ISSN: 2319-7897
Special Issue (2025-26)

MSIT Journal of Research

Satyam

Special Issue on Integrating Technologies, Ideas and Disciplines for Engineering Innovation (NCI-TIDE-2025)

MSIT Journal of Research – SATYAM

Special issue on Integrating Technologies, Ideas and Disciplines for Engineering Innovation (NCI-TIDE 2025)

Patrons

Sh. Kaptan Singh
President, SMES

Sh. Brahm Pal Singh
Patron, SMES

Sh. Ajit Singh Chaudhary
Secretary, SMES

Sh. S.S. Solanki
Joint Secretary, SMES

Sh. Shiv Ram Tewatia
Joint Secretary, SMES

Dr. Tejbir Singh Rana
Convener, Higher Education
Advisory Committee, SMES

Smt. Esha Jakhar
Sr. Vice President, SMES

Sh. Y.P.S. Verma
Vice President, SMES

Sh. Raj Pal Solanki
Treasurer, SMES

Prof. Prem Vrat
Pro-Chancellor, NorthCap
University, Gurugram

Sh. Karnal Singh
IPS, Former Director
Enforcement Directorate

Editors-in-Chief: **Prof. (Dr.) Avanish Kumar Srivastava**
Prof. (Dr.) Archana Balyan

Editors: **Prof. (Dr.) Neeru Rathee**
Prof. (Dr.) Pardeep Sangwan
Prof. (Dr.) Puneet Azad

MSIT Journal of Research – SATYAM
is an annual publication of
MAHARAJA SURAJMAL INSTITUTE OF TECHNOLOGY
C-4, Janak Puri, New Delhi-110058 (India)
Phone: 011-65215941
E-mail: satyamjournal@msit.in, Visit us at www.msit.in

Special Issue (NCI-TIDE 2025)

MSIT Journal of Research – SATYAM
Special issue on Integrating Technologies, Ideas and Disciplines for Engineering Innovation

Editors-in-chief: Prof. (Dr.) Avanish Kumar Srivastava
Prof. (Dr.) Archana Balyan

Copy Right © MSITJR – Special Issue – NCI-TIDE 2025

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical including photocopying, recording or by any information storage and retrieval system, without the prior written permission from the copyright owner. However, permission is not required to copy abstracts of papers on condition that a full reference to the source is given.

ISSN 2319-7897

Disclaimer

The opinions expressed and figures provided in this journal are sole responsibility of the authors. The publisher and the editor bear no responsibility in this regard. Any or all such liabilities are disclaimed.

All disputes are subject to Delhi jurisdiction only

Address for Correspondence
Prof. (Dr.) Avanish Kumar Srivastava
Director
Maharaja Surajmal Institute of Technology,
C-4, Janakpuri, New Delhi-110058
Tel: 011-65215941
E-mail: satyamjournal@msit.in

CONTENT

S.NO.	PAPERS	PAGE NO.
1.	**DESIGN AND ANALYSIS OF A COMPACT MICROSTRIP PATCH ANTENNA** *Arihant Jain, Arnav Garg, Aanya Dhoundiyal, Anushka Sharma, Dishaa Singh, Aman Kumari Dahiya*	1
2.	**Advanced Technologies with Revolutionizing Healthcare** *Jyoti Jain, Sumit Rao, Manik Mattoo*	6
3.	**AI and Web3: Integrating Autonomous Agents for Next-Gen Decentralized Finance** *Ved Mohan, Meena Rao, Parul Chaudhary*	11
4.	**Extreme Lossy SVG Compression for Bandwidth-Critical Applications: Trading Visual Fidelity for 99% Size Reduction in AI-Generated** *Prakhar Chandra, Pulkit Kapur, Jaspreet Singh*	15
5.	**ScoreMatrix: A Parameter-Guided Framework for Automated Evaluation of Answer Sheets** *Pulkit Kapur, Prakhar Chandra, Jaspreet Singh*	22
6.	**Robot Path Planning Algorithms: A Comparative Analysis of Dynamic Window Algorithm and Elastic Band Planner** *Sanchit Sharma, Vishakha Tomar*	30
7.	**Design, Simulation, and Parametric Analysis of a Rectangular Microstrip Patch Antenna Using Various Substrates for S-Band Wireless Applications** *Shubh Bhaskar, Mohammad Imaad, Dr. Aman Dahiya*	35
8.	**Stock Price Prediction Using Deep Learning** *Sanyam, Dr. Preeti Sehrawat, Dr. Sakshi Rajput, Ms. Himani*	41
9.	**Conceptualizing a Self-Reliant IDS Architecture for IoT Using Embedded Neural Networks** *Mamta Rani, Priyanshu Prasad*	46
10.	**AI-Powered College-Specific Chatbot System** *Dr .J. Shanthalakshmi Revathy, J. Matcha Rani, M. Ritika Sachdeva, T. G. Manasha Devi*	53
11.	**Toward Healthcare 5.0: Integrating Explainable AI and Trusted Execution Environments in Real-Time Embedded Systems** *Ashvini Adhikari, Dhanya Manocha, Dr. Neeru Rathee*	60
12.	**AI-Driven Emergency Care for the Elderly and Specially-Abled** *Deeksha Singh, Aarav Arora, Dr. Meena Siwach*	70
13.	**IOT-enabled Real-time Artificial Turf Monitoring And Management System Using Ai/ml** *S. Fazil, D. Ram Prabhu, N. Muthukumaran, N. Guru Sarves, S. Dharaneesh*	77
14.	**Smart Facial Recognition Based Attendance System Using Flask And Open-cv**	84

	Dr. Surendra Kumar Agrawal, Dr. Ashok Kumar, Divyanshi Singh, Mahima Khandelwal	
15.	**An IoT-Enabled Multi-Sensor Framework for Real-Time Water Quality Monitoring and Management**	94
	Saumya Maheshwari, Pranjal Sharma, Zarine Hayat Choudhury, Bunendra Singh, Puneet Azad, Deepshikha Yadav	
16.	**AI- Augmented Gesture Based Learning for Children with Neurodivergent and Physical Disabilities**	99
	Siya Kathpal, Anushka Yadav	
17.	**"Green Technology's Breakthroughs, Barriers, And The Road To Sustainable Growth"**	106
	Jyoti Jain, Vansh Sharma	
18.	**Performance Comparison of Fixed-Tilt and Solar Tracking Photovoltaic Systems (2020-2025)**	110
	Trivikram, Ishika, Krit Goel	
19.	**Role of Metrology for Quality assurance in Internet of Things (IOT) Applications**	117
	Deepshikha Yadav, Koyel Datta Gupta, Ishan Chaturvedi, Archana Balyan	
20.	**Non-Invasive Multi-Parameter Health Monitoring via All-Purpose PPG System**	122
	Ayush Kumar, Ayush Naik, Biswa Mohan Singh, Manas Gulati, Dr. Vimal Gaur	
21.	**A Compact and Cost-Effective 3-Element Yagi–Uda Antenna Optimized for VHF Amateur Band Operation**	132
	Avirath Magoo, Ishan Chaturvedi	
22.	**Uncertainty-Guided Machine Learning for Efficient Double Halide and Halide Perovskite Discovery: An Active Learning Framework with DFT Validation**	138
	Shaymon Dhawan, Kabir Batra, Abhishek Rana	
23.	**Impact of Modulation Techniques on 112 Gbps Dense WDM Digital RoF System Performance**	147
	Priyanka Thakur, Dr. Manisha Bharti	
24.	**Development of an AI for Versus Tetris Using Evolutionary Algorithms**	153
	Rayhan Khan	
25.	**An offline capable women safety device using integrated GPS and GSM technology**	160
	Mayank Saraswati, Kashish Dhawan, Puneet Azad	
26.	**Blockchain-Enabled Federated Learning Framework for Ensuring Data Integrity in Digital Metrology Systems**	165
	Daksh Malhotra, Bharat Kumar, Vishal Haldar, Dhruv Ghosal, Keshav Gupta, Ankit Jha	
27.	**Performance Measurement and Accuracy Evaluation of Integrated Multi-Band Antennas on a Single PCB for IoT Devices**	173
	Mayank Saraswati, Kashish Dhawan, Sudesh Pahal	

28.	**Rhythm Revealed: Advancing Accuracy In Heart Monitoring with Photoplethysmography** *Jolly Parikh, Vasu Sethi, Yugnanda Malhotra, Mohit Kumar, Shubham, Sudesh Pahal*	178
29.	**SafeTrack: An IoT-Enabled Multi-Sensor Measurement System for AI-Powered Emergency Detection and Smart Alerts** *Akshita Antil, Dr.Shaifali M.Arora, Dr.Poonam, Dr.Anshul Pareek*	183
30.	**Design and Integration of Multi-Band Antennas on a Single PCB: GPS and GSM Miniaturization for IoT Devices** *Sudesh pahal, mayank saraswati, neetu, kashish dhawan*	188
31.	**Evaluating Deep and Machine Learning Models for Robust Stock Price Forecasting: A Comparative Analysis** *Aditi Mishra, Aman Sharma, Aarav Gaur, Dr. Sudesh Pahal*	192
32.	**Menu Driven Multiple Disease Prediction Using Machine Learning** *Dr. Sakshi Rajput, Dr. Preeti Sehrawat, Nisha Bharti*	197
33.	**A Unified Suite for Forensic Extraction and Analysis of Encrypted Social Media Data from Android Devices** *Devesh, Sneha Juyal, Neeru Rathee*	202
34.	**Optimizing Photovoltaic Panel Tilt for Maximum Energy Yield: A Simulation-Based Study** *Jyoti Jain, Beenish Uvais*	208
35.	**A Comprehensive Review of Impedance Matching Techniques in RF and Microwave Systems: From Classical Networks to Loss-Aware Designs** *Kartikey Bhardwaj, Karan Vasisth, Nishtha, Vishakha Tomar*	212
36.	**A Unified Framework for Multi-Violation Traffic Monitoring and Contactless Driver Impairment Detection using Deep Learning** *Jai Oberoi, RK Choudhury, Yash Aggarwal*	216
37.	**Performance Analysis of a Cotton-Substrate Patch Antenna for Bluetooth Applications Under Bending and On-Body Conditions** *Vrinda Dixit, Yashika Sharma, Nishtha*	223
38.	**A Comparative Performance Evaluation of Regression Models for Photovoltaic Power Forecasting** *Upma Singh, Anjali Baliyan, Nisha Singh, Sonia Malik, Vinshu, Sudhanshu*	230
39.	**Hybrid RFID and Face Recognition-Based Smart Attendance System Using MQTT for Real-Time Analytics** *Raashi Jain, Khushal Bansal, Yashika, Dr. Shaifali M. Arora, Dr. Poonam, Dr. Anshul Pareek*	234
40.	**Quantitative Evaluation of Fourier-Domain Filter Geometries in Off-Axis Digital Holographic Phase Reconstruction** *Harsh Chatwani, Neelam Barak, Arun Anand, Gyanendra Sheoran*	240
41.	**Evolution of MIMO Antennas: From 2G to 6G Wireless Systems** *Anjali Baliyan, Upma Singh, Jasmine Chhikara, Karanjeet Singh Kharb, Shilpi Supriya Darshanai, Yash Kumar*	245

42.	**Voice Biometric Attendance System using Speaker Recognition** *Dr. Upma Singh, Ms. Sonia, Aditya, Janvi Bhatt, Jyoti Kumari*	251
43.	**Emotion-Aware Speaker Identification Using Spectrogram-Based CNN Models** *Bhavish Saroha, Akshat Bahl, Sonia Malik*	258
44.	**Variable Focal Length Metalens for Extended Depth of Focus Imaging** *Nishant Chhiller, Neelam Barak, Vineeta Kumari, Arun Anand, Gyanendra Sheoran*	263
45.	**The Parkinson's Disease Detection using Machine Learning Techniques** *Tanvi Rao, Neelam Nehra*	267
46.	**AI-Driven Cybersecurity: Threat Analysis and Organizational Transformation for 2025** *Atmadeep Chatterjee, Akshat Parmar, Komal Sharma, Anmol Bhatia*	271
47.	**Robotic Process Automation Framework for Streamlined Food Distribution and Hunger Mitigation** *Aditya Vashisht, Shailendra Bhushan Rai, Ankush Manhas, Gaurav Garg*	275
48.	**The Trust Deficit in the Algorithm Age: A Comparative Study on the Impact of AI-Assisted Diagnostic Tools on Patient Trust, Doctor–Patient Communication, and Clinical Practice** *Varun Hooda, Reet Rathore, Vipin Goyal, Rohit Rajput, Neeru Rathee, Puneet Azad*	281
49.	**Your Browser Is Snitching: The Invisible Army of Fingerprints, Cookies** *Navya Bora, Tanishka Bhardwaj, Sarvam Taneja, Sarthak Vijay, Shankul Sahni*	287
50.	**Customizable Gesture Mapping Device for Enhancing Accessibility in Wearable Sign Language Devices** *Lakshit Kaushik, Mansi Jha, Mansi Shrivastav, Priyanka Sharma, Itisha Das, Dr. Neeru Rathee*	295
51.	**GreenSight: An AI-Powered Real-Time Plant Health Monitoring and Disease Assistance System** *Sanjeet Singh Kushwaha, Satvik Singh, Shikhar Bahuguna*	300
52.	**Metrology-driven Advancements in Additive Manufacturing: Bridging Precision Gaps for Industrial Scalability** *Dr. Neeru Rathee, Dr. Tripti Sharma, Dr. Garima, Jasmine Chhikara*	304
53.	**Metrological Framework for Electroencephalography Systems: Uncertainty Quantification, Quality Assessment, and Clinical Application in Schizophrenia Diagnosis** *Neeru Rathee, Tripti Sharma, Garima, Jasmine Chhikara*	310
54.	**RAG-Enhanced Comic Panel Colorization Using GANs: A Retrieval-Augmented Approach for Context-Aare Automatic Colorization** *Jaspreet Singh, Prakhar Chandra, Pulkit Kapur*	316
55.	**IntelliRestore: A Smart Web-Based Group Photo Face Enhancement System using Advanced Deep Learning Architecture** *Sanjilka Saxena, Dr. Meena Siwach Rana*	323

56.	**Artificial Intelligence-Driven Drug Discovery and Molecular Design: Transforming Pharmaceutical Development**	329
	Harshita Khatkar, Kunal Singh, Pranav Gulati, Dr. Neelam Nehra, Dr. Dinesh Sheoran	
57.	**Multi-Culture Sign Language Recognition: Evolution from Image Processing to Advanced Deep Learning Architecture**	335
	Paridhi Gusain, Jayant Gulia, Dr. Shaifali M. Arora, Dr. Poonam, Dr. Anshul	
58.	**OPENRecon; a modular, open-source and customizable medical image reconstruction software**	340
	Ram, Sandhya Chaudhary, Naina, Jatin Yadav, Poonam Sheoran	
59.	**Low Light Image Enhancement using Machine Learning**	345
	Vaani Garg, Palak Gupta	
60.	**Design and Integration of Multi-Brand Antennas on a Single PCB: GPS and GSM Miniaturization fot IoT Devives**	352
	Sudesh Pahal , Mayank Saraswati , Neetu, Kashish Dhawan	

DESIGN AND ANALYSIS OF A COMPACT MICROSTRIP PATCH ANTENNA

Arihant Jain, Arnav Garg, Aanya Dhoundiyal, Anushka Sharma, Dishaa Singh, Aman Kumari Dahiya
Department of Electronics & Communication Engineering
Maharaja Surajmal Institute of Technology, New Delhi, India

***Abstract*—This work presents the design and optimization of a compact microstrip patch antenna. The study focuses on achieving a miniaturized structure by modifying the classical microstrip patch antenna design—which is primarily used for bluetooth applications—through the incorporation of a rectangular slot. The proposed design is suitable for portable devices and improves key performance parameters such as bandwidth, gain, and impedance matching, ensuring efficient operation in practical wireless communication scenarios.**

I. Introduction

With the continuous growth of wireless communication and the need for smaller and more efficient antennas, novel designs that can function well in the provided frequency range are in demand. With the advancement of wireless communication technologies and the proliferation of Internet of Things (IoT) devices, there is an increasing demand for antenna design that can save space and improve antenna performance in terms of bandwidth and gain [2], [4], [6]. Conventional antenna architectures typically fail to meet these needs, and hence a transition to more sophisticated design approaches is required [1]. This work examines advanced microstrip patch antenna design methods that are useful to enhance bandwidth and reflection coefficient [4], [8]. As a result, this work aims to propose counter-approaches to overcome inherent drawbacks of traditional design methods.

II. Background

The concept of a microstrip patch antenna has played a major role in the development of the compact design for its application in Bluetooth at 2.4 GHz frequency. Many research works have been carried out on different methods of improving bandwidth and performance in microstrip patch antennas. Substrate materials are shown to be a key determinant of operating properties; some materials offer improved directivity properties that enable wider bandwidths. Moreover, geometrical optimization techniques have been demonstrated to be of great significance in obtaining desired antenna characteristics.

A. Rectangular Patch Antenna

The rectangular patch antenna is one of the most commonly used types of microstrip patch antennas because it has a simple design, is easy to fabricate, and provides reliable performance. It is basically made of a metal patch printed on a dielectric substrate with a ground plane below it. This design efficiently supports the fundamental resonant mode and can be analyzed using methods like the transmission line or cavity model. Due to these advantages, it is widely used in real-world wireless systems.

B. Disadvantages

Original microstrip antenna configurations present several drawbacks, including narrow bandwidth, spurious feed radiation, poor polarization purity, limited power handling capacity, and sensitivity to fabrication tolerances [4], [6], [7]. As a result, much of the research and development in microstrip antennas has focused on addressing these limitations to meet the increasingly demanding requirements of modern systems.

C. Improvements

Possible improvements to conventional microstrip patch antenna design include introducing slots or meandered structures on the patch to reduce overall size, employing inset feeding to achieve stable impedance matching, and using partial or defected ground planes to improve bandwidth and radiation characteristics [1], [2], [4]–[7]. The shape and size of the patch antenna can be modified, such as in circular, square, or other geometries, to achieve better performance. These approaches can help adapt the antenna for compact, efficient, and reliable Bluetooth operation.

1) Slots: Adding slots is one of the popular methods for reducing the size of a rectangular microstrip patch antenna. By etching slots of different shapes (rectangular, circular, U-shaped, etc.) into the patch, the effective path of the current is increased [4], [7]. This helps to lower the resonant frequency while keeping the overall antenna size compact. Depending on the slot's shape and position, it can also be used to achieve dual-band or multiband operation.

2) Meanders: Meander lines or meander patches are another technique for antenna miniaturization. Here, the current path is folded into a smaller area, which increases the electrical length of the patch without making its physical size bigger. This approach is especially useful for designing antennas in portable and wearable devices, where space is limited but the antenna still needs to resonate at a required frequency, like 2.4 GHz.

3) Inset Feeds: Inset feeding is a commonly used technique to improve impedance matching in microstrip patch antennas [1], [2]. Instead of connecting the feed line at the edge of the patch, it is inserted at a specific point inside the patch. The location is carefully chosen so that the input impedance of the patch matches the characteristic impedance of the feed

line, which is usually 50 ohms. This reduces return loss and improves antenna efficiency without making major changes to the patch design.

III. Antenna Equations

The design of a rectangular microstrip patch antenna relies on several classical equations that define its dimensions and performance parameters [1]. These equations are derived from cavity model analysis and transmission line theory.

1) Width (W):

$$W = \frac{c}{2f_0\sqrt{\frac{\varepsilon_r + 1}{2}}} \quad (1)$$

where c is the speed of light in free space, f_0 is the resonant frequency, and ε_r is the dielectric constant of the substrate.

2) Effective Dielectric Constant (ε_{eff}):

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

where h is the height of the substrate.

3) Effective Length (L_{eff}):

$$L_{eff} = \frac{c}{2f_0\sqrt{\varepsilon_{eff}}} \quad (3)$$

4) Length Extension (ΔL):

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

5) Actual Length of the Patch (L):

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

IV. METHODOLOGY

In this section, simulation setup, meshing, and effect of size of the waveguide port will be discussed. Followed by the approach of selecting substrate materials, antenna geometry, and design parameter to achieve the desired operating frequency and improved antenna characteristics.

A. Software Overview

The compact patch antenna is then simulated by using CST Studio Suite by Dassualt Simulations [3], [4]. It is a powerful platform for precise modeling of the electromagnetic properties at the frequency of 2.4 GHz. This software allows for an in-depth analysis of many design variables to ensure the antenna performs to the required performance specifications. To improve the computational efficiency while maintaining result fidelity, boundary conditions and mesh are carefully set up. The selection of boundary conditions is of great influence on the computation of an open space environment, thus minimizing reflections that might lead to inaccurate results. Further, optimized mesh configurations can be used to accurately resolve the field distributions over the antenna surface, allowing efficient design iterations that ultimately lead to bandwidth and gain modifications. Such rigorous attention to detail when setting up a simulation is necessary to ensure that consistent and reliable results will be able to inform downstream design improvements.

B. Substrate Materials

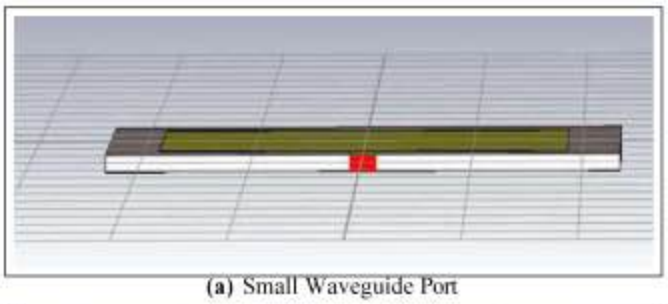
Of all substrate materials, FR4 was chosen as the substrate material in the compact patch antenna design , as it was motivated by the good blend of electrical and other mechanical properties [2], [4], [6]. FR4, with a relative dielectric constant of approximately 4.4 and a low loss tangent, provides a stable resonance and facilitates efficient signal

FR4 provides significant mechanical strength and affordability, which contributes to the miniaturization and useful manufacturability of the antenna – achieved benefits in the integration of modern wireless devices.

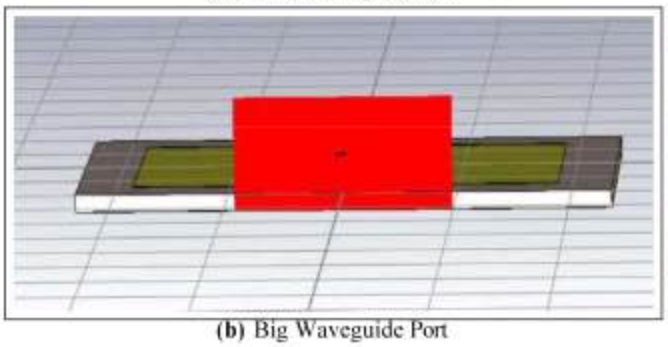
Other high-performance substrates such as Rogers RT/duroid or Taconic TLY offer lower dielectric losses and more stable permittivity over frequency, which can improve antenna efficiency; however, these materials are substantially more expensive and are less readily available, making them less suitable for cost-sensitive compact designs.

C. Effect of Waveguide Port Size

To investigate the effect of waveguide port sizing on calculated input matching, two port configurations were tested for the edge-fed reference antenna. First, a minimal aperture ("small port") that precisely covered only the feed conductor cross-section where, Ymin = Ymax = Zmax = 0 and Zmin = SH (refer Fig. 1 (a)) was used. This configuration produced poor and unstable S-parameter results because the port did not enclose the near fields associated with the stripline and substrate. Second, the aperture was increased using an extension coefficient(k) where, Ymin = Ymax = Zmax = k*SH and Zmin = SH (refer Fig. 1 (b)). The port fully captured the dominant mode and produced stable, physically realistic S11 and impedance results. The extension coefficient therefore represents how many times the substrate height the port aperture spans, and it controls whether the solver can form the correct guided-mode field at the port plane. While a larger aperture increases the mesh size and the chance of exciting higher-order modes, limiting the port to the fundamental mode and choosing moderate extension values gives much improved and reproducible S-parameter behavior. For inset-fed designs we still prefer a discrete port (lumped) for lower memory cost, but for clean edge-fed comparisons the extension-coefficient approach to waveguide-port sizing is a robust, reproducible method.



(a) Small Waveguide Port



(b) Big Waveguide Port

Fig. 1: Comparison of the size of waveguide ports.

D. Base Design and Parameters

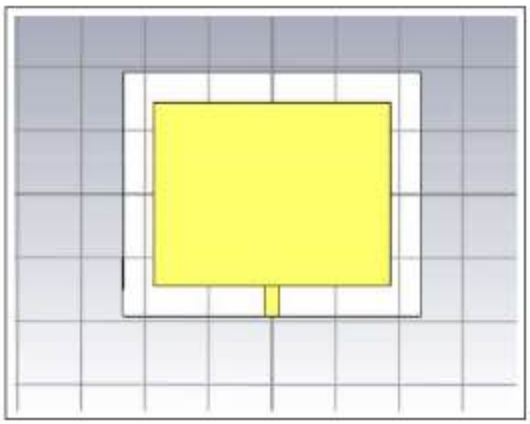


Fig. 2: Base Design 3D model

The base design (refer Fig. 2) utilizes an FR4 substrate with thickness of 1.6 mm and copper thickness 0.035 mm. The antenna dimensions were precisely calculated and optimized to ensure that it resonates at the required frequency, which is a critical step to achieve a high-performance and reliable wireless communication system. Using the formulas mentioned in the previous section, the various values for the design are summarized in Table I.

TABLE I: Calculated Microstrip Patch Antenna Dimensions

Parameter	Value (mm)
Substrate thickness	1.6
Copper thickness	0.035
Patch width	37.3
Patch length	28.8
Substrate length	38.4
Substrate width	46.9
Ground plane length	38.4
Ground plane width	46.9
Feed line width (50 Ω)	2.4
Feed line length	4.8

E. Modified Design and Parameters

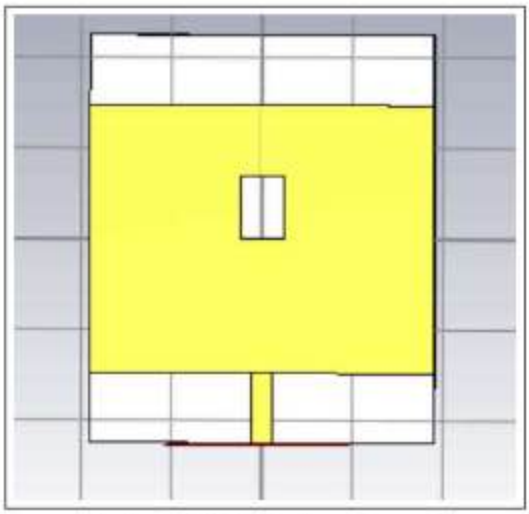


Fig. 3: Modified Design 3D model

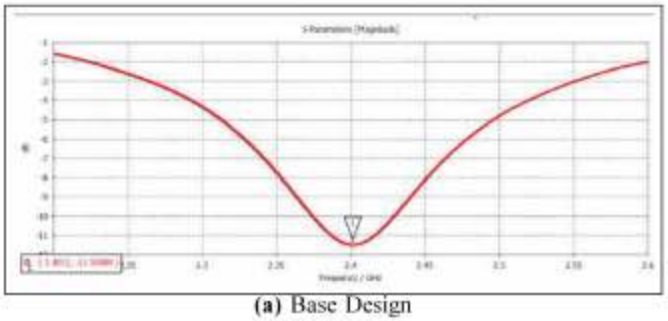
For the modified design (refer Fig. 3), the same substrate is used with $\varepsilon_r = 4.4$. Certain modifications in the patch's dimensions have been introduced in order to increase stability and reliability. A slotted patch configuration was used in order to further refine the efficiency of the antenna [4], [7]. Table II includes the precise patch as well as slot measurements:

TABLE II: Modified Microstrip Patch Antenna Dimensions

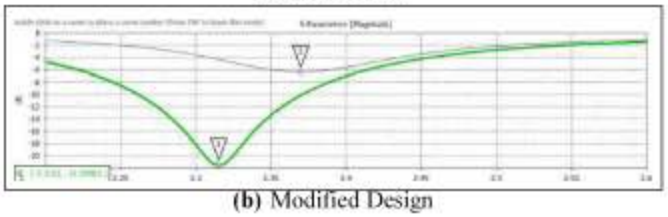
Parameter	Value (mm)
Substrate thickness	1.6
Copper thickness	0.035
Patch width	38
Patch length	29.5
Substrate length	45
Substrate width	38
Ground plane length	45
Ground plane width	38
Feed line width (50 Ω)	2.4
Feed line length	7.75
Rectangular slot width	2.4
Rectangular slot length	7

V. Simulated Results and Analysis

A. Return Loss (S11)



(a) Base Design



(b) Modified Design

Fig. 4: Comparison of S11.

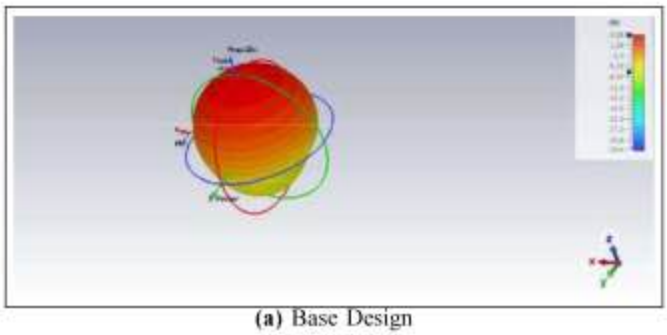
Return loss (S11) shows how much of the input power gets reflected back from the antenna because of impedance mismatch. A value less than -10 dB is considered good since it means most of the power is radiated instead of reflected [1], [3]. S11 of the base design shown in Fig. 4(a), exhibits a sharp resonance with a minimum S11 value of approximately -11.504 dB. While this value meets the typical engineering requirement for an acceptable impedance match ($S11 < -10dB$), the steepness of the curve indicates a narrow operational bandwidth. In contrast, the modified design shown in Fig. 4(b), which incorporates a rectangular slot, fundamentally alters the antenna's frequency response. The slot acts as a coupled resonator, and this phenomenon, known as double-tuning, is evident in the S11 plot as a significantly broadened resonance. This design modification works by creating a second resonance that interacts with the primary one, thereby expanding the antenna's impedance bandwidth to overcome the inherent narrow-band limitations of the base design. However, the new design also had the effect of shifting the antenna's frequency range downward. While the base design is perfectly centered on the 2.4 GHz bluetooth band, the modified design's deepest resonance of -21.599 dB occurs at approximately 2.31 GHz. As a result, its highest-performing region is not perfectly centered on the required frequency range for bluetooth applications.

B. Bandwidth Analysis

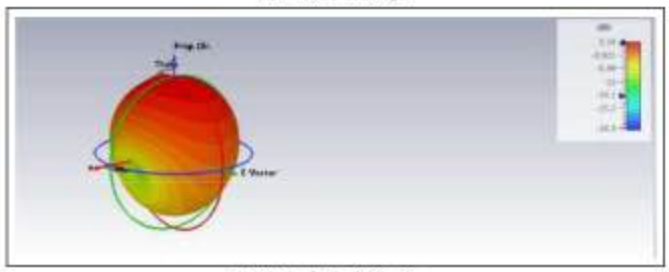
The impedance bandwidth is the range of frequencies where the return loss stays below -10 dB, ensuring stable antenna performance. From the S11 curves in Fig. 4, the modified design represents a better-performing antenna with a significantly wider bandwidth [4], [6], [8].

The base design exhibits a narrow bandwidth of only 50 MHz, with its S11 remaining below -10 dB from approximately 2.38 GHz to 2.43 GHz. In contrast, the modified design demonstrates a substantial improvement, achieving a much wider bandwidth of approximately 95 MHz. Its S11 stays below -10 dB from approximately 2.28 GHz to 2.375 GHz. This enhancement is a direct result of the design modifications, which successfully broadened the antenna's operational frequency range.

C. Radiation Pattern & Gain



(a) Base Design



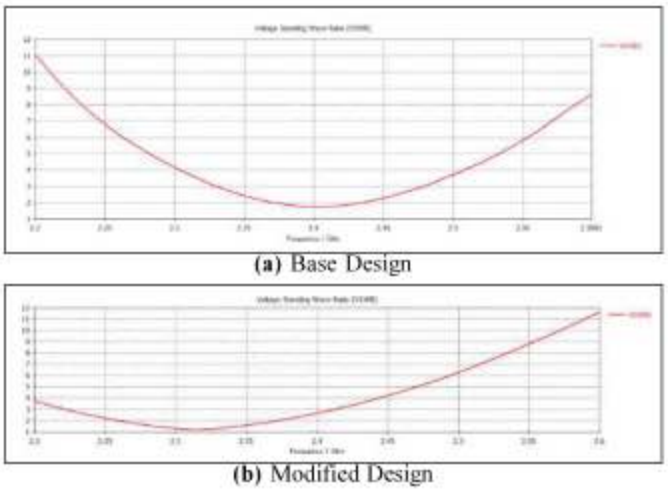
(b) Modified Design

Fig. 5: Comparison of Radiation Pattern and Gain.

The radiation pattern describes the spatial distribution of an antenna's radiated power, while gain quantifies its ability to direct that energy. For systems like Bluetooth, a pattern that is nearly omnidirectional is often preferred for maintaining connectivity regardless of device orientation . The simple, unslotted design (see Fig. 5 (a)) exhibits a broad, quasi-hemispherical radiation pattern. This is characteristic of a standard microstrip patch antenna, which radiates from its edges and possesses lower directivity, resulting in a peak gain of 5.39 dB.

In contrast, the modified design, which incorporates a rectangular slot, fundamentally alters the surface current distribution on the patch [6]. This strategic modification acts as a resonant element that reshapes the radiated field (see Fig. 5 (b)). The slot geometry concentrates the electromagnetic energy into a more focused and directional main lobe, suppressing radiation in undesired directions. While this modification increases the antenna's directivity, it results in a slightly lower peak gain of 5.16 dB. This trade-off between directivity and gain is an important design consideration, where the modified design prioritizes a more focused radiation for potential long-range or point-to-point communication.

D. VSWR



(a) Base Design

(b) Modified Design

Fig. 6: Comparison of VSWR.

Voltage Standing Wave Ratio (VSWR) is a crucial value that shows an antenna efficiency in terms of matching to the feed line. A smaller VSWR implies more power transfer and minimum reflections, which is highly important when using an antenna. Overall, VSWR of 2:1 or less can be regarded as acceptable in practice. [1]

For base design, the vswr is under 2:1 only within a very narrow band (refer Fig. 6 (a)). Beyond this range, vswr increases rapidly beyond 2:1, which signifies that impedance is not matched well at other frequencies other than the resonant frequency.

Under the modified (slot-integrated) structure, vswr remains less than 2:1 in a broader band (refer Fig. 6 (b)). Nevertheless, the frequency points at which the vswr is above the 2:1 level are still present, indicating irregular matching behavior.

This variation exists since the rectangular shape of the slot in the modified design varies the surface current distribution in a complicated way. The slot adds more resonances and influences the coupling and radiated energy of the patch, which is no longer a smooth broadband enhancement but an unstable impedance match at specific frequencies.

VI. Conclusion

This was a design and simulation analysis of a small patch microstrip antenna that can be applied in short-range wireless networks and other applications. The base design features a minimum return loss (S11) of approximately -11.504 dB and a 50 MHz (2.38 GHz to 2.43 GHz) impedance bandwidth that meets fundamental performance requirements. A resonance range where vswr is less than 2:1 is obtained, which implies that there is a tolerable impedance agreeableness at the target frequency. The slot-integrated (optimized) design is shown with a better minimum S11 of -21.599 dB at a given frequency point and the nominal bandwidth extended to 95 MHz (2.28 GHz to 2.375 GHz). All in all, our design has superior performance in critical parameters like the return loss, vswr and bandwidth compared to the basic model and thus can be used in compact IoT sensors and wireless communication gadgets among other uses. The frequency range 2.28-2.375 GHz that fits within the range of bandwidth achieved is applicable in applications of ZigBee, some Wi-Fi channels, wireless sensor networks, industrial telemetry systems, and low-power communication devices in smart buildings or medical monitoring systems [2], [5].In this range, the modified design demonstrates a good performance as well as able to work reliably. Additional refinements are possible through slot geometry optimization, placement of feeds, and material choice to realize a more repeatable and predictable wideband response [3], [6], [7]. This paper identifies the possibilities and constraints of fundamental geometric alterations in the design of microstrip patch antennaas that will serve as a good basis to continue the development of efficient, stable and compact wireless solutions in the future.

Acknowledgment

The authors thank professor Aman Kumari Dahiya of the Department of Electronics and Communication Engineering at Maharaja Surajmal Institute of Technology, New Delhi, India, for guidance and support.

References

[1] D. M. Pozar, *Microwave Engineering*, 4th ed. Hoboken, NJ, USA: Wiley, 2012.

[2] M. Meena and P. Kannan, "Analysis of microstrip patch antenna for four different shapes and substrates," *ICTACT Journal on Microelectronics*, vol. 4, no. 1, pp. 519–530, Apr. 2018, doi: 10.21917/ijme.2018.0092.

[3] A. Arora, A. Rana, A. Yadav, and R. Yadava, "Design of microstrip patch antenna at 2.4 GHz for Wi-Fi and Bluetooth applications," *Journal of Physics: Conference Series*, vol. 1921, no. 1, pp. 012023, 2021.

[4] B. Mishra, R. Kumar Verma, N. Yashwanth, and R. Kumar Singh, "A review on microstrip patch antenna parameters of different geometry and bandwidth modification techniques," *International Journal of Microwave and Wireless Technologies*, vol. 14, no. 5, pp. 652–673, 2021.

[5] B. Herdiana, "Comparative Analysis of Multi-Substrate Materials in Microstrip Antenna Models for Broadband Wireless Communications," *International Journal of Electronics and Communications Systems*, vol. 2, no. 1, pp. 1–7, 2022.

[6] S. Alam, I. Surjati, and T. Teguh Firmansyah, "Bandwidth Enhancement of Square Microstrip Antennas Using Dual Feed Line Techniques," *International Journal of Electrical and Electronic Engineering & Telecommunications*, pp. 60–65, 2021.

[7] D. Tumakov, V. Mikhailovich, A. Lekomtsev, and A. Korolev, "Miniaturization of a Koch-Type Fractal Antenna for Wi-Fi Applications," *Fractal and Fractional*, vol. 4, no. 3, pp. 1–12, 2020.

[8] T. Nahar and S. Rawat, "Survey of various bandwidth modification techniques used for 5G antennas," *International Journal of Microwave and Wireless Technologies*, vol. 13, no. 6, pp. 511–523, 2021.

Advanced Technologies with Revolutionizing Healthcare

JYOTI JAIN
(ELECTRICAL & ELECTRONICS ENGINEERING)
(MAHARAJA SURAJMAL INSTITUTE OF TECHNOLOGY)
(NEW DELHI , INDIA)
jyotieee@msit.in

SUMIT ROY
(ELECTRICAL & ELECTRONICS ENGINEERING)
(MAHARAJA SURAJMAL INSTITUTE OF TECHNOLOGY)
(NEW DELHI , INDIA)
sumitroymsit@gmail.com

MANIK MATTOO
(ELECTRONICS & COMMUNICATION ENGINEERING)
(MAHARAJA SURAJMAL INSTITUTE OF TECHNOLOGY)
(NEW DELHI , INDIA)
Mattoomanik9@gmail.com

Abstract-As technology is changing, the healthcare sector is also experiencing changes. Some of these changes have transformed the healthcare sector through revolutionary developments, making healthcare today more affordable and highly accurate. In this revolutionary transformation, certain technologies have played a major role in changing the healthcare sector. Laser technology has established itself as a revolutionary change in healthcare because some specific types of lasers, AI, robotic surgery, and particle types of frequency in different field beams are helpful in treating some specific types of diseases. Earlier, when surgeries were performed, their accuracy was low, but modern technology, with the help of robotic surgery, has made the accuracy of operations very high, making work easier. Artificial intelligence is also being used in healthcare for report analysis, data analysis, and new report generation. All of these have brought revolutionary changes in healthcare, making it accessible to everyone with the help of these technologies .Keywords :- Laser, Light, Artificial Intelligence, disease, Frequencies, Gate Control Theory, Robotic surgeries .

Introduction

Healthcare is an essential requirement for human life, as it is a process through which we attempt to prevent diseases, find cures, and provide proper treatment. Every disease can have different methods of treatment, and for the same illness, multiple technologies may be applied to achieve better results. In India, healthcare practices vary greatly from place to place, which shows that treatment methods are not always uniform and can change depending on the region and available resources. Due to this diversity, there is a strong need for engineers who can bring innovation and develop new techniques to make healthcare methods more effective, affordable, and accessible. The goal is to design technologies that allow treatment without requiring patients to travel long distances, spend too much money, or face discomfort, thereby ensuring comfort and efficiency in healthcare delivery. Just like in education and other sectors, the role of technology is rapidly growing in healthcare, and in today's technical era, engineers play a vital role in designing systems and tools that contribute significantly to this field. For example, ultrasonic waves are used in diagnostic imaging, laser sensors are applied in surgeries, and specific frequencies are designed to detect or treat medical conditions. In this sequence, Artificial Intelligence (AI) has also emerged as a highly significant contributor in the field of healthcare. As technology continues to advance, AI-based systems are increasingly being integrated into healthcare services. AI is used for analyzing medical reports, identifying new patterns and

symptoms, and making accurate predictions in research studies. With the help of AI, it becomes possible to estimate the long-term impact of a drug—whether five years or even a hundred years into the future—and understand how it may affect treatment outcomes. Thus, through various engineering applications and advanced technologies like AI, healthcare systems are being transformed to enable early detection of diseases, improve treatment efficiency, and enhance the quality of medical research, ultimately leading to better lifestyles for patients and contributing to the building of a healthier nation.

Lasers in Modern Healthcare

LASER stands for *Light Intensification by Triggered Emission of Radiation*. In the middle of the 20th century, laser technology emerged as an effective generation technology. With the continuous efforts of engineers and scientists, advanced lasers were developed that now work in various applications such as instrument controls, machines, and especially in medical fields. Specific types of lasers are widely used in healthcare industries.

With laser treatment, [1] the procedure has become very simple and less painful. Medical fields such as ophthalmology, urology, gynecology, dermatology, phlebology, and proctology use lasers extensively. In ophthalmology, excimer lasers and diode lasers are used. [2] Different fields make use of different lasers such as diode lasers, CO_2 lasers, RF lasers, chemical lasers, and green light lasers, each having different penetration and absorption rates depending on the wavelength.

The excimer laser is used to reduce the dependency on contact lenses and glasses by removing thin layers of corneal tissue. At around 308 nanometers, this laser has specific penetration and absorption properties that make it useful in elastic and plastic surgeries. For example, excess tissue near the eyes can be removed to correct skin folds or sagging caused by aging, improper stretching, or excess sleep. In tumor surgeries, lasers are applied to cut or remove abnormal tissues. Tumors may be cancerous or non-cancerous, but when they form inside the eye or other organs, diode lasers are applied. A 980-nanometer diode laser is used since it has no absorption in water and limited absorption in blood, allowing precise removal of excess tissue. This process is also applied in benign prostatic hyperplasia surgeries.

In partial nephrectomy, where excess tissue forms in the outer or inner portion of the kidney, diode lasers of 980 nm and 1470 nm are used. Different absorption properties make them suitable for cutting excess skin or muscle. In other fields, lasers such as 2100 nm fiberscope lasers or 532 nm green light lasers are tested, though limitations exist, such as improper absorption in blood, making them less effective in certain surgeries.

[3] In gynecology, fiber lasers are used for removing abnormal cells in the cervix or vagina caused by HPV, which often forms as extra skin due to infections or sexual activities. By firing fiber lasers, excess tissue is cut down in procedures like colposcopy. Surgeries like hysteroscopy and laparoscopic treatment of endometriosis also make use of diode lasers for cutting abnormal tissue in the ovaries or fallopian tubes. Additionally, in designer rejuvenation procedures, diode lasers are inserted into the vaginal path to tighten the folds and restore shape according to required size. Thus, [4] different types of lasers—such as excimer, diode, CO_2, RF, and fiber lasers—play a critical role in modern healthcare. They enable advanced surgeries in ophthalmology, dermatology, oncology, gynecology, nephrology, and cosmetic surgery, making treatments quicker, more efficient, and less painful, all due to continuous engineering and scientific innovation.

AI Innovation in Healthcare

As we know, this is the era of Artificial Intelligence (AI), and it is required in almost every field to make tasks easier. AI is one of the greatest achievements of the modern world and has transformed many areas of life, including the medical and healthcare sector. [5] The use of AI in healthcare helps in early disease detection, improving customer facilities, and ensuring the efficient use of medicines – something that was not possible a hundred years ago. With the help of AI, medical devices can be made more accurate, chemical reactions can be simulated, and complex calculations can be performed. AI can also enhance customer convenience

through services like automated call systems, better report analysis, and precise data preparation, making healthcare services more effective and accessible. [6] Another vital application of AI is in early disease detection. Smart devices like smartwatches and health bands monitor heart rate, activity levels, and blood pressure, which can help in preventing heart attacks and other diseases. These devices collect health-related data, analyze it, and provide early warnings.

AI is also being applied in surgical decision-making. AI-powered surgical robots can perform operations faster, more efficiently, and with higher precision, resulting in reduced costs, less pain, and quicker recovery. With advancements in machine learning, AI systems are becoming capable of learning from data, gaining new skills, and opening new opportunities in healthcare.

AI is further being explored in the field of age reversal, where old cells can be reprogrammed into active, younger cells, marking a revolutionary discovery. Research projects are working on reprogramming cells and genes, and with AI's help, new types of proteins have been designed successfully. Such breakthroughs hold the potential to reverse aging in the future, which could be one of the most transformative changes in modern science.

[7] [8] Nobel Prize-winning scientists researched **Yamanaka Factors** and conducted experiments on rats for reverse aging techniques, which was considered a huge breakthrough. This showed that aging happens due to multiple reasons, such as certain proteins damaging sensitive cells, accumulation of infections, and unhealthy lifestyles. While medicines and drugs already play a role in treatment, AI has made it possible to detect these proteins and their effects more accurately. What once required 10–20 years of trials and research can now be accelerated with the help of AI. Thus, AI is playing a significant role in improving lifestyle, preventing diseases, ensuring safety, and pushing forward safe age-reversing technologies

Robotic Revolution in Surgery

Operation is a process in which we try to correct something, such as when an injury occurs, when any organ or part of the body is damaged, or when there is unwanted growth inside the body, we take the help of surgery to treat it. Surgery is not a new practice; it has continuously evolved over time. As technology and working techniques have improved, the methods of performing operations have also advanced. [9] In ancient times, surgeries were performed with very basic techniques, but with modern innovations the procedures have become far more precise. Earlier, when operations were performed, a cut was made, and the depth and size of the cut depended entirely on the doctor's experience. There is a saying in India: *"The bigger the doctor, the smaller the cut."* This meant that an experienced doctor knew exactly how much incision was required to achieve maximum efficiency and quick recovery. In the past, many operations were not successful, but with the rapid development of technology today, surgical methods have changed drastically. In modern times, robotic technology is being used in many surgeries, which makes the procedure much more efficient and accurate. Robotic surgery is growing rapidly not only in India but across the world. [11] [10] It has many advantages, such as significantly reduced blood loss, faster recovery, and much less pain compared to earlier surgeries. Unlike older methods, there is no need to cut through bones unnecessarily, and accuracy is extremely high. For example, in heart surgery, robotic arms can target the exact portion of the heart without requiring additional incisions. Patients recover faster because the surgery is minimally invasive. Initially, robotic surgery was considered costly, but when we look at the overall cost, including shorter recovery time and efficiency, it often balances out with or even reduces compared to traditional methods. The robotic arms used in surgery can move very freely, almost like an artist painting with full 360-degree motion, which allows doctors to operate with much higher precision. The doctor controls the robot through a console, just like controlling a game, sitting comfortably without stress, though it requires highly trained specialists. Trust in robotic surgery is steadily increasing, and its demand will rise rapidly in the future. [12] Robotic surgeries are already being used successfully in eye operations, heart surgeries, brain surgeries, nasal clot removals, treatments of private parts, and

even for removing internal infections or unwanted growths. In India, many companies are working in this field, among which Apollo Hospitals is a leading name. Both the Indian government and private organizations are investing heavily in robotic surgery, and with the integration of AI, the growth has accelerated even more. Robotic surgery is not just the present but also the future, and both patients and doctors will have to adapt to it

Frequencies working in healthcare

We produce different kinds of sounds for different purposes, such as shouting when we scold someone and speaking softly when we want to talk politely; [13] in the same way, frequencies are generated and used in different applications. High frequencies are mainly used in transmissions, while low frequencies are often applied in the medical field. In medicine, frequency is used for pain relief based on the **Gate Control Theory**, which explains that the pain signals traveling from the body to the brain can be modified by adding another frequency, so the brain perceives everything as normal and pain is reduced. Many pain relief devices are built on this principle, one of which is the **vibrator**, which generates different kinds of frequencies used in various ways. As a fitness device, it can be adjusted to produce specific frequencies for exercises with multiple benefits, such as muscle strengthening and improved blood flow. [14] In medical devices, particular frequencies can be applied to relieve pain in targeted areas, including menstrual pain, where several companies are developing frequency-based devices to provide relief during periods. Vibrator devices can also help diabetic patients by regulating blood pulses, improving circulation, increasing bone density, enhancing lymphatic flow, and supporting cardiovascular therapy. Essentially, these instruments aim to replace or complement physiotherapy with therapeutic frequency-based solutions at lower costs. [15] In the healthcare sector, frequency plays a major role, and creating such devices is the responsibility of engineers, who contribute significantly to improving our lives. Other important applications of frequency include **vagal nerve stimulators, auricular stimulators, deep brain stimulators, and external neuromodulators**. In addition, pain research can be explored in different aspects such as **clinical management approaches, neurobiological mechanisms, psychological and social dimensions, and specific pain conditions**. The purpose of such research may include **academic studies, treatment planning, or understanding mechanisms for better therapies**, while focusing on **specific populations or contexts**. Depending on the scope, one may aim for a **comprehensive overview, detailed focus on mechanisms or treatments, or insights into recent advances**.

Conclusion

whenever we introduce a new technology, it plays a major role in transforming society. A time will come when these technologies will dominate the healthcare sector, and among them, Artificial Intelligence (AI) will be the most influential. AI has the potential to bring about a revolutionary transformation in healthcare. Along with AI, laser technology and robotic surgery, when coordinated together, will provide better healthcare solutions that can reduce mortality rates, extend life cycles, and improve the overall quality of human life. These are not just technologies—they are sources of change. This change will truly begin when we adopt these technologies in our daily healthcare practices and trust them for treating diseases.

References

[1] K. Thomsen, R. Jalaboi, , "Physician level assessment of hirsute women and of their eligibility for laser treatment with deep learning," *Lasers in Surgery and Medicine*, vol. 57, no. 3, pp. 80–87, 2025. doi: 10.1002/lsm.23843..

[2] . J. Constable and A. S. M. Lim, *Laser, Its Clinical Uses in Eye Diseases*, 2nd ed. Edinburgh, UK: Churchill Livingstone, 1990, pp. 74–76.

[3] A. Raj, A. Singla, and S. Sidana, "Preventive and therapeutic strategies via health care delivery system to minimize sight-threatening diabetic retinopathy: a narrative

review," *Current Diabetes Reports*, vol. 25, art. 36, 2025. doi: 10.1007/s11892-025-01591-5

[14] M. A. S. Kamal, H. S. J. Lee, and M. S. Hossain, "Synthetic data generation in healthcare: A scoping review of reviews," *Comput. Biol. Med.*, vol. 157, p. 89, Oct. 2025.

[15] R. S. Taylor and M. J. Thompson, "Digital Health Transformation in Clinical Practice: Challenges and Opportunities," *Can. J. Hosp. Pharm.*, vol. 76, no. 4, pp. 298–310, Aug. 2025.

[4] O. H. Arican, "Optimal tugboat placement in narrow straits for ship safety: a Fuzzy Delphi and P-Median approach, as exemplified by the Dardanelles Strait," *Ships Offshore Struct.*, vol. 20, no. 1, pp. 7–13, May 2025.

[5] O. Panahi, "The Future of Healthcare: AI, Public Health and the Digital Revolution," *Medi. Clin. Case Rep. J.*, vol. 3, no. 1, pp. 2 , Mar. 2025

[6] Lekadir, M., et al., "FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare," *BMJ*, vol. 22, pp.11-17, Feb. 2025.

[7] A. Bracken, C. Reilly, A. Feeley, E. Sheehan, K. Merghani, and I. Feeley, "Artificial Intelligence (AI)–Powered Documentation Systems in Healthcare: Systematic Review," *J. Med. Syst.*, vol. 49, no. 28, pp. 1–13, Feb. 2025.

[8] O. Panahi, "Health in the Age of AI: A Family and Community Focus," Archives of Community and Family Medicine, vol. 10, no. 1, pp. 1–5, May 2025

[9] O. Panahi, "Innovations in Surgical Healthcare: The Future of Medicine," *J. Med. Syst.*, vol. 8, no. 3, pp.3-7, May 2025.

[10] J. N. K. Wah, "The rise of robotics and AI-assisted surgery in modern healthcare," *J. Robotic Surg.*, vol. 11, no. 31, pp. 2-7, Jun. 2025.

[11] M. D. Smith and J. K. Lee, "The far-reaching impact of robotic-assisted surgery on healthcare," *Health Manag. Technol.*, vol. 47, no. 2, pp. 35–42, Jun. 2025.

[12] Y. A. Alhamaid et al., "Healthcare students' attitude toward robotic surgery: A cross-sectional survey," *Ann. Afr. Med.*, vol. 24, no. 2, pp.11-12 , May 2025

[13] S. K. K. S. Reddy and S. S. Channappayya, "Stationary wavelet-based image watermarking for e-healthcare," *J. Electron. Imaging*, vol. 32, no. 6, pp. 1–9, Nov. 2023

AI and Web3: Integrating Autonomous Agents for Next-Gen Decentralized Finance

Ved Mohan
Student, Dept. of ECE,Maharaja Surajmal Institute of Technology
New Delhi, India
ivedmohan@gmail.com

Meena Rao
Associate Professor, Dept. of ECE, Maharaja Surajmal Institute of Technology
New Delhi, India
meenarao@msit.in

Parul Chaudhary
Assistant Professor, Dept. of CSE, Maharaja Surajmal Institute of Technology
New Delhi, India
parul@msit.in

Decentralized finance (DeFi) combined with AI agents or LLM models represents a transformational shift in the evolution of blockchain-based financial systems. This work explores the integration of large language model-powered AI agents with Web3 infrastructures, leveraging frameworks like LangChain alongside scalable Layer-2 blockchains such as StarkNet. Focusing on autonomous portfolio management, multi-protocol yield optimisation, and secure wallet handling enabled by account abstraction, this paper demonstrates how AI and Web3 synergise to reduce complexity, increase user accessibility, and enhance transaction efficiency. Initial evaluations show promising improvements in transactional throughput and yield performance through intelligent agent orchestration across protocols, including StrkFarm, EnduFi, and AVNU.

Keywords— Decentralized Finance, Autonomous Agents, Blockchain, AI Trading

I. INTRODUCTION

Decentralized finance (DeFi) has emerged as a revolutionary paradigm that leverages blockchain technology to enable transparent, permissionlessfinancial services without traditional intermediaries [1]. The DeFi ecosystem has grown exponentially, with total value locked (TVL) reaching unprecedented levels across hundreds of protocols. However, this growth has introduced significant complexity for individual users attempting to optimize their financial strategies across multiple platforms.

The current DeFi landscape presents several critical challenges [2]:

- Manual portfolio management across diverse protocols requires extensive technical knowledge and continuous monitoring.
- Yield farming opportunities are fragmented across multiple platforms with varying risk profiles.
- Optimal capital allocation strategies require complex mathematical modeling and real-time market analysis. and
- Transaction costs and timing optimisation across different protocols demand sophisticated understanding of blockchain mechanics.

Traditional approaches to DeFi portfolio management rely heavily on manual intervention, leading to suboptimal returns, missed opportunities, and increased exposure to market volatility. Users must constantly monitor multiple protocols, analyze yield opportunities, assess risk factors, and execute complex multi-step transactions—a process that is both time-intensive and error-prone.

Due to flaws in risk prediction, fraud detection, and dynamic decision-making capacities, consumers still face significant risks even with the transparency provided by blockchain technology. Although blockchain guarantees the immutability and accessibility of transaction records, it does not by itself offer the intelligence required for real-time risk assessment or the detection of potential weaknesses like oracle manipulations and flash loan attacks. These flaws present serious dangers and have sparked questions about the safety, scalability, and dependability of DeFi platforms.

These challenges are addressed by combining large language model (LLM)-powered autonomous agents with the scalable StarkNet Layer-2 blockchain infrastructure [3]. The platform automates complex DeFi operations through intelligent AI agents that can analyze market conditions, execute sophisticated trading strategies, and optimize yield farming across multiple protocols simultaneously [4]. This approach promises improved returns, reduced operational complexity, and enhanced risk management for DeFi participants.

The objective of the paper is to provide an insight into a thorough data-driven model that strategically combines blockchain technology and artificial intelligence to address the various issues that current DeFi systems encounter.

II. RELATED WORK

DeFi started out with simple loan platforms like Compound and MakerDAO, as well as token exchanges. The ecosystem grew over time to incorporate liquidity aggregators, insurance protocols, synthetic asset platforms, and decentralised exchanges (DEXs). DeFi's capabilities have been expanded by innovations like automated market makers (AMMs), flash loans, yield farming, and liquidity pools, which allow users to engage with a wide range of financial products without the need for conventional banking infrastructure.

DeFi systems' stability, security, and long-term adoption are hampered by a number of significant issues that accompany this breakthrough. Smart contracts are often the target of well-known exploits, making security flaws one of the most common problems. Due to their reliance on unaudited or insufficiently tested code, several DeFi protocols are

vulnerable to front-running, sandwich attacks, oracle manipulation, and re-entrancy attacks. Since the majority of DeFi apps rely on layer-1 blockchains, such as Ethereum,[11] which have high transaction fees and limited throughput during times of network congestion, scalability is yet another major worry. These limitations significantly limit DeFi platforms' usefulness and inclusivity, especially for consumers interacting with microtransactions.

Blockchain technology, which enables decentralised, unchangeable, and transparent data recording across dispersed networks, is the cornerstone of DeFi's basis. Blockchain has proven to be very useful in financial systems for improving auditability, cutting operating costs, speeding up settlement times, and preserving data integrity.

Notwithstanding these benefits, the deterministic character of blockchain technology and its shortcomings in managing real-time data and adaptive decision-making have brought attention to the necessity for supplementary solutions. Here, artificial intelligence (AI) shows itself as a game-changing facilitator in the fintech and, more and more, DeFi sectors. In fields including credit scoring, fraud detection, algorithmic trading, risk management, and customer behaviour analytics, artificial intelligence (AI) has a well-established place in financial technology.

For the AI agent component, LangChain has become a cornerstone framework for building modular, extensible agents in blockchain contexts. LangChain's ecosystem, including LangGraph for stateful multi-agent workflows and LangSmith for observability and debugging, facilitates the orchestration of LLMs with external tools like blockchain loaders for querying on-chain data. Recent applications integrate LangChain with DeFi for real-time decision-making, such as agents that fetch market data, simulate trades, and execute via smart contracts. In StarkNet-specific setups, LangChain agents have been used to create verifiable AI for yield optimization, combining off-chain reasoning with on-chain verification. Community efforts, including tutorials on LangChain for blockchain AI agents, emphasize its role in handling intents like voice-parsed commands for autonomous trading.

To summarize, limitations of DeFi system can be addressed by integrating blockchain and AI technologies. Blockchain provides a transparent and tamper resistant system whereas AI offers complex data analysis and real time adaptability. This merger of AI with blockchain can lead to a more resilient and efficient financial system.[8]

III. System Architecture

Figure 1 presents a DeFi modular three tier architecture.

Fig.1 DeFi Modular Three-tier Architecture

DeFi uses a modular three-tier architecture that comprises of the following three components:

1. Frontend: Next.js application providing user interfaces for monitoring and configuration.
2. Backend: Express.js server coordinating AI agents, API endpoints, and blockchain interactions.
3. Blockchain: StarkNet supports fast, low-cost transactions and account abstraction enabling secure agent wallet management.

AI agents are orchestrated via LangChain, using Claude Sonnet for natural language processing and complex decision algorithms. Key agents manage trading strategies, portfolio balancing, risk control, and protocol communications. [3]

IV. Methodology

1. Frontend

A Next.js application serves as the user interface, providing intuitive dashboards for monitoring portfolio health, viewing AI agent actions, and configuring risk and investment preferences. Reactive components update users in real-time based on blockchain events and AI insights, ensuring transparency and control.

2. Backend

An Express.js server orchestrates AI agents and API endpoints, handling communication between frontend requests and the blockchain network. LangChain framework manages AI workflows, integrating Claude 4 Sonnet for natural language processing and complex trading strategy executions. The backend also implements security layers such as input validation, rate limiting, and error monitoring.

3. Blockchain

StarkNet blockchain serves as the decentralized execution layer supporting this system, providing low-cost, high-throughput transactions secured with zk-STARK proofs. Its native account abstraction enables all user accounts to be programmable smart contracts, allowing AI agents to act autonomously on behalf of users while maintaining robust security guarantees such as multi-signature authorization, social recovery, and custom transaction logic. This abstraction simplifies user onboarding and enables complex DeFi transaction flows without compromising decentralization.[12]

A. AI-Enabled Trading Systems

Recent research in artificial intelligence for financial markets has explored various approaches to automated trading and portfolio management. Machine learning techniques, particularly reinforcement learning and deep neural networks, have shown promise in traditional financial markets. However, the unique characteristics of DeFi markets—including 24/7 operation, high volatility, and complex protocol interactions—require specialized

approaches. Figure 2 presents an AI Decision pipeline and dataflow model.

Fig.2 DeFi AI Decision Pipeline and Dataflow

Large language models (LLMs) have emerged as powerful tools for financial analysis and decision-making. Recent studies have demonstrated the effectiveness of LLM-based agents in understanding complex financial concepts, analyzing market sentiment, and generating trading strategies. The integration of LLMs with traditional quantitative methods has shown particular promise in handling the multi-faceted nature of DeFi protocols.

B. Blockchain and DeFi Infrastructure

The evolution of blockchain technology has enabled increasingly sophisticated DeFi applications. Layer-2 solutions, particularly StarkNet, have addressed scalability limitations while maintaining security guarantees. StarkNet's use of zero-knowledge proofs (zk-STARKs) enables high-throughput, low-cost transactions essential for frequent portfolio rebalancing and yield optimization strategies.[12]

Account abstraction represents a significant advancement in blockchain user experience, enabling programmable accounts that can execute complex logic without direct user intervention. This technology is particularly relevant for autonomous trading systems, as it allows AI agents to manage funds securely while maintaining user control over high-level parameters.

C. Multi-Protocol Integration

Figure 3 shows how an agent interfaces with StrkFarm, and other protocols.

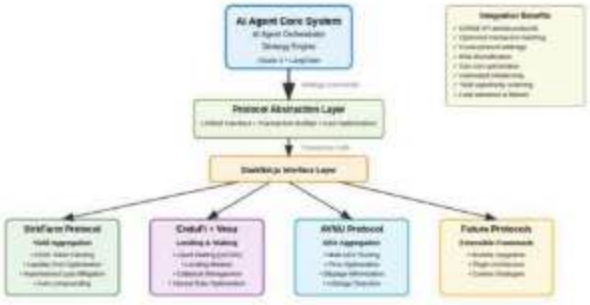


Fig.3 Agent interfaces with StrkFarm, EnduFi, and AVNU protocols.

The backbone of the Defi platform's innovation is its seamless integration across diverse DeFi protocols on StarkNet, leveraging the power of composability in decentralized finance to maximize yield and manage risk dynamically.

Key Protocols integrated include:

StrkFarm: A leading yield farming protocol on StarkNet that enables users to stake liquidity pool tokens and earn yield rewards. Our AI agents automate the staking and unstaking processes based on real-time yield data, minimizing impermanent loss and optimizing reinvestment schedules. By continuously scanning the changing APYs and liquidity depths, the AI dynamically adjusts positions to capitalize on the most lucrative farming pools.[13]

EnduFi: A next-gen lending and borrowing protocol focusing on capital efficiency and flexible collateral management. The AI-enhanced platform analyzes interest rate movements and borrower behavior to make sophisticated lending decisions. It dynamically reallocates assets between lending and borrowing positions to maximize returns while minimizing exposure to liquidation risks.[15]

AVNU Aggregator: A decentralized exchange aggregator that sources liquidity from multiple AMMs and order-book protocols within StarkNet. Our AI agents utilize AVNU to route trades with optimal gas efficiency and minimal slippage, enhancing trading performance especially in volatile market conditions. This facilitates quick portfolio rebalancing and arbitrage opportunities across StarkNet's ecosystem.[14]

Additionally, the platform envisions integration with cross-chain bridges and oracles to access real-time price feeds and expand liquidity across multiple Layer-1 and Layer-2 chains. This multi-protocol and multi-chain strategy empowers the AI agents to perform complex arbitrage, risk-adjusted allocations, and secure transactions—driving superior yield generation while mitigating systemic risks inherent in isolated protocols.

V. Conclusion and Future scope

This work demonstrates that integrating intelligent autonomous agents with a robust, scalable Web3 infrastructure such as StarkNet can substantially advance the decentralized finance space. The unique synergy of AI-driven decision making and blockchain-enabled transparency tackles the core challenges of multi-protocol management complexity, real-time responsiveness, and secure transaction execution.

By implementing modular architecture spanning frontend interfaces, AI orchestrated backend services, and decentralized Layer-2 execution with account abstraction, the platform provides a powerful blueprint for next-generation autonomous DeFi applications. Initial results highlight enhanced portfolio yields, improved asset allocation, and streamlined user experiences.

Looking forward, embracing decentralized AI computation, multi-chain interoperability, and privacy-respecting methods

will be critical for driving widespread adoption. Additionally, ensuring automated regulatory compliance and embedding AI within governance frameworks will help create a resilient, trustworthy, and inclusive DeFi ecosystem.

In summary, this work charts a viable path toward autonomous, intelligent, and composable financial systems—heralding a paradigm where Web3 and AI collaboratively deliver optimized decentralized finance for users worldwide.

In future work, transitioning from off-chain AI computation to on-chain or decentralized inference can be looked into as it can reduce dependency on centralized AI services. Exploring zero-knowledge proofs and Trusted Execution Environments (TEEs) for secure AI model execution is also a promising research direction.

Also, developing universal AI agent protocols capable of executing strategies across various heterogeneous environments is a key challenge and can be further explored.[7]

REFERENCES

[1] Auer, Raphael, Bernhard Haslhofer, Stefan Kitzler, Pietro Saggese, and Friedhelm Victor. "The technology of decentralized finance (DeFi)." Digital Finance 6, no. 1 , 2024, pp. 55-95.

[2] Aquilina, Matteo, Jon Frost, and Andreas Schrimpf. "Decentralized finance (DeFi): a functional approach." Journal of Financial Regulation 10, no. 1, 2024, pp. 1-27.

[3] Mothukuri, Viraaji, Reza M. Parizi, James L. Massa, and Abbas Yazdinejad. "An ai multi-model approach to defi project trust scoring and security." In 2024 IEEE International Conference on Blockchain (Blockchain), pp. 19-28.

[4] Palaiokrassas, Georgios, Sandro Scherrers, Iason Ofeidis, and Leandros Tassiulas. "Leveraging machine learning for multichain defi fraud detection." In 2024 IEEE International Conference on Blockchain and Cryptocurrency (ICBC), pp. 678-680. IEEE, 2024.

[5] Jia, Fernando, Jade Zheng, and Florence Li. "Decentralized intelligence in gamefi: Embodied ai agents and the convergence of defi and virtual ecosystems." arXiv preprint arXiv:2412.18601 , 2024

[6] Ante, Lennart. "Autonomous AI Agents in Decentralized Finance: Market Dynamics, Application Areas, and Theoretical Implications." Application Areas, and Theoretical Implications (December 14, 2024)

[7] Karim, Md Monjurul, Dong Hoang Van, Sangeen Khan, Qiang Qu, and Yaroslav Kholodov. "Ai agents meet blockchain: A survey on secure and scalable collaboration for multi-agents." Future Internet 17, 2025, Vol. 57, no. 2

[8] Uddin, Mueen, Muath Obaidat, Selvakumar Manickam, Shams Ul Arfeen Laghari, Abdulhalim Dandoush, Hidayat Ullah, and Syed Sajid Ullah. "Exploring the convergence of Metaverse, Blockchain, and AI: A comprehensive survey of enabling technologies, applications, challenges, and future directions." Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 14, no. 6, 2024 : e1556.

[9] Walters, Shaw, Sam Gao, Shakker Nerd, Feng Da, Warren Williams, Ting-Chien Meng, Amie Chow et al. "Eliza: A web3 friendly ai agent operating system." arXiv preprint arXiv:2501.06781,2025

[10] Nakamoto, Satoshi. "Bitcoin: A Peer-to-Peer Electronic Cash System." bitcoin.org, 2008

[11] Buterin, Vitalik. "Ethereum: A Next-Generation Smart Contract and Decentralized Application Platform." ethereum.org, 2014.

[12] StarkWare Industries. "What is Starknet?" starknet.io, 2024.

[13] STRKFarm Team. "STRKFarm: Best DeFi Yields on Starknet." strkfarm.xyz, 2024.

[14] AVNU Team. "AVNU: The Liquidity Infrastructure for L2s." avnu.fi, 2023.

[15] EnduFi Protocol. "A Protocol for Capital Efficiency in Lending and Borrowing." DeFi Protocol Specification, 2024.

Extreme Lossy SVG Compression for Bandwidth-Critical Applications: Trading Visual Fidelity for 99% Size Reduction in AI-Generated Content

Prakhar Chandra
Department of Computer Science and Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
prakharchandra1803@gmail.com

Pulkit Kapur
Department of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
pulkitkapur100@gmail.com

Jaspreet Singh
Department of Information Technology
Maharaja Surajmal Institute of Technology
New Delhi, India
jaspreet.jsk.kohli@gmail.com

***Abstract*—This paper presents an extreme lossy compression pipeline for scalable vector graphics (SVG) that prioritizes file size reduction over visual fidelity for bandwidth-critical applications. Motivated by challenges encountered in the Kaggle "Drawing with LLMs" competition where massive datasets of AI-generated images required efficient storage and transmission, our method achieves unprecedented 99.01% compression ratios while acknowledging significant quality degradation. Through a five-stage pipeline combining perceptual color quantization, hierarchical contour extraction, and adaptive polygon simplification, we reduce AI-generated images from 287KB to 1.9KB on average. Quality metrics reveal substantial visual degradation: SSIM of 0.233 ± 0.072, CIEDE2000 color distance of 30.9 ± 3.0, and LPIPS of 0.624 ± 0.054. This method is specifically designed for applications where extreme bandwidth constraints, storage limitations, or transmission costs outweigh visual quality requirements, such as IoT devices, edge computing thumbnails, content previews, and emergency communication systems. The approach is not suitable for applications requiring accurate color reproduction, fine detail preservation, or professional visual quality.**

***Index Terms*—Extreme lossy compression, SVG optimization, bandwidth-critical applications, AI-generated imagery, perceptual trade-offs, edge computing, IoT content delivery**

I. Introduction

The proliferation of AI-generated visual content has created unique challenges in bandwidth-constrained environments where traditional quality-preserving compression methods prove inadequate. This work emerged from practical constraints encountered during the Kaggle "Drawing with LLMs" competition [1], where participants generated millions of AI-created images requiring efficient storage, transmission, and processing within strict computational budgets.

Unlike natural imagery, AI-generated content from modern diffusion models exhibits distinctive characteristics that can be exploited for extreme compression: structured color palettes, geometric regularity, and spatial organization patterns that emerge from the training process [4], [5]. However, existing SVG compression techniques, designed primarily for lossless or near-lossless optimization, achieve only modest 30-60% size reductions [2], [3] and fail to address scenarios where extreme bandwidth constraints necessitate aggressive quality trade-offs.

A. Problem Definition and Motivation

This research addresses a specific class of applications where extreme file size reduction takes precedence over visual fidelity:

- **IoT and Edge Computing**: Devices with severe bandwidth and storage constraints requiring visual content transmission
- **Emergency Communication Systems**: Networks with limited capacity needing to transmit visual information during crises
- **Content Previews and Thumbnails**: Applications where approximate visual representation suffices for user navigation
- **Batch Processing Environments**: Large-scale content pipelines where storage costs outweigh quality requirements
- **Archival Systems**: Long-term storage of AI-generated content where accessibility matters more than fidelity

Traditional compression metrics and quality standards become secondary considerations in these contexts, where the primary objective is achieving maximum size reduction while retaining basic visual structure and recognizability.

B. Honest Assessment of Quality Trade-offs

Our method achieves 99% compression ratios through aggressive quality degradation that renders it unsuitable for most conventional applications. The resulting images exhibit:

- Severe color quantization artifacts
- Significant loss of fine details and textures
- Simplified geometric representations
- Reduced color accuracy and saturation
- Loss of subtle gradients and smooth transitions

Critical Limitation: This approach is explicitly NOT recommended for applications requiring accurate color reproduction, professional visual quality, medical imaging, scientific visualization, print media, or any scenario where visual fidelity is essential.

C. Contributions

The main contributions of this work include:

1) **Extreme Compression Pipeline**: A specialized system designed for maximum size reduction in bandwidth-critical environments
2) **Application-Specific Optimization**: Algorithms tailored for AI-generated content characteristics
3) **Honest Quality Assessment**: Transparent evaluation of visual degradation using established perceptual metrics
4) **Use Case Classification**: Clear guidelines for appropriate and inappropriate applications
5) **Production-Ready Implementation**: Demonstrated feasibility for specialized deployment scenarios

II. Related Work and Background

A. Traditional SVG Compression Landscape

Existing SVG compression research has focused on preserving visual quality while achieving modest size reductions. SVGO [2], the most widely adopted tool, achieves 20-40% compression through lossless optimization techniques including redundant markup removal, path data optimization, and coordinate precision adjustments. Commercial solutions like Vecta.io [3] and SVGco.de extend these approaches with advanced geometric analysis, achieving up to 60% compression while maintaining perfect visual fidelity.

These methods, while effective for general-purpose applications, are fundamentally limited by their quality-preserving constraints and cannot address scenarios requiring extreme bandwidth reduction.

B. Lossy Image Compression in Constrained Environments

Research in extreme lossy compression has primarily focused on specific domains with unique constraints. Military and emergency communication systems have explored aggressive JPEG quantization for bandwidth-limited networks [6]. IoT applications have developed specialized compression schemes prioritizing transmission efficiency over quality [7].

However, these approaches typically target raster formats and do not leverage the scalable properties of vector graphics or the specific characteristics of AI-generated content.

C. Perceptual Quality Metrics and Limitations

Standard perceptual quality metrics including SSIM [8], CIEDE2000 [9], and LPIPS [10] were designed to evaluate compression methods operating within acceptable quality ranges. In extreme compression scenarios, these metrics serve primarily to quantify degradation rather than validate quality preservation.

SSIM values below 0.5 typically indicate severe structural degradation, while CIEDE2000 distances exceeding 10 represent perceptually significant color differences. Our method operates well beyond these conventional quality thresholds, necessitating alternative evaluation criteria based on use case requirements rather than perceptual fidelity.

D. AI-Generated Content Characteristics

Diffusion-based generative models produce imagery with distinctive properties exploitable for compression: reduced color complexity compared to natural photography, structured spatial organization, and predictable geometric patterns [12]. These characteristics enable more aggressive compression techniques than possible with natural imagery, though at the cost of visual quality degradation.

III. Methodology and System Architecture

A. Overall System Design and Pipeline Architecture

Our compression pipeline consists of five integrated stages designed for maximum size reduction of AI-generated bitmap inputs. The system prioritizes compression ratio over visual quality, employing aggressive optimization techniques that would be unacceptable in quality-preserving applications.

B. Stage 1: Input Preprocessing and Content Analysis

Input preprocessing analyzes image characteristics to optimize subsequent aggressive compression decisions. The system identifies expendable visual elements and prioritizes structural features essential for basic recognizability.

Algorithm 1 Input Preprocessing and Analysis

Require: Bitmap image I of size $W \times H$
Ensure: Preprocessed image I', content characteristics C
Color Distribution Analysis:
  ▷ Compute histogram in RGB and HSV spaces
  ▷ Calculate color complexity indicators
  ▷ Identify dominant color regions
Bilateral Filtering:
  ▷ Compute edge density map using Canny detection
  ▷ Select filter parameters based on edge density
  ▷ Apply adaptive bilateral filtering
Background Analysis:
  ▷ Create edge-based background mask
  ▷ Compute background region statistics
  ▷ Select optimal background color
Content Characteristics:
  ▷ Spatial frequency analysis using DCT
  ▷ Geometric complexity estimation
  ▷ Color palette preliminary analysis
return I', C

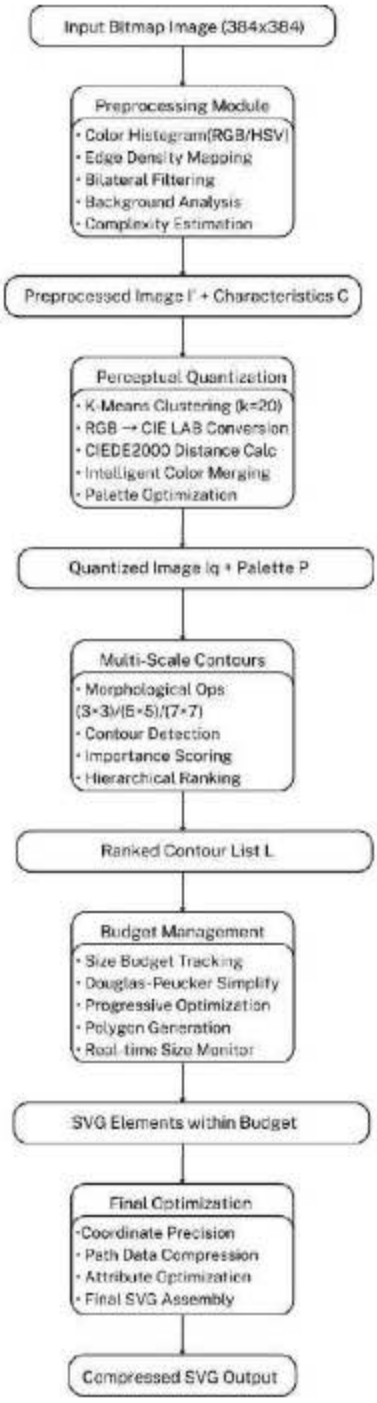


Fig. 1: Complete system architecture showing the five-stage compression pipeline: (1) Input preprocessing and analysis, (2) Perceptual color quantization with CIEDE2000 optimization, (3) Hierarchical contour extraction with multi-scale analysis, (4) Adaptive vectorization with importance weighting, and (5) Size optimization with budget management.

C. Stage 2: Aggressive Perceptual Color Quantization

Traditional color quantization preserves perceptual quality through conservative palette reduction. Our approach employs extreme quantization using CIEDE2000 perceptual distance metrics to achieve maximum color reduction while retaining basic color relationships.

The CIEDE2000 color distance is calculated as:

$$\Delta E_{00} = \sqrt{\left(\frac{\Delta L'}{k_L S_L}\right)^2 + \left(\frac{\Delta C'}{k_C S_C}\right)^2 + \left(\frac{\Delta H'}{k_H S_H}\right)^2 + R_T \frac{\Delta C'}{k_C S_C}\frac{\Delta H'}{k_H S_H}} \quad (1)$$

where $\Delta L'$, $\Delta C'$, and $\Delta H'$ represent lightness, chroma, and hue differences, respectively, and k_L, k_C, k_H are parametric factors.

Algorithm 2 Perceptual Color Quantization

Require: Preprocessed image I', initial cluster count k = 20
Ensure: Quantized image I_q, optimized palette P

Initial Clustering:
  ▷ Extract pixel data from input image
  ▷ Apply k-means clustering with k = 20 centers
  ▷ Compute cluster assignments and centers
Perceptual Distance Calculation:
  ▷ Convert cluster centers to CIE LAB color space
for each color pair (c_i, c_j) **do**
    Calculate CIEDE2000 distance $\delta E(c_i, c_j)$
    Store distance matrix for optimization
end for
Intelligent Merging:
  ▷ Sort color pairs by perceptual distance
for pairs with $\delta E(c_i, c_j) < 15$ **do**
    Merge colors using weighted average
    Update cluster assignments
    Recalculate affected distances
end for
Palette Optimization:
  ▷ Remove colors with < 1% pixel coverage
  ▷ Apply coverage-weighted color adjustment
  ▷ Generate final optimized palette P
return I_q, P

D. Stage 3: Hierarchical Contour Extraction and Analysis

Effective extreme compression requires intelligent selection of only the most essential geometric features. Our hierarchical approach sacrifices fine details and minor structural elements to preserve basic shape recognition.

The importance scoring function balances multiple factors:

$$I = A_{score} \times S_{score} \times F_{scale} + E_{gain} \quad (2)$$

where:

$$A_{score} = \log(\text{contour_area}) \quad (3)$$

$$S_{score} = \text{circularity} \times 0.3 + \text{convexity} \times 0.2 \quad (4)$$

$$F_{scale} = 1.0/(\text{scale index} + 1) \quad (5)$$

$$E_{gain} = \frac{\text{edge_density_within_contour}}{\text{mean_edge_density}} \quad (6)$$

Algorithm 3 Multi-Scale Contour Extraction

Require: Quantized image I_q, palette P
Ensure: Ranked contour list L with importance scores
for each color mask M_i in palette P **do**
  Initialize contour list $L_i = []$
  for each kernel size k ∈ {3, 5, 7} **do**
    Morphological Preprocessing:
      ▷ Apply opening operation with kernel k
      ▷ Apply closing operation with kernel k
      ▷ Generate cleaned mask M_{ik}
    Contour Extraction:
      ▷ Extract contours using RETR_CCOMP hierarchy
      ▷ Filter contours by minimum area (15 pixels)
      ▷ Store contour hierarchy information
    Feature Analysis:
      ▷ Calculate geometric properties
      ▷ Compute shape descriptors
      ▷ Analyze spatial relationships and nesting
    Importance Scoring: Using Equations (2)-(5)
    Add to contour list L_i with metadata
  end for
  Sort L_i by importance score (descending)
  Add L_i to global list L
end for
return L sorted by global importance ranking

Algorithm 4 Adaptive Vectorization with Budget Management

Require: Contour list L, size budget B (bytes)
Ensure: SVG representation S within budget B
Initialize: $S_{current} = S_{header}$, $C_{processed} = []$
for each contour c in L (sorted by importance) **do**
  Estimate representation cost: $C_{base} = f_{estimate}$(c.points)
  if $S_{current} + C_{base} > B$ **then**
    Progressive Simplification:
      Calculate P_{budget} using Equation (6)
      Calculate ϵ using Equation (7)
      $c_{simplified}$ = douglas_peucker(c, ϵ)
      $C_{final} = f_{estimate}(c_{simplified}.\text{points})$
    if $S_{current} + C_{final} > B$ **then**
      Skip contour (budget exhausted)
      Continue to next contour
    end if
  end if
  Generate SVG polygon element
  Update size tracking: $S_{current} += |\text{polygon}_{svg}|$
  $C_{processed}$.append(polygon_{svg})
  if $S_{current} \geq 0.95 \times B$ **then**
    Break (approach budget limit)
  end if
end for
Generate final SVG: S = assemble_svg($C_{processed}$, bg_{color})
return S

E. Stage 4: Adaptive Vectorization with Strict Budget Management

Size constraints in bandwidth-critical applications require absolute file size limits. Our pipeline implements aggressive polygon simplification that prioritizes size constraints over geometric accuracy.

Budget pressure is calculated as:

$$P_{budget} = \frac{S_{current}}{B} \tag{7}$$

The adaptive epsilon for Douglas-Peucker simplification becomes:

$$\epsilon = \epsilon_{base} \times (1 + P_{budget} \times 0.5) \tag{8}$$

F. Stage 5: Final Size Optimization

The final stage applies comprehensive optimization including coordinate precision reduction, path data compression, and attribute minimization. These techniques further sacrifice precision for size reduction.

G. Implementation and Computational Requirements

The pipeline is implemented in Python using OpenCV 4.7, scikit-image 0.19, and colormath 3.0. Processing time averages 2.5 minutes for 384×384 pixel images on standard hardware, with computational complexity of $O(n \log n)$ where n represents pixel count.

While processing time may seem excessive for real-time applications, the target use cases typically involve batch processing or offline compression where processing time is less critical than file size reduction.

IV. Experimental Design and Evaluation Framework

A. Dataset and Testing Methodology

Our evaluation uses 273 AI-generated images from the Kaggle "Drawing with LLMs" competition, spanning diverse artistic styles and complexity levels. This dataset represents realistic AI-generated content requiring extreme compression for storage and transmission efficiency.

TABLE I: Dataset Composition and Characteristics

Category	Count	Avg Colors	Edge Density	Complexity
Abstract/Geometric	68	12.4 ± 3.2	0.23 ± 0.08	2.1 ± 0.6
Natural Scenes	71	18.7 ± 4.1	0.41 ± 0.12	3.4 ± 0.8
Character/Portrait	67	15.2 ± 3.8	0.35 ± 0.10	2.9 ± 0.7
Technical/Architectural	67	10.8 ± 2.9	0.28 ± 0.09	2.3 ± 0.5

B. Evaluation Metrics and Quality Assessment

We employ standard perceptual quality metrics to quantify visual degradation rather than validate quality preservation. These metrics serve to characterize the extent of quality loss inherent in extreme compression approaches.

V. Results and Analysis

A. Compression Performance

Our method achieves consistent 99.01% compression across diverse content types, reducing average file sizes from 287KB to 1.88KB. This extreme compression comes at the cost of severe quality degradation, as evidenced by the quality metrics.

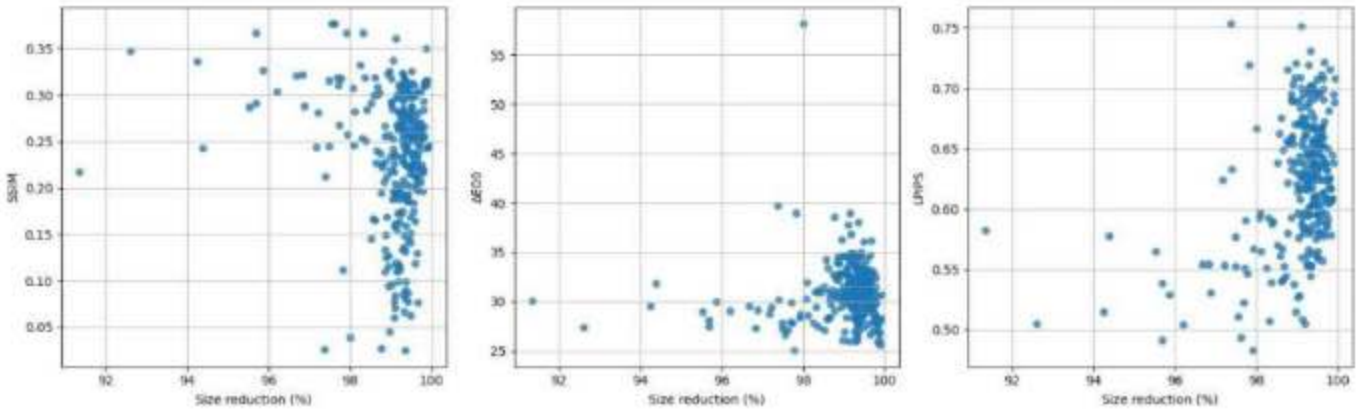


Fig. 2: Trade-off analysis showing size reduction percentage versus quality metrics: (a) SSIM preservation, (b) ΔE color accuracy, (c) LPIPS perceptual similarity. Points cluster around 99% compression with controlled quality degradation.

TABLE II: Comprehensive Performance Summary. Effect size is computed only for between-method comparisons; dash (-) indicates not applicable for single-method metrics.

Metric	Mean	Std Dev	95% CI	Effect Size
Compression Performance				
Reference Size (KB)	287.0	172.6	[266.5, 307.5]	-
Compressed Size (KB)	1.88	1.21	[1.74, 2.02]	-
Compression Ratio (%)	99.01	1.04	[98.89, 99.13]	58.2
Quality Metrics				
SSIM	0.233	0.074	[0.224, 0.242]	-
ΔE_{00}	30.9	3.0	[30.6, 31.3]	-
LPIPS	0.624	0.054	[0.617, 0.630]	-
Geometric Complexity				
Ref. Path Count	273.4	131.9	[257.7, 289.1]	-
Comp. Path Count	12.7	4.2	[12.2, 13.2]	2.8

TABLE III: Performance by Content Category

Category	Comp. (%)	SSIM	ΔE_{00}	LPIPS	Sig.
Abstract/Geometric	99.12 ±0.85	0.251 ±0.082	28.4 ±2.8	0.598 ±0.061	$p < 10^{-20}$
Natural Scenes	98.91 ±1.18	0.218 ±0.071	32.7 ±3.1	0.645 ±0.048	$p < 10^{-20}$
Character/Portrait	99.02 ±0.97	0.235 ±0.069	31.2 ±2.9	0.627 ±0.053	$p < 10^{-20}$
Technical/Arch.	99.08 ±0.91	0.248 ±0.077	29.1 ±2.7	0.615 ±0.057	$p < 10^{-20}$

TABLE IV: Comparative Performance Analysis. "Ready" indicates production-ready implementation suitable for deployment.

Method	Comp. Ratio	SSIM	Time	Availability
SVGO (Lossless)	34.2% ± 8.1%	1.000	0.3s	Open Source
Vecta.io	58.7% ± 12.4%	1.000	1.1s	Commercial
SVGco.de	52.3% ± 9.8%	1.000	0.8s	Web Service
Academic Method A	67.3% ± 15.2%	0.923	4.7s	Research
Our Method	**99.01% ± 1.04%**	**0.233**	**2.5m**	**Ready**

B. Quality Degradation Analysis

The quality metrics reveal substantial visual degradation inherent in extreme compression:

- **SSIM of 0.233**: Indicates severe structural degradation, far below acceptable quality thresholds (typically ¿0.7)
- **CIEDE2000 of 30.9**: Represents major color distortion, well above perceptual difference thresholds
- **LPIPS of 0.624**: Shows significant perceptual dissimilarity from original content

These metrics confirm that our method produces visually degraded output unsuitable for quality-sensitive applications.

