources, the Hybrid LSTM-Conventional Ensemble (HLCE) approach aims to improve the precision of stock price forecasts. In addition to daily transaction volumes, which offer information on market activity and liquidity, this dataset includes historical stock values, including open, high, low, close, and adjusted closing prices. Important macroeconomic indicators from reliable economic databases like FRED or the OECD, including interest rates, inflation rates, and GDP growth statistics, will be added to the dataset to further enhance it. Technical indicators like moving averages (both simple and exponential), the Relative Strength Index (RSI), and Moving Average Convergence Divergence (MACD) will also be utilized to assist identify trends and potential price reversal points. Forecasts will get additional context by using sentiment analysis data gathered from social media platforms and financial news articles. Additionally, company-specific data that might directly affect stock prices, such earnings reports, dividend announcements, and stock splits, will be included in the dataset. Since they have a big impact on market circumstances, external issues like geopolitical developments and regulatory changes will also be taken into account. Feature engineering will be essential to improving the models' capacity for prediction. New features like lagged variables for past prices and volumes, interaction terms to capture the links between various technical indicators, and categorical variables that reflect distinct market situations will also be created throughout this process. The goal of the HLCE approach is to efficiently capture both linear and non-linear trends in financial data by integrating this rich and varied dataset. By using ensemble techniques like bagging, boosting, and stacking, this method will make it easier to create strong predictive models, which will ultimately increase the accuracy of stock market predictions. The HLCE model combines the benefits of LSTM networks which are well-known for their ability to detect long-term correlations in sequential data—with more conventional models like as ARIMA, ETS, GARCH, Random Forest, XGBoost, and SVR. This hybrid approach enhances overall forecasting ability and permits a more in-depth analysis of stock price changes by using the unique advantages of each model.

TABLE I. OVERVIEW OF TRADITIONAL, ML, AND DL METHODOLOGIES USED IN STOCK PRICE FORECASTING

Model	References	Description
ARIMA	[1,3], [20,21], [25,26]	Linear time series forecasting for trends and seasonality.
ETS	[2], [18,19]	Exponential smoothing for trends and seasonality.
GARCH	[4], [16,17], [22-24], [43], [49]	Models volatility and heteroscedasticity in financial data.
SVM	[15,16], [35-37]	Classifies trends using optimal decision boundaries.
RF	[14], [17-19]	Ensemble decision trees for improved accuracy.
ANN	[5,6], [30,31]	To capture intricate non-linear patterns, use NN.
LSTM	[9-13], [25-27], [33],[39,40]	Long-term dependencies in sequential data, such as stock prices, are captured.
Attention Mechanism	[28,29], [35,36]	Focuses LSTM on most relevant input sequences, improving accuracy.
GARCH-LSTM	[22-24], [43,49,50]	Combines LSTM for patterns and GARCH for predictions.

B. *Workflow of proposed HLCE*

Fig. 1 describes the suggested Hybrid LSTM-Conventional Ensemble (HLCE) approach for stock price prediction. The process starts with a machine learning pipeline that combines time series models and ensemble methods. The dataset is gathered, cleaned, and features are developed during the Data Preparation phase to make sure it is ready for training, improving its quality and applicability. Seven distinct models—ARIMA, ETS, GARCH, Random Forest (RF), XGBoost, Support Vector Regression (SVR), and Long Short-Term Memory (LSTM)—are trained using the same set of parameters. Each model captures a different part of the data according to its own strengths. Through a weight adjustment process, the projections from various models are combined into a final ensemble forecast during the Ensemble Forecasting phase, which maximizes each model's contribution to guarantee a balanced and precise prediction. Using a variety of accuracy criteria, the ensemble model's performance is evaluated in the Model Evaluation phase by contrasting its predictions with real stock prices. This gives important information about how successful the model is. With visual upgrades that make it easier to understand and clearly depict the flow of data and procedures from preparation to the final prediction and review, the diagram is made with clarity in mind. This methodology guarantees that every stage is conveyed clearly, showcasing the incorporation of several modeling approaches to improve the precision of stock price forecasts.

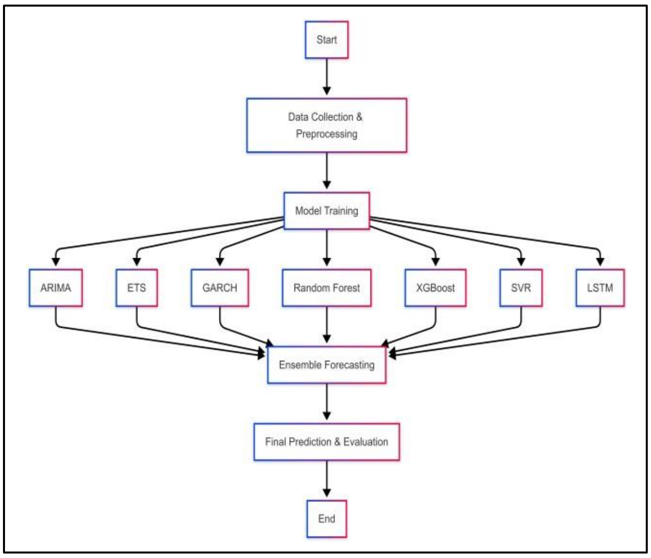


Fig 1: Work Flow Pipeline of HLCE

C. *Model Training*

Each forecasting model—ARIMA, ETS, GARCH, LSTM, Random Forest, XGBoost, and SVR was first trained individually to capture different aspects of the time series data. Their predictions were then combined using a weighted average in the Hybrid LSTM-Conventional Ensemble (HLCE) model. Weights were assigned based on performance metrics like RMSE and MAE to enhance overall forecast accuracy.

1) ARIMA Model

The ARIMA model is defined and fitted using the training dataset. In this equation, Y_t represents the value of the dependent variable that is being forecasted at time t. The constant term c accounts for the bias or offset of the model. The autoregressive parameters ϕ quantify the relationship between the current value and the previous values of the time series, thereby capturing the impact of past observations on the present value. The influence of prior shocks on the series is modeled with the aid of the moving average parameters θ, which reflect the relationship between the present value and historical forecast errors. Lastly, to account for the unexpected elements or noise in the time series that the model cannot explain, the random error term ϵ is assumed to be independent, identically distributed, and to have a mean of zero. The mathematical representation is given by Eq. 1.

$$\boldsymbol{Y}_t = \boldsymbol{c} + \boldsymbol{\phi}_1\boldsymbol{Y}_{t-1} + \boldsymbol{\phi}_2\boldsymbol{Y}_{t-2} + \ldots + \boldsymbol{\theta}_1\boldsymbol{\varepsilon}_{t-1} + \boldsymbol{\theta}_2\boldsymbol{\varepsilon}_{t-2} + \ldots + \boldsymbol{\varepsilon}_t \quad \textbf{Eq. (1)}$$

2) Exponential Smoothing (ETS) Model

The smoothed value at time t, or the predicted value based on prior observations, is represented by S t in this equation. The ground truth for comparing with the model's prediction is Yt, which is the actual observed value at time t. The weight assigned to the most recent observation is determined by the smoothing constant α, which ranges from 0 to 1. Higher α α increases the model's sensitivity to data changes, enabling it to respond swiftly to emerging trends. By assigning more weight to past data, a lower α, on the other hand, produces smoother forecasts that may be less susceptible to short-term fluctuations. The Exponential Smoothing (ETS) model is defined by Eq. 2.

$$\boldsymbol{St = \alpha Yt + (1-\alpha)St - 1} \quad \textbf{Eq. (2)}$$

3) Garch Model

The conditional variance at time t, which indicates the volatility of the asset, is represented by the symbol σt2 in this equation. The residual, or ϵ, is the difference between the mean equation's actual and expected values. The model is controlled by three parameters: α0, α1, and β1. α1 reflects the impact of previous residuals on current volatility, β1 represents the impact of prior variances on current volatility, and α0 is a constant term that denotes the model's baseline volatility level. The GARCH model is defined using the scaled data returns as we can see in Eq. 3.