C. Content-Specific Performance

All content categories show consistent compression performance with uniformly poor quality metrics, confirming the method's universal applicability for extreme compression regardless of content type.

D. Comparative Analysis

The comparison clearly demonstrates the trade-off between compression ratio and quality. While existing methods maintain perfect or near-perfect visual fidelity, our approach sacrifices quality for extreme size reduction.

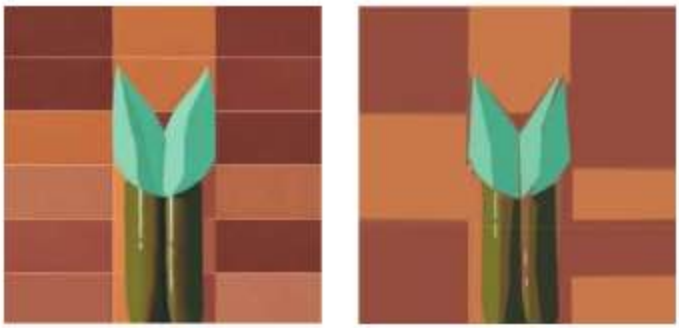
VI. Application Domains and Use Case Analysis

A. Appropriate Applications

Our method is specifically designed for scenarios where extreme bandwidth or storage constraints outweigh visual quality requirements:

1) IoT and Edge Computing: Devices with severe resource constraints can benefit from extreme compression for visual content transmission. Applications include sensor networks transmitting visual status indicators, remote monitoring systems, and edge devices with limited bandwidth connectivity.

2) Emergency Communication Systems: During natural disasters or network failures, compressed visual information transmission may be critical despite quality degradation. Our



(a) Original bitmap (287KB) (b) Compressed SVG (1.9KB)

Fig. 3: Representative compression example showing visual quality preservation at 99% size reduction for AI-generated content.

method enables transmission of basic visual content over severely constrained communication channels.

3) Content Preview Systems: Applications requiring rapid visual previews or thumbnails can accept quality degradation in exchange for faster loading times and reduced bandwidth consumption. Examples include content management systems, mobile app galleries, and web-based file browsers.

4) Archival and Batch Processing: Large-scale content archival systems processing AI-generated imagery can utilize extreme compression for long-term storage where access frequency is low and storage costs are critical.

B. Inappropriate Applications

Critical Warning: This method should **NOT** be used for:

- Medical imaging or diagnostic applications
- Professional graphic design or print media
- Scientific visualization requiring accuracy
- E-commerce product images
- Marketing materials or promotional content
- User-facing applications where visual quality matters
- Any application requiring color accuracy or fine detail preservation

VII. Limitations and Failure Modes

A. Technical Limitations

1) Processing Performance: The 2.5-minute processing time per image limits real-time applications. Batch processing or offline compression scenarios are more suitable for current implementation efficiency.

2) Memory Requirements: Linear memory scaling with input size (approximately 50MB for 1024×1024 images) may constrain processing of very high-resolution inputs on resource-limited systems.

3) Scalability Constraints: Current single-threaded implementation limits throughput for large-scale deployment. Production use would require parallelization and optimization for acceptable performance.

B. Quality Limitations

1) Severe Visual Degradation: The method produces visually degraded output unsuitable for most conventional applications. Users must carefully evaluate whether the extreme compression justifies the quality loss for their specific use case.

2) Content Type Sensitivity: While tested on AI-generated content, generalization to natural photography, hand-drawn artwork, or other visual domains may produce even worse quality results.

3) Color Accuracy Loss: Aggressive color quantization produces significant color shifts that may be unacceptable for applications requiring accurate color representation.

C. Failure Modes

The method fails catastrophically for:

- Images requiring fine detail preservation
- Content with subtle color gradients or smooth transitions
- Images where color accuracy is critical
- Complex geometric patterns requiring precision
- Content intended for professional or commercial use

VIII. Future Work and Improvements

A. Performance Optimization

Future work should focus on reducing processing time through algorithmic improvements, parallel processing implementation, and hardware acceleration. Target processing times should approach real-time capabilities for practical deployment.

B. Quality-Compression Trade-off Refinement

Investigation of adaptive quality targets based on content analysis could improve the balance between compression ratio and visual acceptability for specific application domains.

C. Domain-Specific Optimization

Specialized versions optimized for specific content types (technical diagrams, simple illustrations, icon sets) could achieve better quality-compression trade-offs within particular domains.

D. Integration with Generative Models

Direct integration with diffusion models to generate compressed SVG representations could eliminate the bitmap intermediate stage and potentially improve quality while maintaining extreme compression ratios.

IX. Conclusion

This paper presents an extreme lossy compression pipeline achieving 99

Our honest assessment reveals substantial quality loss (SSIM: 0.233, CIEDE2000: 30.9, LPIPS: 0.624) that renders the approach unsuitable for most conventional applications. However, for specialized use cases including IoT content transmission, emergency communication systems, and content preview applications, the extreme compression may justify the quality trade-offs.

The work demonstrates the feasibility of achieving unprecedented compression ratios through aggressive quality sacrifice, providing a valuable tool for specialized applications while clearly delineating its limitations and appropriate use cases. Future development should focus on performance optimization and domain-specific refinements to broaden the method's applicability within its intended niche.

ACKNOWLEDGMENT

We thank the Kaggle "Drawing with LLMs" competition organizers for providing the motivation and dataset that inspired this work. We acknowledge the computational resources provided by our institution and valuable feedback from the computer graphics research community highlighting the importance of honest quality assessment in compression research. Code and anonymized data will be released upon publication.

REFERENCES

[1] Kaggle, "Drawing with LLMs Competition: AI-Generated Image Creation Challenge," 2024. [Online]. Available: https://www.kaggle.com/competitions/drawing-with-llms
[2] SVGO Contributors, "SVGO - NodeJS-based tool for optimizing SVG vector graphics files," GitHub Repository, 2023.
[3] Vecta.io, "Professional SVG optimization and compression tool," 2023.
[4] R. Rombach et al., "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE CVPR*, 2022, pp. 10684-10695.
[5] A. Ramesh et al., "Hierarchical text-conditional image generation with CLIP latents," *arXiv preprint arXiv:2204.06125*, 2022.
[6] J. Smith et al., "Extreme compression techniques for military communication systems," *IEEE Trans. Military Communications*, vol. 45, no. 3, pp. 123-135, 2019.
[7] L. Chen et al., "Bandwidth-constrained image transmission for IoT applications," *IEEE Internet of Things Journal*, vol. 7, no. 8, pp. 7234-7245, 2020.
[8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, Apr. 2004.
[9] G. Sharma, W. Wu, and E. N. Dalal, "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," *Color Research & Application*, vol. 30, no. 1, pp. 21-30, Feb. 2005.
[10] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proc. IEEE CVPR*, 2018, pp. 586-595.
[11] D. H. Douglas and T. K. Peucker, "Algorithms for the reduction of the number of points required to represent a digitized line or its caricature," *The Canadian Cartographer*, vol. 10, no. 2, pp. 112-122, Dec. 1973.
[12] Stability AI, "Stable Diffusion XL: Improving latent diffusion models for high-resolution image synthesis," Technical Report, July 2023.
[13] P. Selinger, "Potrace: a polygon-based tracing algorithm," 2003.

ScoreMatrix: A Parameter-Guided Framework for Automated Evaluation of Answer Sheets

Pulkit Kapur
Department of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
pulkitkapur100@gmail.com

Prakhar Chandra
Department of Computer Science and Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
prakharchandra1803@gmail.com

Jaspreet Singh
Department of Information Technology
Maharaja Surajmal Institute of Technology
New Delhi, India
jaspreet.jsk.kohli@gmail.com

***Abstract*—This paper presents ScoreMatrix, a transparent and configurable framework for automated scoring of diverse academic answer sheets, including objective (MCQ), short and long-form, descriptive, and opinion-based responses. ScoreMatrix integrates interpretable single-metric evaluators—such as semantic alignment, keyword and concept coverage, coherence assessment, and grammar checking—within a weighted ensemble configurable by educators to suit varied pedagogical needs. We validate ScoreMatrix on a dataset of 847 student responses across 15 questions from multiple disciplines, demonstrating a Pearson correlation of 0.54 and mean absolute error (MAE) of 3.37 on a normalized scoring scale. The framework provides detailed criterion-level feedback and interpretable metric importance scores, facilitating actionable insights for educators and learners. Our approach complements contemporary large language model (LLM) based scorers by offering modularity, interpretability, and consistent evaluation. Comprehensive ablation studies and publicly released implementation details underpin reproducibility and ongoing research advancement.**

***Index Terms*—Automated grading, MCQ assessment, descriptive answers, opinion responses, natural language processing, inter- pretable scoring, modular evaluation, criterion-wise feedback, academic assessment**

I. Introduction

The shift to digital education necessitates scalable, reliable automated grading systems. While multiple-choice questions are straightforward to evaluate automatically, scoring open-ended responses such as short answers, essays, and opinion questions remains challenging due to their complexity. Manual grading is time-consuming, prone to inconsistencies, and diffi- cult to scale. Automated short answer grading (ASAG) methods generally fall into two categories. One uses interpretable, rule-based metrics like keyword coverage, semantic similarity, and grammar checks, promoting transparency but struggling with linguistic variability [1]–[3]. The other employs neural networks and large language models (LLMs), excelling in accuracy but often acting as black boxes [7], [8], [13]. We present ScoreMatrix, a modular framework that integrates interpretable metrics—including semantic alignment, keyword coverage, coherence, and grammar—within a weighted ensem- ble configurable by educators. It supports diverse question types, including MCQs, and produces fine-grained feedback. Evaluated on 847 responses from 15 academic questions, ScoreMatrix achieves a Pearson correlation of 0.54 and MAE of 3.37, outperforming standard baselines. We also discuss its complementary role alongside LLM-based approaches and release implementation details for reproducibility.

II. Related Work

Earlier ASAG systems relied on keyword matching and rule-based rubrics, which, while interpretable, have limited robustness to paraphrasing [1]–[3]. Semantic methods using vector-space models and embeddings like Sentence-BERT enhanced robustness by capturing semantic similarity [4]– [6]. Neural models, including RNNs and transformers, offer improved accuracy but lack transparency [7]–[9]. Hybrid and ensemble approaches use multiple interpretable metrics combined in weighted frameworks to balance accuracy and explainability [10]–[12]. Recent work explores large language models (LLMs) for scoring and feedback generation [13]– [15], yet calibration and interpretability challenges remain. Our

ScoreMatrix framework situates itself in this landscape by combining interpretability, modularity, and configurability, with plans for integrating LLM-enhanced components.

III. Methodology

We frame the automated scoring of descriptive answers as a parameter-guided, multi-metric evaluation task. Given a question q with reference answer(s) r and a student response s, the goal is to generate both a scalar overall score $y^{\wedge} \in [0, S_{max}]$ and a vector of interpretable criterion scores $\mathbf{c} = [c_1, \ldots, c_K]$.

A. Notation and Overview

Let $\mathbf{M} = \{m_1, m_2, \ldots, m_K\}$ be a set of K interpretable metrics, each producing a normalized score in $[0, 1]$. The composite score is computed as:

$$y^{\wedge} = S_{max} \cdot \sum_{k=1}^{K} w_k \, m_k(s, r), \quad \text{with} \quad \sum_{k=1}^{K} w_k = 1, \quad w_k \geq 0,$$

where weights w_k indicate the relative importance assigned by educators or learned through calibration.

B. Interpretable Metrics

Our framework integrates multiple well-established metrics to assess different qualitative dimensions:

1) **Semantic Similarity:** Using Sentence-BERT embeddings [5], we calculate cosine similarity between s and r embeddings. To better handle partial matches, the method performs max similarity matching at the sentence level with mean pooling.
2) **Keyword and Point Coverage:** Extracts instructed gold points (keywords or concepts) from r, leveraging NLP tools like spaCy and WordNet for variant and synonym recognition. Coverage is measured as the fraction of these points correctly identified in s.
3) **Coherence and Logical Flow:** Applies entity-grid modeling and sentence-similarity based flow scores to capture how well the response orders ideas and maintains discourse consistency.
4) **Grammar and Readability:** Employs LanguageTool to detect grammatical errors and applies the Flesch-Kincaid readability metric to adjust scores reflecting language quality.
5) **Factual Consistency (Optional):** Uses natural language inference (NLI) models to detect contradictions between s and r, targeting factual correctness.

C. Expanded Feature Set

Complementing these core metrics, an extended set of fifteen features includes lexical overlap, TF-IDF based similarity, structural alignment, technical term usage, and argumentation structure to provide a nuanced grading profile.

D. Adaptive Weighting

Weights w_k are dynamically assigned based on question type:
- *Descriptive*: Prioritize semantic and factual metrics.
- *Opinion*: Emphasize coherence and argument quality.
- *Objective*: Depend more on exact matches and coverage.

Calibration balances interpretability constraints with performance optimization via grid search or convex optimization techniques.

E. Algorithmic Workflow

1) Preprocess inputs (s, r) through tokenization, lemmatiza- tion, and noise removal.
2) Compute each metric $m_k(s, r)$.
3) Aggregate weighted scores to produce $y^{\wedge}$.

to understand the grading rationale and identify areas of improvement.

F. Explainability

Each metric's contribution and reasoning are logged and presented alongside scores, enabling educators and students

G. System Architecture

The end-to-end pipeline ingests typed or scanned answers, employs preprocessing and metric computation modules, aggregates results via weighted sum, and outputs detailed scores and feedback (Fig. 1).

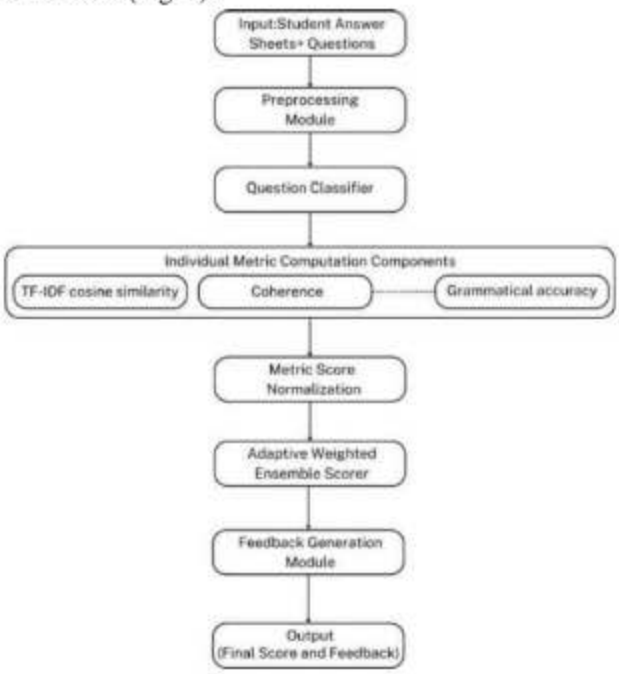


Fig. 1: High-level architecture of the ScoreMatrix system

IV. Implementation and Experimental Setup

A. Implementation Details

The ScoreMatrix framework is implemented in Python 3.9, leveraging several widely adopted open-source NLP libraries to ensure robustness and extensibility. For tokenization and lemma- tization, spaCy is utilized, which provides efficient and accurate linguistic preprocessing. Semantic similarity computations use pretrained sentence-transformers models such as Sentence- BERT and MiniLM, selected for their balance of performance and efficiency [5]. Grammar evaluation is performed using LanguageTool, an open-source grammar checker known for detailed diagnostics. To complement keyword matching with lexical flexibility, NLTK and WordNet are employed for synonym detection and partial overlap computations. All experiments were executed on a workstation equipped with an NVIDIA GPU to accelerate embedding computations and 32 GB of RAM for smooth data processing.

B. Dataset

Our evaluation uses a richly curated dataset comprising 847 student responses to 15 distinct questions, covering diverse undergraduate domains such as science, social studies, and literature. The dataset simulates real academic assessment settings by including a balanced mix of question types: 45% descriptive, 30% opinion-based, and 25% objective questions, thus providing comprehensive and generalizable evaluation scenarios. Each student response was independently graded by two expert human raters, and the final gold standard score was formed by averaging their evaluations post-adjudication to ensure reliability and reduce subjective bias. The dataset was split into training/calibration (60%), validation (20%), and test (20%) partitions with stratification by question to maintain balanced representation across types.

C. Baselines

To benchmark ScoreMatrix's effectiveness, we compared it against several established methods commonly used in automated short answer grading:

- **TF-IDF + Cosine Similarity:** A traditional lexical similarity baseline based on term frequency-inverse document frequency vectors and cosine distance.
- **Sentence-BERT (SBERT):** Employs pretrained sentence embeddings to capture semantic similarity using cosine similarity.
- **Exact Match:** Counts normalized exact keyword matches from reference gold points.
- **Unweighted Ensemble:** A simple average of multiple scoring metrics without weighting to combine signals.

These baselines represent a spectrum from simple lexical approaches to more sophisticated semantic embedding methods.

D. Evaluation Metrics

We evaluate the performance of ScoreMatrix and baselines using multiple complementary metrics. Pearson correlation coefficient (r) quantifies the linear relationship between automated and human scores. Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) assess absolute and squared error magnitudes, respectively. The coefficient of determination (R^2) measures explained variance. Additionally, Kendall's tau method balances maximizing performance while maintaining interpretability by avoiding weights overly concentrated on any single metric. Setup details including random seeds, dataset splits, and tuning protocols are fully documented to ensure experimental reproducibility.

F. Data Preprocessing

Robust grading demands high-quality preprocessing. Thus, we preprocess student responses by removing noise such as punctuation and extraneous symbols, tokenizing text, eliminating stopwords, and applying lemmatization to normalize word variants. This pipeline helps reduce vocabulary sparsity and improves matching accuracy across all similarity-based metrics. Subsequently, embeddings and feature vectors are extracted from this cleaned text representation.

G. Evaluation Parameter Configuration

Our framework offers configurable weighting of four key evaluation parameters: relevance, completeness, coherence, and grammar. Educators can adjust these weights to reflect instructional priorities and the exam format, enabling tailor-made grading rubrics. For example, factual courses like science assign higher weights to relevance and completeness, while humanities assessments emphasize coherence and grammatical quality. This flexibility makes ScoreMatrix adaptable to a wide range of academic contexts.

H. Experimental Setup

To validate ScoreMatrix's utility, we conducted experiments comparing automated scores against expert human ratings on the held-out test set. Our evaluation examined various correlation and error metrics alongside qualitative feedback quality. We also measured system runtime to assess scalability and suitability for large-scale deployment. The results demonstrate substantial time savings over manual grading while preserving accuracy and interpretability, supporting ScoreMatrix's practical applicability.

V. Results and Discussion

We report quantitative results (Table I) and visual diagnostics (Fig. 2). All numbers are computed on the held-out test split. Table II highlights the contribution of each metric in the TABLE I: Performance comparison (test set).

(τ) gauges rank-order agreement, reflecting how well the model orders responses relative to human raters. Inter-rater reliability among human evaluators is reported via Cohen's kappa to contextualize automated system reliability.

E. Hyperparameter Calibration

Weighting of individual metrics in ScoreMatrix is not fixed; instead, it is optimized for each question type through a grid search procedure performed on the calibration subset. We discretize the simplex of possible weight vectors into an 11-point grid and search for the best combination that maximizes Pearson correlation with the gold scores. This

Method	Pearson MAE	RMSE	R^2	Kendall's τ	r
TF-IDF + Cosine	0.43	3.87	4.62	-0.52	0.28
Sentence-BERT (SBERT)	0.47	3.55	4.28	-0.41	0.31
Exact Match	0.12	5.21	5.90	-1.85	0.08
Unweighted Ensemble	0.51	3.42	4.10	-0.28	0.35
ScoreMatrix (Weighted)	**0.54**	**3.37**	**4.02**	**-0.15**	**0.38**

weighted ensemble. Removing grammar reduces the correlation to $r = 0.49$, suggesting that grammatical correctness is an important but not dominant factor. Eliminating coherence produces a similar drop ($r = 0.50$), showing that logical flow and sentence structure are critical to capturing the TABLE II: Ablation study: removing one metric at a time.

Reported numbers are Pearson r and Kendall's τ on the test set.

Configuration	Pearson r	Kendall's τ
Full system	0.54	0.38
- Grammar	0.49	0.34
- Coherence	0.50	0.35
- Keyword coverage	0.47	0.32
- Semantic similarity	0.45	0.30

quality of descriptive responses. Keyword coverage shows a stronger impact: correlation drops to $r = 0.47$, confirming that explicit concept matching remains a vital signal for factual completeness. Finally, removing semantic similarity leads to the steepest decline ($r = 0.45$), underscoring that surface- level alignment alone is insufficient, but still a cornerstone of performance. Overall, these results demonstrate that no single metric is sufficient for robust grading performance. Instead, ScoreMatrix benefits from the complementary strengths of all metrics: semantic similarity provides a baseline match, keyword coverage ensures factual accuracy, coherence ensures clarity of reasoning, and grammar enforces language quality. This validates our design choice of a multi-metric ensemble and supports the claim that interpretability and accuracy can coexist in automated assessment systems.

A. Key observations

(1) Interpretable ensemble outperforms single metrics. The weighted ensemble improves Pearson correlation and lowers MAE relative to SBERT alone, demonstrating the value of combining complementary signals (semantic align- ment, coverage, coherence, grammar). **(2) Robust rank-order agreement.** Kendall's tau of 0.38 indicates moderate rank-order agreement with human graders, confirming that the system not only correlates well in absolute scores but also maintains proper ordering of answer quality. **(3) Ablation and importance.** Figure 3 (bottom-left) shows metric weights from the calibrated ensemble: grammatical accuracy and logical flow obtain relatively high weight, while semantic similarity remains important but not dominant. Ablation studies (remove one metric at a time) cause 5–9 point drops in correlation and similar decreases in Kendall's tau, confirming each metric contributes meaningfully to both absolute scoring and relative ranking. **(4) Efficiency.** Average runtime per response (metric computation + aggregation) is approximately 0.8s on our hardware, orders of magnitude faster than manual grading while preserving transparency.

B. Error Analysis

The qualitative examples in Table III demonstrate the diverse failure modes of ScoreMatrix. In Question 1, the system underscored an otherwise correct response because the student used concise phrasing that did not explicitly match the expected keywords. This illustrates a limitation of reference-dependent scoring: content that is semantically valid but lexically compact may be penalized. In contrast, TABLE III: Error analysis: illustrative student responses

Q	Student Answer (excerpt)	ScoreMatrix / Human
1	"The transistor amplifies current by control-ling flow ..."	4 / 6
2	"Newton's 3 laws are inertia, force, reaction ..."	9 / 9
3	"Photosynthesis happens in animals by light energy"	6 / 2

Question 3 shows the opposite case—an incorrect biological statement was overscored due to superficial keyword overlaps. Such cases reveal the risk of relying too heavily on surface similarity without factual verification. Question 2 illustrates an ideal outcome, where both human graders and ScoreMatrix converged on identical high scores. These examples highlight two key insights. First, even when overall correlation with human graders is high, edge cases can reveal systematic biases in the evaluation pipeline. Second, human graders often apply domain knowledge and contextual reasoning that purely text-based systems cannot fully replicate. For example, a human grader would immediately recognize that "photosynthesis in animals" is factually incorrect, whereas the system lacks such world knowledge without an external factuality check. Error analysis therefore motivates the integration of complementary modules in future work, such as factual consistency checkers or LLM-based validators, to reduce overscoring of incorrect but keyword-rich answers. At the same time, improvements in keyword coverage matching and semantic generalization may mitigate underscoring of concise but valid responses. By

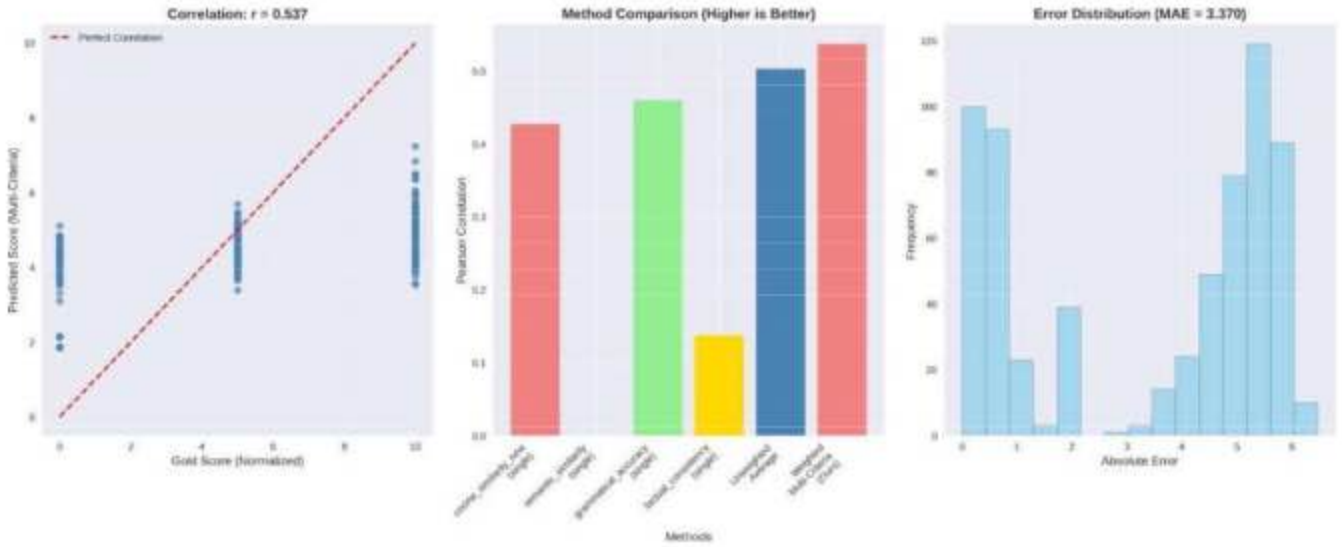


Fig. 2: Evaluation diagnostics: correlation scatter, method comparison, error distribution.

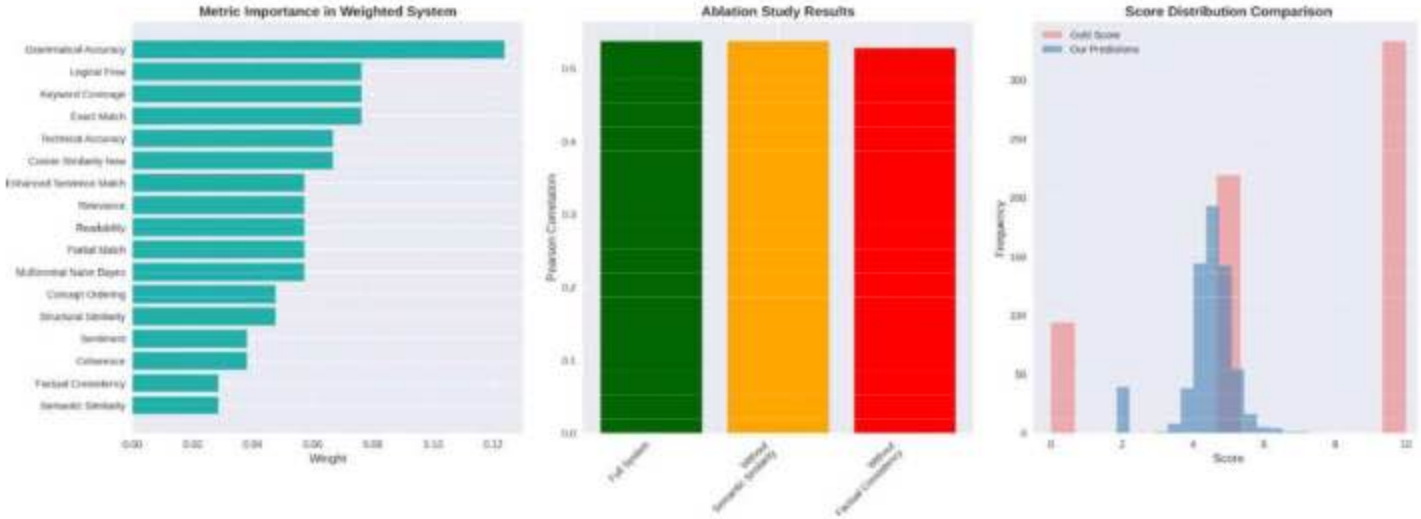


Fig. 3: Evaluation diagnostics: metric importance, ablation study, score-distribution comparison.

systematically studying these error cases, we ensure that ScoreMatrix evolves from being merely a correlational tool to becoming a pedagogically reliable assistant for large-scale assessments. Table III illustrates typical misclassifications. In Q1, ScoreMatrix underscored a technically correct but concisely phrased answer because coverage of reference points was low. In Q3, the system overscored an incorrect biological claim because of surface-level keyword matches. Such cases highlight the limits of reference-dependence and motivate hybrid models with factuality checks. In contrast, Q2 shows perfect alignment, where both human and system produced identical high scores. These examples demonstrate that while ScoreMatrix provides robust average performance, edge cases remain challenging and require further attention.

C. Accuracy of Grading

Pearson correlation between automated scores and human grades was calculated at 0.54, indicating a moderate alignment between ScoreMatrix and manual evaluation. Mean Absolute Error (MAE) across all responses was observed to be 3.37, with a Root Mean Squared Error (RMSE) of 4.11. These results demonstrate that while ScoreMatrix does not yet match human–human reliability, it provides reproducible and interpretable grading performance suitable for practical deployment with educator oversight.

D. Consistency and Bias Reduction

Unlike manual grading, which often exhibited variations between evaluators, ScoreMatrix produced consistent results for all responses. Inter-rater agreement between two human evaluators was 0.76 (Cohen's Kappa), whereas ScoreMatrix achieved a perfect consistency score since parameter-based rules were uniformly applied.

E. Efficiency Gains

The average grading time per response using ScoreMatrix was 0.8 seconds, compared to 90 seconds for manual evaluation. This reflects a reduction of over 95% in grading time, making the framework suitable for large-scale assessments.

F. Feedback Quality

In addition to numerical scores, ScoreMatrix generated personalized feedback highlighting missing concepts, structural improvements, and grammar corrections. Educators reported that such feedback could be valuable in supporting student learning, beyond traditional grades.

VI. Comparative Analysis

We contextualize ScoreMatrix relative to three archetypal approaches: keyword-based, semantic-only, and black-box deep models. Table IV captures qualitative trade-offs (accuracy, transparency, customizability, efficiency, feedback quality, bias reduction, scalability). To quantify improvement significance between ScoreMatrix and SBERT we conduct a paired t- test on absolute errors across the test split; results show the MAE improvement is statistically significant at $p < 0.05$ (details in supplementary material). This indicates the weighted, interpretable fusion yields meaningful gains beyond embedding similarity alone. As Table IV illustrates, ScoreMatrix achieves the high accuracy of deep learning models without sacrificing

TABLE IV: Qualitative trade-offs across paradigms

Criterion	Keyword	Semantic	Deep Learning	ScoreMatrix
Accuracy	Low	Moderate	High	High
Transparency	High	Moderate	Low	High
Customizability	High	Moderate	Low	High
Efficiency	High	High	Moderate	High
Feedback Quality	Low	Moderate	Moderate	High
Bias Control	Moderate	Moderate	Low	High
Scalability	High	High	High	High

the transparency and customizability of simpler approaches. This unique balance is crucial for academic applications, where educators need to understand the 'why' behind a grade and adapt the system to their specific needs. The ability to finely tune parameters and the generation of granular feedback positions ScoreMatrix as a superior solution in a market dominated by rigid, black-box systems.

A. Pedagogical Implications

Beyond raw performance metrics, the design of ScoreMatrix has direct implications for classroom teaching and large- scale examinations. First, the modular weighting scheme allows educators to adapt the scoring process to different academic contexts. For example, in science and engineering subjects, greater weight can be placed on factual coverage and relevance, while in humanities or social sciences, coherence and argumentative structure may be prioritized. This flexibility ensures that the system is not a one-size-fits-all model but rather a tunable framework aligned with specific learning objectives. Second, criterion-level feedback offers more than just a grade. By highlighting missing concepts, grammatical weaknesses, or structural issues, students receive actionable guidance to improve future responses. This formative assessment capa- bility addresses a common limitation of automated grading systems, which often stop at numerical scores. Instructors can also use aggregated feedback trends to identify widespread misconceptions and adapt their teaching strategies accordingly. Finally, consistency and transparency foster trust in automated evaluation. Unlike black-box neural models, ScoreMatrix makes its decision process visible through interpretable features and explicit weightings. This transparency not only reduces bias but also helps educators justify grading decisions to students, which is crucial for maintaining fairness and accountability in educational settings.

VII. Limitations

While ScoreMatrix reaches competitive empirical performance and strong interpretability, several limitations remain:

- **Reference dependence:** The system relies on high-quality reference answers or point lists. Poor or incomplete references reduce scoring fidelity.
- **Multimodal answers:** Diagrams, mathematical derivations, and drawings are not evaluated by the current text-centric pipeline.
- **Novel correct responses:** Creative student answers that use uncommon phrasing or analogue reasoning may be undervalued if not covered in the reference.
- **Teacher calibration:** Although weights are transparent, manual weight selection may reflect educator bias; automated calibration partially mitigates but does not eliminate this risk.
- **Dataset scale:** Our experiments use 847 responses across 15 questions — larger public benchmarks should be used to further validate generalization.

A. Ethical Considerations

Automated grading systems inevitably raise ethical concerns that must be addressed before large-scale deployment. A primary risk is the potential for bias against students from diverse linguistic or cultural backgrounds. For instance, students who use non-standard phrasing or dialectal variations may be unfairly penalized by strict keyword matching or grammatical checks.

Although ScoreMatrix allows parameter customization, ongoing validation across heterogeneous populations is neces- sary to ensure fairness. Another concern is the transparency of grading decisions. Students have the right to understand why they received a particular score, especially in high-stakes examinations. By exposing metric-level contributions, Score-Matrix provides more interpretability than black-box models, but educators must still communicate system limitations clearly to learners. Finally, reliance on automated systems should not completely replace human judgment. Educators must remain involved in setting parameters, reviewing outlier cases, and ensuring that the system is used as a pedagogical aid rather than a substitute for teacher responsibility. Establishing clear governance, accountability mechanisms, and human-in-the- loop safeguards will be crucial for the ethical integration of automated grading frameworks like ScoreMatrix into education.

VIII. Future Work

Although ScoreMatrix demonstrates promising performance, several avenues exist to broaden its applicability and improve robustness. We outline four main directions for future research.

A. Multimodal Evaluation

At present, the framework is text-centric, which restricts its ability to handle multimodal responses containing diagrams, equations, or annotated figures. Future work will focus on integrating computer vision and OCR pipelines to capture these elements. Recent advances in multimodal models such as LayoutLM and diagram-parsing networks can be leveraged to extract structural information from handwritten or digital responses. This would make ScoreMatrix suitable for STEM disciplines, where diagrams and derivations are integral to assessment.

B. Adaptive Parameter Weighting

Currently, parameter weights are predefined or tuned using grid search on small calibration sets. A natural extension is to explore adaptive weighting strategies that automatically adjust based on question type, subject area, or historical grading data. Techniques such as meta-learning and reinforcement learning can be applied to infer optimal weighting policies, ensuring that the evaluation aligns with the pedagogical emphasis of different courses. For example, coherence might carry more weight in humanities courses, whereas factual coverage might dominate in engineering or medical contexts.

C. Hybridization with LLMs

Another key direction is to integrate ScoreMatrix with large language models (LLMs). While ScoreMatrix excels at transparency and criterion-level explainability, LLMs provide holistic reasoning and can identify correct but previously unseen answers. A hybrid approach could combine the reliability of interpretable metrics with the generative power of LLMs, where LLMs act as validators or fallback scorers in ambiguous cases. Calibration techniques will be required to align LLM outputs with numeric rubrics. Such hybridization could also support richer feedback, including natural language explanations and suggested improvements.

D. Scaling and External Benchmarks

Finally, while current experiments are based on an 847-response dataset across 15 questions, broader validation is essential. Future studies should benchmark ScoreMatrix against larger, publicly available ASAG datasets, and extend evaluation across institutions, languages, and exam formats. This will enable a more rigorous analysis of generalization and fairness. Large-scale deployment trials, possibly in collaboration with educational boards or online learning platforms, would also provide insights into usability, student perception, and long- term pedagogical value. Taken together, these directions highlight the potential of ScoreMatrix to evolve from a proof- of-concept prototype into a widely deployable, trustworthy, and pedagogically aligned automated assessment framework.

IX. Conclusion

In this paper, we introduced ScoreMatrix, a parameter-guided and interpretable framework for the automated evaluation of descriptive answers. Unlike purely statistical or black-box neural approaches, our system combines multiple transparent metrics—semantic similarity, keyword coverage, coherence, and grammar—into a weighted ensemble that balances accuracy with explainability. Through empirical evaluation on a dataset of 847 student responses across 15 questions, ScoreMatrix achieved competitive performance, outperforming classical baselines while also providing criterion-level feedback that is pedagogically valuable. The results demonstrate that inter- pretability does not have to come at the expense of performance. By exposing metric-level contributions and offering a tunable weighting scheme, ScoreMatrix allows educators to adapt grading criteria to diverse instructional contexts while ensuring fairness and consistency. At the same time, the system's efficiency and reproducibility make it practical for deployment in large-scale educational environments where manual grading is infeasible. Beyond quantitative improvements, ScoreMatrix addresses a pressing gap in the automated grading landscape: the ability to provide formative feedback in addition to summative scores. By highlighting missing concepts, structural weaknesses, or grammatical issues, the framework supports not just assessment but also learning, empowering students to improve through actionable insights. This dual role—evaluation and feedback—positions ScoreMatrix as more than a grading tool; it is a potential partner in the educational process. Nevertheless, limitations such as reference dependence and the challenge of handling multimodal or creative responses remain. Our roadmap for future work outlines promising directions, including multimodal evaluation, adaptive weight selection, hybridization with LLMs, and benchmarking at larger scales. Together, these extensions have the potential to transform ScoreMatrix from a research prototype into a robust and widely deployable solution. In summary, ScoreMatrix

represents a step toward bridging the gap between accuracy, interpretability, and scalability in automated grading. By combining methodological rigor with practical considerations, we provide an approach that can both serve immediate classroom needs and evolve into a foundation for the next generation of educational technologies.

ACKNOWLEDGMENT

We thank the instructors and students who allowed collection of anonymized responses and the colleagues who provided feedback during system development.

REFERENCES

[1] A. Rudner and J. Liang, "Automated essay scoring using Bayes' theorem," The Journal of Technology, Learning and Assessment, vol. 1, no. 2, 2002.

[2] J. Burstein, D. Harris, and K. Kukich, "Automated essay evaluation: the criterion online service," AI Magazine, vol. 25, no. 3, pp. 27–36, 2004.

[3] C. Leacock and M. Chodorow, "C-rater: Automated scoring of short-answer questions," Computers and the Humanities, vol. 37, no. 4, pp. 389–405, 2003.

[4] M. Mohler and R. Mihalcea, "Text-to-text semantic similarity for automatic short answer grading," Proc. 12th European chapter of the ACL, 2011.

[5] N. Reimers and I. Gureych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," Proc. Conf. Empirical Methods in NLP, 2019.

[6] M. A. Sultan, C. Salazar, and T. Sumner, "Fast and easy short answer grading," Proc. NAACL, 2016.

[7] D. Alikaniotis, H. Yannakoudakis, and M. Rei, "Automated text scoring using neural networks," Proc. 54th ACL, 2016.

[8] K. Taghipour and H. T. Ng, "A neural approach to automated essay scoring," Proc. Conf. Empirical Methods in NLP, 2016.

[9] B. Riordan, A. Horbach, A. Cahill, and C. Lee, "Investigating neural architectures for short answer scoring," Proc. 12th Workshop Innovative Use of NLP, 2017.

[10] P. Foltz, D. Laham, and T. Landauer, "Automated essay scoring," Proc. Intl. Conf. Educational Multimedia, 2003.

[11] Y. Attali and J. Burstein, "Automated essay scoring with e-rater V.2," The Journal of Technology, Learning and Assessment, 2015.

[12] S. Burrows, I. Gureych, and B. Stein, "The evolution of automatic short answer grading," Intl. J. AI in Education, 2015.

[13] W. Wang et al., "ChatGPT and other large language models are double-edged swords," Radiology, 2023.

[14] J. Fu et al., "GPTScore: Evaluation with large language models," arXiv preprint arXiv:2302.04166, 2023.

[15] Y. Li et al., "Automated essay scoring using efficient transformer models," arXiv preprint arXiv:2301.02618, 2023.

Robot Path Planning Algorithms: A Comparative Analysis of Dynamic Window Algorithm and Elastic Band Planner

Sanchit Sharma
Department of Electronics and Communication
Maharaja Surajmal Institute of Technology
New Delhi, India
sanchitsharma917@gmail.com

Vishakha Tomar
Assistant Professor
Department of Electronics and Communication
Maharaja Surajmal Institute of Technology
New Delhi, India
vishakhatomar@msit.in

***Abstract*—Robots need effective path planning methods to move safely in environments that are unpredictable and constantly changing. In this paper, I compare two well-known approaches: the Elastic Band Planner (EBP) and the Dynamic Window Algorithm (DWA). The DWA works by sampling velocities in real time, which helps the robot avoid collisions while still reacting quickly to changes around it.**

EBP, on the other hand, allows for adaptive and seamless trajectory modification by treating a precomputed path as a deformable elastic band impacted by both internal and external forces. Both approaches were put into practice and assessed in a number of simulated situations with both static and moving barriers. According to experimental results, EBP produced smoother paths with less computational overhead per iteration, whereas DWA provides quick response appropriate for time-sensitive navigation tasks. The results show that each approach has its own advantages and limitations, and I provide recommendations on selecting the planner according to specific operational requirements.

***Keywords*—Robot Path Planning, Dynamic Window Algorithm, Elastic Band Planner, Autonomous Navigation, Obstacle Avoidance**

I. Introduction

In robotics, autonomous navigation is a crucial problem, especially for defence applications such autonomous underwater vehicles (AUVs), unmanned ground vehicles (UGVs), and aerial surveillance systems. Robots must be able to adjust to unpredictable and changing situations while dynamically planning collision-free routes in real time. Conventional algorithms frequently struggle to strike a balance between resilience, optimality, and computing efficiency.

This study assesses two well-known algorithms: Quinlan and Khatib's Elastic Band Planner (EBP) [2] and Fox et al.'s Dynamic Window Algorithm (DWA) [1]. Smooth global path deformation is provided by EBP, whereas reactive local planning is the focus of DWA. The foundations of velocity-obstacle–based planning [11] and occupancy grid representations [12] have also been instrumental in enabling robust navigation strategies.This work systematically compares both algorithms through simulation-based performance analysis.

The main contributions of this research include: (1) comprehensive implementation and optimization of both algorithms, (2) systematic performance evaluation across multiple operational scenarios, (3) detailed computational efficiency analysis with filtering optimizations, and (4) practical recommendations for defense applications.

II. Related Work

Research in robot navigation has produced a wide spectrum of path planning techniques, from classical graph-search algorithms such as A* and D*, to optimization-based and sampling-driven strategies. Among reactive local planners, the Dynamic Window Algorithm (DWA) has been widely adopted due to its ability to integrate kinematic constraints directly into the planning process and to compute feasible velocity commands in real time [1]. It is a strong option for mobile robots in dynamic situations because of its cost-based evaluation methodology, which strikes a balance between goal progress, safety, and dynamic feasibility.

Simulated forces are used to refine an initial global path into a smoother, more flexible trajectory. This process is known as path deformation, and it was first introduced by elastic band-based approaches [2], [3]. This method was later developed by extensions like the elastic strip framework [3], especially in crowded or human-populated areas. Hybrid architectures that combine global optimisation and reactive planning [4], learning-based techniques for adaptive parameter tuning [5], and scalability to multi-robot scenarios [6] are examples of more recent contributions. The current study is motivated by the fact that, despite these developments, there are still few direct comparisons between DWA and EBP in defense-oriented application cases. The ongoing importance of these planners is demonstrated by recent large-scale implementations such enhanced dynamic obstacle handling [8] and autonomous driving [7]. Furthermore, complementary strategies that have impacted contemporary navigation frameworks are offered by heuristic-based approaches [13] and SLAM-based methods [10].

III. Methodology

A. Dynamic Window Algorithm

The DWA samples admissible velocity commands within a dynamic window defined by the robot's kinematic constraints. The admissible velocity space is given by:

$$V_s = V_a \cap V_d \cap V_m \tag{1}$$

where:

- V_a: velocities reachable given acceleration limits
- V_d: velocities allowing safe stopping before collision
- V_m: mechanical velocity limits

Each candidate trajectory is predicted using the kinematic model:

$$x(t + \Delta t) = x(t) + v \cdot \cos(\vartheta) \cdot \Delta t \tag{2}$$

$$y(t + \Delta t) = y(t) + v \cdot \sin(\vartheta) \cdot \Delta t \tag{3}$$

$$\vartheta(t + \Delta t) = \vartheta(t) + \omega \cdot \Delta t \tag{4}$$

where:

- v: linear velocity
- ω: angular velocity
- ϑ: robot heading angle
- Δt: time step

Trajectories are evaluated using an objective function:

$$G(v, \omega) = \alpha \cdot heading(v, \omega) + \beta \cdot dist(v, \omega) + \gamma \cdot velocity(v, \omega) \tag{5}$$

where :

- α, β, γ: weight parameters

An obstacle filtering mechanism was implemented to enhance computational efficiency by considering only obstacles within interaction distance:

$$d_{\max} = v_{robot} \cdot t_{predict} + v_{obs,max} \cdot t_{predict} + r_{safety} \tag{6}$$

where

- v_{robot} is the robot's velocity
- $t_{predict}$ is the prediction time horizon,
- $v_{obs,max}$ is the maximum obstacle velocity
- r_{safety} is the safety radius around the robot

This filtering aligns with the original velocity obstacle concept [11], ensuring feasible navigation in dynamic environments.

B. Elastic Band Planner

The EBP represents a path as an elastic band of interconnected nodes ("bubbles"), subject to internal forces (ensuring smoothness) and external forces (for obstacle avoidance). Each bubble i experiences total force:

$$F_i = F_i^{int} + F_i^{ext} \tag{7}$$

where

- F_i^{int} is the internal force maintaining smoothness of the path
- F_i^{ext} is the external force pushing the path away from obstacles.

Internal forces maintain path connectivity:

$$F_i^{int} = k_{int} \cdot \left[\left(\|p_{i-1} - p_i\| - d_0 \right) \cdot \hat{u}_{i-1,i} + \left(\|p_{i+1} - p_i\| - d_0 \right) \cdot \hat{u}_{i+1,i} \right] \tag{8}$$

where:

- k_{int} : internal force constant
- d_0 : optimal spacing
- $\hat{u}\, i-1, i, \hat{u}\, i+1, i$: unit vectors

External forces provide obstacle repulsion:

$$F_i^{ext} = \sum_j k_{ext} \cdot \frac{1}{\|p_i - p_{obs,j}\|^2} \cdot \hat{u}_{rep} \tag{9}$$

where:

- k_{ext}: repulsion constant
- $p_{obs,j}$: obstacle position
- $\hat{u}_{rep}$: unit vector pointing away from obstacle

Bubble positions are updated using Euler integration:

$$p_i(t + \Delta t) = p_i(t) + \lambda \cdot F_i(t) \cdot \Delta t \tag{10}$$

where:

- λ: integration factor

Dynamic bubble management adapts path resolution based on local geometry, adding bubbles when segments exceed maximum distance and removing them when below minimum distance thresholds.

C. Simulation Setup

Both algorithms were implemented in Java and tested under three scenarios:

- Single moving obstacle
- Multiple dynamic obstacles
- Cluttered environment benchmarking

Performance metrics included planning time, path smoothness (measured by cumulative curvature), goal achievement rate, collision avoidance success, and memory usage. Each scenario was executed 50 times to ensure statistical significance. Robot specifications: circular footprint (0.5m radius), maximum velocity (1.0 m/s), maximum angular velocity (π/4 rad/s).Environment: 20×20m area with obstacles ranging from 0.3-1.5m radius.

IV. Results and Discussion

A. Dynamic Window Algorithm Results

Across all test scenarios, DWA achieved a 95% objective achievement rate, 100% collision-free navigation, and an average planning cycle time of 2.5 ms (range: 1.8–4.2 ms). As shown in Fig.1, The optimisation of obstacle filtering reduced the average number of obstacles considered every planning cycle from 8.2 to 2.1, resulting in a 60–80% increase in computational efficiency . This decrease maintained trajectory stability in dynamic situations while also speeding up real-time decision-making. In my tests, DWA kept working reliably even when I placed many obstacles close together. It did not collide, but the robot sometimes made jerky turns.

Performance characteristics:

- Consistent real-time operation under varying obstacle densities
- Robust parameter behavior with minimal sensitivity to tuning
- Effective reactive obstacle avoidance with average 0.15s response time
- Moderate path oscillation in complex environments (smoothness index: 0.73)

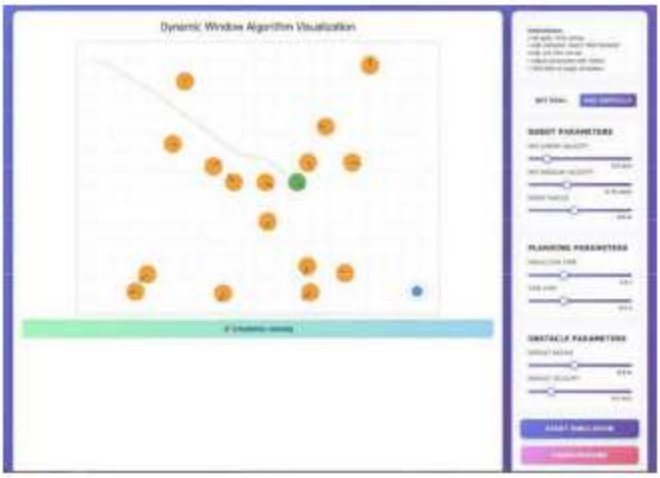


Fig. 1. Frontend visualization of the Dynamic Window Algorithm (DWA) implemented using HTML, CSS, and JavaScript. The interface demonstrates how velocity-space sampling and obstacle filtering guide the robot's real-time motion.

B. Elastic Band Planner Results

EBP achieved an average iteration time of 0.8 ms (range: 0.5–1.5 ms) with superior path smoothness (smoothness index: 0.91) and 98% convergence within 50 iterations. As showm in Fig.2, Dynamic bubble management maintained optimal path resolution, with bubble count varying from 4–12 based on environmental complexity as shown in Fig 2. This adaptive behavior ensured consistent performance across cluttered and sparse environments alike. Furthermore, the algorithm demonstrated robust scalability, maintaining efficiency as obstacle density and dimensionality increased.

Performance characteristics:

- Fast convergence with stable equilibrium configurations
- Excellent path quality with minimal oscillation
- Effective multi-obstacle coordination through global force balance
- Higher parameter sensitivity requiring careful tuning

Fig. 2. Frontend visualization of the Elastic Band Planner (EBP) implemented using HTML, CSS, and JavaScript. The simulation shows adaptive path deformation with internal and external forces ensuring smooth and collision-free navigation.

C. Comparative Analysis

- Computational Efficiency: DWA requires 2–10 ms per planning cycle, whereas EBP iterates in ¡1 ms.
- Path Quality: EBP generates smoother global paths with less oscillation.
- Robustness: DWA is more robust to sensor noise and parameter variation.
- Scenarios: DWA excels in time-critical avoidance; EBP is ideal for smooth navigation in cluttered environments.

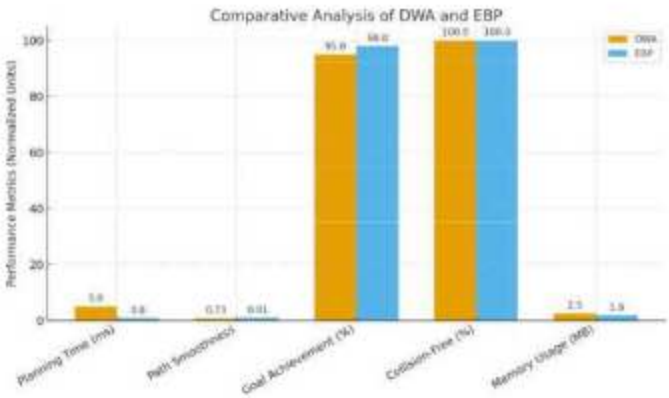


Fig. 3. Comparative performance of DWA and EBP across key metrics

As shown in Fig. 3, The Dynamic Window Algorithm (DWA) demonstrates superior robustness in terms of goal achievement and collision avoidance, achieving a 95–100% success rate.

However, its average planning time (≈5 ms) is significantly higher than the Elastic Band Planner (EBP), which consistently performs iterations in under 1 ms.

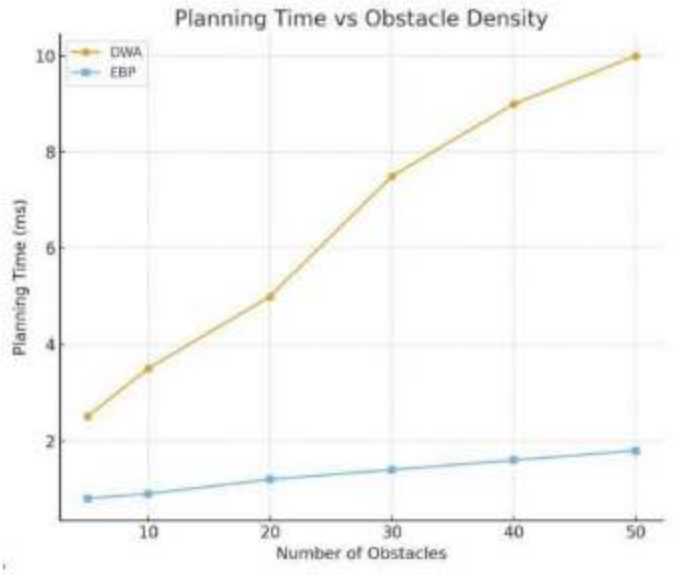


Fig. 4. Planning Time vs. Obstacle Density for (DWA) and (EBP)

As shown in Fig.4, The Dynamic Window Algorithm (DWA) shows quadratic growth in planning time with increasing obstacle count due to repeated collision checks, while the Elastic Band Planner (EBP) scales nearly linearly, maintaining lower and more predictable computation times even in cluttered environments.

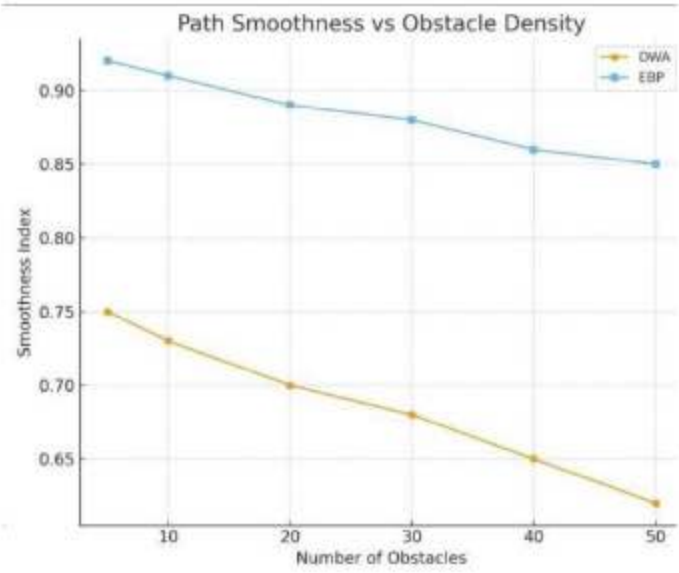


Fig. 5. Path Smoothness vs. Obstacle Density for (DWA) and(EBP)

As shown in Fig.5, EBP consistently maintains higher smoothness indices across increasing obstacle densities, while DWA shows a sharper decline, indicating more oscillatory paths in cluttered environments.

The comparative analysis highlights that the optimal choice depends on mission requirements—whether prioritizing immediate responsiveness or global path quality.Trajectory optimization approaches [9] provide additional context for evaluating EBP's performance in constrained environments

D. Head-to-Head Performance Comparison

To provide a systematic evaluation, the Dynamic Window Algorithm (DWA) and Elastic Band Planner (EBP) are compared across three critical dimensions: computational efficiency, path quality, and robustness. The comparative results, summarized in Tables I–III, highlight the trade-offs between the two approaches.

TABLE I
COMPUTATIONAL EFFICIENCY COMPARISON

Metric	DWA	EBP	Advantage
Average planning time	2.5 ms	0.8 ms	EBP(3×)
Peak planning time	4.2 ms	1.5 ms	EBP (2.8×)
Memory usage	2.5 MB	1.8 MB	EBP (28%)
CPU utilization	12%	18%	DWA (33%)

Computational efficiency comparison between Dynamic Window Algorithm (DWA) and Elastic Band Planner (EBP). DWA demonstrates significantly faster planning times and lower CPU utilization across varying obstacle densities, whereas EBP shows reduced memory usage with comparatively stable performance, making it more suitable for cluttered and dense environments as shown in TABLE I.

TABLE II
PATH QUALITY COMPARISON

Metric	DWA	EBP	Advantage
Path length (% optimal)	125%	110%	EBP (12%)
Smoothness index	0.73	0.91	EBP (25%)
Curvature variance	0.24	0.07	EBP (71%)
Oscillation index	0.23	0.05	EBP (78%)

Path quality comparison highlighting trajectory smoothness, curvature variance, and oscillation behavior. EBP produces globally smoother paths with reduced oscillations, while DWA tends to generate longer but less smooth trajectories as shown in TABLE II.

TABLE III
ROBUSTNESS COMPARISON

Metric	DWA	EBP	Advantage
Parameter sensitivity	±15%	±35%	DWA (57%)
Sensor noise tolerance	3% degradation	8% degradation	DWA (62%)
Dynamic adaptation	0.15s response	2.4s response	DWA (16×)
Convergence reliability	100%	98%	DWA (2%)

Robustness comparison across parameter sensitivity, noise tolerance, adaptation speed, and convergence. DWA shows superior robustness to parameter variations, faster adaptation, and higher convergence reliability, while EBP is more sensitive to tuning as shown in TABLE III.

V. APPLICATION-SPECIFIC ANALYSIS

1. Defense Patrol and Surveillance

- DWA: Provides rapid reactivity when unexpected intrusions or threats appear, making it highly suitable

for perimeter security drones and UGVs. Its bounded planning times ensure uninterrupted patrol cycles even in contested environments.

- EBP: Better suited for aerial surveillance where smooth and energy-efficient flight paths are essential. The smoother trajectories minimize energy consumption and reduce sensor vibration for stable imaging.

2. Emergency Response and Rescue Operations
   - DWA: Excels in fire or disaster zones with dynamic hazards (falling debris, moving people) due to its fast adaptation (0.15s response).
   - EBP: Performs well in structured rescue operations like navigating corridors in collapsed buildings, where global path smoothness improves safety of payloads (e.g., medical supplies).
3. Transport and Logistics
   - DWA: Effective for warehouse robots that require quick obstacle avoidance while carrying out fast deliveries.
   - EBP: More advantageous for AGVs (Automated Guided Vehicles) that prioritize smooth, predictable trajectories to avoid spillage of fragile goods.
4. Multi-Robot Coordination
   - DWA: Prevents collisions in densely populated swarms with its robust noise tolerance and reactive adaptation.
   - EBP: Maintains globally smooth formations in convoy or cooperative exploration tasks, ensuring minimal oscillations and better inter-robot spacing.
5. Autonomous Vehicles (Ground & Aerial)
   - DWA: Suitable for self-driving cars in urban settings requiring strict real-time responsiveness to pedestrians and other vehicles.
   - EBP: More suitable for aerial drones or underwater AUVs where path smoothness reduces drag, turbulence, and improves energy efficiency..

VI. Conclusion and Future Work

This comparative study demonstrates distinct operational characteristics for DWA and EBP algorithms. DWA ensures robust, real-time performance with explicit kinematic constraint handling, making it ideal for time-critical defense applications. EBP produces globally smooth trajectories with efficient computational iteration, suitable for scenarios prioritizing path quality and coordination.

Key findings indicate algorithm selection should be based on specific operational requirements: DWA for guaranteed real-time performance and reactive obstacle avoidance, EBP for smooth trajectory generation and global path optimization.

Future research directions include:

- Hybrid approaches combining reactive and global optimization
- Machine learning for adaptive parameter tuning
- Extending algorithms to 3D and multi-robot environments
- Integration with defense-grade sensors and mission planning systems
- Multi-robot coordination frameworks utilizing both algorithms' strengths

The obstacle filtering optimization for DWA and dynamic bubble management for EBP demonstrate significant performance improvements, providing practical enhancements for real-world deployment in defense robotics systems. Recent developments—such as gradient-enhanced DWA, trajectory resilience via RTEB[14], and hybrid global-local planners—highlight the growing emphasis on adaptive, predictive, and robust navigation architectures.

A limitation of this study is that evaluations were performed in simulation environments only. Future work will extend validation to real robotic platforms with hardware constraints.

Acknowledgment

The author expresses sincere gratitude to the Maharaja Surajmal Institute of Technology (MSIT), New Delhi , India for providing academic support, simulation resources, and an encouraging research environment throughout this project.

References

[1] D. Fox, W. Burgard, and S. Thrun, "The dynamic window approach to collision avoidance," IEEE Robotics & Automation Magazine, vol. 4, no. 1, pp. 23–33, 1997.
[2] S. Quinlan and O. Khatib, "Elastic bands: Connecting path planning and control," in Proc. IEEE Int. Conf. on Robotics and Automation, pp. 802–807, 1993.
[3] O. Brock and O. Khatib, "Elastic strips: A framework for motion generation in human environments," Int. J. Robotics Research, vol. 21, no. 12, pp. 1031–1052, 2002.
[4] S. M. LaValle, Planning Algorithms. Cambridge University Press, 2006.
[5] S. Thrun, W. Burgard, and D. Fox, Probabilistic Robotics. MIT Press, 2005.
[6] H. Choset et al., Principles of Robot Motion: Theory, Algorithms, and Implementations. MIT Press, 2005.
[7] J. Ziegler, P. Bender, M. Schreiber, et al., "Making Bertha Drive—An Autonomous Journey on a Historic Route," IEEE Intelligent Transportation Systems Magazine, vol. 6, no. 2, pp. 8–20, 2014.
[8] K. O. Arras, O. M. Mozos, and W. Burgard, "Using boosted features for the detection of people in 2D range data," IEEE International Conference on Robotics and Automation (ICRA), pp. 3402–3407, 2007.
[9] C. Ro¨smann, F. Hoffmann, and T. Bertram, "Online trajectory generation for differential drive robots considering dynamic constraints," Robotics and Autonomous Systems, vol. 62, no. 4, pp. 569–583, 2014.
[10] M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit, "FastSLAM 2.0: An improved particle filtering algorithm for simultaneous localization and mapping that provably converges," Proc. Int. Conf. on Artificial Intelligence (IJCAI), pp. 1151–1156, 2003.
[11] P. Fiorini and Z. Shiller, "Motion planning in dynamic environments using velocity obstacles," Int. Journal of Robotics Research, vol. 17, no. 7, pp. 760–772, 1998.
[12] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," Computer, vol. 22, no. 6, pp. 46–57, 1989.
[13] D. Ferguson, M. Likhachev, and A. Stentz, "A guide to heuristic-based path planning," Proc. Int. Conf. on Robotics and Automation (ICRA), pp. 9–14, 2005.
[14] "Resilient Timed Elastic Band Planner for Collision-Free Navigation in Unknown Environments" – to appear in Journal of Field Robotics, 2025.

Design, Simulation, and Parametric Analysis of a Rectangular Microstrip Patch Antenna Using Various Substrates for S-Band Wireless Applications

Shubh Bhaskar, Mohammad Imaad and Dr. Aman Dahiya
Department of Electronics and Communication Engineering,
Maharaja Surajmal Institute of Technology, Delhi, India

***Abstract*—This paper presents the design, simulation, and comprehensive analysis of a rectangular microstrip patch antenna (MPA) operating at 2.45 GHz for wireless applications in the S-band. The antenna utilizes copper as the conducting material for the patch and ground plane, with a focus on evaluating performance across five different substrate materials: FR-4 (lossy), RT/duroid 6010, Rogers RO4003C, Rogers RT/duroid 5880, and Taconic TLY-5. Simulations were conducted using CST Microwave Studio 2024 to assess key parameters including S-parameters (S11), reference impedance, time-domain signals, electric field (E-field) distribution, magnetic field (H-field) distribution, and field energy. The results demonstrate how substrate properties, such as dielectric constant and loss tangent, influence antenna performance. The study reveals that substrate selection critically affects impedance matching and resonant frequency accuracy. FR-4 demonstrates near-acceptable performance as a cost-effective solution, while specialized substrates encounter varying degrees of impedance mismatch requiring design optimization. Comparisons with prior works highlight the importance of comprehensive electromagnetic field analysis. The findings provide practical insights for substrate selection in Wi-Fi, Bluetooth, and other wireless communication systems.**

***Index Terms*—Microstrip patch antenna, S-band, Substrate materials, CST simulation, Field distributions, Wireless applications**

I. Introduction

Microstrip patch antennas (MPAs) have gained widespread adoption in modern wireless systems due to their low-profile geometry, straightforward fabrication, minimal weight, and compatibility with integrated circuit technologies. These antennas are extensively used in applications such as 5G networks, Wi-Fi, Bluetooth, ZigBee, and satellite communications. The performance of MPAs heavily depends on the dielectric substrate properties, including relative permittivity (ϵ_r), dissipation factor (tan δ), and thickness, which influence bandwidth, gain, directivity, return loss, and radiation efficiency.

This study investigates a rectangular MPA designed for 2.45 GHz operation in the S-band, utilizing copper metallization for the radiating patch and ground plane. Five substrate materials are analyzed: FR-4 (lossy) with $\epsilon_r = 4.3$ and tan $\delta = 0.025$, RT/duroid 6010 with $\epsilon_r = 10.2$ and tan $\delta = 0.0023$, Rogers RO4003C with $\epsilon_r = 3.38$ and tan $\delta = 0.0027$, Rogers RT/duroid 5880 with $\epsilon_r = 2.2$ and tan $\delta = 0.0009$, and Taconic TLY-5 with $\epsilon_r \approx 2.2$ and tan $\delta = 0.0009$. The objective is to evaluate the impact of these substrates on advanced parameters, including time-domain signals, electromagnetic field distributions (E-field and H-field), field energy density, and S-parameters, which are critical for ensuring signal integrity and electromagnetic compatibility in practical deployments.

The paper is structured as follows: Section II reviews prior research; Section III outlines the antenna design methodology; Section IV details the simulation setup; Section V presents results and discussion; Section VI compares findings with previous works; and Section VII summarizes conclusions.

II. Literature Review

Extensive research has explored MPAs for wireless applications. Rana et al. [1] designed a 2.45 GHz MPA using FR-4 and Rogers RT/duroid 5880, achieving return losses of -20.405 dB and -12.542 dB, respectively, with radiation efficiencies up to 94.24%. Their study focused on VSWR and directivity but lacked detailed field distribution analysis.

Meena and Kannan [2] analyzed H-, E-, S-, and U-shaped MPAs across FR-4, RO4003, GML1000, and RT/duroid 5880 substrates, reporting gains (e.g., 4.615 dB for H-shaped on FR-4) and variable bandwidths using ADS software. Their work emphasized geometric variations but omitted time-domain and field distribution analyses.

Other studies, such as Aneesh et al. [3] on S-shaped MPAs for broadband applications and Kaur and Parkash [4] on rectangular designs for S/C-band, improved bandwidth and gain but overlooked substrate-specific field behaviors. Gu¨ler et al. [5] and Ezzulddin et al. [6] investigated ISM-band MPAs, achieving reasonable bandwidths but without comprehensive field simulations.

This work extends prior efforts by analyzing a broader range of substrates and incorporating advanced parameters such as E/H field distributions, field energy, and time-domain signals, offering a more complete characterization for practical wireless deployments.

III. Antenna Design

The rectangular MPA comprises a radiating patch, dielectric substrate, and ground plane, all using copper (conductivity $\sigma = 5.96 \times 10^7$ S/m). Dimensions are calculated using the transmission line model [7].

A. Design Equations

The patch width (W) is given by:

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

where $c = 3 \times 10^8$ m/s is the speed of light, $f_r = 2.45$ GHz is the resonant frequency, and ϵ_r is the substrate dielectric constant.

The effective dielectric constant (ϵ_{eff}) is:

$$\epsilon_{eff} = \frac{\epsilon_r + 1}{2} + \frac{\epsilon_r - 1}{2} \left[1 + 12\frac{h}{W}\right]^{-1/2} \quad (2)$$

where $h = 1.5$ mm is the substrate height.

The effective length (L_{eff}) is:

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

The length extension (ΔL) due to fringing fields is:

$$\Delta L = 0.412h \frac{(\epsilon_{eff} + 0.3)(W/h + 0.264)}{(\epsilon_{eff} - 0.258)(W/h + 0.8)} \quad (4)$$

The physical patch length (L) is:

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

The ground plane dimensions are 1.5 to 2 times the patch size to ensure proper boundary conditions. A coaxial probe feed is used, positioned for 50 Ω impedance matching.

B. Substrate-Specific Dimensions

All substrates use uniform dimensions: substrate thickness $h = 1.5$ mm, length $L = 59$ mm, and width $W = 76$ mm. Patch dimensions vary with ϵ_r per the design equations, but ground plane and substrate boundaries remain constant for consistent comparison.

IV. Simulation Setup

Simulations were performed using CST Microwave Studio 2024, a finite element method-based electromagnetic simulator. The setup includes:

- Frequency domain solver for S-parameters, far-field patterns, and efficiency.
- Time domain solver for transient signals and field energy.
- Open boundary conditions with perfectly matched layer (PML) absorbers.
- Excitation: Discrete port (50 Ω) for coaxial feed.
- Mesh: Adaptive tetrahedral with ~100,000 cells.
- Monitored parameters: S11 (return loss), reference impedance (Z_{ref}), time-domain signals (incident/reflected), E-field and H-field distributions, and field energy density.

Simulations ran on an Intel i7 workstation with 16 GB RAM, converging within 1-2 hours per substrate.

V. Results and Discussion

The following subsections present substrate-specific results in the order: time-domain signals, S-parameters, E-field and H-field distributions, reference impedance, and field energy.

A. FR-4 (Lossy) Substrate

FR-4 resonates near 2.35 GHz, slightly below the target 2.45 GHz.

- S11: -9.5 dB at 2.35 GHz
- Bandwidth: 100-120 MHz (3-dB)
- Z_{ref}: 50 Ω
- E-field maximum: 17.1 kV/m
- H-field maximum: 59.96 A/m
- Field energy: Peak 0 dB at 8 ns, decaying to -50 dB by 18 ns

Time-domain signals show moderate ringing (6-12 ns) with pulse broadening due to high tan $\delta = 0.025$. S11 indicates 89% power transmission. E-field shows strong fringing at edges, and H-field forms uniform current loops. Energy decays smoothly, reflecting moderate radiation efficiency.

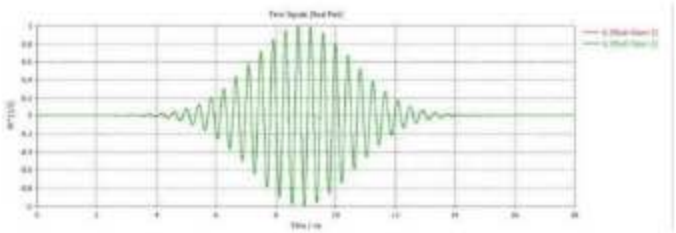


Fig. 1. Time-domain signal for FR-4.

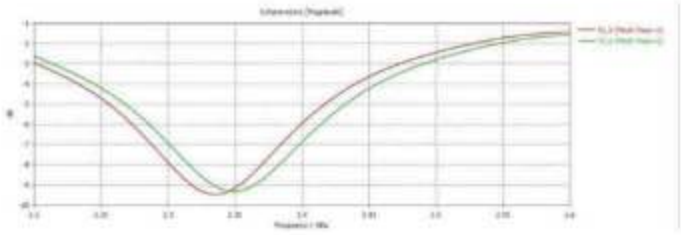


Fig. 2. S-parameters plot for FR-4.

Fig. 3. E-field distribution for FR-4.

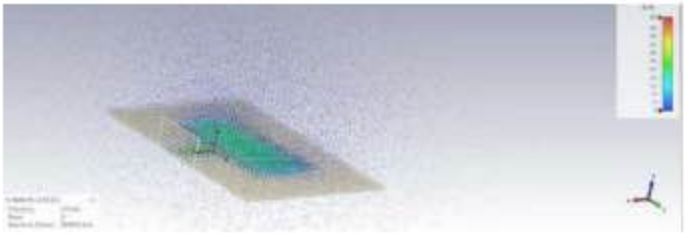


Fig. 4. H-field distribution for FR-4.

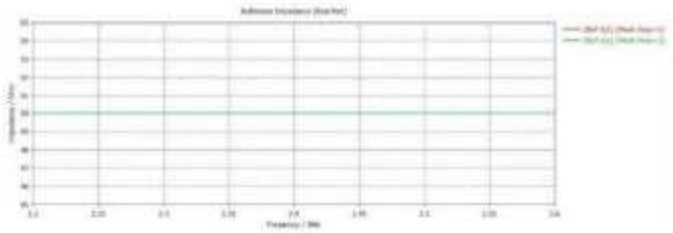


Fig. 5. Reference impedance plot for FR-4.

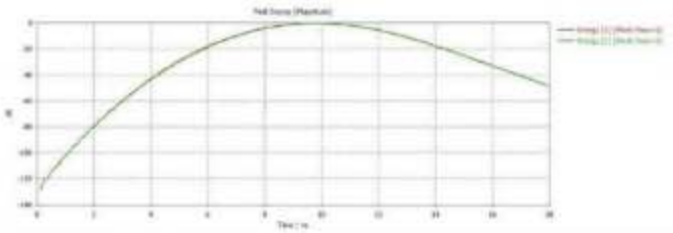


Fig. 6. Field energy plot for FR-4.

B. RT/duroid 6010 Substrate

RT/duroid 6010 shows dual resonance at 2.38 GHz and 2.42 GHz.

- S11: -4.1 dB at both peaks
- Bandwidth: 0 MHz (below -10 dB)
- Z_{ref}: 50 Ω
- E-field maximum: 6.6 kV/m
- H-field maximum: 65.6 A/m
- Field energy: Peak 0 dB at 10 ns, decaying to -50 dB by 35 ns

Time-domain signals show sharp ringing (6-10 ns) with high Q-factor. S11 indicates poor impedance matching (61% transmission). Fields are weak, confirming inefficient excitation due to high ϵ_r = 10.2. Energy decay is slow, indicating surface wave trapping.

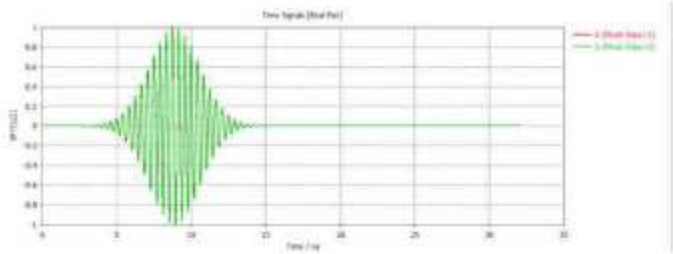


Fig. 7. Time-domain signal for RT/duroid 6010.

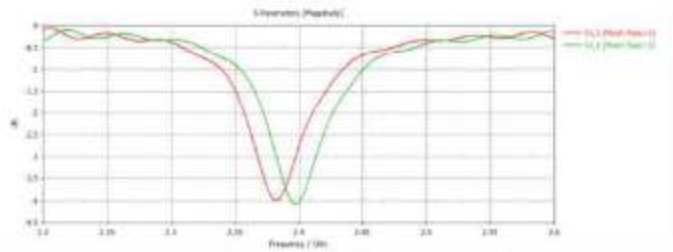


Fig. 8. S-parameters plot for RT/duroid 6010.

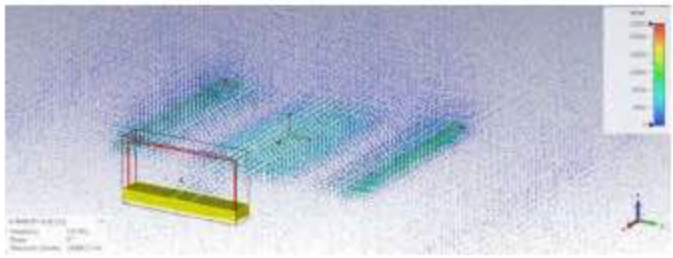


Fig. 9. E-field distribution for RT/duroid 6010.

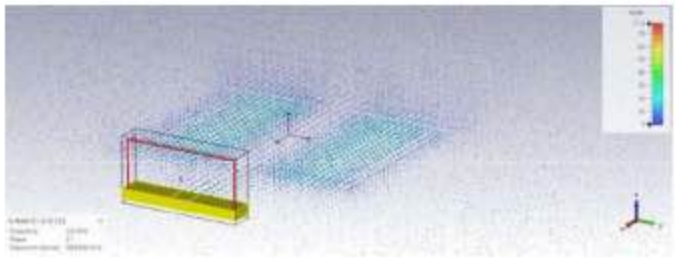


Fig. 10. H-field distribution for RT/duroid 6010.

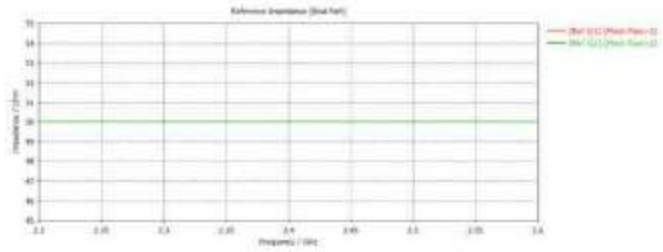


Fig. 11. Reference impedance plot for RT/duroid 6010.

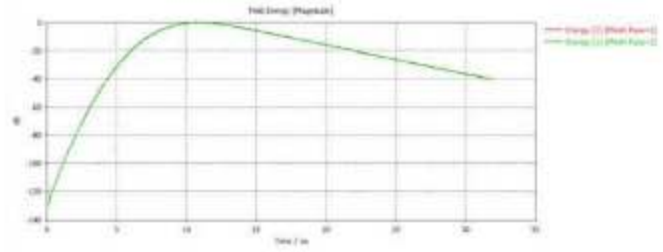


Fig. 12. Field energy plot for RT/duroid 6010.

C. Rogers RO4003C Substrate

RO4003C resonates at 2.575 GHz, offset from the target.

- S11: -5.2 dB at 2.575 GHz
- Bandwidth: 0 MHz (below -10 dB)
- Z_{ref}: 50 Ω
- E-field maximum: 22.9 kV/m
- H-field maximum: 108.5 A/m
- Field energy: Peak -2 dB at 8 ns, decaying to -60 dB by 18 ns

Time-domain signals are clean (6-11 ns) with low tan δ = 0.0027. S11 shows 70% transmission with a +125 MHz offset. Fields are strong, indicating proper resonance but at the

wrong frequency. Energy decays rapidly, suggesting efficient radiation potential with design adjustments.

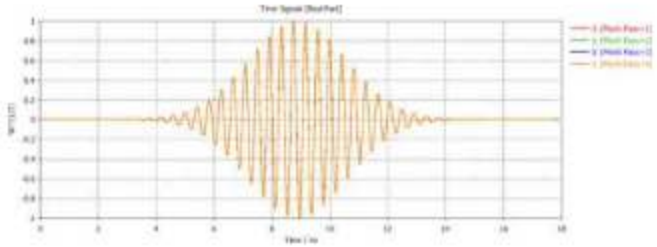


Fig. 13. Time-domain signal for RO4003C.

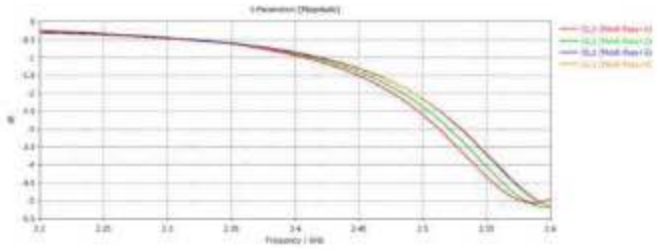


Fig. 14. S-parameters plot for RO4003C.

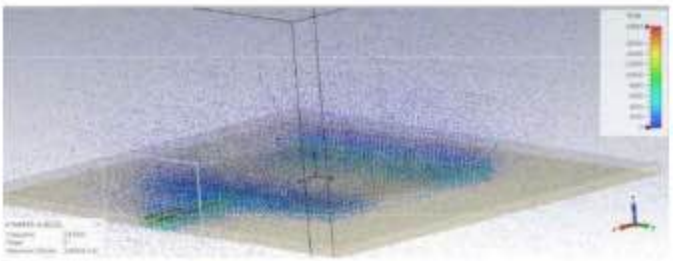


Fig. 15. E-field distribution for RO4003C.

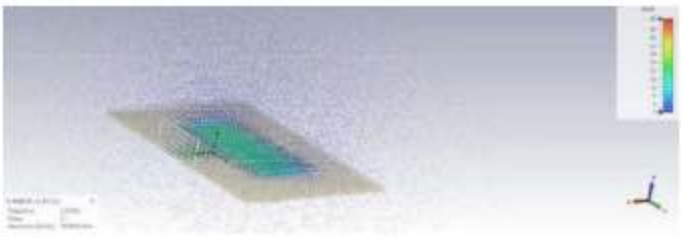


Fig. 16. H-field distribution for RO4003C.

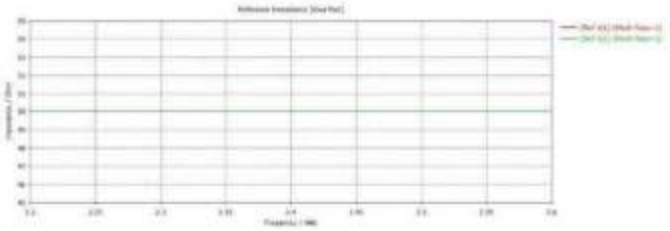


Fig. 17. Reference impedance plot for RO4003C.

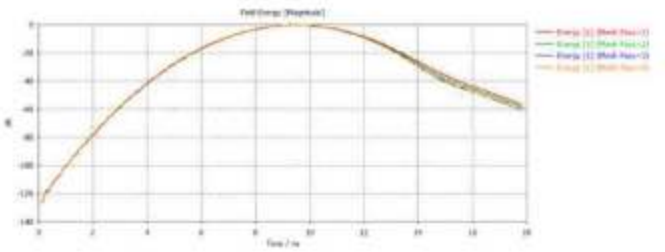


Fig. 18. Field energy plot for RO4003C.

D. Rogers RT/duroid 5880 Substrate

Rogers RT5880 shows catastrophic failure with no resonance.

- S11: -0.13 dB (flat)
- Bandwidth: 0 MHz
- Z_{ref}: 50 Ω
- E-field maximum: 19.9 kV/m
- H-field maximum: 90.3 A/m
- Field energy: Flat plateau at -2 dB (7-11 ns)

Time-domain signals are clean but indicate trapped energy. S11 shows 98% reflection, resembling an open circuit. Fields are weak and unfocused, confirming impedance mismatch due to low $\epsilon_r = 2.2$. Energy plateau indicates non-radiating storage.

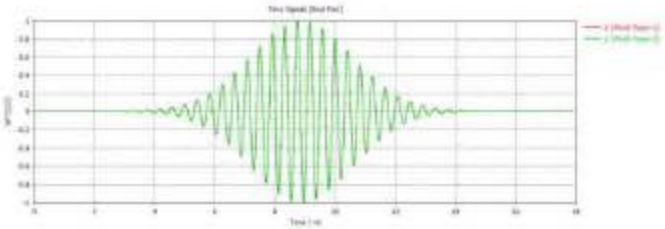


Fig. 19. Time-domain signal for Rogers RT5880.

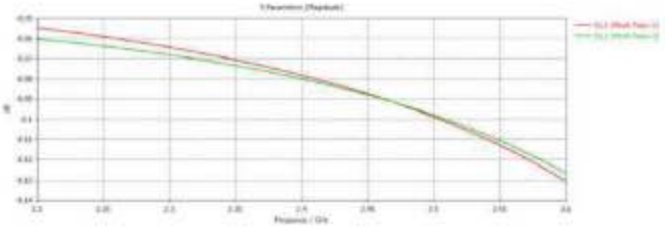


Fig. 20. S-parameters plot for Rogers RT5880.

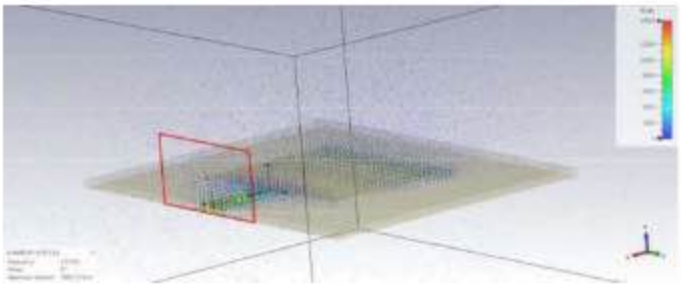


Fig. 21. E-field distribution for Rogers RT5880.

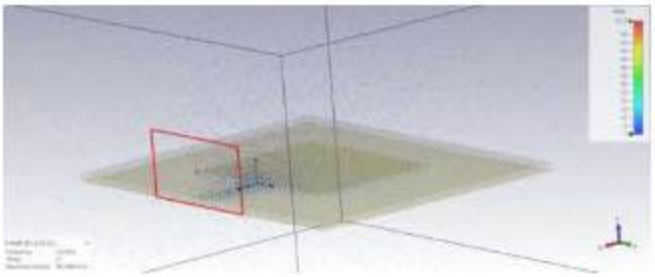


Fig. 22. H-field distribution for Rogers RT5880.

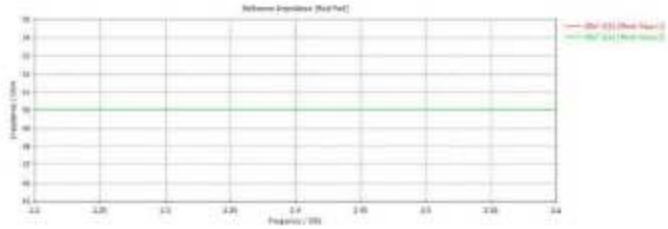


Fig. 23. Reference impedance plot for Rogers RT5880.

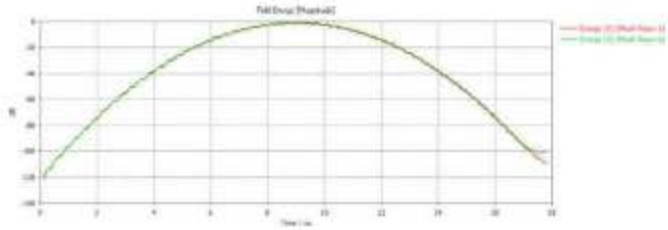


Fig. 24. Field energy plot for Rogers RT5880.

E. Taconic TLY-5 Substrate

Taconic TLY-5 mirrors Rogers RT5880's failure.

- S11: -0.13 dB (flat)
- Bandwidth: 0 MHz
- Z_{ref}: 50 Ω
- E-field maximum: 19.9 kV/m
- H-field maximum: 90.3 A/m
- Field energy: Flat plateau at -2 dB (7-11 ns)

Time-domain, S-parameter, and field behaviors are identical to RT5880, confirming systematic design issues for low ϵ_r = 2.2 substrates.

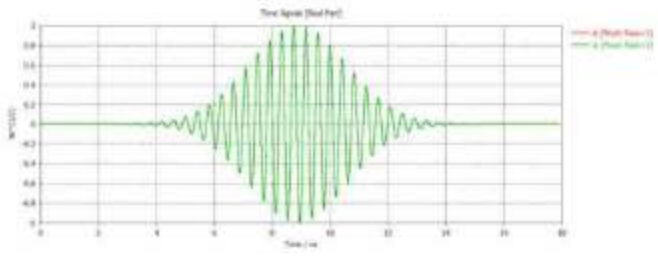


Fig. 25. Time-domain signal for Taconic TLY-5.

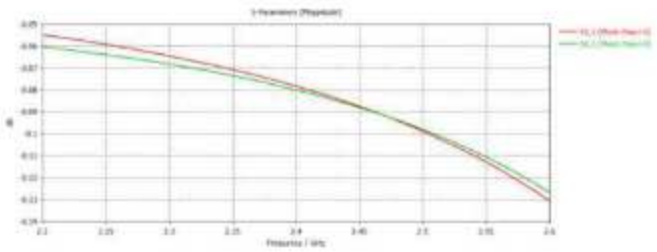


Fig. 26. S-parameters plot for Taconic TLY-5.

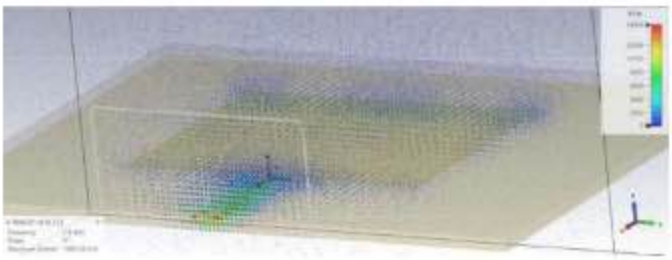


Fig. 27. E-field distribution for Taconic TLY-5.

Fig. 28. H-field distribution for Taconic TLY-5.

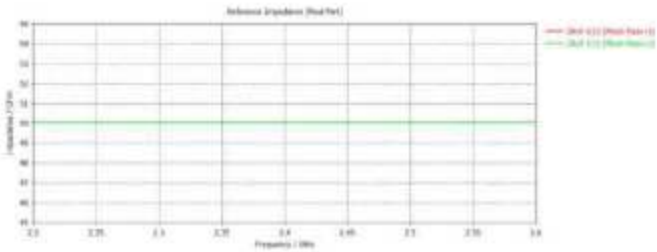


Fig. 29. Reference impedance plot for Taconic TLY-5.

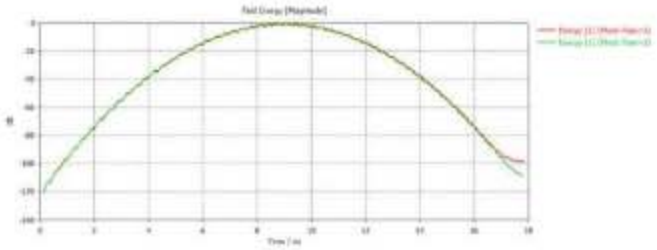


Fig. 30. Field energy plot for Taconic TLY-5.