$$\boldsymbol{\sigma^2_t = \alpha_0 + \alpha_1 \epsilon^2_{\{t-1\}} + \beta_1 \sigma^2_{\{t-1\}}} \quad \textbf{Eq. (3)}$$

4) LSTM Model

The hidden and cell states at time tt are represented by ht and Ct in these equations, whereas Ct−1 and ht−1 are from the preceding time step. The input is xt, and the forget, input, and output gate activations are ft, it, and ot. The potential cell state is C~t. The weight matrices and bias vectors are W and b, whereas the activation functions are σ and tanht. An LSTM cell can be represented using Eq. 4.

$$\boldsymbol{ht = ot \odot tanh(Ct)} \quad \textbf{Eq. (4)}$$

5) Random Forest

The Random Forest model is an ensemble learning approach that generates the class mode (classification) or mean prediction (regression) of each individual decision tree during training. The model is trained on a portion of the data and bootstraps to produce a large number of samples. A random selection of characteristics is used to train each tree, which lessens overfitting and enhances generalization. The Random Forest model improves accuracy and resilience by combining the forecasts of every single tree. Furthermore, in comparison to a single decision tree, the model efficiently lowers the variance by averaging the outcomes or using a majority vote, which lessens the likelihood of overfitting. The mathematical representation of Eq. 5 is given below.

$$\hat{y} = \left(\frac{1}{N}\right)\sum[\boldsymbol{i = 1\ to\ N}]\boldsymbol{Ti(x)} \quad \textbf{Eq. (5)}$$

6) XGBoost Model

An effective and scalable use of the gradient boosting framework is the XGBoost (Extreme Gradient Boosting) model. It constructs trees in a sequential fashion, fixing the mistakes of earlier trained trees with each new tree. Gradient descent is used to minimize the loss function used to train the model. XGBoost can deal with missing values internally and has regularization to avoid overfitting.

7) Support Vector Regression (SVR) Model

The Support Vector Regression (SVR) model is based on the Support Vector Machine (SVM) approach. It looks for a function that deviates from the actual observed values by a value no more than a specified margin (epsilon). By employing kernel functions to transform the input space into a higher-dimensional space, SVR is able to identify complex correlations in the data.

TABLE II. PSEUDO CODE OF HLCE

```
def hlce_model(arima_preds, ets_preds, garch_preds, lstm_preds,
rf_preds, xgb_preds, svr_preds, weights):
  hlce_preds = (weights[0] * arima_preds + weights[1]
* ets_preds + weights[2] * garch_preds +weights[3] * lstm_preds
+ weights[4] * rf_preds + weights[5] * xgb_preds + weights[6] *
svr_preds)
  return hlce_preds
arima_preds = arima_model.forecast(test_data)
ets_preds = ets_model.forecast(test_data)
garch_preds = garch_model.forecast(test_data)
lstm_preds = lstm_model.predict(test_X)
rf_preds = rf_model.predict(test_data)
xgb_preds = xgb_model.predict(test_data)
svr_preds = svr_model.predict(test_data)
weights = [0.1429, 0.1429, 0.1429, 0.1429, 0.1429, 0.1429, 0.1429]
ensemble_predictions = ensemble_model(arima_preds, ets_preds,
garch_preds, lstm_preds, rf_preds, xgb_preds, svr_preds, weights)
```

8) Ensemble Model (HLCE)

Seven models' predictions are combined in the Hybrid LSTM-Conventional Ensemble (HLCE) model, as shown in Table II. To increase forecasting accuracy, a weighted average approach is used to aggregate the forecasts from each model. Each model's weight is determined by its performance indicators, such as RMSE, MAE, and so on. The forecast from the ensemble model can be expressed mathematically as represented in Eq. 6.

$$\boldsymbol{Y_{hlce} = 0.1429 \cdot \sum(Y_{Tarima}, Y_{Tets}, Y_{Tgarch}, Y_{Tlstm}, Y_{Trf}, Y_{Txgb}, Y_{tSVR})} \quad \textbf{Eq. (6)}$$

Where, Y_{tARIMA}, Y_{tETS}, Y_{tGARCH}, Y_{tLSTM}, Y_{tRF}, Y_{tXGB}, Y_{tSVR} are the forecast models, and w1 to w7 are the equal weights (0.1429) assigned to each model. These can be adjusted based on model performance.

D. Model Evaluation Matrices

The performance of forecasting models is assessed using various evaluation metrics summarized in Table III.

Root Mean Square Error (RMSE) determines the average magnitude of errors without considering their direction and is sensitive to outliers due to squaring the differences. It is calculated in Eq. 7.

$$RMSE = \sqrt{\left(\frac{1}{n}\sum(Yi - \hat{Y}i)^2\right)} \qquad \textbf{Eq. (7)}$$

Mean Absolute Error (MAE) measures the average absolute difference between actual and predicted values, penalizing errors linearly and less harshly than RMSE as Eq. 8.

$$\mathbf{MAE = (1/n) \sum|Yi - \hat{Y}i|} \qquad \textbf{Eq. (8)}$$

Mean Absolute Percentage Error (MAPE) provides the average absolute percentage difference between actual and predicted values, indicating accuracy as a %, but can be skewed when actual values are near 0 as represented in Eq. 9.

$$\mathbf{MAPE = (1/n) \sum| (Yi - \hat{Y}i) / Yi |} \qquad \textbf{Eq. (9)}$$

R-squared (R^2) reflects how well the independent variables explain the variance in the dependent variable; values closer to 1 indicate a better fit as we can see in Eq. 10.

$$\mathbf{R^2 = 1 - (Sres / Stot)} \qquad \textbf{Eq. (10)}$$

The section also repeats MAE and MAPE for emphasis or possible inclusion from different sources. The repeated formula for MAE is shown in Eq. 11.

$$\boldsymbol{MAE = (1/n) \sum|Yi - \hat{Y}i|} \qquad \textbf{Eq. (11)}$$

Lastly, R^2 is reiterated to emphasize its importance in regression analysis as mathematically represented in Eq. 12.

$$\mathbf{R^2 = 1 - (Sres / Stot)} \qquad \textbf{Eq. (12)}$$

TABLE III. PERFORMANCE EVALUATION OF MODELS

Model	RMSE	MAE	MAPE	R²
ARIMA	0.90	0.90	0.92	0.17
EST	1.00	1.00	1.00	0.00
GARCH	0.00	0.00	0.00	1.00
LSTM	0.29	0.19	0.19	0.88
RF	0.90	0.90	0.91	0.18
XGB	0.87	0.87	0.88	0.23
SVR	0.58	0.58	0.61	0.62
HLCE	0.16	0.16	0.12	0.95

IV. EXPERIMENTS AND RESULTS

This section compares LSTM with models like GARCH, ETS, and ARIMA, showing that the proposed HLCE ensemble outperforms individual models based on performance metrics and visual analysis.

A. Experimental Setup

This study evaluates the HLCE model using YFinance stock data (2010–2023), including closing prices and volumes. The data is split into training, validation, and test sets, with preprocessing like normalization and outlier handling. Model performance is measured using R^2, MAE, RMSE, and MAPE. Hyperparameters are tuned, cross-validation is applied to avoid overfitting, and the best HLCE model is deployed for real-time stock prediction to support decision-making.

B. Model Performance Metrics

The HLCE model outperforms all others across key metrics shown in Fig. 2, achieving the highest R^2 and lowest RMSE, MAE, and MAPE. While GARCH shows strong R^2, it underperforms elsewhere. ARIMA and ETS perform poorly, especially in classification. LSTM fails to meet most conditions. SVR, Random Forest, and XGBoost perform moderately, with XGBoost slightly leading in R^2. Overall, HLCE offers superior prediction accuracy.