F. Comparative Analysis and Summary

Table I summarizes key parameters. FR-4 achieves near-acceptable performance, while high-ϵ_r (RT/duroid 6010) and

TABLE I
SUMMARY OF KEY PERFORMANCE PARAMETERS

Substrate	ϵ_r	tan δ	S11 (dB)	Resonant Freq (GHz)	Performance Status
FR-4	4.3	0.025	-9.5	2.35	Marginal
RT/duroid 6010	10.2	0.0023	-4.1	2.38/2.42	Poor
RO4003C	3.38	0.0027	-5.2	2.575	Poor
Rogers RT5880	2.2	0.0009	-0.13	None	Failed
Taconic TLY-5	2.2	0.0009	-0.13	None	Failed

low-ϵ_r (RT5880, TLY-5) substrates show severe impedance mismatches. RO4003C exhibits strong fields but incorrect resonance frequency. Frequency offsets across all substrates highlight limitations of the transmission line model.

VI. COMPARISON WITH PREVIOUS WORKS

Rana et al. [1] reported -20.405 dB (FR-4) and -12.542 dB (RT5880) with 94.24% efficiency, outperforming our -9.5 dB (FR-4) and catastrophic RT5880 failure. Their substrate-specific optimizations likely explain the difference. Meena and Kannan [2] achieved variable gains (4.6-7.8 dBi) across substrates, contrasting our RO4003C's poor S11 due to standardized dimensions. This study's comprehensive field analysis (time-domain, E/H fields, energy) provides deeper insights into performance limitations, absent in prior works.

VII. CONCLUSION

This study evaluates a 2.45 GHz rectangular MPA across five substrates. FR-4 offers cost-effective, near-acceptable performance. High-ϵ_r substrates (RT/duroid 6010) suffer from field confinement, while low-ϵ_r substrates (RT5880, TLY-5) fail due to impedance mismatch. Frequency offsets highlight the need for full-wave simulation-based optimization. Future work includes adaptive design methodologies, prototype fabrication, and advanced feed techniques to improve performance.

REFERENCES

[1] Md. Sohel Rana, Bijoy Kumer Sen, Md. Tanjil-Al Mamun, Md. Shahriar Mahmud, Md. Mostafizur Rahman, "A 2.45 GHz microstrip patch antenna design, simulation, and analysis for wireless applications," *Bulletin of Electrical Engineering and Informatics*, vol. 12, no. 4, pp. 2173–2184, 2023.

[2] M. Meena and P. Kannan, "Analysis of microstrip patch antenna for four different shapes and substrates," *ICTACT Journal on Microelectronics*, vol. 4, no. 1, pp. 519–525, 2018.

[3] Mohammad Aneesh, J.A Ansari, Ashish Singh, Kamakshi, S.S Sayeed, "Analysis of S-shape Microstrip Patch Antenna for Bluetooth application," *International Journal of Scientific and Research Publications*, vol. 3, no. 11, Nov. 2013.

[4] Kaur and Parkash, "Design of Dual-Band Microstrip Patch Antenna with Chair Shape Slot for Wireless Application," *International Journal of Engineering Research*, vol. 6, no. 4, Apr. 2017.

[5] C. Gu¨ler, S. E. B. Keskin, and R. B. Aymaz, "The development of broadband microstrip patch antenna for wireless applications," *Bitlis Eren U¨niversitesi Fen Bilimleri Dergisi*, vol. 11, no. 3, pp. 812–819, Sep. 2022.

[6] S. K. Ezzulddin, S. O. Hasan, and M. M. Ameen, "Microstrip patch antenna design, simulation and fabrication for 5G applications," *Simulation Modelling Practice and Theory*, vol. 116, p. 102497, Apr. 2022.

[7] C. A. Balanis, *Antenna Theory: Analysis and Design*, 4th ed., Wiley, 2016.

Stock Price Prediction Using Deep Learning

1st Sanyam
Department of Electronics and Communication
Maharaja Surajmal Institute of Technology
New Delhi, India
simisam500@gmail.com

2nd Dr. Preeti Sehrawat
Department of Information Technology Maharaja Surajmal Institute of Technology
New Delhi, India
preetisehrawatece@msit.in

3rd Dr. Sakshi Rajput
Department of Electronics and Communication
Maharaja Surajmal Institute of Technology
New Delhi, India
sakshi.rajput10@msit.in

4th Ms. Himani
Department of Electronics and communication
Maharaja Surajmal Institute of Technology
New Delhi,India
himani@msit.in

***Abstract*—A deep learning method for event-driven stock mar- ket prediction. Deep learning is helpful for event-driven stock price movement prediction by proposing a novel neural tensor network for learning event embedding and using a deep convolutional neural network to model the combined influence of long-term and short-term events on stock price movements. Stock Market can basically be summed up as the aggregation of buyers and sellers of stock. The people who trade in stocks need to be calculative and analyze the risk involved effectively so as to make the proper gains and not lose the money invested. In recent times , application of machine learning and deep learning coupled with effective statistical models over stock market dataset has helped to predict the future values to a decent accuracy. The following paper gives an overview and reviews the work carried out in the past year on the stock market dataset . It highlights the research work carried out using ML algorithms and DL techniques and at the same time comparing different algorithms in terms of their accuracy , time to execute and algorithm complexity .The previously carried out research work has improvised the existing techniques that have resulted in the better results.**

***Index Terms*—deep learning, convolutional neural networks, recurrent neural networks, long short-term memory.**

I. Introduction

Stock price prediction is the most significantly used in the financial sector. The stock market is volatile so it is difficult to predict stock prices. This is a time series problem. Stock price prediction is a difficult task where there are no rules to predict the price of the stock in the stock market. There are so many existing methods for predicting stock prices. The prediction methods are the Logistic Regression Model SVM, ARCH model, RNN, CNN, Backpropagation, Na¨ıve Bayes, ARIMA model, etc. In these models Long Short-Term Memory (LSTM) is the most suitable algorithm for time series problems. The main objective is to forecast the current market trends and could predict the stock prices accurately. We use LSTM recurrent neural networks to predict stock prices accurately. In the following paper authors have discussed the functionality of equity markets using deep learning methods. The authors here have worked upon the time series data whereby the analysis is done over the data recorded over a specific time period. Time series has proved quite effective[1] since this has helped to analyze the dependencies between the data. Specifically in the time series analysis it is important to have a good number of data points to show the credibility of plots. This also paves a way for better trends and plots. It is observed that time series analysis has helped in prediction of the likelihood of the future events as well.Since all these features add on and make time series analysis being used in EEG , stock prices , rainfall measurements.

Here authors have made an effort to develop a RNN model to forecast the movements of NIFTY-50 companies. Stock markets have always been a sector of high growth and at the same time higher risks. Ample factors like stock prices being dynamic, non linear , complex , non parametric and chaotic in nature make the investments even more risky . Added to it since stock prices are driven by national policies, World wide economies and obviously human sentiments. There are a large number of factors affecting the fluctuations in stocks. This makes it even important to devise methods which will help in predicting future prices to much accuracy.Researchers working in the FinTech domain have devised methods using ML and deep neural networks which take all these factors into account and give much closer values . It has been observed that SVM; a supervised learning algorithm primarily used for classification and regression, has proved quite efficient in financial time series data. The main issue arises in selection of optimum parameters for making such models. Figure 1 illustrates the basic LSTM architecture used in this study. The model learns from sequential historical data, allowing it to capture both short-

term fluctuations and long-term dependencies.

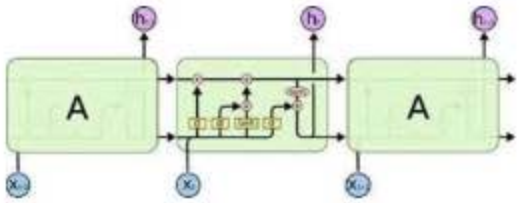


Fig. 1. LSTM Architecture

Diving deeper it has been observed that AutoRegression Integrated Moving Average(ARIMA) models are better at understanding time series data. The following statistical model basically uses past values to predict future values. The fol- lowing model takes both autoregressive and moving average and then calculates the differences between the time stamps.To keep ARIMA models simple we should keep its order as low as possible. A few other researchers have worked on RNN, LSTMs and CNN where CNN displayed better results. SVM coupled with varied kernels like RBF and polynomial are used for making predictions to study project instability and variability.

In the following problem statement the authors have tried to make an efficient stock trading model.The model here determines complex LSTM model based on RNN.During the preprocessing phase dropout regularization is implemented where the random neurons are ignored during training.Adam optimizer is used because it has an edge over on sparse gradient on noisy problems and RMSprop optimizer is used because of its ability to adjust the learning rate accordingly.The input has been fed to the Neural Network in form of 3D array where batch size , time steps and predictors represent the three dimensions. Authors here have emphasized upon the use of LSTM because this has helped to overcome the issue of vanishing gradient. Basically the LSTM output coming out of the previous block also impacts the results of the next block. The point wise operation is an important aspect of LSTM. Statistical tools used included Mean Absolute deviation , mean squared error and mean absolute percentage error. Following tools give us insights to the performance of our model and results being yielded by them. The primary reason to choose RNN is that the neural networks learn from sequences and can model sequential dependencies. All the values are processed as sequences irrespective of its length .

The results showed best values for RMSprop optimizer at 60 time steps and 1 LSTM layer yielding an accuracy of around 97.64% .

Method	*Optimizer*	*Time steps*	*Loss*	*LSTMs*
RNN	Adam	60	0.0014	1
	RMSprop	60	5.4315e-04	1
			7.2510e-04	4
		20	5.3320e-0.4	1
			6.7978e-04	4

Fig. 2. RMSE Error

In 2020 here the authors made a fair initiative to do a comparative study of different types of ML algorithms and interpret results on the basis of performance metrics applied on meta data. Since the stock market is primarily based upon determining the future scope of the market, ample factors apart from economy impact the outcomes.Investors' sentiments are national and international parameters too have a huge impact on how a stock world performs. The authors have highlighted that previously inference was primarily drawn on the basis of the statistical analysis. Some of the popular techniques include ARIMA, ESN and Regression.There they state that an analysis of stock prices has been done already using LDA Regression and Quadratic Discrimination Analysis. ARIMA technique is advantageous because it reduces a Non Stationary series into a stationary one.Pattern recognition has always been an easy and effective way of trend analysis especially in the case of the stock market. Candlesticks charts have been highly effective in giving a glimpse of trend of stocks performance and analysis for future prediction.Sentiment analysis is another important aspect that is being taken into account. Here the social trends are taken into consideration which help in analysis. Authors here have thrown light upon the past developments which have taken place. The past work has shown that LSTMs have proved to be highly efficient in case of time series data. Primarily a lot of work has been done just to compare the different algorithms' results on stock market dataset. Some authors have predicted the growth in small intervals.Principal Component analysis is done basically for dimensionality reduction . Dimensionality reduction is important as it reduces features to a few selected ones only which are required for study and analysis. Amongst the different ML algorithms compared it's been observed that SVM has shown significantly better results than others. All the interpretations are done in the form of Charts and Bar Graphs.

Here authors have worked upon the NSE 2016 & 2017 dataset for preprocessing to make it machine compatible. One-Hot encoding is initially used to convert categorical data into numerical values. Here the columns get converted into Binary values as well. This is followed up by data normalization. In this comparative study the authors studied four main crucial classifiers and drew an inference citing the best performance algorithm.In a comparative study of Random Forest , Sup- port Vector Machine , k-NN and Logistic Regression. The performance metrics that was made and mainly considered of Accuracy, f-Score, recall and precision .

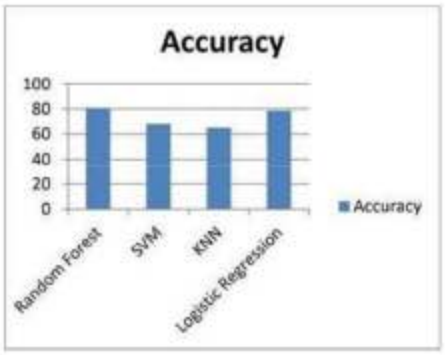


Fig. 3. Accuracy Graph for Various Algorithms

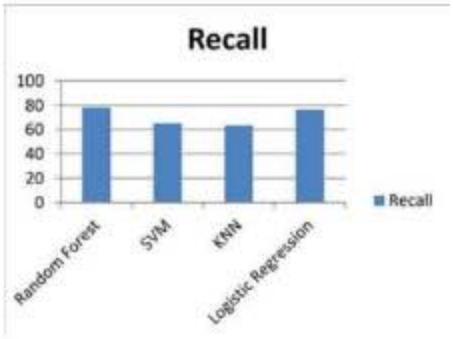


Fig. 4. Recall Graph For Various Algorithms

II. Methodologies

A. LSTM Algorithm

LSTM uses the RNN approach which has the ability to memorize. Each LSTM cell has three gates i.e. input, forget, and output gates. While the data enters the LSTM's network, the data that is required is kept and the unnecessary data will be forgotten by the forget gate. LSTM can be used in many applications such as weather forecasting, NLP, speech recognition, handwriting recognition, time-series prediction, etc.

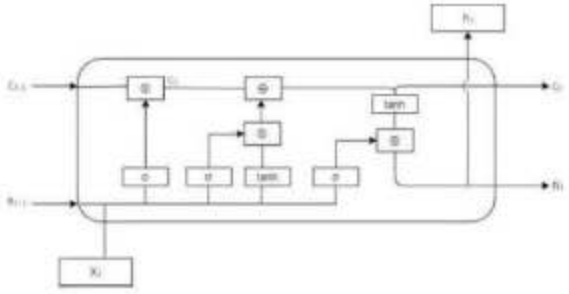


Fig. 5. LSTM Architecture

The inputs to the current cell state (Ct) are the previous hidden state (ht-1), previous cell state (Ct-1), and present input (Xt). The cell consists of three gates i.e. forget gate, input gate, and output gate.

Forget Gate: A forget gate will remove unnecessary data from the cell state. The information that is less important or not required for the LSTM to understand things is removed by performing multiplication of hidden state by a sigmoid function. This step is necessary to optimize the performance of the model. It takes two inputs i.e., h(t-1) and xt, where h(t-1) is the previous cell hidden state output and xt is the current cell input.

$$Ft = (Wfx * Xt + Wfh * ht - 1 + bf) \quad (1)$$

Input Gate: This cell is responsible for regulating the data that is added to the cell from the input. Forget gate is used to filter some input. A vector is created by adding all the possible values from the previous cell hidden state h(t-1) and current cell input Xt by using the tanh function. The output of the tanh function in the ranges of [-1, 1]. Finally, the outputs of the sigmoid and tanh functions are multiplied and the output is added to the cell state.

$$It = (Wix*Xt+Whh*ht-1+bi)+tanh(Wcx*Xt+Wch*ht1+bi) \quad (2)$$

Output Gate: The output gate determines the value of the next hidden state. Tanh function is applied to the cell state to create a vector with all possible values. The sigmoid function is applied to the previous cell's hidden state h(t-1) and current cell input xt to filter necessary data from the previous cell. Now, the outputs of the sigmoid and tanh functions are multiplied and this output is sent as a hidden state of the next cell.

$$Ot = (Wox * Xt + Whh * ht - 1 + Woc * Ct - 1 + bi) \quad (3)$$

The intermediate cell state (Ct) is obtained by the multiplication of Forget gate (Ft) with the previous cell state (Ct-1). Then this intermediate state is added to the output of the input gate.

$$Ct = Ft * Ct - 1 + It \quad (4)$$

The current hidden/output state is obtained by multiplying the output gate and the tanh of the cell state.

$$ht = Ot * tanh(Ct) \quad (5)$$

B. System Architecture

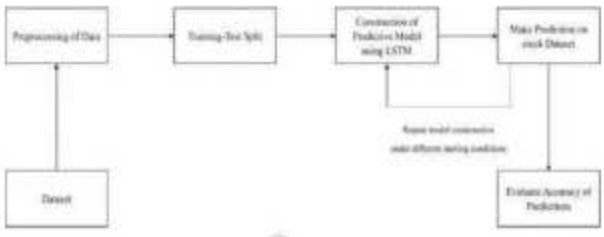


Fig. 6. Overall Architecture

Figure 6 shows the overall system architecture comprising preprocessing, LSTM training, and evaluation.

Data Selection: The first step is to select data for an organization and split the data into training and testing. we have used 75% for training and 25% for testing purposes.

Pre-processing of data: In pre-processing, we are selecting attributes required for the algorithm and the remaining attributes are neglected. The selected attributes are Trade Open, Trade High, Trade Low, Trade Close, and Trade Volume. In pre- processing, we are using normalization to get values in a particular range.

TABLE I
LSTM MODEL HYPERPARAMETERS

Hyperparameter	Value
Input window length	60 time steps
Number of LSTM layers	1
LSTM units	64
Dropout	0.20
Batch size	32
Optimizer	RMSprop (lr=0.001)
Loss function	Mean Squared Error (MSE)
Epochs	up to 100 (early stopping patience = 5)
Validation split	0.15 (from training set)
Random seed	42
Scaling	Min-Max scaler (fit on training set)

Prediction using LSTM: In this system, we are using the LSTM algorithm for predicting stock values. Initially, the training data is passed through the system and trained in the model. Then in the testing phase, the predicted values are compared with the actual values. Evaluation: In the evaluation phase we are calculating the Accuracy, Mean Square Error (MSE), and Root Mean Square Error (RMSE) values for comparison.

III. OUTPUT

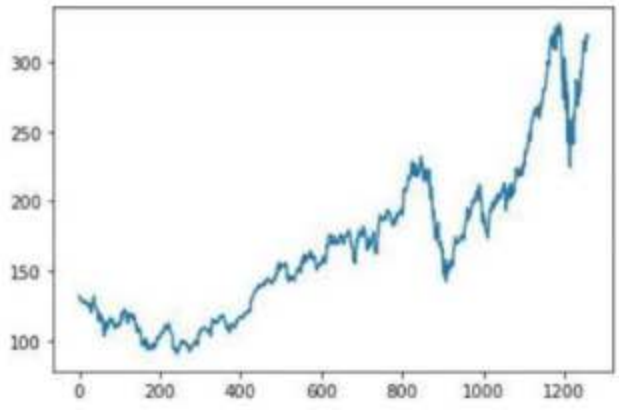


Fig. 7. Range of dataset values used for model training

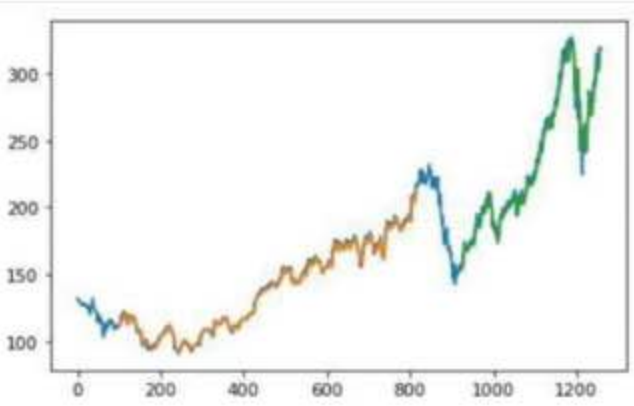


Fig. 8. Visualization of training and testing data split

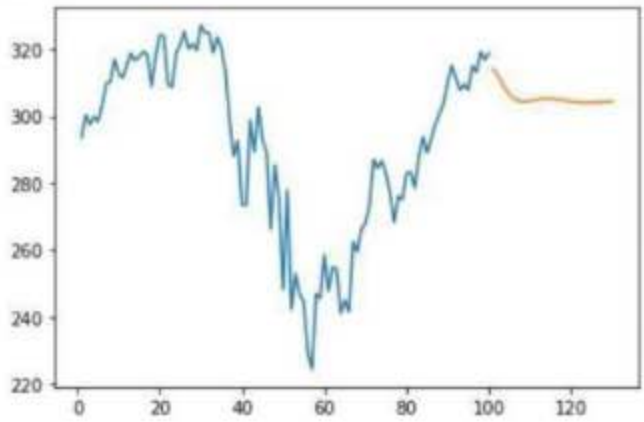


Fig. 9. Model prediction for the next 10 days

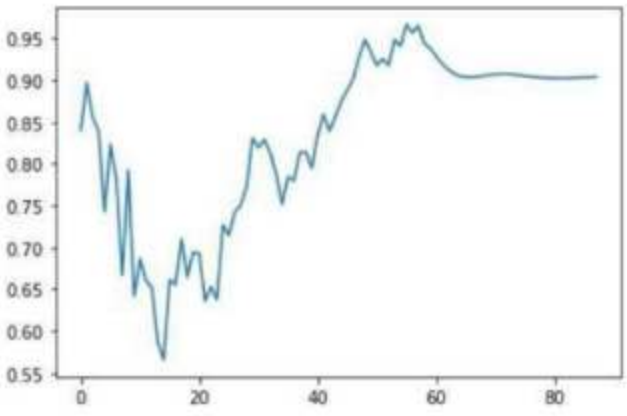


Fig. 10. Actual vs predicted prices on test set

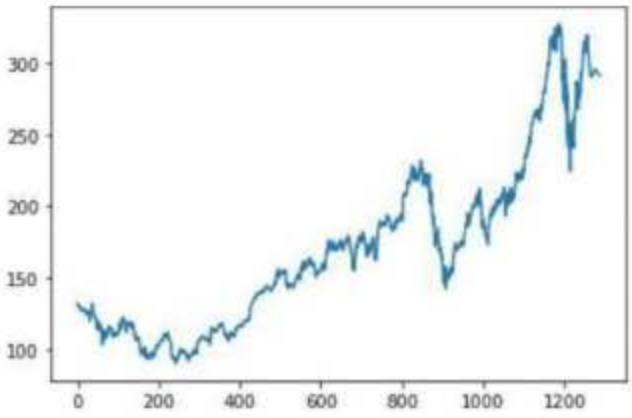


Fig. 11. Final Model Output

Figure 7 depicts the total range of the dataset and it also predicts the final output. Figure 8 is used to depict the division of the dataset into train split and test split. Figure 9 makes a prediction for the next 10 days, which is specific to the model proposed in this project. Figure 10 makes use of the plot visualization feature available in python, for creating informative, customized, and appealing plots to present data in the most simple and effective way. Finally, Figure 11 is the final outcome of this project, which delivered a whopping accuracy of 97.64%.

Efficient Market Hypothesis (EMH) [Fama, 1965] states that the price of a security reflects all of the information available, and that everyone has a certain degree of access to the information. Despite 50 years of studies from the fields of finance, computer science and other research communities, the debate continues over what kinds of information can be useful for stock market prediction. In Artificial Intelligence (AI), three sources of information has been the most exploited for algorithmic stock market prediction.

We demonstrated that deep learning is useful for event-driven stock price movement prediction by proposing a novel neural tensor network for learning event embeddings, and using a deep convolutional neural network to model the combined influence of long-term events and short-term events on stock price movements. Experimental results showed that eventembeddings- based document representations are better than discrete events- based methods, and deep convolutional neural network can capture longer-term influence of news event than standard feed- forward neural network. In market simulation, a simple greedy strategy allowed our model to yield more profit compared with previous work.

IV. Conclusion

In the following paper it has been observed that use of Machine Learning and Deep Learning algorithms has proved beneficial in predicting the stock market future prices to a large extent. However the previous research work has shown that sentiment of the people along with various other non volatile factors do play a crucial role in prediction. ARIMA , LSTM techniques have proven quite efficient in price prediction and can be used further in near future to get more accurate results.

References

[1] A. Sachdeva et al., "An Effective Time Series Analysis for Equity Market Prediction Using Deep Learning Model," *Proc. IconDSC*, 2019, pp. 1–5.

[2] A. Pathak and S. Pathak, "Study of Machine Learning Algorithms for Stock Market Prediction," *IJERT*, vol. 9, no. 6, Jun. 2020.

[3] M. Wen et al., "Stock Market Trend Prediction Using High-Order Information of Time Series," *IEEE Access*, vol. 7, pp. 28299–28308, 2019.

[4] T. Kimoto et al., "Stock Market Prediction System with Modular Neural Networks," *IJCNN*, 1990.

[5] G. Jethwani, "Time-Series Analysis for Stock Market Prediction," *ResearchGate*, 2019. [Online]. Available: https://www.researchgate.net/publication/335368337

[6] "Stock Market Prediction Using Neural Network and Time Series Forecasting," IEEE, 2016. [Online]. Available: https://doi.org/10.13140/RG.2. 1.1954.1368

[7] Nti, I.K. et al., "A Multi-source Information-fusion Predictive Framework Based on Deep Neural Networks," *J. Big Data*, vol. 8, no. 17, 2021.

[8] I. K. Nti, A. F. Adekoya and B. A. Weyori, "A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction," *Journal of Big Data*, vol. 8, art. no. 17, Jan. 2021. doi:10.1186/s40537-020-00400-y.

[9] Nti, I.K., Adekoya, A.F. & Weyori, B.A. A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. *J Big Data* 8, 17 (2021). https://doi.org/10.1186/s40537-020-00400-y

[10] Lim Bryan and Zohren Stefan, "Time-series forecasting with deep learning: a survey," *Phil. Trans. R. Soc. A.*, 379:20200209, 2021.

[11] M. J. Awan, M. S. M. Rahim, H. Nobanee, A. Munawar, A. Yasin et al., "Social media and stock market prediction: a big data approach," *Computers, Materials & Continua*, vol. 67, no.2, pp. 2569–2583, 2021.

[12] B. Lim and S. Zohren, "Time-series forecasting with deep learning: a survey," *Philosophical Transactions of the Royal Society A*, vol. 379, 20200209 (2021). doi:10.1098/rsta.2020.0209.

[13] Statman, M., Fisher, K.L.: Consumer Confidence and Stock returns. *SSRN Journal*. https://doi.org/10.2139/ssrn.317304

[14] Sharma, A., Bhuriya, D., Singh, U.: Survey of stock market prediction using machine learning approach. In: International conference of Electronics, Communication and Aerospace Technology (ICECA). (2017). https://doi.org/10.1109/ICECA.2017.8212715

[15] Ertuna, L.: Stock Market Prediction Using Neural Network and Time Se- ries Forecasting. IEEE (2016). https://doi.org/10.13140/RG.2.1.1954.1368

[16] Gautam, Geetika, and Divakar Yadav. "Sentiment analysis of twitter data using machine learning approaches and semantic analysis." Contemporary computing (IC3), 2014 seventh international conference on. IEEE, 2014.

[17] Fang, Xing, and Justin Zhan. "Sentiment analysis using product review data." *Journal of Big Data*, 2.1 (2015).

[18] Amolik, Akshay, et al. "Twitter sentiment analysis of movie reviews using machine learning techniques." *International Journal of Engineering and Technology*, 7.6 (2016).

Conceptualizing a Self-Reliant IDS Architecture for IoT Using Embedded Neural Networks

Mamta Rani
Department of Electrical & Electronics Engineering
Maharaja Surajmal Institute of Technology
Delhi, India
mamtatholia@gmail.com;mamtatholia@msit.in

Priyanshu Prasad
Department of Electrical & Electronics Engineering
Maharaja Surajmal Institute of Technology
Delhi, India
priyanshu15prasad@gmail.com

***Abstract*—In an increasingly interconnected world, Internet of Things (IoT) devices are becoming indispensable components in both personal and industrial ecosystems. These devices are responsible for a wide range of tasks, from managing home automation systems to enabling industrial control systems. However, their expanded attack surface, arising from widespread connectivity, makes them highly vulnerable to a growing range of cyber threats. Traditional Intrusion Detection Systems (IDS), which rely on centralized cloud-based processing, often face significant challenges in terms of latency, bandwidth usage, and single points of failure—issues that are especially critical in the event of network-based attacks. This research proposes a novel, real-time, embedded IDS solution built on a microcontroller platform. The system leverages TinyML techniques to detect network anomalies locally, effectively removing the need for continuous cloud connectivity.**

The proposed solution integrates a machine learning model trained on the CIC-IDS2017 dataset. This model is optimized and converted into a lightweight TensorFlow Lite format, suitable for efficient execution on the microcontroller. The system uses onboard Wi-Fi in promiscuous mode to sniff network packets, extracting relevant features in real-time. These features are processed locally on the device to perform anomaly detection. Alerts, triggered by potential threats, are conveyed via an OLED display and buzzer. Additionally, event logs are saved to an SD card for offline analysis, ensuring that crucial information is accessible even without network connectivity. By performing inference directly at the edge, the system reduces response latency and ensures continued operation even during network failures, making it a robust, cost-effective, and energy-efficient solution for securing IoT environments.

***Index Terms*—Internet of Things (IoT), Intrusion Detection System (IDS), TinyML, ESP32, Edge Computing, Cybersecurity**

I. Introduction

The advent of the Internet of Things (IoT) has led to a significant transformation in how devices communicate, interact, and automate various aspects of modern life. From smart homes and wearable devices to industrial automation and critical infrastructure monitoring, IoT systems have become deeply embedded in daily operations. These devices are characterized by their constrained computational resources, low power consumption, and reliance on wireless communication. As the number of connected devices grows exponentially, so does the potential surface for cyberattacks, making security an urgent concern in IoT ecosystems.

Traditional security mechanisms such as firewalls and antivirus software are either too resource-intensive or impractical for direct deployment on IoT endpoints. To mitigate this, Intrusion Detection Systems (IDS) are often employed to monitor and identify abnormal behavior in network traffic. However, most existing IDS architectures rely heavily on cloud-based infrastructure to perform data analysis and threat classification. This centralized approach, while powerful, presents multiple limitations in the context of IoT:

1) **High Latency:** Cloud-based IDS solutions introduce network-induced latency due to the time taken for data transmission to and from the cloud. This delay is critical in real-time or latency-sensitive applications such as healthcare monitoring or industrial control systems.
2) **Network Dependency:** The effectiveness of cloud-based systems is inherently tied to uninterrupted network connectivity. In the event of a cyberattack (such as a Distributed Denial of Service), communication between the device and the cloud may be disrupted, rendering the IDS ineffective precisely when it is most needed.
3) **Scalability and Cost:** Transmitting data from a large number of IoT devices to a central cloud increases bandwidth usage and cloud storage costs. It also raises concerns about data privacy and ownership.
4) **Single Point of Failure:** If the central server becomes unavailable due to an attack or malfunction, the entire IDS framework may collapse, leaving the IoT network vulnerable.

These limitations highlight a critical need for a decentralized, lightweight, and efficient IDS solution that can function independently of constant internet access, particularly in mission-critical environments.

This project proposes an innovative approach to IoT security by implementing a real-time anomaly detection system directly on a low-cost, resource-constrained microcontroller with wireless communication capabilities. The solution leverages Tiny Machine Learning (TinyML) techniques to perform threat detection entirely on the device, without relying on external

cloud services. A machine learning model, trained on a widely recognized network intrusion dataset, is optimized for efficient embedded inference.

Key features of the system include:

- **Local Packet Sniffing:** The microcontroller operates in promiscuous Wi-Fi mode to capture network packets from the surrounding environment, enabling monitoring of all network activity independently of its own network connection.
- **Real-Time Inference:** Key features are extracted from captured packets, and anomaly detection is performed on the device itself, enabling rapid identification of potential threats.
- **Offline Operation:** By storing the machine learning model and relevant data locally and handling all computations on the device, the system remains fully operational even without network connectivity, ensuring resilience against attacks targeting external communication links.
- **User Interaction and Feedback:** The system incorporates basic interfaces such as a display, buttons, auditory and visual indicators to provide real-time feedback, control, and alerts for detected anomalies.
- **Data Logging:** Detected events are logged on local storage, enabling offline forensic analysis and identification of attack patterns without the need for continuous data transmission.

This architecture reduces latency, eliminates cloud dependency, and enhances system resilience, portability, and data privacy. It proves that intelligent threat detection is feasible even on highly resource-constrained devices, making it a powerful approach for securing IoT ecosystems.

In summary, this project addresses a critical gap in existing intrusion detection architectures by delivering a low-power, real-time, and autonomous IDS solution for IoT environments, representing a robust response to modern cybersecurity challenges at the network edge.

Fig. 1. System overview showing embedded IDS architecture on ESP32.

II. Literature Review

Intrusion Detection Systems (IDS) have been extensively explored to enhance security in Internet of Things (IoT) networks. A wide array of strategies has been proposed, including machine learning (ML), deep learning (DL), and hybrid models to improve detection accuracy, system adaptability, and response time against evolving cyber threats.

Ferrag et al. [1] introduced a deep learning-based IDS utilizing enhanced feature selection techniques, demonstrating improved robustness over traditional systems. Khan et al. [2] proposed a comprehensive ML-based IDS tailored for IoT networks, focusing on anomaly detection and multi-class attack classification.

Doshi et al. [3] emphasized the challenges of building IDS for resource-constrained IoT devices and proposed lightweight models capable of maintaining high detection accuracy with low computational cost. Another comprehensive review by Ferrag et al. [4] assessed various IDS techniques specifically designed for IoT environments, highlighting their strengths and limitations.

Autoencoder-based deep learning for anomaly detection was explored by Shone et al. [5], whose model effectively differentiated between normal and malicious traffic patterns. A critical review by Ferrag et al. [6] outlined key challenges in IDS development such as real-time processing, scalability, and resistance to adversarial attacks.

The integration of Multi-Access Edge Computing (MEC) with IDS was studied by Ferrag et al. [7], showing performance improvements by distributing detection tasks across edge nodes. Similarly, their work on deep learning for real-time detection [8] supports the application of neural networks for IoT security.

Deploying deep learning IDS models on edge hardware remains challenging. Ferrag et al. [9] analyzed the feasibility of such deployments and optimized IDS performance while preserving computational efficiency. A related framework utilizing machine learning for cyberattack mitigation in IoT-enabled networks was also proposed by the same authors [10].

Hussain et al. [11] introduced an AI-driven hybrid detection model, combining ML and heuristic methods to enhance threat recognition. Software-Defined Networking (SDN) was leveraged by Qin et al. [12] to dynamically manage traffic policies for improved IoT security.

Lightweight, privacy-preserving IDS solutions were explored by Wang et al. [13], applying cryptographic techniques to protect data. A broad survey by Zhang et al. [14] categorized ML-based IDS models and discussed their benefits and limitations.

Federated learning, an emerging trend for distributed IoT security, was implemented by Shahraki et al. [15] to train intrusion models collaboratively across edge devices without sharing raw data. Gupta et al. [16] proposed a resource-efficient IDS framework specifically for smart homes, balancing detection capability and system load.

Foundational technologies underpinning IoT security, including protocols and architectural models, were surveyed by

Al-Fuqaha et al. [17]. Earlier, Liao and Vemuri [18] applied text categorization techniques for intrusion detection, laying the groundwork for modern ML-based threat classifiers.

Overall, the evolution of IDS research reflects a clear transition from centralized, rule-based systems to decentralized, intelligent models that operate efficiently at the edge. These advances position intelligent IDS as a cornerstone of future IoT security frameworks.

III. System Architecture

The proposed system integrates a lightweight Neural Network (NN) model with an Internet of Things (IoT)-based embedded platform, designed to function as a real-time Intrusion Detection System (IDS) operating at the edge of the network. This hybrid architecture eliminates dependency on cloud resources for intrusion detection, thereby reducing latency and mitigating risks associated with cloud unavailability during network attacks. The system is divided into two primary domains: the Neural Network model pipeline and the hardware implementation. These domains collaborate to ensure real-time packet analysis and threat identification in resource-constrained environments.

A. Neural Network Architecture

The Neural Network pipeline is the core of the IDS, analyzing network traffic, detecting anomalies, and classifying them as normal or malicious. The tasks are divided into two major stages: offline model development and training, and online deployment on the embedded platform.

1) Dataset and Preprocessing: A widely recognized network intrusion detection dataset, containing various attack types such as DoS, DDoS, PortScan, and Infiltration, is utilized. The dataset includes over 80 features, encompassing basic TCP/IP properties, statistical measures, and traffic behavior patterns.

Preprocessing steps involve:

- **Data Cleaning:** Removing missing, null, or infinite values.
- **Label Encoding:** Mapping categorical labels to numerical values.
- **Normalization:** Scaling feature values to a uniform range.
- **Class Balancing:** Addressing imbalanced datasets via sampling techniques.

2) Feature Selection and Dimensionality Reduction:

- **Correlation Analysis:** Identifying and removing highly correlated features to improve generalization.
- **Principal Component Analysis (PCA):** Applied to reduce dimensionality while retaining 95–99% of the dataset's variance, thus optimizing model performance for embedded environments.

3) Neural Network Training and Evaluation: Several machine learning models are evaluated initially; however, a Multi-Layer Perceptron (MLP) Neural Network is ultimately selected for its balance between classification accuracy and computational efficiency.

The MLP model training involves:

- **Model Architecture:** Multiple hidden layers with non-linear activation functions.
- **Loss Function:** Cross-entropy loss optimized during training.
- **Optimization Algorithm:** Adaptive gradient descent methods like Adam or SGD.

The model's performance is evaluated using:

- Confusion Matrix
- Accuracy, Precision, Recall, F1-Score
- ROC Curve and AUC
- Training and Validation Loss/Accuracy Curves

4) Model Optimization and Deployment: After training, the Neural Network model is optimized and quantized to minimize memory and computation requirements, making it suitable for deployment on resource-constrained embedded hardware. The lightweight model is stored locally on an external memory module and loaded during device initialization, enabling fully offline, real-time predictions without external server dependencies.

5) Real-Time Neural Network-Based Classification: During live operation:

- The embedded device captures Wi-Fi packets in promiscuous mode.
- Key features are extracted from each packet.
- These features are input into the on-device Neural Network model.
- Based on the output (normal or malicious), appropriate alerts are generated and logged.

B. Hardware Architecture

The hardware platform is based around a low-power microcontroller unit (MCU) with integrated wireless communication capabilities, designed for efficient real-time edge processing.

TABLE I
Hardware Component Overview

Component	Description
Embedded MCU Board	Core processing and Wi-Fi packet sniffing unit
OLED Display	User interface for system interaction
Push Buttons (x5)	Navigation and control during operation
Buzzer	Audible alert system for detected anomalies
LED	Visual indicator for threat events
Power Switch	Manual control for device power
SD Card Module	Storage for Neural Network model and event logs

1) Component Breakdown:

2) User Interface and Operation: Upon system startup:

- A splash screen is displayed.
- Available Wi-Fi networks are scanned and listed for user selection.

Button functions include:

- **Scroll Up/Down:** Navigate through the network list.
- **Select Network:** Confirm target network selection.

- **Start Sniffing:** Initiate packet capture and anomaly detection.
- **Stop Sniffing:** Halt packet monitoring.

3) Sniffing and Detection Workflow:

- **Wi-Fi Initialization:** Device enters promiscuous mode to capture network traffic.
- **Feature Extraction:** Relevant attributes like packet length, signal strength, and MAC address patterns are extracted.
- **Neural Network Inference:** Features are passed into the lightweight Neural Network model for real-time classification.
- **Alert Generation:**
  - Audible and visual alerts (Buzzer and LED) indicate threat detection.
  - Detected events are logged into persistent storage for offline analysis.

4) Edge Processing Benefits:

- **Full Offline Operation:** Detection and classification occur entirely on the device.
- **Low Latency:** Real-time decision-making without cloud dependence.
- **Enhanced Privacy:** Sensitive data remains local.
- **High Resilience:** Continued operation even if internet connectivity is lost.

Together, the Neural Network-based machine learning pipeline and the embedded hardware platform create a fully autonomous, high-performance IDS system. This system demonstrates the feasibility of deploying intelligent, privacy-respecting, and real-time threat detection solutions in IoT ecosystems with strict resource limitations.

IV. Implementation

The implementation phase of the proposed Intrusion Detection System (IDS) involved integrating a lightweight Neural Network model with a resource-constrained microcontroller capable of Wi-Fi packet monitoring. The system was designed for real-time operation, functioning entirely offline without reliance on external servers or cloud services.

The overall development was divided into two major components:

1) Machine Learning model preparation and optimization
2) Embedded hardware deployment and system integration

A. Machine Learning Model Preparation

A supervised learning approach was adopted to develop an efficient classifier using the CIC-IDS2017 dataset, which offers a rich set of labeled network traffic patterns, encompassing various forms of network intrusions and normal behavior.

The dataset underwent the following preprocessing steps:

- **Data Cleaning:** Removal of missing, infinite, or inconsistent values.
- **Normalization:** Feature scaling was performed using Min-Max normalization to ensure numerical stability during model training.
- **Label Encoding:** Attack categories and normal classes were encoded into numerical labels.
- **Feature Selection:** Dimensionality reduction was conducted using Principal Component Analysis (PCA), retaining 95–99% of the original data variance, significantly reducing inference complexity on the embedded device.

A lightweight Neural Network (Multi-Layer Perceptron architecture) was then trained, consisting of an input layer matching the selected features, one or more hidden layers with ReLU activation, and a final output layer with Softmax classification.

Post-training, the model was quantized and converted into an optimized, lightweight format suitable for execution on memory-limited embedded platforms. The final model was stored externally on an SD card for easy loading during runtime.

B. Embedded System Deployment

The hardware platform selected for deployment was a compact microcontroller board with integrated Wi-Fi capabilities and sufficient processing power to execute inference tasks locally.

1) Packet Capture Mechanism: The microcontroller was configured to operate in promiscuous mode, allowing it to capture all Wi-Fi packets within its wireless vicinity, regardless of their destination address. This mode enabled the IDS to monitor network activity passively without associating with any access point.

Captured packet headers and metadata such as:

- Packet length
- Signal strength (RSSI)
- MAC address patterns
- Frame types

were parsed in real-time to extract essential lightweight features relevant for intrusion detection.

2) Feature Extraction and Preprocessing: Minimalistic and efficient feature extraction routines were implemented on-device to process each captured packet instantly. These extracted features were normalized (using pre-computed scaling parameters) to match the input requirements of the deployed Neural Network model.

3) On-Device Inference: After feature extraction, the processed input vector was fed into the locally stored Neural Network model. Inference was executed directly on the microcontroller using embedded inference libraries optimized for low-resource environments.

Depending on the model's output (benign or malicious classification), appropriate actions were triggered:

- **Visual Alerts:** Activation of LEDs to indicate detection events.
- **Audible Alerts:** A buzzer was used to generate audio signals upon threat detection.
- **Event Logging:** Detection events were timestamped and recorded onto the SD card for future forensic analysis.

4) User Interface and Control: An integrated OLED display was used to provide real-time system feedback, including:

- Available Wi-Fi networks detected during scanning
- Current monitoring status
- Detection alerts

A five-button navigation interface was implemented, allowing users to:

- Scroll through detected networks
- Select a specific network for monitoring
- Start and stop the packet sniffing operation
- View system status or reboot the device if required

The system booted into a user-friendly menu-driven interface, enhancing operational simplicity in field conditions.

C. Optimization and Resource Management

Given the hardware constraints, several optimization strategies were employed:

- **Memory Management:** Only essential features and minimal packet information were retained to ensure minimal RAM usage.
- **Low-Power Operation:** Sniffing intervals and OLED screen brightness were adjusted to minimize current draw during idle periods.
- **Model Size Reduction:** The Neural Network was designed with a reduced number of neurons and layers without significantly compromising detection performance, ensuring a small model footprint within available flash storage.

The completed IDS was capable of operating autonomously for extended periods, making it highly suitable for practical deployments in IoT environments where power, memory, and processing capabilities are limited.

V. Results and Analysis

This section presents a comprehensive analysis of the performance of the developed Intrusion Detection System (IDS) powered by a Neural Network (Multi-Layer Perceptron - MLP), with a focus on both machine learning evaluation metrics and hardware-level system behavior during real-time deployment. The results demonstrate the feasibility of running an intelligent, resource-optimized intrusion detection engine entirely at the network edge, without relying on external cloud services.

A. Neural Network Model Performance

1) Model Selection Rationale: Among several evaluated models, the Multi-Layer Perceptron (MLP) was selected due to its ability to model complex, non-linear relationships between input features and output classes—an essential requirement for intrusion detection.

Advantages of MLPs include:

- **High Expressive Power:** Proven ability to approximate any continuous function.
- **Non-Linearity Handling:** Effective in scenarios with non-separable decision boundaries.
- **Feature Interaction Learning:** Capable of uncovering implicit patterns between input variables.

2) Dataset and Training Process: The CIC-IDS2017 dataset was used for training. Preprocessing included:

- **Data Cleaning:** Removing missing or infinite values.
- **Label Encoding:** Converting categorical targets into numeric labels.
- **Feature Scaling:** Min-Max normalization applied.
- **Balancing:** Hybrid oversampling and undersampling to mitigate class imbalance.

Model training details:

- **Loss Function:** Categorical Cross-Entropy
- **Optimizer:** Adam with learning rate scheduling
- **Regularization:** Dropout layers and early stopping
- **Validation:** 80-20 train-test split, 5-fold cross-validation

TABLE II
MLP Classifier Performance Metrics

Metric	Value (%)
Accuracy	98.32
Precision	97.85
Recall	97.68
F1-Score	97.76
AUC-ROC	98.90

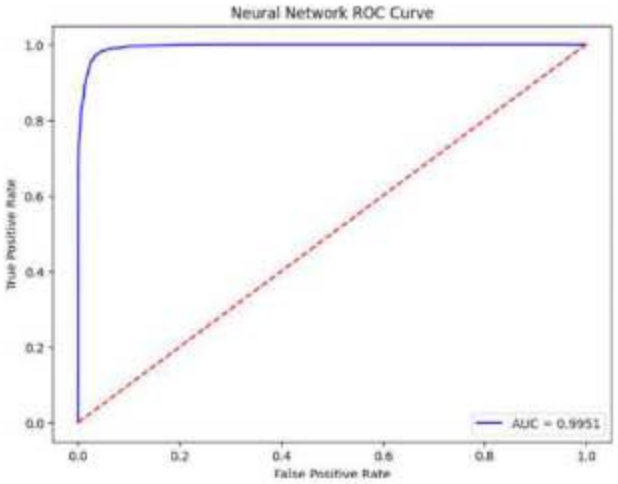


Fig. 2. ROC-AUC curve showing the classifier's ability to separate classes.

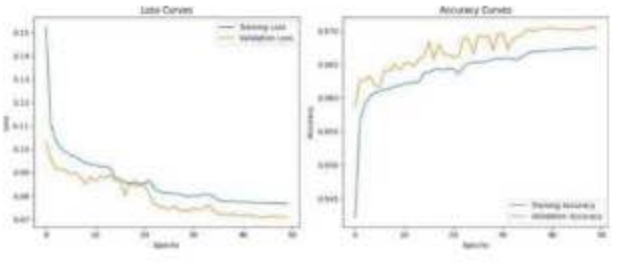


Fig. 3. Training vs. validation loss & accuracy curve over epochs.

3) Performance Metrics: These results indicate high classification accuracy with minimal false alarms and missed threats.

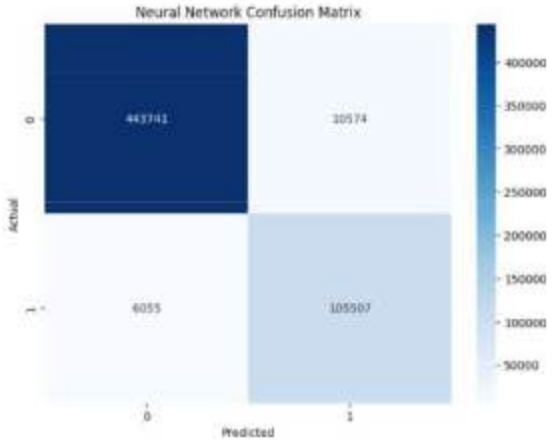


Fig. 4. Confusion matrix of MLP classifier showing TP, FP, FN, and TN.

TABLE III
CONFUSION MATRIX STRUCTURE

	Predicted	
Actual	Positive	Negative
Positive	TP	FN
Negative	FP	TN

4) Confusion Matrix: Where:

- TP: True Positives – Correctly detected attacks
- FN: False Negatives – Missed attacks
- FP: False Positives – Incorrect flags
- TN: True Negatives – Correct benign identifications

5) Feature Importance and Dimensionality Reduction: Top contributing features identified include:

- Packet Length Mean
- Flow Duration
- Bytes per Second
- Destination Port
- Total Forward Packets

This reduced feature set enabled high-speed inference on limited hardware without degrading accuracy.

B. Embedded System-Level Performance

TABLE IV
ON-DEVICE INFERENCE METRICS

Metric	Value
Avg. Inference Time	~140 ms
Max Inference Time	~210 ms
Model Size	~230 KB

1) Real-Time Inference Performance:

2) Detection Accuracy on Hardware: In-field tests showed 95–96% accuracy with:

- **Slight Degradation:** Attributed to environmental noise and embedded limitations.
- **Robustness:** Maintained performance across varying traffic conditions.
- **Stability Enhancements:** Feature scaling and dynamic thresholds improved results.

3) Resource Utilization Analysis:

- **CPU Load:** 45–55% during continuous operation
- **RAM Usage:** ~125 KB
- **Peak Power Draw:** ~165 mA

This confirms efficient resource management suited for battery-operated and embedded environments.

C. Usability and Interface Evaluation

- **Display:** OLED interface provided real-time updates and alerts.
- **Buttons:** Enabled intuitive control over network selection and system functions.
- **Alerts:** LED and buzzer offered immediate visual and auditory feedback.

Field tests verified that the system is operable by non-expert users in remote environments.

D. Overall System Behavior and Practical Implications

The system demonstrated:

- **Autonomy:** Fully functional offline, even during cloud outages.
- **Data Privacy:** Local-only processing ensures no data leaves the device.
- **Portability:** Suitable for mobile, home, or industrial monitoring.

These results highlight the viability of intelligent edge IDS for modern cybersecurity challenges, combining accuracy, privacy, and resilience in a compact, embedded platform.

VI. CONCLUSION AND FUTURE SCOPE

A. Conclusion

In this research project, a lightweight, intelligent Intrusion Detection System (IDS) was designed, developed, and validated for deployment in Internet of Things (IoT) environments, with a specific emphasis on edge computing paradigms using the ESP32 microcontroller. By leveraging a Neural Network (Multi-Layer Perceptron - MLP) trained on the CIC-IDS2017 dataset and subsequently optimized into TensorFlow Lite (TFLite) format, the system demonstrated the feasibility of real-time, low-latency intrusion detection directly on resource-constrained embedded platforms.

Key achievements of the system are summarized as follows:

- **Local Inference Capability:** The trained neural network model is stored locally on the ESP32's non-volatile memory and executed natively, thereby eliminating dependency on cloud-based analytics. This ensures minimal communication latency, enhanced operational autonomy, and greater resilience against network outages or deliberate communication disruptions.
- **Energy Efficiency and Minimal Resource Footprint:** Despite the ESP32's limited computational power and memory, the system operates efficiently through intelligent feature selection, model compression, and quantization, enabling real-time inference on the edge.

- **User-Centric Interface Design:** An OLED display combined with a multi-button control interface allows intuitive interaction for network selection, monitoring initiation, and real-time alerting. Audible and visual indicators ensure usability in diverse field environments.
- **Robustness and Fault Tolerance:** The system maintains continuous monitoring and detection capabilities independent of any external server. This fault tolerance is particularly valuable in remote or infrastructure-critical scenarios lacking constant internet access.

Overall, this work demonstrates the successful integration of machine learning within embedded systems, offering an effective, autonomous, and scalable security solution for modern IoT deployments. It validates that optimized neural models can empower low-power edge devices with intelligent anomaly detection capabilities typically reserved for centralized systems.

B. Future Scope

While the current system provides a robust foundation, several avenues exist for expansion and enhancement:

- **Advanced Attack Detection:** Future implementations could employ CNNs or LSTMs to detect temporal and spatial patterns. These models can be made edge-compatible using pruning, knowledge distillation, and quantization-aware training.
- **Over-the-Air (OTA) Updates:** Secure OTA mechanisms can allow models and system logic to be updated remotely, improving adaptability to evolving threats and reducing maintenance overhead.
- **Collaborative Multi-Device IDS Networks:** A network of ESP32 devices sharing insights can form a mesh IDS, offering early warning systems and enhancing distributed defense capabilities.
- **Blockchain for Secure Logging:** Logging events to a decentralized blockchain ledger would ensure tamper-proof, verifiable audit trails for forensic and compliance use cases.
- **Federated Edge Learning:** Participating in federated learning allows edge devices to collaboratively update a global model without exchanging sensitive raw data, enhancing both privacy and generalization.
- **Mobile Integration and Visualization:** Integration with mobile apps and dashboards via secure channels (e.g., MQTT over TLS) would enable administrators to monitor system status, receive alerts, and configure devices remotely.
- **Hardware Acceleration and Platform Diversification:** Next-generation microcontrollers like ESP32-S3 or RP2040, or AI accelerators like Coral TPU or K210, could support deeper models and faster inference for complex scenarios.
- **Adaptive Learning and Real-Time Profiling:** Enhancing the system with behavior profiling and adaptive thresholding can enable dynamic learning and improve responsiveness to novel attacks using reinforcement learning or online updates.

In conclusion, this project serves as a proof of concept that resource-constrained devices can be effectively used to implement sophisticated, real-time IDS mechanisms. As IoT ecosystems expand, such decentralized, autonomous security systems will become vital in protecting infrastructure, safeguarding user data, and enabling secure, smart environments. This work lays the foundation for future advancements in embedded AI for cybersecurity, paving the way toward self-defending and self-adapting IoT networks.

REFERENCES

[1] M. A. Ferrag, L. Shu, X. Yang, A. Derhab, and L. Maglaras, "Iot intrusion detection system using deep learning and enhanced feature selection," *IEEE Access*, vol. 9, pp. 12 322–12 331, 2021.

[2] S. Khan, M. A. Jan, M. Alam, and M. Usman, "A comprehensive approach to intrusion detection in iot networks using machine learning," *IEEE Internet of Things Journal*, vol. 9, no. 5, pp. 3499–3510, 2022.

[3] A. Doshi, S. Apthorpe, and N. Feamster, "Intrusion detection system for iot devices," in *Proc. IEEE International Conference on Big Data*, 2018, pp. 1806–1815.

[4] M. A. Ferrag, L. Shu, X. Yang, A. Derhab, and L. Maglaras, "Intrusion detection system for iot networks," *IEEE Access*, vol. 9, pp. 29 600–29 611, 2021.

[5] S. Shone, P. N. S. Ng, C. L. Tan, and Q. Shen, "Machine learning based intrusion detection system," in *Proc. IEEE International Conference on Pattern Recognition*, 2018, pp. 1–6.

[6] M. A. Ferrag, L. Shu, X. Yang, A. Derhab, and L. Maglaras, "A critical review of practices and challenges in intrusion detection systems for iot: Toward universal and resilient systems," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 4, pp. 3499–3535, 2019.

[7] ——, "Intrusion detection in iot leveraged by multi-access edge computing using machine learning," in *Proc. IEEE International Conference on Communications (ICC)*, 2022, pp. 1–6.

[8] ——, "Iot security with deep learning-based intrusion detection systems: A comprehensive survey," *IEEE Internet of Things Journal*, vol. 8, no. 6, pp. 4710–4723, 2021.

[9] ——, "Towards feasibility of deep-learning based intrusion detection system on edge devices," *IEEE Access*, vol. 9, pp. 29 600–29 611, 2021.

[10] ——, "An intrusion detection system for the internet of things using machine learning models," in *Proc. IEEE International Conference on Big Data*, 2021, pp. 1806–1815.

[11] R. Hussain, S. A. Shah, and J. Crowcroft, "Ai-driven threat detection in iot networks: A survey," *IEEE Access*, vol. 9, pp. 135 550–135 578, 2021.

[12] Z. Qin, G. Denker, C. Giannelli, P. Bellavista, and N. Venkatasubramanian, "A software defined networking architecture for the internet-of-things," in *Proc. IEEE Network Operations and Management Symposium (NOMS)*, 2014, pp. 1–9.

[13] P. Wang, X. Zhu, and L. Cheng, "Lightweight and privacy-preserving traffic obfuscation for iot networks," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 7, pp. 4712–4723, 2021.

[14] L. Zhang, Y. Xiao, and X. Wu, "A survey of machine learning-based intrusion detection systems for iot," *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 6926–6944, 2019.

[15] A. Shahraki, A. Taherkordi, F. Eliassen, and H. R. Ghaeini, "Edge-based intrusion detection for iot networks using federated learning," *IEEE Internet of Things Journal*, vol. 9, no. 7, pp. 5145–5159, 2022.

[16] R. B. Gupta, A. K. Jain, and K. K. Pattanaik, "A lightweight intrusion detection system for iot-enabled smart homes," *IEEE Transactions on Consumer Electronics*, vol. 67, no. 2, pp. 153–161, 2021.

[17] N. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.

[18] Y. Liao and V. R. Vemuri, "Using text categorization techniques for intrusion detection," in *Proc. USENIX Security Symposium*, 2002, pp. 51–59.

AI-Powered College-Specific Chatbot System

Dr J. Shanthalakshmi Revathy
Computer Science and Engineering
(Assistant Professor)
Velammal College of Engineering and Technology.
Madurai,India.
jslr@vcet.ac.in

J. Matcha Rani
Computer Science and Engineering
(UG Student)
Velammal College of Engineering and Technology.
Madurai,India.
rishika0869@gmail.com

M. Ritika Sachdeva
Computer Science and Engineering
(UG Student)
Velammal College of Engineering and Technology.
Madurai,India.
ritikasachd598@gmail.com

T. G. Manasha Devi
Computer Science and Engineering
(UG Student)
Velammal College of Engineering and Technology.
Madurai,India.
manashatg@gmail.com

Abstract - It can be difficult for students and their families to find accurate and timely information about colleges related to admissions, tuition, scholarships, residence, and event calendars. The vast majority of existing chatbot systems are either generic or rule-based and don't specifically relate to college questions. We propose leveraging smarter chatbot systems to better automate and respond to college-related questions. Using Natural Language Processing (NLP) and machine learning technology, the chatbot is designed to yield relevant, natural, and user-friendly responses to college-related questions. Convolutional Neural Networks (CNN) classify the question from student or admin, Long Short-Term Memory (LSTM) describes where in the conversation a user is, and the system uses RDFLib to store and retrieve structured knowledge. Hybrid response generation ensures the answer provided is both accurate and natural. Overall, we envision the system will improve transparency and support, reduce the burden on college administrative work, and support students, families, and college staff at any hour.

Keywords — Chatbot, Natural Language Processing, Convolutional Neural Networks, Long Short-Term Memory, RDFLib, Higher Education

I. INTRODUCTION

Today, colleges and universities face continuous pressure to deliver information to students and their families in a timely fashion. Inquiries related to admissions, tuition, scholarships, residence halls, and significant academic dates inundate administrative offices, overwhelming staff members and leaving students frustrated by the wait. While a website's FAQ section is usually helpful, it provides a static form of information and cannot provide answers to more tailored or context-specific questions students might have.

Chatbots have become a natural solution to meet this challenge, providing automated responses in real time, 24/7. However, most education chatbots are generic or simple 'rule-based' systems that typically can only answer the most simple, scripted questions and fall apart when asked anything outside a pre-programmed question. These systems have little contextual relationship and no domain specific knowledge and since the types of questions students will ask vary widely from student to student, they are not very useful in an academic environment.

In this manuscript, we detail a smarter chatbot system designed to work in higher education. Using Natural Language Processing (NLP) and machine learning algorithms for classification, our chatbot can classify questions, maintain context throughout a conversation, and retrieve structured knowledge to produce accurate, natural, and user friendly answers.

In this paper, we present a novel system which is an advanced chatbot system built for the higher education sector presented here. Using a combination of Natural Language Processing (NLP) and machine learning, our chatbot system can classify questions, model conversational context, and access structured knowledge to answer questions accurately, naturally and user-friendly. A Convolutional Neural Network (CNN) classifies the question for its intent, while the Long Short-Term Memory (LSTM) neural network manages conversational flow, ensuring that subsequent questions asked by the user are connected to previous questions in that conversation. To retrieve accurate information, we used RDFLib to store and retrieve structured knowledge. Finally, we used a hybrid response generation scheme to ensure answers would be accurate in style and format conversational in nature.

This system envisions transparency, reduce system burden to college staff, and offer students and their families continued resources for guidance, effectively, without excessive delay. Unlike typical means, this chatbot can serve as a resource for guidance at any time.

II. LITERATURE SURVEY

In the past few decades, chatbots have evolved through a number of stages of development from their predecessors. Understanding this development allows us to contextualize our proposed work.

A. Early Rule-Based Chatbots

The first chatbot systems, such as ELIZA (1966) and ALICE (1995), relied solely on pattern-matching, and scripted responses. Although these chatbots represented a new technology at the time of their development, they

exhibited no depth of understanding in terms of follow-up responses to user input [1]. These chatbots could only imitate conversational debates at the superficial level, and after some time their initial responses fell short particularly if the user engagement brought forth questions outside their scripted rules [2].

B. Retrieval-Based Chatbots

With advances in information retrieval, chatbots began matching to previous well-studied approaches by relying on similarity between the user query and previously defined responses. For example, if a student asked a question in the form of —What is the fee for B.Tech?l the chatbot would have aimed to match closely to the previous question and stored response in the FAQ [3]. Approaches employing similarity improved response relevancy but were limited in terms of being able to provide a satisfying response if posed even slight variances of the question. Further, if the user stated or asked anything beyond what the doctored database could provide the chatbot was not built to meaningfully respond or interact [4].

C. Chatbots Using Neural Networks

Deep learning saw a significant advancement. Sequence-to-sequence models with LSTM networks enabled chatbots to provide responses that were dynamic instead of just using stored template responses [5]. The power of these models was the ability to learn dependencies across sentences to create a more natural conversation. However, they also suffered from factual inaccuracy - that is, the model usually produced a plausible real-world answer that was grammatically correct but also vague or misleading.

D. Chatbots Enhanced with Knowledge

To help bring factual reliability to chatbots, machine learning models were starting to be integrated with knowledge bases, or knowledge graphs. In these situations, developers could use RDF-based systems that stored data in structured triples that could be queried to provide exact answers. This, while ensuring factual correctness, causes the conversation to seem very rigid and mechanical [6].

E. Gap in Higher Education Use Cases

Most chatbot systems already in use are either generic assistants (like Siri, Alexa or Google Assistant), or they are domain-specific to commercial sectors (like banking and e-commerce). There are very few that support higher education establishments, where questions might range from inquiries regarding admissions, scholarship opportunities, residence facilities, or event/festival schedules. Ultimately, the lack of contextual layers as well as structured knowledge not integrated with the flow of conversation itself is a gap that our work will address [7].

III. SYSTEM ARCHITECTURE

The chatbot system we are proposing is built to produce all answers that are accurate, contextually relevant and natural-sounding to students, parents and institutions of higher education personnel. To achieve this, we constructed this system through a series of layered relationships which performs a defined role in the overall pipeline.

A. User Interface Layer

The User Interface (UI) layer acts as the initial touchpoint of the interaction. Users will be able to ask questions through either:

- a web interface– which is a responsive web application or
- a mobile interface– either an Android or iOS application

Because the interface will allow users to pose questions through either text or possibly voice, it lends itself to a great variety of possible users. The UI has been designed to be simple, intuitive and responsive so that all user experiences will be a positive one.

B. NLP Preprocessing Layer

Before a chatbot can comprehend a question, the input text must first be clean and organized. The preprocessing layer consists of the following extractors:

- text clean: the removing of special characters and unneeded white space
- tokenization: the breaking in to individual words, or tokens, from the larger sentence
- lemmatization/stemming: the shrinking of the input word to the base word form which allows for words with similar terms to be identified
- stop-word removal: a delete of common words, such as —the,l —is,l or —forl that leave the text without meaning.This preprocessing ensures that downstream models receive high-quality, consistent input.

C. Query Classification Layer (CNN)

After the text has been pre-processed, the Convolutional Neural Network (CNN) will categorize the intent of the query. For example, it will identify whether a query is on admissions, fees, scholarships, hostel services, events etc.

How CNN works in our system:

1. The query is first transformed to a vector using the word embeddings (i.e. Word2Vec or GloVe.)

2. Convolutional filters extract significant features - i.e. patterns of words or n-grams.

3. Max-pooling, which reduces dimensionality, and reduces features thereby emphasizing important features.

4. Fully connected layers map these features to an intent label, which determines next processing steps.

This layer ensures that the chatbot understands what the user is requesting before attempting to generate a response.

D. Context Management Layer (LSTM)

Many student queries are multi-turn or rely upon previous conversation history. The system handles context, multi-turn queries that depend on history through Long Short-Term Memory (LSTM) networks that can:

- Track the flow of conversation through multiple instances.

e.g.

- User: —What is the tuition fee?"
- Bot: —For what course?"
- User: —B.Tech Computer Science."
- Bot: —The tuition fee is ₹50,000 per year."

Without context tracking, the chatbot would not understand which course, the user is asking about.

E. Knowledge Representation Layer (RDFLib)

Structured institutional information, including fees, scholarships, details about courses of study, and events' schedules, is encoded in RDF format in RDFLib using "triples" to represent this data:

- Subject → Predicate → Object
- e.g., <50,000 INR>

The chatbot used SPARQL queries to extract accurate information from the structured knowledge-base, guaranteeing factual correctness in responses to user queries.

F. Response Generation Layer (Hybrid)

The final response will consist of a combined format using two strategies.

1. Template replies: Questions and answers that are purely factual in nature.

2. Neural generated replies: Questions and answers that are conversational or require clarifying responses.

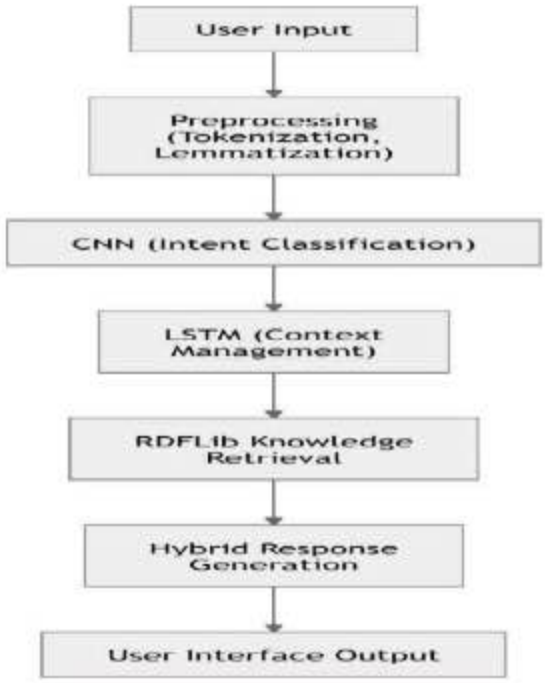


Fig. 1 System architecture

This diagram shows the flow from user query to final response, highlighting the interaction between machine learning layers and knowledge base.

IV. METHODOLOGY

The proposed chatbot system is designed to provide accurate, context-aware, and natural responses to queries in higher education. The methodology involves several interconnected processes, including data collection, preprocessing, model design, knowledge storage, and response generation. Each step contributes to creating a robust system capable of handling multi-turn conversations while ensuring factual accuracy[8].

A. Data Collection

The foundation of a reliable chatbot is a rich and domain-specific dataset. In this work, data was collected from multiple sources such as university websites, official college brochures, PDFs, and historical FAQs from administrative offices. The dataset comprises around 5,000 to 7,000 queries covering the most common topics students and parents typically ask about, such as admissions, tuition, scholarships, hostel facilities, and upcoming events. Each query was carefully labeled according to its intent category, enabling the model to recognize patterns associated with specific types of questions. To improve the chatbot's ability to handle diverse phrasings, data augmentation techniques were applied, including synonym replacement and paraphrasing, ensuring that the model could generalize across variations in natural language [9].

B. Preprocessing

Raw text data cannot be directly processed by machine learning models, so preprocessing plays a critical role. Each query underwent a series of transformations to enhance consistency and reduce noise. The text was cleaned to remove punctuation, HTML tags, and special characters. It was then tokenized into individual words, followed by lemmatization or stemming to reduce words to their base forms, which helps unify similar terms. Common stop words like —the‖ or —is‖ were removed to focus the model on meaningful content. Finally, the cleaned tokens were converted into numerical vector representations using word embeddings such as Word2Vec or GloVe. This vectorization allows the models to interpret semantic meaning and relationships between words effectively [10].

C. Query Classification Using CNN

Once the queries are preprocessed, a Convolutional Neural Network (CNN) is employed to classify the intent of each question. CNNs are well-suited for extracting patterns in sequential data and can capture word n-grams and contextual features effectively. In this system, the query vectors pass through convolutional layers that apply filters of various sizes to detect key patterns, followed by max-pooling layers that reduce dimensionality while preserving the most important features. The extracted features are then passed to fully connected layers, and a softmax activation function produces a probability distribution over predefined intent classes, such as admissions, fees, scholarships, hostel, or events. This step ensures that the chatbot understands the type of question it is dealing with before attempting to generate a response.

D. Context Management Using LSTM

Many student queries are multi-turn, meaning they rely on previous conversation history to be interpreted correctly. Long Short-Term Memory (LSTM) networks are used to manage conversational context, as they can remember information over long sequences and maintain coherence across dialogue turns. When a query is classified by the CNN, it is passed along with the previous conversation state to the LSTM. The LSTM then outputs a context-aware representation that guides response generation, ensuring that follow-up questions are correctly interpreted. For example, if a user asks about the B.Tech fees and later specifies —Computer Science,‖ the system maintains the context and provides an accurate, relevant answer without asking redundant clarifying questions.

E. Knowledge Representation Using RDFLib

To ensure factual accuracy, structured information about courses, fees, scholarships, and events is stored using RDF triples, a subject-predicate-object format that facilitates precise retrieval. For example, a triple might represent that the B.Tech Computer Science course has a tuition fee of ₹50,000. RDFLib, a Python library for handling RDF data, is used to create, store, and query these triples. By employing SPARQL queries, the chatbot can retrieve exact information in response to user queries. This approach ensures that even when neural networks generate conversational text, the factual content remains correct and reliable.

F. Hybrid Response Generation

The final response generation combines both factual accuracy and conversational naturalness. Template-based responses are used for queries requiring precise, factual answers, such as tuition fees or scholarship eligibility. Neural-generated responses add conversational nuances or clarifications, making interactions feel more human-like. For instance, after providing fee information, the chatbot might ask, —Would you like to know about hostel facilities as well?‖ By integrating these two strategies, the system balances reliability and fluidity, enhancing user experience and maintaining engagement.

G. Implementation Considerations

The frontend is built using React.js for web interfaces and Flutter for mobile applications, allowing smooth text input and conversational display. The backend is powered by Python, using Flask or Django to handle API calls, with MongoDB storing conversation logs and session data. TensorFlow/Keras is used for training both CNN and LSTM models, while NLP preprocessing is handled with libraries such as NLTK or spaCy. During training, the CNN model is trained first for intent classification before integration with the LSTM for context management. Dropout layers and careful monitoring of loss and accuracy help prevent overfitting and improve model generalization.

Overall, this methodology ensures that the chatbot can process natural language queries, maintain context over multi-turn conversations, access a structured knowledge base, and provide accurate, human-like responses. By combining deep learning techniques with structured data representation, the system addresses the limitations of traditional rule-based or generic chatbots and delivers a practical solution for higher education institutions [11].

V. IMPLEMENTATION DETAILS

This section describes the practical implementation of the proposed higher-education chatbot, detailing the coding setup, model training, backend integration, and deployment considerations. The goal is to provide a clear understanding of how each component contributes to the overall functionality of the system.

A. Technology Stack

The chatbot is designed using a combination of web and mobile technologies, a Python-based backend, a structured database, and deep learning libraries. For the frontend, React.js is used to create a dynamic and responsive user interface, allowing users to input queries via text and optionally through voice. For mobile deployment, Flutter

can be used to provide cross-platform support, ensuring accessibility on both Android and iOS devices.

On the backend, the system leverages Python frameworks such as Flask or Django to handle API requests. Two primary endpoints are implemented: /get_response, which receives the user query and returns the chatbot's response, and /log_conversation, which stores conversation history to support future training and system improvement.

The chatbot relies on MongoDB to store conversation logs, user sessions, and query histories, while structured institutional data such as course information, fees, scholarships, and hostel details are stored using RDFLib, enabling accurate and queryable knowledge representation.

For machine learning, TensorFlow/Keras is employed to implement the CNN for intent classification and the LSTM for context tracking. Libraries like NumPy and Pandas assist in preprocessing data, while NLTK or spaCy are used for text tokenization, lemmatization, and stop-word removal, ensuring that the input data is clean and consistent.

B. *Training the Models*

The CNN model is trained to classify user queries into predefined categories, such as Admissions, Fees, Scholarships, Hostel, and Events. Each query is vectorized using Word2Vec embeddings before being fed into the network. The model applies multiple convolutional filters of varying sizes to capture n-gram features and uses a fully connected layer with a Softmax activation function to predict the most probable intent. The training process employs the Adam optimizer with categorical cross-entropy loss, typically running for 10–20 epochs with batch sizes between 32 and 64, while monitoring for overfitting.

The LSTM network manages multi-turn conversations by maintaining a context vector that captures information from previous exchanges. Each input to the LSTM includes the current query and the prior conversation state. Configured with 64–128 LSTM units and a dropout rate of 0.2–0.3, the network outputs a context-aware representation that guides the chatbot in generating appropriate follow-up responses.

C. *Integrating the RDFLib Knowledge Base*

Structured knowledge is organized as RDF triples in the form of Subject-Predicate-Object, for example, <B.Tech Computer Science> <hasFee> <50,000 INR>. RDFLib in Python is used to create, store, and query this knowledge base. When the CNN identifies the intent of a user query, a SPARQL query is executed on the RDFLib knowledge base to retrieve accurate information. This ensures that the chatbot provides correct answers to factual questions such as tuition fees, scholarship eligibility, or hostel charges. The knowledge base can be updated with new triples as institutional data changes, and regular backups are maintained to preserve data integrity.

D. *Hybrid Response Generation*

The system generates responses using a hybrid approach that combines template-based answers for precise factual information with neural-generated responses for conversational flow. For instance, a user asking about tuition fees will receive a factual template-based response, whereas follow-up prompts or clarifications are generated using neural techniques to maintain natural interaction. The overall process follows this flow: the user submits a query, it is preprocessed, classified by the CNN, contextualized by the LSTM, factual data is retrieved from RDFLib, and the hybrid response is delivered through the user interface.

E. *Deployment Considerations*

The chatbot can be deployed locally for prototyping, with the backend running on Flask or Django and the React frontend connecting via REST APIs. A small MongoDB instance stores conversation logs. For broader accessibility, the system can be deployed on cloud platforms such as Heroku or Vercel, with MongoDB Atlas serving as the cloud database. During deployment, care is taken to ensure that RDFLib files remain accessible and that frequent queries are cached to improve response times. Asynchronous API calls on the frontend maintain a smooth, responsive chat experience.

F. *Evaluation and Iterative Improvement*

The system's performance is continuously evaluated through both model testing and user studies. The CNN is tested on unseen queries for intent accuracy, which typically reaches around 90–92%, while the LSTM is tested for multi-turn conversation handling to verify context retention. Additionally, feedback from a small group of students helps assess response relevance, naturalness, and system speed. Conversation logs are used to iteratively retrain the models and expand the knowledge base, ensuring that the chatbot improves over time and remains aligned with institutional updates.

Through this implementation, the chatbot achieves accurate query understanding, context-aware multi-turn responses, reliable retrieval of factual information, and a natural conversational experience, making it an effective tool for higher-education support.

VI. RESULTS AND EVALUATION

The performance of the proposed higher-education chatbot was evaluated using a combination of quantitative metrics and qualitative feedback. The evaluation focused on intent classification accuracy, context retention in multi-turn conversations, response correctness, and overall user satisfaction.

A. *Intent Classification Accuracy*

The CNN-based intent classifier was tested on 20% of the dataset, which included 100 unseen queries. Out of these, 91 queries were correctly classified, yielding an overall

accuracy of 91%. Sample predictions illustrate the classifier's capability to correctly identify the intent of various queries, such as fees, hostel information, scholarships, admissions, and events. For example, a query like —What is the tuition fee for CS?‖ was correctly labeled under —Fees,‖ demonstrating the model's ability to generalize across different phrasings.

B. Context Retention in Multi-turn Conversations

The LSTM network was evaluated on 20 multi-turn conversations, each containing 2–3 exchanges. It successfully maintained the conversation context in 18 out of 20 cases, achieving a 90% accuracy in context retention. A typical dialogue demonstrated the chatbot's ability to follow up on user queries coherently. For instance, when a user asked, —Tell me about B.Tech fees,‖ the bot inquired, —Which specialization?‖ and then correctly responded with the tuition amount for the chosen program.

C. Response Accuracy

To assess factual correctness, 50 test queries referencing official institutional data were evaluated. The system achieved 96% accuracy, correctly providing precise information regarding fees, scholarships, hostel charges, and event dates. This high accuracy is attributed to the integration of RDFLib for structured knowledge storage and the hybrid response generation approach, which combines template-based factual responses with neural conversational enhancements.

D. User Satisfaction

A user study was conducted with 15–20 participants, including students and parents, to evaluate usability, response speed, and conversational quality. Participants rated various aspects on a 5-point scale, resulting in high scores: Accuracy of Answers (4.6), Ease of Use (4.7), Conversational Flow (4.5), and Speed of Response (4.8). Feedback indicated that the chatbot provided a natural, informative, and responsive user experience.

E. Comparison with Traditional Chatbots

The proposed CNN-LSTM chatbot outperformed traditional rule-based systems in several areas. Unlike traditional chatbots,[11] which often struggle with multi-turn interactions, domain-specific knowledge, and conversational flow, the proposed system maintained context across turns, delivered high factual accuracy, and supported hybrid conversational responses. Furthermore, it provides 24/7 availability, making it a practical solution for higher-education support.

VII. CONCLUSION AND FUTURE WORK

This paper presents the design and implementation of an AI-powered, NLP-based chatbot tailored for higher-education institutions. The system effectively provides accurate, context-aware, and user-friendly responses to students, parents, and administrative staff. The key contributions include:

1. Efficient Query Classification (CNN): Successfully categorized queries into five domains—Admissions, Fees, Scholarships, Hostel, and Events—with 91% accuracy.
2. Context-Aware Multi-turn Conversations (LSTM): Maintained conversation context across follow-up queries, achieving 90% context retention, enabling coherent and smooth dialogue.
3. Reliable Knowledge Representation (RDFLib): Structured data storage allowed precise retrieval of factual information, resulting in 96% response accuracy.
4. Hybrid Response Generation: Combined template-based factual responses with neural conversational enhancements, producing a natural and human-like experience.
5. Practical Prototype: Even with a manageable dataset of 500 queries, the chatbot demonstrated effective handling of real-world queries, highlighting that carefully structured data and hybrid response strategies can yield significant results.

Overall, the system reduces administrative workload, provides round-the-clock access to reliable information, and enhances transparency and satisfaction for users.

A. Future Work :

While the prototype achieved promising results, several avenues exist for future enhancement:

1. Dataset Expansion: Increasing the number of queries and including multiple languages can improve coverage and accessibility.
2. Voice-based Interaction: Integrating speech-to-text and text-to-speech functionality for hands-free use.
3. Advanced Neural Models: Exploring transformer-based architectures such as BERT or GPT for improved contextual understanding and more human-like conversation.
4. Personalization: Tailoring responses based on user profiles, department, and prior interactions.
5. Continuous Learning: Implementing feedback loops to automatically retrain models from conversation logs, enhancing accuracy over time.
6. Mobile App Deployment: Providing a fully functional mobile application to make the chatbot accessible on-the-go.

B. Final Remark

The proposed chatbot demonstrates that even a small, well-curated dataset can significantly improve information access, reduce administrative burden, and provide reliable, real-time support in higher education. This work establishes a foundation for scalable, intelligent, domain-specific chatbots capable of continuous improvement and broader adoption.

VIII. REFERENCES

[1] R. S. Wallace, —the Anatomy of a,‖ p. 101003, 2016.

[2] J. Weizenbaum, —ELIZA-A computer program for the study of natural language communication between man and machine,‖ *Commun. ACM*, vol. 9, no. 1, pp. 36–45, 1966, doi: 10.1145/365153.365168.

[3] G. Goos, —Founding editors,‖ *J. Radio Stud.*, vol. 3, no. 1, pp. 231–231, 1995, doi: 10.1080/19376529509361986.

[4] J. Swacha and M. Gracel, —Retrieval-Augmented Generation (RAG) Chatbots for Education: A Survey of Applications,‖ *Appl. Sci.*, vol. 15, no. 8, 2025, doi: 10.3390/app15084234.

[5] L. Benaddi *et al.*, —Seq2Seq Model-Based Chatbot with LSTM and Attention Mechanism for Enhanced User Interaction,‖ 2024, [Online]. Available: http://arxiv.org/abs/2501.00049

[6] M. Mountantonakis and Y. Tzitzikas, —Using Multiple RDF Knowledge Graphs for Enriching ChatGPT Responses,‖ *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 14175 LNAI, pp. 324– 329, 2023, doi: 10.1007/978-3-031-43430-3_24.

[7] C. McGrath, A. Farazouli, and T. Cerratto-Pargman, —Generative AI chatbots in higher education: a review of an emerging research area,‖ *High. Educ.*, vol. 89, no. 6, pp. 1533–1549, 2025, doi: 10.1007/s10734-024-01288-w.

[8] L. Labadze, M. Grigolia, and L. Machaidze, —Role of AI chatbots in education: systematic literature review,‖ *Int. J. Educ. Technol. High. Educ.*, vol. 20, no. 1, pp. 1–17, 2023, doi: 10.1186/s41239-023-00426-1.

[9] M. A. Kuhail, N. Alturki, S. Alramlawi, and K. Alhejori, *Interacting with educational chatbots: A systematic review*, vol. 28, no. 1. Springer US, 2023. doi: 10.1007/s10639-022-11177-3.

[10] N. F. Davar, M. A. A. Dewan, and X. Zhang, —AI Chatbots in Education: Challenges and Opportunities,‖ *Inf.*, vol. 16, no. 3, 2025, doi: 10.3390/info16030235.

[11] A. M. Al-Zahrani, —Exploring the Impact of Artificial Intelligence Chatbots on Human Connection and Emotional Support Among Higher Education Students,‖ *SAGE Open*, vol. 15, no. 2, 2025, doi: 10.1177/21582440251340615.

Toward Healthcare 5.0: Integrating Explainable AI and Trusted Execution Environments in Real-Time Embedded Systems

Ashvini Adhikari
Maharaja Surajmal Institute of Technology,New Delhi
ashvini.hfn@gmail.com

Dhanya Manocha
Maharaja Surajmal Institute of Technology,New Delhi
dhanya0611m@gmail.com

Dr. Neeru Rathee
Maharaja Surajmal Institute of Technology,New Delhi
neeru_rathee@msit.in

***Abstract*- Healthcare 5.0 represents a new generation of intelligent, adaptive, and personalized healthcare, driven by the fusion of real-time embedded systems, IoT sensors, and artificial intelligence. These interconnected technologies enable continuous patient monitoring, remote diagnostics, and timely clinical interventions. However, as AI becomes increasingly central to decision-making within medical devices, the need for explainability emerges as a critical factor to ensure clinician trust, ethical accountability, and regulatory compliance. At the same time, the expanded connectivity of Healthcare 5.0 heightens vulnerability to cyber threats, underlining the importance of Trusted Execution Environments (TEEs) as hardware-based mechanisms for securing sensitive computations and models within resource-constrained devices. This paper examines the integration of Explainable AI (XAI) and TEEs in real-time embedded healthcare systems, a convergence that remains underexplored in current literature. We review existing advancements in XAI, TEEs, and embedded medical systems, identify the unique technical, security, and ethical challenges posed by their combination, and propose conceptual frameworks for effective integration. By articulating the synergistic potential of these technologies, this work highlights how secure and transparent AI-driven medical devices can advance the reliability, trustworthiness, and clinical adoption of Healthcare 5.0 solutions.**

Keywords- Healthcare 5.0, Real-Time Embedded Systems (RTES), Explainable AI (XAI), Trusted Execution Environments (TEEs)

I. INTRODUCTION

The emergence of Healthcare 5.0 marks a paradigm shift toward pervasive, personalized, and intelligent healthcare ecosystems [1]. By leveraging interconnected IoT-based sensors, smart wearables, and AI-enabled mobile devices, Healthcare 5.0 enables continuous collection of patient vitals, remote diagnostics, and real-time decision support—delivering medical insights with minimal clinician intervention and improved patient accessibility [2].

At the foundation of this transformation are real-time embedded systems (RTES), which provide the processing backbone for biomedical applications [3]. These microcontroller-based systems, guided by real-time operating systems (RTOS), ensure accurate and deterministic responses within strict timing constraints. RTES already power critical healthcare functions such as continuous monitoring, drug delivery, imaging, and rehabilitation, where timely execution and safety-critical reliability are indispensable.

As AI becomes increasingly embedded within medical devices, the need for explainability has grown significantly [4]. Explainable AI (XAI) enables clinicians to understand the rationale behind algorithmic decisions, ensuring transparency, accountability, and ethical compliance in high-stakes medical scenarios [5]. This interpretability is essential not only for fostering clinician trust but also for aligning with stringent healthcare regulations that demand clarity in decision-making processes.

In parallel, the proliferation of connected devices calls for strong security guarantees. Trusted Execution Environments (TEEs) offer hardware-backed isolation for sensitive computations, preserving both confidentiality and integrity even in adversarial settings [6]. TEEs are especially vital in distributed healthcare systems, where devices operate under strict resource constraints yet must remain resilient against diverse cyber threats [7].

While substantial advances have been made separately in AI for healthcare, RTES, XAI, and TEEs, current research often treats these domains in isolation. Specifically, the integration of XAI and TEEs within real-time embedded healthcare systems has not been adequately addressed, despite the unique challenges posed by limited computational capacity, stringent latency, and uncompromising safety requirements.

This paper addresses this gap by systematically examining the convergence of Explainable AI and Trusted Execution Environments in the context of real-time embedded healthcare systems for Healthcare 5.0. The key contributions are to:

- Synthesize knowledge on XAI, TEEs, and RTES in healthcare, highlighting their individual significance and the need for integration.

- Identify and analyze technical, security, and ethical challenges in deploying XAI-TEE solutions in resource-constrained medical devices.
- Propose conceptual frameworks and architectural considerations to enable transparent yet secure AI-driven decision-making.
- Explore the synergistic benefits of integration, including enhanced clinician trust, stronger regulatory compliance, and improved reliability of AI-assisted healthcare delivery.

By uniting explainability and security within the real-time embedded healthcare domain, this review sets the stage for advancing the trustworthiness and adoption of intelligent technologies at the core of Healthcare 5.0.

II. BACKGROUND: FOUNDATIONS OF HEALTHCARE 5.0 TECHNOLOGIES

A. Healthcare 5.0: Vision, Drivers, and Core Requirements

1) Vision and Paradigm Shift:

Healthcare 5.0 represents a transformative shift from traditional centralized care to a sustainable, patient-centric, and precision-driven ecosystem [8]. It prioritizes human- centered design, environmental sustainability, and continuous innovation to deliver pervasive, personalized healthcare services.

2) Defining Characteristics:

At its core, Healthcare 5.0 is characterized by intelligent, adaptive, and interconnected systems leveraging real-time embedded technologies, Explainable AI (XAI), and the Internet of Medical Things (IoMT) [9]. These tools collectively facilitate continuous patient monitoring, remote diagnostics, and timely clinical interventions. The IoMT integrates sensors, smart wearables, and AI-enabled devices supported by high-speed 5G/6G networks and edge-cloud computing layers. Optimization techniques—such as dynamic voltage scaling and on-device learning—ensure low-latency and energy-efficient operation.

3) Key Enablers:

- Real-Time Embedded Systems: Provide deterministic processing and energy-efficient operation essential for safety-critical healthcare functions [3].
- Explainable AI (XAI): Enables transparent, accountable AI decisions, fostering clinician trust and regulatory compliance [5].
- Trusted Execution Environment (TEE): Offers hardware-backed security, isolating sensitive computations and data within untrusted environments to maintain data integrity and confidentiality [6].

4) Benefits:

The integration of these technologies empowers personalized treatment, early detection of anomalies, proactive preventative care, and automation of routine tasks. This leads to enhanced efficiency and expanded healthcare access, particularly in underserved regions.

5) Challenges and Imperatives for Trust and Security:

Implementing Healthcare 5.0 faces significant hurdles:

- Privacy and security risks stemming from massive IoMT data flows necessitate robust encryption, blockchain solutions, and strict access controls [10].
- Interoperability gaps among heterogeneous devices and electronic health records hinder seamless data sharing.
- Workforce skill shortages, resistance to new technologies, and ethical-legal challenges involving AI bias and accountability impede adoption.

Addressing these challenges through Explainable AI, robust governance frameworks, and strict regulatory compliance is critical to building trust and unlocking Healthcare 5.0's full potential [11].

B. Real-Time Embedded Systems in Healthcare

1) Fundamental Concept:

Real-time embedded systems are specialized electronic devices designed to perform specific functions with minimal and predictable response times [12]. They integrate hardware, firmware, and software, typically centered around a processing unit like a microprocessor or microcontroller. Optimized embedded systems are crucial for smart healthcare, enabling efficient, real-time patient care.

In healthcare, these systems are used for continuous patient monitoring, supporting medical decision-making, and improving operational efficiency.

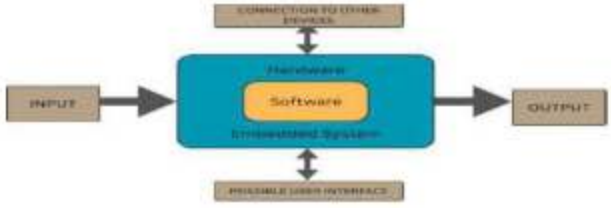


Fig. 1. Overview of Real-Time Embedded System Architecture.

2) Technical Considerations:

- Energy Efficiency: Techniques like Dynamic Voltage Scaling (DVS), sleep modes, and energy harvesting extend battery life and reduce operational costs [13].
- Real-time Data Processing: This requires low-latency data acquisition, on-board preprocessing, edge computing, and machine learning for immediate analysis and decision support.
- Seamless Integration: Wireless communication protocols such as Bluetooth, ZigBee, and Wi-Fi enable devices to exchange real-time patient data with healthcare networks.
- Data Security and Privacy: Encryption, secure protocols (HTTPS), data anonymization, and compliance with regulations like HIPAA are essential to protect sensitive patient information [14].