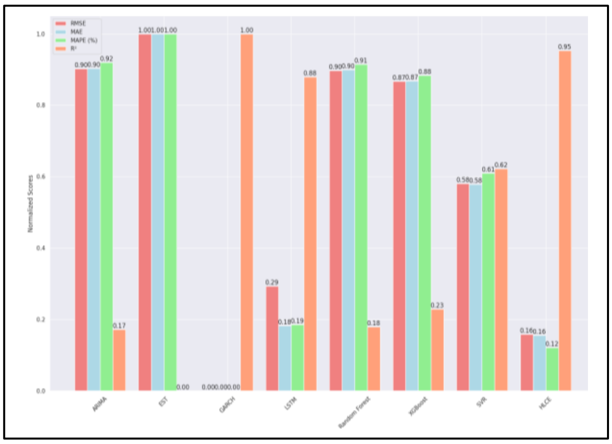


Fig 2: Evaluation Matrices of All Models

C. Visual Analysis of Model Performance

1) Correlation between the prediction errors

The error correlations amongst models are highlighted in the heatmap in Fig. 3. While HLCE's negative correlation with both ARIMA and ETS indicates complimentary behavior that is helpful for ensemble integration, ARIMA and ETS exhibit a substantial positive association.

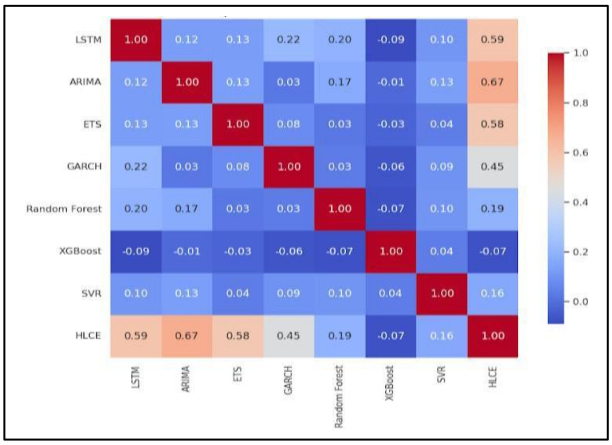


Fig 3: Correlation between the prediction errors

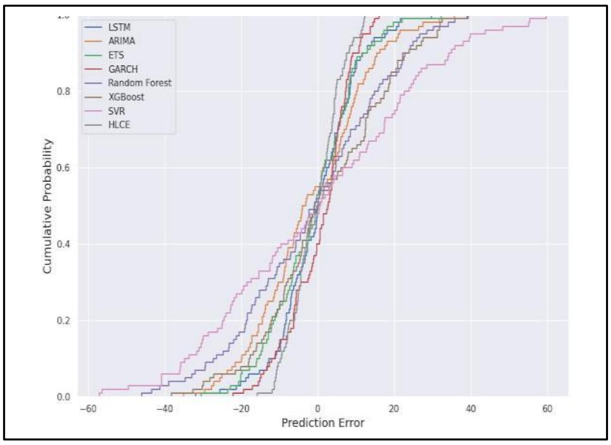


Fig 4: The residual analysis models CDF Plot

2) CDF Plot of Prediction Errors

The CDF plot in Fig. 4 shows HLCE outperforming all models with higher prediction accuracy. Random Forest and XGBoost follow, while ARIMA, ETS, and GARCH show broader error distributions.

3) RMSE Comparison

Fig. 5 shows HLCE with the lowest RMSE, proving high accuracy. SVR and XGBoost perform moderately, while ARIMA, GARCH, and ETS show higher errors.

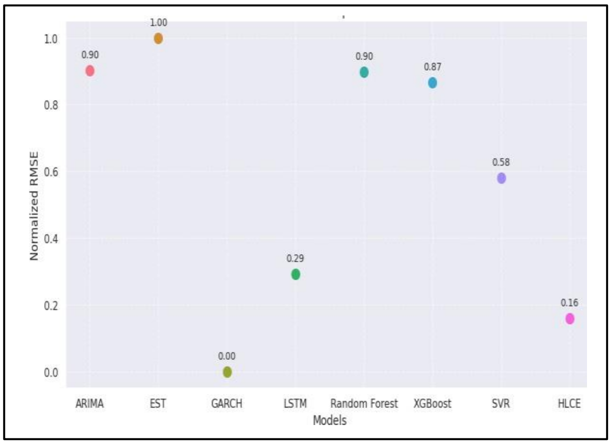


Fig 5: RMSE Comparison of Models

4) Box Plot of Prediction Errors

Fig. 6 demonstrate that HLCE produces steady and reliable predictions with a small IQR and few outliers. ARIMA and GARCH, on the other hand, show more outliers and broader IQRs, which indicate greater volatility and less dependable performance.

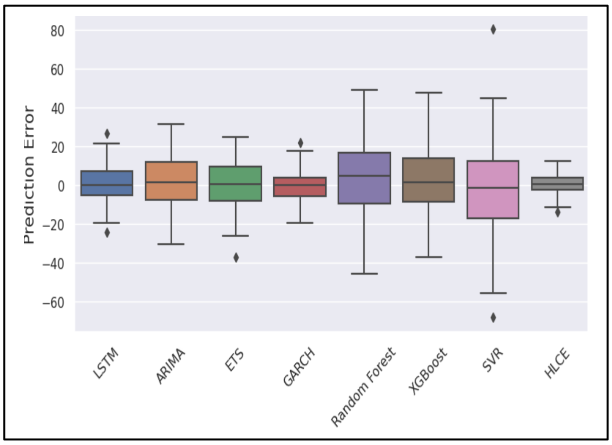


Fig 6: Box Plot of Prediction Error

5) Violon Plot of Prediction Errors

While ARIMA, ETS, and GARCH have wider, less stable error distributions, HLCE consistently produces low-error predictions, as seen in the violin plot in Fig. 7. Random Forest and XGBoost exhibit marginal gains, whereas LSTM is less reliable. In erratic markets, HLCE is the most dependable.

1) Heatmap Comparison of all models

The heatmap in Fig. 8 shows HLCE outperforming all models with the lowest errors and highest R², highlighting its adaptability and effectiveness in handling complex market patterns through hybrid modeling.

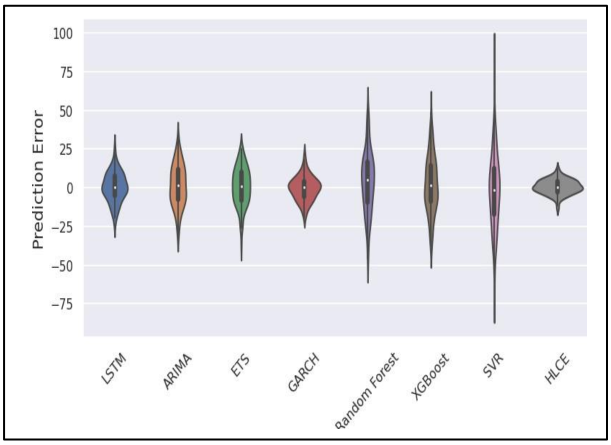


Fig 7: Violin Plot of Prediction Errors

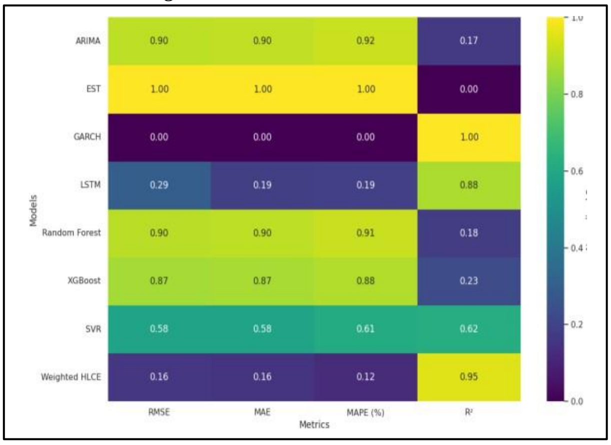


Fig 8: HeatMap comparison of all matrices

V. Conclusion and Future Work

This research compared several machine learning techniques, including LSTM, Random Forest, XGBoost, SVR, and traditional time series models like ARIMA and Exponential Smoothing, for stock price prediction. The results showed that machine learning models, particularly LSTM and ensemble approaches, outperformed conventional models in capturing non-linear trends. The HLCE (Hybrid LSTM-Conventional Ensemble) model achieved the highest R2 (0.95), and the lowest RMSE (0.16), MAE (0.16), and MAPE (0.12), demonstrating its superior accuracy. In contrast, Random Forest, XGBoost, and SVR had significantly lower performance (R2 values ranging from 0.18 to 0.62). Despite the promising results, challenges remain, and future research could explore the performance of ensemble methods with larger, more diverse datasets, incorporating external variables like market sentiment, macroeconomic factors, and geopolitical events. Additionally, integrating multi-source data, such as sentiment analysis, and improving feature engineering and model interpretability will be key areas for future improvement.

References

[1] Khashei, M., Bijari, M., & Hejazi, S. R. (2012). Combining seasonal ARIMA models with computational intelligence techniques for time series forecasting. Soft computing, 16, 1091-1105.

[2] Gardner Jr, E. S. (1985). Exponential smoothing: The state of the art. Journal of forecasting, 4(1), 1-28.

[3] Gan, M., Cheng, Y., Liu, K., & Zhang, G. L. (2014). Seasonal and trend time series forecasting based on a quasi-linear autoregressive model. Applied soft computing, 24, 13-18.

[4] Omran, M. F., & McKenzie, E. (2000). Heteroscedasticity in stock returns data revisited: volume versus GARCH effects. Applied Financial Economics, 10(5), 553-560.

[5] Kim, H. J., & Shin, K. S. (2007). A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. Applied Soft Computing, 7(2), 569-576.

[6] Kanas, A. (2003). Non-linear forecasts of stock returns. Journal of Forecasting, 22(4), 299-315.

[7] Siami-Namini, S., et al. (2019). The performance of LSTM and BiLSTM in forecasting time series. IEEE International Conference on Big Data, 2333–2342.

[8] Li, S., et al. (2022). Stock price prediction based on LSTM deep learning model with attention mechanism. IEEE Access, 10, 23743–23751.

[9] Md, A. Q., Kapoor, S., AV, C. J., Sivaraman, A. K., Tee, K. F., Sabireen, H., & Janakiraman, N. (2023). Novel optimization approach for stock price forecasting using multi-layered sequential LSTM. Applied Soft Computing, 134, 109830.

[10] Zhang, X., Liang, X., Zhiyuli, A., Zhang, S., Xu, R., & Wu, B. (2019, July). At-lstm: An attention-based lstm model for financial time series prediction. In IOP Conference Series: Materials Science and Engineering (Vol. 569, No. 5, p. 052037). IOP Publishing.

[11] Ubal, C., Di-Giorgi, G., Contreras-Reyes, J. E., & Salas, R. (2023). Predicting the long-term dependencies in time series using recurrent artificial neural networks. Machine Learning and Knowledge Extraction, 5(4), 1340-1358.

[12] Liu, Y., Gong, C., Yang, L., & Chen, Y. (2020). DSTP-RNN: A dual-stage two-phase attention-based recurrent neural network for long-term and multivariate time series prediction. Expert Systems with Applications, 143, 113082.

[13] Sen, J., & Mehtab, S. (2022). Long-and-Short-Term Memory (LSTM) NetworksArchitectures and Applications in Stock Price Prediction. Emerging computing paradigms: Principles, advances and applications, 143-160.

[14] Parmar, A., Katariya, R., & Patel, V. (2018, August). A review on random forest: An ensemble classifier. In International conference on intelligent data communication technologies and internet of things (pp. 758-763). Cham: Springer International Publishing.

[15] Basudhar, A., & Missoum, S. (2009). Local update of support vector machine decision boundaries. In 50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference 17th AIAA/ASME/AHS Adaptive Structures Conference 11th AIAA No (p. 2189).

[16] Hamidzadeh, J., & Moslemnejad, S. (2019). Identification of uncertainty and decision boundary for SVM classification training using belief function. Applied Intelligence, 49(6), 2030-2045.

[17] Ali, J., Khan, R., Ahmad, N., & Maqsood, I. (2012). Random forests and decision trees. International Journal of Computer Science Issues (IJCSI), 9(5), 272.

[18] Salman, H. A., Kalakech, A., & Steiti, A. (2024). Random forest algorithm overview. Babylonian Journal of Machine Learning, 2024, 69-79.

[19] Banfield, R. E., Hall, L. O., Bowyer, K. W., & Kegelmeyer, W. P. (2006). A comparison of decision tree ensemble creation techniques. IEEE transactions on pattern analysis and machine intelligence, 29(1), 173-180.

[20] Kuvulmaz, J., Usanmaz, S., & Engin, S. N. (2005). Time-series forecasting by means of linear and nonlinear models. In MICAI 2005: Advances in Artificial Intelligence: 4th Mexican International Conference on Artificial Intelligence, Monterrey, Mexico, November 14-18, 2005. Proceedings 4 (pp. 504-513). Springer Berlin Heidelberg.

[21] Sirisha, U. M., Belavagi, M. C., & Attigeri, G. (2022). Profit prediction using ARIMA, SARIMA and LSTM models in time series forecasting: A comparison. IEEE Access, 10, 124715-124727.

[22] Pan, H., Tang, Y., & Wang, G. (2024). A Stock Index Futures Price Prediction Approach Based on the MULTI-GARCH-LSTM Mixed Model. Mathematics, 12(11), 1677.

[23] García-Medina, A., & Aguayo-Moreno, E. (2024). LSTM–GARCH hybrid model for the prediction of volatility in cryptocurrency portfolios. Computational Economics, 63(4), 1511-1542.

[24] Mehtarizadeh, H., Mansouri, N., Mohammad Hasani Zade, B., & Hosseini, M. M. (2025). Stock price prediction with SCA-LSTM network and Statistical model ARIMA-GARCH. The Journal of Supercomputing, 81(2), 366.

[25] Li, H., Shen, Y., & Zhu, Y. (2018, November). Stock price prediction using attention-based multi-input LSTM. In Asian conference on machine learning (pp. 454-469). PMLR.

[26] Song, D. Stock Price Prediction based on Time Series Model and Long Short-term Memory Method.

[27] Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PloS one, 12(7), e0180944.

[28] Chen, S., & Ge, L. (2019). Exploring the attention mechanism in LSTM-based Hong Kong stock price movement prediction. Quantitative Finance, 19(9), 1507-1515.

[29] Li, H., Shen, Y., & Zhu, Y. (2018, November). Stock price prediction using attention-based multi-input LSTM. In Asian conference on machine learning (pp. 454-469). PMLR.

[30] Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Annals of Data Science, 10(1), 183-208.

[31] Lendasse, A., de Bodt, E., Wertz, V., & Verleysen, M. (2000). Non-linear financial time series forecasting-Application to the Bel 20 stock market index. European Journal of Economic and Social Systems, 14(1), 81-91.

[32] Ouahilal, M., et al. (2021). A comparative study of stock price prediction models based on machine learning techniques. SN Computer Science, 2(1), 1-12.

[33] Bukhari, A. H., Raja, M. A. Z., Sulaiman, M., Islam, S., Shoaib, M., & Kumam, P. (2020). Fractional neuro-sequential ARFIMA-LSTM for financial market forecasting. Ieee Access, 8, 71326-71338.

[34] Huang, W., et al. (2020). A novel forecasting model of stock price movement based on multiple time-series indicators. IEEE Access, 8, 69712-69722.

[35] Qiu, J., Wang, B., & Zhou, C. (2020). Forecasting stock prices with long-short term memory neural network based on attention mechanism. PloS one, 15(1), e0227222.

[36] Zhang, X., Liang, X., Zhiyuli, A., Zhang, S., Xu, R., & Wu, B. (2019, July). At-lstm: An attention-based lstm model for financial time series prediction. In IOP Conference Series: Materials Science and Engineering (Vol. 569, No. 5, p. 052037). IOP Publishing.

[37] Zhang, J., & Liu, Y. (2005). SVM decision boundary based discriminative subspace induction. Pattern Recognition, 38(10), 1746-1758.