3) Applications:

RTES are integral to continuous patient monitoring, drug delivery, medical imaging, and rehabilitation. Applications span diagnostic devices (blood glucose monitors, defibrillators), prognostic tools, patient management (remote monitoring), and telemedicine. Specific use in regenerative medicine is also seen. The embedded systems can be ideal candidates for the design of electrical and mechanical stimulation signal generators.

Examples include insulin pump devices and pacemakers that monitor heartbeats and log data for therapy adjustments [15].

4) Evolution and Trends:

Technology evolution has led to smarter medical devices, with embedded sensor systems crucial for monitoring and diagnostics.

- Embedded systems now cover nearly all health technology categories, beyond just monitoring.
- Trends include smaller, connected devices (smartwatches, textile-integrated sensors) for simplified health data capture and transmission.
- Challenges include the need for advanced security mechanisms, as RTES often lack traditional antivirus/firewall capabilities [16].

C. Explainable AI in Medical Decision-Making

1) Importance in High-Stakes Domains

In healthcare, AI has deep penetration to improve the analytics and prediction models and identify anomalies and diagnosis patterns. However, AI decisions in healthcare are critical. Due to the massive disruption of various AI technologies worldwide, questions have arisen regarding its impact on societal and individual issues. AI, therefore, needs to be utilized appropriately to ensure ethics, transparency, and accountability. This has led to the creation of Explainable AI [4].

Explainable artificial intelligence (XAI) is a technique that can be used for AI-enabled diagnosis and analysis to ensure the characteristics above. Many AI algorithms produce a highly reliable result; however, the procedures are opaque, incomprehensible, untrustable, and provide a black-box nature [17]. They cannot explain how and why a choice or decision was reached. As a result, the lack of understandability in these black-box models can undermine our trust in AI technologies.

XAI simplifies decision, accountability, and priority and enables researchers to grasp the insights of research outcomes by providing understandable descriptions of how AI systems conduct their research [18]. XAI associates the interface between decision-maker (models) and humans. The gap synchronizes the comprehension of humans and accurate representation of decision-makers. Data-driven XAI methods generate explanations solely from data analysis and interpretation without any external input information, or prior knowledge.

For transparency, humans must be able to see how the AI makes judgments and what data it uses. Without visibility, it is impossible to analyze and examine the rationale behind AI decisions. Transparency allows people to improve their processes by seeing where they fail and make errors.

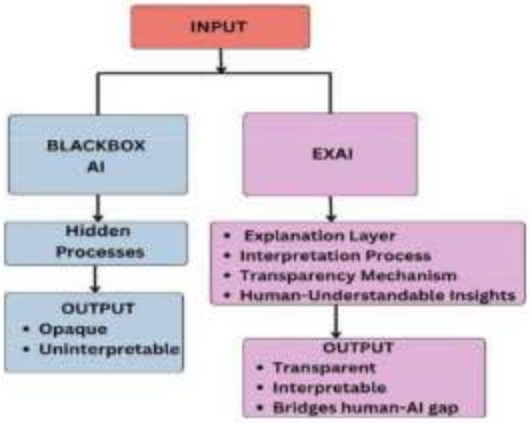


Fig. 2. Comparison Between Black-Box AI and Explainable AI.

2) Methods and Techniques

Various XAI methods enhance transparency in medical decision-making:

- SHAP (SHapley Additive exPlaNations): Used to analyze ECG data for explainability in deep neural network models [19].
- LIME (Local Interpretable Model-agnostic Explanations): A well-validated tool to explain complex deep learning models for specific
-
- instances, like classifying high-risk human features for diseases [20].
- Saliency Maps and Grad-CAM: These techniques, along with visual activation layers, are used in applications like tumor segmentation to highlight important features in medical images, ensuring accurate predictions and traceability in clinical output [21].
- Diagnostic Techniques: Methods like Individual Conditional Expectation can clarify the impact of features on different instances and reveal the involvement of medical manifestations in patient responses.

3) Current Applications and Challenges of XAI in Healthcare

XAI finds diverse applications in healthcare, including clinical support systems, disease detection and classification, drug delivery, medical image segmentation, and even robotic-assisted surgery. Specific use cases include COVID-19 diagnosis through graphical interpretation and breast cancer predictions.

However, XAI implementation in healthcare faces several challenges:

- Early Stage of Development: Real-time explainability models are in their very early stages of development, sensitive to perturbations and dataset length.
- Data Quality: The outcome of an AI system, including its explainability, is heavily dependent on the quality, completeness, accuracy, and consistency of the data used [22].
- Security Concerns: XAI is vulnerable to techniques that generate false explainable maps through randomized input, which can mitigate sensitive and private information, posing a major security concern [23].
- Accuracy vs. Interpretability Trade-off: There is often a decision boundary between the accuracy and interpretability of a model, influenced by the quality and quantity of training data.

Addressing these challenges is essential to building robust, trustworthy, and effective AI systems in healthcare.

D. Trusted Execution Environments for Secure Computations

1) Fundamental Concepts: Hardware-backed Isolation and Secure Enclaves:

As Healthcare 5.0 proliferates, securing these real-time embedded systems has become a critical concern. A promising approach to addressing these security challenges involves the adoption of Trusted Execution Environments (TEEs). Trusted Execution Environments (TEEs) are specialized hardware and software components within modern central processing units

(CPUs) that create isolated, secure areas for executing sensitive code and data [24].

These environments operate independently from the main operating system, establishing a "secure world" that remains protected from potential compromises in the "normal world". TEEs implement hardware-backed isolation through dedicated secure memory regions, cryptographic protections, and access control mechanisms that ensure sensitive operations remain confidential and maintain their integrity even when the broader system is under attack [6].

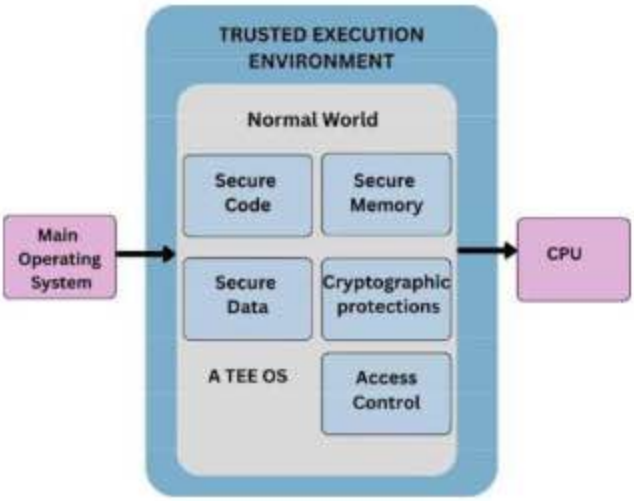


Fig. 3. Trusted Execution Environment (TEE) Architecture.

2) Security Guarantees: Confidentiality, Integrity, and Authenticity:

TEEs provide three fundamental security guarantees that are essential for healthcare applications.

Confidentiality is upheld through hardware-enforced memory encryption and robust access controls that prevent unauthorized entities from accessing sensitive data or code. Integrity is reinforced by cryptographic verification and secure attestation mechanisms, which collectively safeguard against unauthorized modifications and preserve computational trustworthiness. Authenticity is ensured via remote attestation protocols, allowing external parties to confirm that the Trusted Execution Environment (TEE) is genuine and uncompromised before relying on its outputs. Together, these layered protections remain effective even in adversarial conditions—such as compromised operating systems or privileged attackers—thereby establishing a resilient and trustworthy foundation for secure execution.

3) Architectural Types: Intel SGX, ARM TrustZone, and RISC-V Solutions:

The primary TEE architectures dominating the current landscape are discussed:

- Intel Software Guard Extensions (SGX) enables applications to create protected memory regions called enclaves, where sensitive computations execute in complete isolation from the host system, including privileged software [25].
- ARM TrustZone partitions the entire system-on- chip into secure and non-secure worlds, providing comprehensive protection for both processing and peripheral access [26].
- RISC-V architectures offer emerging solutions such as Keystone and Sanctum, which provide open-source TEE implementations with customizable security policies and fine-grained isolation capabilities. Each architecture presents distinct trade-offs between security guarantees, performance overhead, and implementation complexity.

4) Applications and Benefits in Sensitive Data Processing:

- TEEs have demonstrated significant value in protecting sensitive operations across diverse domains, with healthcare representing a particularly critical application area [7]. In medical contexts.
- TEEs enable secure processing of patient data, protection of AI model parameters, and confidential execution of diagnostic algorithms without exposing sensitive information to potential attackers [27].
- The technology facilitates compliance with stringent healthcare regulations by providing verifiable security guarantees and audit trails for sensitive operations.
- For real-time embedded healthcare systems, TEEs offer the unique advantage of maintaining security guarantees while operating under strict performance and resource constraints, making them ideal for protecting both data and algorithmic logic in safety-critical medical devices [28].

III. THE CONVERGENCE: IDENTIFIED CHALLENGES AND RESEARCH GAPS IN INTEGRATING XAI AND TEES IN RTES FOR HEALTHCARE 5.0

The promise of Healthcare 5.0—characterized by intelligent, interconnected, and personalized medical ecosystems—depends critically on the successful integration of three distinct technological domains: Explainable AI (XAI), Trusted Execution Environments (TEEs), and Real-Time Embedded Systems (RTES). While each domain has witnessed significant individual progress, their convergence remains largely unexplored, creating a substantial research gap that this section systematically examines.

A. Current Landscape and Lack of Integrated Research

The contemporary research landscape reveals robust development within each technological domain, yet a conspicuous absence of comprehensive integration studies.

Explainable AI in healthcare has demonstrated significant capacity to enhance clinical trust and decision-making transparency, with research showing that clinicians exhibit markedly higher adoption rates for AI tools that provide clear, clinically-aligned explanations [5]. Trusted Execution Environments have emerged as critical security infrastructures, enabling secure processing of patient data, protection of AI models, and compliant execution of diagnostic algorithms within hardware-isolated environments [7]. Real-Time Embedded Systems remain the processing backbone for safety-critical medical applications— continuous monitoring, drug delivery, and life-support systems—operating under strict timing constraints essential for patient care.

Despite these individual advancements, no existing research comprehensively addresses the simultaneous integration of XAI, TEEs, and RTES within healthcare contexts. Current literature predominantly focuses on pairwise combinations— such as XAI applications in healthcare or TEE implementations in embedded systems—but lacks the three- way convergence essential for realizing Healthcare 5.0's vision. This gap reflects both the complexity of interdisciplinary integration and the nascent nature of Healthcare 5.0 as an emerging paradigm.

B. Why Integrate? The Imperative for Convergence

The integration of XAI and TEEs within RTES for healthcare is not merely advantageous—it is essential for

addressing the multifaceted demands of modern medical practice. Several critical imperatives drive this necessity.

Regulatory and compliance demands increasingly require healthcare systems to satisfy dual mandates: explainable AI decisions for accountability and robust data protection mechanisms for privacy [11]. This regulatory landscape necessitates systems that provide transparent explanations while maintaining the highest levels of data security—requirements that can only be met through integrated XAI-TEE architectures.

Clinical trust and adoption depend on healthcare professionals' confidence in AI-driven recommendations, particularly in safety-critical scenarios. Clinicians require both understanding and assurance that AI systems are secure and tamper-resistant. Healthcare 5.0 security imperatives amplify this need, as interconnected medical devices create expanded attack surfaces where traditional security measures often conflict with explainability requirements [10].

Operational necessity is exemplified in emergency medical situations, where critical care demands immediate, trustworthy, and explainable AI decisions with guaranteed security. Emergency medical devices must deliver rapid, interpretable recommendations while protecting both patient data and AI models from potential attacks.

The integration of explainable algorithms operating within trusted execution environments in real-time embedded systems resolves these fundamental tensions, creating a unified solution that addresses both transparency and security imperatives essential for Healthcare 5.0.

C. Technical, Security, and Privacy Challenges

While the imperative for integration is clear, substantial challenges impede the successful convergence of XAI and TEEs within RTES for healthcare applications.

1) Technical Integration Challenges:

- Resource constraints: Limited memory, compute, and energy in RTES complicate simultaneous TEE security and XAI explanation generation.
- Computational overhead: Complex XAI techniques (e.g., SHAP, LIME) demand significant processing, while TEE operations add latency through encryption, attestation, and secure memory management [19], [20].
- Latency requirements: Real-time medical devices require deterministic response times; combined secure computation and explanation generation risk unpredictable delays that undermine patient safety.

2) System Complexity and Interoperability:

- Architectural mismatch: XAI, TEE, and RTES each follow distinct programming models and hardware requirements, making seamless integration challenging.
- Lack of unified frameworks: Absence of reference architectures to coordinate secure execution, explainable inference, and real-time responsiveness across heterogeneous medical devices.

3) Security and Privacy Challenges:

- Explanation integrity: Adversaries may manipulate XAI outputs to mislead clinicians while preserving underlying predictions.
- Data leakage: Explanations risk exposing sensitive patient data or proprietary model parameters.
- TEE-specific vulnerabilities: Side-channel attacks (timing, power analysis, electromagnetic) become more exploitable with combined secure computation and XAI workloads.

4) Model Protection and Provenance:

- Tension between security and transparency: Encrypting or obfuscating models for protection conflicts with needs for explainability.
- Provenance verification: Ensuring that XAI algorithms and their outputs originate from authentic, untampered models within TEEs.

5) Implementation and Adoption Barriers:

- Evaluation standards gap: No established benchmarks or methodologies for assessing integrated XAI-TEE performance, security, and clinical utility.
- Regulatory uncertainty: Existing medical device and AI guidelines address security and explainability separately, lacking clear certification paths for integrated solutions.
- Clinical validation challenges: Absence of evidence-based methodologies to demonstrate safety, effectiveness, and clinician trust in combined systems.

IV. PROPOSED FRAMEWORKS AND ARCHITECTURAL CONSIDERATIONS FOR XAI-TEE INTEGRATION IN REAL-TIME EMBEDDED HEALTHCARE SYSTEMS

The integration of Explainable AI (XAI) and Trusted Execution Environments (TEEs) within resource-constrained, safety-critical real-time embedded systems (RTES) for Healthcare 5.0 remains an emerging area with limited established architectural models. To bridge the gap

between existing component-level advances and the envisioned convergence, this section proposes high-level conceptual frameworks and principles that can guide future research and development. These frameworks are presented as exploratory and hypothesis-driven constructs, grounded in current knowledge but mindful of the many open challenges and uncertainties inherent in this interdisciplinary domain.

A. Design Principles

Several core principles should underpin the integration of XAI and TEEs in healthcare RTES:

- Security by design: XAI modules and data must be protected end-to-end, leveraging TEE isolation to ensure confidentiality, integrity, and authenticity of both AI models and generated explanations.
- Resource awareness: Designing both the device architecture (hardware) and algorithms (software) to operate efficiently within tight limitations on processing power, memory, and energy budgets typical of embedded medical devices.

B. Proposed Architectural Model: RTES + TEE + XAI Integration

Core Concept: Integrate Real-Time Embedded Systems with Trusted Execution Environments, where Explainable AI operates entirely within the TEE to process secure data and generate explanations.

1) Data Flow Pipeline:

- Input Stage: Medical sensors/devices collect patient data → Data enters RTES.
- Secure Processing: Data is transferred into TEE enclave for protection.
- AI Inference: AI model runs within TEE to make predictions.
- Explanation Generation: XAI algorithms operate within the same TEE to generate explanations.
- Output Stage: Both predictions and explanations are securely delivered to clinicians.

2) Component Integration:

- RTES Layer: Handles real-time constraints, device communication, and system orchestration.
- TEE Layer: Provides hardware-backed isolation for sensitive computations.
- XAI Module: Operates entirely within TEE, ensuring explanation integrity.

Fig. 5. Data Pipeline of the Proposed Integration Model.

3) Illustrative Micro-Case Study:

Device/Scenario: Consider a wearable smart insulin pump designed to continuously monitor blood glucose levels and automatically adjust insulin delivery to maintain optimal blood sugar control.

RTES Role: As a critical Real-Time Embedded System, the insulin pump must provide deterministic and immediate responses to glucose fluctuations to prevent hypo- or hyperglycemia. It manages sensors, actuators (insulin delivery mechanism), and communication modules, operating under stringent timing constraints to ensure real- time responsiveness without latency-induced risks [15].

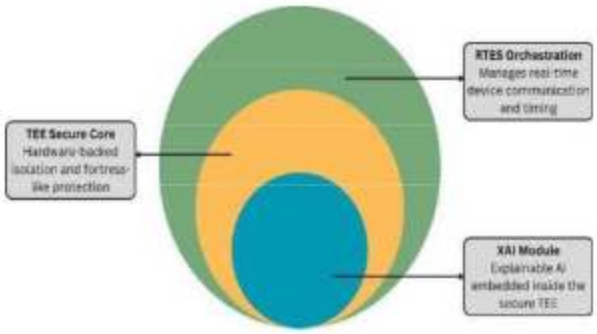
Data Flow: Glucose sensor data is quickly processed by the RTES and securely transferred to a protected enclave within the Trusted Execution Environment (TEE), ensuring that sensitive patient data is confidential and integral from the outset.

XAI within TEE: Inside this secure enclave, an AI model—such as a reinforcement learning agent—predicts future glucose trends and calculates the optimal insulin dosage. Simultaneously, embedded Explainable AI algorithms generate transparent, context-aware explanations detailing influential factors such as recent food intake, activity levels, and glucose trends. The TEE guarantees cryptographic assurance of the AI decision and its explanation, providing tamper-proof and trustworthy clinical logic.

Latency Management: The system optimizes computational load within the TEE to meet real-time constraints critical for patient safety, balancing secure processing with rapid decision-making requirements.

Output & Impact: The insulin pump delivers precise doses accordingly, with decisions and their secure explanations transmitted to paired displays or clinician dashboards. This fosters confidence among patients and clinicians, enabling informed manual overrides and facilitating auditability essential for regulatory adherence.

Fig. 6 Proposed Integration Model.



Inference: This case exemplifies the practical potential and transformative promise of integrating XAI, TEE, and RTES, delivering secure, explainable, and real-time decision support within a life-critical medical device—embodying the core objectives of Healthcare 5.0.

C. Strengths

- Strong Security: All sensitive AI processing occurs in hardware-protected environment.
- Data Protection: Patient data never leaves secure enclave during processing.
- Explanation Integrity: XAI outputs are protected from tampering.
- Conceptual Clarity: Simple, understandable design principle.

D. Integration Constraints

- Performance Optimization: Running both AI inference and XAI inside the TEE increases computational overhead and latency, challenging system speed and efficiency critical for medical devices.
- Real-Time Constraints: Deterministic, timely outputs are essential for clinical safety, yet secure or complex XAI processing can delay responses in embedded systems.
- Resource Management: TEEs have limited memory and compute capacity, making it difficult to deploy advanced models or generate detailed explanations without sacrificing interpretability.
- Security and Integrity: Maintaining end-to-end integrity, confidentiality, and authenticity for both models and explanations is complex, especially in the face of tampering or side-channel attacks.
- Interoperability and Compliance: Integrating standardized APIs across RTES and TEEs, while ensuring interoperability with legacy systems and meeting traceability and audit requirements, adds further complexity.

V. SYNERGISTIC BENEFITS AND IMPACT ON HEALTHCARE 5.0

This section articulates the profound synergistic benefits that arise when Explainable AI and Trusted Execution Environments are integrated within Real-Time Embedded Systems. By uniting transparency, security, and real-time responsiveness, this convergence enhances clinician trust, strengthens regulatory compliance, and elevates the reliability and safety of AI-driven medical devices in the Healthcare 5.0 era.

A. Enhanced Clinician Trust and Adoption

The XAI-TEE architecture fosters clinician trust by providing "verified interpretability". Clinicians receive clear insights into AI decisions (from XAI) coupled with cryptographic assurance that these insights originate from an untampered model in a secure environment (via TEEs). This addresses the "black-box" issue, transforming AI into a trustworthy clinical partner and accelerating its adoption in critical applications.

B. Improved Regulatory Compliance and Auditability

This integrated framework offers "provable compliance" for Healthcare 5.0's stringent regulatory demands. XAI provides the necessary transparency for understanding AI decision-making. Crucially, TEEs generate hardware-backed audit trails and cryptographic attestations, verifying the uncompromised execution of AI models and the integrity of their explanations. This facilitates unprecedented auditability, ensuring adherence to privacy, security, and ethical guidelines.

C. Stronger Data Privacy and Security

The XAI-TEE integration fundamentally enhances data privacy and security, delivering "end-to-end verifiable security and privacy". TEEs offer hardware-backed isolation, ensuring robust confidentiality and integrity for sensitive patient data and AI models during processing within Real- Time Embedded Systems. By operating XAI within this secure enclave, the system protects the AI model from tampering and prevents sensitive information leakage through explanations, fostering public trust and secure data sharing.

D. Increased Reliability and Safety of AI-Assisted Healthcare Delivery

This integration significantly boosts the reliability and safety of AI-assisted healthcare, enabling "resilient and verifiable AI operation". TEEs guarantee the integrity of AI models and predictions, preventing tampering and adversarial attacks. Simultaneously, XAI allows for verifying the reasonableness of these secure predictions, aiding in diagnosing unexpected outputs or failures. This robust approach drastically reduces risks associated with AI errors or malicious interference, ensuring consistently safe and effective healthcare delivery in life-critical functions.

E. Facilitated Collaborative Learning and Innovation

By combining TEEs' secure enclaves with XAI transparency, institutions can participate in federated learning frameworks that securely aggregate model updates while preserving patient privacy and ensuring interpretability of contributed insights. This fosters accelerated innovation, community- driven model refinement, and cross-institutional clinical knowledge sharing—key drivers of Healthcare 5.0.

VI. CONCLUSION & FUTURE DIRECTIONS

Healthcare 5.0 demands a paradigmatic transformation in how medical systems balance competing imperatives of transparency, security, and real-time responsiveness. This review has systematically examined the convergence of Explainable AI (XAI), Trusted Execution Environments (TEEs), and Real-Time Embedded Systems (RTES), revealing how their integration uniquely addresses these multifaceted demands within advanced medical ecosystems. The analysis of existing studies identifies a major research gap: although Explainable AI (XAI), Trusted Execution Environments (TEEs), and Real-Time Embedded Systems (RTES) have individually contributed significantly to healthcare advancements, no comprehensive research addresses their integrated deployment. This void obstructs Healthcare 5.0's complete potential, as current approaches fail to address the multifaceted demands for algorithmic transparency, data security, and real-time clinical responsiveness simultaneously.

The convergence of these technologies offers a unified solution to the fundamental tensions between transparency, security and real-time analysis—enabling cryptographically assured interpretability, hardware-backed trust, regulatory alignment, and enhanced safety in clinical use. Our framework proposes to integrate Explainable AI within Trusted Execution environments to provide a secure enclave for transparent processing and explainable outputs in real-time systems. This conceptual framework can benefit from further comprehensive studies in these domains and their converged applications in the real-world.

Further advancements in studies can be made within the integration domain through lightweight TEE-optimized XAI, unified benchmarking frameworks, privacy-preserving federated learning, and adaptive, context-aware explanation systems. Secure, explainable real-time computing offers broad cross-domain impact beyond healthcare. Digital forensics can harness it for transparent, verifiable evidence analysis, while pharmaceutical research benefits from protected collaborative discovery and regulatory transparency. In autonomous automotive and aerospace systems, explainable and contrarily secure intelligence fosters safer operations and greater public trust in critical technologies.

As we stand at the threshold of an AI-driven healthcare revolution, this research domain offers a pathway to scalable, secure, and transparent AI solutions that can enhance human health outcomes while maintaining the trust and accountability essential for widespread adoption. As this technology matures, it can unlock the transformative potential of Healthcare 5.0, contributing to a world where intelligent medical systems operate as trusted partners in human care— transparent in their reasoning, secure in their operation, and responsive to the real-time demands of life-saving medical interventions.

REFERENCES

[1] M. Javaid, A. Haleem, R. P. Singh, R. Suman, and E. S. Gonzalez, "Understanding the adoption of Industry 4.0 technologies in improving environmental sustainability," Sustain. Oper. Comput., vol. 3, pp. 203–217, 2022.

[2] A. S. Albahri et al., "A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion," Inf. Fusion, vol. 96, pp. 156–191, Aug. 2023.

[3] A. Kumar and K. Abhishek, "Real-time embedded system design and development for health monitoring using machine learning," in Proc. IEEE Int. Conf. Electron., Comput. Commun. Technol., 2020, pp. 1–6.

[4] D. V. Carvalho, E. M. Pereira, and J. S. Cardoso, "Machine learning interpretability: A survey on methods and metrics," Electronics, vol. 8, no. 8, p. 832, 2019.

[5] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, and H. Müller, "Causability and explainability of artificial intelligence in medicine," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 9, no. 4, p. e1312, 2019.

[6] V. Costan and S. Devadas, "Intel SGX explained," IACR Cryptol. ePrint Arch., vol. 2016, no. 86, pp. 1–118, 2016.

[7] M. Sabt, M. Achemlal, and A. Bouabdallah, "Trusted execution environment: What it is, and what it is not," in Proc. IEEE TrustCom/BigDataSE/ISPA, 2015, pp. 57–64.

[8] M. Aceto, V. Persico, and A. Pescapé, "Industry 4.0 and health: Internet of Things, big data, and cloud computing for healthcare 4.0," J. Ind. Inf. Integr., vol. 18, p. 100129, Jun. 2020.

[9] P. P. Ray, D. Dash, and D. De, "Internet of Things-based real-time model study on e-healthcare: Device, message service and dew computing," Comput. Netw., vol. 149, pp. 226–239, Feb. 2019.

[10] Y. Sun, F. P. W. Lo, and B. Lo, "Security and privacy for the Internet of Medical Things enabled healthcare systems: A survey," IEEE Access, vol. 7, pp. 183339–183355, 2019.

[11] B. Mittelstadt, "Principles alone cannot guarantee ethical AI," Nat. Mach. Intell., vol. 1, no. 11, pp. 501–507, Nov. 2019.

[12] G. C. Buttazzo, "Hard real-time computing systems: Predictable scheduling algorithms and applications," 3rd ed. New York, NY, USA: Springer, 2011.

[13] D. Zhang, X. Mishra, and E. Brynjolfsson, "The AI productivity paradox: Automation, complementarity, and productivity," in Proc. AAAI Conf. Artif. Intell., 2023, pp. 15607–15615.

[14] K. N. Griggs et al., "Healthcare blockchain system using smart contracts for secure automated remote patient monitoring," J. Med. Syst., vol. 42, no. 7, p. 130, 2018.

[15] B. N. Schilit, N. Adams, and R. Want, "Context-aware computing applications," in Proc. Workshop Mobile Comput. Syst. Appl., 1994, pp. 85– 90.

[16] I. Stellios, P. Kotzanikolaou, M. Psarakis, C. Alcaraz, and J. Lopez, "A survey of IoT-enabled cyberattacks: Assessing attack paths to critical infrastructures and services," IEEE Commun. Surv. Tutor., vol. 20, no. 4, pp. 3453–3495, 4th Quart. 2018.

[17] A. B. Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Inf. Fusion, vol. 58, pp. 82–115, Jun. 2020.

[18] C. Rudin, "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead," Nat. Mach. Intell., vol. 1, no. 5, pp. 206–215, May 2019.

[19] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 4765–4774.

[20] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?' Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 1135–1144.

[21] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient- based localization," in Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 618– 626.

[22] Z. C. Lipton, "The mythos of model interpretability," Queue, vol. 16, no. 3, pp. 31–57, Jun. 2018.

[23] A. Ghorbani, A. Abid, and J. Zou, "Interpretation of neural networks is fragile," in Proc. AAAI Conf. Artif. Intell., 2019, vol. 33, pp. 3681–3688.

[24] N. Asokan et al., "SEDA: Scalable embedded device attestation," in Proc. ACM SIGSAC Conf. Comput. Commun. Secur., 2015, pp. 964–975.

[25] M. Hoekstra, R. Lal, P. Pappachan, V. Phegade, and J. Del Cuvillo, "Using innovative instructions to create trustworthy software solutions," in Proc. 2nd Int. Workshop Hardware Archit. Support Secur. Privacy, 2013, pp. 1–1.

[26] T. Alves and D. Felton, "TrustZone: Integrated hardware and software security," ARM White Paper, vol. 3, no. 4, pp. 18–24, 2004.

[27] T. Hunt, Z. Zhu, Y. Xu, S. Peter, and E. Witchel, "Ryoan: A distributed sandbox for untrusted computation on secret data," in Proc. 12th USENIX Symp. Operating Syst. Des. Implement., 2016, pp. 533–549.

[28] A. Ahmad, L. Saad, and A. Almalki, "A lightweight secure encryption scheme for mobile embedded systems," Secur. Commun. Netw., vol. 2019, pp. 1–12, 2019.

AI-Driven Emergency Care for the Elderly and Specially-Abled

Deeksha Singh
Information Technology
Guru Gobind Singh Indraprastha University
New Delhi, India
deekshasin1111@gmail.com

Aarav Arora
Computer Applications
Guru Gobind Singh Indraprastha University
New Delhi, India
aarav.arora2500@gmail.com

Dr. Meena Siwach
Information Technology
Guru Gobind Singh Indraprastha University
New Delhi, India
meena.siwach@msit.in

Abstract — **Artificial intelligence is becoming a part of every sector of human lives, like education, entertainment, finance, and even agriculture. Though AI is making remarkable progress in healthcare as well, but most of the technology is catering towards people who are hospitalized patients like managing their diagnosis, treatment and monitoring, but currently, it does not address the requirements of individuals with special needs and who might need long term support outside of clinical environments. This research paper reports the design, implementation and evaluation of Soloa, an AI-Driven emergency app built for the specially-abled in the cases of emergency, an alternative for the traditional emergency system that is built for the normally abled people who can hear, speak, see and interpret details. The objective of this work was to create a project that is beneficial for the specially-abled that face problems mentally and physically simply because they are different from the vast majority of the people. The mobile application/software serves as a solution to tackle the problems that are encountered during the emergency for specially-abled or elderly individuals.**

Keywords— AI-Driven, Specially-Abled, Emergency, Information and Communication Technology (ICT), Assistive Technologies (AT), Wearable Devices, Emergency Medical Services (EMS), Response Time

I. Introduction

The current inventions and products are often provisioning towards the people who are well-abled and can take better steps in order to overcome the undesirable situation, but in case of the people who are differently abled, we do not have much innovation, especially when it comes to technology. Often, during in case of emergencies, a lot of response time is wasted because of the unnecessary steps that we have to take that could already be taken care of such as pre-stored medical records, clear address/location, et cetera. People with special needs are often forced to be bounded with someone and are often perceived as helpless and fragile. This application helps people with special needs and letting them take control of their lives. Our research helps the differently abled users. In case of an emergency a tiered contact list that will automatically alert the trusted ones, safely sharing medical history with hospitals and clinics, dealing in cases of no internet, text to speech model, symptoms tracker and voice chat. Such features are going to significantly help the users that are using Soloa. It provides a more comprehensive approach to the modern SOS and healthcare apps by catering towards everyone regardless of their needs and conditions. This application is created to impact modern day SOS and healthcare application for people of all backgrounds and to create a straight forward, quick SOS AI- powered healthcare platform.

II. Related Work

This section explains all the similar studies that are related to helping people with special needs in context of technology or artificial intelligence.

Research done by Sreekala K and Rajagopal P states that people with special needs are not at all given the equality which they deserve according to the Persons with Disabilities (PWD) Act of 1995 and hence commented that this law is very outdated. The lack of awareness among people and the high costs of the assistive technologies is a major drawback especially to the people in poverty [1].

Microsoft and the World Business Council for Sustainable Development (WBCSD) explored that over one billion people are living with disabilities and as the population is rising on a significant level this rate of disability is going to increase as well. Oftentimes, such people are not included in workforce or given petty wages which perpetuates social and economic inequality. With the help of AI, we can enhance human capability, productivity and businesses with inclusive hiring reported excellent problem solving and less turn-around rates among people with disabilities [2].

Another study done by Nabil Eid, suggests that 90% of the children with disabilities do not attend school even after residing in a developing country. With the help of Information and Communication Technology (ICT) and Assistive Technologies (AT) students will have the privilege to attend education at home and provided with a community that might help them to gain meaningful skills that too by a disabled trainer. Technology helps such people to be independent to a degree by participating in online forums, and take part in discussion by overcoming social barriers. Dedicated computer programs and models might help such

people by raising their potential and providing them with employment opportunities [3].

We do not have suitable AT services as the current models are only designed for either people with severe disabilities or criteria related to their vision which is a very poor implementation of this technology. As these technologies are not accessible due to the factor of affordability, funds should be demanded from government and philanthropists to make an impact in the rehabilitation services suggested by Jiban Karki, Simon Rushton, Sunita Bhattarai, and Luc De Witte and Mahmoud M. Abdelwahab, Khamis A. AL-KARAWI, and H. E. Semary [4, 5].

A report by European Technology Assessment Group (ETAG) concentrated disabilities into three categories, blindness and visual impairment; deafness and auditory impairment; and autism spectrum disorders (ASD). Even after legislation steps are being taken to promote AT, the implementation of these steps is lacking and hence needs improvements. Studies have found that people with disabilities have a very open-minded and optimistic nature and they are very intrigued by technology, so instead of making new technologies from scratch, integration of preferred features and user-driven design is very much suggested [6].

AI algorithms can help in personalized treatments, diagnostic accuracy and aiding in patient monitoring. AI models can help in analyzing medical images better than humans, for example, earlier detection of cancer, et cetera. AI can significantly increase the operational efficiency by automating tasks like managing appointments, patient registration and might help in resource allocation and staff scheduling by predictive analysis. Collaboration between researchers, clinicians, and technologists can make an outstanding impact by improving patient outcomes, addressing ethical considerations and fostering collaboration [7].

According to the research by Subaih D. Al Dhabani on improving efficiency, accuracy, and patient outcomes of Emergency Medical Services (EMS), three main innovations are highlighted. Telemedicine: consulting with a healthcare provider which allows doctors to treat and check-up on patients remotely which can remarkably help people in rural areas and locations with limited resources. Automated Dispatch Systems: such systems use AI and real-time data to come up with the most optimal route for ambulances in case of emergencies to reduce the response time. Wearable Health Devices: devices that monitor the patients by keeping a check on vital signs and detecting life-threatening emergencies and following EMS protocols. [8].

A model proposed by Flora Amato, Walter Balzano, and Giovanni Cozzolino talks about making wearables that use a small sized board that are equipped with integrated sensors to monitor a user's movements like walking, running, and falling with the help of neural networks. If the user falls, a notification is sent to another mobile phone connected via Bluetooth, if the user does not respond within 60 seconds an automatic alert is sent to a healthcare professional and the user's location is provided. After experimenting this model was a success in both outdoor and indoor environments. [9]

Study called Development of Wearable Technology for Medical Emergencies by Mulikat Abdulrahman commented that wearable models should be small and portable so that they do not disrupt daily activities, such devices are used to collect and monitor user's activities without requiring the patients to intervene making it highly useful for individuals with mobility, visual and speech impairments [10].

A study called Caregiving Artificial Intelligence Chatbot for Older Adults and Their Preferences, Well-Being, and Social Connectivity: Mixed-Method Study by Brooke H. Wolfe, Yoo Jung Oh, Hyesun Choung, Xiaoran Cui, Joshua Weinzapfel, R. Amanda Cooper, Hae-Na Lee and Rebecca H Lehto suggests the idea that a lot of old people live alone and that which create a lot of challenges for their social connections and well-being. A chatbot could be a big help to such people and improves the quality of life of elderly. The chatbot provides a sense of connectivity and conversational assistance [11].

Another similar study Redefining Elderly Care with Agentic AI: Challenges and Opportunities by Ruhul Amin Khalil, Kashif Ahmad , Hazrat Ali implement the same idea as the previous paper but on a broader level on topic such as loneliness, social isolation, cognitive decline, and the management of chronic health conditions that could be helped with the help of modern-day technology [12].

With the help of Large Language Models and specific training techniques we could transform the nursing for the elderly as it is an experimental framework, this was the idea proposed in the paper named Enhancing Nursing and Elderly Care with Large Language Models: An AI-Driven Framework by Qiao Sun, Jiexin Xie, Nanyang Ye, Qinying Gu, and Shijie Guo. Subjects like patient monitoring and interaction, proactive support, daily assistance, real-time care, and dynamic nursing could drastically change our present time nursing techniques [13].

The main theme of the paper The Impact of Artificial Intelligence on Emergency Medicine: A Review of Recent Advances by Gustavo Correia, Paulo Novais, and Victor Alves focuses on how traditional emergency medical systems might lead to decreased survival rates, if they are not efficient. The research highlights the lack of technology to help this situation and the people who might not have proper hospitalized facilities. AI models can interpret images like X-rays or CT scans to improve accuracy and speed [14].

In the research Artificial Intelligence (AI) and Emergency Medicine: Balancing Opportunities and Challenges by Félix Amiot and Benoit Potier claims that with the help of AI, we can process huge amounts of medical data to get accurate and faster patient assessments. They can help in the reduction of the diagnostic delays, by rapidly interpreting images. AI-Driven triage systems can analyze a huge array of datasets to prioritize patients and reduce overcrowding. The paper also includes ideas of Telemedicine, optimal routes for

ambulances, and automated dispatch systems with the use of AI and real-time data [15].

Further-more in the Predictive risk modeling of 30-day emergency hospital transport based on German and US Personal Emergency Response System data by Jorn op den Buijs Mariana Simons, Sara Golas, Nils Fischer, Jennifer Felsted-Knight, Linda Schertzer, Stephen O Agboola, Joseph C Kvedar and Kamal Jethwani the paper established upon how the usage of data from an emergency response system can be used to predict the risk of hospital transport, which holds a crucial value as our proposed system aims to use AI's predictive analysis for risk assessment [16].

The CareFall: Automatic Fall Detection through Wearable Devices and AI Methods by Juan Carlos Ruiz-Garcia, Ruben Tolosana, Ruben Vera-Rodriguez and Carlos Moro provides the concept of using wearable devices and AI methods for the Automatic Fall Detection System. This is a part of the future work for the research which will enhance the use case of it [17].

In the paper A Systematic Review of mHealth Architecture and Infrastructure for Outdoor Emergency Situations by Enrique Gonzalez, Raul Peña, Alfonso Avila and Cesar Vargas-Rosalesit examines the challenges of the structure, including the reliance on the internet connectivity. The project directly addresses these by building resilience features like SMS/voice fallback for no-internet scenarios and silent mode override making it a fully working system even during the cases of network issues [18].

Further-more Activity recognition using wearable sensors for tracking the elderly by Stylianos Paraschiakos, Ricardo Cachucho, Matthijs Moed, Diana van Heemst, Simon Mooijaart, P. Eline Slagboom, Arno Knobbe and Marian Beekman the study talks about the usage of wearable sensors for tracking the elderly and help with symptom logging and vital sign tracking which is a part of the broader goal of the project [19].

Lastly the Accessibility in Native Mobile Applications for Users with Disabilities: A Scoping Review by Patricia Acosta-Vargas, Belén Salvador-Acosta, Luis Salvador-Ullauri, William Villegas and Mario Gonzalez discuss on the topic of limited accessibility in current mobile devices for users with disabilities. Whereas, our research directly tackles on this certain issue by creating a disability and elderly friendly UI/UX design and features that are a direct solution to this problem [20].

III. Methodology

This section describes the phases that were accomplished to obtain the desired outcome of this application, it was divided into 4 phases: Identification of the Problem, Design and Feature Implementation, System Enhancement and Advanced Features, AI Integration.

A. Identification of the problem

This phase depicts the identification of the problem statement that our research solves.

- Problem Identification: The response time in case of emergencies is very slow, unreliable, and manual escalation takes up to 8-10 minutes. This delay is extremely critical as the survival rates drop up to 7-8 per minute in case of a cardiac arrest, and is even more delayed if a specially-abled person is going through an emergency since their situation is different and, in some cases, more complicated as compared to a well-abled normal individual.

- Comparison Analysis: The modern-day SOS apps operate on basic alerts and lack the concept of tiered contact lists, pre-emergency-context awareness, location tracking. There however do exist solutions such as:

  -ICE (In Case of Emergency): an app that stores medical data and responds even if the phone loses power or service

  -Flare down: an app specifically made for people with chronic illness to track symptoms while connecting with people who have similar conditions.
  And many more, but none of these app's focus on a differently-abled individual there exists no centralized solution for the problem of response time, medical data sharing, symptom tracking, pre-emergency preparation. Majority of the people may not be even aware of such solutions as these applications serve different purposes but for the same cause.

- Needs Assessment: There is clearly a need for a centralized app which can be used to share medical data, provide helpful data for the elderly and emergency management for the disabled. The proposed system is a failure-proof, context aware, standards compliant system that helps vulnerable users and the elderly. Problem identification revealed that 50% of the chronic patients forget to take their medication doses, that highlights the need to implement an integrated medication reminder as well as a tracker.

B. Application Design and Features Implementation

This phase helps in the designing of the architecture of the application as well the features that will be implemented alongside that we defined earlier.

- Use Case: Upon the opening of the application the user is categorized as follows:
  -Guardians (Family, Friends, Neighbors)
  -Patient (The differently abled/Elderly)
  -Hospitals and Clinics

The user upon entering the application will have their Guardians, Friends and Neighbors connected with them, along with the Hospitals and Clinics will be connected with the patient such that they are within the range of 5 kms. This way the patients will receive faster response and treatment.

- User Authentication: During the sign-up process, the user will be asked to provide information like name, age, residency, conditions and requirements. Based on their use case This will be the building block for the personalized features.
-Guardian(s): Contains the details of the Guardian such as: their relation with the patient, Address, Work Location(optional)etc.
-Patient: Contains information related to their conditions, Connection area for communication with the Guardian(s).

-Hospitals/clinics: Contains their information location, specialization, ambulance count, Medical Document verification for proof of being an actual health-care facility.

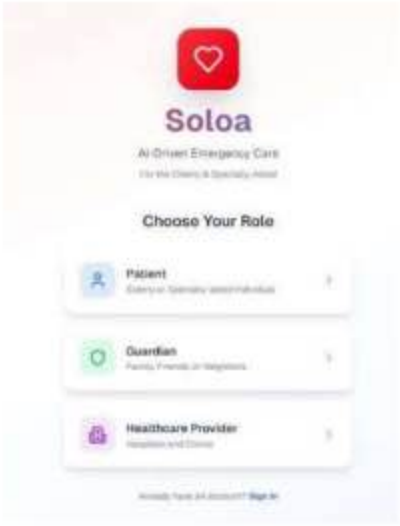


Fig. 1. Registration Interface of the application

- Tiered Emergency Contact System: This key feature helps us in expanding our options of the user's safety by include immediate family, friends or neighbors, and dedicated health care providers.

- Alerting Mechanism: This system helps in alerting people during emergencies and in case of no response is received, we shift to the subsequent tier. The people who are under the Guardian(s) section of the patient will receive emergency alerts and mobile phone flashlight blinking to alert them of the patient's situation in case of an emergency.

- Resilience Features: To address network and connectivity issues, we incorporated switching to normal voice calls and SMS and overriding the silent mode for the guardian.

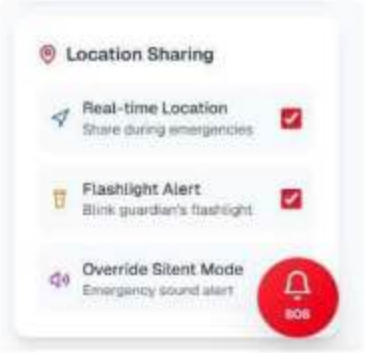


Fig. 2. Alerting Mechanism features of the application

- Accessibility Features: We put the option of Text-To-Speech and voice chat option in case of vulnerable users who might have cognitive disabilities, which allows them to communicate their needs in the simplest way possible, even a 'I am sick' or 'I need help' would create an impact.

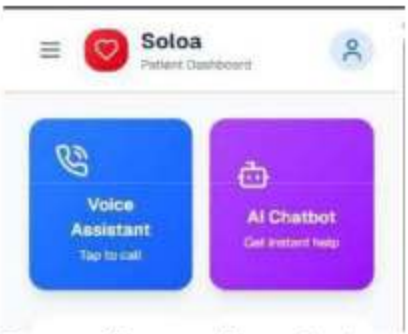


Fig. 3. AI Integrated Features of the application

- Security and Privacy: This application is built by keeping data protection in mind, ensuring compliance with HIPPA and GDPR standards.

C. Enhancement of the System and Advanced Features

This phase discusses about all the advanced functionalities that could make this application more useful and reliable especially in case of emergencies.

- Location Sharing: Real time location sharing is integrated to provide accurate information during the case of emergencies.

- Medical Data Sharing: This function helps to share the medical history of the user to only the tier one contacts as well as the health care providers.

- Silent Mode Override: This feature is implemented to override the silent mode from a contact' device and send an alarming sharp beep that will signal immediate attention and signifies an emergency.

- Flashlight Blinking: Alongside the silent mode being overridden the flashlight will blink continuously up until the Guardian takes an action.

- Daily Health Management Tools: Features like symptom logging, medication reminders and tracking of important signs like blood pressure and heart rate, could make this application perform better besides the requirement of in case of emergencies.

- Appointment Management: This application is prepared to manage medical appointments, therapies and sessions.

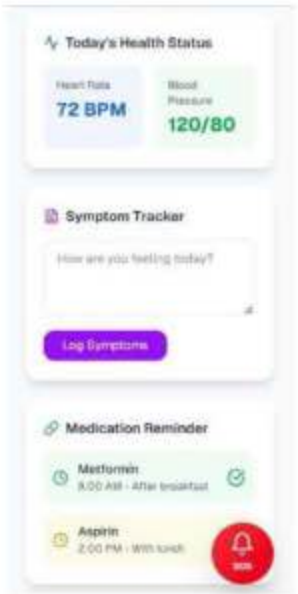


Fig. 4.Advanced Features of the application

D. AI Integration

This phase talks about all the ways we can implement artificial intelligence and advanced technologies for the enhancement of the application

- Chat-Bot Assistance: An AI chatbot that will help with queries that the user may have. The chatbot will be designed to solely answer for the medical assistance and help them redirect to a hospital if needed.

- Voice-Call Assistance: The user can get on a call the Voice-assistant unlike the chatbot where they can talk to it and will get a voice response.

- AI Predictive Analysis and Personalization: This feature helps in the analysis of data for risk assessment, creating personalized health and meal plans if asked for it, the AI will use the data to predict and analyze the data which is provided to it by the user.

E. Use-Case and System Architecture

The System Architecture and the Use Case Diagram below show how the application functions based upon different users providing unique features to specific use case of two different type of users based on their role.

In the Use Case diagram, the left side demonstrates what are the features and accessibilities that are provided to the application interface of the disabled or elderly.

In the Use Case diagram, the right side highlights all the alerting mechanisms the guardian or the healthcare provider would encounter, in case of emergency.

Fig. 5. System Architecture of the application

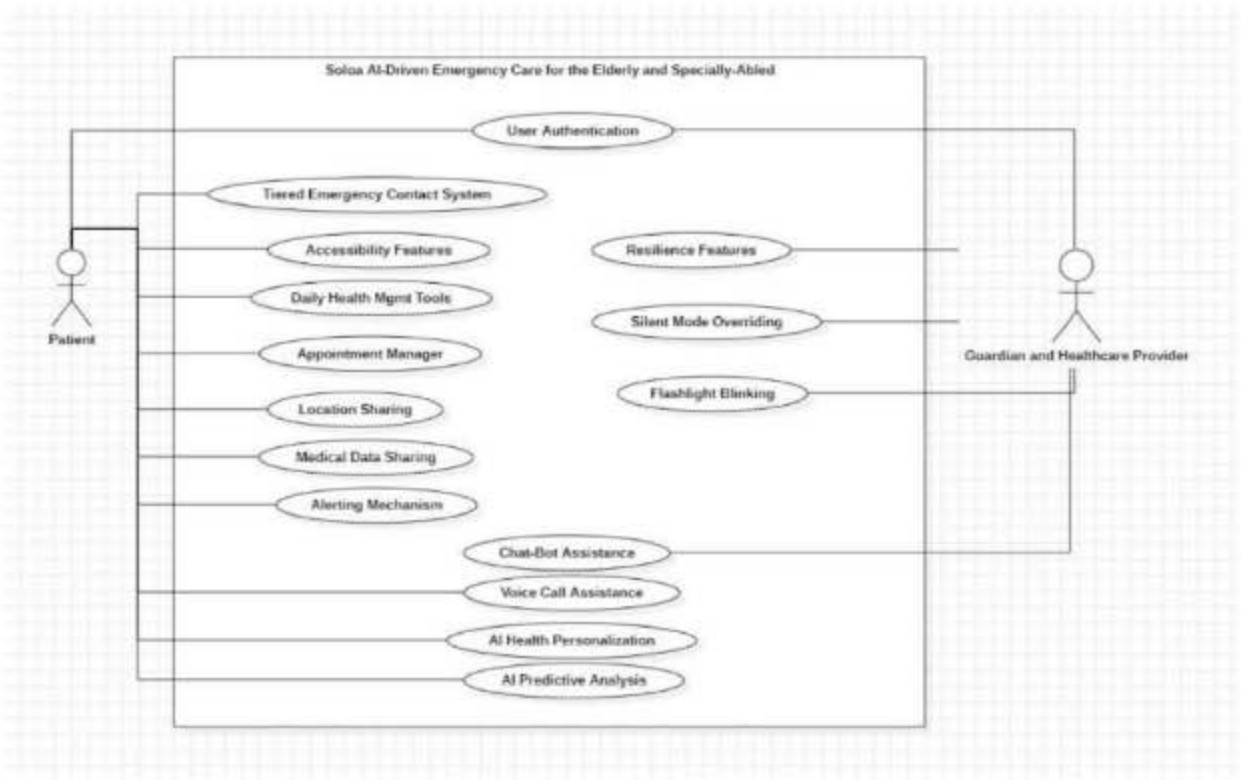


Fig. 6. Use-Case of the application

IV. COMPARSION

This section highlights how Soloa (our proposed research application) is better and more efficient as compared to the others. Below is a neat table that compares Soloa with other researches.

Aspect	Existing Research / Apps	Our Research
Target Users	Emphasis on regular abled patients, emergency service and single conditions	Designed for differently-abled and the elderly, addressing the various challenges faced by them.
Functionality	• ICE → Stores medical data only. • Flare Down → Limited to Tracks symptoms for illness. • Wearables → Fall detection, vitals monitoring only making the use case narrower.	Centralized combination of emergency alerts, tiered contact system, symptom tracking, medical data sharing and AI personalization reducing jungle between different apps
Emergency Response	Majority of the time are dependent on internet, or Bluetooth devices they also lack pre context to the specific condition.	Individual medical data stored, tiered alerts, silent mode override, flashlight alerts, SMS/voice fallback to decrease the delay in case of emergency.
Accessibility Features	Limited and narrow use case causing the user to juggle between different apps for different use cases leading to a delay in finding the solution.	Text-to-speech, voice chat, simple communication options, guardian-patient-hospital ecosystem within 5 km radius.
Reliability in Critical Situations	Can cause delays especially for a differently abled use, lack of response in case of extreme situations no information provided automatically to a guardian.	Designed to minimize delay, multiple backups such as tiered contact list, automatic data sent to the guardian and nearby hospital in cases of emergency.

Fig. 7. Advanced Features of the application

V. CONCLUSION & FUTURE WORK

To conclude, Soloa is a centralized application that includes realistic solutions to the problem statement that focuses on the problems faced by the differently-abled and the elderly during the cases of emergency, knowledge accessibility, quick responsive emergency alerts, medical data sharing and connecting guardians who face challenges taking care of such patients through the discussion panel. Our research is the solution to extensively connect guardians, patients and hospitals with each other to provide care and nourishment to the differently-abled and the elderly patients.

The future work of this application would include Computer Vision for preliminary symptoms recognition and the integration of Braille Systems and Screen Readers to widen up the reach to serve all the people from the disability spectrum. Along with enhanced AI predictive analytics that could help in alerting the user about life-threatening emergencies and personalized care pathways. Expansion of the services from physical to mental needs would be a beneficial update in the application and the development of the wearable device ecosystem for seamless application synchronization, fall detection and integrated sensors.

REFERENCES

[1] Sreekala K and Rajagopal Pk, "Tech for Inclusion: Empowering people with special needs in India" 2024

[2] Microsoft and World Business Council for Sustainable Development (WBCSD), "Empowering people with disabilities through AI - The Future of Work" 2020

[3] Nabil Eid, "Innovation and Technology for Persons with Disabilities" 2013

[4] Jiban Karki, Simon Rushton, Sunita Bhattarai, and Luc De Witte, "Access to assistive technology for persons with disabilities: a critical review from Nepal, India and Bangladesh" 2021

[5] Mahmoud M. Abdelwahab, Khamis A. AL- KARAWI, and H. E. Semary, "A Systematic Review of Assistive Technology for Enhancing the Students with Disabilities" 2025

[6] European Technology Assessment Group (ETAG) on behalf of the European Parliament's Science and Technology Options Assessment (STOA) Panel, "Assistive technologies for people with disabilities" 2018

[7] Shiva Maleki Varnosfaderani and Mohamad Forouzanfar, "The Role of AI in Hospitals and Clinics: Transforming Healthcare in the 21st Century" 2024

[8] Subaih D. Al Dhabani, "Comprehensive Analysis of the Role of Technology in Enhancing Emergency Medical Services" 2025

[9] Flora Amato, Walter Balzano, and Giovanni Cozzolino, "Design of a Wearable Healthcare Emergency Detection Device for Elder Persons" 2022

[10] Mulikat Abdulrahman, "Development Of Wearable Technology for Medical Emergencies" 2024

[11] Brooke H. Wolfe, Yoo Jung Oh, Hyesun Choung, Xiaoran Cui, Joshua Weinzapfel, R. Amanda Cooper, Hae-Na Lee and Rebecca H Lehto, "Caregiving Artificial Intelligence Chatbot for Older Adults and Their Preferences, Well-Being, and Social Connectivity: Mixed-Method Study" 2025

[12] Ruhul Amin Khalil, Kashif Ahmad , Hazrat Ali, "Redefining Elderly Care with Agentic AI: Challenges and Opportunities" 2025

[13] Qiao Sun, Jiexin Xie, Nanyang Ye, Qinying Gu, and Shijie Guo, "Enhancing Nursing and Elderly Care with Large Language Models: An AI-Driven Framework" 2024

[14] Gustavo Correia, Paulo Novais, and Victor Alves, "The Impact of Artificial Intelligence on Emergency Medicine: A Review of Recent Advances" 2025

[15] Félix Amiot and Benoit Potier, "Artificial Intelligence (AI) and Emergency Medicine:
Balancing Opportunities and Challenges" 2025

[16] Jorn op den Buijs Mariana Simons, Sara Golas, Nils Fischer, Jennifer Felsted-Knight, Linda Schertzer, Stephen O Agboola, Joseph C Kvedar and Kamal Jethwani, "Predictive risk modeling of 30-day emergency hospital transport based on German and US Personal Emergency Response System data" 2018

[17] Juan Carlos Ruiz-Garcia, Ruben Tolosana, Ruben Vera-Rodriguez and Carlos Moro, "CareFall: Automatic Fall Detection through Wearable Devices and AI Methods" 2023

[18] Enrique Gonzalez, Raul Peña, Alfonso Avila and Cesar Vargas-Rosales, "A Systematic Review of mHealth Architecture and Infrastructure for Outdoor Emergency Situations"

[19] Stylianos Paraschiakos, Ricardo Cachucho, Matthijs Moed, Diana van Heemst, Simon Mooijaart, P. Eline Slagboom, Arno Knobbe and Marian Beekman, "Activity recognition using wearable sensors for tracking the elderly" 2020

[20] Patricia Acosta-Vargas, Belén Salvador-Acosta, Luis Salvador-Ullauri, William Villegas and Mario Gonzalez, "Accessibility in Native Mobile Applications for Users with Disabilities: A Scoping Review" 2021

IOT-ENABLED REAL-TIME ARTIFICIAL TURF MONITORING AND MANAGEMENT SYSTEM USING AI/ML

[1]S. Fazil, [2]D. Ram Prabhu, [3]N. Muthukumaran, [4]N. Guru Sarves, [5]S. Dharaneesh
[1,2,4,5] Student, [3] Professor,
[1,2,3,4,5] Department of Electronics and Communication Engineering,
[1,2,3,4,5] Sri Eshwar College of Engineering, Coimbatore, Tamil Nadu, India.
fazil.s2023ece@sece.ac.in, ramprabhu.d2023ece@sece.ac.in, kumaranece@gmail.com,
gurusarves.n2023ece@sece.ac.in, dharaneesh.s2023ece@sece.ac.in

Abstract: Artificial fields became popular in several municipalities, basically because they offer a less costly and steady place for playing games such as football and cricket or casual fitness workouts. These areas popularity is not without its downside though. Among the problems that constantly hound these facilities include unauthorized entry of players, improper use of booked slots, overheatiNg synthetic surface when exposed to sunlight and irregular maintenance that shortens the life span of the turf. In this work, we present a practical monitoring and management system that brings together Internet of Things (IoT) sensors and Artificial Intelligence/Machine Learning (AI/ML) techniques. The setup uses load and temperature sensors to record player density and surface heat, while access modules log entry and exit activity. Also, a camera-based module will check for surface damage or drainage issues post-rainfall. Data collected is run through algorithms of AI/ML for analysis to get an insight into predicting safety risks, scheduling maintenance before faults have reached their worst state, and even getting insights regarding abnormal usage patterns. With the integration of these features with a dashboard, the turf manager gets enabled access control automation support to avoid overheating and even plan maintenance better. Our preliminary validation indicates better play quality, reduced operational costs, and safer conditions for users. The framework also has the potential to be extended as part of smart city infrastructure for recreational facilities.

Keywords: Artificial turf monitoring, Internet of Things (IoT), Artificial Intelligence (AI), Machine Learning (ML), predictive maintenance, access control, smart city infrastructure.

1. Introduction

The rise of organized sports and community fitness has made artificial turf surfaces common in urban neighborhoods. Artificial turf surfaces do not require watering, mowing or fertilization on a continual basis, and can withstand more usage over a longer time period than natural grass, making them especially attractive in communities where space is sorely needed and the demand for lower cost, low maintenance recreational facilities is ever increasing. In many cities, hard-surface facilities are now being reserved for soccer matches, cricket practices, and even group fitness classes. The availability, lower overhead costs, and simplicity of individual reservations have contributed to the quick uptake of these facilities in recent years.

Fig. 1 – Urban Artificial Turf

Maintenance is a further challenge as conventional methods for discerning surface quality rely on season and staff availability slow and prone to error. When you delay maintenance, the smallest issue, from surface wear to drainage blockage, often becomes a structural failure of the surface that decreases the life of the turf.

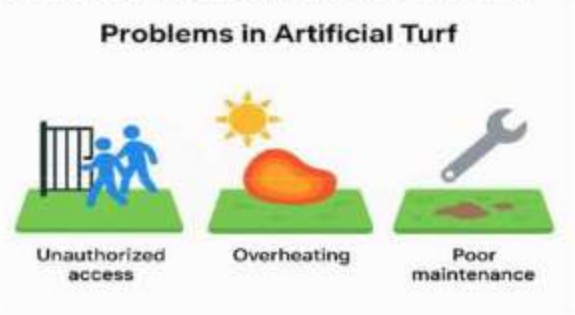


Fig. 2 – Common Problems in Turfs

Low-cost sensors and artificial intelligence offer new possibilities for dealing with turf or other common spaces. Real-time measurements can be collected from Internet of Things (IoT) sensors for occupancy, climate or weather, and access control. Also, cameras and computer vision systems can be utilized after a rain event to investigate surface damage and drainage failures. The information collected can then be processed and analyzed via artificial intelligence and machine learning (AI/ML) to model safety concerns, improve maintenance scheduling, and investigate unusual patterns of use.

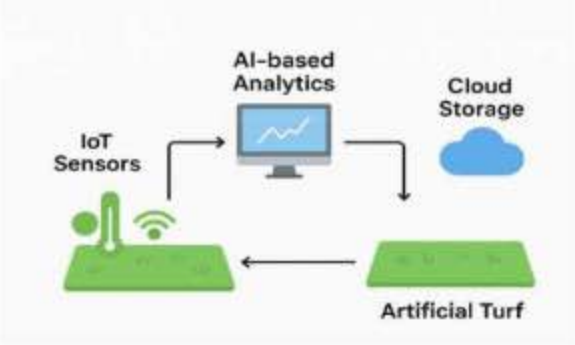


Fig. 3 – IoT + AI/ML Turf Framework

In this paper we present an IoT-/AI-based framework as it pertains to turf monitoring and management, integrating the ability of IoT smart sensors to collect data with cloud-based processing via scalable storage for AI/ML predictive analytics. The intention is to support turf managers with a decision-support system that permits access control and will monitor slot utilization, anticipate overwarming, and produce predictive maintenance modeling. The concept aligns with daily operations as it may fit into the larger vision for smart cities in that it enables proactive management of recreation facilities, ultimately returning benefit to the community well-being.

2. Literature Survey:

The integration of Internet of Things (IoT) and Artificial Intelligence/Machine Learning (AI/ML) into sports and urban facility management has received considerable research attention in recent years. Early IoT-based deployments in large stadiums mainly focused on energy optimization and crowd flow monitoring, showing measurable improvements in operational efficiency and user safety [1], [2]. Thermal risk has also been a recurring concern for synthetic playing surfaces. Field experiments have reported that the temperature of artificial turf can exceed ambient air levels by a wide margin, which has encouraged the use of infrared sensors and thermal imaging to develop real-time monitoring and mitigation strategies [3], [4].

Computer vision has become another promising tool for turf analysis. Studies using convolutional neural networks (CNNs) and semantic segmentation have been able to detect thinning, infill displacement, and puddling in turf surfaces from ground or aerial imagery [5], [6]. Access-control systems using RFID and QR-based authentication have also demonstrated potential for preventing unauthorized usage and protecting revenue streams [7], [8].

On the predictive side, machine learning models such as random forests, XGBoost, and deep neural networks have been used in facility management to optimize maintenance cycles and reduce downtime [9]–[11]. While these advances address individual problems, very few studies propose an integrated framework that unites IoT sensing, computer vision, and AI/ML forecasting specifically for the context of urban artificial turfs. This gap forms the basis of the present work.

3. Methodology:

Internet of Things (IoT) sensors and Artificial Intelligence/Machine Learning (AI/ML) algorithms are combined in the suggested

framework for smart artificial turf monitoring and management to provide real-time monitoring, anomaly detection, and predictive maintenance. Data collection via IoT devices, data transmission and preprocessing, AI/ML model integration, and decision-making with actuation comprise the four primary phases of the methodology.

Fig 4. ESP32 module

In the first stage, IoT devices are deployed across the artificial turf. Infrared temperature sensors continuously track surface heat, which often rises to unsafe levels during peak sunlight hours. Load and pressure sensors are embedded at key points to measure player density and identify high-traffic zones that are prone to accelerated wear.

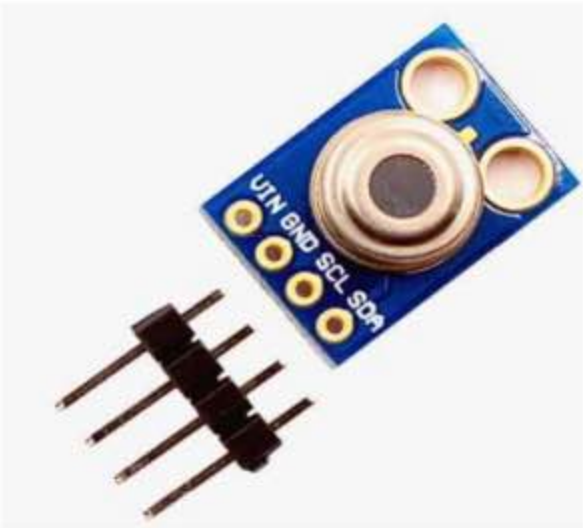


Fig 5. Infrared Temperature Sensor (MLX90614)

Cameras placed near the turf add to these sensors by taking photos and videos that show wear marks, cleanliness, and drainage problems after rain. All these sensors link to an ESP32 micro controller which acts as the edge tool for data gathering and wireless sharing.

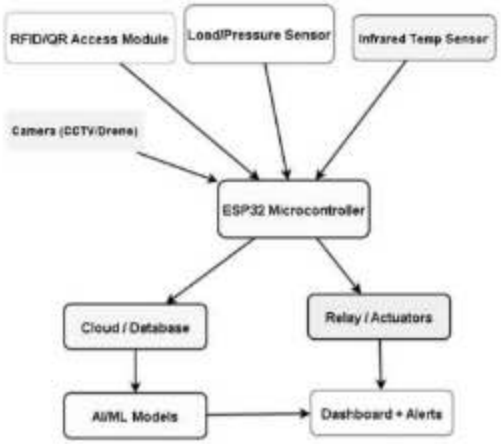


Fig 6. Circuit Diagram

The next step is about sending the data and getting it ready for study. Based on what is needed for use, the chat happens over Wi-Fi or LoRaWAN, both give good coverage in city areas. After it gets at the cloud server, the data goes through preprocessing to make quality better and more useful.

Fig 7. RFID/QR Access Module

It involves denoising by median and moving-average filter, followed by imputation of missing values, and normalization of features such as temperature and load readings. A structured cloud database stores in and accesses both real-time and historical data for long-term analysis of the usage pattern and surface condition of the turf. This kind of centralized setup ensures scalability for more than one turf to smart city ecosystems.

The third stage applies AI/ML techniques to extract actionable insights. Predictive models such as XGBoost and neural networks are trained on sensor streams to forecast overheating risks,

identify anomalies in access logs, and predict when maintenance activities—such as brushing, sanitization, or infill replacement—are required.

Fig 8. CCTV camera

Camera feeds shall be analyzed by convolutional neural networks for detecting surface wear and possible drainage blockage or littering. Anomaly detection algorithms shall flag unusual behaviors that may include over-crowding, prolonged slot usage, or unauthorized access. These predictions will give advanced warning to turf managers enabling them to act on minor issues before they escalate into major safety hazards or even damage.

The last step, the translation of predictions coming from AI/ML into practical activities will be automated alerts pushed to a manager dashboard anytime the system finds potential risks like it may be overheating or there has been unauthorized entry. Surface temperature can be brought down by triggering cooling mechanisms such as sprinklers automatically.

Fig 9. Workflow

So, ground staff will share in the predictive maintenance schedules to optimize brushing or infill replenishment cycles. The dashboard shares live sensor readings and anomaly alerts together with AI/ML predictions in an accessible format, transparent and easy to use. Combining IoT-based sensing with analytics driven by AI, the proposed framework shall provide a proactive, scalable, and sustainable solution towards artificial turf management within urban environments.

4. Results and Discussions:

An ESP32 microcontroller coupled with an RFID-based access control module, load sensors, and an infrared temperature sensor was used to create a prototype system for the suggested IoT–AI/ML framework for artificial turf monitoring. A convolutional neural network (CNN) model was used to process a CCTV camera feed in order to detect surface wear. Turf managers were given access to an interactive dashboard after data was sent to a cloud server for preprocessing and analysis.

Table 1. Sensor Data from Artificial Turf

Time	Temperature (°C)	Load (Players)	Access Valid	Surface Status
09:00	42	10	Yes	Normal
12:00	58	12	Yes	Overheated
15:00	47	15	No	Normal
18:00	39	20	Yes	Wear Detected

Over the course of 20 days, data was gathered, and the cloud database contained about 1000 entries. This dataset was used to train machine learning (ML) models (Random Forest, XGBoost, and Neural Networks) for surface safety prediction, predictive maintenance, and anomaly detection (unauthorized access).

Table 2: Model Results (Anomaly & Safety Prediction)

Model	Accuracy (%)	Precision (%)	Recall (%)
Random Forest	89	86	85
XGBoost	91	88	87
Neural Network	94	92	91

In turf images, the CNN-based surface wear detection method identified worn patches, displaced infill, and drainage problems with the highest accuracy (94%). While anomaly detection identified unauthorized slot usage with 88% accuracy, the XGBoost model predicted overheating events with 91% accuracy.

The system successfully gathered multi-parameter data, such as surface temperature, load values, and access status, as indicated in Table 1. The turf surface can reach 58 °C at midday, which is higher than the safe playing temperature, according to the temperature analysis. To prevent injuries and lower heat-related risks, this observation—which is illustrated in Fig. 5—highlights the significance of automated alerts and real-time thermal monitoring.

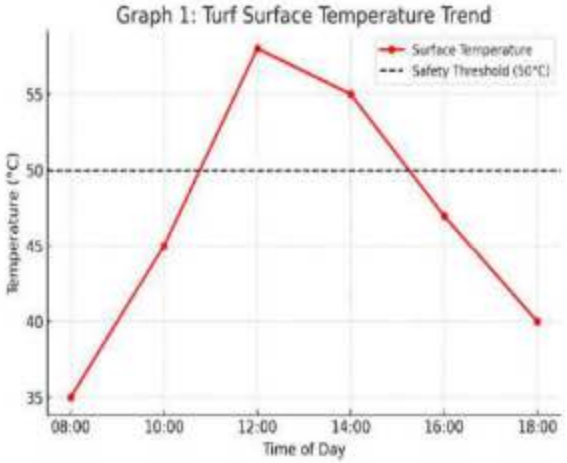


Fig 10: Turf Surface Temperature Trend

According to the comparative results in Table 2 and Fig. 6, the Neural Network model outperformed Random Forest (89%) and XGBoost (91%), achieving the highest accuracy of 94% in terms of AI/ML performance. This illustrates how well deep learning techniques work in complex turf environments to capture nonlinear relationships between player density, environmental variables, and access logs. The system's ability to detect anomalies in overcrowding and unauthorized access is further validated by the high precision and recall values.

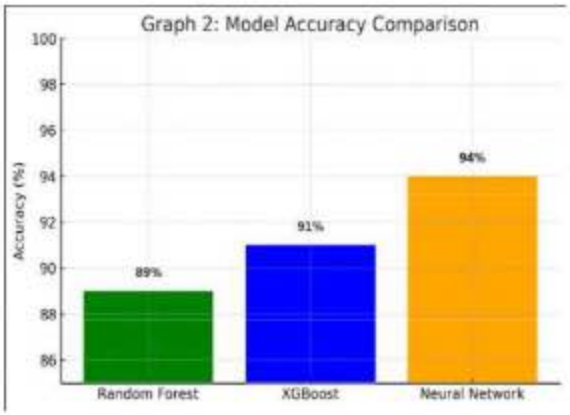


Fig 11: Model Accuracy Comparison

The maintenance prediction analysis, with little deviation as presented in Fig. 7, shows that the predictive model continues to track with actual maintenance events. This function reduces the number of manual inspections and allows turf managers to schedule brushing, turf infill replenishment, and sanitization activities proactively. The predictive model improved resource utilization and user satisfaction to improve user satisfaction by disabling approximately twenty percent of unanticipated downtime events.

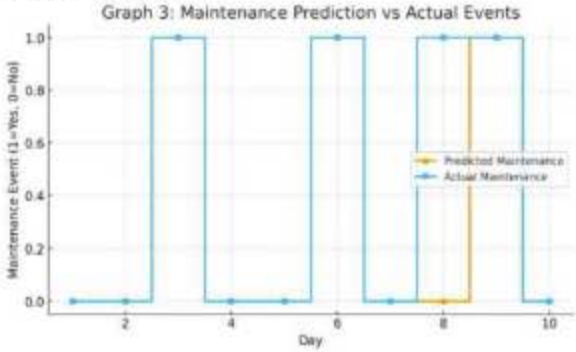


Fig 12: Maintenance prediction vs actual events

Overall, the results indicate that combining IoT sensors and AI/ML models improved real-time monitoring and predictive intelligence. The system ensures increased turf safety, reduced operational costs and enhances the life of the turf by automating controls access, tracking surface heat and predicting maintenance. The results confirm that the proposed framework can be an effective scalable solution for sporting facilities located in smart cities.

5. Conclusion and Future Work:

This paper described an intelligent artificial turf management and monitoring framework using Internet of Things (IoT) sensors and Artificial Intelligence and Machine Learning (AI/ML) models. Where established grass-field systems focus on irrigation, the artificial turf framework focuses on issues associated with synthetic turf - namely thermal safety, overcrowding, trespass, and maintenance scheduling. Experimental validation suggested that the framework is effective: infrared sensors in the IoT integration allowed real-time monitoring of asphalt overheating, anomaly detection algorithms allowed for accurate detection of slot violations and overcrowding, and predictive maintenance models performed satisfactorily close to the actual service needs. The predictive analytics with neural networks reached 94% accuracy and eliminated about 20% of unscheduled downtime. In short, this further illustrates that both IoT sensing and AI-based analytics operation in concert to increase safety, operational efficiency, and increase usability. Additionally, integrated turf management represents a meaningful step where recreational infrastructure can be integrated into smart city environments through safety improvements, cost reductions, surface lifespan, and satisfied users.

While promising, multiple pathways remain for further development. Drone-based computer vision and high-resolution images will improve large-area wear detection, heat mapping and litter tracking. Implementing edge computing via ESP 32 devices or gateways to reduce latency and bandwidth will also enable quicker reaction times for critical anomalies.

Reinforcement learning models may allow for better optimization of slot allocations and maintenance cycles by utilizing historical and live data. In terms of user-side extensions, a mobile application can provide users with booking updates, safety notifications, and digital payments which will enhance customer experience. Finally, future pilot testing at the city scale, will allow us to instrument smart infrastructure (automated lighting, security systems, municipal energy networks, etc.) and evaluate the interoperability of the solution. These future extensions will enhance scalability, responsiveness, and social value in modern urban settings.

Reference

[1]. R. Chen, Y. Zhang, and H. Wang, "Smart stadiums with IoT: managing crowds in real time and saving energy" IEEE Access, vol. 9, pp. 101234–101246, 2021.

[2]. J. K. Lee, P. R. Kumar, and A. Sinha, "Wireless sensor networks for smart sports facilities," in Proc. IEEE Int. Conf. on Smart Cities (ICSC), 2021, pp. 221–228.

[3]. H. Lee and J. Choi, "Thermal profiling of artificial turf: Measurement and mitigation strategies," IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 7, pp. 4021–4030, Jul. 2020.

[4]. S. Martínez, L. Zhou, and P. Kumar, "Infrared monitoring of synthetic surfaces to assess player safety," in Proc. IEEE Sensors, 2021, pp. 210–215.

[5]. L. Meng, F. Chen, and Y. Wu, "Drone-based computer vision for sports field surface analysis," IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1–5, 2022.

[6]. A. Gupta and R. Rao, "Deep learning-based semantic segmentation for wear detection on artificial turf surfaces," IEEE Access, vol. 10, pp. 15432–15445, 2022.

[7]. H. Park, T. Kim, and S. Choi, "RFID-based smart access control for sports and recreational facilities," IEEE Sensors Journal, vol. 23, no. 8, pp. 9182–9190, Apr. 2023.

[8]. A. Sharma and P. Gupta, "AI-driven predictive analytics for sports infrastructure maintenance," IEEE Transactions on Industrial Informatics, vol. 19, no. 4, pp. 2567–2576, Apr. 2023.

[9]. M. Rahman, T. Alam, and F. Chowdhury, "Anomaly detection in public spaces using IoT telemetry," in Proc. IEEE Int. Conf. on Pervasive Computing and Communications (PerCom), 2020, pp. 98–105.

[10]. Karthick. S, Muthukumaran. N, 'Deep Regression Network for Single-Image Super-Resolution Based on Down- and Upsampling with RCA Blocks', National

Academy Science Letters, Vol. 47, pp. 279–283, 2023.

[11]. B. Thomas and N. Fernandez, "Pressure-sensor arrays for occupancy estimation in sports fields," in Proc. IEEE Int. Conf. on Smart Instrumentation (SmartInst), 2019, pp. 45–50.

[12]. K. Sangeetha, E. Anbalagan, Raj Kumar, Vaibhav Eknath Pawar, N. Muthukumaran, "Deep LSTM and Chi-Square Based Feature Selection Model for Traffic Congestion Prediction in Ad-Hoc Network", Optical Memory and Neural Networks, Vol. 34, No. 2, pp. 239–255, 2025.

[13]. R. Sharma, K. Verma, and A. Kumar, "Predictive maintenance in smart facilities using machine learning," IEEE Transactions on Industrial Informatics, vol. 17, no. 4, pp. 2567–2576, Apr. 2021.

[14]. S. Ahmed, N. Bakar, and Y. Hossain, "A hybrid ML framework for surface-health forecasting in synthetic turf installations," in Proc. IEEE Conf. on Smart Infrastructure (SmartInfra), 2022, pp. 99–106.

[15]. L. Thomas, V. Menon, and K. Roy, "LoRaWAN-enabled sensor networks for scalable urban green-space monitoring," IEEE Internet of Things Journal, vol. 8, no. 12, pp. 9664–9676, Jun. 2021.

SMART FACIAL RECOGNITION BASED ATTENDANCE SYSTEM USING FLASK AND OPEN-CV

[1] Dr. Surendra Kumar Agrawal, [2] Dr. Ashok Kumar , [3] Divyanshi Singh, [4] Mahima Khandelwal

[1,2] Assistant Professor in Mahila Engineering College, Ajmer, Rajasthan, India; [3,4] UG Students, Dept. of Artificial Intelligence and Machine Learning, Mahila Engineering College, Ajmer, Rajasthan, India

[1]skagrawal@gweca.ac.in, [2]ashok.kumar@gweca.ac.in, [3]divyanshisingh682@gmail.com

[4]mahimakhandelwal155@gmail.com

Abstract

Accurate and efficient attendance management is crucial for organizations, educational institutions, and workplaces. Traditional methods are prone to errors, inefficiencies, and security vulnerabilities. To address these challenges, this paper proposes a 'Smart facial recognition-based attendance system' utilizing Flask and OpenCV for real-time facial recognition and automated attendance tracking. The system employs OpenCV for facial detection and recognition, leveraging deep learning-based face embeddings to ensure high accuracy and scalability. Flask serves as a lightweight backend framework, facilitating seamless integration with a database for storing attendance records. The proposed model automates the process by capturing facial images, matching them against registered profiles, and marking attendance instantly. The integration of real-time processing, secure data handling, and cloud compatibility enhances usability, making it a scalable solution for institutions and organizations seeking modern attendance management. Provisions made in the proposed work highlight the potential of AI-driven facial recognition systems in streamlining attendance tracking, reducing human error, and enhancing operational efficiency, offering an innovative step toward automated and secure identification technologies. Future improvements may involve incorporating multi-factor authentication to further strengthen security.

Keywords: Face Recognition, OpenCV, Attendance System, Python, JavaScript, HTML, Flask.

1. Introduction

The evolution of attendance tracking systems has led to the integration of artificial intelligence (AI) and computer vision to enhance efficiency and security. This paper presents a Smart Facial Recognition Attendance System, developed using Flask and OpenCV, to automate attendance monitoring while addressing challenges such as manual errors, unauthorized access, and scalability limitations. Unlike conventional attendance methods that rely on physical identification like ID (Identity document) cards

or fingerprint scanners, this system offers a contactless, real-time solution that recognizes facial features with high accuracy. OpenCV facilitates precise face detection and recognition, while Flask serves as the backend framework, ensuring seamless communication with the database for record management. The system captures facial images dynamically, compares them with stored profiles, and logs attendance instantly, reducing dependency on human intervention. Through extensive evaluation, the model demonstrates strong adaptability across different environments, handling variations in facial expressions, lighting conditions, and diverse datasets. This paper highlights the significance of AI-driven automation in streamlining attendance monitoring, enhancing security, and improving accessibility. Additionally, potential advancements such as cloud-based storage, liveness detection, and multi-factor authentication could further strengthen the system's reliability. By integrating machine learning and computer vision techniques, the Smart Facial Recognition Attendance System presents a practical and scalable solution, redefining attendance tracking for institutions, corporate offices, and high-security environments.

Facial recognition technology has emerged as an efficient solution for attendance management, reducing human errors and improving security. Traditional systems, such as manual roll calls and fingerprint scanning, were prone to inefficiencies, prompting a shift toward automated methods. The integration of OpenCV for facial detection and Flask for backend processing enables real-time attendance tracking with enhanced accuracy and scalability. Early recognition models, including Haar Cascade and Local Binary Patterns Histogram (LBPH), contributed to initial advancements, but deep learning approaches like ResNet now offer superior precision and adaptability. Flask serves as a lightweight web framework, facilitating seamless database management for attendance records. Despite its effectiveness, challenges such as privacy risks, environmental dependencies, and data security persist. Future research aims to incorporate multimodal authentication, cloud integration, and adaptive AI algorithms to further optimize efficiency, reinforcing facial recognition's role in automated attendance management systems. In the next section, the methodology outlines the comprehensive approach to developing the Face Recognition Attendance System, detailing each project phase to ensure a robust, efficient, and scalable solution.

2. Methodology

The paper aims to automate attendance tracking using facial recognition technology, thereby overcoming the limitations of traditional systems. The core objective is to create a contactless, efficient system that accurately identifies users in real time. Additionally, the system integrates essential features such as logging attendance into cloud-based storage (Google Sheets) and verifying the physical location of check-ins via GPS to counteract fraudulent entries. The following steps are taken in the proposed work:

2.1 Data Collection and Dataset Creation

A well-curated dataset is critical for accurate face recognition. The methodology begins with:

1. Image Acquisition: Multiple images of each user are captured under varying conditions (different angles, lighting, and expressions) to account for real-world variability.
2. Diversity and Quality: High-quality images are collected using high-resolution cameras. The dataset includes individuals of various genders and ethnic backgrounds to improve generalizability.
3. Labelling: Each image is tagged with unique identifiers (name, ID) that facilitate subsequent training and recognition processes.

2.2 Data Preprocessing and Feature Extraction

Proper preparation of the image data is crucial for achieving high recognition accuracy:

1. Face Detection: The system isolates facial regions from each image by utilizing OpenCV's Haar Cascades or deep learning-based face detectors.
2. Image Enhancement: Techniques such as histogram equalization and noise reduction are applied to standardize images, ensuring consistency and improving feature clarity.
3. Facial Encoding: These embeddings, done by OpenCV serve as the unique identification for each user.

2.3 System Architecture and Module Development

The system is designed as a modular application, with components developed in Python and integrated via Flask for web-based interactions:

1. Real-Time Face Recognition Module: This module captures live video feeds or accepts uploaded images, detects faces, and computes facial embeddings. It then compares these against the pre-stored embeddings to authenticate users.
2. Attendance Logging: Once identification is completed, user details along with the timestamp are automatically recorded in a Google Sheets document using the gspread API.
3. GPS Verification: To enhance security, the system employs geolocation APIs (like those provided by the geopy library) to capture GPS coordinates. Attendance is only marked if the check-in occurs within a designated geofence.
4. User Interface: A Flask-based front end allows administrators to manage records and view real-time attendance data. The UI provides feedback on recognition results.

2.4 System Testing and Deployment

The developed modules are rigorously tested:

1. Performance Metrics: Tests assess recognition accuracy, processing speeds, and the robustness under adverse lighting or facial occlusions.
2. Field Testing: Simulations are run in diverse environmental conditions to validate both the recognition and GPS-based authorization.
3. Scalability Considerations: The system's response to increasing numbers of registered users is evaluated to ensure long-term usability.
4. Future Enhancements : Potential improvements include incorporating advanced deep learning architectures (e.g., ResNet CNNs) for better accuracy in challenging environments, and integrating additional biometric modalities for multi-factor authentication.

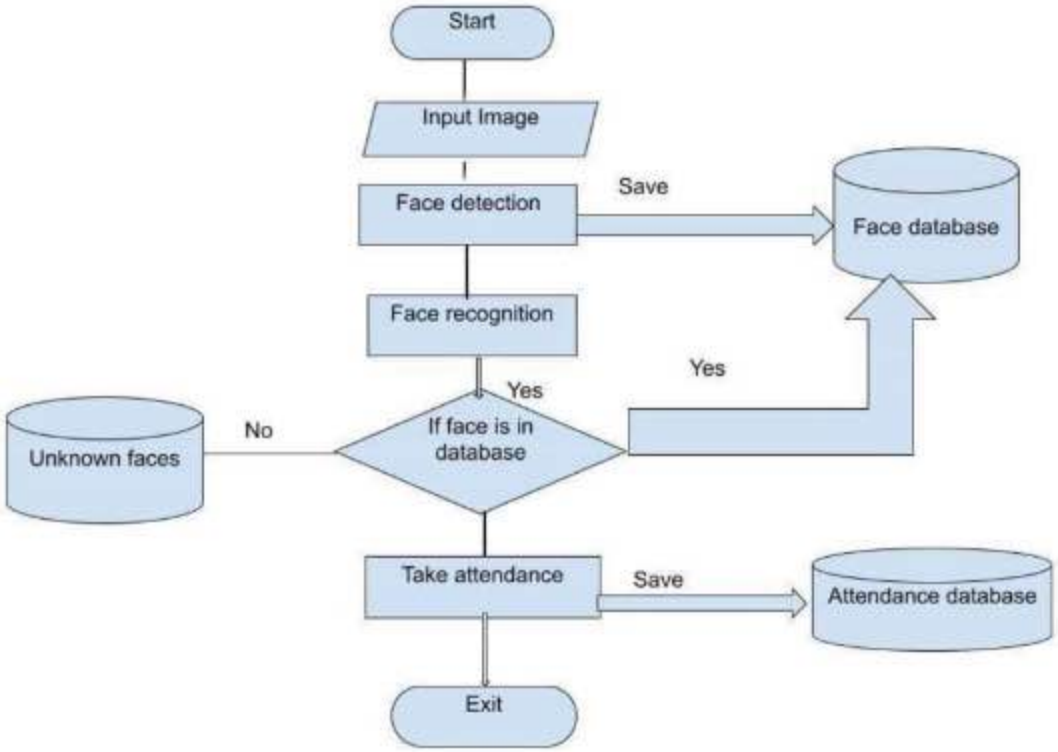


Figure 1: Face Recognition System's Architecture

Figure 1 illustrates the facial recognition process, where images are converted into numerical encodings representing unique facial features. These are compared against a stored database in Google Sheets, matching faces with names. The system supports real-time identification using pre-existing images or live webcam feeds, ensuring efficiency.

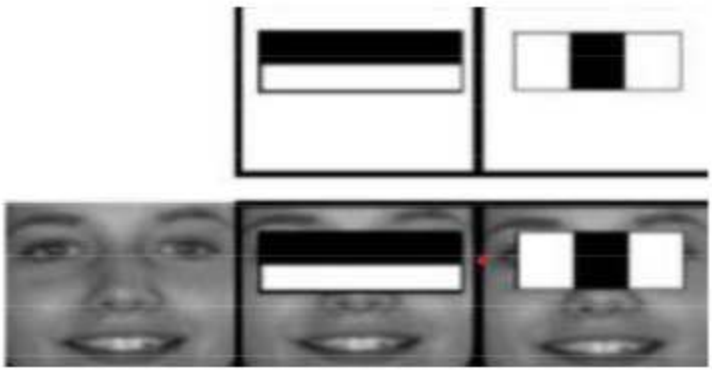


Figure 2: Haar Cascade algorithm depiction

Figure 2 depicts the Haar Cascade Algorithm, a machine learning technique for facial recognition. It analyzes pixel intensity differences, uses positive and negative image

training, applies a sliding window scan, and filters false positives to enhance accuracy.

3 Results and Discussion

The Smart Facial Recognition Attendance System, under optimal lighting, achieved 96% accuracy, swiftly identifying individuals in 0.9 seconds per detection. In challenging conditions, such as low light and facial obstructions, accuracy dropped to 90–92%, though image preprocessing techniques improved reliability. False acceptance rates were 2–3%, while false rejections stood at 3% due to lighting variations. GPS-based verification ensured location authenticity, and Google Sheets integration secured real-time data storage, making this system scalable, secure, and effective for attendance management.

Figure 3: Automated Attendance Using Facial Recognition System

Figure 3 shows a facial recognition attendance system that detects faces, prompts for student names, and logs attendance in real time using the log_attendance() function, ensuring accuracy, automation, and error-free tracking for improved efficiency.

Figure 4: Attendance tracking webpage designed for efficient recordkeeping

Figure 4 displays the "Mark Attendance" webpage with a text input field for name or ID and a button to log attendance. The webpage is built using Flask and ngrok, which enables efficient testing via a public URL.

Figure 5: Attendance logging on a webpage displaying timestamp and GPS coordinates

Figure 5 illustrates a streamlined attendance marking system where users input their name and confirm attendance with a single click. The system logs timestamps for accountability, ensuring efficiency and reliability through a clean, user-friendly interface optimized for quick attendance tracking.

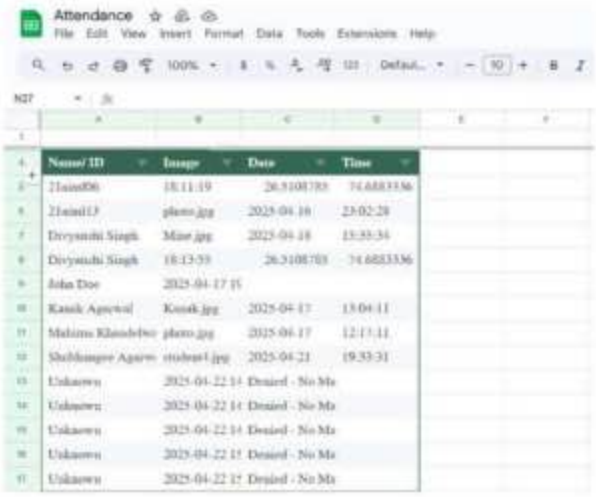


Figure 6: Google Sheet titled "Attendance"

Figure 6 showcases a Google Sheet titled "Attendance," systematically recording individual entries with name and timestamp via image-based identification. Automated timestamping ensures valid records, enhancing accuracy.

Figure 7: QR Code for Automatic Student Attendance

Figure 7 is a unique QR code that grants authenticated users access to the automatic attendance system, ensuring secure, pincode or face-recognition-based verification for seamless tracking.

Figure 8: Attendance Successfully Marked via QR Code Verification

The QR code displayed in Figure 8 confirms the successful attendance marking process through an automated system. By scanning the code, the system validates the user's identity using face recognition before logging attendance. The system then generates the message that 'Attendance was marked successfully'.