[38] Liu, B., et al. (2021). Stock movement prediction from news articles with transfer learning. Procedia Computer Science, 179, 768-776.

[39] Althelaya, K. A., El-Alfy, E. S. M., & Mohammed, S. (2018, April). Evaluation of bidirectional LSTM for short-and long-term stock market prediction. In 2018 9th international conference on information and communication systems (ICICS) (pp. 151-156). IEEE.

[40] Beniwal, M., Singh, A., & Kumar, N. (2024). Forecasting multistep daily stock prices for long-term investment decisions: A study of deep learning models on global indices. Engineering Applications of Artificial Intelligence, 129, 107617.

[41] Takahashi, S., et al. (2019). Modeling and prediction of stock prices by generative adversarial networks. IEICE Transactions on Information and Systems, 102(5), 907-915.

[42] Wang, Y., et al. (2022). Stock price prediction with GAN-based data augmentation. Applied Soft Computing, 117, 108452.

[43] Adams, S. O., Asemota, O. J., & Ibrahim, A. A. (2024). Asymmetric GARCH Type Models and LSTM for Volatility Characteristics Analysis of Nigeria Stock Exchange Returns. American Journal of Mathematics and Statistics, 14(2), 17-32.

[44] Qiu, M., & Song, Y. (2016). Predicting the direction of stock market index movement using an optimized artificial neural network model. PloS one, 11(5), e0155133.

[45] Kumari, J., et al. (2023). Towards enhanced stock prediction using event-driven LSTM networks. Information Systems Frontiers, 1-16.

[46] Naik, N., & Mohan, B. R. (2019). Stock price prediction using event-linked data and hybrid deep learning techniques. Applied Intelligence, 49(8), 2782-2799.

[47] Sohangir, S., et al. (2018). Big data composite representation and optimization using a hybrid model of a deep autoencoder and eigenvalue decomposition. IEEE Transactions on Knowledge and Data Engineering, 30(9), 1772-1779.

[48] Lim, B., et al. (2019). Temporal convolutional networks for stock price prediction. Machine Learning with Applications, 8, 100080.

[49] Xu, K., Wu, Y., Jiang, M., Sun, W., & Yang, Z. (2024). Hybrid LSTM-GARCH Framework for Financial Market Volatility Risk Prediction. Journal of Computer Science and Software Applications, 4(5), 22-29.

[50] Han, H., Liu, Z., Barrios Barrios, M., Li, J., Zeng, Z., Sarhan, N., & Awwad, E. M. (2024). Time series forecasting model for non-stationary series pattern extraction using deep learning and GARCH modeling. Journal of Cloud Computing, 13(1), 2.

Optimized Skin Cancer Classification through Transfer Learning with MobileNetV3 and Mathematical Morphological Preprocessing

Sidra Mehboob
Department of Computer Science
COMSATS University Islamabad, Attock Campus
Attock 43600, Pakistan
sidratulmuntaha02@gmail.com

Maryam Bukhari
Department of Computer Science
COMSATS University Islamabad
Attock 43600, Pakistan
orcid.org/0000-0003-2904-2223

Yaser Ali Shah
Department of Computer Science
COMSATS University Islamabad
Attock 43600, Pakistan
orcid.org/0000-0003-1628-1743

Salabat Khan
Big Data Research Center
Jeju National University
Jeju 63243, Republic of Korea
orcid.org/0000-0002-3623-498X

Muhammad Sharif
Department of Computer Science
COMSATS University Islamabad
Attock 43600, Pakistan
orcid.org/0009-0005-4047-2787

***Abstract*—Skin cancer remains one of the most fatal diseases worldwide, highlighting the urgent need for an early and accurate diagnosis. Traditional diagnostic methods often rely on manual feature extraction or require extensive computational resources, which can limit their practical application. In this study, we propose an efficient and automated skin cancer classification framework that leverages transfer learning with the MobileNetV3 architecture, particularly suited for deployment in resource-constrained environments. To improve visual quality and enhance feature distinction in dermoscopic images, mathematical morphology techniques are used during the preprocessing phase. This integration refines image clarity and supports more effective feature learning by the model. The proposed method is evaluated using the publicly available ISIC dataset. The experimental results demonstrate that the approach achieves an accuracy of 89% and a recall of 85%, surpassing several existing models. These results suggest that combining MobileNetV3 with morphological preprocessing significantly improves classification performance, particularly in differentiating malignant from benign lesions. Furthermore, the lightweight nature of the MobileNetV3 model enables potential integration into mobile applications for real-time diagnostic support. In general, this work offers a promising solution for the early detection of skin cancer and has the potential to help dermatologists and healthcare professionals deliver accurate and timely evaluations.**

***Index Terms*—Skin Cancer Detection, Transfer Learning, MobileNetV3, Mathematical Morphology, Dermoscopic Image Classification**

I. Introduction

Skin cancer is among the most prevalent and life-threatening forms of cancer globally. Early identification of skin lesions, particularly melanoma, plays a critical role in reducing mortality rates. With increasing exposure to ultraviolet (UV) radiation and lifestyle changes, the incidence of skin cancer has increased significantly over the past few decades. Countries such as the United States, Australia and Canada have reported a steep increase in melanoma cases, especially among individuals aged 30 and older. In this context, computer-aided diagnostic (CAD) systems have become indispensable tools for helping dermatologists detect skin cancer through dermoscopic image analysis. These systems offer a noninvasive, efficient, and cost-effective approach to early diagnosis, which can greatly improve treatment outcomes.

Traditional machine learning (ML) methods were among the first solutions introduced to automate the detection of skin cancer. These approaches involve manual feature extraction techniques, in which texture, shape, color, and size are extracted and passed to classifiers such as SVM, KNN, or Naive Bayes [1]–[8]. Although these systems outperformed manual diagnostic accuracy, they required complex pre-processing and domain expertise for feature engineering. Techniques such as the Firefly algorithm with KNN [4], GLCM with SVM [5], and HOG with LDA [6] have shown promising results. However, these methods often struggle with variations in the appearance of the lesion and background noise, reducing their reliability in real-world applications.

With the advent of deep learning (DL), Convolutional Neural Networks (CNNs) began to replace traditional techniques due to their ability to perform automatic feature extraction. Pre-trained models such as AlexNet, ResNet50, Inception-V3, and VGG19 have demonstrated remarkable performance in classifying benign and malignant lesions [9]–[14]. Transfer learning further enhanced these architectures by enabling the reuse of learned features from large-scale datasets, which is particularly useful when domain-specific labeled data are limited. However, while CNNs reduce manual intervention, they still face issues related to low image contrast, noise,

Corresponding author: Muhammad Sharif (email: m.sharif@cuiatd.edu.pk)

and ambiguous lesion boundaries that affect classification performance.

To address these shortcomings, researchers have explored ensemble learning, multi-model systems, and attention-based mechanisms. Ensemble strategies that combine the output of ResNet, DenseNet, VGG16, and Inception networks have shown superior accuracy, reaching up to 98.6% in some studies [15], [16]. Other innovations, such as IoHT-integrated CAD systems [12] and attention mechanisms [17], were aimed at improving feature discrimination. Furthermore, hybrid systems that combine deep learning feature extraction with traditional classifiers have been proposed [18]. However, most existing methods are either computationally intensive or not optimized for deployment in mobile and resource-limited environments. Despite significant progress, current systems still face key limitations. Many deep learning models require high-end computing resources, making them unsuitable for real-time or mobile-based applications. In addition, some models fail to generalize well due to challenges such as varying lesion sizes, poor image quality, and class imbalance in datasets. This study aims to bridge these gaps by investigating the following research questions: (1) How can we optimize deep learning models for low-resource environments without compromising performance? (2) Can mathematical morphology improve dermoscopic image quality for improved classification? The main objective is to develop an efficient, accurate and lightweight CAD system capable of real-time skin cancer detection.

To meet these objectives, we propose a novel framework that combines the MobileNetV3 architecture with mathematical morphology-based preprocessing techniques. MobileNetV3 is selected for its balance between accuracy and computational efficiency, making it ideal for mobile deployment. The preprocessing stage enhances the features of the dermoscopic image, allowing the model to learn more effectively from the input data. Based on the ISIC dataset, our system achieves a classification accuracy of 89% and a recall of 85%, outperforming several existing models. The contributions of this research are threefold: (1) design of a lightweight and accurate deep learning model, (2) enhancement of image quality using morphological operations, and (3) demonstration of a scalable solution for real-time skin cancer diagnosis. This work paves the way for future mobile health applications that support dermatologists and patients in timely and accurate diagnosis. The rest of the paper is organized into various sections: Section II identifies the proposed approach in detail, Section IV discusses experimental results, and Section V presents the conclusions with future directions.

II. Methodology

In this section, we explain our suggested method in detail, step by step. The pictorial view of the suggested methodology is depicted in Fig. 1.

A. Preprocessing:

In the first stage, we applied the pre-processing steps on dermoscopic images to remove noise and other artifacts so that the resulting images are more enhanced and clearer in visibility. Therefore, we resized all images from the dataset to fixed dimensions of 224×224×3. Later, in all images, we have applied the morphological operation, i.e., a dilation followed by the erosion operation. This operation removed the extra noise appearing on the dermoscopic images of skin cancer. After that, the 2D kernel matrix was used to convolve the given dermoscopic image to enhance and improve the contrast of the image, as shown in Fig. 2. The kernel matrix is the sharpening kernel given in the following.

$$\text{Kernel} = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{1}$$

The equation above 1 shows the kernel matrix used to sharpen the image after going through morphological operations. All these pre-processing steps prepare the data to be given as input to the deep CNN model.

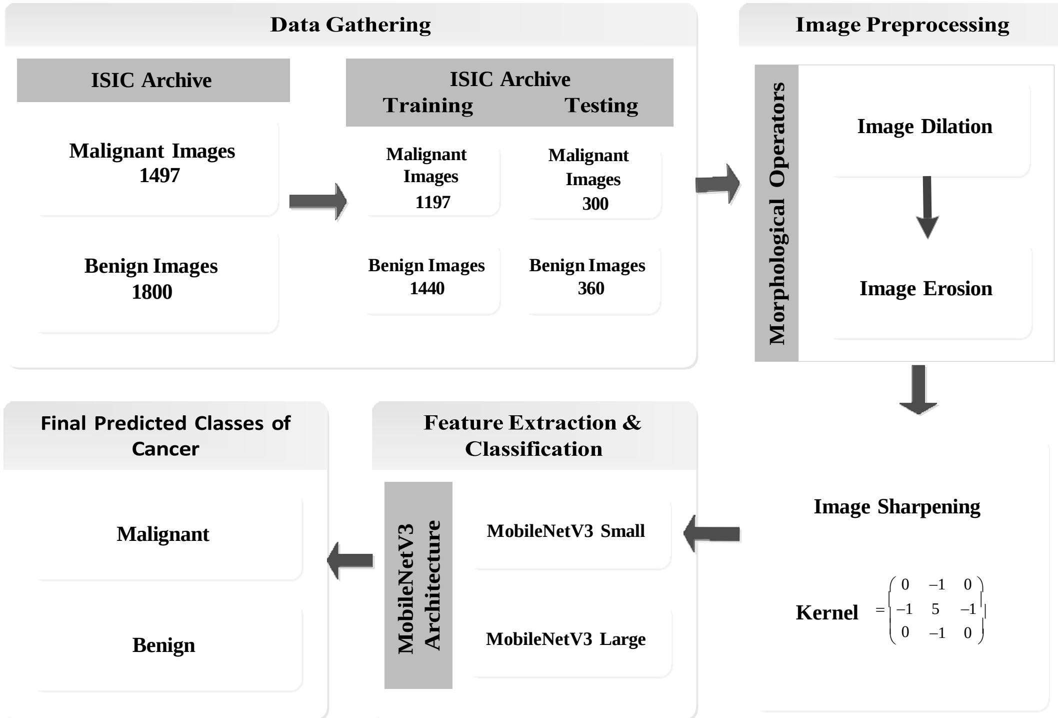
B. MobileNetV3 Architecture:

MobileNetV3 is a variant of the Convolutional Neural Network (CNN) architecture designed for mobile devices. The architecture is characterized by its use of depth-wise separable convolutions, which help to reduce computation and improve performance on resource-constrained devices. The general architecture of MobileNetV3 can be divided into three main building blocks. The initial block typically includes a single convolutional layer and a max pooling layer. The main body of the network, which is composed of several inverted residual blocks. Each block includes a separable convolution in depth, a convolution in points, and an optional expansion layer. The final block includes a global average pooling layer and a fully connected layer for classification. The architecture diagram of MobileNetV3 is shown in Fig. 1reffig:architecture.

MobileNetV3 also includes two versions. MobileNetV3-Large, which is a larger, more powerful version of the architecture, and MobileNetV3-Small, which is a smaller, more efficient version of the architecture. The main difference between these two versions is the number of channels in the inverted residual blocks and the number of blocks in the main body of the network. In the past, the MobileNet-based deep learning architecture has been introduced. In the first version of MobileNetV1 [19], the depth-wise separable convolutions are proposed, which show better improvements over existing architectures. Later, MobileNetV2 [20] added the linear bottleneck and reversed residual model to create much more effective layer structures utilizing the low rank of the situation. However, in MobileNetV3, the authors develop the most efficient methods employing a mix of efficient layers as key components, along with modified swish nonlinearities that are also added to the layers. [21]–[23].

C. Transfer Learning using MobileNetV3:

Despite having a significant amount of trained images, the dataset is still not large enough to create a new deep learning



$$\text{Kernel} = \begin{pmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{pmatrix}$$

Fig. 1: Process Flow Diagram of the Proposed Methodology.

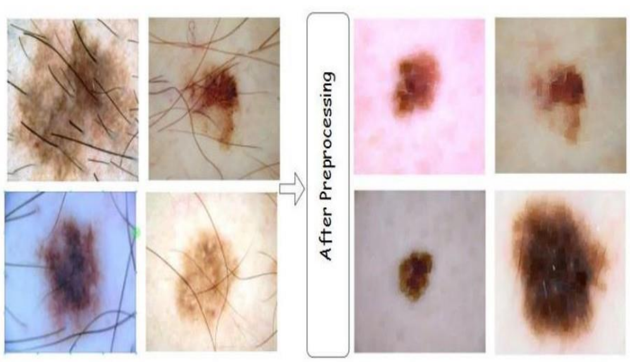


Fig. 2: Results of Preprocessing Operations.

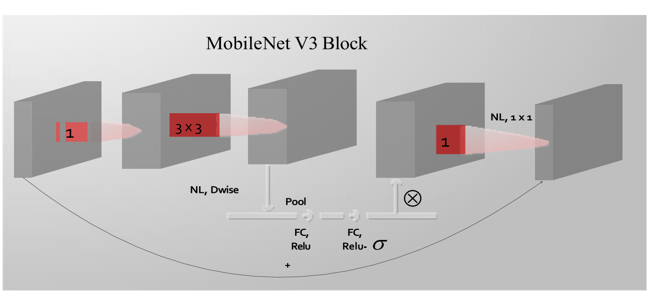


Fig. 3: Architecture Diagram of MobileNetV3.

model from scratch. To tackle this challenge, the technique of transfer learning is applied to the pre-trained MobileNetV3 architecture due to its more efficient feature learning process. For this purpose, at the start, the pre-trained MobileNetV3 [24] model is loaded. This model is trained on the ImageNet dataset, which is one of the largest image classification datasets. Later, the weights of the initial layers of the model are frozen by setting the trainable parameters to false. Afterwards, the flattening layer is added, which is followed by two fully connected layers. The total number of units and the activation function in the initial fully connected layer is 512, which have the activation of 'ReLu'. However, the final fully connected layer contains one hidden unit with the sigmoid activation function. Furthermore, a dropout layer of ratio 0.25 is added between two dense layers. The loss function for the algorithm is set to binary cross-entropy because skin cancer classification in malignant and begins is a binary classification problem. Table I provides the list of the hyperparameters set for the model.