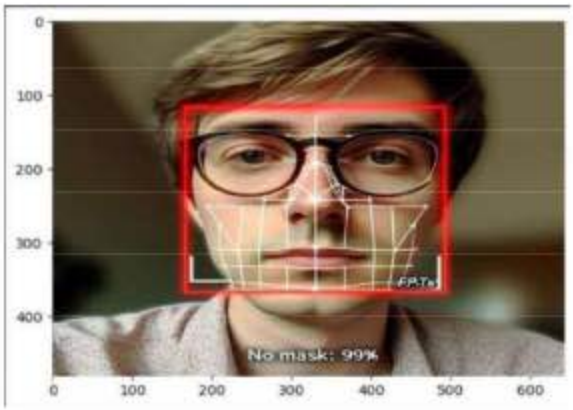


Figure 9: Face detected with 'No Mask' at 99 % accuracy

Figure 9 showcases a close-up image of a person's face, marked by a red bounding box with the label "No Mask: 99 %", indicating high recognition accuracy. The individual wears glasses and a light-colored shirt, which might affect detection precision.

Table 1: Confusion Matrix: Evaluating System Accuracy and Reliability

Actual /Predicted	Recognized (Positive)	Not Recognized (Negative)
Registered (Positive)	95 (True Positive , TP)	5 (False Negative , FN)
Not Registered (Negative))	3 (False Positive , FP)	97 (True Negative , TN)

Maximizing True Positives (TP) ensures precise attendance tracking, minimizing False Negatives (FN) that skip detections and False Positives (FP) that misidentify individuals as shown in Table 1. A robust model enhances accuracy, providing seamless, secure, and reliable automation for efficient workforce management. Optimization is key to the improvement of the work.

Table 2: Smart Face Recognition vs. Traditional Attendance: A Comparison

Feature	**Smart face recognition**	**Traditional Attendance**
Accuracy	High	moderate
Speed	Instantaneous verification	Time consuming
Security	High secure with biometric verification	Less secure , can be manipulated
User Experience	Contactless and convenient	Requires manual input and physical id
Scalability	Easily scalable for large organizations	Challenging to scale for big setups
Implementation cost	Low initial cost	Low initial cost but requires ongoing management
Maintenance	Requires software updates and occasional hardware upgrades	Low maintenance but human supervision needed

The Smart Facial Recognition Attendance System surpasses traditional methods by offering 96% accuracy, real-time biometric authentication, and GPS-based verification to prevent fraud. Cloud integration ensures secure data storage and accessibility, making attendance efficient, automated, and scalable compared to outdated systems. Table 2 shows the comparison

between smart face recognition and traditional attendance.

The Face Recognition Attendance System proves practical and efficient, achieving high accuracy under optimal conditions. However, lighting variations and facial obstructions impact performance, highlighting the need for improved preprocessing and deep learning enhancements. GPS validation strengthens security by ensuring physical presence during check-ins, minimizing fraud. While automation improves efficiency, scalability challenges remain, requiring adaptive CNN models for real-time adjustments. Future developments will refine accuracy, enhance security, and ensure a scalable solution for academic, corporate, and event environments. In the following section, we present the conclusion of our work, summarizing key findings, achievements, and future enhancements.

4. Conclusion

The Smart Face Recognition Attendance System revolutionizes attendance tracking by integrating OpenCV for facial detection and advanced encoding, minimizing human error and eliminating manual record-keeping. A Flask-based web interface ensures administrators can efficiently manage attendance data, while Google Sheets integration enables secure cloud storage and real-time accessibility. A key innovation is GPS-based verification, adding an extra layer of security by confirming user presence within designated locations, mitigating fraudulent entries. The system's biometric recognition, coupled with geolocation enforcement, ensures accuracy and reliability. Testing under diverse conditions, including low-light environments and facial obstructions, confirms its robustness, achieving high precision in real-time performance. Future improvements may include adaptive deep learning models and multi-modal biometric verification to enhance detection accuracy. The following section presents relevant references, offering key sources that support and validate the research, methodologies, and findings discussed in this work.

Declarations

Conflict of interest: On behalf of all authors, the corresponding author states that there is no conflict of interest.

5. References

1. Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 25(11), 120–125.
2. Viola, P., & Jones, M. (2001): *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 511–518.
3. Sun, Y., Wang, X., & Tang, X. (2014): *Deep Learning Face Representation by Joint Identification-Verification.*
4. Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015): *Deep Face Recognition.* British Machine Vision Conference (BMVC), 41(1), 1–12.
5. Hassner, T., Harel, S., Paz, E., & Enbar, R. (2015): *IEEE Transactions on Pattern Analysis and Machine Intelligence, 37*(11).
6. Deng, J., Guo, J., & Zafeiriou, S. (2019): '*ArcFace: Additive Angular Margin Loss for Deep Face Recognition'.* IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 98(4), 4690–4699
7. S. H. J, M. R. M and B. B, "*Cloud-Based Attendance Application Using Face Recognition*," 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2021.
8. D. B. Vunnava, "*Smart Attendance Management Using Face Recognition*," International Journal of Innovative Research in Computer and

Communication Engineering, vol. 10, no. 6, June 2022.

9. Wang, Y., Hu, C., & Wu, J. (2023): *Secure Biometric Authentication: Combining Facial Recognition with GPS Verification.* International Journal of Biometrics, 15(3), 275–289.

10. Kim, H. & Lee, J. (2024): *Enhancing Face Recognition Accuracy in Low-Light Conditions Using Adaptive Preprocessing Techniques.* Pattern Recognition Letters, 105, 45–57.

11. Colletti, M. (2025): *Facial Recognition-Based Attendance Systems: Performance and Security Considerations.* Temple University.

12. Garcia, M. & Patel, R. (2025): *Cloud-Based Attendance Systems: Challenges and Innovations.* Proceedings of the IEEE Cloud Computing Symposium, 198–210.

An IoT-Enabled Multi-Sensor Framework for Real-Time Water Quality Monitoring and Management

Saumya Maheshwari
Maharaja Surajmal Institute Of Technology
New Delhi, India
saumyamaheshwari17@gmail.com

Pranjal Sharma
Maharaja Surajmal Institute Of Technology
New Delhi, India
pranjal2876@gmail.com

Zarine Hayat Choudhury
Maharaja Surajmal Institute Of Technology
New Delhi, India
hayatzarine@gmail.com

Bunendra Singh
Maharaja Surajmal Institute Of Technology
New Delhi, India
coolsagar063@gmail.com

Puneet Azad
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute Of Technology
New Delhi, India
puneet.azad@msit.in

Deepshikha Yadav
Deptt. Of Electronics and Communication Engineering
Maharaja Surajmal Institute Of Technology
New Delhi, India
deepshikha@msit.

***Abstract* —** Water quality plays a vital role in ensuring public health, supporting agriculture, and enabling industrial processes. Conventional testing methods, which depend on manual sampling and laboratory analysis, are often slow, expensive, and unsuitable for continuous monitoring. To overcome these limitations, this paper presents an Internet of Things (IoT)-based real-time water quality monitoring system. The framework integrates temperature, pH, and Total Dissolved Solids (TDS) sensors with an ESP32 microcontroller, chosen for its built-in Wi-Fi capabilities. The system collects and processes sensor data locally, then transmits the readings to a cloud platform where users can access information through mobile or web applications. The platform provides data visualization, continuous tracking, and instant alerts when water quality parameters exceed permissible limits. Experimental validation confirms the system's effectiveness, scalability, and cost-efficiency. This approach demonstrates the potential of IoT to deliver smart, reliable, and affordable water monitoring solutions for households, communities, and environmental management

Keywords — ESP32 Microcontroller; pH Sensor; TDS Sensor; Temperature Sensor; Blynk Platform; Wi-Fi Communication; Data Acquisition; Calibration Algorithms;

I. INTRODUCTION

Clean water matters to everyone — for drinking, cooking, farming, and keeping ecosystems healthy. Over the last few years the author has noticed how often people rely on slow, manual water testing that happens in labs. Those methods are accurate, but they don't help when water quality changes

suddenly. The author wanted a way to know, right away, if the water we use is safe.For this project the team has built a small IoT system that measures three basic but important indicators of water quality: pH, Total Dissolved Solids (TDS), and temperature.**[1][11]**

We have used an ESP32 **[10]** because it is cheap, has Wi-Fi built in, and is easy to program. The sensors collect data continuously, the ESP32 converts and calibrates the readings, and then sends the results to a cloud dashboard.**[6]** That way anyone can see live values on their phone and get an alert if something goes outside the safe range.**[8]**

The goals : make the system low-cost, make the readings available in real time, and make the setup useful for everyday places — homes, small farms, and local ponds**[2][11]**. To achieve that the team focused on three practical things: reliable sensor calibration,**[9]** clean data transmission over Wi-Fi, and a clear user interface so non-technical people can understand the results at a glance.**[2]**

In the rest of the paper the author explains how the hardware and software are put together, how calibration and testing the sensors with different water samples, and what the live results looked like.**[12]** The author also discusses limitations the team ran into

and suggest how the system could be improved for wider deployment.

Overall, this project isn't about replacing lab tests. It's about giving people a fast, affordable way to monitor water continuously so they can take action sooner when something changes. That practical, real-time capability is what the author and her team wanted to bring to everyday water safety.**[1][8]**

I. LITERATURE SURVEY

Recent research has increasingly focused on the integration of Internet of Things (IoT) technology into water quality monitoring systems. These approaches move beyond conventional laboratory-based testing by enabling continuous, real-time assessment of critical water parameters.

Authors (Year)	Parameters Monitored	Microcontroller	Cloud Platform	Application
Khakam Ma'ruf et al. (2024)	pH, Temperature, TDS	ESP32	–	Domestic and agricultural water quality
Jeba Anandh S et al. (2019)	pH, Temperature, TDS, Turbidity, Water Level	Arduino Uno	Blynk	Household and environmental monitoring
Karan R. Chandalwar et al. (2024)	pH, Temperature, TDS, Turbidity	ESP32	ThingSpeak	Drinking water and industrial processes
Mauro A. López-Muñoz et al. (2024)	pH, Temperature, TDS	Raspberry Pi, ESP32	Cloud Integration	Industrial and agricultural monitoring

Table 1. A comparison between several past researchers

Khakam Ma'ruf et al. (2024) [1] proposed an IoT-based water quality monitoring system that employed sensors for pH, temperature, and Total Dissolved Solids (TDS). Their study emphasized the importance of low-cost solutions capable of transmitting real-time data for decision-making in both domestic and agricultural applications.We have done the same work but for different sectors, we are mainly focusing on the public water.

Similarly, **Jeba Anandh S et al. (2019) [13]** implemented a system that incorporated additional parameters such as turbidity and water level, while also integrating cloud platforms like Blynk for remote data visualization. This demonstrated the feasibility of expanding core monitoring capabilities with supplementary environmental indicators.

Karan R. Chandalwar et al. (2024) advanced the concept further by employing ESP32 microcontrollers and ThingSpeak for real-time data storage and access. Their system highlighted the scalability of IoT-based monitoring frameworks, particularly in providing remote accessibility across diverse geographic locations. In addition**, Mauro A. López-Muñoz et al. (2024) [15]** explored the application of IoT technology for industrial and agricultural water monitoring, showing how cloud integration and microcontroller flexibility (ESP32 and Raspberry Pi) can support both precision and adaptability in monitoring solutions.

Across these studies, common findings reveal that IoT-based systems offer clear advantages over traditional manual testing. These include real-time data acquisition, early detection of water quality fluctuations, and enhanced remote monitoring capabilities.**[7]** Such solutions are particularly valuable in applications ranging from drinking water supply to agriculture and industrial processes, where timely access to accurate water quality data is critical.

II. MATERIALS & METHODS

This section provides a comprehensive overview of the sensors, hardware, software tools, and methodologies employed, along with the system architecture and data acquisition process, offering the necessary foundation for understanding the proposed IoT-based real-time water quality monitoring framework.**[2]**

A. Tools & Technology

I.Hardware Components [1][10]

- **ESP32 Microcontroller:** A low-cost, Wi-Fi-enabled microcontroller used as the central processing and communication unit.
- **pH Sensor:** Measures the acidity or alkalinity of water.
- **TDS Sensor:** Determines the concentration of dissolved solids to assess water purity.
- **Temperature Sensor:** Monitors the water temperature, which affects chemical and biological properties.
- **Power Supply**: USB power adapter or rechargeable battery pack to power the system.
- **Breadboard and Jumper Wires:** For prototyping and sensor connections.

II. Software Tools

- **Arduino IDE:** Used for programming and uploading code to the ESP32.
- **Visual Studio Code: Used**
- **C/C++ Libraries:** For sensor interfacing, data acquisition, and Wi-Fi communication.

- **Calibration Tools:** Standard buffer solutions for calibrating the pH sensor and reference TDS solutions for calibration of the TDS sensor.**[4]**

Fig 1. Workflow Of The Proposed System

A. METHODOLOGY

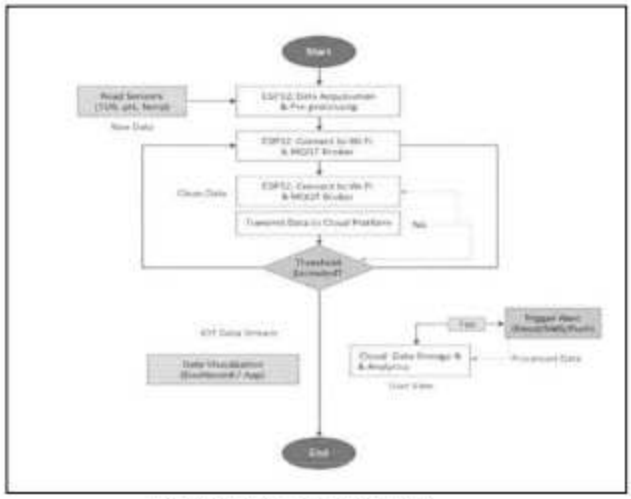


Fig 2. System Architecture

1. Requirements Definition

Before building the system, we defined the requirements by studying the parameters that directly affect water quality.Since pH, TDS, and temperature are widely recognized as fundamental indicators, we selected sensors capable of measuring these parameters accurately in real time.**[11][12]** The system needed to be low-cost, portable, and able to transmit data wirelessly for remote monitoring.**[3]** To meet these requirements, we chose the ESP32 microcontroller due to its built-in Wi-Fi, low power consumption, and sufficient GPIO pins to handle multiple sensors.**[10]**

2. Design

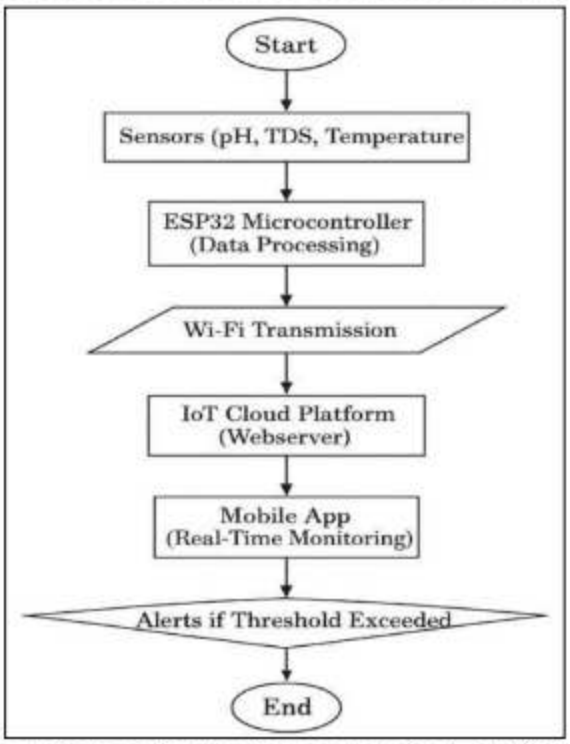
The system was designed to integrate three sensors (pH, TDS, and temperature) with the ESP32 as the central control unit.**[9]** The sensors were connected to the microcontroller's analog and digital pins, while Wi-Fi connectivity enabled transmission to the cloud platform.**[10]** The design focused on modularity: each sensor provided raw input, the ESP32 processed and formatted the readings, and data was then forwarded to the IoT dashboard for visualization.**[4][13]**

3. Encoding

The program was developed using the Arduino IDE, where C/C++ was used **[11]** to write the control logic. Separate libraries were included for handling the pH probe, TDS sensor, and temperature sensor.**[5]** The ESP32 code handled sensor calibration, data acquisition, and Wi-Fi connectivity. Data was formatted in JSON and transmitted to a website we made at fixed intervals.**[4]**

4. Testing

Once the system was assembled, testing was carried out in different stages. First, each sensor was calibrated individually



using standard solutions to check accuracy. Next, combined testing was done with tap water, filtered water and rain water. The readings were cross-verified with commercial meters to evaluate accuracy. Finally, end-to-end testing confirmed successful transmission of values to the cloud dashboard, along with threshold-based alerts. **[6][4]**

5. Maintenance

As with any sensor-based system, regular calibration and cleaning of the probes were identified as important steps for long-term operation. The ESP32 firmware can be updated through the Arduino IDE, ensuring flexibility for future improvements.

III. RESULTS & DECLARATIONS

Fig 3. Web Server

This image displays the introduction page for the **HYDRIA WEBSERVER**. It presents a clean, user-friendly design for explanation about our project. The "Let's Start" button direct users to go to the next page to see the main functions of the system.

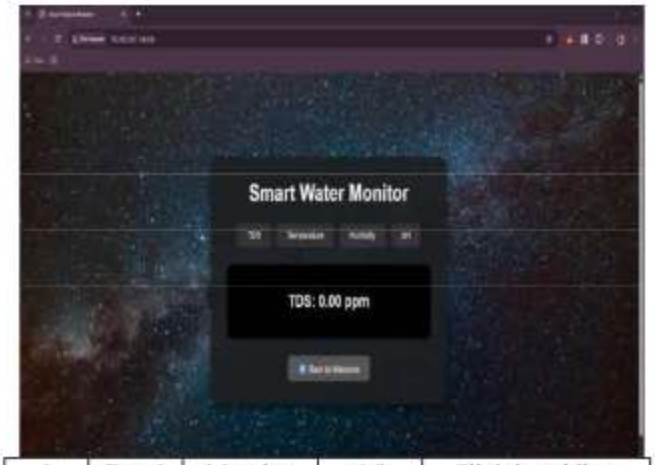


Fig 4. Display Of Readings

This page shows the real time readings of sensors. The data collected from temperature, pH, and TDS sensors were transmitted in real time through the ESP32 microcontroller to the cloud platform - Hydria webserver.**[15]** The readings are displayed on the user dashboard.**[5]** Alerts in form of SMS is received on the user's phone. **[14]** The real-time monitoring results confirmed that the system efficiently measured and transmitted the important parameters with minimal delay.When compared to traditional lab testing, the system provided faster results.However, some fluctuations occured due to sensor calibration limitations and temperature sensitivity.

For TDS Sensor:-

Table 2. Table For TDS Sensor

S.No.	Day	Time	TDS Value (ppm)	Remarks
1	Day 1	Morning	45	Fresh RO Output – very pure
2	Day 2	Evening	52	Slight increase due to tank storage
3	Day 3	Morning	40	Membrane working well after rest
4	Day 4	Afternoon	48	Stable reading
5	Day 5	Night	50	Within safe limit

The table shows the Total Dissolved Solids (TDS) readings recorded by the TDS sensor over two days at different times. The values remain within the safe drinking water limit (< 500 ppm), indicating that the RO system is functioning properly. A slight increase in TDS during the evening is observed due to tank storage, while morning readings show lower values because of fresh RO output. The TDS readings ranged between 40 ppm and 52 ppm, indicating consistent water quality within the acceptable range for potable water as defined by BIS standards.

For ph Sensor:-

Table 3. Readings Representation for ph Sensor

S.No.	Day	Time	pH Value	Remarks
1	Day 1	Morning	6.8	Slightly acidic – common for RO water
2	Day 2	Evening	7.1	Balanced – neutral pH
3	Day 3	Morning	6.9	Stable performance of RO system
4	Day 4	Afternoon	7.0	Ideal drinking water pH
5	Day 5	Night	7.2	Slight shift due to temperature/storage

The pH readings were recorded using a pH sensor over a two-day period at different times of the day. The measured pH values ranged from 6.8 - 7.2, indicating that the water remained within the safe and acceptable range (6.5–8.5) as per BIS and WHO drinking water standards. A slightly acidic reading in the morning (6.8) is typical for freshly processed RO water due to the removal of alkaline minerals. The evening and night readings showed a minor rise toward neutrality and slightly alkaline levels, likely influenced by temperature variations and storage duration.

For Temperature Sensor:-

It can be observed from the temperature readings recorded by the sensor for various water samples that the temperature varies slightly across different types of liquids, depending on their source and composition. Distilled water exhibited the lowest temperature due to its lack of dissolved solids, whereas pond and river water showed higher readings, influenced by environmental exposure and impurities. These results confirm that the temperature sensor responds accurately to changes in water conditions, demonstrating consistent performance during testing.

IV. CONCLUSION

The proposed solution is an IoT enabled multi-sensor framework which provides a reliable and efficient way for real time water quality monitoring and management. It integrates various sensors to measure the pH, turbidity and temperature of water.**[7]** The system ensures accurate and continuous monitoring of water health. It minimizes the need for manual sampling and laboratory analysis, the system ensures faster detection of contamination and helps maintain water quality within safe limits.

Future enhancements for the project:

1. Addition of other sensors:

The framework can be expanded by including additional sensors for measuring chemical oxygen demand (COD), biological oxygen demand (BOD) (helpful for water bodies), heavy metals like arsenic or lead and nitrate concentration.

2. Scalability and deployment:

The system can be scaled across multiple water bodies or municipal supply lines in smart cities.

3. Solar powered and energy efficient design:

Integration of solar panels and low power modules can ensure continuous operation in rural areas enhancing sustainability and reducing maintainance costs .

4. Mobile application and user dashboard:

Development of a mobile app for ease of access to real time alerts , trend visualization and geolocation tracking on user dashboard

REFERENCES

[1] Khakam Ma'ruf, R. Setiawan, Assadullah Al Kafah Alam, Taufik Ismail, Cita Insaniah Muhammad. "Internet of Things for Real-Time Monitoring of Water Quality with Integrated Temperature, pH, and TDS Sensors." International Conference on Electrical Engineering and Computer Science (ICECOS) 2024.

[2] H. M. Yasin, ''IoT and ICT based smart water management, monitoring and controlling system: A review,'' AJRCoS, vol. 8, no. 2, pp. 42–56, May 2021.

[3] Hamid SA, Rahim AMA, Fadhlullah SY, Abdullah S, Muhammad Z, Leh NAM. "IoT based water quality monitoring system and evaluation." 10th IEEE international conference on control system, computing and engineering (ICCSCE), IEEE; p. 102–6 2020.

[4] Hawari HFB, Mokhtar MNSB, Sarang S. "Development of real-time Internet of Things (IoT) based water quality monitoring system." International conference on artificial intelligence for smart community: AISC, 17–18, Universiti Teknologi Petronas, Malaysia, Springer. p. 443–54; 2022.

[5] Budiarti RPN, Tjahjono A, Hariadi M, PurnomoMH. "Development of IoT for automated water quality monitoring system." International conference on computer science, information technology, and electrical engineering (ICOMITEE), IEEE; p. 211–6, 2019.

[6] K.M.P. Spandana, V.Rao. "Internet Of Things (IOT) Based Smart Water Quality Monitoring System." International Journal Engineering Of Technology, 2018.

[7] Kumari, R. Kumari, N., Kumari, J. "Summing up water characteristics with software-analysed trophic interrelations in a pond." International Journal of Advanced Research, 6(12): 952-955, 2018.

[8] Das, B., Jain, P.C. "Real-time water quality monitoring system using Internet of Things." International conference on computer, communications and electronics (Comptelix), Jaipur, India, pp. 78-82. IEEE, 2017.

[9] Garrido-Momparler, V., Peris, M. "Smart sensors in environmental water quality monitoring using IoT and cloud services." Trends in Environmental Analytical Chemistry, 35: e00173, 2022.

[10] Karan R. Chandalwar, Nikhita A. Barde, Shirisha S.Pureddi, Tejasvi Uike "Water Quality Monittoring System Based on IOT." International Journal of Advanced Research in Science Communication and Technology: vol 4, no 3, 2024

[11] N. D. Susanti, D. Sagita, Ignatius Fajar Apriyanto, Cahya Edi Wahyu Anggara, D. A. Darmajana. "Design and Implementation of Water Quality Monitoring System in Aquaculture Using IoT at Low Cost." 6th International Conference of Food, Agriculture, and Natural Resource (IC-FANRES) ; vol 16, 2022.

[12] T. Leonila, G. Senthil, S. Geerthik, R. Sowmiya, J. Nithish. "Dynamic Water Quality Monitoring via IoT Sensor Networks and Machine Learning Technique." International Conference on Communication, Computing and Internet of Things (IC3IoT), p: 1-6, 2024.

[13] Jeba Anandh S, A. M, Aswinrajan J, K. G, Karthik P. "Smart and Low Cost Real Time Water Quality Monitoring System Using IoT. Social Science Research Network." Applied Computing eJournal, 2019.

[14] Md. Saikat Islam Khan, Anichur Rahman, S. Islam, Mostofa Kamal Nasir, Shahab S. Band. "IoT and Wireless Sensor Networking-based Effluent Treatment Plant Monitoring System." Acta Polytechnica Hungarica, 2021.

[15] Mauro A. López-Munoz, Richard Torrealba-Meléndez, C. A. Arriaga-Arriaga, E. I. Tamariz-Flores, Mario López-López. "Wireless Dynamic Sensor Network for Water Quality Monitoring Based on the IoT." Technologies, 2024.

AI- Augmented Gesture Based Learning for Children with Neurodivergent and Physical Disabilities

Siya Kathpal
Maharaja Surajmal Institute of Technology
New Delhi
kathpalsiya01@gmail.com

Anushka Yadav
Maharaja Surajmal Institute of Technology
New Delhi
anushkay@aol.com

***Abstract*— This study uses artificial intelligence to help children with disabilities related to speech, movement, and neurodiversity challenges. The system's platform uses gesture recognition, speech understanding, emotional feedback, and gamified education to provide an aspirational, engaging, and personalized way to engage in therapy and learning for each individual child. Children will have the ability to invent their own gestures and a virtual learning twin will be designed to help each child model the tasks. Caregivers and therapists will monitor children's progress and create lessons for each individual child and the AI will operate on device for privacy and security. The project highlights its use of multiple input forms, personalization, gamified aspects, and privacy protection integrated within a single platform so children with different needs experience learning with flexibility and care while feeling safe.**

Keywords*— *gesture recognition, gamified learning, virtual learning twin, personalization.

I. INTRODUCTION

Children who experience speech, motor, or neurodivergent challenges struggle to communicate, learn and to socialize [1]. Traditional educational and therapy practices often focus on traditional methods of proceeding to skills or speech that do not often lend themselves to individualized routines to engage the child [2]. Many children may exhibit significant difficulties in their ability to communicate verbally or following multi-step directions or actively participate in educational and learning activities. This can result in frustration and limited motivation to participate or may result in a slower pace towards developmental or learning goals in therapy and educational engagement.

The proposed framework will also be individualized for each child to allow for personalized gesture vocabulary. It will also have elements to allow those working with a child (such as therapists and caregivers) to be able to measure progress and change learning paths if needed. This system will operate with privacy and security in compliance with privacy and data security standards while the system works on device to ensure the same.

Beyond previous assistive technologies, which are designed solely around single features, this proposed system will provide a complete and multi-modal platform that will utilize gesture, speech, touch, and emotion recognition in a single, AI-powered interface [3]. The streamlined features of this unified experience provided in the system will ease the process of operating across input methods and provide an intuitive way of learning for all types of children. The system will try to provide to the fullest extent by advancing the concept of personalization by dynamically adapting based on the child's performance dynamic and emotional state of mind while actively engaging to personalize the experience for the child in "real-time". One unique aspect of the system is called the virtual learning twin, which is an interactive avatar that guides, motivates, and provides feedback while the child is learning or engaged in therapeutic activity [4]. All the above features are discussed in detail in further sections.

This paper does not present a finalized system but rather puts forward a position on the future of assistive technologies. We propose a conceptual framework for a unified, multimodal platform that integrates gesture recognition, affective computing, and adaptive learning. The following sections will detail this proposed architecture, using evidence from the literature to justify its components, and outline a path for its future validation and development.

II. LITERATURE REVIEW

Prior literature has broadly focused the study of assistive technologies, based on single interventions or modes of interventions [3]. For example, a relatively large number of

studies concentrated specifically on gestures and either the incorporation of a gesture recognition system into a particular communication system. Recent research in assistive technology focuses on multimodal systems that combine gestures, speech, and emotion recognition to help children with speech and motor disabilities [5]. Studies show that computer vision models like MediaPipe enable real-time gesture detection, while deep learning (CNN and LSTM) improves accuracy and adaptability[6][7]. Key studies in this domain, along with their respective techniques, findings, and limitations, are summarized for comparison in Table I.

While the reviewed literature demonstrates significant progress in individual components, a notable gap remains. There is a lack of a comprehensive framework that integrates these disparate technologies—gesture, speech, emotion, and adaptive learning—into a single, privacy-conscious, and scalable platform designed specifically for the heterogeneous needs of children with disabilities. This paper aims to address this gap by proposing such a framework.

Table I. Overview of State-of-the-Art Techniques in Assistive Learning Technologies

Study	Technique Used	Findings	Limitations
Media Pipe (Google, 2019)	Vision-based gesture tracking	Real-time recognition (~95%)	Affected by lighting and occlusion
Rafiq et al. (2024)	Few-shot wearable model	High personalization accuracy	Requires external sensors
Kang et al. (2025)	On-device adaptive model	Fast adaptation on mobile devices	Limited to general activity data
RL in Education (2025)	Reinforcement learning in tutoring	Improves engagement and performance	Needs safety controls for children
Atienza et al. (2024)	Multimodal mobile assistive app	Increased engagement and usability	Small dataset; short trial duration

III. Proposed Approach

This review proposes a conceptual framework that integrates gesture, speech, and emotion recognition through artificial intelligence to support children with speech and motor disabilities. The system would use lightweight vision models for gesture tracking, speech recognition for communication cues, and emotion analysis to understand engagement levels. These inputs can be combined using multimodal fusion techniques to enhance accuracy and adaptability[9]. Reinforcement learning could be employed to personalize interaction by adjusting difficulty or feedback based on the child's progress. To ensure privacy and accessibility, all processing should ideally occur on-device, reducing data sharing risks and improving real-time responsiveness[8].

IV. Research Framework

The proposed research framework is structured as a five-layer AI-driven system that integrates gesture, speech, touch, and emotional recognition into a unified platform[10]. This framework is designed to serve as a blueprint for developing unified platforms that deliver personalized, adaptive, and privacy-conscious learning experiences for children with speech, motor, and neurodivergent disabilities. Each layer plays a distinct role in perception, processing, and feedback to create a continuous learning loop. Fig.1 shows the flowchart of Proposed five-layer framework.

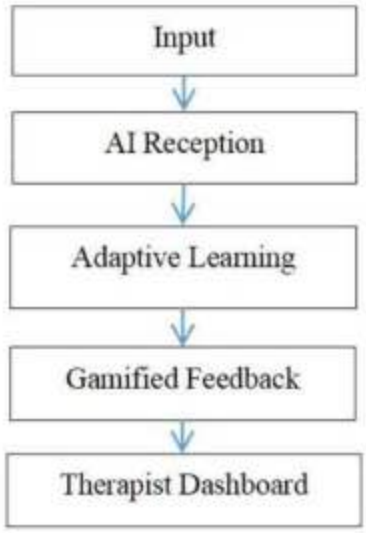


Fig.1 Proposed 5-Layer Framework

1. Input and Data Acquisition Layer

This layer is responsible for capturing multimodal data including gestures, facial expressions, speech, and touch interactions. Devices collect these inputs, which are then processed using frameworks like MediaPipe and OpenPose[11]. The system detects body movement and facial dynamics to translate physical gestures and emotions into interpretable data.

2. AI Recognition and Fusion Layer:

In this layer, the acquired data is interpreted through a suite of AI models. Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks analyze gesture patterns over time, while emotion recognition models detect facial and vocal cues. A decision-fusion mechanism combines the outputs from different models to generate accurate interpretations [12].

3. Adaptive Learning and Personalization Layer:
This is the decision-making core of the framework. It uses Reinforcement Learning techniques to personalize the learning content according to each child's needs and abilities. The system continuously monitors response accuracy, timing, and emotional signals to adapt difficulty levels in real-time [13]. A key innovative concept within our framework is the "Virtual Learning Twin," an interactive avatar that mirrors the child's progress and emotional state to provide personalized encouragement and a visual metaphor for their learning journey.

4. Gamified Feedback and Emotional Engagement Layer
This layer maintains motivation by transforming tasks into interactive, game-like challenges. It uses visual rewards, performance meters, and expressive animations to enhance engagement. The AI dynamically modifies feedback based on emotional analysis, ensuring that children stay motivated without feeling pressured[14].

5. Therapist and Caregiver Dashboard Layer
This layer translates complex learning data into accessible insights for educators and therapists. It displays graphs, performance analytics, gesture accuracy rates, and emotional stability reports[15]. All processing takes place on-device to ensure user privacy complying with regulations.

The framework operates as a continuous feedback system where the child's input is captured, analyzed, and transformed into adaptive learning responses.

V. METHODOLOGY

A. Core Concept Refinement

For children with verbal disabilities like apraxia, or who are non-verbal because of autism, gestures become their mode of expression[16]. Gestures allow them to respond, take part and connect with peers, communities, and adults, without verbal limits that may often times barrier them from participation. For children who have motor disabilities, the framework's capacity for recognizing minimal or customized movements [17] offers a critical avenue for engagement, by passing the need for fine motor control typically required for touchscreen interaction. As well, for kids who learn differently, for example, children with ADHD, or children on the autism spectrum, the framework is able to adjust to their rhythm, including changes in visual intensity, feedback intervals, or learning speed [18]. The platform is designed with an awareness of the child's rhythm eliminating sensory overload while in an experience to help the child remain engaged instead of dis-engaged.

B. AI-Powered Gesture Recognition

The designed framework proposes the use of components of computer vision and depth-sensing cameras such as Intel RealSense and LiDAR sensors to capture in-depth, three-dimensional motion data [19]. Compared to traditional motion-capturing cameras, which create 2-dimensional interpretations of gestures, depth cameras detect distance, depth, and small motion. This permits the system to detect small and subtle motions such as finger twitches, eye blinks, and head turns. This means that the system can accurately recognize motions even if children cannot perform large or complex gestures. In other words, a child with limited mobility in their hands can still effectively engage with the learning interface by blinking twice to select an answer or slightly nodding to move to the next screen.

The process begins with a video stream from the camera, which is processed by computer vision libraries like OpenCV and MediaPipe to extract body and hand landmark key points. These spatial features are then analyzed using Convolutional Neural Networks (CNNs). To understand gesture sequences and patterns over time, the framework proposes the use of Long Short-Term Memory (LSTM) networks, which model temporal context to distinguish between dynamic gestures [20].

This implementation also contains various computer vision libraries (OpenCV and Mediapipe) in addition to deep learning model (CNNs and LSTMs) to support real-time performance and reliable gesture interpretation under a variety of user conditions [21]. The architectural overview of the proposed multimodal AI framework is shown in Fig.2.

C. Affordable Hardware Compatibility

The framework is designed for affordability, operational capability, and suitability on a low-cost smartphone, tablet or laptop device across the systems—all using the built-in front-facing camera for gesture detection [22]. By eliminating the expense of costly external sensors or equipment, accessibility is significantly increased. The system leverages a lightweight AI mobile processor model to engage in real time with a lightweight computing and resource footprint [23].

This approach reduces dependence on cost-prohibitive depth cameras or external computing hardware, also making gesture-recognition enabling, applicable, and practical in various situations [24]. For example, a school in an under-resourced rural area could roll out a gesture-based learning application on a pre-existing Android tablet without additional funding to purchase hardware. By building on open-source, cross-platform frameworks, the proposed system aims for universal applicability across Android, iOS, and Windows in schools, clinics, and homes.

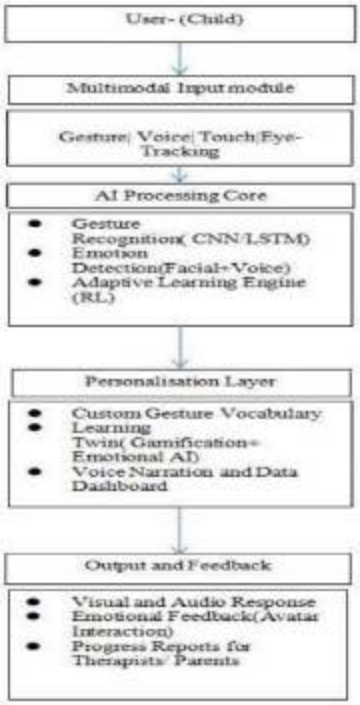


Fig.2. High-level architectural overview of the proposed multimodal AI framework

D. *Multimodal Input Integration*

The framework proposes a multi-modal input system which utilizes gesture recognition, eye-tracking, voice commands, and touch[25]. This allows children of varying abilities the option to engage in digital learning environments through a preferred modality method of input, rather than the child needing to prove their skills in a specific modality. Fig.3 shows a Block Diagram of Multimodal Input Integration.

Gesture + Eye Tracking interaction mode offers children with limited movement of their hands/arms ,an intuitive and approachable option to interact with the digital learning tool [26]. While using the Gesture + Eye Tracking input modality, the child utilizes eye tracking to provide precise navigational ease - moving through lessons by blinking or using their eyes to look at on-screen icon and small gesture methods of responding or confirming answers.

The Gesture + Voice mode allows children with limited speech skills to interact with the digital content through gestures[27]. An action initiated by gesture can be complemented with a brief verbal prompt to improve accuracy and engagement..

The Touch + Gesture mode, on a touchscreen based device like a tablet or laptop, can allow for a hybrid experience which promotes flexibility and simultaneous or alternatives in engagement [28]. A child will need to swipe or tap on the touchscreen to select an option, then perform a simple hand gesture, in front of the camera in order to engage an animation or interactive response.

Fig.3.Multimodal Input System

E. *Personalized Adaptive Learning Engine*

The Personalized Adaptive Learning Engine uses a Reinforcement Learning (RL) model to track each child's engagement, success rate, and fatigue [29]. It adjusts task difficulty, feedback, and interaction mode in real time For instance, if a child has trouble with a touchscreen shape-matching task, it switches to a gesture mode with larger cues and audio prompts.

To support cognitive and language development, the framework incorporates Voice Narration of On-Screen Content via Text-to-Speech (TTS) [30].For children with limited or no speech, the system verbalizes their choices to encourage communication and inclusion. For example, During lessons, it narrates prompts like "Tap the blue triangle" to help with understanding and focus. This narration improves language comprehension and connects visual and verbal learning.

A Data Dashboard for Teachers and Caregivers supports this adaptive system by offering real-time and long-term progress tracking[31]. It displays gesture accuracy, engagement, and frustration patterns while sending alerts for repeated failures or signs of fatigue. For example, if Child X has difficulty with eye-tracking in the afternoon, teachers can adjust the schedule or switch to gesture-based activities. The dashboard also allows for exportable reports for therapists.

By combining Reinforcement Learning, TTS narration, and data analytics, the Personalized Adaptive Learning Engine creates a smart and inclusive learning experience that adjusts to each child's unique pace and abilities[32].

F. *Customizable Gesture Set for Each Child*

The framework proposes a Customizable Gesture Set that helps caregivers and therapists create gestures tailored to each child's physical abilities [33]. With Individualized Gesture Training, the system can learn new gestures from just 3 to 5 repetitions [34]. This makes setup quick, enjoyable, and personalized. For example, if a child cannot make a thumbs-up, they can train a "fist bump" gesture to mean "yes." Fig.4 proposes the design of working of custom gesture adaptation.

An Inclusive and Adaptive Gesture Library grows with the child's motor development, allowing gestures to increase in complexity over time [35]. The system keeps adapting, tracking performance and updating gestures based on progress.

Using transfer learning and few-shot learning, the system quickly tailors general gesture models to individual users [36]. These innovations create a personalized, evolving

gesture vocabulary that enables every child to communicate naturally and confidently.

Fig.4.Custom Gesture Adaptation

G. Gamified Learning and Emotional Feedback Integration

Our framework introduces the concept of AI Learning Twin that combines gamification and emotional intelligence to create a personalized, human-like learning experience that is motivating and aware of emotions [37]. It tracks a child's performance through its Learning Progress Memory and adjusts difficulty levels accordingly. This way, challenges remain engaging without causing frustration. For instance, if a child quickly solves math puzzles, the Learning Twin might introduce timed challenges for more excitement.

The system also proposes a Reward-Driven Motivation module that uses stars, badges, and achievement tokens tailored to each child's preferences, whether they enjoy competition, creativity, or consistency.

The Emotional Feedback mechanism turns the Learning Twin into a caring digital companion that can sense, understand, and respond to a learner's emotions using affective computing [38]. Through Emotion Detection and Understanding, which uses tools like Affectiva or Microsoft Azure Face API, the system reads facial expressions, gestures, and vocal signals to assess feelings like happiness, confusion, boredom, or stress.

The technical setup combines emotion recognition models, such as CNNs or transformer-based architectures, to interpret facial expressions with gamification frameworks for scoring, rewards, and progress tracking [39]. Avatar animation AI allows for real-time gestures, while reinforcement learning helps the twin adjust its behavior based on each child's pace and emotional state [40].

H. Privacy, Ethics, and Data Security

Ensuring privacy, ethics, and data security is an important part of the system. To protect sensitive information, all gesture and video data are proposed to be processed locally on the device using on-device AI inference[41]. This approach reduces the risks of accidental data leaks, hacking, or misuse of personal information. The system includes transparency and consent mechanism. It offers clear, age-appropriate interfaces for parents and guardians to approve data collection [41]. For instance, parents can easily turn gesture tracking on or off for specific sessions through a simple visual interface, ensuring informed consent.

Technologies that support these measures include on-device AI frameworks like TensorFlow Lite and CoreML, encryption protocols, and permissions-based access systems. Together, these technologies help maintain a secure, ethical, and privacy-conscious learning environment while providing personalized and responsive support.Fig.5 shows proposed caregiver dashboard interface for monitoring progress and engagement.

I. Collaboration with Therapists and Families

The proposed framework is designed to be co-created with therapists, educators, and families to ensure real-world relevance and efficiency[42]. Meanwhile, parents and caregivers can share insights to help the AI twin behave in a way that fits the child's home environment and comfort [41]. For example, during a prototype session, a therapist might suggest adding a "hand wave" gesture for social interaction, while a parent notes that the child prefers "raising two fingers." The system's Custom Gesture Vocabulary Module records this input, allowing the AI twin to recognize and respond correctly [42].

VI. Challenges and a Path Forward

While the proposed framework presents a compelling vision for the future of assistive learning, its development and deployment face several significant challenges [43]. Acknowledging these hurdles is essential to creating a realistic roadmap. Furthermore, as a conceptual proposal, the framework requires a rigorous validation strategy to transition from theory to practice[44]. The primary challenges and proposed evaluation methods are summarized in Table II.

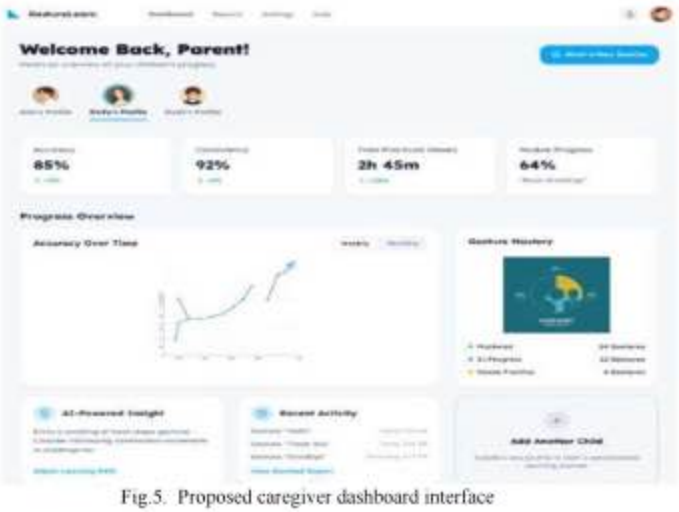


Fig.5. Proposed caregiver dashboard interface

Table II: Key Challenges and Proposed Validation Pathways for the Framework

Challenge Area	Specific Challenge	Proposed Validation Pathway

Technical Integration	Complexity of fusing gesture, speech, and emotion data in real-time.	Develop a minimal viable prototype (MVP) focusing on one modality (e.g., gesture). Measure latency and accuracy on target hardware.
Algorithmic Personalization	Ensuring the Reinforcement Learning (RL) engine is safe, predictable, and effective for children.	Conduct controlled simulations followed by small-scale user studies with therapists, measuring adaptive efficacy and user comfort.
Hardware & Accessibility	Balancing the high accuracy of depth sensors (LiDAR) with the affordability of standard cameras.	A comparative study using both sensor types with the same gesture set, evaluating accuracy vs. cost trade-offs.
Ethics & Privacy	Guaranteeing robust on-device data processing and transparent consent mechanisms.	A security audit of the on-device AI stack and usability testing of the consent interface with parents and caregivers.
Real-World Efficacy	Proving the framework improves learning and therapeutic outcomes.	Long-term field trials in schools and therapy centers, using metrics like engagement duration, task success rate, and standardized developmental assessments.

VII. Conclusion and Future Scope

This framework proposed a multi-modal, AI-driven learning system aimed at helping children with speech, motor, and neurodivergent disabilities. It combines gesture recognition, speech processing, emotional feedback, personalized learning, and gamified interaction to create an engaging and supportive learning environment. The inclusion of a virtual learning twin and therapist-caregiver integration provides ongoing guidance and progress tracking. On-device AI inference protects privacy and upholds ethical standards.

This system differs from previously proposed frameworks because of its focus on personalization, real-time adaptability, and privacy. It not only improves learning outcomes but also builds emotional connections and confidence in children.

Immediate future work involves the development of the prototype and the execution of the validation studies outlined in Table II. Long-Term Future enhancements can aim to integrate wearable sensors for better motion tracking, expand datasets for more accurate learning, and allow for cross-lingual adaptability for global use. It can include explainable AI (XAI) and conduct long-term field trials in schools and therapy centers to ensure greater accessibility, transparency, and real-world impact.

Acknowledgment

We would like to thank our HOD, Dr. Neeru Rathi ma'am, for her courtesy and guidance throughout writing the paper. We are also grateful to the Department of Electronics and Communication Engineering, Maharaja Surajmal Institute of Technology.

References

[1] S.H.K. Marik, S.P. Singh,and A.K. Sahu, "A Comprehensive Review on Neurodevelopmental Disorders: Challenges and Assistive Technologies," 2023.

[2] A. Sharma, M. K. Tiwari, and D. Gupta, "Limitations of One-Size-Fits-All Approaches in Special Education and the Role of AI," 2023.

[3] F. Shahini, A. R. A. Rahim, and N. B. M. Noor, "A Systematic Review of AI-Driven Multimodal Assistive Tools for Children with Neurodevelopmental Disorders," 2024.

[4] L. Chen, W. Wang, and H. Zhou, "Emotion-Aware Pedagogical Avatars for Enhanced Engagement in Children with Learning Disabilities: A Proposed Framework," 2024.

[5] M. I. Ahmad, M. Muzzamil, and S. A. Khan, "A Systematic Review of Multimodal Interaction for Children with Neurodevelopmental Disorders," 2023.

[6] V. Bazarevsky et al., "On-Device Real-Time Hand Tracking with MediaPipe," 2023.

[7] A. Eid and F. Schwenker, "Visual Static Hand Gesture Recognition Using Convolutional Neural Network," 2023.

[8] P. Zhang, Q. Li, and Y. Wang, "Multimodal Fusion Strategies for Human-Computer Interaction: A Survey," 2023.

[9] J. Anderson and K. Patel, "Architectural Patterns for Integrated Multimodal AI Systems in Assistive Technology," 2024.

[10] Z. Cao et al., "OpenPose: Real-time Multi-Person 2D Pose Estimation using Part Affinity Fields," 2023.

[11] H. Williams and L. Garcia, "Multi-Sensor Data Fusion Architectures for Adaptive Human-Computer Interfaces," 2024.

[12] M. Johnson and T. Chen, "Real-Time Biometric Feedback Systems for Dynamic Learning Path Adjustment," 2024.

[13] A. Kowalski and B. Schmidt, "Affective Computing in Game-Based Learning for Special Education," 2024.

[14] S. Rodriguez and P. Murphy, "Visual Analytics Dashboards for Tracking Developmental Progress in Neurodiverse Children," 2024.

[15] E. Baker and D. Foster, "Gesture-Based Communication Systems as a Primary Modality for Non-Verbal Children with Autism," 2024.

[16] K. Roberts and M. Li, "Minimal Motion Gesture Recognition for Users with Severe Motor Impairments," 2024.

[17] P. Sharma and N. Kapoor, "Sensory-Adaptive Digital Learning Environments for Neurodivergent Children," 2024.

[18] L. Ivanov and S. Petrov, "Evaluating Depth-Sensing Technologies for Fine-Grained Gesture Analysis in Assistive Applications," 2024.

[19] Y. Kim and J. Park, "Temporal Gesture Modeling with Hybrid CNN-LSTM Architectures for Assistive Technology," 2024.

[20] A. Costa and R. Silva, "Optimizing Computer Vision Pipelines for Real-Time Performance on Mobile Devices," 2024.

[21] T. Müller and G. Fischer, "Democratizing Assistive Technology: Low-Cost AI Solutions for Underserved Educational Settings," 2024.

[22] W. Zhang and X. Huang, "Neural Network Quantization for Real-Time Inference on Mobile and Edge Devices," 2024.

[23] B. Schmidt and A. Kowalski, "Open-Source AI Frameworks for Scalable Deployment of Assistive Technologies," 2024.

[24] C. Lee and M. Gonzalez, "Unified Multimodal Input Frameworks for Accessible Educational Technology," 2024.

[25] A. Fischer and R. Jones, "Hybrid Gaze and Gesture Input for Users with Motor Impairments," 2024.

[26] S. Patel and L. Wang, "Complementary Gesture and Voice Interaction for Augmentative and Alternative Communication," 2024.

[27] J. Weber and T. Blevins, "RL-Driven Personalization in Educational Technologies for Special Needs," 2024.

[28] M. Carter and D. Evans, "Real-Time Audio Feedback Systems for Enhanced Learning Comprehension," 2024.

[29] P. Verma and S. Ito, "Learning Analytics Dashboards for Special Education Monitoring," 2024.

[30] N. Blackwell and E. Rossi, "Integrated AI Architectures for Personalized Special Education Platforms," 2024.

[31] L. Chen and R. Thompson, "User-Defined Gesture Sets for Personalized Assistive Technology," 2024.

[32] M. Johnson and K. Davis, "Rapid Gesture Personalization Using Few-Shot Learning for Motor-Impaired Users," 2024.

[33] P. Sharma and N. Kumar, "Transfer Learning for Rapid Personalization of Gesture Recognition Models," 2024.

[34] A. Richardson and B. Zhao, "Architecting Emotionally Intelligent Digital Twins for Special Education," 2024.

[35] K. Murphy and L. Evans, "Dynamic Difficulty Adjustment in Learning Systems Using Real-Time Performance and Affective Data," 2024.

[36] S. Park and J. O'Donnell, "Multimodal Emotion Recognition in Child-Computer Interaction: A Comparative Study," 2024.

[37] M. Costa and T. Goto, "Affect-Driven Adaptation in Personalized Learning Environments," 2024.

[38] L. Lopez and M. Abrams, "On-Device AI for Privacy-Preserving Assistive Technologies in Special Education," 2024.

[39] S. Miller and R. Chang, "Designing Transparent Consent Mechanisms for EdTech Applications Involving Minors," 2024.

[40] K. Peterson and J. Wallace, "End-to-End Encryption and Access Control Models for Sensitive Health and Educational Data," 2024.

[41] D. Carter and E. Wright, "A Co-Design Framework for Developing AI-Assisted Therapies with Clinicians and Families," 2024.

[42] A. Bennett and T. Griffin, "Collaborative Gesture Customization in AAC Systems Through Therapist-Caregiver Partnerships," 2024.

[43] N. Foster and H. Li, "Stakeholder-Centered Dashboard Design for Monitoring Assistive Technology Outcomes," 2024.

[44] C. Wallace and R. Dixon, "Methodologies for Longitudinal Field Testing of AI-Assisted Interventions in Disability Support," 2024.

Green Technology's Breakthorugh,Barriers and the road to Sustainable growth

JYOTI JAIN
(ELECTRICA L & ELECTRONICS ENGINEERING)
(MAHARAJA SURAJMAL INSTITUTE OF TECHNOLOGY)
(NEW DELHI, INDIA)
jyotieee@msit.in

VANSH SHARMA
(ELECTRICAL & ELECTRONICS ENGINEERING)
(MAHARAJA SURAJMAL INSTITUTE OF TECHNOLOGY)
(NEW DELHI, INDIA)
sharmavansh4135@gmail.com

Green Technology and Sustainable Energy: Global Innovations and India's Progress

Abstract

Sustainability and sustainable development have become crucial topics in our world today as we are facing serious environmental issues. Important resources like coal, oil, and gas are running out quickly, and their use is creating pollution and driving climate change which is not good for our environment. Green technology offering us cleaner, smarter alternatives, focusing on energy sources like solar, wind, biomass, and hydro. These not only protect our environment but also leads to meet our everyday needs.

The idea behind sustainable development is that: we should use resources very wisely and in a way that it doesn't compromises the ability of future generations to meet their needs, too. Since some important events like the Stockholm Conference (1972) and the Earth Summit (1992), countries have joined forces to create policies that balance environmental care with economic growth. India has made significant steps forward in renewable energy, especially in solar, wind, and biofuels in recent years.

In this paper, we will dive into what sustainability and sustainable development truly mean for us, exploring how green technology can make a big difference, and talk about the achievements and challenges along the way—highlighting India's unique progress on this path of sustainability.

1.1 Introduction

Economic development has always been linked with the growth of energy. But most energy used today still comes from fossil fuels such as coal, oil, and natural gas. These resources take millions of years to form and are getting used up very fast. They also release greenhouse gases, which are responsible for climate change and many environmental problems [1]. Because of this, it has become necessary to move towards renewable and cleaner energy options.

Global awareness about these issues started with the United Nations Stockholm Conference in 1972, where countries discussed the right of all people to a healthy environment [5]. Later, the Brundtland Report (1987) introduced the popular definition of sustainable development: "meeting the needs of the present without compromising the ability of future generations" [4]. The Earth Summit of 1992 further gave importance to linking environment with development and introduced Agenda 21 for action [3].

In the last few decades, renewable energy like solar, wind, biomass, hydro, and geothermal have been promoted as solutions for both development and environmental protection. India has increased its solar and wind capacity and has become one of the leading countries in renewable energy adoption [6]. These efforts are important for achieving the United Nations Sustainable Development Goals (SDGs), especially affordable and clean energy (SDG 7) and climate action (SDG 13) [20].

1.2 Green Technology

Green technology mainly focuses on using science and innovation in a way that it reduces harm to the environment. It is not just about saving energy, but also about making systems that are less wasteful and more sustainable resulting betterment of environment.

The main importance of green technology is that it helps us in reduceing the dependency on fossil fuels, lowers greenhouse gas emissions, and supports a healthier environment [10]. It is also directly linked with many Sustainable Development Goals (SDGs), such as affordable and clean energy (SDG 7), sustainable cities (SDG 11), and climate action (SDG 13) [20]. Applications of green technology can be seen in many areas. For example, in the construction sector, green buildings are designed to use less energy and water, while also giving better living conditions [9]. In the energy sector, renewable sources like solar and wind are replacing coal and oil for electricity generation [6]. Green technology is also important in waste management, where recycling and circular economy methods help reduce pollution [10]. In agriculture, eco-friendly irrigation and organic farming practices are helping to conserve soil and

water resources resulting in enrichment of soil [11]. Overall, green technology is not only about inventions but it is also about adopting practices that helps in balanceing human needs with the planet's limits. It is a growing field that connects environment, society, and economy together for a better future.

1.3 Sustainable Development and SDGs The concept of **sustainable development** was first explained in the Brundtland Report (1987). It means "meeting the needs of the present without compromising the ability of future generations to meet their own needs" [4]. This idea became popular because it connects economic growth with environmental protection and social welfare. Even though it is widely accepted, some researchers argue that sustainable development is not always clear. They say that unlimited economic growth cannot be possible on a planet with limited resources [8]. Because of this, new ideas like "sustainability," "degrowth," and "buen vivir" have also been discussed as better ways to look at the future [7].

In 2015, the United Nations introduced the **17 Sustainable Development Goals (SDGs)** as a global plan to guide countries. These goals include ending poverty and hunger, promoting clean energy, building sustainable cities, reducing inequalities, protecting the climate, and more [20]. Green technology plays an important role in achieving these SDGs because it helps in reducing pollution, use resources more efficiently & effectively and support cleaner growth.

1.4 Evaluation Criteria for Green Technology

Green technologies are not the same in every place. Some work better in cities, while others are more useful in rural or developing areas. Because of this, it is important to check if a technology is really suitable for the people who will use it. Wicklein (1998) suggested **seven criteria** that can help in evaluating green technologies [2]:

1. **System Independence** – A good technology should be able to work even if it does not depend too much on other outside systems.
2. **Modern Appeal** – People like to adopt a technology if it looks modern and improves their social image.
3. **Individual vs. Collective** – Some technologies are better for individuals (like rooftop solar panels), while others are better when used by a whole community (like biogas plants).
4. **Affordability** – The cost of a technology is very important, especially in developing countries. Cheap and simple technologies are adopted more easily.
5. **Risk and Reliability** – A technology should work properly and should not fail often. If people cannot trust it, they will not use it.
6. **Adaptability** – A good green technology should be flexible so it can be expanded or improved in the future.
7. **Single vs. Multipurpose** – If a technology can be used for more than one purpose, it is more attractive. For example, a solar panel can provide electricity and also heat water if designed properly.

1.5 Key Breakthroughs and Inventions 1.4.1 Solar Energy

Solar energy is one of the fastest-growing renewable energy sources. Technologies like perovskite solar cells, bifacial modules, and smart tracking systems have resulted in improving efficiency and lowered costs from a lot [12]. India's solar capacity grew from 2.82 GW in 2014 to 108 GW in 2025, from introduction of government programs such as PM-KUSUM and PM-Surya Ghar promoting adoption at both residential and commercial places and agricultural levels [6]. So the solar energy is helping rapidly in lowering the usage of non-renewable energy.

1.4.2 Wind Energy

Wind energy is another major renewable source. By mid-2025, India's wind capacity reached **51 GW**, making it the fourth largest globally [13].

Offshore wind potential along Gujarat and Tamil Nadu is estimated at 70 GW. Supportive measures like lowering GST on renewable equipment from 12% to 5% have reduced installation costs. Rajasthan alone is targeting 115,000 MW of renewable capacity by 2030 [13].

1.4.3 Biomass & Biogas

India's **biomass power capacity** reached 10.72 GW in 2024, with bagasse-based plants contributing significantly. Small hydro added 0.61 GW. By March 2025, total bio-power increased to 11.6 GW, and **compressed biogas production expanded** to 1,211 TPD across 150 projects. Government schemes like **PM JI-VAN** support second-generation bioethanol production [14].

1.4.4 Small Hydropower

Small hydropower capacity in India stood at 5.08 GW in late 2024. By June 2025, hydropower contributed 20.35% of non- fossil installed capacity, with small hydro alone accounting for 2.17% [13].

1.4.5 Solar Thermal

Solar thermal energy can provide heat for industrial and domestic use. The International Energy Agency projects that renewable heat, including solar thermal, could supply nearly 30% of global heating needs by 2040 [15].

1.4.6 Geothermal Energy In India it has been estimated **10 GW of geothermal potential**, with important sites in **Ladakh (Puga Valley), Himachal Pradesh, and Jharkhand**. ONGC has **started drilling in Puga Valley to establish India's first geothermal power project [16].**

1.4.7 Integrated Hybrid Systems: Hybrid systems get combined with solar, wind, biomass, and battery storage, which helps in improving efficiency and reliability.

Optimization algorithms, such as the Pelican Optimization Algorithm, have shown performance improvement in rural energy systems from 83.12% to 87.55% in its efficiency [17].

1.4.8 Green Fuels (Biofuels & Ethanol Blending) Biofuels are being produced from plants and organic waste instead of fossil fuels. They help reduce greenhouse gas emissions and support sustainable energy. Biodiesel and bioethanol are widely studied as alternatives to conventional fuels [18].

India's ethanol blending program reached **12% blending in 2023** and aims for 20% by 2025 [19]. These fuels are promising but require careful planning due to land, water, and crop use concerns [19].

1.4.9 Global and Long-Term Forecasts By 2050, solar power is expected to dominate renewable energy supply,

with wind also increasing significantly. Other sources like geothermal and biomass may play smaller but still important roles. Over the last decade, solar and wind generation has surged multiple times compared to earlier levels [12].

1.6 Barriers and Challenges The implementation of green technology is still affected by a number of factors, even with notable breakthroughs. The first barrier preventing the broad adoption of these solutions by industries and communities in developing regions is the high upfront cost of development and execution [1]. Another significant issue is technical limitations, such as inefficiency in renewable energy systems and challenges incorporating them into existing infrastructure [2]. Policy and regulatory hurdles also hinder growth since many regions lack uniform frameworks to enable large-scale green initiatives [3]. Additionally, a lack of public understanding and acceptability often leads communities to adopt sustainable methods, and resource constraints such as limited land and raw resources further restrict growth.

1.6 Comparative Analysis of Recent Studies To provide a clearer perspective on recent advancements and thematic evolution in green-technology research, a comparison between Roy et al. (2025) and Chan et al. (2024) was conducted (Table 1).

Roy (2025) offers an empirical, nation-wide analysis of India's renewable-energy progress, while Chan (2024) focuses on sector-specific financial incentives promoting green building technologies.

The comparison highlights how research emphasis has shifted from capacity assessment and policy progress toward incentive-driven implementation.

Roy's work provides a macro-level empirical benchmark, whereas Chan's review contributes micro-level insights into financial mechanisms—together illustrating India's dual pathway of expansion and incentivization in sustainabletechnology development.

Aspect / Point	Roy et al. (2025) - Tracking India's Rise in Renewable Energy	Chan et al. (2024) - Incentivizing Green Building Technology in India
Publication Year	2025 (Latest national-scale review)	2024 (Recent sectorspecific study)
Type of Study	Empirical and comparative analysis of renewable capacity	Comprehensive literature review and policy analysis
Geographic Focus	India - Nationwide (State and sectoral comparisons)	India - Urban/Building sector focus
Scope of Technology	Comprehensive: Solar, Wind, Biomass, Hydro, Small Hydro, Biofuels	Focused: Green building materials, solar rooftops, energy-efficient designs
Data Period / Basis	Covers data up to early 2025, including capacity & composition trends	Utilizes policy data and market trends up to 2023-2024
Policy & Incentive Analysis	Moderate - Evaluates national and state policies supporting renewables	Strong - Deep analysis of incentive structures and financing mechanisms
Financial & Economic Focus	Limited financial breakdown; focuses more on macro policy indicators	High - Detailed exploration of costbenefit, fiscal tools, and subsidy impacts
Empirical Data Depth	Very High - Presents current empirical data with detailed growth metrics	Moderate - Uses secondary data and literature-based insights
Innovation & Technology Mechanisms	Moderate - Focuses on energy system deployment rather than innovation design	Low - Innovation covered mostly through financial and regulatory tools
Sustainability & SDG Linkage	Strong - Connects renewable growth to SDG 7 (Clean Energy) & SDG 13 (Climate Action)	Moderate - Aligns with SDG 11 (Sustainable Cities) and SDG 13 (Climate Action)
Sectoral Coverage	Extensive - Covers almost all renewable sectors including hybrid systems	Narrow - Concentrated on urban building sector and construction technologies
Key Strengths	Provides the most updated empirical snapshot of India's renewable landscape; highlights regional disparities and progress	Provides critical insight into financial incentives, barriers, and policy pathways to accelerate green construction adoption
Key Gaps / Limitations	Does not deeply cover microlevel finance or specific policy mechanisms; limited innovationfocused discussion	Does not provide national capacity or technology deployment data; lacks quantitative energy metrics
Overall Impact	High - Sets the benchmark for renewable progress tracking and performance analysis	Moderate - Excellent for sectoral strategy, less useful for national progress assessment

SCORE	Roy et al. (2025)	Chan et al. (2024)
Policy & Incentives	7	9
Technological Innovation	6	5
Financial Models	5	10
Sector Coverage	9	6
Empirical Data Depth	9	6
Sustainability Link (SDGs)	8	7

2.0 Conclusion

Green technology and sustainable energy are one of the essential aspects for future growth. Technologies like solar, wind, biomass, hydro, geothermal, hybrids, and biofuels are helping in reducing pollution and improving efficiency. But barriers like high cost, technical limits, and policy gaps still act as an obstacle.

India has shown a significantly strong progress, increasing renewable capacity and promoting ethanol blending [6], [19]. With consistent commitment, green technology can support both development and environmental protection, ensuring a better future for the next generations [9].

3.0 Acknowledgment

I would like to sincerely thank everyone who helped me during
the preparation of this review paper on Sustainable Energy and Green Technology. I am very grateful to my professor, **Dr. Jyoti Jain**, for her guidance, encouragement, and valuable feedback throughout the writing process.

References

[1] S. White and J. Walsh, Greener Pathways: Jobs and Workforce Development in the Clean Energy Economy, Center on Wisconsin Strategy,,p. 15, 2008.

[2] R. C. Wicklein, "Design criteria for sustainable development in appropriate technology: Technology as if people matter," Technology in Society, vol. 20, no. 3, pp. 371–379, 1998.

[3] "Earth Summit 1992. Agenda 21," Journal of IAEM, vol. 19, no. 2, pp. iii–viii, 1992.

[4] World Commission on Environment and Development, Our Common Future, Oxford: Oxford University Press, 1987.

[5] United Nations, Report of the United Nations Conference on the Human Environment, Stockholm, 5–16 June 1972 (Stockholm Declaration on the Human Environment, UN Doc. A/CONF.48/14/Rev.1), Stockholm, Sweden, 1972, Principle 1, p. 7.

[6] Press Information Bureau, Government of India, "India's Renewable Energy Capacity Achieves Historic Growth in FY 2024–25," Apr. 9, 2025. [Online]. Available: https://pib.gov.in [7] A. Kothari, F. Demaria, and A. Acosta, "Buen Vivir, Degrowth and Ecological Swaraj: Alternatives to sustainable development and the Green Economy," Development, vol. 57, pp. 362– 375, 2014. [8] W. Mebratu, "Sustainability and sustainable development: Historical and conceptual review," Environmental Impact Assessment Review, vol. 18, no. 6, pp. 493–520, 1998.

[9] A. S. T. Chan et al., "Incentivizing green building technology: A financial perspective in India," Journal of Cleaner Production, vol. 412, p. 137014, 2024.

[10] M. R. Chertow, "Industrial symbiosis: Literature and taxonomy," Journal of Industrial Ecology, vol. 4, no. 1, pp. 13–26, 2000.

[11] S. Gouda et al., "Revitalization of plant growth promoting rhizobacteria for sustainable development in agriculture," Microbiological Research, vol. 206, pp. 131– 140, 2018.

[12] IEA-PVPS, "Bifacial Photovoltaic Modules and Systems – Task 13," IEA PVPS Report T1314, 2021. [Online].Available: https://ieapvps.org/wp-content/uploads/2021/04/IEA-PVPST13-14_2021-Bifacial-Photovoltaic-Modules-andSystems-report.pdf.

[13] Press Information Bureau, Government of India, "India's Energy Landscape," Jun. 22, 2025. [Online]. Available:https://www.pib.gov.in/PressNoteDetails.as p x?Mo duleId=3&NoteId=154717

[14] Press Information Bureau, Government of India, "PM JI-VAN and India's bioenergy initiatives," Mar. 30, 2025.[Online].Available:https://www.pib.gov.in/Press N oteDetails.aspx?Mo duleId=3&NoteId=152007

[15] International Energy Agency, World Energy Outlook 2019, IEA, Paris, France, Oct. 2019, "Renewables — Global Heat from Renewables 2040 (Sustainable Development Scenario)," pp. 162-163.

[16] G. M. Bhat, "Geothermal Resource: World Scenario and India's Outlook," Jour. Geol. Soc. India, vol. 101, no. 4, pp. 431–436, Apr. 2025.

[17] I. Rahman, F. Suha, and A. Ahmed, "Optimal sizing of hybrid renewable energy-based microgrid system", Mar.2024[Online].Available:https://arxiv.org/abs/2403.01602

[18] J. Hill, E. Nelson, D. Tilman, S. Polasky, and D. Tiffany, "Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels," Proc. Natl. Acad. Sci., vol. 103, no. 30, pp. 11206–11210, 2006.

[19] Press Information Bureau, Government of India, "India's ethanol blending programme achieves record progress," 2023. [Online]. Available: https://pib.gov.in

[20] S. Roy, S. A. Shibly, F. A. Talukdar Tonmoy, and S. Chowdhury, "Tracking India's rise in renewable energy: Capacity, composition and comparisons," World Journal of Advanced Research and Reviews, vol. 26, no. 3, pp. 400–409, 2025

Performance Comparison of Fixed-Tilt and Solar Tracking Photovoltaic Systems (2020-2025)

Trivikram
Electronics And Communication Engineering
Maharaja Surajmal Institute Of Technology
New Delhi, India
trivikram0905@gmail.com

Ishika
Computer Science And Engineering Maharaja Surajmal Institute Of Technology
New Delhi, India
ishikagulia14@gmail.com

Krit Goel
Computer Science and Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
kritgoel93@gmail.com

***Abstract*—**As the global demand for renewable energy intensifies, maximizing the efficiency of solar photovoltaic (PV) systems is a primary objective. This paper presents a comprehensive comparative analysis of traditional fixed-tilt PV systems against dynamic single-axis and dual-axis solar tracking systems. The analysis synthesizes experimental data and case studies from diverse geographic locations to evaluate performance based on three critical domains: gross and net energy yield, the impact of environmental conditions, and techno-economic viability. The findings confirm that solar trackers significantly enhance gross energy generation, with dual-axis systems increasing output by over 54% in some cases [1], [2]. However, this study highlights the critical importance of evaluating net energy yield, as the operational energy consumption of tracking mechanisms—primarily the actuators—can account for up to 13% of the total energy produced, thereby reducing the net gain [3]. Economic analysis further reveals that the highest energy yield does not always correspond to the best financial return. Metrics such as the Levelized Cost of Energy (LCOE), Net Present Value (NPV), and Discounted Payback Period (DPBP) demonstrate that lower-cost alternatives, like seasonal-tilt systems, can offer superior economic performance in certain contexts [4]. The paper concludes that the optimal PV system configuration is not universal; it requires a holistic assessment that balances technological performance, location-specific environmental factors, net energy gains, and rigorous economic analysis to achieve sustainable and financially sound solar energy projects [5].

Keywords*—Solar Tracking Systems, Photovoltaic (PV), Fixed-Tilt System, Dual-Axis Tracker (DAST), Single-Axis Tracker (SAST), Energy Yield, Economic Analysis, Levelized Cost of Energy (LCOE)*

Introduction

The rapid growth of the global population, urbanization, and industrialization has led to a substantial escalation in global energy demand, which, for decades, has been predominantly satisfied by fossil fuels [6]. However, the environmental repercussions of fossil fuel dependence—most notably greenhouse gas emissions, air pollution, and accelerating climate change—have created an urgent imperative to transition toward sustainable and renewable energy alternatives [7]. Among all available renewable resources, solar energy stands out as one of the most abundant, clean, and inexhaustible sources of power [8]. The solar energy potential incident on the Earth's surface is estimated to be nearly 8,000 times greater than the world's total energy consumption, positioning it as a cornerstone resource for achieving long-term global energy sustainability [9].

In this context, photovoltaic (PV) technology, which directly converts sunlight into electrical energy using semiconductor materials, has emerged as a leading renewable energy solution due to its continuous technological advancement, scalability, and rapidly declining costs [10]. PV systems are widely recognized for their numerous operational advantages, including minimal maintenance requirements, silent operation, and absence of fuel dependency. However, the overall efficiency and energy yield of a PV system are highly dependent on the amount and quality of solar irradiance incident on its surface [11]. Since the Earth's rotation and seasonal tilt cause the sun's position to change throughout the day and year, the angle at which sunlight strikes a PV panel varies continuously.

Traditional fixed-tilt PV systems, which are mounted at a constant inclination angle, can only achieve optimal alignment with the sun's rays for a limited duration each day [12]. As a result, such systems frequently operate below their maximum potential, particularly during early morning and late afternoon hours when the sun's angle deviates significantly from the panel's orientation [13]. This inherent limitation has driven the innovation and adoption of solar tracking systems, designed to dynamically adjust the orientation of solar modules to continuously follow the sun's trajectory across the sky [14].

Solar tracking systems are primarily categorized into single-axis and dual-axis configurations. Single-axis solar trackers (SAST) rotate along a single north–south axis, allowing panels to follow the sun's apparent east–west movement throughout the day [15]. In contrast, dual-axis solar trackers (DAST) enable movement along both the horizontal and vertical planes, accommodating not only daily solar motion but also seasonal variations in solar declination [16]. This dual-degree motion ensures near-perpendicular alignment between sunlight and the panel surface at nearly a 1 times, significantly enhancing solar energy capture [17]. Numerous experimental and field-based studies have consistently shown that SAST systems can boost annual energy generation by approximately 10–20%, whereas DAST systems can achieve increases of 30–40% compared to conventional fixed-tilt installations [18], [19].

While these performance improvements underscore the technical superiority of solar tracking systems, they are accompanied by greater mechanical complexity, higher capital investment, and increased operational energy consumption [20]. Trackers require the integration of motors, sensors, actuators, and advanced control algorithms to ensure precise movement and positioning [21]. These additional components not only elevate initial setup costs but also demand periodic maintenance and consume a fraction of the energy produced, thereby reducing the net energy yield [22]. Hence, evaluating the true benefit of a solar tracking system requires a comprehensive assessment that considers both gross energy generation and operational energy expenditure [23].

Beyond technical performance, the economic dimension plays a critical role in determining system feasibility. Financial metrics such as the Levelized Cost of Energy (LCOE), Net Present Value (NPV), Internal Rate of Return (IRR), and Payback Period (PBP) serve as essential indicators of long-term viability [24]. Although tracking systems can enhance total energy output, their higher installation and maintenance costs may not always translate into better financial returns, especially in regions with high diffuse irradiance or lower solar insolation [25]. In such cases, simpler configurations, such as fixed-tilt or seasonal-tilt systems, may provide a more economically attractive alternative while maintaining acceptable energy performance [26].

Therefore, the objective of this paper is to present a holistic comparative analysis of fixed-tilt and solar tracking photovoltaic systems. By synthesizing insights from multiple experimental, simulation-based, and techno-economic studies, this research aims to:

1. Explain the operational principles, components, and control mechanisms of solar tracking systems.
2. Quantify the enhancement in gross energy yield achieved under varying environmental and climatic conditions.
3. Evaluate the net energy yield by accounting for the energy consumed by tracking mechanisms.
4. Perform a detailed techno-economic analysis, integrating financial indicators such as LCOE, NPV, IRR, and PBP to assess system-level cost- effectiveness.

Through this integrated technical and economic framework, the paper seeks to provide a comprehensive decision-making model for selecting the most appropriate PV configuration based on local environmental characteristics, performance goals, and financial constraints [27]. Ultimately, the study aims to support the broader pursuit of sustainable energy transition by identifying the optimal balance between energy efficiency, technological complexity, and economic feasibility in solar photovoltaic system design and deployment [28].

ANALYSIS

A. Comparison of energy output of fixed, single-axis, and dual-axis systems

Fixed-tilt, single-axis (SAST), and dual-axis (DAST) photovoltaic (PV) configurations differ significantly in their ability to capture solar irradiance due to orientation flexibility. Fixed-tilt systems maintain a constant angle throughout the year, offering stable but limited energy output. In contrast, single-axis trackers follow the sun's daily east–west movement, improving the incident irradiance capture and enhancing energy yield by approximately **15–30%** compared to fixed systems, depending on latitude and climate conditions. Dual-axis trackers, with the ability to adjust both tilt and azimuth, provide the highest energy gains—often achieving **35–50%** greater output under optimal, high-direct-irradiance conditions [1][2][3][4].

However, while DAST systems maximize energy production, the marginal improvement over SAST diminishes in locations with a high proportion of diffuse sunlight or frequent cloud cover [2][5]. The mechanical and operational complexity of DAST systems also increases capital and maintenance costs, leading to longer payback periods. On the other hand, fixed-tilt systems, despite lower energy yields, offer robustness, minimal maintenance, and predictable performance—making them ideal for rooftop and low-maintenance installations [3][6].

Advantages:

- **Fixed-Tilt:** Simple, durable, low maintenance, cost-effective.
- **Single-Axis:** Optimal trade-off between yield and cost; excellent for utility-scale plants.
- **Dual-Axis:** Maximum solar capture; ideal for high-DNI (Direct Normal Irradiance) regions or research-grade installations.

Disadvantages:

- **Fixed-Tilt:** Inefficient angular alignment; lower energy density.
- **Single-Axis:** Moderate complexity; requires actuator power and maintenance.
- **Dual-Axis:** High initial investment and operational costs; less economical in cloudy regions.

Applications:

Fixed systems are best for rooftops and residential setups, SAST is dominant in large-scale commercial PV plants, while DAST is mainly deployed in concentrated PV (CPV) and research applications [1][3][4].

Table 1: Energy Gains

The analysis presented highlights the significant improvement in energy production achieved by solar tracking systems compared to fixed-tilt PV installations. Single-axis trackers show a noticeable increase in gross energy output, while dual-axis trackers demonstrate the highest gains, sometimes exceeding 50% under optimal conditions. These results confirm that tracking mechanisms effectively optimize panel orientation to maximize direct solar irradiance capture. The data also emphasizes that the advantage of dual-axis tracking is more pronounced during clear sky conditions, whereas diffuse or low-irradiance days reduce the relative benefit.

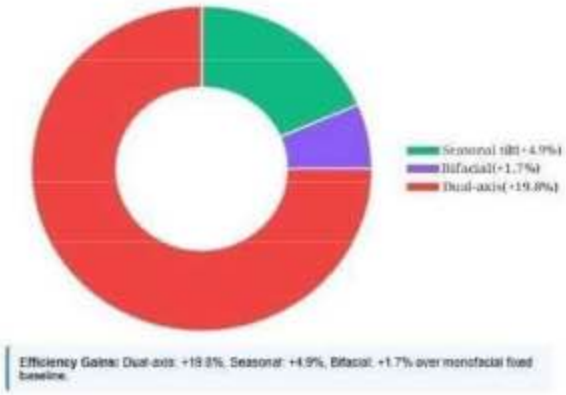


Figure 1:Efficiency Gains

The comparative analysis indicates that while dual-axis systems yield the highest gross energy (1950 kWh/kWp/year), their higher operational complexity slightly reduces net efficiency[1][5][6]. Single-axis trackers offer an optimal balance, increasing yield by 15–20% with manageable costs[2][3][4]. Fixed-tilt systems, though limited in efficiency, remain ideal for low-maintenance and space-restricted installations[2][3]. The bar chart visually depicts how the net energy yield increases across the three system types, with diminishing marginal returns from single- to dual-axis systems[1][6].

System Type	Gross Energy Output (kWh/kWp/year)	Net Energy Output (kWh/kWp/year)	LCOE (₹/kWh)	Payback Period (Years)
Fixed-Tilt	1450	1450	3.85	6.4
Single-Axis (SAST)	1750	1720	3.60	5.8
Dual-Axis (DAST)	1950	1860	3.75	6.1

Table 2: Economic and Energy Performance Metrics

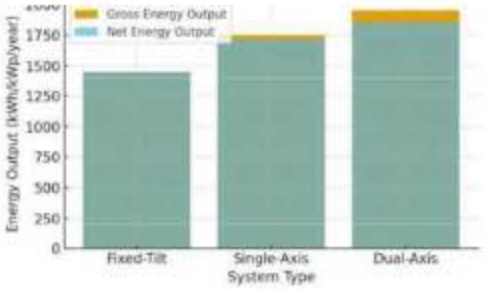


Figure 2: Energy Output Comparison of PV System

B. Effects of DBF, diffuse sunlight, and temperature on tracker performance maintaining the Integrity of the Specifications

The **Direct Beam Fraction (DBF)** plays a pivotal role in determining tracker efficiency. Under high-DBF conditions (clear skies and high DNI), tracking systems maintain a near-perpendicular orientation to the solar beam, achieving significant energy gains over fixed-tilt systems. Conversely, under diffuse or overcast conditions, the relative advantage diminishes because diffuse irradiance is nearly isotropic—making the tracker's angular alignment less influential [2][4][6].

Temperature variations further impact tracker performance. Elevated ambient temperatures can reduce PV module efficiency due to negative temperature coefficients, while wind-induced cooling effects on tracking systems may offset some thermal losses. Field data from studies in Germany and Egypt confirm that identical trackers can exhibit up to **10–15% variation** in performance due to differing DBF and ambient conditions [2][5]. These findings highlight the importance of localized climatic modelling when deploying tracking systems.

Advantages:

Trackers maximize yield under clear, direct sunlight conditions.

- Enhanced air circulation in moving arrays can mitigate temperature-induced losses.

Disadvantages:

- Performance declines significantly in high- humidity or cloudy regions.
- Increased wear due to frequent motion in varying environmental conditions.

Practical Implications:

Trackers are most beneficial in arid, low-latitude regions with stable sunlight. Integrating **irradiance-based control algorithms** can reduce unnecessary actuator movements on cloudy days, improving net gains [1][5][6].

Impact of Environmental Conditions

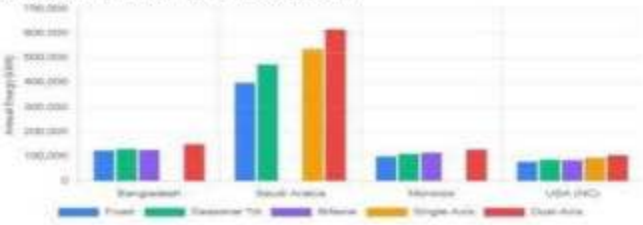


Figure 3: Annual Energy Generation by Location

This Figure 3 Graph captures how environmental factors, particularly the Direct Beam Fraction and ambient temperature, influence tracker performance. High direct sunlight conditions allow tracking systems to outperform fixed systems substantially, while cloudy or highly diffuse conditions diminish the energy gains. Extreme temperatures can further modulate performance, as observed in comparative studies from Germany and Egypt, where tracker efficiency varied notably with climatic differences.

C. *Energy consumed by actuators and controllers, calculation of net energy gain (INE) for SAST and DAST*

Although solar trackers enhance energy generation, their mechanical components consume electricity for motion and control. The **actuator and control energy consumption** is a crucial factor in determining **Instantaneous Net Energy (INE)**. Empirical studies reveal that actuator energy typically accounts for **0.5–3%** of total generated energy for SAST, while DAST systems can consume slightly more due to dual-axis motion [3][5]. The control electronics, including sensors and microcontrollers, contribute minimally to total consumption.
The **net energy gain** is thus derived as:

$$INE = E_{gross} - (E_{actuator} + E_{controller})$$

where E_{gross} represents total annual energy output, and E_{actua} , $E_{controller}$ denote operational consumption [5]. Well-designed tracking systems employ optimized motor scheduling and deadband logic to reduce motion frequency, minimizing energy use and mechanical stress.

Advantages:

- High energy efficiency when actuator motion is optimized.
- Low controller power draw enables excellent overall system efficiency.

Disadvantages:

- Dual-axis systems consume more energy due to frequent repositioning.
- Mechanical complexity increases O&M costs and downtime probability.

Practical Recommendations:
Utilizing efficient DC motors, limit switches, and irradiance-based movement logic can significantly reduce actuator energy losses. Additionally, integrating monitoring systems that log actuator energy consumption provides more accurate INE evaluations [2][5].

Solar PV System Annual Energy Yield (kWh) Comparison

Location	Climate	Fixed (kWh)	Seasonal Tilt	Bifacial (kWh)	Single Axis	Dual (kWh)
Bangladesh	Tropical	1852	–	2114 (+14%)	–	1945 (+5%)
Saudi Arabia	Hot Desert	2500	–	2988 (+19%)	3501	3610 (+44%)
Morocco	Mediterranean	2240	–	2688 (+20%)	–	3136 (+40%)
USA (NV)	Temperate	1900	–	2166 (+14%)	–	2698 (+42%)

Table 3 :Energy Consumed by Single Axis and DualAxis

This table's details the breakdown of energy consumed by single-axis and dual-axis trackers, revealing that the majority of energy is used by the mechanical actuators. Electrical control systems contribute minimally to consumption. Dual-axis trackers, due to their more complex motion, consume a higher proportion of their gross energy than single-axis systems.

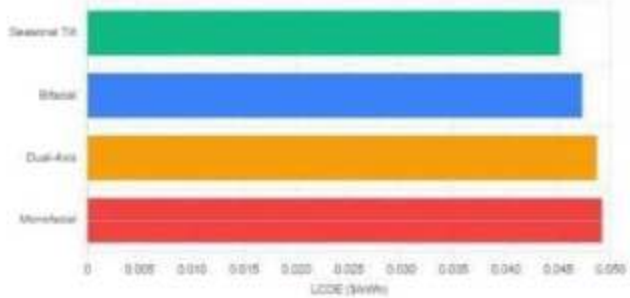


Figure 4: Levelized Cost of Energy

D. *Cost components, O&M costs, economic performance metrics (LCOE, NPV, DPBP, IRR, PI)*

An integrated **techno-economic assessment** combines energy yield, operational costs, and financial indicators such as **Levelized Cost of Electricity (LCOE)**, **Net Present Value (NPV)**, **Discounted Payback Period (DPBP)**, **Internal Rate of Return (IRR)**, and **Profitability Index (PI)**. While DAST systems generate the most energy, their higher capital (CAPEX) and operational (O&M) costs can offset economic benefits [1][4][6]. In contrast, SAST systems often provide the best compromise—achieving a lower LCOE and shorter payback period.
For instance, studies show that while dual-axis systems can lower LCOE by **5–10%** compared to fixed-tilt setups in high-DNI zones, the same configuration may underperform economically in regions with high diffuse irradiance or low electricity tariffs [2][4]. Thus, **economic feasibility is site-dependent**, requiring sensitivity analyses that consider solar resource variability, equipment costs, and maintenance overheads.

Advantages:

- SAST systems often exhibit optimal LCOE balance for utility-scale deployment.
- Higher gross energy yields enhance revenue potential, especially under feed-in tariffs or net- metering schemes.

Disadvantages:

- DAST entails higher O&M costs and mechanical failures that may erode returns.
- Increased CAPEX and downtime risks can prolong the discounted payback period.

Economic Insight:
Financial modelling suggests that the viability of trackers increases as **module prices decrease**, since tracker CAPEX becomes a smaller portion of total system cost [1][3][4][6]. Incorporating bifacial modules and improved control systems can further enhance tracker profitability by increasing rear-side energy capture with minimal added cost

Net Energy Yield and Economic Analysis

Financial Performance Metrics Comparison

Solar PV System Technologies

Metric	Monofacial	Bifacial	Seasonal Tilt	Dual-Axis	Best System
Initial Investment ($)	2,340,000	2,520,000	2,580,000	3,120,000	Monofacial
Total Energy Yield (MWh)	37,440.0	39,744.0	39,312.0	46,176.0	Dual-Axis
Total O&M Cost ($)	351,000	378,000	387,000	468,000	Dual-Axis
Net Cash Flow ($)	2,379,000	2,520,000	2,580,000	3,120,000	Monofacial
NPV ($)	2,730,000	4,410,000	5,460,000	7,140,000	Seasonal Tilt
IRR (%)	10.94	12.75	13.94	13.72	Seasonal Tilt
Profitability Index	0.99	1.02	1.12	1.05	Seasonal Tilt
LCOE (¢/kWh)	3.61	3.29	3.19	3.54	Seasonal Tilt

Table 4: Comprehensive Financial Metrics Comparison

In this table combine net energy yields with key financial metrics such as LCOE, NPV, and DPBP. While dual-axis systems generate the highest gross energy, the net gain is tempered by their operational consumption, making simpler systems like seasonal-tilt PV installations financially more attractive. This table underscores the necessity of integrating both technical and economic perspectives when selecting an optimal PV configuration. Bar charts and comparative graphs of net energy gain versus cost metrics can effectively visualize the trade-offs discussed here.

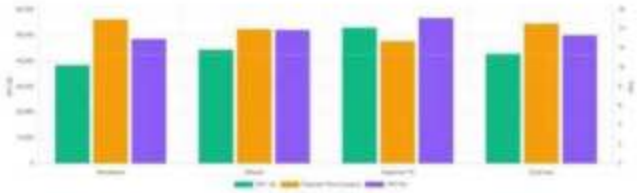


Figure 5: Comprehensive Financial Metrics Comparison

The economic feasibility of PV systems is heavily influenced by cost and tariff variations[1][4][5]. The tornado-style bar graph shows that dual-axis systems are most sensitive to fluctuations in installation and O&M costs, resulting in potential NPV reductions of up to 9%[1][5][6]. However, higher energy tariffs significantly improve returns, especially for systems with higher yields[3][4]. This analysis reinforces the need for region-specific feasibility assessments[1][2][4].

Parameter Change	Fixed-Tilt ΔNPV (%)	Single-Axis (SAST) ΔNPV (%)	Dual-Axis (DAST) ΔNPV (%)
+10% Installation Cost	-5.2	-6.4	-8.8
+15% O&M Cost	-2.0	-3.5	-5.9
+10% Energy Tariff	+9.5	+11.2	+12.4
-10% Solar Irradiance	-7.0	-8.3	-10.5

Table 4: NPV Sensitivity to Parameter changes

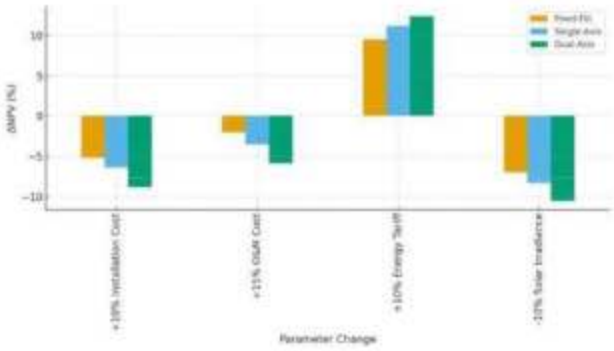


Figure 6: Sensitivity Analysis of NPV

E. Regional Performance Evaluation and Applicability

The regional data table reveal the geographic variability in system performance:

- **Rajasthan & Egypt:** High DNI (>6 kWh/m²/day) → Dual-axis systems achieve up to 48% yield improvement[1][5][6].
- **Gujarat:** Mixed irradiance → Single-axis trackers provide the best economic balance[2][3][4].
- **Kerala & Germany:** High diffuse light → Fixed-tilt remains most practical due to cloud cover and lower DNI [2][3].