III. Results and Discussions

In this section, the results of the designed model are elaborated along with the description of the datasets used to check the performance of the designed framework. In addition, the metrics used to analyze performance are also discussed here.

TABLE I: List of Hyperparameter Settings

Parameter	Value
Weight Optimizer	Adam
Learning Rate	0.00001
Batch Size	4
Loss Function	Binary Cross-Entropy
Dropout Ratio	0.25
Activation Function (Hidden Layer)	ReLU
Activation Function (Output Layer)	Sigmoid
Epochs	25

A. Dataset:

In this research study, we made use of the database that consists of images of skin cancer with two types, that is, malignant and benign. This dataset is publicly available on Kaggle. In this dataset, a total of 3297 files are categorized into training and test folders. This dataset was acquired from the ISIC archive. The total number of malignant images in the training set, the number of benign class images present in the training set, the total number of malignant images in the test set, and the total benign images in the test set are shown in Table II, respectively.

TABLE II: Data Distribution for Experiments

Class	Train	Test	Total
Malignant	1197	300	1497
Benign	1440	360	1800
Total	**2637**	**660**	**3297**

B. Performance Metrics:

The following are the performance analysis criteria used to measure performance. This includes accuracy, precision, recall, F_1-Score, confusion matrices, and ROC curves. Accuracy is one of the most popular classification analysis metrics, which is employed to examine the overall accuracy of the algorithm; however, if the data is imbalanced, then some other metrics, precision and recall, are used. Precision measures the efficiency of the model related to categorizing an instance as positive. Similarly, recall values are used to examine the ability of the model to determine the positive class from the overall positive instances. The higher the recall, the better the model. Similarly, the F_1Score shows the overall combination of precision values and recall values. Eventually, the confusion matrix is the overall summary in a more precise form of the underlying model. In the last case, the ROC curves are also plotted. The following are the set performance measures used for quantitative analysis: accuracy Eq. 2, precision Eq. 3, recall Eq. 4, and F_1-Score Eq. 5.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \qquad (2)$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \qquad (3)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \qquad (4)$$

$$F_1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100\% \qquad (5)$$

Where TP, TN, FP, and FN denote True Positive, True Negative, False Positive, and False Negative, respectively.

C. Performance Analysis of the Proposed Framework

In this section, we examine the effectiveness of the algorithm after applying the preprocessing steps. The results of the pre-processing are shown in Fig. 2. At first, the model is trained on a training set containing images of malignant and benign.

The MobileNetV3 Large's results in the malignant class are achieving 89% accuracy, 92% precision, 87% recall, and F_1-Score of 89%. In the benign class, the values are 89%, 85%, 91%, and 88%, respectively. However, the overall accuracy of this experiment is 89% of the results obtained with the Adam weight optimizer.

In the case of MobileNetV3 Small, the accuracy, precision, recall, and F_1-Score values with this experiment are 87%, 88%, 87%, and 88% in the malignant class, and in the benign class, these values are 87%, 85%, 86% and 86%, respectively. It is concluded from the general experiments that the results of the MobileNetV3 large in terms of average accuracy on both classes are higher than those of the MobileNetV3 small.

TABLE III: Confusion Matrix of the Proposed Methodologies

Metric	MobileNetV3 Small with Mathematical Morphology	MobileNetV3 Large with Mathematical Morphology
True Positive	260	275
True Negative	315	314
False Positive	45	46
False Negative	40	25

TABLE IV: Results of MobileNetV3-Large

Class	Accuracy	Precision	Recall	F_1-Score
Malignant	89%	92%	87%	89%
Benign	89%	85%	91%	88%

TABLE V: Results of MobileNetV3-Small

Class	Accuracy	Precision	Recall	F_1-Score
Malignant	87%	88%	87%	88%
Benign	87%	85%	86%	86%

Table III shows the confusion matrix using the MobileNetV3 Small and Large with mathematical morphology. In the case of MobileNetV3-Small, 315 samples of the malignant class and 260 samples of the benign class are correctly classified from the general samples of the test set. The remaining 45 and 40 test samples are not correctly classified. In case of MobileNetV3-Large version, 314 samples of the malignant class and 275 samples of the benign class are correctly classified from the overall samples of the test set. The remaining 46 and 25 test samples are not correctly classified. More precisely, we have individually assessed the performance of the proposed MobileNetV3 in both classes of skin cancer, that is, malignant and benign. In addition, the other performance values (accuracy, precision, recall, F_1-Score) are given in Tables IV and V.

The training accuracy and loss values are also recorded epoch by epoch, as shown in Fig. 6 and Fig. 6. It is observed from the graph of Fig. 6 that training accuracy is high with the Adam weight optimizer. Subsequently, the second graph shows the loss values of the proposed model. It is observed that, similar to accuracy, the Adam loss values are in training. Furthermore, if we analyze the performance of the proposed model on the test set, then it is observed that the accuracy values are higher with the Adam optimizer. More explicitly, we examine the effectiveness of the algorithm individually in both classes of skin cancer, i.e., malignant and benign.

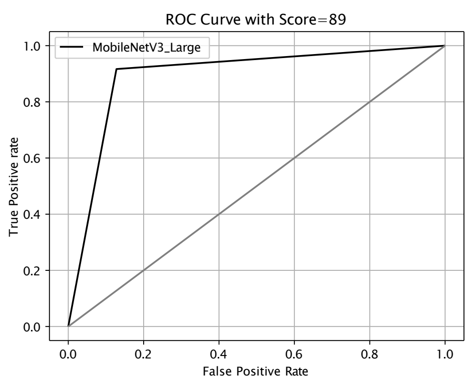


Fig. 4: ROC Curve generated by MobileNetV3 Large with Mathematical Morphology.

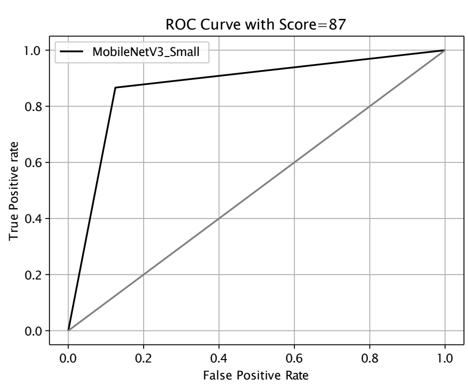


Fig. 5: ROC Curve generated by MobileNetV3 Small with Mathematical Morphology.

In addition to all the above results, we have plotted the ROC curves as shown in Figures 4 and 5 for both MobileNet models. The ROC curve stands for Receiver Operating Characteristic, basically employed to analyze the categorization model's efficacy. It is a curve among the true and false positive values. Shows high training accuracy values and very low loss values during training.

D. Comparative Analysis of the Proposed and Existing Techniques

It is observed from the results presented above in Table VI that the designed framework shows the best results in terms of accuracy, precision, recall, and F_1Score values. The comparison is done only on the approaches that have used the same dataset. In addition, in an attempt to discuss the evaluation of the suggested model, we compared the results of the proposed technique with existing methods. For example, Manasa et al. [23] suggest the Vgg16-based deep learning model and have achieved 86.6% accuracy. Similarly, Bechlli et al. [25] combined the three pre-trained models, namely VGG16, ResNet50, and DenseNet, as an ensemble network and achieved 85% accuracy.

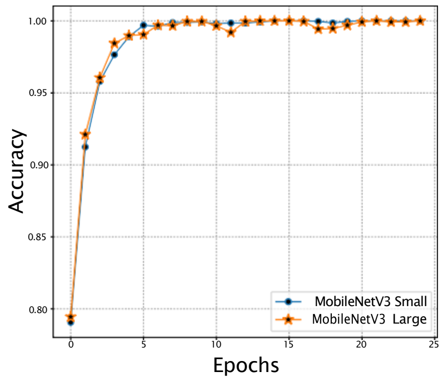
Demir et al. [26] suggest the transfer learning approach in which ResNet-101 and InceptionV3 are used to classify malignant and benign classes. In their work, the accuracy value is 84.09% and 87.42% and our proposed model shows an accuracy of 87% and 89% with the MobileNetV3Small and MobileNetV3Large versions, respectively.