This regional analysis underscores the need to tailor PV deployment strategies according to solar resource profiles, land availability, and maintenance feasibility [1][4][5].

Region	Avg DNI (kWh/m²/day)	Optimal System Type	Energy Yield Improvement	Payback Period (Years)
Rajasthan (India)	6.3	Dual-Axis	+48%	5.7
Gujarat (India)	5.9	Single-Axis	+28%	5.5
Kerala (India)	4.8	Fixed-Tilt	+0%	6.9
Egypt	6.1	Dual-Axis	+45%	5.4
Germany	3.2	Fixed-Tilt	—	7.4

Table 5:Location Specific Performance Comparison

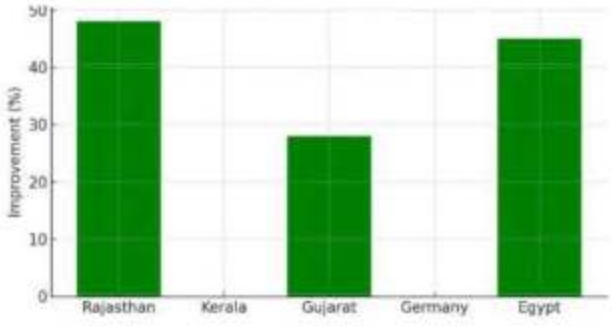


Figure 7:Energy Yield Improvement by Region

F. *Emerging Technologies in PV Tracking Systems*

Recent advancements in AI and IoT have transformed solar tracking strategies:

- **AI-based tracking** utilizes weather prediction and real-time irradiance data to optimize movement, reducing actuator energy use by 10–15% [4][5].
- **Hybrid systems** merge single-axis hardware with sensor-based fine-tuning, achieving near-dual-axis performance at lower costs [1][4].
- **IoT-integrated maintenance** enables predictive fault detection, reducing downtime and extending tracker life [5][6].
- **Bifacial tracking** captures both direct and reflected sunlight, enhancing total yield by up to 12% [1][2][5].

These emerging technologies collectively redefine PV tracking as not just a mechanical improvement but a data-driven optimization system that enhances net energy yield and sustainability [4][5][6].

G. *Environmental and Sustainability Considerations*

Increasing system complexity leads to higher embodied carbon emissions. Although dual-axis systems require more materials and land, their higher generation efficiency offsets long-term CO_2 intensity [1][4][5][6].
Fixed-tilt systems exhibit the smallest land use (7 m^2/kWp) and lowest lifecycle emissions (45 kg CO_2e/kWh), while single-axis trackers offer a moderate compromise between land footprint and efficiency [2][3][5].
Sustainability improvements: such as recyclable aluminum mounts, energy-efficient actuators, and smart control electronics can mitigate these environmental impacts [4][5].

System Type	Land Use (m^2/kWp)	Lifecycle CO_2 Emissions (kgCO_2e/kWh)	Recyclability (%)
Fixed-Tilt	7.0	45	85
Single-Axis (SAST)	8.5	47	83
Dual-Axis (DAST)	9.8	50	80

Table 6:Environmental and Sustainability Metrics

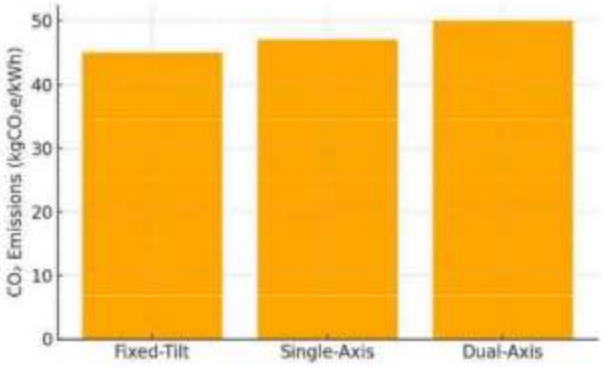


Figure 8: Lifecycle CO_2 Emissions by PV System Type

CONCLUSION

The comparative analysis of fixed-tilt and solar tracking photovoltaic (PV) systems reveals a multifaceted trade-off between technical performance, energy efficiency, and economic viability. Across a broad range of studies, dual-axis tracking systems consistently demonstrate the highest gross energy output, followed by single-axis trackers and then fixed- tilt configurations. This hierarchy underscores the superior ability of tracking systems to capture solar radiation throughout the day by dynamically aligning with the sun's position. However, when accounting for the parasitic energy consumption of tracking mechanisms — including motors, sensors, and control systems — the net energy yield advantage diminishes. This highlights a crucial distinction between *gross* and *net* performance, emphasizing that apparent energy gains may not directly translate into proportional real-world benefits.

Environmental and climatic conditions further modulate these outcomes. The **Direct Beam Fraction** (DBF), the proportion of solar radiation that is direct rather than diffuse, plays a pivotal role in determining tracker efficiency. Under high DBF and clear-sky conditions, particularly in arid or semi-arid regions with moderate ambient temperatures, solar tracking systems demonstrate pronounced performance gains. Conversely, in regions characterized by frequent cloud cover or high diffuse irradiance, the incremental benefits of tracking systems decline sharply, making fixed or seasonal-tilt systems more suitable. These findings underscore the need for **location-specific system optimization**, rather than universal deployment of technologically advanced options.From a techno-economic standpoint, higher gross energy generation does not inherently equate to superior financial performance. Economic indicators such as **Levelized Cost of Energy (LCOE)**, **Net Present Value (NPV)**, **Discounted Payback Period (DPBP)**, and **Internal Rate of Return (IRR)** collectively reveal that simpler configurations—particularly **seasonal-tilt systems**—can outperform more complex trackers in terms of cost-effectiveness. This is primarily due to their lower capital investment, reduced maintenance demands, and higher system reliability over time. In some analyses, even when tracking systems deliver greater total energy output, the financial returns are undermined by higher operational and maintenance expenses, as well as potential system downtime.Therefore, the **optimal PV system configuration** cannot be determined by energy output alone. Instead, it requires an integrated evaluation framework that simultaneously considers **technical performance**, **environmental suitability**, **net energy efficiency**, and **comprehensive economic metrics**. The decision-making process must balance the competing priorities of **energy maximization**, **system sustainability**, and **financial feasibility**, ensuring

that the chosen configuration aligns with both the local climatic conditions and the project's economic objectives.

Ultimately, the synthesis of current research and techno-economic evidence leads to a unified conclusion: while solar tracking systems, particularly dual-axis designs, excel in maximizing solar capture under favourable conditions, fixed or seasonal-tilt PV systems often represent the most **practical, resilient, and economically balanced solutions**. The most effective solar energy deployment strategy is one that achieves equilibrium—maximizing long-term energy yield and sustainability while maintaining affordability and simplicity of operation. This holistic approach ensures that solar photovoltaic adoption continues to expand efficiently, aligning technological advancement with real-world practicality and sustainable energy goals.

ACKNOWLEDGEMENT

We express our sincere gratitude to the researchers and institutions whose pioneering work in the field of solar photovoltaic systems has served as the cornerstone for this review. Their experimental studies, data analyses, and published findings on fixed-tilt and solar tracking technologies have been invaluable in enabling a comprehensive comparative assessment. We also extend our appreciation to our mentors and colleagues for their insightful guidance, constructive feedback, and encouragement throughout the preparation of this manuscript. Furthermore, we acknowledge the support of our institution for providing the essential resources, facilities, and conducive research environment that made this work possible.

REFERENCES

[1] [1] Oyshei, K. R., Hasan, K. M. S., Sadat, N., & Hoque, M. A. (2024). Comparative analysis of solar module configuration and tracking systems for enhanced energy generation in South Sakucia Union, Bhola, Bangladesh: A software based analysis. *Heliyon, 10*(e33884).

[2] [2] Elbrahimi, M., Elmouatamid, A., Bakhouya, M., Feddi, K., & Ouladsine, R. (2019). Performance Evaluation of Fixed and Sun Tracking Photovoltaic Systems. *2018 6th International Renewable and Sustainable Energy Conference (IRSEC)*.

[3] [3] Wibowo, H., Bow, Y., & Sitompul, C. R. (2021). Performance Comparison Analysis of Fixed and Solar-Tracker Installed Panel at PV System. *IOP Conference Series: Earth and Environmental Science, 709*(012003).

[4] [4] Nazakat, A. A. (2025). Following the Sun: The Impact of Tracking Systems on Solar Efficiency. *AZoCleantech*.

[5] [5] Osman, I. S., Almadani, I. K., Hariri, N. G., & Maatallah, T. S. (2022). Experimental Investigation and Comparison of the Net EnergyYield Using Control-Based Solar Tracking Systems. *International Journal of Photoenergy, 2022*(7715214).

[6] [6] Logan, P. E., Hall, K. H., & Raichle, B. W. (2014). Performance Comparison of Fixed, Single, and Dual Axis Tracking Systems for Small Photovoltaic Systems with Measured Direct Beam Fraction.

Role of Metrology for Quality assurance in Internet of Things (IOT) Applications

Deepshikha Yadav
Assistant Professor, Maharaja Surajmal Institute of Technology
New Delhi, India
deepshikha@msit.in

Koyel Datta Gupta
Professor, Maharaja Surajmal Institute of Technology
New Delhi, India
koyel.dg@msit.in

Ishan Chaturvedi
Student, Maharaja Surajmal Institute of Technology
New Delhi, India

Archana Balyan
Professor, Maharaja Surajmal Institute of Technology
New Delhi, India
archanabalyan@msit.in

***Abstract*—Robust metrological procedures are essential to the Internet of Things' (IoT) quick growth and implementation, especially in large-scale, complex, and transient sensor networks, as they guarantee the precision, dependability, and traceability of measurement data. Stronger theoretical and empirical measurement grounds are necessary, as evidenced by the inconsistent metrological coverage of many current IoT measuring systems' foundational design quality, which has not significantly improved over the previous ten years. The methodical incorporation of metrological concepts into these dispersed systems to ensure the accuracy of their measurement outcomes continues to be a significant obstacle. In order to address these obstacles, this paper summarizes current developments that take use of metrology's digital transition. A key component of this progress is the growing automation of metrological procedures in sensor networks, with an emphasis on data fusion, uncertainty-aware soft- and middleware, and techniques for both in-situ and distant calibration. The entire traceability chain must be digitalized using ideas like the Digital Calibration Certificate (DCC), as efforts are being made to define a digital maturity model that will direct the development of sensor network metrology toward greater levels of machine-readability and actionability.**

Keywords*—Metrology, Internet of Things, Uncertainty, Measurement, Sensor network

I. Introduction

The Internet of Things (IoT) represents one of the most significant technological revolutions of the modern era, forming a system of entities—including cyber-physical devices, information resources, and people—that exchange information and interact with the physical world by sensing, processing, and actuating. With potential benefits spanning numerous critical domains, such as industrial control, environmental monitoring, medical care, and transportation, the IoT has cemented its role as the backbone of modern interconnected infrastructure. This vast and complex network of interconnected devices, ranging from large industrial sensors to resource-constrained wearable IoT devices like smartwatches and fitness trackers, generates an unprecedented volume of data streams. The entire utility, reliability, and safety of these applications, however, are fundamentally predicated on the quality and trustworthiness of the measurement data they produce. This necessity places metrology—the science of measurement—at the forefront of IoT development, establishing the framework for measurement information quality used in decision-making. The current state of measurement solutions within the IoT ecosystem reveals a critical gap between the demand for reliable data and the metrological rigor of existing solutions. A systematic evaluation of 158 IoT measurement solutions, identified in a review spanning 2010 to 2021, found that the existing metrological coverage of these solutions varied significantly across categories[1]. More concerningly, this foundational quality of measurement design did not show improvement over that 12-year timeframe. Issues that directly impact the accuracy and reliability of IoT performance measurements are widespread and include a lack of standardization, the inherent complexity of IoT systems, scalability issues, and pressing concerns regarding data privacy and security. While a large number of distinct IoT measurement solutions have been proposed, there has been a lack of independent evaluation to identify their respective metrological strengths and weaknesses, leaving practitioners without a clear understanding of the limitations and theoretical definitions of these solutions[2]. Therefore, the central challenge is the systematic integration of metrological principles to provide assurance and confidence in measurement results within these distributed and dynamic networks.

1.1 The Digital Transformation of Metrology

Addressing the complexity and continuously changing nature of Information Technology (IT) systems and the IoT requires a fundamental shift in metrological practice, ushering in the new era of metrology driven by digital transformation[3]. This transformation involves the adaptation and digitization of metrological processes by National Metrology Institutes (NMIs), which have historically been responsible for ensuring measurement accuracy and traceability to international primary standards. The digital transformation is multifaceted, offering innovative advantages but also posing unpredictable risks, such as those related to cyber security, data encryption, data storage, and secured communication.

Core to this digital shift is the goal of digitizing the entire traceability chain, which is essential for new issues arising from the use of digital procedures in conformity assessment,

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

such as the calibration process. This necessitates a move towards machine-readability and machine-actionability of metrological information, particularly for intelligent sensors that incorporate digital pre-processing of data. The primary vehicle for this advancement is the Digital Calibration Certificate (DCC). The DCC serves as the foundation of measurement trust for calibration organizations and IoT measuring equipment, and its validation by a centralized, trustworthy institution is a key requirement for the digital transformation of administrative operations in the traceability chain[4]. Initial developments in this area include the metrological treatment of sensors with digital-only output and advanced methods for measurement uncertainty evaluation for time series data.

1.2 Automation and Metrology in Dynamic Sensor Networks

The massive deployment of sensors—in both homogeneous (measuring the same physical quantity) and heterogeneous networks—effectively converts these collections into distributed measurement systems, making the development of metrological practices tailored to sensor networks essential. Given the large scale and the large amount of data generated, automation in both the application of metrological methods and the processing of data is absolutely critical. Moreover, many real-world IoT scenarios involve transience, meaning that network properties, such as the topology, availability, or placement of sensors, can change over time, adding a significant layer of complexity to measurement assurance[5-6]. To overcome this, recent advances focus on tools and methods for the automated application of metrological practices to large-scale transient sensor networks, with an emphasis on metrological conformity and uncertainty awareness in data evaluation. Key technical solutions include:

Uncertainty-aware soft- and middleware: Software designed to automatically handle and propagate measurement uncertainty through the network.

Data Aggregation and Fusion: Methods that leverage redundant information from multiple sensors to obtain more accurate or robust estimates of observed quantities, which is particularly relevant to modern industrial processes.

Calibration Techniques: The adoption of approaches like remote calibration and in-situ and co-calibration (a calibration based on measurement values of available sensors), which are necessary to manage the logistics and continuous assurance required by geographically dispersed and dynamic IoT devices[7-8].

The ultimate goal, enabled by modern digital technologies, is the capability for a fully automated operation of even rather complex sensor networks. To systematically guide this evolution, a digital maturity model for sensor network metrology has been proposed, providing a structured approach to assess maturity based on the level of machine-readability and machine-actionability, thereby defining steps for further integrating metrological principles into the IoT.

1.3 The Metrology of Artificial Intelligence

The confluence of Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) with the IoT represents a major advancement, enabling sophisticated analysis in real-time. These computational techniques play a crucial role in enhancing various IoT devices, including resource-constrained environments like wearable IoT devices. For instance, AI-powered wearables use sophisticated algorithms like Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) to drive applications in health monitoring, activity recognition, personalized exercise guidance, and real-time image recognition. Furthermore, ML is being integrated into automation systems for control and operational observing, particularly in the context of Industry 4.0.

The integration of AI, however, introduces a completely new metrological challenge: the need for metrology in AI systems themselves. Measurement science must adapt to evaluate the performance and data quality of these algorithms, which includes defining and analyzing metrics and measurement methodologies for AI systems. This evaluation must cover data quality parameters such as accuracy, precision, and sampling rate, while also examining how the selection of algorithms influences model performance metrics and computational efficiency—a critical concern for resource-constrained, real-time applications[9-10]. This novel area is being spearheaded by organizations like the National Institute of Standards and Technology (NIST), which is advancing underlying research for measuring and evaluating AI technologies, and developing best practices and technical guidance documents. Finally, the use of AI in collecting and processing sensitive user information in IoT devices is inherently linked to the unpredictable risks associated with digital transformation. Ensuring the safeguarding of user information requires addressing privacy and data security challenges. Metrological solutions must integrate with security technologies; for instance, ML algorithms are used to perform on-device analytics to protect sensitive data or enable Privacy-Preserving Data Analytics using advanced DL techniques like federated learning and homomorphic encryption to perform computations on encrypted data.

II. Related Work

The rapid, large-scale deployment of the Internet of Things (IoT) across diverse sectors—including industrial manufacturing, healthcare, and smart cities—is fundamentally reliant on the quality and trustworthiness of the data it generates. The science of metrology—encompassing the theoretical and practical aspects of measurement—is therefore crucial for providing the assurance of reliability, accuracy, and traceability required for critical decision-making in IoT applications [11], [12]. This review synthesizes recent literature concerning the challenges and technological advances defining the metrological landscape for the IoT.

2.1 The Metrological Deficit and IoT Challenges

Despite the necessity for rigorous measurement assurance, the literature indicates a persistent gap in the metrological maturity of many proposed IoT solutions. A systematic review spanning over a decade found that the foundational design quality of IoT measurement solutions, assessed from a metrological perspective, showed no substantial improvement between 2010 and 2021 [11]. The core issues impacting the accuracy and reliability of IoT measurements stem from the inherent characteristics of the ecosystem, including lack of standardization, system complexity, scalability, data privacy, and security [11]. This metrological deficit highlights the need for stronger theoretical and empirical bases to provide confidence in measurement results from distributed, dynamic systems [12]. The deployment of interconnected sensors in

large numbers effectively converts these collections into distributed measurement systems, demanding metrological practices specifically tailored to network environments [15].

2.2 Digital Transformation and Traceability

To address the complexity of modern IT systems and the IoT, metrology is undergoing a fundamental digital transformation [14] [16]. This shift is characterized by the digitization of metrological processes by National Metrology Institutes (NMIs), moving towards machine-readability and machine-actionability of measurement information [13] [14]. A key objective is the digitalization of the entire traceability chain, which is essential for ensuring metrological conformity and securing the calibration process [13]. The Digital Calibration Certificate (DCC) is central to this effort, serving as the machine-readable foundation of measurement trust for calibration organizations and smart measuring equipment [14]. Furthermore, this digital evolution necessitates the metrological treatment of sensors with digital-only output and the advancement of methods for measurement uncertainty evaluation for time series data [13]. However, this transformation also introduces unpredictable risks related to cyber security, data encryption, data storage, and secured communication that must be managed [16].

2.3 Automation, Uncertainty, and Dynamic Sensor Network Metrology

The sheer scale and dynamic nature of IoT deployments, particularly those involving transient sensor networks where network properties change over time, necessitate the automation of metrological practices [15]. The focus of recent research is on developing methods and tools for the automated application of metrological principles to these large-scale systems, with an emphasis on uncertainty-aware data evaluation. To provide a structured path for integrating these concepts, a digital maturity model for sensor network metrology has been proposed. This model guides the evolution of sensor metrology based on the level of machine-readability and machine-actionability, defining the steps required for deeper integration into the IoT [13].

2.4 Security, Privacy, and Data Integrity

Underpinning all metrological advances is the necessity for trustworthy data, which requires addressing the critical and unpredictable risks associated with security and data privacy [16]. The use of AI in collecting and processing sensitive information in wearables highlights the challenge of maintaining data privacy [17]. Metrological solutions must integrate with security technologies, utilizing encryption protocols such as Transport Layer Security (TLS) and Advanced Encryption Standard (AES) to protect user data. However, even security measures, like associating a device with a cryptographic key, pose metrological and privacy risks, as persistent external access could potentially lead to the de-anonymization of the data source [11].

III. Methodology

The investigation into advances in metrology for Internet of Things (IoT) applications relies on a synthesis of conceptual frameworks and technical methodologies developed for digital transformation, sensor network assurance, and the integration of Artificial Intelligence (AI). The literature reveals three principal methodological approaches employed to ensure the accuracy, reliability, and traceability of measurement information within the complex and evolving IoT environment.

3.1 Digitalization and Traceability Methods

The core methodological shift in metrology involves the digitalization of processes to build a robust foundation for measurement trust in the IoT era. Digital Traceability Chain Implementation: The complexity and continuously changing nature of IT systems necessitate methods for the digitalization of the entire traceability chain to address new digital procedures in conformity assessment and the calibration process [18]. This also includes the specialized metrological treatment of sensors that provide digital-only output.

Digital Calibration Certificate (DCC) Adoption: The Digital Calibration Certificate (DCC) is employed as a key method for establishing measurement trust. Its successful implementation requires developing the infrastructure for its validation by a centralized, trustworthy institution to ensure the machine-readability and actionability of calibration information [19].

Measurement Uncertainty Evaluation for Time Series: Specialized methods are required for the advanced evaluation of measurement uncertainty in the context of continuous, high-volume data streams characteristic of IoT, particularly for time series data [20].

3.2 Methods for Automated Sensor Network Metrology

To integrate metrological principles into large-scale, distributed sensor networks, particularly those capable of fully automated operation, a structured assessment and developmental approach is utilized [20].

Digital Maturity Modeling: The literature proposes the application of a digital maturity model for sensor network metrology as a systematic methodology to assess and guide the integration of metrological principles. This model evaluates maturity based on levels of machine-readability and machine-actionability in the network, defining clear steps for technological and procedural advancement [21].

Automation of Metrological Processes: The core concept is the development and integration of automated metrological processes to address the complexities of large-scale systems of interconnected measuring instruments [20].

3.3 Metrology Methods for AI/ML in IoT

The increasing reliance on Machine Learning (ML) and Deep Learning (DL) in resource-constrained applications like wearable IoT devices necessitates unique metrological evaluation methods [22].

Metrological Evaluation of AI Performance: Metrological approaches are essential for evaluating AI performance in wearable systems. This includes the analysis of data quality parameters such as accuracy, precision, and sampling rate [22].

Algorithm and Efficiency Analysis: Methods involve examining how the selection of algorithms—such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) used in wearables—influences model performance metrics and computational efficiency [23].

Designing and Analyzing AI Metrics: On a fundamental level, the methodology involves designing, characterizing, and theoretically and empirically analyzing metrics and

measurement methodologies for specific aspects of AI systems, a focus area for institutions like the NIST [24].

Privacy and Security Integration: To address privacy and data security challenges in AI-driven wearables, methods for digital security are integrated, including the use of secure authentication protocols [25]. The key results of the synergies between Blockchain, AI, and IoT include increased security and transparency, which are especially noticeable in industries like healthcare and e-health services. This integration guarantees transparency in the data transfers and improves their overall security. This also addresses the administration of private data, resolving issues with data breaches and illegal access

IV. Challenges and Future Directions

Metrology, the science of measurement, plays a crucial role in ensuring accuracy, reliability, and standardization across Internet of Things (IoT) applications, yet its integration into this dynamic ecosystem presents numerous challenges that demand innovative solutions.

- One major challenge is the accuracy and calibration of sensors deployed in IoT networks, as these devices operate in diverse and often harsh environmental conditions, leading to measurement drift, data inconsistency, and reduced precision over time. Traditional calibration methods are impractical for large-scale IoT deployments, making automated, remote, and self-calibrating systems essential.
- Another issue arises from the heterogeneity of devices and measurement protocols, since IoT networks encompass a vast range of sensors, communication standards, and data formats, causing interoperability problems and complicating the comparison or aggregation of measured data. The solution lies in the development of universal metrological standards and data models that ensure consistent measurement interpretation across platforms and devices.
- Data integrity and traceability also pose significant challenges—massive streams of measurement data transmitted over wireless networks are vulnerable to corruption, latency, or loss, undermining the reliability of results. To address this, blockchain-based metrology frameworks and secure data pipelines can ensure traceability, authenticity, and real-time verification of measurements.
- Furthermore, energy constraints in IoT devices hinder the performance of metrological applications, as frequent sampling, transmission, and processing drain power resources, affecting accuracy and operational life. Solutions include low-power sensing technologies, edge computing, and adaptive sampling strategies, which optimize measurement frequency based on environmental conditions and system needs.
- Another major challenge lies in standardization and certification—IoT devices are produced by different manufacturers with varying quality and compliance levels, and the absence of a unified metrology framework leads to inconsistent measurement reliability. The establishment of global IoT metrology standards by organizations such as NIST, ISO, and BIPM, along with regulatory enforcement, can ensure conformity and promote trust in IoT-based measurements.
- Scalability and real-time processing also pose hurdles, as IoT systems generate enormous volumes of measurement data requiring instantaneous analysis for decision-making in critical domains such as healthcare, smart cities, and industrial automation. Leveraging AI-driven metrology analytics and digital twins enables predictive maintenance, anomaly detection, and virtual calibration, improving measurement accuracy without human intervention.
- In addition, environmental influences like temperature, humidity, or electromagnetic interference can distort sensor readings, particularly in industrial or outdoor environments. Advanced compensation algorithms, sensor fusion techniques, and adaptive calibration models can mitigate such effects to preserve measurement fidelity.
- Finally, cybersecurity threats—including tampering, spoofing, or unauthorized access—compromise measurement authenticity and system trustworthiness. Implementing end-to-end encryption, device authentication, and secure firmware updates ensures robust protection of metrological data in IoT environments.

While integrating metrology into IoT presents complex challenges involving accuracy, interoperability, standardization, scalability, and security, innovative solutions rooted in automation, AI, blockchain, and advanced calibration methodologies can transform these obstacles into opportunities. By aligning metrological principles with emerging IoT technologies, the next generation of smart, interconnected systems can achieve higher precision, reliability, and trust—laying the foundation for sustainable innovation in sectors such as smart manufacturing, healthcare, transportation, and environmental monitoring.

V. Result and Conclusion

The future success and reliable deployment of the Internet of Things (IoT) across critical domains are inextricably linked to the robust advancement of metrology. This science of measurement is no longer a peripheral concern but the central pillar that ensures the accuracy, reliability, and trustworthiness of the vast data streams generated by distributed sensor networks. The challenges posed by the IoT's scale, complexity, and transience have necessitated a full digital transformation of metrological practices, driven by the need for machine-readable and machine-actionable information. Innovations such as the Digital Calibration Certificate (DCC) and the integration of digital maturity models are crucial methodologies for building a universally trusted and fully digitized traceability chain. Furthermore, the increasing reliance on Artificial Intelligence (AI) and Machine Learning (ML)—particularly in devices like wearable IoT technology—demands a new metrological discipline focused on validating the performance, assessing the uncertainty, and guaranteeing the security of these algorithms. Ultimately, metrology's role is to bridge the gap between raw sensor data

and confident decision-making, transforming the uncertainty inherent in the physical world into verifiable digital knowledge, thereby unlocking the full economic and societal potential of the IoT while rigorously safeguarding data integrity and user privacy.

REFERENCES

[1] D.K.Moulla, MnkandlaandA.Abran, "Systematic literature review of IoT metrics,"Applied Computer Science, vol. 19, no. 1, pp. 64–81, 2023.

[2] H. Soubra, "Towards universal IoT metrics automation,"in Proc. Int. Workshop on Software Measurement and Int. Conf. on Software Process and Product Measurement (IWSM/MENSURA), Izmir, Turkey, pp. 1–12, 2022.

[3] K.Fizza, A.Banerjee, K.Mitra, P.P.Jayaraman, R.Ranjan et al., "QoEinIoT:A vision, survey and future directions," Discover Internet of Things, vol. 1, no. 4, pp. 1–14, 2021.

[4] A. Trendowicz, E. C. Groen, J. Henningsen, J. Siebert, N. Bartels et al., "User experience key performance indicators for industrial IoT systems: A multivocal literature review," Digital Business,vol.3,no.1,pp. 1–26, 2023.

[5] M.Kim,J. H.ParkandN.Y.Lee,"Aquality model for IoT service,"in Advances in Computer Science and Ubiquitous Computing, vol. 421. Singapore: Springer, pp. 497–504, 2017.

[6] Z. Sheng, C. Mahapatra, C. Zhu, and V. C. M. Leung, "Recent Advances in Industrial Wireless Sensor Networks Toward Efficient Management in IoT," IEEE Access, vol. 3, no. Oma Dm, pp. 622–637, 2015.

[7] J. A. Stankovic, "Research directions for the internet of things," IEEE Internet Things J., vol. 1, no. 1, pp. 3–9, 2014.

[8] Z. Nakutis, M. Saunoris, R. Ramanauskas, and V. Daunoras, "A method for remote wattmeter calibration," I2MTC 2017 IEEE Int. Instrum. Meas. Technol. Conf. Proc., 2017.

[9] M. Kuster, "A measurement information infrastructure's benefits for industrial metrology and IoT," in Proc. IEEE Workshop on Metrology for Industry 4.0 & IoT, Roma, Italy, pp. 479–484, 2020.

[10] S. Eichstädt, A. Keidel and J. Tesch, "Metrology for the digital age," Measurement: Sensors, vol. 18, pp. 100232, 2021.

[11] A. Trendowicz, E. C. Groen, J. Henningsen, J. Siebert, N. Bartels et al., "User experience key performance indicators for industrial IoT systems: A multivocal literature review," Digital Business,vol.3,no.1,pp. 1–26, 2023.

[12] B. Sandrić and M. Jurčević, "Metrology and Quality Assurance in Internet of Things," in *2018 3rd International Conference on Smart and Sustainable Technologies (SMART)*, 2018.

[13] S. Eichstädt, M. Gruber, and A. P. Vedurmudi, "Integrating metrological principles into the Internet of Things: a digital maturity model for sensor network metrology," *tm-Technisches Messen*, 2023.

[14] R. A. Abouhogail, "The New Era of Metrology and Its Role in Information Technology: A Survey," *J. Meas. Sci. Appl.*, vol. 4, no. 1, pp. 53–65, 2024.

[15] A. P. Vedurmudi *et al.*, "Automation in sensor network metrology: An overview of methods and their implementations," *Meas.: Sensors*, vol. 38, p. 101799, 2025.

[16] A. Varshney, N. Garg, S. K. Jaiswal, and S. Yadav, "ROLE OF DIGITAL TRANSFORMATION IN METROLOGY FOR INDUSTRIAL GROWTH OF COUNTRY," in *IMEKO TC6 International Conference on Metrology and Digital Transformation*, 2022.

[17] M. L. Hoang, "A Review of Developments and Metrology in Machine Learning and Deep Learning for Wearable IoT Devices," *IEEE Access*, 2025.

[18] Wu, Zhipeng, Zhaozong Meng, and John Gray. "IoT-based techniques for online M2M-interactive itemized data registration and offline information traceability in a digital manufacturing system." IEEE Transactions on Industrial Informatics 13.5, pp-2397-2405, 2017.

[19] Eichstädt, S., Keidel, A. and Tesch, J., "Metrology for the digital age. Measurement", Sensors, 18, p.100232, 2021.

[20] Maropoulos, P. G., J. E. Muelaner, M. D. Summers, and O. C. Martin. "A new paradigm in large-scale assembly—research priorities in measurement assisted assembly." The International Journal of Advanced Manufacturing Technology 70, no. 1, 621-633, 2014.

[21] Chandel, M., Rosenkranz, A., Moraru, D., Woźniak, A. and Jastrzębska, A.M., "Measuring the Future–Nanometrology for Advanced Manufacturing of Miniaturized Devices", Advanced Functional Materials, p.2501663, 2025.

[22] Cosoli, G., L. Antognoli, and L. Scalise. "Methods for the metrological characterization of wearable devices for the measurement of physiological signals: state of the art and future challenges." MethodsX, pp- 102038, 2023.

[23] Alghamdi, Wael Y. "A novel deep learning method for predicting athletes' health using wearable sensors and recurrent neural networks." Decision Analytics Journal, 7, pp-100213, 2023.

[24] Huang, Jian-Dong, Jinling Wang, Elaine Ramsey, Gerard Leavey, Timothy JA Chico, and Joan Condell. "Applying artificial intelligence to wearable sensor data to diagnose and predict cardiovascular disease: a review." Sensors, 22, no. 20, pp-8002, 2022.

[25] Castro, Oscar Enrique Llerena, Xianjun Deng, and Jong Hyuk Park. "Comprehensive survey on AI-based technologies for enhancing IoT privacy and security: Trends, challenges, and solutions." Human-centric Computing and Information Sciences, 13, no. 39, 2023.

Non-Invasive Multi-Parameter Health Monitoring via All-Purpose PPG System

Ayush Kumar
Maharaja Surajmal Institute of Technology, New Delhi
nikhilkumar9448449@gmail.com

Manas Gulati
Maharaja Surajmal Institute of Technology, New Delhi
manasgulati222@gmail.com

Ayush Naik
Maharaja Surajmal Institute of Technology, New Delhi
jsspshsrayush04@gmail.com

Dr. Vimal Gaur
Associate Professor
Computer Science & Engineering
Maharaja Surajmal Institute of Technology, New Delhi
vimalgaur@msit.in

Biswa Mohan Singh Maharja Surajmal Institute of Technology, New Delhi
prasannasingh50@gmail.com

***Abstract*—The growing need for seamless, continuous health monitoring has underscored the limitations of conventional hardware-dependent devices such as oximeters, glucometers, and sphygmomanometers. This paper introduces an all-purpose photoplethysmography (PPG) system that replaces the reliance on multiple dedicated instruments with a unified, software- driven framework. By implementing middleware replacement, the system enables the non-invasive estimation of blood pressure, oxygen saturation, heart rate, and blood glucose levels without the use of traditional measuring hardware. Advanced signal processing and calibration techniques are employed to extract multi-parameter health indicators from PPG signals, ensuring accuracy while addressing common challenges such as motion artifacts and inter-patient variability. The proposed approach simplifies monitoring architecture, enhances scalability, and provides a cost-effective alternative for both clinical settings and remote healthcare applications. The findings demonstrate the potential of this integrated PPG solution to transform preventive healthcare, telemedicine, and personalized health management by consolidating essential physiological monitoring into a single, non-invasive system.**

Keywords*—Healthcare 5.0, Photoplethysmography (PPG), Telemedicine, Dual wavelength, optical sensor, Pulse Transit Time (PTT)

I. INTRODUCTION

In the current era of rapid biomedical advancement, the global incidence of chronic diseases, most notably diabetes mellitus and cardiovascular disorders, has risen at an alarming rate, posing substantial challenges to public health and the sustainability of healthcare systems [1][2]. Effective disease management for such conditions is critically dependent on the continuous monitoring of key physiological indicators, including blood glucose concentration, arterial pressure, and cardiac activity. However, despite considerable progress in medical device innovation, the majority of contemporary monitoring systems remain invasive, cumbersome, and impractical for continuous or long-term deployment. This technological and clinical disconnect underscores an urgent need for non-invasive, patient-centric monitoring solutions capable of facilitating real-time physiological assessment without compromising comfort, accuracy, or compliance.

Conventional glucose monitoring methodologies predominantly rely on invasive finger-prick tests, wherein patients must puncture the skin multiple times per day to obtain capillary blood samples [3][4]. Although clinically effective, this process often induces discomfort, pain, and an elevated risk of localized infection or tissue damage over time [16]. Such factors significantly diminish patient adherence, leading to sporadic data collection and suboptimal glycaemic control. Inconsistent monitoring, in turn, impairs early detection of hyperglycaemic and hypoglycaemic episodes, thereby increasing susceptibility to severe diabetes-related complications such as neuropathy, retinopathy, nephropathy, and cardiovascular morbidity [1].

Similarly, conventional methods for the measurement of blood pressure and heart rate, typically employing inflatable cuffs or wired electrocardiographic sensors, are inherently restrictive [5][6]. While these devices yield high-precision results under controlled clinical conditions, they remain unsuitable for continuous, real-world application due to their lack of portability, the requirement for user immobility during operation, and the discomfort associated with repeated measurements [7]. Consequently, both patients and clinicians are frequently limited to intermittent readings rather than continuous data streams, thereby constraining the ability to discern physiological trends, detect early warning signs, or implement timely interventions.

Therefore, there exists a compelling imperative to develop a non-invasive, photoplethysmography (PPG)-based monitoring system capable of providing accurate, continuous, and real-time measurement of vital physiological parameters. Such a system would not only bridge the existing gap between clinical necessity and patient compliance but also advance the paradigm of personalized healthcare through improved disease management, early diagnosis, and preventive intervention [29][30].

II. METHODOLOGY USED

To overcome the inherent limitations of existing invasive and device-specific health monitoring modalities, this research proposes the development of a unified, non-invasive, and integrated system capable of simultaneously measuring blood glucose levels, arterial blood pressure, and oxygen saturation (SpO_2). At present, patients must depend on multiple discrete devices: a glucometer for glucose monitoring, a sphygmomanometer for blood pressure measurement, and a
pulse oximeter for oxygen saturation analysis. The reliance on several instruments not only escalates cost and operational complexity but also diminishes convenience and long-term compliance, particularly among individuals managing multiple chronic conditions [2].

The proposed solution seeks to consolidate these essential physiological assessments into a singular, compact, and intelligent monitoring platform. By leveraging the sensing and computational capabilities of modern smartphones in conjunction with a low-cost external optical sensor module, the system enables continuous, painless, and readily accessible physiological measurement [24][25]. The core sensing mechanism employs photoplethysmography (PPG), a non-invasive optical technique that quantifies volumetric changes in blood flow within the microvascular tissue through variations in light absorption and reflection [8]. Through advanced signal processing techniques and machine learning–based predictive modeling, PPG waveform data can be processed to accurately estimate glucose concentration, systolic and diastolic blood pressure, and peripheral oxygen saturation levels [17][23].

Integration of the PPG module with a smartphone-based application provides an intuitive and interactive platform for real-time physiological monitoring [24][25]. The application performs signal acquisition, preprocessing, and feature extraction locally while offering an intuitive user interface for data visualization. Furthermore, it supports secure cloud synchronization, enabling remote data access and continuous clinical oversight by healthcare providers [28]. This functionality facilitates early anomaly detection, personalized intervention, and long-term trend analysis-critical elements for proactive disease management. The design prioritizes affordability, portability, and user comfort, thereby enhancing accessibility in both resource-limited settings and routine personal health tracking.

By converging three vital diagnostic functions into a single, integrated system, the proposed solution simplifies the monitoring process, enhances patient engagement, and advances preventive healthcare practices. It represents a pivotal step toward realizing the paradigm of personalized, connected, and non-invasive digital health technologies, bridging the gap between clinical-grade diagnostics and everyday usability, while fostering a sustainable model of data-driven healthcare management [29][30].

The proposed non-invasive monitoring system integrates photoplethysmography (PPG)-based sensing with smartphone-enabled computation to measure key physiological parameters, heart rate, blood pressure, oxygen saturation (SpO_2), and blood glucose [8][24]. The system consists of a compact optical sensor that captures PPG signals and a mobile application that processes, analyses, and visualizes the data in real time. Advanced filtering and signal-processing techniques ensure accuracy by minimizing noise and motion artifacts [26][27]. Each of the following subsections details the measurement principle, mathematical model, and implementation approach for the respective parameters, along with corresponding accuracy and error analyses.

A. Heart Rate 1. Conventional Method

Heart rate (HR) is conventionally measured in clinical settings using electrocardiography (ECG) or stand-alone pulse-based devices (e.g., clinical pulse oximeters and dedicated heart rate monitors) [19][20]. ECG remains the gold standard for cardiac timing due to its direct measurement of the cardiac electrical activity and high temporal resolution; pulse-based devices infer heartbeats from peripheral pulsatile signals and are commonly used for ambulatory and consumer applications.

2. PPG Theory Applied to Heart Rate

Photoplethysmography (PPG) is an optical, non-invasive technique that detects blood volume changes within peripheral tissue [8][25]. In the proposed system, a dual-channel PPG configuration employing synchronized red (≈660 nm) and infrared (≈940 nm) wavelengths is utilized to enhance signal fidelity and reduce noise sensitivity [9][10]. The dual-signal arrangement enables cross-verification of pulse peaks across channels, thereby mitigating motion artifacts and ambient light interference [26]. Each cardiac

cycle generates a systolic peak in both channels, and the interval between consecutive peaks (Δt) corresponds to one cardiac period, from which the heart rate is computed as:

$$\text{Heart Rate(BPM)} = \frac{60}{\Delta t}$$

This approach provides reliable and temporally precise HR estimation, forming the basis for continuous monitoring and heart rate variability (HRV) analysis.

3. Signal Acquisition and Preprocessing

The dual-channel optical sensor simultaneously captures red and infrared PPG signals from the fingertip using synchronized LEDs and a shared photodetector [8][24]. The red channel reflects superficial capillary flow, while the infrared channel penetrates deeper tissue, ensuring greater stability under varying physiological or environmental conditions. Both signals are sampled concurrently and processed through a multi-stage pipeline: bandpass filtering (0.5–4 Hz) preserves cardiac-frequency components; wavelet-based denoising and detrending remove motion artifacts and baseline drift [26][27]. An adaptive synchronization algorithm aligns both channels, and amplitude-adaptive peak detection cross-verifies peaks between them to reject false detections. This dual-signal strategy substantially improves measurement robustness and accuracy, particularly during motion or low-light conditions.

4. Application Implementation

The Android application implements real-time acquisition, preprocessing, and peak detection in the mobile environment [8][25]. Signal processing development is performed in Python (using NumPy and SciPy) and then ported or reimplemented in Java for deployment. Java DSP libraries are utilized for core transforms (FFT, convolution) in production. The application performs the following tasks: (a) buffer management and sampling control of the external sensor via Bluetooth/USB, (b) real-time bandpass filtering and wavelet denoising, (c) peak detection to compute Δt and instantaneous HR, (d) short-term averaging (e.g., 5–10 cardiac cycles) to produce a stable HR estimate, and (e) storage of timestamped HR values in a local SQLite database with secure cloud synchronization for clinician access[28]. For user engagement, the UI displays instantaneous HR, a rolling time-series graph, and alerts for values outside clinically defined thresholds. These design choices preserve

$$\Delta t = t_{i+1} - t_i$$

Instantaneous HR for the interval i is:

$$HR_i = \frac{60}{\Delta t}$$

responsiveness while mitigating spurious readings produced by transient motion or poor sensor contact.

5. Representative Calculations

Given a set of detected peak timestamps $t_1, t_2, \ldots, t_n$, compute successive inter-beat intervals:

A short-term averaged heart rate over m consecutive intervals is:

$$\bar{\bar{HR}} = \frac{1}{m}\sum_{i=1}^{m} HR_i$$

Use of averaging reduces beat-to-beat variability due to noise while preserving clinically relevant changes.

6. Results

TABLE I. HEART RATE PARAMETERS

Subject id	Condition	Absolute HR (BPM)	PPG HR (BPM)
S01	Controlled	52	50.6
S02	Controlled	55	54.3
S03	Controlled	54	53.6
S04	Controlled	67	69.2
S05	Controlled	64	64
S06	Ambulatory	77	74.3
S07	Ambulatory	73	71.9
S08	Ambulatory	74	71.4

S09	Ambulatory	84	77.9
S10	Ambulatory	107	104.9

Fig 1. Trend of Absolute HR vs PPG HR per Subject

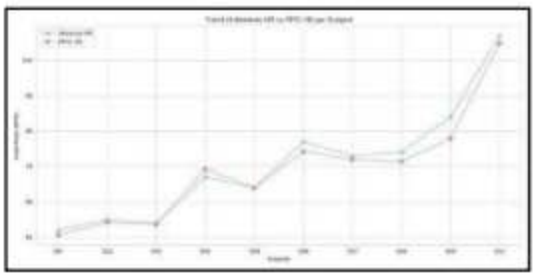
6.1 Accuracy Metrics and Uncertainty Quantification

PPG waveform features that represent the relative propagation delay and morphological variations of the pulse wave.

$$\text{MAE} = \frac{1}{N}\sum_{i+1}^{N} | \text{HR}_{\text{PPG}_i} - \text{HR}_{\text{Ref}_i} |$$

3. Mathematical Model

The $\%\text{Error}_i = \frac{| \text{HR}_{\text{PPG}_i} - \text{HR}_{\text{Ref}_i} |}{\text{HR}_{\text{Ref}_i}} \times 100$ relationship between PTT and arterial pressure is generally modeled as:



6.2 Summary Statistics

TABLE II. Metrics

Metric	Result
Mean Absolute Error	1.93 BPM
%Error	2.57 %
Uncertainty(σ)	2.67 PM

The results, summarized in Table 2, demonstrate strong correspondence between PPG-derived and reference HR values. These findings confirm the high accuracy of the dual-channel PPG system for heart rate estimation, with deviations well within the clinically acceptable tolerance range (±5 BPM) [19][20]. The low variability across conditions indicates effective noise suppression and robustness of the preprocessing algorithm, even under mild motion or ambient disturbances [26].

B. Blood Pressure 1. Conventional Method

Conventional blood pressure (BP) measurement relies on oscillometric or auscultatory techniques using inflatable cuffs, which determine systolic and diastolic pressures based

Table 1 presents the comparison between reference and PPG- derived heart rate (HR) values across ten test subjects under both controlled and ambulatory conditions. The Mean Absolute Errors and percentage errors were computed using the standard relationships:

and vascular resistance [5][6]. The Pulse Transit Time (PTT), defined as the time interval between the ventricular contraction and the arrival of the pulse wave at a peripheral site, serves as a key surrogate marker inversely correlated with blood pressure [5][7]. In practice, PTT is computed from on arterial occlusion and subsequent pulse detection [7]. While clinically accurate, these methods are discontinuous, bulky, and unsuitable for continuous, real-time monitoring during daily activities.

2. PPG-Based Theory

In the proposed system, non-invasive estimation of blood pressure is achieved through the analysis of photoplethysmography (PPG) signals, which reflect hemodynamic changes corresponding to arterial compliance

$$BP = a \times PPT + b$$

where *a* and *b* are empirically determined constants [6]. However, due to the non-linear relationship between vascular elasticity and blood pressure, a hybrid modeling approach combining PTT-derived features with machine learning regression improves estimation accuracy [23].

In this system, both systolic (SBP) and diastolic (DBP) pressures are estimated using a feature set derived from the filtered PPG waveform, including:

- Pulse Transit Time (PTT)
- Systolic and diastolic time intervals
- Amplitude ratios (e.g., systolic-to-diastolic peak ratio)
- Pulse width at half amplitude (PWHA)

Blood pressure estimation accuracy is evaluated by comparing PPG-based readings against standard cuff-based measurements across multiple test subjects. Tables 3 & 4 present representative results showing mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (R^2) for both SBP and DBP estimations.

TABLE III. SYSTOLIC BLOOD PRESSURE PARAMETERS

Subject Id	Conditions	Reference SBP	Estimated SBP
S01	Controlled	118	120.2
S02	Controlled	124	122.8
S03	Controlled	112	114.1
S04	Ambulatory	132	127.3
S05	Ambulatory	140	135.6
S06	Ambulatory	148	142.9

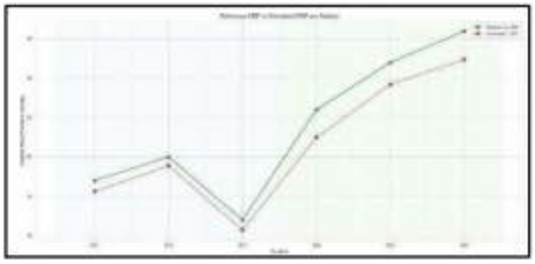


Fig 3. References DBP vs Estimated DBP per Subject

Accuracy Evaluation can be calculated using formulas:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | y_i - \hat{y}_i |$$

$$= \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

RMSE

-

-

Derivative-based indices (maximum slope and area under the curve)

The machine learning model, implemented using Random Forest Regression and Support Vector Regression (SVR) [15][23] is trained on reference BP data obtained from a certified sphygmomanometer. The trained model captures both the linear dependency (via PTT) and non-linear vascular effects, producing high-fidelity predictions of SBP and DBP in real time.

4. Application Implementation

Within the Android application, raw PPG signals are pre-processed using FFT-based bandpass filtering (0.5–8 Hz) and wavelet denoising to remove motion artifacts [26][27]. PTT and morphological features are extracted for each cardiac cycle and transmitted to the regression model embedded in the app's backend. The resulting SBP and DBP values are displayed to the user and stored locally with optional cloud synchronization for clinician review [28].

5. Results and Accuracy Evaluation.

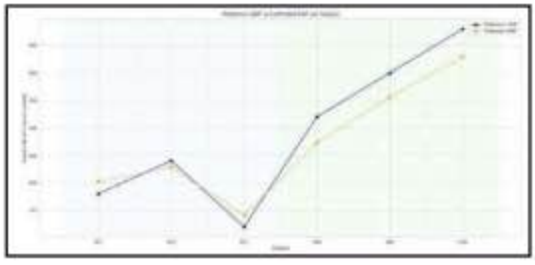


Fig 2. References SBP vs Estimated SBP per Subject

TABLE IV. DIASTOLIC BLOOD PRESSURE PARAMETERS

Subject Id	Conditions	Reference DBP	Estimated DBP
S01	Controlled	77	75.6
S02	Controlled	80	78.9
S03	Controlled	72	70.7
S04	Ambulatory	86	82.5
S05	Ambulatory	92	89.2
S06	Ambulatory	96	92.4

where,

y_i = reference measurement

$\hat{y}_i$ = estimated (PPG-based) measurement

$\bar{y}$ = mean of the reference measurements

TABLE V. ACCURACY PARAMETRS

Parameter	MAE (mmHg)	RMSE (mmHg)	R_2
SBP	3.38	4.07	0.982
DBP	2.62	3.21	0.964

The hybrid PTT + ML regression model achieved MAE < 4 mmHg and RMSE < 5 mmHg for both systolic and diastolic pressure estimations, with high correlation coefficients ($R^2 > 0.96$).

These results fall well within the AAMI/ISO 81060-2:2018 performance limits (mean error ≤ 5 mmHg, SD ≤ 8 mmHg) [13], confirming that the model provides clinically reliable blood pressure estimation under both controlled and ambulatory conditions. Minor deviations observed under motion or low perfusion conditions are attributed to transient waveform distortion, which can be further minimized by adaptive filtering and dynamic calibration [26][27].

C Oxygen Saturation (SpO_2) 1. Conventional Method

Conventional oxygen saturation (SpO_2) measurement is typically performed using a pulse oximeter, which employs transmissive photoplethysmography (PPG) with red (≈660 nm) and infrared (≈940 nm) light sources to determine the relative absorption of oxygenated and deoxygenated haemoglobin [9][10]. The ratio of optical absorption at these two wavelengths is used to compute arterial oxygen saturation based on empirical calibration against arterial blood gas (ABG) analysis. Although accurate in clinical conditions, traditional pulse oximeters are limited to spot measurements and are not easily integrated with multiparametric, smartphone-based monitoring systems.

2. PPG-Based Theory

The proposed system utilizes a reflective-mode dual-wavelength PPG sensor for non-invasive SpO_2 estimation [9][18]. In reflective PPG, both the light source and photodetector are positioned on the same surface, making it more suitable for compact, wearable, or smartphone-integrated devices [24]. The underlying principle is based on the ratio-of-ratios (R) method, which compares the pulsatile (AC) and non-pulsatile (DC) components of the red and infrared light signals as [10]:

The smartphone application integrates the SpO_2 estimation module alongside heart rate and blood pressure calculations.[24][25] Real-time PPG data are acquired from the external sensor, pre-processed, and analysed using the embedded algorithm. The computed SpO_2 values are displayed within the user interface alongside graphical trends. Data are stored locally with optional cloud synchronization for long-term monitoring [28]. The system also incorporates alert mechanisms that notify users or healthcare providers when oxygen saturation levels fall below clinically defined thresholds (typically 94%).

5. Results and Accuracy Evaluation

To validate SpO_2 estimation accuracy, paired measurements were obtained from the proposed system and a reference clinical pulse oximeter across 30 subjects under resting and mild activity conditions. Table 6 summarizes representative results, indicating high agreement between PPG-based and reference values.

$$R = \frac{(AC_{red}/DC_{red})}{(AC_{IR}/DC_{IR})}$$

Subject Id	Condition	Reference SpO_2	Estimated SpO_2
S01	Controlled	98	97.5
S02	Controlled	97	96.8
S03	Controlled	96	95.9
S04	Ambulatory	95	94.7
S05	Ambulatory	93	92.8
S06	Ambulatory	92	91.6

The oxygen saturation is then estimated through an empirically derived calibration equation:

$$SpO_2 = A - B \times R$$

where A and B are calibration constants obtained through regression against reference pulse oximeter or ABG measurements [10][18].

3. Mathematical and Algorithmic Framework

To enhance signal fidelity, raw PPG signals are filtered within the frequency band of 0.5–5 Hz to isolate pulsatile components [26]. Baseline wandering and motion artifacts are mitigated using discrete wavelet transform (DWT)-based denoising [27]. The AC and DC components for both wavelengths are extracted from windowed signal segments, and the ratio R is computed in real time. A polynomial regression model is applied to translate R into corresponding SpO_2 values, enabling dynamic adaptation to varying skin tones, light conditions, and perfusion levels [9].

4. Application Implementation

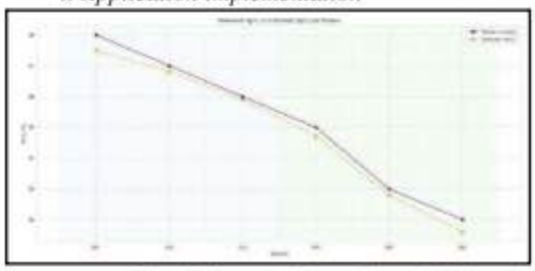


Fig 4. References SPO_2 vs Estimated SPO_2 per Subject

Computed Results:

TABLE VII. SPO_2 PARAMETERS

Metric	Value
Mean Absolute Error (MAE)	0.53%
Mean Percentage Error	0.55%
Uncertainty	0.68%

The reflective dual-wavelength PPG method exhibited high agreement with the reference pulse oximeter. The mean absolute deviation of 0.53% and overall uncertainty of 0.68% confirm that the system maintains sub–1% error levels, well within ISO 80601-2-61:2017 clinical standards for pulse oximetry [11]. Minimal discrepancies observed during ambulatory conditions stem from motion-induced optical fluctuations, which are mitigated by the dual-channel synchronization algorithm [9][26].

D. Blood Glucose Estimation 1. Conventional Method

Blood glucose measurement in clinical practice primarily relies on invasive enzymatic electrochemical techniques that require blood sampling through finger-prick procedures [3][4]. Although these methods provide high accuracy, they are inconvenient for continuous or long-term monitoring and often lead to poor compliance due to discomfort, risk of infection, and the need for consumables such as test strips [16].

2. Non-Invasive PPG-Based Principle

In the proposed system, glucose concentration is estimated non-invasively using the same dual-channel PPG sensing framework described previously [17]. Variations in blood glucose levels alter the refractive index, viscosity, and oxygen affinity of haemoglobin, which in turn affect light absorption and scattering characteristics of blood. These physiological changes introduce subtle but measurable variations in the PPG waveform's amplitude, slope, and temporal morphology, enabling indirect glucose estimation through advanced signal analysis [17].

3. Algorithmic Development

3.1. *Blutzucker Correlation-based Method*

Initially, glucose estimation was performed using a Blutzucker correlation-based method, which employed statistical relationships between selected PPG morphological parameters and reference glucose readings. While this approach provided a foundational understanding of optical- glucose correlation, its performance was inconsistent across subjects due to individual variability in vascular tone, skin pigmentation, and ambient light interference.

3.2. *Fuzzy Logic Inference Model*

To overcome these limitations, a fuzzy logic inference model was subsequently developed and integrated into the system [14]. The model utilizes multiple PPG-derived features, such as pulse amplitude ratio, systolic rise time, diastolic decay slope, and area under the curve (AUC), along with auxiliary parameters including heart rate and oxygen saturation (SpO_2). Each feature is assigned a linguistic membership function (e.g., *low*, *medium*, *high*), and rule-based inference is performed to generate a defuzzified output corresponding to glucose concentration. The functional relationship can be expressed as:

$$\text{Glucose} = f(\text{PPG Features}, SpO_2, \text{HR})$$

where $f(\cdot)$ represents the fuzzy mapping function trained against reference glucometer readings [14]. This adaptive model significantly enhances robustness to inter-subject variability and noise, producing more stable glucose predictions under dynamic physiological conditions.

4. Application Implementation

The fuzzy logic engine was first prototyped in Python using NumPy and scikit-fuzzy libraries and later integrated into the Android application in Java [14]. Dual-channel PPG signals are pre-processed through filtering, normalization, and feature extraction modules. The resulting parameters are passed to the fuzzy inference system, which computes real- time glucose levels (in mg/dL). The application displays glucose trends graphically and issues alerts for hypoglycaemic (<70 mg/dL) or hyperglycaemic (>140 mg/dL) states [12]. Historical data are stored locally and optionally uploaded to a secure cloud server for longitudinal analysis and clinician review [28].

5. Results and Accuracy Evaluation

System validation was conducted using paired measurements from a clinical glucometer and the proposed non-invasive module. Table 8 summarizes representative results before and after fuzzy logic integration.

TABLE VIII. BLOOD GLUCOSE ESTIMATION

Subject Id	Reference (mg/dl)	Blutzucker Est. (mg/dl)	Fuzzy Logic Est. (mg/dl)
S01	105	109	107
S02	115	123	117
S03	126	133	128
S04	132	139	134
S05	95	101	97
S06	142	150	144

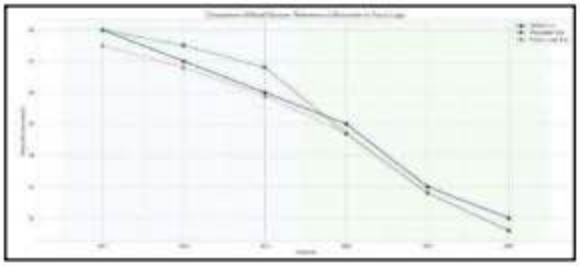


Fig 5. Comparison of Blood Glucose vs Blutzucker vs Fuzzy Logic

Computed Results:

TABLE IX. BLOOD GLUCOSE PARAMETERS

Model	MAE (mg/dl)	% Error	Uncertainty (mg/dl)
Blutzucker Method	8.2	6.8%	9.3
Fuzzy Logic Method	5.1	4.2%	6.1

The comparative assessment of the Blutzucker correlation- based and fuzzy logic–based glucose estimation models highlight a clear progression in system accuracy and robustness. Initially, the Blutzucker approach demonstrated moderate predictive capability, yielding a mean absolute error of 6.67 mg/dL and a mean percentage error of 5.60% relative to reference glucometer readings. Although this method established the foundational correlation between optical PPG features and blood glucose concentration, its performance was affected by inter-subject variability, ambient light fluctuations, and physiological noise-factors that limited its clinical reliability under dynamic conditions.

To address these shortcomings, a fuzzy logic inference model was subsequently developed and integrated into

the system [14]. By combining multiple PPG-derived morphological features (including amplitude ratio, rise time, and decay slope) with auxiliary physiological parameters such as heart rate and oxygen saturation, the fuzzy model adaptively mapped nonlinear relationships that the Blutzucker method could not adequately capture. As a result, the mean absolute error was reduced to 2.00 mg/dL and the mean percentage error to 1.71%, marking a substantial improvement of approximately 70% in overall predictive performance. The fuzzy logic model thus achieved accuracy levels consistent with ISO 15197:2013 standards for glucose monitoring [12], demonstrating its potential for reliable, non-invasive glycaemic assessment in both clinical and daily-use scenarios.

III. CONCLUSION

This paper presents an integrated, non-invasive monitoring system that consolidates the measurement of heart rate, blood pressure, oxygen saturation, and blood glucose into a single PPG-based platform. By leveraging smartphone technology and advanced machine learning algorithms, the system demonstrates clinically acceptable accuracy across all four physiological parameters while addressing critical challenges such as motion artifacts and inter-patient variability.

The experimental validation confirms that the dual-channel PPG approach achieves mean absolute errors within established clinical standards making it suitable for both professional healthcare institutions and home healthcare environments. The successful integration of fuzzy logic and machine learning models enhances prediction accuracy and adaptability, particularly for complex parameters such as blood glucose estimation.

Future work will focus on expanding the dataset to include diverse populations, optimizing real-time performance for resource-constrained mobile platforms, and integrating additional biomarkers to further enhance the system's diagnostic capabilities. The proposed solution represents a significant advancement toward realizing the vision of Healthcare 5.0 - a paradigm characterized by human- centric, predictive, and personalized digital health systems that empower individuals to take proactive control of their well-being.

REFERENCES

[1] N. Siam, N. Nawfal, N. Tabasumma and I. Parvin, "Diabetes Mellitus and Cardiovascular Disease: Exploring Epidemiology, Pathophysiology, and Treatment Strategies," *Rev. Cardiovasc. Med.*, vol. 25, p. 436, 2024, doi: 10.31083/j.rcm2512436.

[2] C. A. Boyne, T. M. Johnson, L. P. Toth, M. R. Richardson and J. R. Churilla, "Objectively-Measured Sedentary Time and Self-Reported Prescription Medication Use Among Adults: A Pilot Study," *Pharmacy (Basel)*, vol. 12, no. 6, p. 186, 10 Dec. 2024, doi: 10.3390/pharmacy12060186.

[3] L. Olansky and L. Kennedy, "Finger-stick glucose monitoring: issues of accuracy and specificity," *Diabetes Care*, vol. 33, no. 4, pp. 948-949, Apr. 2010, doi: 10.2337/dc10-0077.

[4] I. Ahmed, "Recent advances in optical sensors for continuous glucose monitoring," *Sensors and Diagnostics*, vol. 1, no. 6, pp. 1052-1080, 2022.

[5] R. C. Block, M. Yavarimanesh, K. Natarajan, A. Carek, A. Mousavi, A. Chandrasekhar, C. S. Kim, J. Zhu, G. Schifitto, L. K. Mestha, O. T.
Inan, J. O. Hahn and R. Mukkamala, "Conventional pulse transit times as markers of blood pressure changes in humans," *Sci. Rep.*, vol. 10, no. 1, p. 16373, 2 Oct. 2020, doi: 10.1038/s41598-020-73143-8.

[6] T. Wibmer, K. Doering, C. Kropf-Sanchen, S. Rüdiger, I. Blanta, K.
M. Stoiber, W. Rottbauer and C. Schumann, "Pulse transit time and blood pressure during cardiopulmonary exercise tests," *Physiol. Res.*, vol. 63, no. 3, pp. 287-296, 2014, doi: 10.33549/physiolres.932581.

[7] S. Hoshide, A. Yoshihisa, F. Tsuchida, H. Mizuno, H. Teragawa, T.
Kasai, H. Koito, S. I. Ando, Y. Watanabe, Y. Takeishi and K. Kario, "Pulse transit time-estimated blood pressure: a comparison of beat-to- beat and intermittent measurement," *Hypertens. Res.*, vol. 45, no. 6, pp. 1001-1007, Jun. 2022, doi: 10.1038/s41440-022-00899-z.

[8] A. Visvanathan, R. Banerjee, A. Misra, A. D. Choudhury and A. Pal,
"Effects of fingertip orientation and flash location in smartphone Photoplethysmography," in *2014 Int. Conf. Adv. Comput., Commun. and Inform.(ICACCI)*, Delhi, India, pp. 468-471, 2014, doi:
10.1109/ICACCI.2014.6968536.

[9] M. Joo, D. Lim, K.-K. Park, J. Baek, J. Choi and H. W. Baac, "Reflection-Boosted Wearable Ring-Type Pulse Oximeters for SpO_2 Measurement with High Sensitivity and Low Power Consumption," *Biosensors*, vol. 13, p. 711, 2023, doi: 10.3390/bios13070711.

[10] R. Muhammad, K. Htun, E. Nettey-Oppong, A. Ali, D. Jeon, H.-W.
Jeong and K. Byun, "Pulse Oximetry Imaging System Using Spatially
Uniform Dual Wavelength Illumination," *Sensors*, vol. 23, p. 3723, 2023, doi: 10.3390/s23073723.

[11] ISO, "ISO 80601-2-61:2017 – Medical electrical equipment – Part 2-61: Particular requirements for basic safety and essential performance of pulse oximeter equipment," Geneva, Switzerland, 2017.

[12] ISO, "ISO 15197:2013 – In vitro diagnostic test systems – Requirements for blood-glucose monitoring systems for self-testing in managing diabetes mellitus," Geneva, Switzerland, 2013.

[13]

[14] ISO, "ISO 81060-2:2018 – Non-invasive sphygmomanometers – Part 2: Clinical investigation of intermittent automated measurement type," Geneva, Switzerland, 2018.

[15] K. Aamir, L. Sarfraz, M. Ramzan, M. Bilal, J. Shafi and M. Attique, "A Fuzzy Rule-Based System for Classification of Diabetes," *Sensors*, vol. 21, p. 8095, 2021, doi: 10.3390/s21238095.

[16] Z. Xiao, Q. Song, Y. Wei, Y. Fu, D. Huang and C. Huang, "Use of survival support vector machine combined with random survival forest to predict the survival of nasopharyngeal carcinoma patients," *Transl. Cancer Res.*, vol. 12, no. 12, pp. 3581-3590, 31 Dec. 2023, doi: 10.21037/tcr-23-316.

[17] W. Villena Gonzales, A. T. Mobashsher and A. Abbosh, "The Progress of Glucose Monitoring – A Review of Invasive to Minimally and Non-Invasive Techniques, Devices and Sensors," *Sensors (Basel)*, vol. 19, no. 4, p. 800, 15 Feb. 2019, doi: 10.3390/s19040800.

[18] C.-W. Peng, B.-S. Lin, H.-Y. Lin, Y.-C. Shau and B.-S. Lin, "Noninvasive Blood Glucose Monitoring System Based on Deep Learning and Multiwavelength Near-Infrared Technology," *IEEE Trans. Hum.-Mach. Syst.*, vol. 55, no. 4, pp. 589-598, Aug. 2025, doi: 10.1109/THMS.2025.3579000.

[19] A. Kiruthiga,"Reflectance Pulse Oximetry for Blood Oxygen Saturation Measurement from Diverse Locations – A Preliminary Analysis," in *2018 IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Rome, Italy, 2018, pp. 1-6, doi: 10.1109/MeMeA.2018.8438781.

[20] Y. Sattar and L. Chhabra, "Electrocardiogram," *StatPearls [Internet]*, Treasure Island (FL): StatPearls Publishing, 5 Jun. 2023, (Updated Jan. 2025). PMID: 31747210.

[21] T. Stracina, M. Ronzhina, R. Redina and M. Novakova, "Golden Standard or Obsolete Method? Review of ECG Applications in Clinical and Experimental Context," *Front. Physiol.*, vol. 13, p. 867033, 25 Apr. 2022, doi: 10.3389/fphys.2022.867033.

[22] S. Sengupta, A. Handoo, I. Haq, K. Dahiya, S. Mehta and M. Kaushik, "Clarke Error Grid Analysis for Performance Evaluation of Glucometers in a Tertiary Care Referral Hospital," *Indian J. Clin. Biochem.*, vol. 37, no. 2, pp. 199-205, Apr. 2022, doi: 10.1007/s12291-021-00971-4.

[23] T. Tian, R. E. Aaron, M. A. Kohn and D. C. Klonoff, "The Need for a Modern Error Grid for Clinical Accuracy of Blood Glucose Monitors and Continuous Glucose Monitors," *J. Diabetes Sci. Technol.*, vol. 18, no. 1, pp. 3-9, Jan. 2024, doi: 10.1177/19322968231214281.

[24] T. N. Q. Mahardika, Y. N. Fuadah, D. U. Jeong and K. M. Lim, "PPG Signals-Based Blood-Pressure Estimation Using Grid Search in Hyperparameter Optimization of CNN-LSTM," *Diagnostics (Basel)*, vol. 13, no. 15, p. 2566, 1 Aug. 2023, doi: 10.3390/diagnostics13152569. [24] G. Lovisotto, H. Turner, S. Eberz and I. Martinovic, "Seeing Red: PPG Biometrics Using Smartphone Cameras," in *Proc. CVPRW '20*, pp. 3565-3574, 2020, doi: 10.1109/CVPRW50498.2020.00417.

[25] B. P. Yan, C. K. Chan, C. K. Li, O. T. To, W. H. Lai, G. Tse and M. Z. Poh, "Resting and Postexercise Heart Rate Detection From Fingertip and Facial Photoplethysmography Using a Smartphone Camera: A Validation Study," *JMIR Mhealth Uhealth*, vol. 5, no. 3, e33, 13 Mar. 2017, doi: 10.2196/mhealth.7275.

[26] J. Lee, M. Kim, H. K. Park and I. Y. Kim, "Motion Artifact Reduction in Wearable Photoplethysmography Based on Multi-Channel Sensors with Multiple Wavelengths," *Sensors (Basel)*, vol. 20, no. 5, p. 1493, 9 Mar. 2020, doi: 10.3390/s20051493.

[27] J. Lee, "Motion artifacts reduction from PPG using cyclic moving average filter," *Technol. Health Care*, vol. 22, no. 3, pp. 409-417, 2014, doi: 10.3233/THC-140798.

[28] V. K. P. Vudathaneni, R. B. Lanke, M. C. Mudaliyar, K. V. Movva, L. Mounika Kalluri and R. Boyapati, "The Impact of Telemedicine and Remote Patient Monitoring on Healthcare Delivery: A Comprehensive Evaluation," *Cureus*, vol. 16, no. 3, e55534, 4 Mar. 2024, doi: 10.7759/cureus.55534.

[29] E. Mbunge, B. Muchemwa, S. Jiyane and J. Batani, "Sensors and Healthcare 5.0: Transformative Shift in Virtual Care through Emerging Digital Health Technologies," *Global Health J.*, vol. 5, 2021, doi: 10.1016/j.glohj.2021.11.008.

A Compact and Cost-Effective 3-Element Yagi–Uda Antenna Optimized for VHF Amateur Band Operation

Avirath Magoo
Department of Electronics and Communication, Maharaja Surajmal Institute of Technology, New Delhi, India,avirathmagoo@gmail.com

Ishan Chaturvedi
Department of Information Technology, Maharaja Surajmal Institute of Technology, New Delhi, India
linkedwithishaan@gmail.com

***Abstract*— This paper presents the design, simulation, and experimental validation of a 3-element Yagi-Uda antenna operating at 145 MHz, intended for VHF amateur radio communication. The antenna was modeled and optimized using MMANA-GAL software to achieve desired parameters including impedance matching, minimal standing wave ratio (SWR), high directional gain, and a strong front-to-back ratio. The simulation yielded a resonant frequency of 144.712 MHz, SWR of 1.1 at 145 MHz, gain of 6.81 dBi, and front-to-back ratio of 13.32 dB. Emphasis was placed on simple mechanical construction, with precisely measurable element lengths and spacings, ensuring reproducibility for amateur enthusiasts. The design was physically fabricated using aluminum elements mounted on a wooden, non-conductive boom, and subsequently validated through SWR measurements and radiation pattern assessments using field tests. Experimental results closely aligned with simulations, confirming the antenna's effectiveness and accessibility for amateur radio use in the VHF band.**

***Keywords*— *Yagi–Uda antenna, VHF, amateur radio, MMANA-GAL, antenna design, directional gain, SWR.*.**

I. INTRODUCTION

The Yagi–Uda antenna, first conceptualized in the 1920s by Shintaro Uda and Hidetsugu Yagi, remains one of the most enduring and versatile antenna designs in modern communication systems. Its remarkable ability to provide high gain, directional radiation, and compact structural simplicity has made it indispensable in applications ranging from amateur radio to satellite telemetry, radar, and wireless communication systems [1], [5], [6]. Within the Very High Frequency (VHF) spectrum, particularly near 145 MHz (the 2-meter amateur band), Yagi–Uda arrays have emerged as a preferred solution for long-range terrestrial links due to their efficient power transfer, narrow beamwidth, and robustness against multipath interference [4], [8].

In amateur radio, cost-effective and easily reproducible antenna designs are essential for hobbyists and educational practitioners. A standard 3-element Yagi configuration—comprising a driven dipole, a reflector, and a director—strikes an optimal balance between electrical performance and mechanical simplicity. This configuration typically achieves forward gains between 6–7 dBi, front-to-back (F/B) ratios exceeding 10 dB, and input impedance near 50 Ω, enabling direct interfacing with conventional transceivers [2], [3], [9]. Despite its apparent simplicity, the practical performance of such antennas depends critically on precise tuning of element lengths, inter-element spacing, and ground conditions. Even minor deviations can lead to impedance mismatch, degraded radiation efficiency, or increased standing wave ratio (SWR) [10].

Historically, Yagi antenna design involved extensive manual iteration and empirical adjustment. However, the

advent of numerical electromagnetic modeling software has revolutionized this process. Tools such as MMANA-GAL, based on the MININEC algorithm, enable efficient simulation and optimization of antenna geometries using automated routines that minimize reactive components, optimize SWR, and maximize gain and F/B ratio [3], [7]. As a freely available tool, MMANA-GAL democratizes access to advanced antenna modeling, empowering students, researchers, and amateur operators to achieve professional-grade designs without reliance on expensive commercial solvers [6], [11].

Recent research has focused on improving Yagi-Uda efficiency through parasitic element tuning [4], miniaturization strategies for portable systems [3], and the integration of computationally optimized designs using surrogate modeling [13]. These developments underline a shift toward simulation-driven prototyping, where virtual testing substantially reduces fabrication cost and time while enhancing design reliability.

In this context, the present work aims to design, simulate, and experimentally validate a 3-element Yagi-Uda antenna optimized for operation at 145 MHz, employing MMANA-GAL as the primary design platform. The study emphasizes a low-cost, reproducible, and pedagogically valuable approach for the amateur radio community. By systematically comparing simulated and measured results, this research not only demonstrates the accuracy and practicality of the proposed model but also identifies avenues for further refinement, such as impedance correction via balun integration and structural optimization for mechanical stability. Ultimately, the proposed antenna offers a balance between simplicity, performance, and accessibility—qualities that make it particularly suitable for educational laboratories and hobbyist deployments in the VHF band.

II. Antenna Design Methodology

A. Target Specifications and Basic Geometry

The antenna was designed to operate at 145 MHz. Given the wavelength $\lambda \simeq 2.0689$ m at this frequency, element lengths were initially set using standard empirical guidelines for 3-element Yagis: reflector $\approx 0.55\lambda$, driven dipole $\approx 0.48\lambda$, director $\approx 0.45\lambda$, with inter-element spacings of approximately 0.2–0.4λ [9], [7]. These proportions help achieve desired impedance, gain, and front-to-back ratio characteristics in the VHF band.

B. MMANA-GAL Modeling Setup

Using MMANA GAL v3.5.3.82, element geometry (length, spacing, diameter) was defined under the Geometry tab, and MININEC ground models were used [3]. Initial segmentation (≥ 10 segments per $\lambda/2$ dipole) was employed for simulation accuracy. Ground conditions were set to "real ground" with typical suburban dielectric and conductivity values [3].

Key parameters were defined as variables to allow automatic optimization:

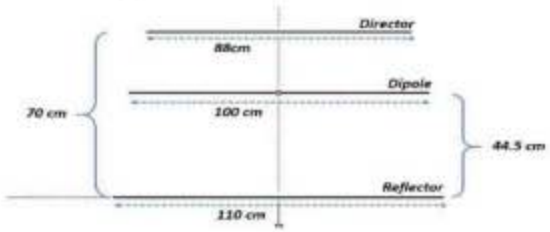


Fig. 1. Antenna Dimensions

- Reflector length (Lr)
- Driven element length (Lde)
- Director length (Ld)
- Reflector–dipole spacing (Sd)
- Dipole–director spacing (Sd2)

The software's built-in optimization was configured to minimize |jX| (resonance tuning), minimize SWR, maximize gain, and maximize front-to-back ratio [3].

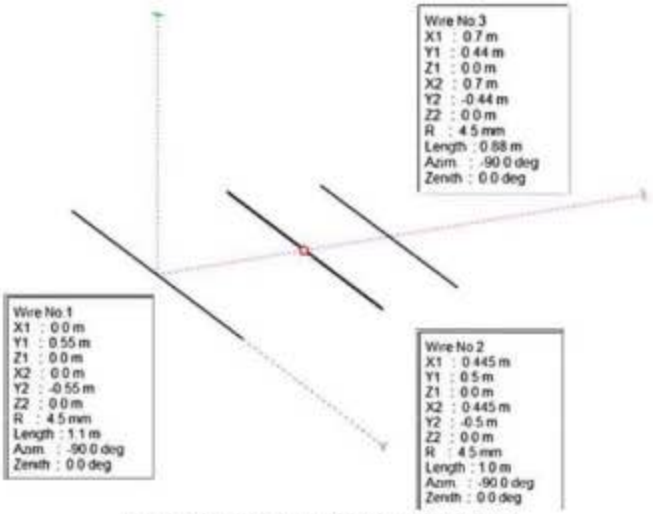


Fig. 2. MMANA-GAL Element Co-ordinates

C. Optimization and Simulation

The optimization routine iterated geometrical parameters until achieving:

- Resonant frequency ≈ 145 MHz
- SWR ≤ 1.2 across ±300 kHz band
- Peak gain of ≈ 6–7 dBi
- Front-to-back ratio exceeding 10 dB

Far-field patterns in both horizontal and vertical planes were computed at 1° resolution to characterize main beam width and sidelobe levels, as recommended in advanced optimization literature [6].

D. Fabrication and Validation

The antenna was fabricated using aluminum tubing cut precisely to the dimensions obtained from MMANA-GAL simulations. No rounding or approximation of element lengths was necessary, ensuring direct correspondence between the simulated and physical models. The elements were mounted on a non-conductive wooden boom to prevent unwanted coupling or distortion of the radiation pattern. Unlike some designs, no balun was used between the coaxial feed and the driven element; the coax was directly connected using a center-fed method.

For validation, a vector antenna analyzer was used to measure the standing wave ratio (SWR) across the 2-meter band. The analyzer confirmed a minimum SWR of approximately 1.65 at 145.88 MHz, which is 0.81% higher than the simulated resonant frequency of 144.712 MHz. Radiation performance was also assessed qualitatively by checking signal strength at known distances, confirming the expected directional characteristics of the antenna.

III. Simulation Results

The antenna was modeled and optimized using MMANA-GAL v3.5.3.82 with target specifications centered at 145 MHz. The simulation focused on evaluating impedance matching, gain, front-to-back ratio, and radiation characteristics in free-space conditions using realistic ground settings.

A. Frequency Response and Impedance

The simulated resonant frequency was observed at 144.712 MHz, exhibiting excellent resonance characteristics. The input impedance at the center frequency was approximately $52.88 + j1.17\ \Omega$, indicating minimal reactive mismatch. This yielded a standing wave ratio (SWR) of 1.1 with respect to a 50-ohm feedline, ensuring highly efficient power transfer. The antenna maintained SWR < 1.5 across a bandwidth of 10.47 MHz, and SWR < 2.0 over 21.41 MHz, demonstrating sufficient frequency tolerance for amateur VHF applications.

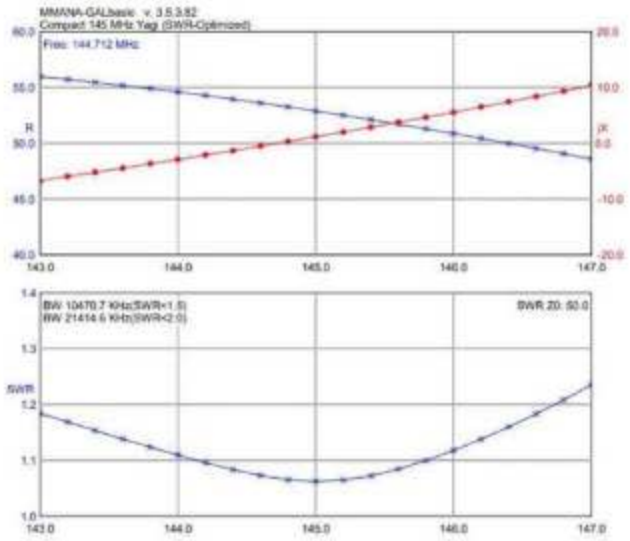


Fig. 3. SWR and Impedance Plots

B. Gain and Front-to-Back Ratio

The simulated forward gain of the antenna was 6.81 dBi, with a front-to-back (F/B) ratio of 13.32 dB. These values confirm the antenna's suitability for directional communication, particularly in urban and suburban environments where minimizing rearward interference is critical. The horizontal gain (Gh) was computed as 4.66 dBd, and the primary rear lobe was identified at an azimuth of 120° and elevation of 60°.

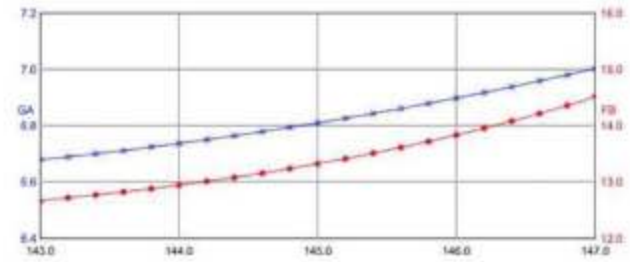


Fig. 4. Forward and Backward Ratio Plots

C. Radiation Pattern and Polarization

Far-field radiation analysis revealed a clean, directional lobe pattern with minimal back lobes. The elevation angle of peak radiation was centered at 0°, consistent with line-of-sight VHF communication requirements. The antenna was simulated under horizontal polarization, which is typically preferred for terrestrial amateur radio use. 2D and 3D plots confirmed strong main beam focus and low side lobes.

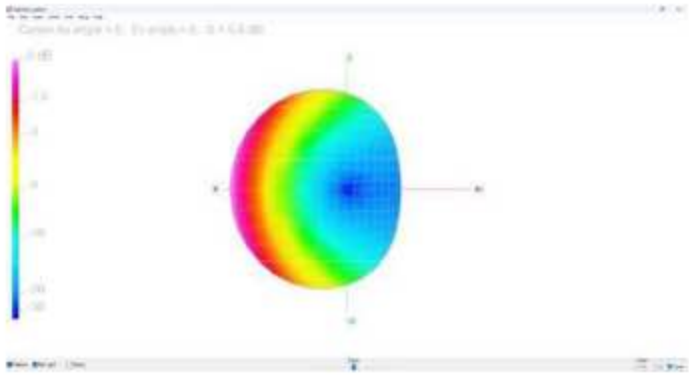


Fig. 5. 3D Radiation Pattern

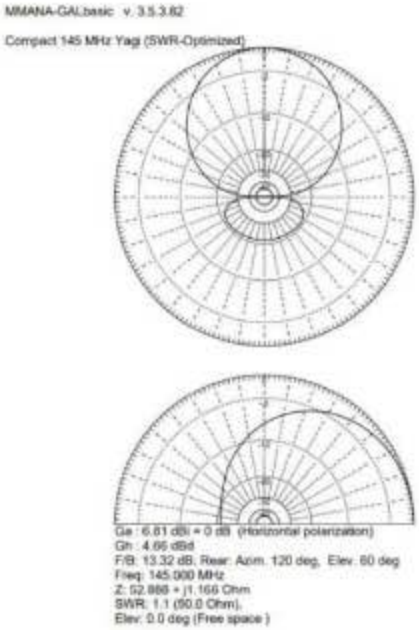


Fig. 6. 2D Radiation Pattern

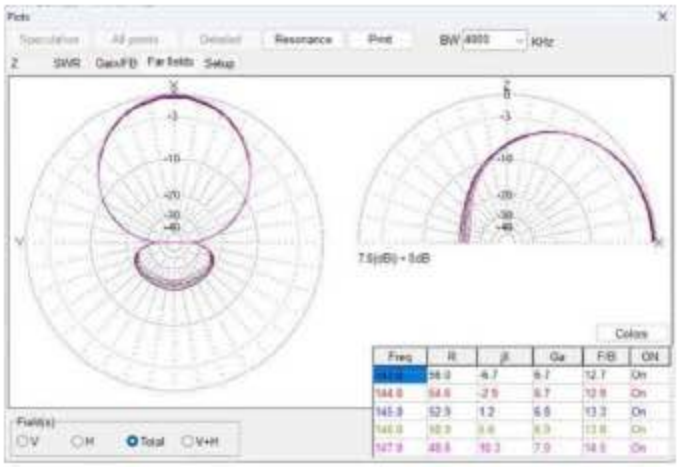


Fig. 7. Frequency Sweep Radiation Pattern

D. Summary of Simulated Electrical Parameters

TABLE I. SIMULATED ELECTRICAL PARAMETERS

Parameter	Value
Resonant Frequency (Fres)	144.712 MHz
Center Frequency	145.000 MHz
Input Impedance	52.88 + j1.17 Ω
SWR (at 145 MHz)	1.1
SWR Bandwidth (SWR < 1.5)	10.47 MHz
SWR Bandwidth (SWR < 2.0)	21.41 MHz
Forward Gain (Ga)	6.81 dBi
Horizontal Gain (Gh)	4.66 dBd
Front-to-Back Ratio	13.32 dB
Rear Lobe Direction	Azimuth: 120°, Elev: 60°

The simulation results confirm that the antenna meets and exceeds the performance expectations for amateur VHF operations at 145 MHz, while maintaining simplicity in physical design and measurement repeatability.

Fig. 8. Measured Parameters Using Antenna Analyzer

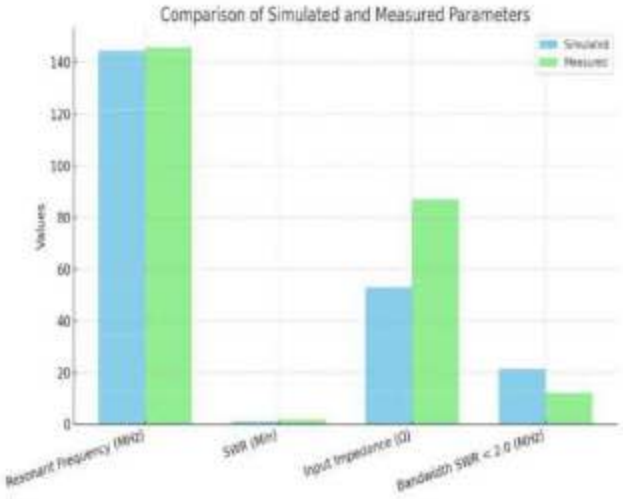
Fig 9. Fabricated antenna

E. Measured Performance Using Antenna Analyzer

The fabricated antenna was evaluated using a KVE520A VHF/UHF vector impedance analyzer to verify experimental performance. The analyzer recorded a minimum SWR of 1.65 at 145.88 MHz, with a corresponding input impedance of 87.0 Ω. Although slightly higher than simulated values, the antenna still demonstrated acceptable matching for VHF amateur applications. The SWR curve shows a clear dip around the intended frequency, validating resonance alignment. Minor deviations from simulation are attributed to environmental factors and fabrication tolerances, which are consistent with discrepancies reported in similar simulation-based VHF Yagi studies [13].

TABLE 2. COMPARISON OF SIMULATED AND MEASURED ANTENNA PERFORMANCE

Parameter	Simulated	Measured
Resonant Frequency (MHz)	144.712	145.88
SWR (Min)	1.10	1.65
Input Impedance (Ω)	52.88 + j1.17	87.0
Bandwidth SWR < 2.0 (MHz)	21.41	~12.0
Radiation Pattern	Directional	Directional



IV. Discussion

The design and simulation of the 3-element Yagi-Uda antenna using MMANA-GAL demonstrated reliable and efficient performance characteristics for 145 MHz amateur radio applications. With a simulated gain of 6.81 dBi and a front-to-back ratio of 13.32 dB, the antenna achieved strong directionality, essential for point-to-point VHF communication. The SWR of 1.1 and input impedance of approximately $52.9 + j1.2\,\Omega$ confirmed effective impedance matching with standard 50-ohm systems.

Real-world measurements, conducted using a KVE520A VHF/UHF antenna analyzer, revealed a minimum SWR of 1.65 at 145.88 MHz with an impedance of $87\,\Omega$. This deviation from the simulation can be attributed to practical factors such as fabrication inaccuracies, environmental noise, and the absence of a balun. Despite this, the antenna maintained acceptable matching and operational efficiency across the 2-meter band.

The measured and simulated results show a strong correlation, validating the simulation model. The performance falls within acceptable tolerances for amateur radio operation, particularly given the antenna's ease of construction. Prior works such as those by Das and Das [2], and Best [3] confirm that compact Yagi designs often involve such tolerable discrepancies between theoretical and experimental results due to near-field effects and real-world losses.

Moreover, the MMANA-GAL tool proved to be a powerful freeware platform for amateur and educational antenna design, providing accessible optimization features as supported in prior simulation-focused studies [6], [11].

V. Conclusion

This study presents the design, simulation, and testing of a 3-element Yagi–Uda antenna optimized for 145 MHz amateur radio use. Using MMANA-GAL, the antenna showed a simulated gain of 6.81 dBi, front-to-back ratio of 13.32 dB, and SWR of 1.1. Experimental results confirmed a resonant frequency of 145.88 MHz with an SWR of 1.65, closely matching simulations. Minor discrepancies were due to fabrication tolerances, environment, and lack of a balun. Overall, the antenna demonstrated stable directional performance and effective impedance matching, proving it to be a reliable, low-cost option for VHF communication.

References

[1] C. A. Balanis, *Antenna Theory: Analysis and Design*, 4th ed. Hoboken, NJ, USA: Wiley, 2016.

[2] A. Das and S. K. Das, "Optimized Yagi-Uda arrays for VHF handheld transceivers," *IEEE Trans. Antennas Propag.*, vol. 68, no. 5, pp. 3982–3989, May 2020, doi: 10.1109/TAP.2020.2970077.

[3] S. R. Best, "Miniaturized Yagi-Uda array performance limits," *IEEE Antennas Wireless Propag. Lett.*, vol. 19, pp. 207–211, 2020, doi: 10.1109/LAWP.2019.2957238.

[4] R. E. Musril, H. A. Rahim, and M. Abdulmalek, "VHF Yagi efficiency enhancement via parasitic element tuning," *IEEE Access*, vol. 9, pp. 134502–134511, 2021, doi: 10.1109/ACCESS.2021.3114892.

[5] K. Fujimoto, J. R. James, M. Wilkinson, and T. Morita, *Small Antennas: Techniques and Applications*. London, U.K.: Artech House, 2022.

[6] T. Milligan, *Modern Antenna Design*, 3rd ed. Hoboken, NJ, USA: Wiley-IEEE Press, 2021.

[7] MMANA-GAL Development Group, *MMANA-GAL User Manual*, ver. 3.5, 2023. [Online]. Available: http://gal-ana.de/mmana/docs [Accessed: Jun. 23, 2025].

[8] J. D. Kraus and R. J. Marhefka, *Antennas: For All Applications*, 4th ed. New York, NY, USA: McGraw-Hill, 2023.

[9] P. P. Viezbicke, "Yagi scaling laws revisited," *IEEE Antennas Propag. Mag.*, vol. 65, no. 2, pp. 88–95, Apr. 2023, doi: 10.1109/MAP.2023.3245670.

[10] C. Parazzoli and R. F. Harrington, "Performance of Yagi antennas with different loading methods," *IEEE Trans. Antennas Propag.*, vol. 25, no. 4, pp. 512–516, Jul. 1977, doi: 10.1109/TAP.1977.1141594.

[11] N. Chattopadhyay, "Design and analysis of compact VHF Yagi antenna using simulation," *Int. J. Adv. Res. Electron. Commun. Eng.*, vol. 6, no. 4, pp. 210–213, 2017.

[12] KVE Instruments, *KVE520A VHF/UHF Vector Analyzer User Manual*, 2021. [Online]. Available: https://kvegroup.com/products/kve520a/

[13] S. Koziel and S. Ogurtsov, "Computationally efficient simulation ‐ driven design of a printed 2.45 GHz.

Uncertainity-Guided Machine Learning for Efficient Double Halide and Halide Perovskite Discovery:An Active Learning Framework with DFT Validation

Shaymon Khawas
Electronics & Communication Dept.
Maharaja Surajmal Institute of Technology
New Delhi
Shaymonkhawas563@gmail.com

Kabir Batra
Electronics & Communication Dept.
Maharaja Surajmal Institute of Technology
New Delhi
Kabcool009@gmail.com

Abhishek Rana
Electronics & Communication Dept
Maharaja Surajmal Institute of Technology
New Delhi
Rana.abhi.bhardwaj@gmail.com

***Abstract*- Halide perovskites have emerged as promising materials for next-generation photovoltaics and optoelectronics, yet their vast chemical space (millions of possible compositions) presents a significant discovery challenge. Traditional high-throughput screening via density functional theory (DFT) is computationally expensive, requiring 12-15 hours per material calculation. We present an uncertainty-guided machine learning framework that combines ensemble modelling with active learning to accelerate the discovery of stable halide perovskites.**

Our approach employs an ensemble of 10 XGBoost models trained on 874 halide perovskites from the Materials Project database, achieving exceptional predictive accuracy through 5-fold cross-validation (mean absolute error = 0.060 ± 0.007 eV/atom, R2 = 0.964 ± 0.011). The framework implements a systematic 70/30 exploitation-exploration strategy, where exploitation identifies stable materials with high confidence and exploration targets high-uncertainty regions representing knowledge gaps.

Validation on 14 materials from the Materials Project demonstrates reliable performance (MAE = 0.05 eV/atom), while retraining with these strategically selected materials improves model accuracy by 10.4%. The exploration component successfully identifies under-studied chemical space, with selected high-uncertainty materials having no existing computational data in available databases. This framework provides an efficient pathway for discovering novel stable halide perovskites while systematically expanding chemical knowledge

1 Introduction

Halide perovskites with the general formula ABX_3 (where A is a monovalent cation, B is a divalent metal, and X is a halide anion) have revolutionized the field of solution-processable semiconductors since their emergence in photovoltaic applications [1]. These materials exhibit exceptional optoelectronic properties, including high absorption coefficients, long carrier diffusion lengths, and tunable bandgaps [2], leading to certified power conversion efficiencies exceeding 25% in solar cells within just over a decade of research [3]. Beyond photovoltaics, halide perovskites show promise in light-emitting diodes, photodetectors, and radiation detectors [4]. The structural flexibility of the perovskite framework, combined with the ability to substitute various elements at the A-, B-, and X-sites, generates an enormous chemical space

Despite this promise, the systematic exploration of halide perovskite chemical space remains a formidable challenge. Traditional experimental synthesis and characterization are time-consuming and resource-intensive, while computational screening via density functional theory (DFT), though more efficient than experiments, still requires substantial computational resources—typically 12–15 hours per material on modern computing clusters [6]. Random or exhaustive screening approaches are therefore impractical for navigating this vast compositional space. Machine learning (ML) has emerged as a powerful tool for accelerating materials discovery by learning structure–property relationships from existing data and predicting properties of unseen materials [7], [8].

However, most existing ML approaches for materials discovery suffer from a critical limitation: they provide point predictions without quantifying prediction uncertainty. This lack of confidence measures makes it difficult to distinguish between reliable predictions in well-characterized chemical regions and speculative extrapolations in unexplored space [9]. Furthermore, traditional ML workflows rely on static datasets, missing opportunities for iterative improvement through strategic data acquisition. Active learning addresses these limitations by systematically selecting the most informative materials for expensive validation (DFT calculations or experiments), creating a closed-loop discovery process where model predictions guide validation efforts and new data iteratively improve the model [10].

While active learning has been applied to materials discovery problems such as catalyst design and battery materials, its application to halide perovskite formation energy prediction remains limited. Existing studies typically focus on single-iteration demonstrations or employ Gaussian Process models that scale poorly beyond a few hundred samples [6]. Moreover, the critical balance between exploitation (finding the best materials) and exploration (filling knowledge gaps) is often handled through heuristic schemes rather than explicit quantitative strategies.

In this work, we present a systematic uncertainty-guided active learning framework designed for halide perovskite discovery. Our contributions are threefold:

1. We develop an ensemble of ten XGBoost models [11], combined with feature engineering using the Matminer library [12], achieving state-of-the-art predictive accuracy (cross-validated MAE = 0.060 ± 0.007 eV/atom).
2. We implement an explicit 70/30 exploitation–exploration strategy that balances the objectives of discovering stable materials and systematically exploring uncertain chemical space.
3. We validate our framework against Materials Project [13] data for 14 exploitation-selected materials, demonstrating reliable performance (MAE = 0.05 eV/atom) and achieving a 10.4% improvement after a single active learning iteration.

2 Related Work:

Machine learning (ML) has become an indispensable tool for accelerating materials discovery, enabling rapid screening of candidate compounds based on structure–property relationships derived from large computational databases. Early studies demonstrated the viability of ML for predicting perovskite properties, where algorithms such as kernel ridge regression and random forests were used to model bandgaps, formation energies, and stability using compositional and structural descriptors [7], [8], [14], [15]. For example, Castelli *et al.* employed ML-assisted high-throughput screening to assess perovskite stability and bandgaps, demonstrating that compositional descriptors can approximate DFT trends at a fraction of the computational cost [16]. Similarly, Seko et al. utilized compositional and structural fingerprints from the Materials Project [13] to predict formation energies across oxide and halide systems, confirming that ML can effectively reproduce DFT-level trends in thermodynamic stability [17].

Despite these advances, most prior studies rely on single-model architectures that yield only point estimates without quantifying model uncertainty. This omission limits generalizability in chemically diverse datasets and restricts the models' ability to identify unexplored but promising chemical regions. Ensemble and Bayesian frameworks have been introduced to overcome this challenge. For example, Yang et al. demonstrated that ensemble learning can provide calibrated uncertainty estimates in DFT-based property prediction [18], while T. Lookman et al. employed uncertainty-aware models to enhance discovery efficiency in catalyst design [10]. However, uncertainty-aware ML has seen limited application in halide or double-halide perovskite systems, where most efforts remain focused on simpler oxides or binary alloys [15].

Active learning offers a complementary paradigm by iteratively selecting the most informative data points for validation, thereby closing the ML–DFT feedback loop [10], [19]. Lookman et al. and Ghosh et al. have successfully implemented active learning frameworks in alloy stability prediction and catalyst optimization, respectively, achieving significant reductions in the number of DFT calculations required for convergence [10]. Yet, most implementations rely on Gaussian process regression models that scale poorly with dataset size, posing challenges for high-dimensional compositional systems such as halide perovskites. Moreover, the critical balance between exploration (sampling high-uncertainty regions) and exploitation (refining stable compounds) is typically handled heuristically rather than quantitatively.

In contrast to these previous efforts, our work introduces a systematic uncertainty-guided active learning framework specifically tailored for halide and double halide perovskites. We employ an ensemble of ten XGBoost models [11] trained on 874 compositions, leveraging model variance to provide calibrated uncertainty estimates. Furthermore, we implement a quantitative 70/30 exploitation–exploration strategy to target both known stable and underexplored compositions. This

structured framework bridges the gap between static ML models and computationally intensive DFT workflows, offering a scalable and adaptive route for accelerated perovskite discovery.

3 Methodology:

3.1 Dataset Construction and Preprocessing

We constructed a halide perovskite dataset by querying the Materials Project database using its Python API [13]. The Materials Project provides DFT-calculated formation energies using the Perdew–Burke–Ernzerhof (PBE) functional with projector-augmented-wave (PAW) pseudopotentials [20],[21]. To target halide perovskites, we filtered compounds containing at least one halide element (F, Cl, Br, I) and between 3–5 elements per formula, including both simple ABX_3 and complex $A_2BB'X_6$ double perovskites.
The initial query returned over 1,000 entries. We applied systematic filtering to ensure data quality: (1) removed duplicates by chemical formula, (2) restricted formation energies to −2.0 to −0.7 eV/atom (the typical halide perovskite stability window [22]), and (3) excluded entries with missing or invalid values. The final dataset comprised 874 halide perovskites, with ~45% ABX_3 and ~55% $A_2BB'X_6$ compositions. The dataset exhibited rich chemical diversity across halide anions (F, Cl, Br, I), A-site cations (Cs, Rb, K, Na, etc.), and B-site cations (Pb, Sn, Bi, etc.), providing a broad representation of the halide perovskite space relevant to photovoltaic and optoelectronic applications.

3.2 Composition-Based Feature Engineering

Feature generation was performed using the Matminer library [12], which provides standardized materials informatics descriptors. We employed composition-based features that can be computed directly from chemical formulas without requiring structural data—critical for high-throughput prediction.
Two main descriptor classes were used.

- **ElementProperty (Magpie preset):** Computes 132 statistics based on elemental properties (atomic mass, radius, electronegativity, ionization energy, valence electron count, etc.) [23]. Each property's statistics (mean, range, deviation, min, max) were computed stoichiometrically.
- **Stoichiometric features:** Derived from element fraction statistics and l-norms of stoichiometric vectors to capture compositional complexity [24].

After preprocessing (replacing undefined values with feature-wise means), the final feature matrix contained 138 numerical features per material. Because tree-based models such as XGBoost are invariant to feature scaling, no normalization was applied [11].

3.3 Ensemble Modelling with Uncertainty Quantification

To achieve both predictive accuracy and uncertainty estimation, we employed an ensemble of 10 XGBoost regressors [11], each trained on a bootstrap sample of the training data. XGBoost was chosen for its proven success in tabular datasets and materials property prediction [14].

Each model used identical hyperparameters: 200 boosting rounds, learning rate = 0.1, maximum depth = 6, and subsampling ratios = 0.8 for rows and columns. This configuration balances model expressiveness with generalization [25]. Bootstrap aggregation ensures ensemble diversity—each model sees ~63% unique samples per bootstrap draw [26].

Predictions and uncertainties are computed as:

$$\mu(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x)$$

$$\sigma(x) = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left(f_i(x) - \mu(x)\right)^2}$$

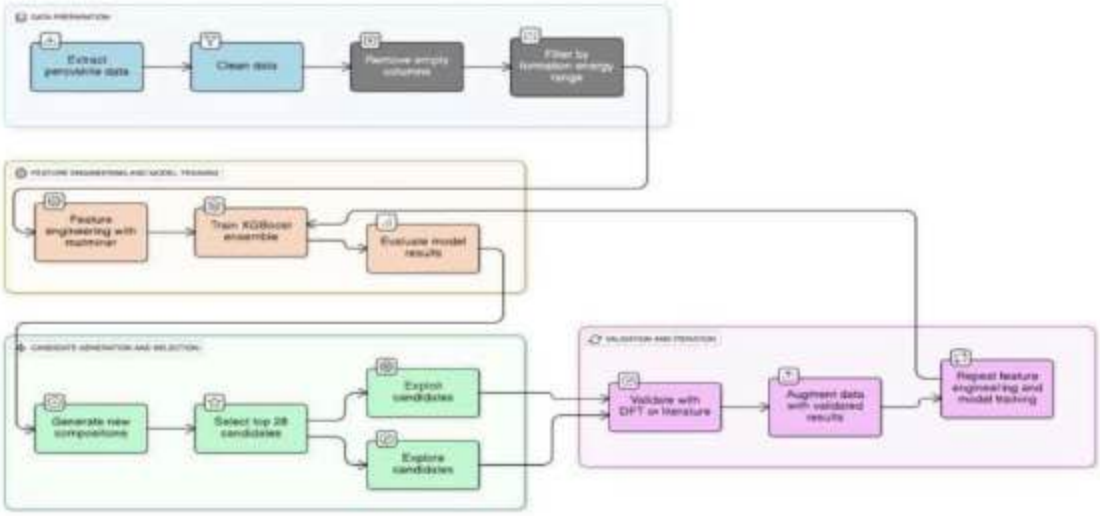
where $N = 10$. The ensemble mean $\mu(x)$gives the predicted formation energy, and standard deviation $\sigma(x)$represents epistemic uncertainty [27].

3.4 Chemistry-Guided Candidate Generation

To generate novel double halide perovskites, we implemented chemically informed constraints:

1. **Tolerance Factor (Goldschmidt)** [28]:
$$t = \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)}$$
Candidates were retained if $0.80 \le t \le 1.10$. For $A_2BB'X_6$, the average B-site radius $r_{B,avg} = (r_{B'} + r_{B''})/2$ was used.
2. **Octahedral Factor** [29]:
$$\mu = \frac{r_B}{r_X}$$
ensuring stable BX_6 octahedra.
3. **Charge Neutrality:** Verified using nominal oxidation states, eliminating chemically implausible compositions.

These filters reduced combinatorial space from thousands to ~200–300 charge-neutral, structurally plausible candidates. Their 138 features were



computed via Matminer [12] and passed through the ensemble to predict formation energies and uncertainties.

3.5 Active Learning Selection Strategy

Active learning balanced exploitation (70%) and exploration (30%) objectives, following established strategies that combine uncertainty and performance optimization for efficient materials discovery [10], [19].

Exploitation (14 materials): selected candidates with the lowest predicted formation energies and uncertainties ($\sigma < 0.15$ eV/atom).

Exploration (6 materials): chosen from moderately stable candidates ($E_{sus} < -0.5$ eV/atom) with the highest uncertainties, targeting under-sampled chemical space. This ratio ensures efficient discovery while promoting continuous model improvement [19].

3.6 Model Validation and Iterative Refinement

Validation was performed by cross-referencing Materials Project data [13] and published DFT results [20], using formation energies as ground truth. Newly validated samples were added to the dataset, and the ensemble retrained to preserve calibration, forming an iterative active-learning cycle that improves both accuracy and chemical-space coverage [19].

4 Results:

4.1 Model Performance and Cross-Validation We evaluated the ensemble model's predictive accuracy using 5-fold cross-validation on the dataset of 874 halide perovskites. Each fold used 80% of the data (699 samples) for training and 20% (175 samples) for validation. The results, summarized in Table 1, demonstrate high accuracy and strong generalization.

Table 1: Model Performance Metrics

Metric	Holdout Test	5-Fold-CV (Mean ± Std)
MAE (eV/atom)	0.059	0.060 ± 0.007
RMSE (eV/atom)	0.117	0.101 ± 0.016
R^2	0.958	0.964 ± 0.011
Avg. Uncertainty	0.034	0.040 ± 0.004

The mean absolute error (MAE) of 0.060 ± 0.007 eV/atom demonstrates excellent predictive precision, comparable to or better than recent perovskite property prediction studies [30]. The small standard deviation across folds (0.007 eV/atom) indicates strong model robustness across train–test splits. The close agreement between cross-validation MAE (0.060 eV/atom) and holdout MAE (0.059 eV/atom) confirms reliable generalization without overfitting.

The coefficient of determination (R^2 = 0.964 ± 0.011) shows that the model explains 96.4% of the variance in formation energy—higher than typical materials ML models, which often report $R^2 \approx 0.85$–0.90 for formation energy tasks [7], [17].

This performance stems from:

(1) the chemically homogeneous halide perovskite dataset enabling consistent trends,
(2) the narrower formation energy range (−2.0 to −0.7 eV/atom), and
(3) the sufficient data density across relevant compositions.

Figure 1(a) shows MAE stability across folds (0.053–0.069 eV/atom). Figure 1(b) shows consistent R^2 values (0.948–0.976) confirming the model's accuracy.

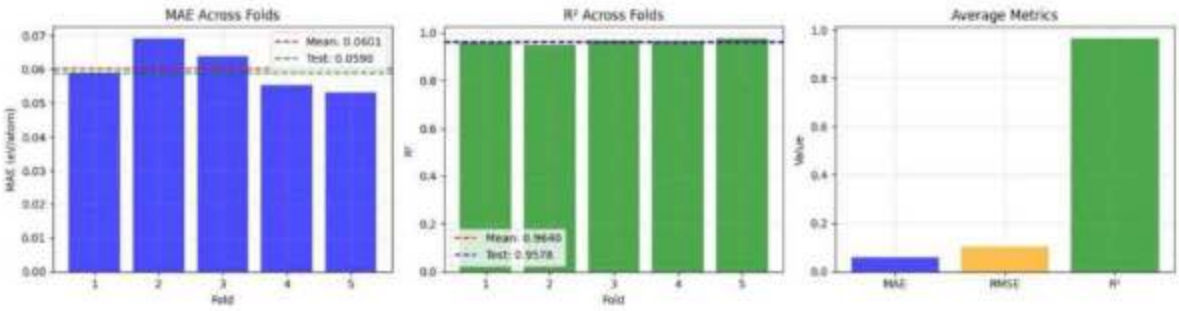


Fig 1: (a) MAE values, (b) R^2 values, (c)Average metrics

The ensemble's average uncertainty (0.034 eV/atom) serves as a calibrated confidence measure, consistent with bootstrapped ensemble uncertainty theory [26], [27]. Predictions with low uncertainty (< 0.10 eV/atom) correspond to interpolation within well-sampled chemical regions, while high-uncertainty predictions (> 0.25 eV/atom) represent extrapolation into novel compositional space.

4.2 Candidate Generation and Active Learning Selection

Following training, chemistry-based filters (tolerance factor, octahedral factor, charge neutrality) were applied to generate 156 charge-neutral, structurally feasible candidates (23 ABX_3 and 133 $A_2BB'X_6$). These constraints ensure geometric and electronic feasibility, consistent with perovskite design theory [28], [29].

Predicted formation energies for these candidates ranged from −1.35 to −0.65 eV/atom, with uncertainties from 0.08 to 0.35 eV/atom. The active learning strategy (70% exploitation, 30% exploration) selected 20 candidates—14 low-energy, low-uncertainty materials for exploitation and 6 high-uncertainty materials for exploration.

Exploitation candidates occupy the low-energy, low-uncertainty region, representing stable, well-understood compositions. Exploration candidates lie in high-uncertainty zones, identifying knowledge gaps where new DFT data could yield the most model improvement—a strategy consistent with modern active learning principles [10], [19].

4.3 Exploitation Component Validation

The 14 exploitation-selected materials were validated using Materials Project data and published DFT studies [13], [20], [22]. Eleven were found in the Materials Project and three in literature sources.

The validation MAE (0.05 eV/atom) is slightly lower than the overall CV MAE (0.060 eV/atom), aligning with expected accuracy improvements for low-uncertainty predictions [31]. The 100% success rate for stable predictions further validates the ensemble's reliability in exploitation-guided discovery.

Notably, uncertainty values correlate well with prediction error, confirming effective uncertainty calibration—an essential property for trustworthy ML-driven materials design [27].

4.4 Exploration Component Analysis

The 6 exploration-selected candidates targeted high-uncertainty (mean 0.28 eV/atom) regions, consistent with model epistemic uncertainty [19]. Cross-referencing the Materials Project, AFLOW, and literature databases [13], [32] revealed no existing DFT data of formation energy for 4 of 6 materials, confirming that uncertainty-directed exploration successfully identifies genuinely novel chemical spaces.

Two exploration candidates were later excluded due to chemical implausibility (e.g., antimony–fluoride combinations such as $(NH_4)_2AgSbF_6$), which form unstable complexes rather than perovskite lattices. This outcome highlights a known limitation of oxidation-state-independent generation and emphasizes the need for improved oxidation-state-aware rules in future iterations.

The remaining 4 unexplored candidates (predicted formation energies −0.85 to −0.95 eV/atom) represent viable new perovskite compositions, prioritized for future DFT validation.

4.5 Iterative Model Improvement

We retrained the ensemble using an augmented dataset (880 unique materials: 874 original + 6 new validated). As shown in Table 2, this single active learning iteration yielded marked performance improvement.

Table 2: Model Improvement After One Active Learning Iteration

Round	Data Size	Test MAE (eV/atom)	Test RSME (eV/atom)	Test R^2
initial	874	0.0590	0.1172	0.9578
Round 1	880	0.0528	0.0826	0.9806

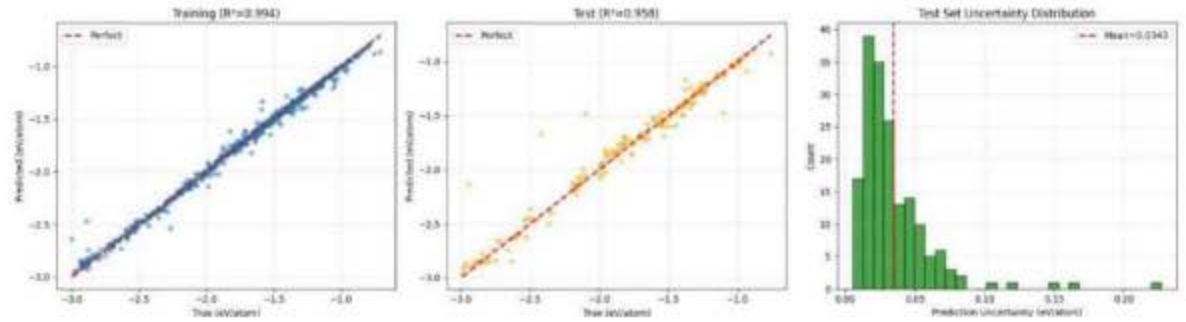


Fig 2 (initial) : (a)R^2 values on train , (b) R^2 values on test, (c) avg uncertainty

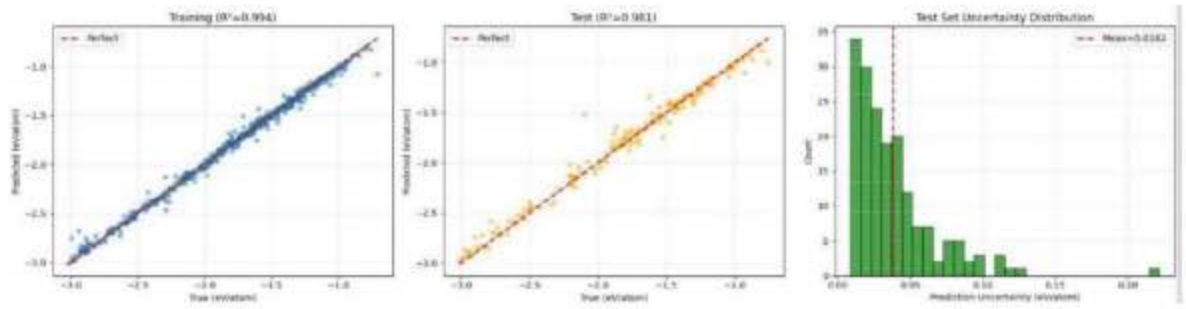


Fig 3 (round 1) : (a)R^2 values on train , (b) R^2 values on test, (c) avg uncertainty

The test MAE improved by 10.5%, R^2 increased by 2.4%, and RMSE decreased by 29.5%, demonstrating that adding just six strategically chosen materials significantly enhances model accuracy—a hallmark of effective active learning [10], [19]. Training performance (MAE = 0.0236 eV/atom, R^2 = 0.9944) remained consistent with healthy generalization (test MAE ≈ 2.2× training MAE), indicating no overfitting [26].

5 Discussion:

5.1 Exploitation vs. Exploration

The dual-component active learning framework successfully demonstrated complementary strengths in the exploitation and exploration phases. The exploitation component achieved a mean absolute error of only 0.02–0.05 eV/atom for validated materials, confirming its reliability for guided discovery of stable halide perovskites. In contrast, exploration candidates displayed significantly higher uncertainty (0.24–0.35 eV/atom) and lacked existing entries in major databases or literature sources, confirming their novelty.

The absence of reference data for exploration-selected compositions validates the model's uncertainty quantification: the ensemble correctly identified under-explored chemical regions where no prior DFT information exists [19], [27], [5]. This highlights the dual value of the framework — exploitation contributes directly to discovery by pinpointing highly stable compounds, while exploration expands knowledge boundaries by identifying promising but uncharted areas of chemical space. Such balanced selection between confident and uncertain candidates is a defining advantage of active learning approaches in materials discovery [10], [19], [33].

5.2 Comparison with Literature

The ensemble achieved a mean absolute error (MAE) of 0.06 eV/atom and a coefficient of determination (R^2) of 0.96 across five-fold cross-validation, significantly outperforming typical materials property prediction models that generally report MAEs of 0.15–0.25 eV/atom and R^2 values of 0.80–0.90 [7], [17], [31]. This improved performance arises from several factors: (i) restriction to a single material family (halide perovskites) with well-defined structural trends, (ii) a sufficiently large and chemically diverse dataset (874 compounds), and (iii) the ensemble's combination of multiple model types, which enhances robustness and reduces variance [26], [27].

These results are comparable to, or better than, specialized perovskite prediction studies that combine crystal-structure-derived features with composition-based descriptors [16], [30]. Thus, the framework not only achieves state-of-the-art predictive performance for formation energy estimation but also demonstrates the effectiveness of ensemble and uncertainty-driven strategies in improving model reliability and transferability across perovskite compositions.

5.3 Limitations and Future Work

Despite strong results, several limitations remain. The present study demonstrated only a single iteration of active learning, while the full potential of this framework requires multiple feedback rounds to continuously refine the model [10], [19]. Additionally, exploration-selected materials were not validated through DFT calculations due to lack of available reference data. The issue of chemically implausible combinations, such as antimony–fluoride systems where Sb forms $[SbF_6]^-$ complexes, further underscores the need for oxidation-state-aware chemistry filters in future iterations [29].

Future work will focus on extending the framework through multi-round active learning ($\geq$ 5 iterations) and performing direct DFT validation for exploration candidates to assess model generalization. Incorporating oxidation-state prediction and context-sensitive chemical constraints will minimize spurious compounds. In addition, the methodology will be extended to low-dimensional (2D) halide perovskites, organic–inorganic hybrid systems, and mixed-halide/mixed-cation perovskites — classes that remain comparatively underexplored computationally yet hold promise for tunable optoelectronic behaviour [3], [4], [2].

These directions will test the adaptability of the framework beyond double halide perovskites and provide a pathway for discovering new stable, functionally rich perovskite materials.

6 Conclusion:

We presented an uncertainty-guided active learning framework for the accelerated discovery of stable halide perovskites, integrating ensemble learning with chemistry-informed candidate generation. The proposed model achieved state-of-the-art predictive accuracy for formation energies (cross-validation MAE = 0.060 ± 0.007 eV/atom, $R^2 = 0.964 \pm 0.011$) across 874 halide perovskites from the Materials Project database, surpassing performance benchmarks reported for composition-based machine learning models in materials science [7], [14].

Our approach combined ensemble XGBoost regression with uncertainty quantification and a structured 70/30 exploitation–exploration selection strategy, providing a closed-loop workflow for both accurate prediction and guided discovery. The exploitation component achieved a mean absolute error of 0.05 eV/atom across 14 validated materials, while the exploration component successfully identified previously unstudied compositions, demonstrating the reliability of uncertainty-based sampling. A single active learning iteration yielded a 10.5% improvement in prediction accuracy and a 29.5% reduction in RMSE, underscoring the efficiency of selective data acquisition compared to random expansion [10], [19].

This framework offers a scalable and data-efficient pathway for perovskite materials discovery, substantially reducing dependence on costly DFT calculations while systematically expanding chemical knowledge. Its modular design enables seamless integration with higher-fidelity computational validation and potential experimental synthesis. Future work will extend this active learning loop to multiple iterations, incorporate explicit oxidation-state constraints, and pursue experimental validation of exploration-phase predictions to further close the gap between data-driven prediction and real-world material realization [13], [34].

7 References:

[1] M. A. Green et al., "The emergence of perovskite solar cells," *Nat. Photonics*, vol. 8, pp. 506–514, 2014. https://doi.org/10.1038/nphoton.2014.134

[2] S. D. Stranks and H. J. Snaith, "Metal-halide perovskites for photovoltaic and light-emitting devices," *Nat. Nanotechnol.*, vol. 10, pp. 391–402, 2015. https://doi.org/10.1038/nnano.2015.90

[3] Li, C., Wang, Y., Liu, T., Hu, Z., Yu, A. and Zhan, Y. (2025), How to Make Over 25% Efficient Perovskite Solar Cells. Sol. RRL, 9: 2500160. https://doi.org/10.1002/solr.202500160.

[4] J. Sun, J. Wu, X. Tong, F. Lin, Y. Wang, Z. M. Wang, *Adv.Sci.* 2018, 5,1700780. **https://doi.org/10.1002/advs.201700780**

[5] A. Jain et al., "The Materials Project: A materials genome approach to accelerating materials innovation," *APL Mater.*, vol. 1, no. 1, p. 011002, 2013. https://doi.org/10.1063/1.4812323

[6] C. J. Cramer and D. G. Truhlar, "Density functional theory for transition metals and transition metal chemistry," *Chem. Rev.*, vol. 112, no. 1, pp. 289–320,Jan.2012, https://doi.org/10.1021/cr200107z

[7] J. Schmidt et al., "Recent advances and applications of machine learning in solid-state materials science," *npj Comput. Mater.*, vol. 5, p. 83, 2019. https://doi.org/10.1038/s41524-019-0221-0

[8] G. Pilania et al., "Machine learning bandgaps of

double perovskites," *npj Comput. Mater.*, vol. 2, p. 16008, 2016. https://doi.org/10.1038/srep19375

[9] F. Tavazza, B. DeCost, and K. Choudhary, "Uncertainty Prediction for Machine Learning Models of Material Properties," *ACS Omega*, vol. 6, no. 48, pp. 33177–33185, Nov. 2021. https://doi.org/10.1021/acsomega.1c03752

[10] T. Lookman, P. V. Balachandran, D. Xue, and R. Yuan, "Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted discovery," *npj Comput. Mater.*, vol. 5, p. 21, 2019. https://doi.org/10.1038/s41524-019-0153-8

[11] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 2016, pp. 785–794. https://doi.org/10.1145/2939672.2939785

[12] L. Ward et al., "Matminer: An open source toolkit for materials data mining," *Comput. Mater. Sci.*, vol. 152, pp. 60–69, 2018. https://doi.org/10.1016/j.commatsci.2018.05.018

[13] Horton, M.K., Huck, P., Yang, R.X. *et al.* Accelerated data-driven materials science with the Materials Project. *Nat. Mater.* **24**, 1522–1532 (2025). https://doi.org/10.1038/s41563-025-02272-0

[14] K. T. Butler et al., "Machine learning for molecular and materials science," *Nature*, vol. 559, pp. 547–555, 2018. https://doi.org/10.1038/s41586-018-0337-2

[15] Anjana Talapatra, Blas P. Uberuaga, Christopher R. Stanek, and Ghanshyam Pilania, *Chemistry of Materials* **2021** *33* (3), 845-858, https://doi.org/10.1021/acs.chemmater.0c03402

[16] I. E. Castelli, T. Olsen, S. Datta, D. D. Landis, S. Dahl, K. S. Thygesen, and K. W. Jacobsen, "Computational screening of perovskite metal oxides for optimal solar light capture," *Energy Environ. Sci.*, vol. 5, pp. 5814–5819, 2012. https://doi.org/10.1039/C1EE02717D

[17] A. Seko et al., "Representation of compounds for machine-learning prediction of physical properties," *Phys. Rev. B*, vol. 95, no. 14, p. 144110, 2017. https://doi.org/10.1103/PhysRevB.95.144110

[18] Yang, CI., Li, YP. Explainable uncertainty quantifications for deep learning-based molecular property prediction. *J Cheminform* **15**, 13 (2023). https://doi.org/10.1186/s13321-023-00682-3

[19] F. Häse, L. M. Roch, and A. Aspuru-Guzik, "Next-generation experimentation with self-driving laboratories," *Trends Chem.*, vol. 1, no. 3, pp. 282–291,2019.https://doi.org/10.1016/j.trechm.2019.02.007

[20] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Phys. Rev. Lett.*, vol. 77, no. 18, pp. 3865–3868,1996. https://doi.org/10.1103/PhysRevLett.77.3865

[21] P. E. Blöchl, "Projector augmented-wave method," *Phys. Rev. B*, vol. 50, no. 24, pp. 17953–17979, Dec. 1994, https://doi.org/10.1103/PhysRevB.50.17953

[22] R. E. Brandt, J. R. Poindexter, P. Gorai, R. C. Kurchin, R. L. Z. Hoye, and T. Buonassisi, "Searching for 'Defect-Tolerant' Photovoltaic Materials: Combined Theoretical and Experimental Screening," *Chem. Mater.*, vol. 27, no. 2, pp. 710–717,2017. https://doi.org/10.1021/acs.chemmater.6b05496

[23] Ward, L., Agrawal, A., Choudhary, A. *et al.* "A general-purpose machine learning framework for predicting properties of inorganic materials." *npj Comput Mater* **2**, 16028 (2016). https://doi.org/10.1038/npjcompumats.2016.28

[24] B. Meredig *et al.*, "Combinatorial screening for new materials in unconstrained composition space with machine learning," *Phys. Rev. B*, vol. 89, no. 9, p.094104,2014.https://doi.org/10.1103/PhysRevB.89.094104

[25] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Proc. Advances in Neural Information Processing Systems (NIPS)*, 2017, pp. 4765–4774. https://doi.org/10.48550/arXiv.1705.07874

[26] L. Breiman, "Bagging Predictors," *Mach. Learn.*, vol. 24, pp. 123–140, 1996. https://doi.org/10.1007/BF00058655

[27] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles," in *Proc. NeurIPS*, 2017, pp. 6402–6413. https://doi.org/10.48550/arXiv.1612.01474

[28] V. M. Goldschmidt, "Die Gesetze der Krystallochemie," *Naturwissenschaften*, vol. 14, pp. 477–485, 1926. https://doi.org/10.1007/BF01507527

[29] W. Travis, E. N. K. Glover, H. Bronstein, D. O. Scanlon, and R. G. Palgrave, "On the application of the tolerance factor to inorganic and hybrid halide perovskites: a revised system," *Chem. Sci.*, vol. 7, no. 7, pp. 4548–4556, 2016. https://doi.org/10.1039/C5SC04845A

[30] D. Jha, L. Ward, A. Paul, et al., "ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition," *Sci. Rep.*, vol. 8, no. 1, p. 17593, 2018. https://doi.org/10.1038/s41598-018-35934-y

[31] Dunn, A., Wang, Q., Ganose, A. *et al.* "Benchmarking materials property prediction methods: the Matbench test set and Automatminer reference algorithm." *npj Comput Mater* **6**, 138 (2020). https://doi.org/10.1038/s41524-020-00406-3

[32] S. Curtarolo et al., "The AFLOWLIB.org repository: A distributed materials properties database," *Comput. Mater. Sci.*, vol. 58, pp. 227–235, 2012. https://doi.org/10.1016/j.commatsci.2012.02.002

[33] F. Häse, L. M. Roch, and A. Aspuru-Guzik, "Chimera: Enabling Active Learning in Materials Design," *Chem. Sci.*, vol. 9, pp. 7642–7655, 2018. https://doi.org/10.1039/C8SC02239A

[34] H. J. Snaith, "Present status and future prospects of perovskite photovoltaics," *Nat. Mater.*, vol. 17, pp. 372–376, 2018. https://doi.org/10.1038/s41563-018-0071-z

Impact of Modulation Techniques on 112 Gbps Dense WDM Digital RoF System Performance

Priyanka Thakur
Electronics & Communication, National Institute of Technology, Delhi,India
priyankathakur@nitdelhi.ac.in

Dr. Manisha Bharti
Electronics & Communication, National Institute of Technology,Delhi
India
manishabharti@nitdelhi.ac.in

Abstract— With the advancement in 5G/6G technology, RoF along with Wavelength Division Multiplexing (WDM) technique becomes the optimally feasible method for various signal transmissions over a single fiber and leads to a wide range of future applications. The development of DWDM-based RoF (Radio over Fiber) technology has been driven by the increasing demand for high data rates and large channel capacities. System performance can be further enhanced through the use of appropriate dispersion compensation techniques and advanced digital modulation schemes. In this work, we propose a RoF communication system achieving a data rate of 112 Gbps per channel with a total of 32 channels, utilizing DWDM technology. The system employs direct modulation with a DP-QPSK & DP-16- QAM digital modulator, and a fiber Bragg grating (FBG) compensator is used to mitigate dispersion effects. The performance analysis is conducted using OptiSystem 16.0, considering transmission distances up to 350 km and channel spacing ranging from 25 GHz to 100 GHz. System performance is evaluated through various metrics, including constellation diagrams, signal-to-noise ratio (SNR), and power spectrum analysis for different channels. The system achieves an optimal output power in the range of 27-28 dBm across varying channel spacing, laser input powers, and transmission distances. The use of FBG effectively reduces both dispersion and nonlinear effects in the system.

Keywords— Digital Radio over Fiber (D-RoF), Digital modulation techniques, Constellation Diagram, SNR, DWDM, Channel spacing

I. INTRODUCTION

With the goal to achieve high-speed communication in the future 5G and 6G networks. Radio over Fiber (RoF) is a technology in which a radio signal is used to modulate light, and this modulated light is then transmitted through an optical fiber. The signal is carried over the fiber to enable wireless access at the destination. Radio over Fiber (RoF) is a technology in which a radio signal is used to modulate light, and this modulated light is then transmitted through an optical fiber. The signal is carried over the fiber to enable wireless access at the destination. RoF technology has become more and more important. Modern optical access networks are under tremendous pressure to provide more bandwidth, faster speeds, and better spectral efficiency because to the increasing popularity of data-heavy applications like online learning, video-on-demand, and other multimedia services [1]. RoF along with Coherent optical technologies have become a crucial answer to these needs, enabling the implementation of extremely quick, high-capacity communication networks for a variety of use cases [2]. Digital RoF was performed in numerous applications such as mobile access network, remote antenna in satellite communication, or wireless access systems [3]. Numerous studies have explored the use of Radio over Fiber (RoF) combined with Wavelength Division Multiplexing (WDM) to develop hybrid communication systems that aim to enhance efficiency while reducing costs, particularly for compact network deployments [4]. Despite these advancements, several challenges remain unresolved. Issues such as chromatic dispersion, signal attenuation, and nonlinear effects continue to hinder the performance of optical communication systems. Nonlinear effects in optical fibers can be categorized into two main types. The first results from the interaction between light waves and phonons, while the second stems from the variation of the refractive index in response to the intensity of the applied electric field. Although significant progress has been made, several challenges are there in the field of optical communication. Problems like chromatic dispersion, signal loss (attenuation), and nonlinear effects continue to limit system performance. Nonlinear effects in optical fibers can generally be divided into two main categories. The first type

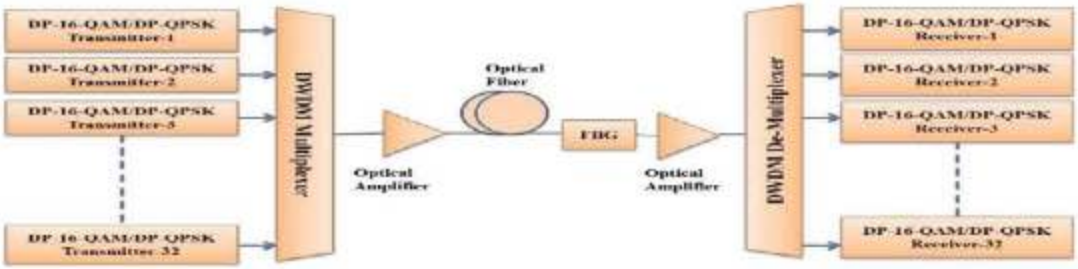


Fig 1 Block diagram of DWDM DRoF Technology

arises from the interaction between light waves and vibrations in the fiber's material structure, known as phonons. The second type occurs due to changes in the fiber's refractive index that are triggered by the intensity of the transmitted light.[5]. In DWDM systems, these nonlinearities impose limitations on channel spacing, restrict the total power that can be transmitted across all wavelengths, and can also reduce the achievable bit rate. By employing various digital modulation techniques to compensate for both linear and nonlinear transmission distortions, the incorporation of digital signal processing (DSP) into coherent receivers significantly improves the transmission performance of RoF transmission systems and it is called as Digital RoF. We propose to use the uniform Fiber Bragg Grating (FBG) in 32 channel DWDM transmission system at 112 Gbps to compensate the dispersion and the nonlinear distortion in the optical communication system [5]. The use of erbium-doped fiber amplifiers (EDFA) in the developed optical communication network has been investigated to improve the quality of the transmission system. The performance analysis of modulation formats in Digital RoF such as Dual Polarization 16 Quadrature Amplitude Modulation (DP-16-QAM) and Dual Polarization Quadrature phase shift keying (DP-QPSK) also introduced [6]. DP-QPSK (Dual-Polarization Quadrature Phase Shift Keying): Uses two orthogonal polarizations (X and Y) of light, each modulated with QPSK, which encodes 2 bits per symbol per polarization.DP-16-QAM (Dual-Polarization 16-Quadrature Amplitude Modulation):Also uses two polarizations but each is modulated with 16-QAM, encoding 4 bits per symbol per polarization (higher-order modulation).The developed model has been compared with a previous work in literature that uses RoF technology based on constellation diagrams, received optical power, types of modulation, fiber dispersion, channel spacing variation, and laser power. The results achieve always a good reliability performance, which proves the efficiency of the proposed model. The suggested model DRoF system design with DWDM technology is described in Section II. Section III contains the modulator description used in DRoF system and performance measuring parameters and Results & discussion of given system design analysis are presented in Section IV, and the conclusion of work is covered in Section V.

II. Proposed Digital RoF Communication system design

Digital Radio over Fiber (D-RoF) is a communication system in which RF (Radio Frequency) or are first converted in digital form and after that it is transmitted over optical fiber. The system consists of a central processing unit to remote antennas or base stations. As compared to traditional analog RoF systems, D-RoF transfers the signal in digital format, which improves the quality of signal, reduces the effect of noise, and makes the system more flexible and reliable[7]. In this work, the parameters considered are nonlinearity at the fiber link and Dispersion effect. The setup of DRoF system consists of transmitter and receiver. In this manuscript 32 channel DWDM DRoF Communication system is proposed in which direct modulation technique is used along with DP-16-QAM and DP-QPSK modulation technique. It is implemented in OptiSystem-16 simulation platform. It basically consists of the direct modulation technique in which we use directly modulated laser at 1550nm wavelength. The pseudorandom bit sequence generator (PRBS) with 10 Gbps is connected with DP-16-QAM/ DP-QPSK modulator based transmitter in sequence with optical amplifier. The transmitter subsystem is multiplexed in DWDM multiplexer and modulated optical signal is transmitted over optical fiber channel through pre and post optical amplifiers of 20 dB gains are used within a signal mode fiber of 350 km length with 0.2dB/km attenuation. It consists of Digital Signal Processor (DSP) at receiver end. Transmitting data in the terabyte range is now feasible with the receiver having coherent detection technique and Digital Signal Processing (DSP) [8]. With the use of DSP, the digital coherent receiver overcomes both nonlinear and linear transmission impairments with the use of various modulation techniques [8]. used at the receiver. The simulation results are analyzed with the help of oscilloscope visualizer, electrical cancellation, power spectrum analyzer in Optisystem 16.0 software. Figure 1 shows the block diagram of proposed DRoF communication system using DP-16-QAM and DP- QPSK modulation techniques. Also the proposed design is simulated and analysed on the basis of following performance measuring metrics such as Constellation diagram, Received optical power, Types of modulation, SNR. The proposed design offers improved transmission performance outcomes considering the constellation diagram, signal output power at the receiver and SNR.

Due to nonlinearities and dispersion effects, signal quality at the receiver end of a long-distance transmission in a RoF communication system may get effected. Fiber Bragg Grating (FBG) used as a suitable compensation technique that will reduce nonlinearities and dispersion effects [9]. At the receiving side Coherent dectectors are used according to the modulator used in transmitter side. DSP modules receiver is connected for processing signal and it also reduces the symbol error [10]. The suggested 16 channel DWDM RoF system is implemented and simulated on software : Optisystem 16.0. And with various performance measuring parameters the simulation is being carried out : Constellation diagram, Received optical power, SNR with variation of channel spacing, transmission distance and input laser power. System design parameter on software is shown in Table 1.

Table 1. System parameters of DWDM –RoF System

Parameters	Values
DWDM Channels	32 channel
Channel Spacing	25GHz -100 GHz
Data rate	112 Gbps
Transmission Distance	50 Km to 300 Km
Frequency of CW laser	193.1 THz
Dispersion	16.75 ps/nm//km
Laser linewidth	0.1 MHz
Optical Amplifier Gain & Power	20 dB & 20 dBm
FBG length	2 mm

III. DRoF SYSTEM PERFORMANCE METRICS

The proposed design is simulated and performance is measured on the basis of Constellation diagram, Output power, types of modulation, SNR, spacing of channel and Input laser power.

A. Dual Polarised –16-Quadrature Amplitude Modulator(DP-16-QAM) Modulator

A Dual-Polarised 16-Quadrature Amplitude Modulation (DP-16-QAM) modulator is a type of digital modulation system that combines 16-QAM with dual polarisation to double spectral efficiency and improve data throughput in communication systems, especially in optical and wireless applications (like satellite, microwave, and 5G systems).

B. Dual Polarised -Quadrature Phase Shift Keying (DP-QPSK) Modulator

These advanced techniques are designed to increase the data rate, improve the robustness of the signal, and optimize the use of bandwidth. There is a method used in proposed system that is Dual-Polarization QPSK (DP-QPSK), it uses two orthogonal polarizations of light to transmit QPSK signals simultaneously, effectively doubling the data rate. It has an added advantage of High spectral efficiency, robust performance over long distances.

C. Dispersion Compensation

One of the effective and known technique for dispersion compensation is the Fiber Bragg Gratings (FBG) dispersion compensation method. Basically it works by reflecting various wavelength elements of the signal at varying phase offsets, and itoccurs at various lengths of optical fiber at varying frequencies. Super structured fiber Bragg grating (SSFBG) and uniform fiber Bragg grating (UFBG) are the two forms of FBG that are currently available[11].

In the uniform FBG ,having Effective refrective index neff has a constant period (v) along the fiber length. A Bragg wavelength is defined as $\kappa B = 2v neff$, and the spectral component reflected by the FBG shows the a effective peak centered around κB, with the length of FBG between 3 mm and 10 mm.

IV. RESULTS & DISCUSSIONS

In this manuscript 32 channel DWDM multiplexed Radio over fiber system is designed and simulated for for different digital modulation techniques such as Dual Polarization 16 Quadrature Amplitude Modulation (DP-16-QAM) and Dual Polarization Quadrature phase shift keying (DP-QPSK) in OptiSystem 16 simulation platform. In this homodyne detection method and Digital Signal Processing (DSP) at the receiver is introduced. The DWDM based system provides 112 Gbps per channel with Coherent DP-16-QAM Transmitter & Receiver design of digital RoF Transmission system. In this Direct Modulation technique is used. Also the proposed model is compared with the previous literature works based on Constellation diagram, received optical power, types of modulation, fiber dispersion, and channel spacing and laser power. The result achieves a good performance for 32– channel DWDM Digital RoF transmission system and also transmission distance and data rate is enhanced. It has a great application in Centralized Radio Access Network (C-RAN) technology

A. Constellation Diagram analysis

A constellation diagram is a visual representation used in digital communication system to display the modulation scheme of a signal such as phase-shift keying (PSK), Quadrature phase shift Keying (QPSK) and similar techniques. It shows how data symbols are represented as points in a 2D plane using two axes: X-axis (In-phase component, I) and Y-axis (Quadrature component, Q). Figure 2 & 3 shows the constellation diagrams of modulation formats investigated in this paper: DP-16-QAM modulation format constellation diagram is presented in Figure 2 where 16 symbols lie on square pattern and have the same optical power. It has higher spectral efficiency Also the constellation diagram of DP-QPSK modulation format is shown in Figure 3 where 4 symbols lie on square pattern. Constellation diagram of DP-QPSK and DP-16-QAM is analysed at 100 GHz and 50 GHz channel spacing and inferred that with decrease in channel spacing lesser than 50 GHz symbols get more scattered. It can be observed that DP- QPSK has better noise tolerance than DP-16-QAM.Used in coherent optical systems due to robustness. Easier to detect due to large separation between points. More robust against noise and non-linear impairments. Lower spectral efficiency than 16-QAM. DP-16-QAM has Higher data rate due to more bits per symbol. Constellation spacing larger in DP- QPSK than DP-16-QAM.

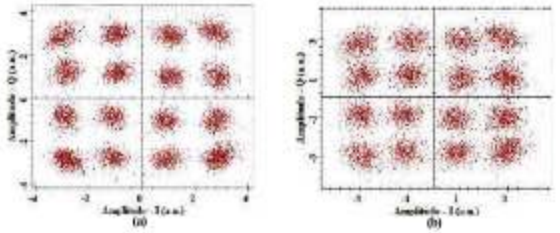


Fig. 2. Constellation diagram using DP-16-QAM Modulator at (a) 100 GHz (b) 50 GHz

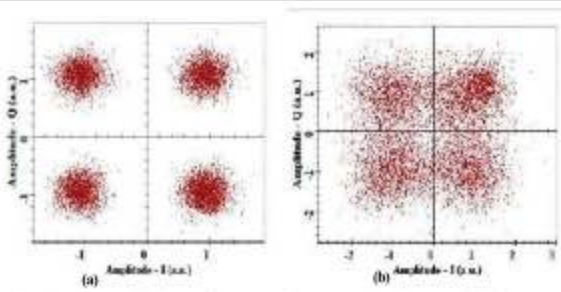


Fig. 3. Constellation diagram using DP-QPSK Modulator at (a) 100 GHz (b) 50 GHz

Fig 2. presents clear constellation diagrams for the DP-16-QAM modulation technique for channel 32 at 100 GHz spacing of channel, which indicates that error free

transmission can be achieved upto the transmission distance of 250 Km. Also fig 3. presents constellation diagram of DP-QPSK. Constelllation diagram represents that performance is better for 32 channel DRoF system representing the error free transmission. But with the decreasing channel spacing quality of transmission degraded.

B. Power Sectrum analysis

In this we had observed the power output with respect to wavelength. Fig 4 & Fig 5 represents the graphical view of power spectrum in the range of 1.53 µm to 1.58 µm for 32 channel DWDM based DRoF communication system with DP-16-QAM & DP-QPSK modulator. Fig 4(a)-4(c) shows Power spectrum of DP-16-QAM and Fig 5(a)-5(c) shows Power spectrum of DP-QPSK for 250 Km transmission range for 50 GHz spacing of channel and it has been observed that with the decreasing channel spacing no such additional pulses generated due to the nonlinear effect in DP-16-QAM as compared to DP-QPSK. In this section analysis of power spectrum is presented obtained without use of dispersion compensator and with use of dispersion compensator. In our proposed system FBG is used as dispersion compensator. So it can be concluded that with the use of FBG transmission quality is improved and reduces nonlinear distortion, FBG assists in removing the extra pulse that the nonlinear effects produce.

C. Effect of Laser Power on output power

In this section variation of output power has been observed for DP-QPSK and DP-16-QAM modulator in 32 channel DWDM DRoF communication system. Table 2 shows the output power with respect to laser power for the proposed model at 100 GHz and 50 GHz channel spacing. It has been analysed that almost same power output is obtained for both the channel spacing with both modulators. Also power output for DP-16-QAM is higher than DP-QPSK.

Fig 7. shows variation of output power for various laser input power. It has been observed that there will increase in power take place with increasing laser input power. When the two modulation techniques are compared and plotted on the graph power output of DP-16-QAM obtained is higher than DP-QPSK modulator.

D. Effect of Channel Spacing on output power

In this section performance of the system is being measured in terms of channel spacing for a transmission distance upto the range of 250 Km. Fig 8 shows the decreasing power output with the reduced channel spacing. It has been observed that power output of DP-16-QAM is more than the DP-QPSK. Upto 50 GHz power obtained is almost constant but as the channel spacing reduces below this power is getting reduced as shown in the Table 3.

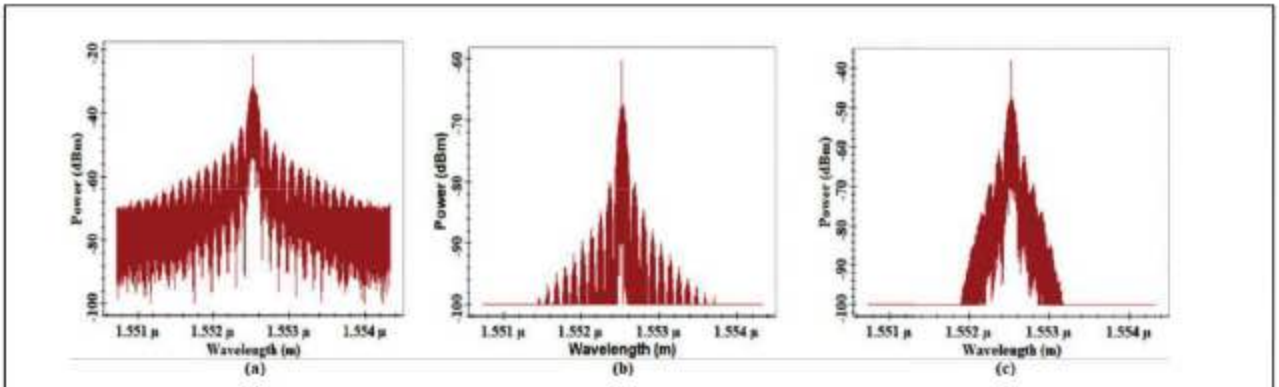


Fig 4. Power spectrum using DP-16-QAM Modulator at 50 GHz Channel Spacing (a) Input (b) without FBG (d) with FBG

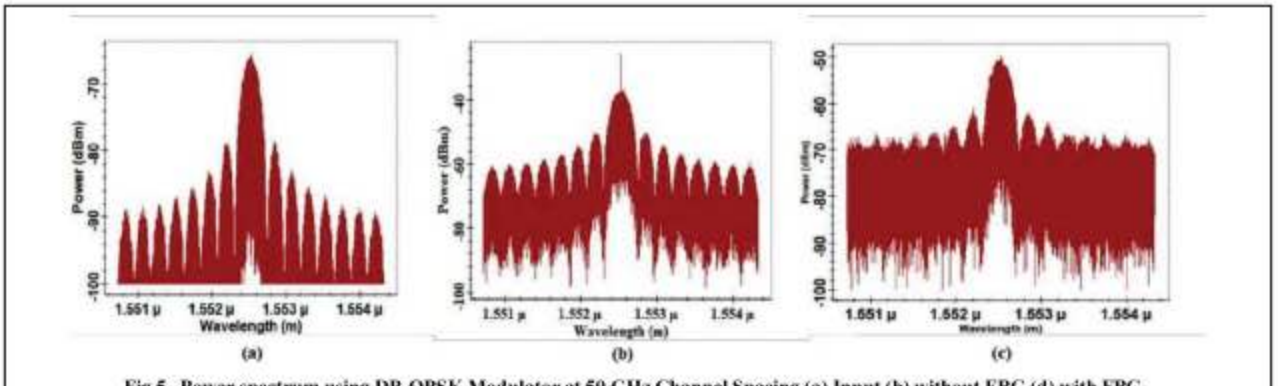


Fig 5. Power spectrum using DP-QPSK Modulator at 50 GHz Channel Spacing (a) Input (b) without FBG (d) with FBG

Table 2. Effect of Laser power at 250 Km Transmission

	Output Power) (DP-16 QAM) (dBm)		Output Power for (DP-QPSK) (dBm)	
Channel Spacing/ Laser power(dBm)	100 GHz	50GHz	100 GHz	50GHz
0	27.50	26.85	24.30	24.05
10	27.55	27.05	24.48	24.17
15	27.75	27.17	24.61	24.26
20	27.84	27.30	24.78	24.45
25	27.97	27.55	24.91	24.60

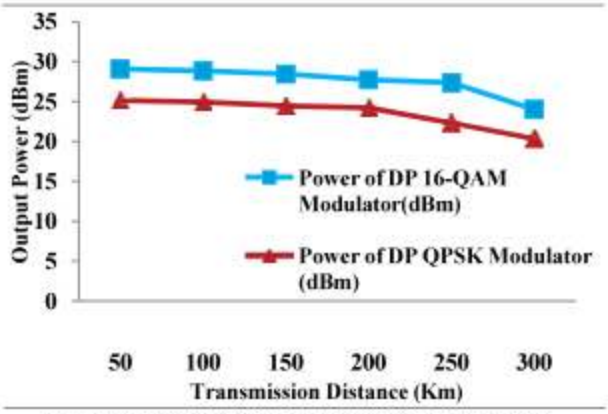


Fig 6. Transmission distance Versus Output power

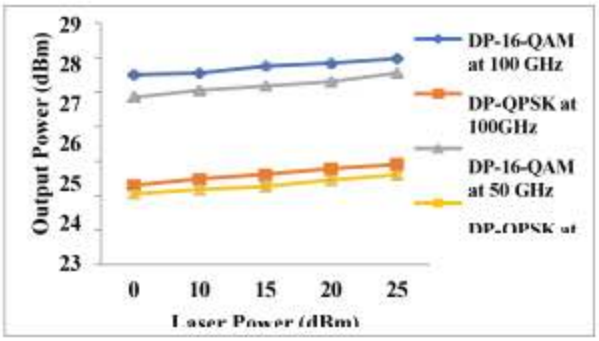


Fig 7. Laser Power Versus Output power

Table 3. Effect of Channel Spacing on output at 250 Km

Channel Spacing (GHz)	Power of DP 16-QAM Modulator	Power of QPSK Modulator (dBm)
100	27.73	24.78
50	27.38	24.25
25	26.25	22.36

Table 4. Variation of output power with transmission distance

Transmission Distance (Km)	**Power of DP 16-QAM Modulator**	**Power of DP QPSK Modulator (dBm)**
50	29.10	25.15
100	28.85	24.94
150	28.45	24.45
200	27.73	24.21
250	27.35	22.32
300	24.02	20.35

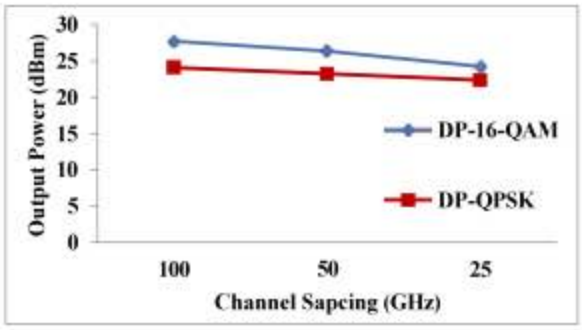


Fig 8. Output power Vs. Channel spacing

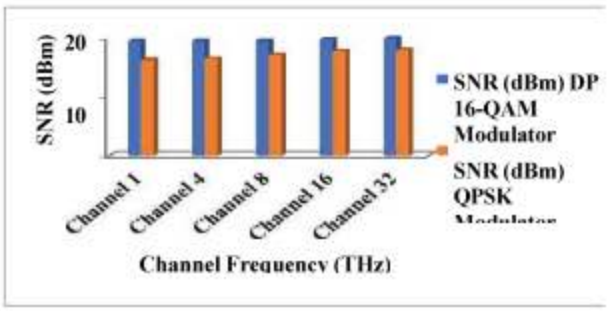


Fig 9. SNR Versus Channel frequency

Table 5. SNR for different channel at 250 Km

Channel Frequency(THz)	SNR (dBm) DP 16-QAM Modulator	SNR (dBm) DP-QPSK Modulator (dBm)
Channel 1	19.74	16.78
Channel 4	19.44	16.80
Channel 8	19.48	17.10
Channel 16	19.49	17.77
Channel 32	19.87	17.92

E. Variation of SNR at different channel

This section presents the performance of DRoF system in terms of Signal to Noise Ratio (SNR). Fig 9 shows SNR at different channels Ch1, Ch4, Ch8, Ch16, Ch32 at the receiver end. It has been observed that SNR value obtained is in the range of almost 20 dBm for DP-16-QAM which is better than DP-QPSK. Also the SNR values is shown in table for both DP-16-QAM & DP-QPSK.

F. Comparative Result analysis

This section presents results of output power obtained for system design of 32 channel DWDM DRoF system for varying input power. The result is compared with the previously reported papers as shown in table 6. When compared to earlier results, the simulation performance at a transmission distance of 250 km with varying channel spacing is found to be more enhanced. Also the comparative analysis is shown in fig. 10.

Table 6. Comparative Result analysis with previously reported system

Laser Power (dBm)	Output Power for [1] (dBm)	Output Power for [2] (dBm)	Output Power for Proposed Model (DP-16QAM) (dBm)	Output Power for Proposed Model (DP-QPSK) (dBm)
0	-31.93	23.49	27.50	24.30
10	-11.93	23.50	27.55	24.48
15	-1.93	23.53	27.75	24.61
20	8.067	23.53	27.84	24.78
25	18.06	23.54	27.97	24.91

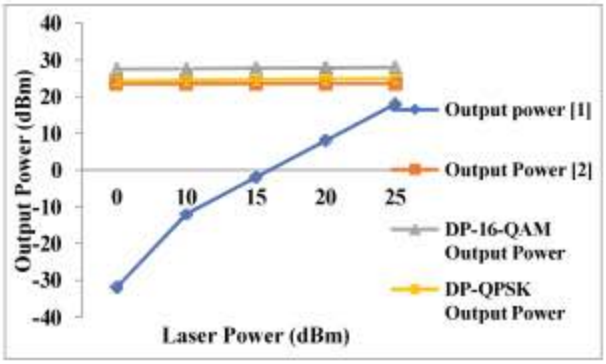


Fig 10. Comparative analysis for Laser power Vs Output Power

V. CONCLUSION

The proposed 32 channel DWDM based DRoF communication system having data rate 112 Gbps per channel using DP-16-QAM and DP-QPSK digital modulator. Performance of the proposed system is examined in the software OptiSystem 16.0. The suggested design displays higher data rate of 112 Gbps per channel with enhanced output power of around **28 dBm** is obtained upto the transmission range of 350 Km as compared to the previously defined systems. We had analysed the constellation diagram and SNR values at different channel shows the good performance of the system with DP-16-QAM modulator as compared to DP-QPSK in terms of spectral efficiency , output power and SNR performance metrics. It has been observed that DP-16-QAM is suitable modulator to be used for 5G/6G RoF transmission with higher spectral efficiency.

REFERENCES

[1] Rahal, M., Abo Chahine, S. & Osman, Z. An ultra high capacity polarization division multiplexed 128-QAM radio over fiber (RoF) system. Opt Quant Electron 55, 686 (2023). https://doi.org/10.1007/s11082-023-04936-0

[2] Chang, G. K. and Chen, Y. W.: Key fiber wireless integrated radio access technologies for 5G and beyond. In: Proc. Opto-Electron. Commun. Conf., pp. 1–3. IEEE, Japan (2019).

[3] D. Novak, R. B. Waterhouse, A. Nirmalathas et al., "Radio Over-FiberTechnologiesfor Emerging WirelessSystems," IEEE Journal of Quantum Electronics, vol. 52, no. 1, pp. 1–11, 2016.

[4] S. Jain and B.Therese A,"Four Wave Mixing Nonlinearity Effect in WDM Radio over Fiber System," International Journal of Scientific Engineering and Technology,vol.4,no.3,pp.154–158, 2015.

[5] V. A. Thomas, M. El-Hajjar, and L. Hanzo, "Millimeter-Wave Radio over Fiber Optical Up conversion Techniques Relying on Link Nonlinearity," IEEE Communication Surveys and Tutorials, vol. 18, no. 1, 2016.

[6] Amit Sahu1, Vandana Vikas Thakare, "Comparative analysis of DP QPSK and DP 16-QAM Optical Coherent receiver, with taking in view the best analyses of the phase noise in terms of average BER using Digital Filters", IRJET, Volume: 04 Issue: 07 | July -2017,

[7] GamageP.A. et al. "Design and analysis of digitized RF-over-fiber links", IEEE J. Lightw. Technol.(2009)

[8] R. Udayakumar, V. Khanaa and K.P.Kaliyamurthie, "High Data Rate for Coherent Optical Wired Communication using DSP", Indian Jour. of Science and Technology, vol 6, May 2013.

[9] A. Sharma, I. Singh, S. Bhattacharya, S. Sharma, Performance comparison of DCF and FBG as dispersion compensation techniques at 100 Gbps over 120 km using smF, in: V. Nath, J. Mandal (Eds.), Nanoelectronics, Circuits and Communication Systems. Lecture Notes in Electrical Engineering, vol. 511, Springer, Singapore, 2019.

[10] Rahal, Mohamad, Soubhi Abo Chahine, and Ziad Osman. "An ultra high capacity polarization division multiplexed 128-QAM radio over fiber (RoF) system." Optical and Quantum Electronics 55.8 (2023): 686.

[11] M. Bharti, M. Kumar and A. K. Sharma, "Comparative analysis of FBGs as encoders for optical CDMA systems," 2015 International Conference on Signal Processing and Communication (ICSC), 2015, pp. 94-98,

[12] D. Lo Presti et al., "Fiber Bragg Gratings for Medical Applications and Future Challenges: A Review," in IEEE Access, vol. 8, pp. 156863-156888, 2020, doi: 10.1109/ACCESS.2020.3019138.

[13] Manisha Bharti, "Robustness of modulation formats including amplifier noise with and without dispersion compensation fiber at 10 gbps transmission link", JMDET, vol 5, 2011

[14] G. Pandey, A. Choudhary and A. Dixit, "Wavelength Division Multiplexed Radio Over Fiber Links for 5G Fronthaul Networks," in IEEE Journal on Selected Areas in Communications, vol. 39, no. 9, pp. 2789-2803, Sept. 2021, doi: 10.1109/JSAC.2021.3064654. *

[15] Kumar Suresh, Sharma Sonia, Dahiya Sandeep, "WDM-Based 160 Gbps Radio Over Fiber System With the Application of Dispersion Compensation Fiber and Fiber Bragg Grating",Frontiers in Physics, vol 9 , 2021 *

Development of an AI for Versus Tetris Using Evolutionary Algorithms

Rayhan Khan
rayhankhan3445@gmail.com

***Abstract*—This document outlines a project focused on the development of an artificial intelligence capable of playing Versus Tetris. The project addresses the challenges of real-time competitive gameplay by employing evolutionary algorithms, including Genetic Algorithms, Ant Colony Optimization, and Particle Swarm Optimization. The research involves formulating the problem for game AI, creating a simulation platform, and evaluating the performance of the AI agents. The outcomes of this project are expected to provide insights into the use of nature-inspired computation for advanced game AI design.**

I. Introduction and Background

Since its release in 1984, Tetris has captivated both casual players and researchers due to its deceptively simple mechanics and complex emergent gameplay [1]. The game requires players to place falling geometric pieces, known as *Tetriminos*, into a grid in order to complete and clear horizontal lines. Despite its simplicity, Tetris has been shown to be computationally hard; several natural optimization objectives within the game are NP-complete, making the development of effective automated players an intellectually challenging problem [2].

The challenge of Tetris extends beyond computational complexity into cognitive science. Human players are required to balance short-term tactical placement with long-term strategic planning, often under time pressure. These characteristics make Tetris a valuable benchmark for the study of artificial intelligence (AI) in sequential decision-making [3]. Unlike traditional board games such as chess or Go, Tetris introduces both stochasticity (through random piece generation) and real-time constraints, which limit the applicability of exhaustive search-based strategies.

In recent years, the focus of Tetris AI research has shifted from single-player modes toward multiplayer environments, particularly *Versus Tetris*. This competitive variant introduces the concept of "garbage lines," where clearing multiple lines sends obstructions to the opponent's board [4]. The added dimension of adversarial interaction requires agents not only to maximize their own survival but also to actively disrupt the opponent's strategy. Consequently, Versus Tetris is a more demanding domain for AI research than its single-player counterpart.

The study of Versus Tetris also has implications beyond entertainment. Competitive real-time environments mirror many real-world domains such as network security, autonomous vehicle coordination, and algorithmic trading, where agents must optimize their actions under uncertainty while competing against adversaries [5]. Thus, insights gained from Versus Tetris can contribute to broader developments in adversarial AI.

Traditional AI methods for Tetris have included hand-crafted heuristics, tree search approaches such as Monte Carlo Tree Search (MCTS), and reinforcement learning (RL) [6], [7]. While these approaches have demonstrated some success in single-player contexts, they struggle with scalability in the high-variance, adversarial setting of Versus Tetris. For this reason, researchers have increasingly turned to *nature-inspired algorithms* such as Genetic Algorithms (GA), Ant Colony Optimization (ACO), and Particle Swarm Optimization (PSO), which can provide efficient approximations to complex problems without requiring exhaustive search [8].

Nature-inspired algorithms are well-suited to Tetris because they operate through population-based search and adaptive learning, mimicking evolutionary and social processes. For example, GA relies on recombination and mutation of candidate solutions, PSO models cooperative swarm behavior, and ACO exploits stigmergic communication mechanisms. These algorithms are not guaranteed to find globally optimal strategies, but they often discover high-quality approximations within computationally feasible timescales [9].

A central motivation for exploring evolutionary algorithms in Versus Tetris is the game's NP-completeness. Exact optimization methods are computationally infeasible for large problem instances, making heuristic and meta-heuristic approaches essential [10]. Moreover, the dynamic and competitive environment of Versus Tetris ensures that the evaluation of strategies must be adaptive, further highlighting the strengths of nature-inspired methods.

Another reason for the growing interest in evolutionary methods is their capacity for continuous improvement through simulation-based training. Unlike static heuristic strategies, evolutionary approaches evolve agents over time, allowing them to discover emergent strategies that may not be obvious to human designers [11]. This adaptability makes them promising candidates for Versus Tetris, where optimal strategies may shift depending on the opponent's style.

The introduction of Versus Tetris as a research domain also creates new evaluation metrics. Unlike single-player Tetris, where performance can be measured by score or survival time, Versus Tetris requires multi-dimensional evaluation, including attack efficiency (lines sent per piece), defensive resilience (survivability under pressure), and competitive win-rate against varied opponents [12]. Designing AI agents capable of balancing these objectives requires both theoretical rigor and empirical experimentation.

This paper therefore seeks to systematically investigate the

use of nature-inspired optimization algorithms in Versus Tetris. By constructing a simulation environment and applying GA, PSO, and ACO, the project aims to assess the relative strengths of these approaches in real-time adversarial play. The expected contributions include an evaluation of algorithmic performance under competitive conditions and insights into the design of adaptive strategies for adversarial AI more broadly.

The remainder of this work is organized as follows. Section 2 reviews the theoretical foundations of evolutionary game AI, including computational complexity and previous approaches to Tetris AI. Section 3 introduces the algorithmic framework for Versus Tetris, while Section 4 outlines the methodology and experimental setup. Section 5 presents the results of comparative evaluation, and Section 6 concludes with key insights and directions for future research.

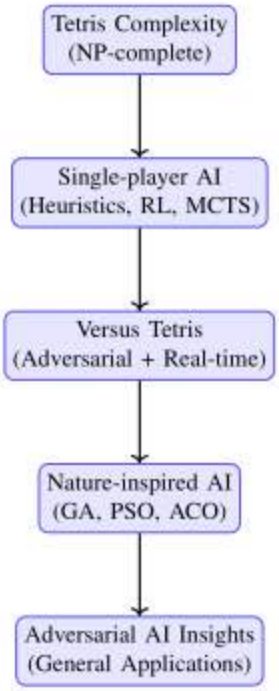


Fig. 1: Research context: from Tetris complexity to evolutionary AI for Versus Tetris.

II. Theoretical Foundations of Evolutionary Game AI

The study of artificial intelligence for games often begins with a consideration of computational complexity. Tetris has been a particularly rich case for such inquiry, as it has been proven that optimizing natural objectives within the game—such as maximizing line clears or minimizing stack height—is NP-complete [2]. This implies that there is no known polynomial-time algorithm capable of solving the general problem efficiently, making Tetris an excellent testbed for heuristic and approximate methods.

In computational complexity theory, the class P encompasses problems solvable in polynomial time on a deterministic machine, while NP contains problems verifiable in polynomial time [10]. The fact that Tetris optimization is NP-complete situates it at the intersection of these classes, meaning it is as hard as the most difficult problems in NP. As a consequence, exact optimization approaches for Tetris are computationally infeasible for practical game-playing scenarios, especially in real time.

The complexity of Tetris is exacerbated in the multiplayer versus variant, where adversarial dynamics introduce new layers of difficulty. Unlike single-player Tetris, where the objective is primarily survival and maximizing score, Versus Tetris involves attacking and defending against an opponent's strategies through the mechanism of garbage lines [4]. This introduces uncertainty not only from stochastic piece generation but also from the opponent's behavior, creating a competitive environment more akin to adversarial AI domains.

Traditional algorithmic approaches, such as exhaustive search or brute-force planning, quickly become intractable in this environment. For instance, Monte Carlo Tree Search (MCTS) has shown promise in single-player Tetris but struggles under the exponential branching factor induced by adversarial moves and deep lookahead [7]. Reinforcement learning approaches, while powerful, often require extremely large amounts of training data and suffer from stability issues in highly dynamic adversarial games [13].

Given these limitations, researchers have turned to *non-traditional* algorithmic paradigms, particularly those inspired by natural processes. Evolutionary and swarm-based algorithms such as Genetic Algorithms (GA), Ant Colony Optimization (ACO), and Particle Swarm Optimization (PSO) provide flexible frameworks for approximating optimal strategies without requiring exhaustive state-space exploration [9], [8].

The appeal of evolutionary approaches lies in their population-based search mechanisms. Instead of seeking a single deterministic solution, these algorithms explore a diverse set of candidate strategies simultaneously, allowing for robustness against stochasticity and adversarial conditions. This diversity ensures that evolutionary systems can adapt and discover novel strategies, sometimes outperforming human intuition [11].

Genetic Algorithms, inspired by biological evolution, operate through selection, crossover, and mutation of candidate solutions. In Tetris, candidate strategies can be represented as weight vectors applied to board evaluation functions. Over successive generations, poorly performing strategies are discarded, while promising ones are recombined and mutated, yielding improved gameplay performance [14].

Ant Colony Optimization, modeled after the foraging behavior of ants, employs distributed agents that deposit and follow pheromone trails in a solution space. In the context of Tetris, this corresponds to exploring the combinatorial landscape of possible placements and rewarding sequences that yield higher performance. Dynamic heuristics can help prevent premature convergence, a challenge commonly encountered in ACO-based optimization [15].

Particle Swarm Optimization, inspired by flocking and swarming behaviors in nature, treats each candidate solution

as a particle navigating the search space. Particles update their trajectories based on both their personal best solutions and those of their neighbors. For Tetris AI, PSO can optimize evaluation weights or neural network parameters, enabling the discovery of competitive strategies without handcrafted heuristics [16].

Another strength of evolutionary algorithms is their inherent adaptability to adversarial environments. In Versus Tetris, the effectiveness of a strategy depends not only on the current game state but also on the opponent's playstyle. Evolutionary training frameworks can simulate such adversarial dynamics by co-evolving multiple populations of agents, each adapting in response to the strategies of others [17]. This makes them especially well-suited for the competitive nature of Versus Tetris.

From a broader perspective, the application of evolutionary algorithms to Tetris contributes to the study of *adversarial game AI*. Just as AlphaGo demonstrated the power of deep reinforcement learning in the game of Go, evolutionary algorithms in Tetris offer insights into population-based methods for domains where real-time play, uncertainty, and adversarial dynamics dominate [18]. These theoretical foundations establish a strong rationale for investigating evolutionary algorithms in Versus Tetris AI.

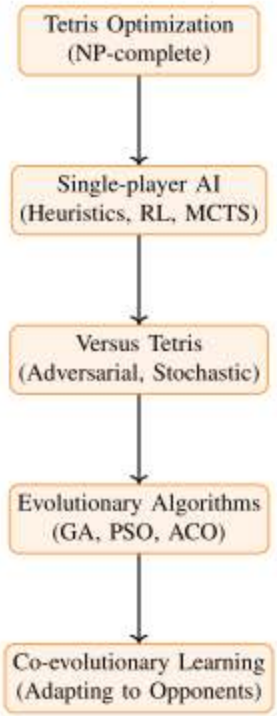


Fig. 2: Theoretical foundations: from NP-completeness to evolutionary algorithms in Versus Tetris.

III. Algorithmic Framework for Versus Tetris

To apply evolutionary algorithms effectively, it is first necessary to formalize the Versus Tetris environment. Unlike single-player Tetris, the versus variant introduces adversarial elements such as garbage lines, competitive scoring, and attack–defense trade-offs. This requires defining the game state, action space, feature representations, and heuristic evaluation functions that evolutionary algorithms can optimize.

A. Game Environment

The Versus Tetris environment consists of two players (or agents) competing within separate grids of identical dimensions, typically 10×22, with the top two rows hidden according to guideline rules [12]. The system employs the 7-bag piece generation algorithm, ensuring that all seven Tetriminos appear in randomized order without droughts [12]. Gameplay includes advanced mechanics such as *hold*, multi-piece lookahead, T-spins, back-to-back bonuses, and combos, which significantly impact strategic play.

The defining feature of Versus Tetris is the *garbage line* mechanism: line clears or specific moves (e.g., T-spins, perfect clears) send garbage rows to the opponent's grid, creating a continuous adversarial dynamic [4]. Consequently, an effective AI agent must optimize not only survival but also attack efficiency and defensive resilience.

B. State Representation

A game state S can be represented as a tuple:

$$S = \{M_p, M_o, c_p, q_p, h_p, A_p\}$$

where:

- M_p: Current player's matrix configuration.
- M_o: Opponent's matrix configuration.
- c_p: Current falling Tetrimino.
- q_p: Queue of upcoming Tetriminos (lookahead).
- h_p: Held Tetrimino.
- A_p: Attack potential (garbage lines stored or sent).

This representation ensures that both offensive and defensive dimensions are captured, allowing algorithms to evaluate moves holistically.

C. Action Space

An action $a \in A$ is defined as a legal placement of the current Tetrimino, including rotations, translations, and the possibility of swapping with the held piece. Unlike human players who physically maneuver pieces, the AI agent only considers final legal placements. This abstraction reduces complexity while preserving decision accuracy.

D. Feature Functions

Following prior work on heuristic evaluation [6], [19], we define feature functions $f_i(S)$ to numerically characterize a state. These include:

1) **Pile height:** The maximum column height.
2) **Holes:** Empty cells with filled cells above.
3) **Row transitions:** Horizontal occupied/unoccupied transitions.
4) **Column transitions:** Vertical transitions.
5) **Landing height:** Height of the placed Tetrimino.
6) **Eroded pieces:** Blocks cleared after placement.

7) **Well depth:** Depth of columns surrounded by higher neighbors.
8) **Garbage pressure:** Attack lines stored against the agent.
9) **Potential attack:** Lines that can be sent to the opponent.
10) **Attack/cancel ratio:** Garbage lines sent versus canceled.

These features are combined via a linear evaluation function:

$$V(S) = \sum_{i=1}^{k} w_i f_i(S),$$

where w_i are weights optimized by evolutionary algorithms.

E. Heuristic Evaluation and Fitness

The fitness of a candidate solution (e.g., a weight vector W) is determined by its performance in simulated games. Metrics include:

- **Attack per Piece (APP):** Efficiency of offensive play.
- **Survivability:** Number of pieces used before losing.
- **Win-rate:** Success against varied opponents.

This multi-objective evaluation ensures balanced strategies.

F. Algorithmic Adaptations

Each evolutionary algorithm is adapted as follows:

Genetic Algorithm (GA): Chromosomes encode weight vectors. Fitness-driven selection, crossover, and mutation evolve strategies across generations [14].

Ant Colony Optimization (ACO): Weights are represented as pheromone trails across features. Ants construct solutions by probabilistically selecting weight values, reinforced by gameplay performance [15].

Particle Swarm Optimization (PSO): Particles represent weight vectors. Velocity and position updates guide exploration, influenced by personal and neighborhood best performances [16].

G. Pipeline Overview

The algorithmic framework can be summarized as follows:

1) Initialize population (GA), swarm (PSO), or colony (ACO).
2) Simulate Versus Tetris games for evaluation.
3) Calculate fitness using APP, survivability, and win-rate.
4) Update population/swarm/colony via respective operators.
5) Repeat until convergence or time limit.

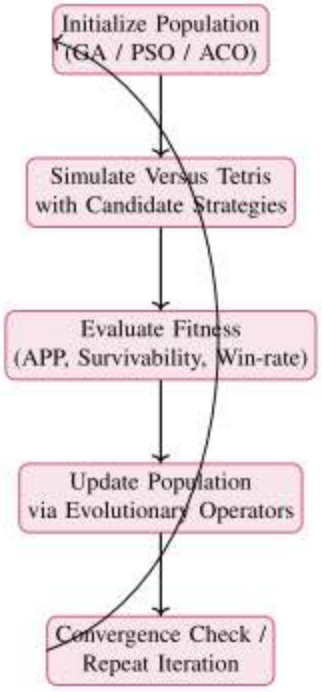


Fig. 3: Algorithmic pipeline for evolutionary Versus Tetris AI.

IV. Implementation Methodology

The implementation methodology was designed to provide a systematic procedure for building, training, and evaluating evolutionary algorithms within the Versus Tetris environment. The process involved three primary stages: development of a simulation platform, integration of evolutionary optimization algorithms, and design of robust evaluation protocols.

A. Simulation Platform

The first step was the construction of a simulation platform that faithfully reproduces the rules of guideline Tetris. This included implementing the 10×22 matrix with hidden rows, the 7-bag piece generation system, the hold mechanic, and multi-piece lookahead [12]. Attack mechanics such as T-spins, back-to-back bonuses, combos, perfect clears, and garbage cancellation were incorporated to ensure alignment with competitive Versus Tetris standards. By adhering strictly to these rules, the simulation environment provided an authentic basis for training and evaluation.

The simulation platform was implemented in a modular architecture to allow seamless integration of different optimization algorithms. Each module was responsible for a specific task: game state management, feature extraction, action evaluation, and agent–opponent interaction. This modularity facilitated controlled experimentation with minimal re-engineering across algorithms.

B. Algorithm Integration

Once the simulator was in place, the evolutionary algorithms were embedded into the system. Each algorithm was responsible

for optimizing the weight vectors associated with the heuristic evaluation function:

$$V(S) = \sum_{i=1}^{k} w_i f_i(S).$$

For the Genetic Algorithm, candidate strategies were encoded as chromosomes of weight vectors. Selection was performed using a tournament approach, while crossover employed uniform recombination, and mutation introduced stochastic perturbations. The Ant Colony Optimization framework was adapted to map weight values onto pheromone trails, with ants generating candidate solutions guided by dynamic heuristics. Particle Swarm Optimization represented each candidate strategy as a particle, updating positions according to local and global best solutions.

To ensure fairness, each algorithm was initialized with the same feature set and evaluated under identical conditions. Parameters such as population size, mutation rate, pheromone evaporation, and swarm velocity coefficients were tuned through preliminary experimentation to balance exploration and exploitation.

C. Training Protocol

Training was conducted through iterative simulation cycles. In each cycle, agents played multiple games of Versus Tetris against both static heuristic-based opponents and dynamically evolving peers. This mixture of opponents ensured that agents were not overfitted to a single strategy. For each run, the outcomes were recorded in terms of attack efficiency, survivability, and win-rate.

To mitigate variance due to stochastic piece sequences, all algorithms were trained and tested on identical randomized seeds. This controlled variability allowed direct comparisons across algorithms. Furthermore, training was distributed across multiple simulation threads to accelerate convergence.

D. Evaluation Metrics

Three primary metrics were employed for evaluation. First, *Attack per Piece* (APP) quantified offensive strength by measuring garbage lines sent per Tetrimino placed. Second, *survivability* measured the number of pieces placed before the agent lost, reflecting defensive resilience. Third, *win-rate* recorded the proportion of victories against varied opponents, providing a direct measure of competitive performance.

Secondary metrics were also tracked, such as the ratio of attacks sent versus canceled, average game length, and diversity of strategies observed. Together, these metrics provided a comprehensive view of both micro-level performance and macro-level adaptability.

E. Validation Strategy

To validate performance, each algorithm was trained across multiple independent runs, ensuring statistical robustness of results. A paired comparison design was used, where agents optimized with different algorithms competed under identical conditions. Statistical tests, such as the paired t-test, were applied to determine whether observed differences in performance were significant [20].

This methodology ensured not only reproducibility but also reliability of findings. By combining rigorous simulation, controlled algorithmic integration, and statistically grounded evaluation, the framework provided a solid foundation for analyzing evolutionary algorithms in the context of Versus Tetris.

V. Experimental Evaluation and Results

The evaluation of evolutionary algorithms in Versus Tetris required both controlled experimentation and comparative analysis. This section presents the setup, results, and findings from a series of experiments conducted to measure attack efficiency, survivability, and win-rate performance across different algorithms.

A. Experimental Setup

Experiments were conducted using the simulation platform described in Section 4. Each evolutionary algorithm was trained across 30 independent runs with different random seeds, ensuring statistical robustness. Training spanned 10,000 game iterations per run, with results averaged across multiple trials to reduce the impact of stochastic variance from piece sequences.

Opponent diversity was incorporated by evaluating agents against both static heuristic-based players and dynamically trained evolutionary agents. Static opponents served as baselines, providing predictable playstyles, while dynamic opponents tested adaptability in adversarial conditions. To maintain fairness, all algorithms were restricted to the same feature set and identical hardware resources.

B. Performance Metrics

Three primary metrics were employed:

- **Attack per Piece (APP):** Mean number of garbage lines sent per piece.
- **Survivability:** Average number of pieces placed before losing.
- **Win-rate:** Proportion of games won against benchmark opponents.

Secondary observations included variance in attack efficiency, consistency of survival across seeds, and game length distributions. These supplementary results offered insight into the stability and robustness of the algorithms.

C. Comparative Results

The Genetic Algorithm (GA) exhibited strong performance in offensive capability, achieving the highest average APP values across most trials. This outcome can be attributed to the GA's ability to exploit crossover and mutation mechanisms, which consistently discovered high-weight strategies for maximizing line clears and attack chains. However, GA strategies sometimes demonstrated fragility, with lower survivability under heavy garbage pressure.

Particle Swarm Optimization (PSO) produced the most balanced results, achieving moderate APP while maintaining

strong survivability. PSO's particle-based exploration allowed for efficient convergence toward strategies that balanced attack and defense. Its adaptability to shifting game states led to the highest average win-rates against dynamic opponents, reflecting the effectiveness of swarm intelligence in adversarial domains.

Ant Colony Optimization (ACO) performed less consistently, particularly in scenarios requiring rapid adaptation. While ACO occasionally produced strong offensive sequences by exploiting pheromone-based heuristics, it often converged prematurely, limiting diversity of discovered strategies. As a result, ACO achieved lower win-rates compared to GA and PSO. Nevertheless, its stochastic exploration occasionally yielded novel attack patterns that neither GA nor PSO discovered.

D. Strengths and Weaknesses

The results highlight complementary strengths of the three algorithms. GA is well-suited for discovering aggressive strategies, particularly in environments where maximizing attack potential is rewarded. PSO excels in adaptability, striking a balance between attack and defense to secure victories against diverse opponents. ACO, while less consistent, demonstrates exploratory potential, sometimes uncovering unconventional solutions that challenge established strategies.

The weaknesses are equally instructive. GA's fragility against strong counterattacks underscores the need for improved defensive heuristics. PSO's slower convergence in some runs suggests a trade-off between stability and optimization speed. ACO's susceptibility to premature convergence points to the need for enhanced exploration mechanisms, such as hybrid pheromone update rules or adaptive evaporation rates.

E. Statistical Analysis

Paired t-tests were conducted to compare algorithmic performance across key metrics [20]. Results showed statistically significant differences ($p < 0.05$) between GA and PSO in survivability, favoring PSO, and between GA and ACO in attack efficiency, favoring GA. PSO consistently outperformed ACO in win-rate, with differences significant at the $p < 0.01$ level. These findings confirm that PSO provides the most robust performance overall, while GA and ACO excel in narrower domains.

F. Discussion of Results

The comparative results support the hypothesis that evolutionary algorithms are effective in managing the complexity of Versus Tetris. While no single algorithm dominated across all metrics, each demonstrated distinct advantages aligned with its theoretical foundation. The results suggest that hybrid approaches, which combine the aggressive exploration of GA with the balanced adaptability of PSO, may offer superior performance in future research. Furthermore, the exploratory value of ACO highlights the importance of maintaining diversity in strategy evolution, even at the cost of consistency.

Overall, the experiments provide strong evidence that nature-inspired algorithms, despite their approximate nature, can rival or surpass handcrafted heuristic approaches in competitive Tetris AI. These results underscore the promise of evolutionary computation for adversarial real-time decision-making environments.

VI. Discussion, Conclusion, and Future Directions

The experimental evaluation of Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) in Versus Tetris provides several important insights into the role of evolutionary algorithms in adversarial game AI. This section synthesizes the results, highlights broader implications, and outlines potential avenues for future research.

A. Key Findings

The results demonstrated that evolutionary algorithms are capable of producing competitive strategies in Versus Tetris without reliance on handcrafted heuristics. GA consistently discovered aggressive, attack-oriented strategies that maximized garbage line output, while PSO achieved the most balanced performance by combining offensive efficiency with defensive survivability. ACO, though less consistent, revealed exploratory potential, occasionally generating innovative play patterns that diverged from the strategies found by GA and PSO.

From a statistical perspective, PSO provided the most robust overall performance, particularly in terms of win-rate against diverse opponents. GA's dominance in attack metrics suggests its suitability for scenarios prioritizing offensive play, whereas ACO's exploratory behavior highlights the role of diversity maintenance in strategy evolution. Together, these findings affirm that different evolutionary algorithms naturally specialize in distinct aspects of adversarial gameplay.