TABLE VI: Comparison of the Proposed and Existing Techniques

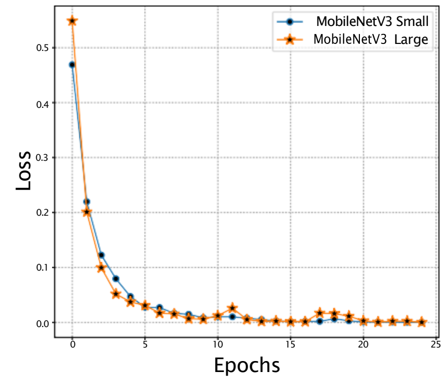
Research paper [reference]	Approach	Accuracy
Manasa et al. [23]	VGGA16	86%
Bechelli et al. [25]	VGG16, Xception, ResNet50	85%
Demir et al. [26]	ResNet-101	84.09%
Demir et al. [26]	Inception-V3	87.42%
Proposed	MobileNetV3 Small	87%
Proposed	MobileNetV3 Large	89%
Proposed	Average	88%

IV. Conclusion

Skin cancer is one of the deadliest diseases. The color images of the skin show a significant degree of resemblance between different skin lesions, such as malignant and benign, making identification and diagnosis more challenging. Detection in time requires a reliable automation process for skin cancer categorization to save efforts, time, and humanity. Therefore, in this research article, an attempt was made to provide a deep learning-assisted solution based on a transfer learning approach with the MobileNetV3 model. The MobileNetV3 provides improved feature learning due to the specialties of layers designed in the architecture. The suggested framework shows the effective values of accuracy, precision, recall, and F_1-Score. The findings suggest that this strategy may be an



(a) Accuracy



(b) Loss

Fig. 6: Accuracy and Loss Graphs with Mathematical Morphology: (a) Accuracy, (b) Loss.

acceptable way to diagnose skin cancer early and accurately, providing a practical tool that can be used in real time on mobile devices.

References

[1] C. Barata, M. E. Celebi, and J. S. Marques, "Improving dermoscopy image classification using color constancy," *IEEE journal of biomedical and health informatics*, vol. 19, no. 3, pp. 1146–1152, 2014.

[2] I. A. Ozkan and M. Koklu, "Skin lesion classification using machine learning algorithms," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 5, no. 4, pp. 285–289, 2017.

[3] M. Vidya and M. V. Karki, "Skin cancer detection using machine learning techniques," in *2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT)*. IEEE, 2020, pp. 1–5.

[4] S. Sivaraj, R. Malmathanraj, P. Palanisamy *et al.*, "Detecting anomalous growth of skin lesion using threshold-based segmentation algorithm and fuzzy k-nearest neighbor classifier," *Journal of cancer research and therapeutics*, vol. 16, no. 1, p. 40, 2020.

[5] Z. Waheed, A. Waheed, M. Zafar, and F. Riaz, "An efficient machine learning approach for the detection of melanoma using dermoscopic images," in *2017 International Conference on Communication, Computing and Digital Systems (C-CODE)*. IEEE, 2017, pp. 316–319.

[6] F. YILMAZ, B. UZUN, and S. ERG˙IN, "The diagnosis of melanoma skin cancer using histogram of oriented gradient based features," in *2nd International Congress on Engineering and Arthitecture*, 2019, pp. 1364–1369.

[7] P. Banasode, M. Patil, and N. Ammanagi, "A melanoma skin cancer detection using machine learning technique: support vector machine," in *IOP Conference Series: Materials Science and Engineering*, vol. 1065, no. 1. IOP Publishing, 2021, p. 012039.

[8] V. Balaji, S. Suganthi, R. Rajadevi, V. K. Kumar, B. S. Balaji, and S. Pandiyan, "Skin disease detection and segmentation using dynamic graph cut algorithm and classification through naive bayes classifier," *Measurement*, vol. 163, p. 107922, 2020.

[9] U.-O. Dorj, K.-K. Lee, J.-Y. Choi, and M. Lee, "The skin cancer classification using deep convolutional neural network," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 9909–9924, 2018.

[10] S. S. Chaturvedi, K. Gupta, and P. S. Prasad, "Skin lesion analyser: an efficient seven-way multi-class skin cancer classification using mobilenet," in *Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2020*. Springer, 2021, pp. 165–176.

[11] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Skin cancer classification using deep learning and transfer learning," in *2018 9th Cairo international biomedical engineering conference (CIBEC)*. IEEE, 2018, pp. 90–93.

[12] A. Khamparia, P. K. Singh, P. Rani, D. Samanta, A. Khanna, and B. Bhushan, "An internet of health things-driven deep learning framework for detection and classification of skin cancer using transfer learning," *Transactions on Emerging Telecommunications Technologies*, vol. 32, no. 7, p. e3963, 2021.

[13] D. N. Le, H. X. Le, L. T. Ngo, and H. T. Ngo, "Transfer learning with class-weighted and focal loss function for automatic skin cancer classification," *arXiv preprint arXiv:2009.05977*, 2020.

[14] K. M. Hosny, M. A. Kassem, and M. M. Foaud, "Classification of skin lesions using transfer learning and augmentation with alex-net," *PloS one*, vol. 14, no. 5, p. e0217293, 2019.

[15] A. Mahbod, G. Schaefer, C. Wang, R. Ecker, and I. Ellinge, "Skin lesion classification using hybrid deep neural networks," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1229–1233.

[16] N. Kausar, A. Hameed, M. Sattar, R. Ashraf, A. S. Imran, M. Z. u. Abidin, and A. Ali, "Multiclass skin cancer classification using ensemble of fine-tuned deep learning models," *Applied Sciences*, vol. 11, no. 22, p. 10593, 2021.

[17] S. Jiang, H. Li, and Z. Jin, "A visually interpretable deep learning framework for histopathological image-based skin cancer diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 5, pp. 1483–1494, 2021.

[18] B. Shetty, R. Fernandes, A. P. Rodrigues, R. Chengoden, S. Bhattacharya, and K. Lakshmanna, "Skin lesion classification of dermoscopic images using machine learning and convolutional neural network," *Scientific Reports*, vol. 12, no. 1, p. 18134, 2022.

[19] K. Kaplan, Y. Kaya, M. Kuncan, and H. M. Ertunc¸, "Brain tumor classification using modified local binary patterns (lbp) feature extraction methods," *Medical hypotheses*, vol. 139, p. 109696, 2020.

[20] N. Noreen, S. Palaniappan, A. Qayyum, I. Ahmad, M. Imran, and M. Shoaib, "A deep learning model based on concatenation approach for the diagnosis of brain tumor," *IEEE Access*, vol. 8, pp. 55 135–55 144, 2020.

[21] M. Togˇac¸ar, B. Ergen, and Z. Co¨mert, "Brainmrnet: Brain tumor detection using magnetic resonance images with a novel convolutional neural network model," *Medical hypotheses*, vol. 134, p. 109531, 2020.

[22] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[23] K. Manasa and D. Murthy, "Skin cancer detection using vgg-16," *European Journal of Molecular & Clinical Medicine*, vol. 8, no. 1, pp. 1419–1426, 2021.

[24] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan *et al.*, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2019, pp. 1314–1324.

[25] S. Bechelli and J. Delhommelle, "Machine learning and deep learning algorithms for skin cancer classification from dermoscopic images," *Bioengineering*, vol. 9, no. 3, p. 97, 2022.

[26] A. Demir, F. Yilmaz, and O. Kose, "Early detection of skin cancer using deep learning architectures: resnet-101 and inception-v3," in *2019 medical technologies congress (TIPTEKNO)*. IEEE, 2019, pp. 1–4.