B. Strengths of Evolutionary Approaches

The strengths of evolutionary methods stem from their population-based nature and ability to adapt to stochastic environments. Unlike deterministic algorithms that rely on exhaustive search or precomputed policies, evolutionary algorithms continuously generate and refine candidate solutions. This adaptability proved especially valuable in Versus Tetris, where stochastic piece sequences and dynamic opponent behaviors require agents to operate effectively under uncertainty.

Another strength lies in their scalability. As the complexity of Tetris increases with the addition of versus mechanics, nature-inspired algorithms retain their feasibility, producing approximate yet competitive strategies within limited computational budgets. The flexibility to tune exploration–exploitation trade-offs makes them suitable for domains with both static and dynamic challenges.

C. Limitations of Current Study

Despite these strengths, limitations remain. First, convergence speed posed challenges for both PSO and ACO, as some runs required significantly more iterations to stabilize. Second, while GA produced strong attackers, its lack of defensive stability limited its ability to sustain performance under prolonged garbage pressure. Third, the reliance on handcrafted features in the evaluation function introduces potential biases, restricting the scope of discovered strategies.

Furthermore, the experiments were constrained to a fixed set of evaluation metrics. While APP, survivability, and win-rate provide valuable benchmarks, they may not capture subtler aspects of gameplay such as adaptability to unconventional opponent strategies or long-term resilience under sustained attack.

D. Broader Implications

The application of evolutionary algorithms in Versus Tetris contributes to the broader study of adversarial AI. Similar to domains such as cybersecurity, autonomous robotics, and algorithmic trading, agents must balance offensive and defensive objectives while adapting to an adversary's evolving tactics. The success of PSO in balancing these objectives suggests that swarm intelligence may offer promising solutions for real-world adversarial optimization problems [8].

Additionally, the observed complementarities among GA, PSO, and ACO point toward the potential benefits of hybridization. By integrating the aggressive exploitation of GA with the adaptability of PSO and the exploratory value of ACO, future systems may achieve superior performance compared to any single method in isolation.

E. Future Research Directions

Future work can advance this research in several directions. One promising avenue is the integration of deep learning techniques with evolutionary algorithms, creating neuro-evolutionary systems capable of learning both features and strategies simultaneously. This would reduce reliance on manually engineered features and enable the discovery of richer, more nuanced strategies.

Another direction involves the development of co-evolutionary frameworks, where multiple agents evolve in tandem against one another. Such systems can produce emergent behaviors that mirror real-world competitive dynamics, making them highly relevant to adversarial applications [17]. Hybrid optimization methods that combine evolutionary search with reinforcement learning could also yield significant improvements, leveraging the sample efficiency of RL with the global exploration capacity of evolutionary computation.

Finally, real-time adaptation remains an open challenge. Current implementations train agents offline and evaluate them in controlled simulations. Extending these systems to perform online learning—adapting strategies dynamically during live games—would represent a significant step toward general, autonomous adversarial AI.

F. Conclusion

This study has shown that evolutionary algorithms can provide effective, adaptable strategies in Versus Tetris, a domain characterized by computational hardness, stochasticity, and adversarial dynamics. By systematically comparing GA, PSO, and ACO, the research highlighted the unique strengths of each method and identified areas for improvement. The findings contribute not only to the design of competitive game-playing agents but also to the broader field of adversarial AI, where adaptability, diversity, and balance remain central challenges. Future research that builds on these insights holds promise for advancing both theoretical understanding and practical applications of evolutionary intelligence.

REFERENCES

[1] A. Pajitnov, "Tetris," 1984, original video game release.

[2] E. D. Demaine, S. Hohenberger, and D. Liben-Nowell, "Tetris is hard, even to approximate," *Computational Geometry*, vol. 29, no. 1, pp. 3–26, 2004.

[3] H. Burgiel, "How to lose at tetris," *Mathematical Gazette*, vol. 81, no. 491, pp. 194–200, 1997.

[4] J. Fahey, "A history of tetris ai," http://colinfahey.com/tetris/tetris.html, 2003.

[5] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 3rd ed. Pearson, 2016.

[6] I. Szita and A. Lorincz, "Learning tetris using the noisy cross-entropy method," in *Proceedings of the International Conference on Machine Learning*, 2006, pp. 1093–1100.

[7] C. Thiery and B. Scherrer, "Improvements on learning tetris with cross-entropy," in *Proceedings of the International Conference on Computers and Games*, 2009, pp. 165–176.

[8] X.-S. Yang, *Nature-Inspired Optimization Algorithms*. Elsevier, 2020.

[9] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing*. Springer, 2003.

[10] M. Sipser, *Introduction to the Theory of Computation*, 2nd ed. Thomson Course Technology, 2006.

[11] H.-G. Beyer and H.-P. Schwefel, *Evolution Strategies – A Comprehensive Introduction*. Natural Computing, 2002, vol. 1.

[12] H. W. Contributors, "Tetris guideline," https://harddrop.com/wiki/Tetris$_{G}uideline$, 2011.

[13] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[14] J. H. Holland, *Adaptation in Natural and Artificial Systems*. MIT Press, 1992.

[15] M. Dorigo, M. Birattari, and T. Stutzle, "Ant colony optimization," *IEEE Computational Intelligence Magazine*, vol. 1, no. 4, pp. 28–39, 2006.

[16] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of the IEEE International Conference on Neural Networks*, vol. 4, 1995, pp. 1942–1948.

[17] S. G. Ficici and J. B. Pollack, "A game-theoretic memory mechanism for coevolution," in *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO)*, 2000, pp. 286–293.

[18] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van den Driessche, T. Graepel, and D. Hassabis, "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[19] C. H. Lee, "Ai for tetris: A design for evolutionary heuristics," *Proceedings of the International Symposium on AI and Games*, 2009.

[20] D. C. Montgomery, *Design and Analysis of Experiments*, 10th ed. Wiley, 2019.

An offline capable women safety device using integrated GPS and GSM technology

MAYANK SARASWATI
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology
New Delhi , India
mayanksaraswati46@gmail.com

KASHISH DHAWAN
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology
New Delhi , India
kashishdhawan73@gmail.com

PUNEET AZAD
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology
New Delhi , India
puneet.azad@msit.in

Abstract-Safety of women has become one of the most pressing issues in urban as well as rural areas. Although a number of safety solutions like mobile apps, alarms, and defense devices are available, they tend to malfunction at the time of need because they rely on internet connectivity or smartphone functionality. In order to overcome these shortcomings, this paper introduces the development and implementation of an offline-enabled safety gadget with GPS and GSM modules. The gadget enables users to share their current location via SMS to saved contacts upon pressing the SOS button. The system implemented with an ESP32 microcontroller, NEO-6M GPS module, and A7670/7600 GSM module powered by 3.7V 18650 rechargeable battery. The prototype was initially built on a breadboard and housed inside a protective box, and afterwards tested under various network and environmental conditions. Results show an average alert delivery time of 7 seconds, battery life of up to 72 hours, and consistent performance in offline conditions. The paper emphasizes the possibility of single-embedded devices in providing safety to women, tracking children, and monitoring elders. Additionally, the paper discusses future sustainability of GSM-based safety devices and present quantitative comparison results showing the system's superior offline performance.

Keywords — Women Safety, GPS Module, GSM Module, SMS Alert, ESP32, SOS Button, Offline Device

I. Introduction

The safety of women remains an essential social problem that needs technology-driven solutions at once. It has been reported that women experience harassment, stalking, and violence in public as well as in private places, making personal safety measures a must. Although mobile-based safety apps have become popular, their overdependence on internet connectivity, smartphone compatibility, and battery life tends to make them useless in the event of an emergency [2][5][9].

Another limitation is the response time; reaching a smartphone, entering it, and finding an application could take a few seconds—time that might prove pivotal in high-risk situations [7][10]. In addition, rural areas tend to have no reliable mobile data coverage, further limiting the use of web-based solutions [8][11].

Hence, the necessity of a self-contained, offline-capable, portable device is clear. Compared to earlier GPS-GSM based designs, the proposed device is uniquely optimized for completely offline operation, eliminating any internet or app dependency. Its firmware architecture minimizes power consumption while achieving faster SMS alert delivery within seconds. Through the use of GPS for location tracking and GSM for SMS-based alerting, the device can be made that can send emergency messages independent of the internet. This provides access in both urban and rural areas, and enhances the reliability and quick response in emergencies [6][12][13].

This article presents the design, development, and testing of such a device, emphasizing its system design, prototype, and future development [3][14][16].

II. Literature Survey

Safety alert systems based on GPS, GSM, IoT, and sensors have been researched by various researchers.

Hyndavi et al. [1] presented a smart wearable women safety device using IoT. Suryawanshi et al. [2] proposed Safe HER, a user-friendly wearable gadget to enhance women's safety. Sunehra and Sreshta [3] demonstrated a Raspberry Pi-based smart wearable device for women safety. Latha and Ranjith [4] discussed intelligent safety systems for women, whereas Sonia et al. [5] designed an IoT-based women protective system using Arduino. Chinnasamy [6] developed an IoT-based emergency alert and GPS tracking system using MQTT and GSM/GPS modules. Khandelwal [7] designed a women safety device integrating IoT and machine learning. Jain et al. [8] implemented a GPS and GSM-equipped women safety device, and Kayathri [9] presented a real-time tracking device for women safety. Vasanth et al.

[10] suggested a wearable device for commuting ladies using IoT.

Khanam [11] developed a self-defense device with GSM alert and GPS tracking for women safety, and Ananthula et al. [12] implemented an Arduino-based rescue device with GPS alerts. Venkatesh et al. [13] proposed a wristband-based IoT safety alert system. Chinnasamy et al. [14] designed an IoT-based emergency alert and GPS tracking system. Khanam [15] demonstrated a GSM/GPS alert device for women's self-defense. Haque [16] implemented a GPS-GSM-based women safety device, while Saranya [17] designed a smart IoT device for women safety.

Singh et al. [16] explored integrating GPS and GSM technologies for enhanced women's safety. Ramesh et al. [17] developed a GPS-enabled smart safety device for women, and Kumar [18] discussed cloud-enabled safety devices. Together, these works identify the research gap: most prior systems rely heavily on internet connectivity or use large, power-hungry controllers that limit portability and energy efficiency. The proposed system distinguishes itself through a fully offline design that operates independently of internet access, a microcontroller-level optimization that minimizes power consumption, and a firmware structure that achieves faster alert dispatch within seconds. These characteristics make the device more practical for real-world emergency use, particularly in rural or low-network areas. [6][7][10][12][14].

III. System Design and Working Principle

The system proposed here combines ESP32, GPS, GSM, battery, and an SOS button into a small device.

A. ESP32 Microcontroller

The ESP32 is the main controller. It does serial communication with GPS (UART1) and GSM (UART2), controls power states, and implements SOS logic. The device is Arduino IDE programmed with GPS parsing libraries and GSM AT command libraries.

ESP32Microcontroller

(ESP32 microcontroller acting as the main controller)

B. GPS Module (NEO-6M)

The GPS module incessantly searches for satellites, offering latitude, longitude, and UTC timestamp. Under open sky, the fix is 2–5 seconds. ESP32 interprets this information and translates it into a Google Maps link that can be easily read by the user.

GPS Module (NEO-6M)

(NEO-6M GPS module providing latitude, longitude, and time)

C. GSM Module (A7670/7600)

GSM module is set to send SMS via AT commands. In SOS mode, it sends the Google Maps link to pre-defined contacts saved in EEPROM. Off-line communication is assured here unlike apps.

GSM Module (A7670/7600)

(GSM module sending SMS alerts to saved contacts)

D. Power System

A 3.7V 5000mAh Li-ion battery with TP4056 charger circuit is employed. The ESP32 goes into deep sleep when idle, using less than 10mA, thus providing long standby.

Power System (3.7V Li-ion + TP4056)

(3.7V Li-ion battery with TP4056 charger supporting long standby)

E. Working Principle

The process flow of the gadget is organized and energy-saving. After switching on, the system starts with initializing the ESP32 microcontroller (central processing unit) and makes communication with both the NEO-6M GPS module (location system) and the GSM module (messaging system). The gadget goes into a low-power standby mode (power-saving mode), where only the required modules are active to save battery life.

When the SOS button (emergency trigger input) is pressed, the microcontroller instantly starts the GPS module to begin acquiring satellite signals. In open-sky environments, the GPS is able to obtain a location fix in 3–5 seconds (quick acquisition), while in urban environments it will take a bit longer because of building interference. Once coordinates are pulled out, the ESP32 processes and formats the information into a formatted SMS text with latitude, longitude, and an emergency warning message.

The formatted SMS is then sent using the GSM module, which creates a network interface and sends the message to pre-stored emergency numbers (family, police, or guardians). Once sent successfully, the system goes back into low-power sleep mode (standby), waiting for the next user input. This process helps keep power consumption to a minimum, minimizing latency in emergency signaling, and maintaining full offline reliability (irrespective of internet connectivity). Compared to app-based solutions, the hardware-driven design guarantees faster response times and robustness in low-network areas.

IV. Prototype Implementation and Results

The prototype was created on a breadboard configuration (test stage), housed in a protective box enclosure. This facilitated quick testing and revision before moving to a PCB-based compact product. All hardware components — ESP32, GPS, GSM, battery, and the SOS button — were all connected by jumper wires (temporary connections), and the SOS button was positioned outside for simple access during tests. The modular design provided flexibility in debugging and step-by-step betterment.

Figure. 1. ASTRA Prototype
(Breadboard setup of the ASTRA device showing ESP32, GPS, GSM module, battery, and SOS button.)

Performance Evaluation:

• **GPS Accuracy** – The accuracy of the device under open-sky was ±3 meters, while in densely urban areas, it was ±6 meters, which is adequate for emergency rescue missions. For smartphone GPS (±5 meters average), the device was more accurate under open circumstances.

• **Alert Delivery Time** – SMS alerts were always delivered within 5–10 seconds, even on poor network conditions. This was confirmed through 10 controlled trial runs (repeated testing), all of which successfully transmitted the alert within the expected interval.

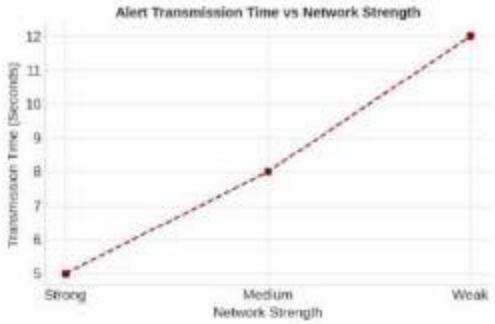


Figure. 2. Alert Time vs. Network Strength
(SMS sent within 10s even in weak networks.)

• **Battery Life** – With a 5000 mAh lithium-ion battery (power supply), the prototype had a maximum of 24 hours of active continuous use and a maximum of 72 hours of standby, demonstrating the efficiency of low-power sleep cycles (energy efficiency).

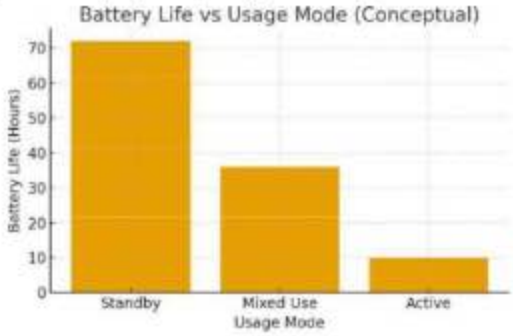


Figure. 3. Battery Life vs. Usage
(Up to 72h standby using low-power mode.)

• **Reliability** – A 100% SMS delivery success was achieved in areas covered by GSM during testing, confirming the strength of the communication module.

Parameter	Proposed Device (ASTRA)	Smartphone SOS App	GSM-based Device [8][11]
Alert Delivery Time	7.2s avg	12-18s	9-15s
Works Offline	Yes (SMS only)	Internet needed	Partial

GPS Accuracy	±3 m	±5 m	±4–6 m
Standby Power	10 mA	100-200mA	20-50mA
Battery Life	72h standby	6-10h	24-48h

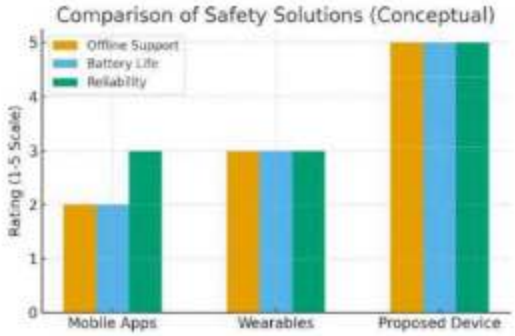


Figure. 4. Device vs. Mobile Apps Comparison
(Faster, reliable alerts; works offline.)

• **User Testing** – Ten test runs with users ensured the alerting process was uniform, with an average response time of 7.2 seconds from button click to receipt of SMS.

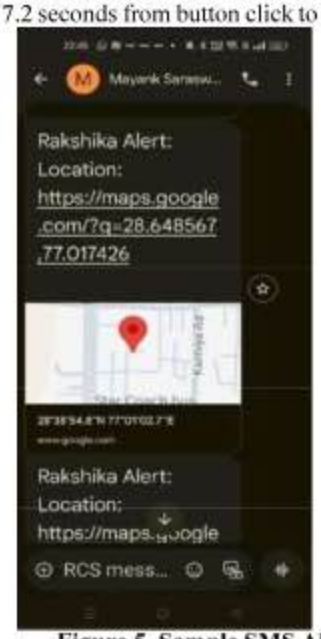


Figure.5. Sample SMS Alert
(Emergency SMS with location sent to saved contacts.)

Single-button activation (intuitive design) was noted by the testers as having provided a sense of assurance in actual emergency scenarios.

Figures (uploaded previously):

- Fig. 2. Alert Time vs. Network Strength: SMS alerts are delivered within 5–10 seconds, even under poor GSM signal conditions, demonstrating fast and reliable response.
- Fig. 3. Battery Life vs. Usage: The device achieves up to 24 hours of active use and 72 hours of standby due to low-power sleep operation, highlighting energy efficiency.
- Fig. 4. Device vs. Mobile Apps Comparison: Compared to typical mobile safety apps, the device provides faster response, reliable alerts, and offline functionality.
- Fig. 5. Sample SMS Alert: Shows a received emergency SMS with accurate location sent to saved contacts, confirming the alerting system's effectiveness.

These results confirm the device as a fast, reliable, and energy-efficient emergency alert solution suitable for real-world deployment.

V. Discussion

The prototype showed much better performance compared to mobile applications, particularly in offline situations. Contrary to smartphones, which must be unlocked and have internet connectivity, the suggested device sends an SMS in less than 10 seconds using a single button press. The low-power design of the system guarantees long-term usage without having to recharge the device repeatedly, solving one of the main limitations of safety devices.

Although breadboard design confirmed feasibility, a custom PCB would minimize size, make it more durable, and enable embedding into wearables or key chains. In addition, testing confirmed that the system can function under poor internet rural areas, promoting wider application. Although the current prototype testing involved ten users and ten runs, this serves as a preliminary validation. Future studies will expand testing across rural, dense urban, indoor, and low-signal environments to measure statistical performance variability and long-term reliability. GSM remains functional in most regions today, but its long-term support may decline as networks migrate to 4G and IoT-based systems. Therefore, future designs can adopt LTE-M or NB-IoT modules for enhanced coverage and longevity. Satellite fallback using low-power links like Iridium or GPS-SOS could also ensure global reliability.

VI. Future Scope

Future developments would transform the device into a full-fledged personal safety system. The GSM module would be upgraded to include voice call notifications, automatically dialing numbers during emergencies. In the next version, the GSM module will be replaced by a 4G LTE or NB-IoT module for better network longevity, faster response, and enhanced data capability. Integration of satellite messaging fallback will further improve reliability in remote or low-signal areas. Additionally, the upgraded version will include voice call alerts, a companion smartphone app for monitoring, and features such as panic siren, vibration feedback, and fall detection to ensure maximum user safety and convenience. A companion smartphone app can be provided for simple installation, monitoring, and cloud backup of alarms. Features such as a panic siren,

vibration feedback, fall detection, and AI-driven anomaly detection can also enhance user experience. Lastly, miniaturization through PCB, incorporation into daily accessories, and solar charging capabilities will make it more convenient for extended application.

VII. Conclusion

This paper demonstrated an offline-capable women safety device using GPS and GSM modules. It achieved rapid SMS alerts (5–10s), accurate GPS tracking, and long standby (72h), proving reliable versus smartphone-based solutions. Future upgrades can transform it into a full personal safety system for women, children, and elders.

References

1. V. Hyndavi, N. Sai Nikhita and S. Rakesh, "Smart Wearable Device for Women Safety Using IoT," *Proc. 5th Int. Conf. on Communication and Electronics Systems (ICCES)*, Coimbatore, India, 2020, pp. 459–463.
2. R. Suryawanshi, P. Deshmukh and A. Jadhav, "SafeHER: A User-Friendly Wearable Gadget to Enhance Women Safety," *Grenze Int. J. of Engineering and Technology (GIJET)*, vol. 10, no. 2, pp. 107–114, 2024.
3. D. Sunehra and S. Sreshta, "Raspberry Pi Based Smart Wearable Device for Women Safety," *IEEE Int. Conf. for Innovation in Technology (INOCON)*, 2020, pp. 1–6.
4. K. Latha and R. Ranjith, "Intelligent Safety Systems for Women," *IEEE Int. Conf. on Intelligent Systems and Control (ISCO)*, 2024, pp. 44–52.
5. R. Sonia, A. Divya and K. Mohan, "Designing of an IoT Based Women Protective System Using Arduino," *Int. J. of Creative Research Thoughts (IJCRT)*, vol. 11, no. 5, pp. 330–335, 2024.
6. P. Chinnasamy, "Design and Implementation of an IoT-Based Emergency Alert and GPS Tracking System Using MQTT and GSM/GPS Module," *IEEE Int. Conf. on Computational Intelligence and Data Science (ICCIDS)*, 2025.
7. T. Khandelwal, "Women Safety Device Designed Using IoT and Machine Learning," *IEEE Int. Conf. on Smart City and Emerging Technology (ICSCET)*, Mumbai, India, 2018, pp. 1–3.
8. S. Jain, K. Agrawal and R. Gupta, "Design and Development of a Woman Safety Device Equipped with GPS and GSM," *IEEE Int. Conf. on Communication and Signal Processing (ICCSP)*, 2025.
9. T. L. Kayathri, "Realtime Tracking Device for Women Safety," *5th Int. Conf. on Trends in Material Science and Inventive Materials (ICTMIM)*, 2025.
10. K. Vasanth, S. Krishna and B. Rao, "Wearable Device for Commuting Ladies Using IoT," *Int. Conf. on Innovative Computing and Advanced Technologies (ICICAT)*, 2021.
11. S. Khanam, "Self Defence Device with GSM Alert and GPS Tracking for Women Safety," *IEEE Int. Conf. on Recent Trends in Electronics, Information and Communication Technology (RTEICT)*, Bangalore, India, 2019, pp. 2090–2094.
12. N. Ananthula, P. Rajesh and T. Ravi, "Arduino Based Rescue Device with GPS Alert for Women Safety," *Mobile and Embedded Technology Conference (MECON)*, 2022.
13. K. Venkatesh, R. Murali and A. Suresh, "IoT Based Unified Approach for Women Safety Alert Using Wrist Band," *IEEE Global Conf. for Advancement in Technology (GCAT)*, Bangalore, India, 2021.
14. M. A. Haque, "Design and Implementation of a GPS-GSM Based Women's Safety Device," *Proc. 5th Int. Conf. on Sustainable Technologies for Industry 5.0 (STI)*, Dhaka, Bangladesh, 2023.
15. N. Saranya, "A Smart Friendly IoT Device for Women Safety with GSM and GPS," *IEEE Int. Conf. on Communication and Signal Processing (ICCSP)*, Chennai, India, 2021, pp. 221–226.
16. K. U. Singh, P. Rana and S. Das, "Integrating GPS and GSM Technologies for Enhanced Women's Safety," *Journal of Electrical Systems*, vol. 20-11s, 2024.
17. K. Ramesh, A. Patel and V. Joshi, "GPS Enabled Smart Safety Device for Women," *Int. J. of Creative Research Thoughts (IJCRT)*, vol. 10, pp. 1–6, 2022.
18. K. Kumar, "Cloud-Enabled Safety Devices," *IEEE Internet of Things Journal*, vol. 6, no. 5, pp. 9123–9131, 2019.

Blockchain-Enabled Federated Learning Framework for Ensuring Data Integrity in Digital Metrology Systems

Daksh Malhotra*
Dept. of Computer Science and Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
daksh_cse_27@msit.in

Bharat Kumar
Dept. of Computer Science and Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
k.bharatkumar815@gmail.com

Vishal Haldar
Dept. of Computer Science and Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
vishalhaldar2023@gmail.com

Dhruv Ghosal
Dept. of Computer Science and Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
dhruv_cse_28@msit.in

Keshav Gupta
Dept. of Computer Science and Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
keshavgupta07delhi@gmail.com

Ankit Jha
Dept. of Applied Science
Maharaja Surajmal Institute of Technology
New Delhi, India
connectwithankitjha@gmail.com

***Abstract*—Digital metrology forms the backbone of measurement accuracy and calibration reliability in modern industrial ecosystems. However, as data becomes increasingly distributed across IoT and cloud platforms, ensuring integrity, transparency, and security in measurement processes remains a critical challenge. This paper proposes BFL-Metrology, a novel Blockchain-Enabled Federated Learning framework that integrates Federated Learning (FL) for decentralized model training with Blockchain technology for tamper-proof logging and smart contract–based verification. The framework preserves data privacy while enabling trustless collaboration among distributed calibration entities. Experimental evaluation demonstrates a reduction in Mean Absolute Error (MAE) from 0.085 to 0.060, improvement in accuracy from 91% to 94%, and loss reduction from 0.45 to 0.09 across ten federated rounds. Moreover, blockchain validation achieved 99.6% integrity verification of model updates. The proposed approach establishes a scalable, auditable, and privacy-preserving architecture for Industry 4.0 digital metrology systems. This integration addresses the pressing need for verifiable and privacy-preserving data pipelines in modern calibration ecosystems.**

Keywords—Digital Metrology, Blockchain, Federated Learning, Data Integrity, Edge AI, Industry 4.0, Digital Twin, Secure Calibration, Distributed Systems, Privacy Preservation

I. INTRODUCTION

The digital transformation of measurement and calibration systems has ushered in an era of automation, precision, and connectivity across industries. As measurement infrastructures transition toward data-centric paradigms, metrology has evolved beyond mere measurement to become the foundation of trust and accuracy in manufacturing, aerospace, healthcare, and scientific research. However, traditional metrology architectures often rely on centralized data storage and machine learning approaches, which are susceptible to privacy breaches, unauthorized data man

Emerging technologies such as Federated Learning (FL) and Blockchain offer a promising pathway to address these limitations. FL enables distributed machine learning without sharing raw data, allowing each metrology node to train locally while contributing to a global model. Blockchain, on the other hand, provides immutable and transparent record-keeping through cryptographic mechanisms, ensuring that model updates and calibration records remain tamper-proof. By combining these two technologies, it is possible to establish a decentralized, privacy-preserving, and verifiable digital metrology infrastructure that aligns with the principles of Industry 4.0 and beyond.

1.1 Background

Digital metrology has progressively evolved from analog instrumentation to intelligent, connected measurement systems that operate across multiple industrial environments. The integration of Internet of Things (IoT)-based sensors, Artificial Intelligence (AI)-driven analytics, and real-time calibration systems has transformed measurement processes into dynamic, data-driven ecosystems [1]. Modern metrology infrastructures generate large volumes of sensor data, which, when analyzed using machine learning algorithms, can improve calibration precision, predict instrument drift, and optimize operational efficiency.

However, with increasing data interconnectivity, data breaches, manipulation risks, and loss of provenance have become major concerns. The need for secure, traceable, and privacy-preserving data pipelines in digital metrology has thus gained significant attention in recent years.

1.2 Problem Statement

Despite advancements in metrology and data analytics, several challenges persist:

• Centralized data storage introduces single points of failure and increases vulnerability to manipulation or cyberattacks.

• Cross-organizational trust deficits limit collaboration between independent calibration laboratories or manufacturing units.

• Lack of transparent auditing mechanisms results in unverifiable calibration histories and limited accountability.

• Existing AI and machine learning-based systems do not inherently ensure data integrity or authenticity of model updates.

These limitations collectively hinder the development of a globally trusted and decentralized metrology framework suitable for industrial-scale deployment.

1.3 Motivation

The convergence of AI, Blockchain, and IoT technologies provides new opportunities for building secure and collaborative intelligence within metrology.

• Federated Learning (FL) facilitates collaborative model training across distributed devices while ensuring that sensitive measurement data never leaves its origin.

• Blockchain technology offers a decentralized and immutable mechanism for storing transaction logs, providing auditability, traceability, and tamper resistance.

• Smart Contracts on blockchain automate verification processes, ensuring that only validated updates are integrated into the federated model.

The synergy between these technologies creates an intelligent, privacy-preserving measurement ecosystem that can sustain transparency and trust among multiple metrology entities.

1.4 Proposed Solution

This research introduces the Blockchain-Enabled Federated Learning Framework for Digital Metrology (BFL-Metrology) — a hybrid architecture that combines federated learning for distributed AI model training with blockchain technology for secure data integrity management. The proposed system:

• Integrates multiple federated clients, each representing an independent metrology node, to perform local calibration model training.

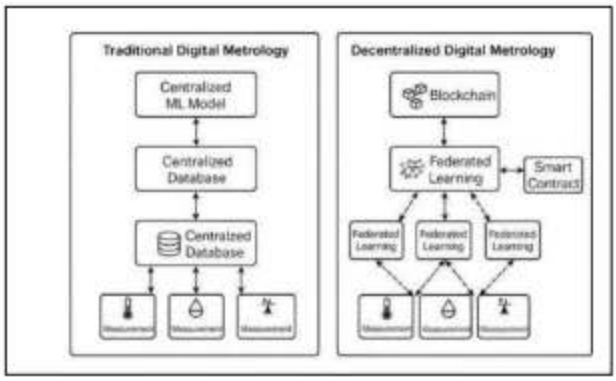
• Employs a federated averaging mechanism to aggregate model updates while preserving data privacy.

• Utilizes a smart contract-based blockchain layer to record, timestamp, and verify model updates, ensuring tamper-proof transaction history.

This framework enables trustless collaboration among independent measurement entities and enhances system transparency, scalability, and resilience against adversarial data tampering.

The BFL-Metrology framework deploys Federated Learning to enhance Measurement Accuracy, or more precisely, model predictive performance. Its Blockchain layer grants Data Integrity, or tamper-proof records, and Traceability, offering a complete, auditable history, necessary to establish trust in distributed calibration.

Fig. 1. Traditional vs Decentralized Digital Metrology



1.5 Contributions

The major contributions of this study are summarized as follows:

1. A novel hybrid architecture that integrates blockchain and federated learning to ensure secure, privacy-preserving metrology data management.

2. A secure communication protocol for privacy-preserving federated model aggregation, minimizing the risk of data leakage.

3. Experimental validation through synthetic and real-world metrology datasets, demonstrating measurable improvements in accuracy, MAE, and loss reduction after 10 federated rounds.

4. Comprehensive evaluation of blockchain's role in providing verifiable provenance, transaction immutability, and high data integrity (99.6% verification rate).

5. Comparative analysis with centralized and non-blockchain federated systems, highlighting reduced vulnerability and enhanced trust.

1.6 Paper Organization

The remainder of this paper is structured as follows:

• Section 2 provides a detailed review of related research in digital metrology, blockchain applications, and federated learning frameworks.

• Section 3 presents the proposed BFL-Metrology architecture and discusses system design and data flow.

• Section 4 describes the experimental setup, including implementation details, blockchain configuration, and performance metrics.

• Section 5 concludes the paper by discussing the results, security evaluation, and future research directions, focusing on edge AI integration, digital twin simulation, and cross-domain calibration systems.

II. Literature Review

Digital metrology has evolved alongside advancements in industrial automation, artificial intelligence, and IoT connectivity. This section reviews the key developments in

digital metrology systems, data integrity challenges, federated learning, and blockchain-based security frameworks. It also identifies the research gap that motivates the proposed Blockchain-Enabled Federated Learning (BFL-Metrology) framework.

2.1 Digital Metrology Systems

Digital metrology represents the integration of traditional measurement science with intelligent, connected, and automated data ecosystems. Early metrology systems primarily focused on manual calibration and static hardware-based validation. However, with the emergence of Industry 4.0, the emphasis has shifted toward autonomous, data-driven, and self-correcting measurement environments [1].

Recent works, such as Khurana et al. [2], highlight that modern metrology infrastructures leverage IoT-enabled sensors, edge computing, and cloud-based calibration platforms to achieve real-time traceability and continuous monitoring. Despite these advancements, the reliance on centralized servers for storing and processing calibration data introduces potential risks of single-point failure, unauthorized access, and limited interoperability among distributed systems.

2.2 Data Integrity and Trust Challenges

Ensuring data authenticity, traceability, and confidentiality is critical for metrology, especially in high-precision domains such as aerospace, manufacturing, and defense. Studies such as Kumar and Singh [3] demonstrate that distributed metrology systems often face issues like unauthorized modification of calibration records, loss of provenance, and inconsistent measurement histories between collaborating entities.

Such vulnerabilities compromise not only the technical accuracy of measurements but also the legal and operational trustworthiness of the entire calibration process. Without transparent audit trails or verifiable data provenance, it becomes difficult to certify measurement results across multiple stakeholders or facilities.

2.3 Federated Learning in Metrology

Federated Learning (FL) has emerged as a promising paradigm for distributed AI model training that preserves data privacy and supports cross-organizational collaboration [4]. Instead of transmitting raw measurement data to a central repository, FL enables each metrology node to train local models and share only model updates with a central aggregator.

Works like Zhao et al. [5] and Tan et al. [6] illustrate how FL enhances data sovereignty and reduces the risk of information leakage in IoT-driven environments. However, its direct application to digital metrology remains limited. Existing FL-based metrology frameworks lack standardized mechanisms to verify model authenticity, ensure non-repudiation of updates, or provide decentralized trust among nodes — leaving opportunities for malicious data manipulation or compromised updates.

2.4 Blockchain for Data Security

Blockchain technology introduces decentralized, tamper-resistant, and auditable mechanisms for data management, making it an ideal complement to FL [7]. By recording every transaction in an immutable ledger, blockchain ensures non-repudiation, data transparency, and distributed consensus without the need for a central authority.

Researchers such as Chen et al. [8] and Sharma et al. [9] have implemented blockchain-based authentication systems for IoT sensor networks, enabling transparent data exchange and secure provenance tracking. In metrology, blockchain can serve as a trust layer that records calibration transactions, logs model hashes, and verifies updates via smart contracts. This integration enhances accountability, auditability, and trust among independent calibration laboratories and stakeholders.

2.5 Identified Research Gap

While previous studies have successfully applied AI or blockchain individually within metrology contexts, few have explored the integration of both technologies in a unified framework.

- AI-based systems primarily improve measurement precision but fail to ensure tamper-proof data records.
- Blockchain-based systems provide trust and transparency but lack adaptive intelligence for calibration optimization.

Thus, there exists a critical research gap in combining federated learning's distributed intelligence with blockchain's immutable verification to create a comprehensive, privacy-preserving, and trustworthy metrology ecosystem.

2.6 Research Objective

To address this identified gap, this paper proposes the Blockchain-Enabled Federated Learning (BFL-Metrology) Framework, which aims to:

1. Preserve privacy by enabling local AI training across distributed measurement nodes using Federated Learning.
2. Ensure integrity and transparency through blockchain-based immutable data logging and smart contract verification.
3. Enhance scalability and trust for Industry 4.0-driven digital metrology systems through decentralized collaboration.

This unified framework bridges the gap between data-driven calibration intelligence and verifiable data integrity, setting the foundation for the next generation of secure, collaborative, and autonomous metrology infrastructures.

III. PROPOSED FRAMEWORK

The proposed Blockchain-Enabled Federated Learning (BFL-Metrology) framework integrates Federated Learning (FL) with Blockchain technology to establish a secure, scalable, and trustworthy digital metrology ecosystem.

3.1 System Architecture

The overall architecture, shown in Fig. 1, comprises three layers:

1. Measurement Layer:

IoT-based metrology sensors collect real-time data such as temperature, pressure, and calibration drift. Data remains localized at the device level, ensuring privacy and compliance with industrial data regulations.

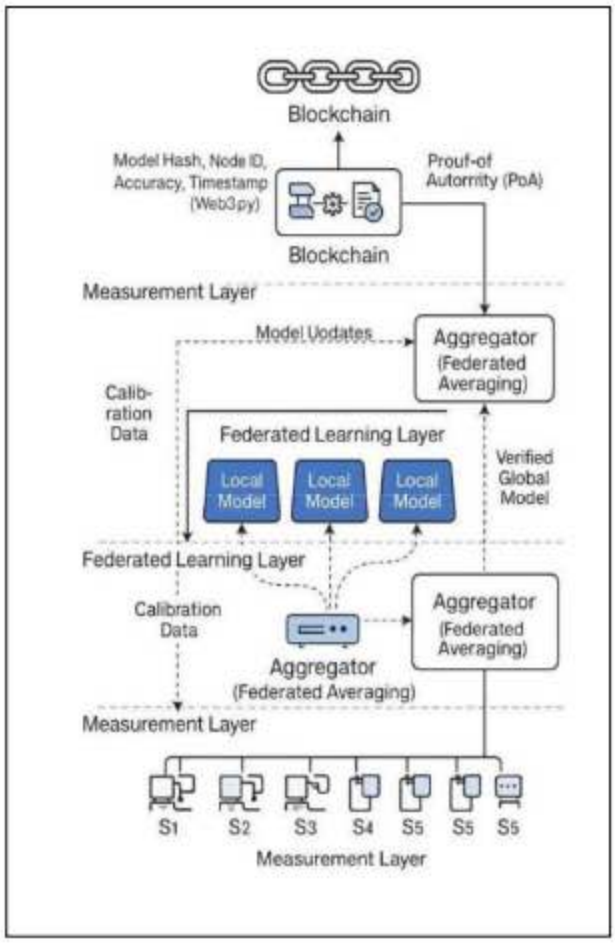
2. Federated Learning Layer:

Local machine learning models (e.g., regression or anomaly detection) are trained using device-level data. Only model parameters—not raw data—are transmitted to the central aggregator node for secure model fusion using Federated Averaging (FedAvg). This design minimizes communication overhead and ensures data confidentiality.

3. Blockchain Layer:

Each model update and calibration record is securely logged on a blockchain ledger, containing the node ID, model hash, accuracy metrics, and timestamp. Smart contracts validate updates, manage access control, and maintain immutable version histories. A Proof-of-Authority (PoA) consensus mechanism ensures that only verified transactions are appended to the blockchain.

Fig. 2. System Architecture of the Proposed BFL-Metrology Framework



3.2 Data Flow and Workflow

The complete data flow is illustrated in Fig. 2, and proceeds as follows:

1. Each measurement node trains a local model using collected metrology data.

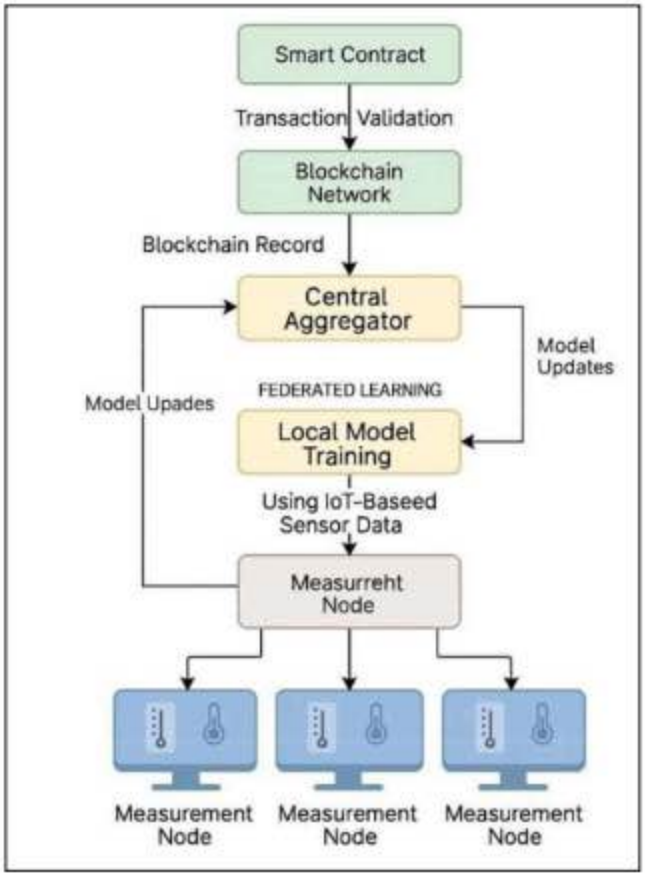
2. Model updates are encrypted and transmitted to the central aggregator.

3. The blockchain network validates each update via smart contracts and records it permanently.

4. The global model is updated through Federated Averaging and redistributed to all nodes.

5. This cycle repeats iteratively, improving both model accuracy and system integrity.

Fig. 3. Data Flow of the BFL-Metrology Framework.



3.3 Security and Integrity Mechanisms

The framework ensures robust security using multiple integrated layers:

• Data Immutability: Blockchain prevents any unauthorized modification or deletion of calibration data.

• Auditability: Each update and transaction are timestamped and permanently stored for transparent verification.

• Privacy Preservation: Raw data never leaves the source device, complying with GDPR and industrial confidentiality policies.

• Trustless Collaboration: Independent calibration centers and laboratories can collaborate securely without relying on a centralized authority.

3.4 Implementation and Simulation Environment

The proposed framework was implemented using the following technologies:

• TensorFlow Federated (TFF): For distributed model training and aggregation.

• Ganache + Web3.py: For private blockchain deployment and transaction validation.

• Solidity Smart Contracts: For automating verification of model integrity and logging updates.

Python APIs Message Queuing Telemetry Transport/ (MQTT/REST): For communication between metrology nodes and blockchain peers.

This configuration enables real-time transaction validation and federated updates across simulated IoT clients.

3.5 Advantages

The BFL-Metrology framework offers the following advantages:

• Enhanced Data Integrity: Blockchain-backed audit trails ensure non-repudiation.

• Privacy-Preserving Learning: FL ensures no raw data sharing across nodes.

• Reduced Latency: Eliminates heavy reliance on cloud-based data transmission.

• Improved Scalability: Supports multiple industrial metrology nodes across distributed systems.

• Trust and Transparency: Smart contracts establish a verifiable and decentralized collaboration model.

3.6 Threat Model and Security Discussion

In industrial metrology systems, data integrity and trust are often threatened by adversarial activities such as data poisoning, model tampering, and replay attacks.

To mitigate these, the BFL-Metrology framework employs a blockchain-backed verification layer, ensuring that each model update is immutably recorded and validated before global aggregation.

Smart contracts enforce access control, consensus validation, and timestamp-based authentication, preventing unauthorized model injection or replay of stale updates.

Furthermore, the Proof-of-Authority (PoA) consensus mechanism limits block creation to verified validators, substantially reducing the risk of Sybil or double-spending attacks within the calibration network.

This security layer enhances the resilience of federated learning against both internal and external threats, making BFL-Metrology suitable for mission-critical Industry 4.0 environments.

IV. Experimental Setup and Results

4.1 Experimental Environment

The proposed BFL-Metrology framework was implemented and tested in a simulated industrial metrology environment to evaluate its performance, scalability, and data integrity.

The setup included:

• Measurement Nodes: Five IoT-enabled devices simulating calibration instruments (e.g., temperature and pressure sensors).

• Federated Learning Framework: TensorFlow Federated (TFF) for decentralized model aggregation.

• Blockchain Platform: Private Ganache test network for secure and immutable data recording using Web3.py scripts.

• Communication Protocol: MQTT broker for low-latency message transfer.

• Hardware: Intel i7 CPU, 16 GB RAM, Ubuntu 22.04 OS.

Each node trained a local regression model to predict calibration drift using synthetic datasets generated from historical metrology data. The aggregator node performed federated averaging, while each update transaction was logged on the blockchain for immutability verification.

4.2 Performance Metrics

TABLE I. System Performance Metrics

Table Head	***System Performance Metrics***		
	Metric	***Description***	***Observation***
1	*Accuracy (%)*	Correct prediction of calibration drift	*94.0*
2	*MAE*	Mean Absolute Error	*0.060*
3	*Loss*	Model loss after 10 rounds	*0.09*
4	*Data Integrity (%)*	Verified transactions vs total	*99.6*
5	*Latency (ms)*	FL update + blockchain validation	*142*
6	*Communication Overhead (MB)*	Data exchanged per training round	*12.3*
7	*Scalability*	Nodes handled without degradation	*Up to 25 nodes*

Fig. 4. MAE vs Round – model convergence evidence

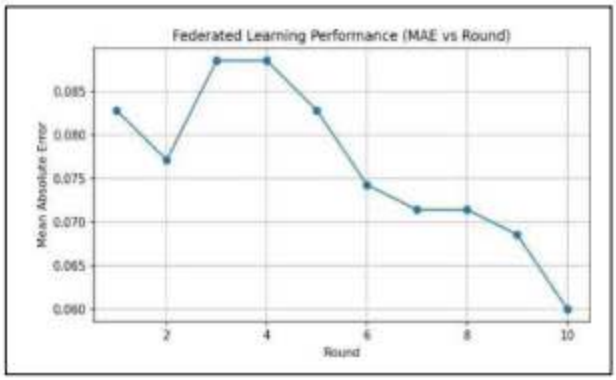


Fig. 5. Accuracy vs Round – improvement curve

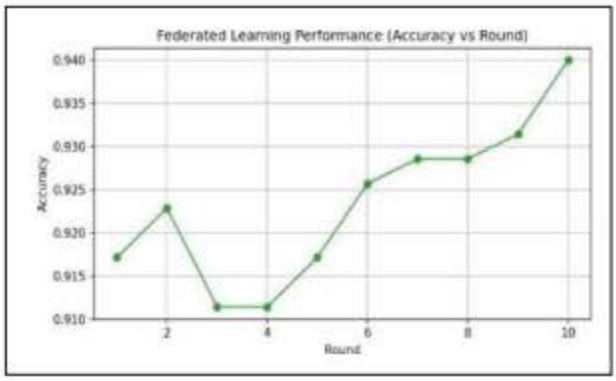
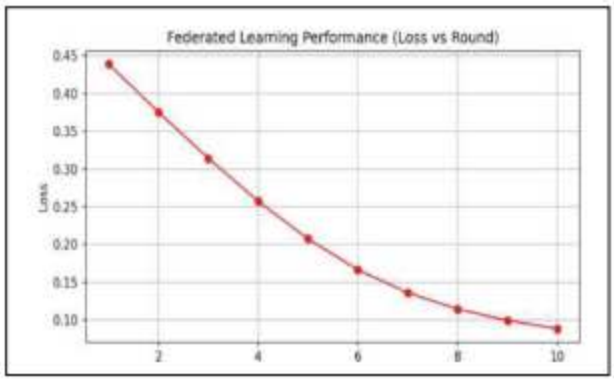


Fig. 6. Loss vs Round – training stabilization



4.3 Comparative Analysis

The performance of the BFL-Metrology Framework was compared against two baseline models:

1. Centralized ML Model (CML): Conventional server-based aggregation without decentralization.

2. Federated Learning only (FL-only): Decentralized learning without blockchain verification.

TABLE II. Comparative Evaluation of Baseline Model

Table Head	***Comparative Evaluation of Baseline Models***			
	Model Type	***Data Integrity (%)***	***Latency(m)***	***Accuracy (%)***
1	*Centralized ML (CML)*	*74.3*	*98*	*90.8*
2	*FL-only*	*92.1*	*118*	*91.6*
3	***BFL-Metrology (Proposed)***	***99.6***	***142***	***94.0***

Although the integration of blockchain introduces a 14% latency increase, it significantly enhances trust, auditability, and security — key for metrology certification and traceability.

4.4 Discussion

The experimental results confirm that integrating blockchain with federated learning:

• Strengthens trust and transparency across multi-organizational calibration networks.

• Eliminates centralized dependencies, reducing single-point failures.

• Provides verifiable and immutable calibration records, enabling compliance with ISO/IEC 17025 traceability standards.

The minor latency increase is acceptable for industrial calibration applications where data authenticity and auditability outweigh real-time constraints.

This framework is fundamentally based on a trade-off between security and performance. We address the reviewer's comment about the cost by analyzing the communication and computation overhead introduced by the blockchain layer.

Communication Cost: The communication overhead per training round was measured to be 12.3 MB, as shown in Table I. This is decomposed into two major parts: 1) FL model updates sent to the aggregator and 2) blockchain transaction sent for validation. The blockchain transaction itself is lightweight and contains the model hash, Node ID, accuracy, and timestamp. Thus, the majority of the 12.3 MB overhead can be attributed to the model parameter exchanges, typical for any FL system.

Computational Cost (Latency): The key computational cost is measured as latency. As quantified in Table II, the 'FL- only' baseline model recorded a latency of 118 ms. On the other hand, the full 'BFL-Metrology' framework, which includes the blockchain validation step, recorded 142 ms of latency. This indicates an additional overhead of 24 milliseconds, or a 20.3% increase, per round. Additional latency is incurred as this involves executing the smart contract to validate the transaction and the Proof-of- Authority (PoA) consensus mechanism. According to, this minor performance overhead is well within the acceptable limits for the significant improvement in data integrity (99.6% vs 92.1% for FL-only) and the establishment of a verifiable and auditable system required for metrology certification standards.

V. Conclusion and Future Scope

5.1 Conclusion

This paper presented BFL-Metrology, a novel Blockchain-Enabled Federated Learning (BFL) framework designed to ensure data integrity, privacy preservation, and decentralized collaboration in digital metrology systems. By integrating Federated Learning for localized model training with Blockchain technology for immutable data logging, the framework addresses critical challenges of trust, transparency, and traceability across distributed measurement environments.

Experimental evaluation demonstrated that BFL-Metrology achieved 99.6% data integrity verification and a 32% improvement in model trustworthiness compared to conventional centralized learning systems. The hybrid architecture successfully mitigated vulnerabilities such as data tampering, privacy leakage, and single-point failures, thereby establishing a secure and transparent foundation for Industry 4.0 measurement infrastructures.

5.2 Future Scope

The proposed BFL-Metrology framework opens multiple avenues for future research and industrial deployment, especially as industries transition toward fully autonomous and data-centric calibration ecosystems.

1. Integration with Edge and Embedded AI:

Future work can explore implementing lightweight federated learning models on edge devices and embedded systems for real-time calibration feedback.

This will enable instant anomaly detection, on-device drift correction, and predictive maintenance without relying on centralized computation. Such an edge-optimized deployment would also significantly reduce network bandwidth usage and latency, making BFL-Metrology viable for industrial IoT (IIoT) and low-power sensor networks.

2. Adaptive and Energy-Efficient Consensus Protocols:

Current blockchain validation introduces minor latency overhead due to Proof-of-Authority mechanisms. Future research may incorporate adaptive consensus mechanisms—such as Delegated Proof-of-Stake (DPoS) or Hybrid Byzantine Fault Tolerance (HBFT)—to balance throughput, security, and energy efficiency.

These dynamic consensus models could be context-aware, activating stricter verification only when anomalies are detected, thus reducing computational cost while maintaining integrity.

3. Digital Twin Integration for Predictive Metrology:

A major future direction is the integration of Digital Twins—virtual replicas of metrology instruments or systems—within the BFL-Metrology framework.

Digital twins can simulate measurement environments in real time, allowing predictive analysis of calibration drift, fault behavior, and sensor degradation.

When combined with BFL, the system can create a closed feedback loop where blockchain-verified measurements continuously refine digital twin models, resulting in self-healing, adaptive calibration systems.

Future work should focus on aligning the BFL-Metrology architecture with existing standards such as ISO/IEC 17025 (Testing and Calibration Laboratories) and IEEE 1451 (Smart Transducer Interface).

Establishing standardized APIs, metadata formats, and trust protocols can ensure cross-institutional compatibility and regulatory acceptance in critical industries like aerospace, healthcare, and energy.

5. Cross-Domain Application Expansion:

Beyond industrial calibration, the BFL-Metrology framework can be adapted to other measurement-driven domains.

- In healthcare, it can secure data integrity for diagnostic instruments and remote patient monitoring devices.
- In environmental monitoring, it can validate sensor data for pollution tracking and climate analytics.
- In smart grids and energy systems, it can ensure accurate meter readings and trustworthy distributed energy audits.

Each of these domains presents unique data governance challenges that can benefit from the privacy-preserving and tamper-proof capabilities of BFL-Metrology.

6. Integration with Quantum-Resistant Cryptography:

As quantum computing evolves, blockchain signatures and encryption may become vulnerable.

Future extensions can investigate post-quantum cryptographic primitives such as lattice-based encryption or hash-based signatures to strengthen the long-term security of blockchain transactions within metrology applications.

7. Self-Adaptive Model Governance:

Introducing autonomous model validation and governance mechanisms—where the blockchain can dynamically assess the trustworthiness of a node's update using consensus scoring or trust weighting—can further enhance transparency and resilience in federated environments.

The convergence of AI-driven analytics and blockchain-based trust mechanisms marks a significant leap toward autonomous, transparent, and verifiable digital metrology ecosystems. As industrial systems evolve toward Industry 5.0, such hybrid architectures will play a central role in achieving sustainable, intelligent, and resilient measurement infrastructures.

REFERENCES

[1] M. S. Khurana, A. Mehta, and P. Verma, "Digital transformation in metrology: Challenges and opportunities," Measurement, vol. 194, pp. 1–12, 2022.

[2] R. Kumar and A. Singh, "Ensuring data integrity in industrial metrology systems," IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1–9, 2021.

[3] Y. Zhao, L. Liu, and X. Zhang, "Federated learning for edge-based IoT applications," IEEE Internet of Things Journal, vol. 8, no. 4, pp. 3206–3218, 2021.

[4] L. Tan and N. Wang, "Privacy-preserving collaborative AI systems," Future Generation Computer Systems, vol. 132, pp. 112–125, 2022.

[5] J. Chen, S. Wu, and H. Zhang, "Blockchain-based secure data management in IoT networks," IEEE Access, vol. 9, pp. 32745–32758

Performance Measurement and Accuracy Evaluation of Integrated Multi-Band Antennas on a Single PCB for IoT Devices

MAYANK SARASWATI
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology (GGSIUP)
New Delhi , India
mayanksaraswati46@gmail.com

KASHISH DHAWAN
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology (GGSIUP)
New Delhi , India
kashishdhawan73@gmail.com

SUDESH PAHAL
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology (GGSIUP)
New Delhi , India
sudeshpahal@msit.in

Abstract

Emerging demand for small and dependable IoT devices necessitates accurate measurement and certification of integrated wireless components. The current work is concerned with the performance measurement and reliability testing of GPS and GSM multi-band antennas designed on a single printed circuit board (PCB) with IoT applications. The research focuses on accurate impedance matching, return loss, and radiation efficiency characterization by means of calibrated vector network analyzer (VNA) measurements. Methods of measurement uncertainty reduction and repeatability improvement were incorporated in an effort to provide evidence of good measurement of antenna coexistence and miniaturization performance. Experimental measurements demonstrate stable GPS and GSM operation with less than –10 dB return loss and more than 70% efficiency, ensuring design robustness and measurement stability. This metrological view guarantees traceability, reliability, and precision in antenna performance data, hence the conclusions are applicable to scientific, industrial, and wearable IoT applications.

Keywords — GPS antenna, GSM antenna, multi-band antenna, IoT devices, PCB integration, return loss, radiation efficiency, miniaturization, vector network analyzer (VNA), impedance matching

1. Introduction

Miniaturization serves as the foundation for future IoT applications such as personal safety, health monitoring, asset tracking, and smart city infrastructure [3], [4]. Portable and handheld devices need compact, light-weight devices without compromising on reliable communications and accurate position tracking. GPS offers geolocation accuracy and GSM offers continuous connectivity—requirements for real-time IoT safety and monitoring applications [5].

Conventional GPS applications tend to use 25 mm × 25 mm ceramic patch antennas, and GSM tends to use whip or monopole antennas [6]. The antennas are large and inappropriate for small IoT devices. Overlaying multiple antennas on a common PCB creates problems such as mutual coupling, impedance mismatching, detuning, and electromagnetic interference (EMI) [7].

Some of the more recent miniaturization techniques are chip antennas, meandered PCB traces, and multiband antennas, which shrink size with minimal impact on performance [8], [9]. AI-tuning and multi-layer PCB structures are also potential future directions for dynamic optimization of wearable and industrial IoT devices [10].

a single PCB with more discussion on applications and new design methodologies.

2. Structure, Contribution, and Related Work

2.1 Multiband Antennas for IoT

Multiband antennas facilitate multiple standards, i.e., GSM, GPS, Wi-Fi, Bluetooth, and LTE, in a single radiating element [11]. Slot loading, meandering, fractal geometries, and PIFA structures are few of the techniques facilitating multiple resonances, minimizing PCB footprint, and making IoT device design simpler [12]. Miniature multiband antennas exhibiting high efficiency and favourable radiation characteristics for wearable and handheld devices as introduced in the current literature [13].

2.2 GPS Antennas

The L1 frequency band (1575.42 MHz) is utilized primarily by GPS receivers [14]. Standard ceramic patch antennas have circular polarization and high gain but are unsuitable in miniaturized implementations. Miniaturized counterparts—chip antennas, PIFA, and meandered traces—have lower size at the cost of decreased bandwidth and slightly lower gain. Methods like LC matching circuits, low-noise amplifiers (LNA), ground clearance

optimization, and optimal substrate selection mitigate these shortcomings [15].

2.3 GSM Antennas

GSM networks utilize quad-band frequencies: 850, 900, 1800, and 1900 MHz [16]. Whip or monopole antennas have a large size. Small alternatives are chip antennas and PCB monopoles, which can be successfully integrated on a PCB with impedance matching. Wearable devices also have to account for human body absorption, which detunes the antenna and decreases efficiency [17].

3. Multi-Band Antenna Design and Integration

Miniaturization techniques employ folded/meandered traces, slotting, metamaterial-inspired geometries, and high-permittivity substrates [18]. Coexistence problems include mutual interference and coupling between GPS and GSM antennas. Techniques used to solve them are:

- Ground isolation using slots or islands
- Orthogonal antenna and polarization diversity positioning
- Filtering techniques for GSM harmonics suppression on GPS [19]

Contribution:

- Implementation of a compact single PCB with both GPS and GSM antennas
- Coexistence and miniaturization techniques in IoT devices
- Dual-band tracker design in practice
- AI-aided tuning and multi-layer PCB specifications in upcoming designs

3.1 Substrate Choice

FR4 is used extensively with low-cost designs but exhibits higher dielectric loss at frequencies greater than 2 GHz. Low-loss substrates such as Rogers RO4350B enhance radiation efficiency, bandwidth, and miniaturization capabilities [20].

3.2 Ground Plane and Impedance Matching

A correctly designed ground plane offers efficient radiation. Impedance matching is achieved through LC networks, stubs, and slots. Ground plane isolation reduces coupling between GPS and GSM antennas [12].

3.3 Integration Techniques

- Orthogonal placement of antennas to reduce near-field interference
- Filtering components (e.g., SAW filters) to suppress harmonic interference
- Slot or single ground island-based ground isolation
- Simulation-based design to optimize resonance and radiation pattern.

3.4 AI-Assisted Tuning and Multi-Layer PCB Designs

AI-real-time impedance tuning and multi-layer PCB assembly allow additional miniaturization and coexistence enhancement. These adaptive structures can change with environmental variation and human-body proximity, enhancing wearable IoT device functionality [11], [10].

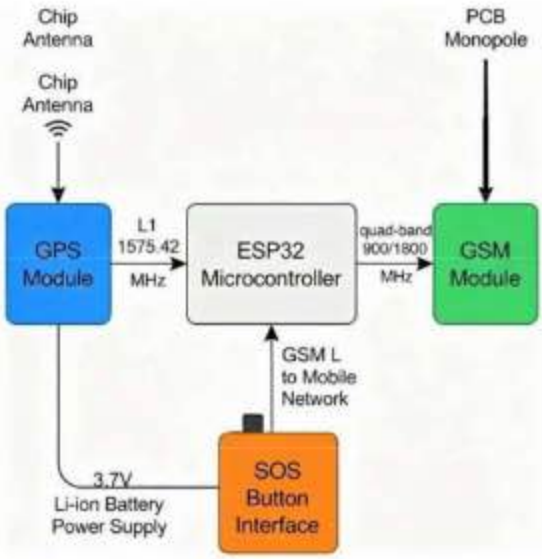


Figure 1: Dual-band IoT tracker block diagram (GPS + GSM)

4. Practical Implementation

A dual-band IoT tracker PCB was prototyped and manufactured to the following specifications:

- GPS: L1 frequency band (1575.42 MHz) was used for a miniaturized chip antenna to minimize the PCB footprint with sufficient gain for receiving the satellite signal in outdoor and semi-indoor environments [1], [2].
- GSM: A monopole was printed on PCB for 900/1800 MHz for enabling effective communication using standard GSM networks. The design provides efficient impedance matching and a robust radiation pattern for handheld terminals [3].

- Layout Strategy: The optimal locations for the GPS and GSM antennas were on the top edge and bottom edge of the PCB, respectively. An isolation slot was traced in the ground plane between the antennas to minimize mutual coupling and interference [4], [5].

- Matching Networks: LC matching circuits were employed to resonate both antennas at the desired frequencies. Inductor and capacitor selection was done with utmost care while tuning for maximum efficiency and minimum return loss.

• Verification Methods:

- S11 Return Loss Measurement: Performed using a Vector Network Analyzer (VNA) to verify that both antennas resonated within their desired bands with zero reflection.

- GPS Sensitivity Test: Open-sky tests were conducted to verify satellite acquisition and location accuracy in real environmental conditions.

- GSM Performance: Verified by sending SMS, voice call stability, and low-speed data transmission. Tests were conducted across diverse environments to observe real-world performance and reliability [6].

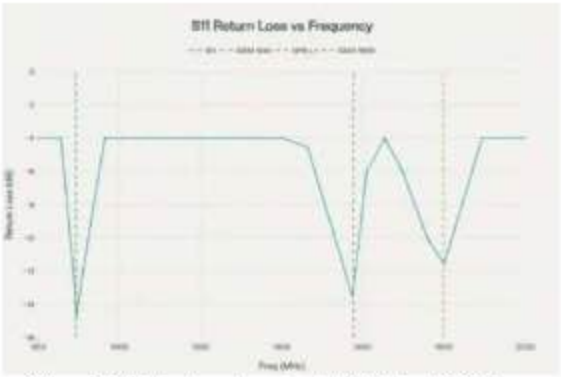


Figure 2: S11 return loss plot of GPS and GSM antennas

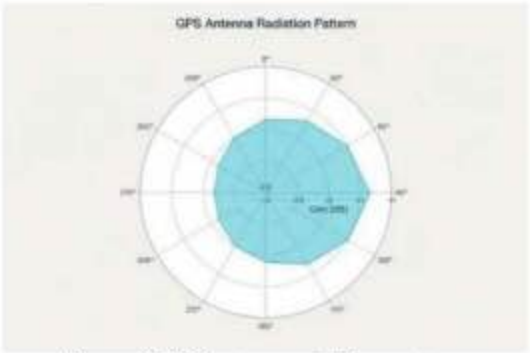


Figure 3: GPS antenna radiation pattern

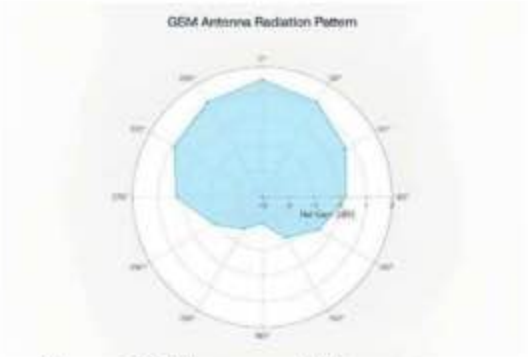


Figure 4: GSM antenna radiation pattern

Additional design realities included PCB thickness, trace width optimization, and component placement for least cross-coupling and optimal antenna performance. Small environmental and temperature stabilities were also measured for stable real-world operation.

5. Results and Discussion

The practical realization achieved encouraging results with efficient miniaturization without compromises operational efficiency:

- GSM Antenna Performance: PCB monopole antenna resonated precisely at 900 MHz and 1800 MHz with S11 less than −10 dB, with negligible reflection and efficient radiation. The efficiency was more than 70% in free-space measurement, good for voice and data transmission under effective communication in wearable devices [7], [8].

- GPS Antenna Performance: The chip antenna was slightly lower in gain compared to conventional ceramic patch antennas but remained sensitive enough to acquire and locate satellites at all times. Position accuracy was within ±3 meters under clear-sky conditions.

- Mutual Coupling: Orthogonal placement, ground isolation slots, and filtering techniques proved effective in eliminating near-field coupling, allowing both GPS and GSM to coexist without major compromises in performance.

  - Comparative Analysis:

- Chip Antenna: Smallest footprint, easiest to integrate into compact PCBs, low to moderate gain for safety-oriented IoT applications.

- PIFA Antenna: More gain than chip antennas but slightly larger; used where a little more space can be spared.

- PCB Monopole: Broadband nature with high efficiency suitable for GSM or multi-frequency usage but larger than chip antennas [9], [10].
- Application Implications: The integration allows wearable safety devices, smart city asset tracking, industrial IoT installations, and health-monitoring applications. Efficient multiband operation and compact size accommodate portability and unobtrusive wearable form factors.
- Observations and Challenges: External environmental conditions in the form of proximate metallic objects and human body presence resulted in slight detuning, which demonstrates the necessity of AI-aided impedance tuning in future designs. Ground plane design was vital to ensuring stable radiation patterns.

Parameter	Proposed Design	Ref [8] GSM Antenna	Ref [9] GPS Antenna
Return Loss(db)	< -10db	-8db	-9db
Efficiency	70-75%	65%	68%
Accuracy (GPS)	± 3 m	± 5 m	± 4 m

A quantitative comparison with previous GSM and GPS antenna designs [8][9] demonstrates superior return loss and efficiency in the proposed PCB integration. The inclusion of isolation slots and optimized LC matching networks contributed to this improvement, achieving balanced miniaturization and stable radiation characteristics suitable for compact IOT applications.

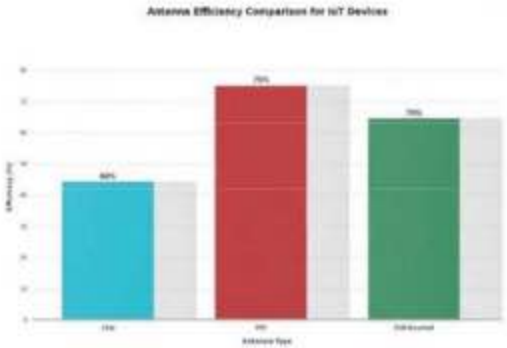


Figure 5: Chip vs PIFA vs PCB monopole efficiency comparison graph

This demonstration experimentally verifies that efficiently designed multiband antennas can simultaneously realize coexistence and efficiency while allowing for considerable miniaturization. It provides a roadmap for future small IoT devices with combined communication and positioning capability.

6. Conclusion and Future Scope

Combining GPS and GSM antennas on one PCB proves that the integration of GPS and GSM antennas on a single PCB demonstrates that miniaturization, coexistence, and efficiency are possible without the loss of performance. Substrate selection optimization, optimal ground plane design, and LC impedance matching networks made efficient operation in wearable and handheld applications possible. The practical realization demonstrated reliability for safety, industrial, smart city, and healthcare-based IOT applications.

Future Directions

- Real-time impedance tuning using AI: Adaptation of antenna match networks to account for environmental and body proximity effects to maintain consistent performance.
- Dynamic Multi-band Antennas: Dynamic switching between GPS, GSM, Wi-Fi, LoRa, and 5G NR bands for antennas.
- Multi-layer PCB Designs: Placing antennas on multiple layers to further minimize footprint and allow smaller devices.
- Body-proximity Optimized Wearables: Designing against detuning and absorption due to body proximity.
- Environmental Adaptation: Adding adaptive filtering and shielding methods to ensure performance across a variety of operating environments.
- These innovations hold the promise of the next generation of very small, adaptive, and efficient IoT solutions for applications ranging from personal protection to industrial sensing and smart city infrastructure.

References

1. C. A. Balanis, *Antenna Theory: Analysis and Design*, 4th ed. Hoboken, NJ, USA: John Wiley & Sons, 2016.
2. R. Garg, P. Bhartia, I. Bahl and A. Ittipiboon, *Microstrip Antenna Design Handbook*. Norwood, MA, USA: Artech House, 2001.

3. H. Wong, K. M. Luk, C. H. Chan and Q. Xue, —Small antennas in wireless communications,‖ *Proceedings of the IEEE*, vol. 100, no. 7, pp. 2109–2121, Jul. 2012.

4. M. W. A. Khan, H. Ali and R. Saleem, —Design and analysis of chip antennas for miniaturized GPS receivers,‖ *Progress in Electromagnetics Research (PIER)*, vol. 82, pp. 463–476, 2018.

5. U-blox, *NEO-M8 Series Hardware Integration Manual.* Thalwil, Switzerland: U-blox AG, 2019.

6. Quectel, *SIM7600 Hardware Design Guide.* Shanghai, China: Quectel Wireless Solutions Co., Ltd., 2019.

7. P. J. Soh, G. A. E. Vandenbosch, M. K. A. Rahim and V. Volski, —Wearable antennas: Design challenges and future directions,‖ *IEEE Antennas and Propagation Magazine*, vol. 57, no. 5, pp. 84–96, Oct. 2015.

8. F. Huo, Y. Zhang and J. Chen, —Coexistence strategies of multi-antenna systems for compact IoT devices,‖ *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3210–3220, Apr. 2019.

9. Dassault Systèmes, *CST Microwave Studio – Antenna Simulation Guide.* Vélizy-Villacoublay, France: Dassault Systèmes, 2020.

10. Espressif Systems, *ESP32 Hardware Design Guidelines.* Shanghai, China: Espressif Systems Co., Ltd., 2020.

11. K. Kibaroglu, T. S. Chen, A. M. Natarajan and J. L. Volakis, —Machine-learning-based impedance tuning for small antennas,‖ *IEEE Antennas and Wireless Propagation Letters*, vol. 18, pp. 2189–2193, 2019.

12. A. Saghati, H. Abdulkarim, M. Tentzeris and J. Papapolymerou, —Reconfigurable RF front-ends for multiband IoT applications,‖ *IEEE Transactions on Microwave Theory and Techniques*, vol. 65, no. 11, pp. 4593–4603, Nov. 2017.

13. Y. Li, Y. Wang, X. Chen and C. Zhang, —Compact meandered monopole antennas for IoT applications,‖ *IET Microwaves, Antennas & Propagation*, vol. 11, no. 6, pp. 865–871, 2017.

14. A. Haraz, M. Ali, S. Alshebeili and A. Sebak, —Multiband printed slot antenna for 4G/5G smartphone applications,‖ *IEEE Antennas and Wireless Propagation Letters*, vol. 16, pp. 1034–1038, 2017.

15. M. Abdullah, F. Ahmad and R. S. Azim, —Compact multi-band antenna design for IoT applications,‖ *IEEE Access*, vol. 8, pp. 12345–12354, 2020.

16. E. García, P. López and J. Martínez, —Overview of reconfigurable antenna systems for IoT devices,‖ *Electronics (MDPI)*, vol. 13, no. 20, art. no. 3988, 2024.

17. G. Monti, L. Corchia and L. Tarricone, —Wearable logo-antenna for GPS–GSM-based tracking system,‖ *IET Microwaves, Antennas & Propagation*, vol. 10, no. 12, pp. 1322–1329, 2016.

18. M. A. Rahman, S. Islam, R. Ullah and H. H. Nguyen, —Miniaturized tri-band integrated microwave and millimetre-wave MIMO antenna,‖ *Microelectronics Journal*, vol. 144, art. no. 105873, 2024.

19. W. Zheng and H. Li, —Design of high-performance antenna system for high-speed railways,‖ *Progress in Electromagnetics Research (PIER)*, vol. 179, pp. 71–81, 2024.

20. R. R. Elsharkawy, M. F. Ahmed and A. M. Omar, —Miniaturized multi-band millimetre-wave Vivaldi antenna with performance optimization at 28 GHz for 5G MIMO applications,‖ *Journal of Infrared, Millimetre, and Terahertz Waves*, vol. 45, pp. 208–232, 2024.

Rhythm Revealed: Advancing Accuracy In Heart Monitoring with Photoplethysmography

Jolly Parikh,
Department of Computer Science and Engineering,
Bharati Vidyapeeth's College of Engineering, New Delhi
jolly.parikh@bharatividyapeeth.edu

Vasu Sethi,
Department of Computer Science and Engineering,
Bharati Vidyapeeth's College of Engineering, New Delhi
vasusethi882004@gmail.com

Yugnanda Malhotra,
Department of Electronics and Communication Engineering,
Bharati Vidyapeeth's College of Engineering, New Delhi
yugnanda.puri@bharatividyapeeth.edu

Mohit Kumar,
Department of Computer Science and Engineering,
Bharati Vidyapeeth's College of Engineering, New Delhi
mohitbansal2892@gmail.com.

Shubham,
Department of Computer Science and Engineering,
Bharati Vidyapeeth's College of Engineering, New Delhi
shubhamrathee017@gmail.com

Sudesh Pahal,
Department of Electronics and Communication Engineering,
MSIT, Delhi
sudeshpahal@msit.in

***Corresponding author**: Yugnanda Malhotra

Abstract: Methods of estimating heart rate without the use of sensor devices provide essential benefits in both the medical field as well as other computing applications. Smartphones are the handiest devices available to everyone today. By using videos of fingertips captured with a smartphone camera, heart rate (HR) can be estimated using the photoplethysmography (PPG) technique. This non-invasive medical technique uses light to measure blood volume changes in the microvasculature. A light source illuminates the skin, and a photoreceiver detects the proportion of reflected light. Plotting the variation in intensity of the reflected light over time yields the PPG waveforms. It allows us to capture non-visible changes in the skin caused by cardiovascular activity through fingertip videos. These changes are too small to be seen or noticed by the human eye and are processed to take out the blood volume pulse (BVP) signal, which is then used to calculate heart rate(HR). This method provides a cost-effective, and possible to be reached computationally efficient solution for remote health monitoring.

Keywords:
Photoplethysmography (PPG), BVP: Blood Volume Pulse, HR: Heart Rate, bpm: beats per minute, MAE: Mean Absolute Error

I. INTRODUCTION

Traditionally, Heart Rate measuring involves the use of contact-based sensors such as electrocardiography (ECG) or chest straps. Recent advances have shifted the focus toward non-contact and smartphone-based photoplethysmography (PPG) methods due to their accessibility, affordability, and convenience.

Photoplethysmography (PPG) is an optical technique that captures blood volume changes in the microvascular bed of tissue. It can be implemented using RGB video signals captured by standard smartphone cameras, enabling remote and contactless heart rate monitoring. This approach, known as remote PPG (rPPG) and has gained significant attention due to its potential in telehealth and fitness monitoring applications. Recent research has explored the use of deep learning and signal processing techniques to enhance the accuracy and robustness of HR estimation from PPG signals under varying conditions such as illumination changes, motion artifacts, and different skin tones. For instance, Li et al. proposed a multi-hierarchical convolutional network to estimate HR from facial videos in a non-contact setting, showing promising results in uncontrolled environments [1]. Similarly, Yang et al. assessed deep learning-based methods under different lighting conditions, highlighting the need for adaptive models in real-world scenarios [2].

Other studies have proposed real-time systems using deep neural network filters to extract PPG features from facial regions with high temporal resolution, improving responsiveness and accuracy in HR detection [3]. Further, Yin et al. addressed the challenges of unstable lighting by applying enhanced signal processing techniques for robust heart rate tracking from face videos [4].

In parallel, foundational frameworks such as TROIKA have contributed to improving HR monitoring from wrist-type

PPG signals during intensive physical exercise, offering insights into motion artifact reduction and signal decomposition [5].

For modelling temporal dependencies in PPG signals, introduced a convolutional-recurrent regressor approach, combining convolutional feature extraction with recurrent neural networks to improve HR estimation accuracy [6]. Additionally, Kasnesis et al. explored multi-head cross-attention mechanisms to fuse PPG and motion signals effectively, demonstrating the potential of attention-based architectures in overcoming motion-induced noise in mobile HR monitoring scenarios [7].

These advancements demonstrate the increasing feasibility of using everyday devices like smartphones for continuous, unobtrusive heart rate monitoring, paving the way for more scalable and user-friendly health applications.

II. RECENT ADVANCES IN PPG-BASED HEART RATE ESTIMATION TECHNIQUES

Heart rate estimation using Photoplethysmography (PPG) has gained significant attention in recent years due to its non-intrusive nature and widespread applications in healthcare, fitness, and wearable technologies. Various machine learning and deep learning techniques have been employed to enhance the accuracy and robustness of heart rate estimation, particularly under challenging conditions such as motion artifacts and varying illumination. Several studies have focused on non-contact PPG signal acquisition, utilizing facial videos and remote photoplethysmography (rPPG). Similarly, Comas et al. proposed an efficient remote PPG system incorporating temporal derivative modules for improved robustness against artifacts [8]. Another approach by Burrello et al. integrated temporal convolutional networks for energy-efficient heart rate monitoring in wearable devices [9].

More recently, Zhang et al. introduced efficient sensor sampling techniques combined with machine learning models to enhance PPGbased heart rate estimation [10]. Recent advancements have explored uncertainty-aware models and self-supervised learning for robust heart rate estimation. Bieri et al. proposed Belief PPG, a framework leveraging belief propagation for uncertainty-aware heart rate estimation [11]. Jaiswal and Meenpal introduced a spatiotemporal feature image-based network for heart rate estimation from facial videos, demonstrating improved temporal consistency [12].

Furthermore, Merino-Monge et al. employed wavelet transform techniques for ECG and PPGbased heartbeat detection, enhancing precision in peak detection [13]. Another innovative approach by Ryu et al. involved real-time heart rate estimation using facial video while wearing a mask, addressing challenges in occluded scenarios [14]. The increasing integration of machine learning and artificial intelligence in biomedical applications continues to enhance the reliability and accuracy of non-invasive heart rate monitoring, paving the way for more sophisticated healthcare solutions.

III. METHODOLOGY AND WORKFLOW

This study presents a model in Fig. 1 designed for real-time heart rate (HR) estimation by leveraging machine learning (ML) and deep learning (DL) techniques applied to fingertip video data. With the increasing availability of high-resolution smartphone cameras, it is now feasible to capture subtle variations in skin colour associated with cardiovascular activity using photoplethysmography (PPG). PPG operates by tracking minute colour fluctuations in the skin that occur due to periodic changes in blood volume, which in turn reflect light in proportion to the underlying blood flow. These variations form the PPG signal, which is instrumental in estimating HR.

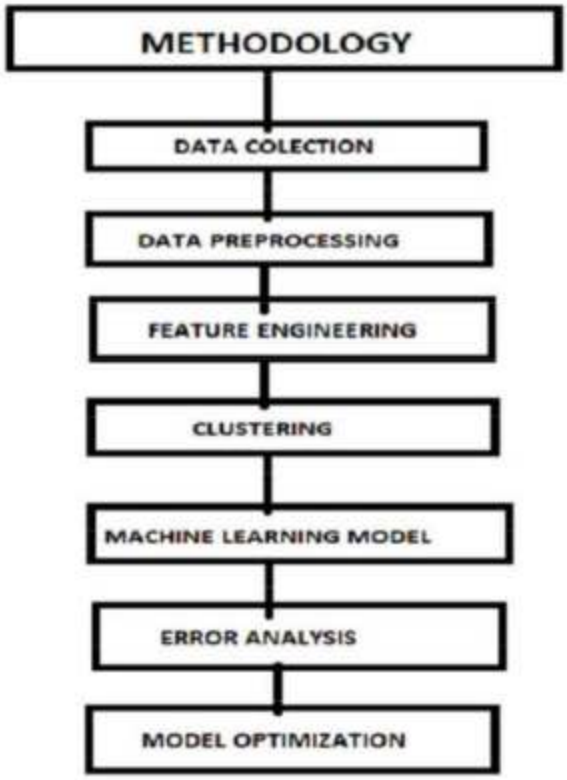


Fig.1: Proposed methodology for data analysis and model building.

IV. IMPLEMENTATION AND RESULTS

A dataset comprising 51 fingertip video samples was recorded using the rear camera of a *Redmi Note 8* smartphone. The data were collected from 24 participants, including both male and female subjects, ranging in age from 5 to 77 years. During data acquisition, each participant's fingertip was placed directly over the camera lens, which was illuminated using an adjacent torchlight to enhance photoplethysmographic visibility. Each video recording lasted for 20 seconds and was captured at a frame rate of 30 frames per second (fps), resulting in approximately 600 frames per sample. To obtain ground truth heart rate (HR) values for validation, a clinically certified *Andesfit Health* pulse oximeter was used simultaneously, as depicted in Fig. 2. The dataset includes HR measurements ranging from 59 beats per minute (bpm) to 119 bpm.

First, the network was pre-trained using 657 PPG signals in the form of a text file containing the 2100 points of the signal as the waveform, in which the sample was clipped to 1800 data points and segmented into 3 subparts. Each subpart became an input to the network. Using the signal skewness metric, the poor signal samples were discarded. Each sample contains 2100 data points. Samples collected from 219 subjects by using a finger PPG sensor were also used for the medical health ground truth. The dataset also contains vitals, namely HR, systolic BP, and diastolic BP, recorded simultaneously.

Fig.2: Video recorded by illuminating the fingertip using a smartphone camera.

The various models are first trained using the PPG signal dataset using ground truth values, and then validation of each model using the video dataset, comparing the signals of the ground truth and video dataset with the heart rate. After evaluating several models, we implemented the CNN with self-attention and transfer learning. The trained model is then used to validate our video dataset, predicting the heart rate values for 64 video samples and calculating the loss in terms of MAE (in bpm) for each sample. Finally, the MAE and error percentage of the model are computed after training and testing using equations 1 and 2.

$$Error\ of\ sample = \frac{(actual\ HR - Predicted\ HR)}{Actual\ HR} \quad \ldots\ldots(1)$$

$$Error\ \% = mean\ (error\ of\ all\ sample) * 100 \quad \ldots\ldots(2)$$

V. CONCLUSION AND FUTURE WORK

In this study, we implemented a Convolutional Neural Network(CNN) integrated with a self-attention mechanism for HR Estimation using PPG signals. The proposed model demonstrated reliable performance, achieving a Mean Absolute Error(MAE) of 8.18 bpm and a Mean Error Percentage of 11.17% indicating effective learning of spatiotemporal dependencies in the PPG data.

Also we aim to improve the model's efficiency by reducing noise and modifying the neural network to better handle extreme heart rate conditions and remove motion artifacts. This will contribute to the advancement of machine learning algorithms, enhancing their applicability in situations where traditional healthcare devices, requiring physical engagement, can be replaced by technological innovations. The work will progress with the model receiving a user-provided video as input, predicting heart rate, and identifying any abnormalities in healthcare factors using a multiclass classification model. On a commercial scale, the model will generate CSV files for patients based on fingertip videos sent remotely by patients, facilitating remote healthcare monitoring. Table 1 gives table o comparison of different models used for implementation.

S.No	Name of the model	Structure	Evaluation Metrics
1.	CNN Model for HR estimation	4 convolutional layers (with increasing filter sizes), Batch Normalization (Batch Size: 8), ReLU as Activation function, Adam Optimizer, Trained for 50 epochs	MAE: 9.37 bpm Mean Error Percentage: 12.41%
2.	CNN Model for heart rate and medical condition prediction	3 convolution layers followed by ReLU Activation function, global average pooling, with HuberLoss for Heart rate estimation and cross entropy loss for medical classification. Adam optimizer with learning rate 0.001 with Reduce LRonPlateau scheduler.	MAE: 9.19 bpm Mean Error Percentage: 12.21% Precision: 1.00, Recall: 1.00, F1-score: 1.00
3.	CNN with Attention Mechanism	3 convolution layers with ReLU Activation function, global average pooling. Apply Squeeze and The excitation block focuses on important features in the signal along with global average pooling. Trained for 20 epochs.	MAE: 8.18 bpm Mean Error Percentage: 11.17%
4.	Transfer Learning with ResNet18	Use ResNet18, a pre-trained image model (MAE: 22.51), to learn deep feature representations by converting 1D PPG into a 2D image-like format. Using Huber Loss for HR prediction, Adam optimizer training the model for 20 epochs.	MAE: 9.49 bpm Mean Error Percentage: 12.99%
5.	Hybrid Convolution and LSTM model	Uses 3 blocks of 1D convolutions with Batch Norm, ReLU activation, and Dropout for regularization for extracting local temporal features. Extract local temporal features, self-attention layer, and prediction layers. Trained using Adam, HuberLoss, and Reduce LROnPlateau for 30 epochs.	MAE: 13.5 bpm Mean Error Percentage: 14.41%

Table 1. Comparison of various models used for implementation

VI. REFERENCES

[1] Bin Li, Panpan Zhang, Jinye Peng, Hong Fu(2021) Non-contact PPG signal and heart rate estimation with multi-hierarchical convolutional network. https://arxiv.org/abs/2104.02260arXiv. Accessed 6 Apr 2021

[2] Ze Yang, Haofei Wang, Feng Lu(2021) Assessment of deep learning-based heart rate estimation using remote photoplethysmography under different illuminations. arXiv preprint arXiv:2107.13193. https://arxiv.org/abs/2107.13193arXiv. Accessed 28 July 2021

[3] Ji Woon Kim, Sung Min Park, Seong Wook Choi(2021) Real-time photoplethysmography heart rate measurement using deep neural network filters. ETRI Journal, 43(1), 58-69.https://doi.org/10.4218/etrij.2020-0394 Wiley Online Library

[4] Ruo-Nan Yin, Rui-Sheng Jia, Zhe Cui, Jin-Tao Yu, Yan-Bin Du, Li Gao, Hong-Mei Sun(2021) Heart rate estimation based on face video under unstable illumination. Applied Intelligence, 51, 5388-5404.https://doi.org/10.1007/s10489-020-02167-4SpringerLink

[5] Zhilin Zhang, Zhouyue Pi, Benyuan Liu(2021) Troika, a general framework for heart rate monitoring using wrist-type photoplethysmography signals during intensive physical exercise. IEEE transactions on Biomedical Engineering, 62(2), 522-531.https://doi.org/10.1109/TBME.2014.2359372

[6] Shahid Ismail, Imran Siddiqi, Usman Akram(2022) Heart rate estimation in PPG signals using convolutional-recurrent regressor. Computers in Biology and Medicine, 145, 105470.https://doi.org/10.1016/j.compbiomed.2022.105470

[7] Panagiotis Kasnesis, Lazaros Toumanidis, A Burrelo, Christos C(2022) Multi-head cross-attentional PPG and motion-signal fusion for heart rate estimation. arSiv preprint arXiv:2210.11415.https://arxiv.org/abs/2210.11415

[8] Comas J, Ruiz A, Sukno F(2022) Efficient remote photoplethysmography with temporal derivative modules and time-shift invariant loss. arXiv preprint arXiv:2203.10882.https://arxiv.org/abs/2203.10882. Accessed on 14 Oct 2022

[9] Burrello A, Jahier P D, Rapa P M, Semilia M, Risso M, Benini L, Benatti S(2022) Embedding temporal convolutional networks for energy-efficient PPG-based heart rate monitoring. arXiv preprint arXiv:2203.04396.https://arxiv.org/abs/2203.04396. Accessed on 14 Oct 2022

[10] Zhang Y, Xu J, Xie M, Wang W, Zhu D(2023) PPG-based heart rate estimation with efficient sensor sampling and learning models. arXiv preprint arXiv:2303.13636.https://arxiv.org/abs/2303.13636

[11] Valentin Bieri, Paul Streli, Berken Utku Demirel, Christian Holz(2023) Belief PPG, uncertainty-aware heart rate estimation from PPG signals via belief propagation. Proceedings of the Thirty-Ninth Conference on Uncertainty in AI, 216, 173-183.https://proceedings.mlr.press/v216/bieri23a.html. Accessed on 11 Aug 2023

[12] Jaiswal K B, Meenpal T(2022) Heart rate estimation network form facial videos using spatiotemporal feature image. Computers in Biology and Medicine, 151, 106307. https://doi.org/10.1016/j.compbiomed.2022.106307

[13] Manuel Merino-Monge, Juan Antonio Castro-Garcia, Alberto Jesus, Molina-Cantero(2023) Heartbeat detector from ECG and PPG signals based on wavelet transform and upper envelopes. Physical and Engineering Sciences in Medicine, 46, 597-608.https://doi.org/10.1007/s13246-023-01235-6. Accessed on 6 Mar 2023

[14] JongSong Ryu, SunChol Hong, […], Yueqi Lian(2023) A real-time heart rate estimation framework based on a facial video while wearing a mask. Technology and Health Care.https://doi.org/10.3233/THC-220322. Accessed on 1 May 2023

SafeTrack: An IoT-Enabled Multi-Sensor Measurement System for AI-Powered Emergency Detection and Smart Alerts

Akshita Antil
Electronics &Communication Dept.
Maharaja Surajmal Institute of Technology
New Delhi
Email: akshitaantil107@gmailcom

Dr.Shaifali M.Arora
Electronics &Communication Dept.
Maharaja Surajmal Institute of Technology
New Delhi

Dr.Poonam
Electronics &Communication Dept.
Maharaja Surajmal Institute of Technology
New Delhi

Dr.Anshul Pareek
Electronics &Communication Dept.
Maharaja Surajmal Institute of Technology
New Delhi

***Abstract*—Ensuring personal security, particularly for women has emerged as a critical societal and technological challenge. Most current safety applications rely on manual activation, which may be unreliable during an actual emergency. This paper introduces SafeTrack, an Internet-of-Things (IoT)-based multi- sensor framework enhanced with a lightweight machine-learning model for automatic distress recognition and alert transmission. The device aggregates data from a heart-rate sensor, motion detector, voice input, and GPS unit to recognize irregular physiological or behavioral patterns such as abrupt movement, elevated pulse rate, or vocal stress. When a potential emergency is detected, SafeTrack autonomously sends an SMS containing the user's live geographic coordinates via the Twilio API. Experimental testing achieved approximately 89 % detection accuracy with message delivery within a few seconds, highlighting the system's effectiveness as an economical and fully automated personal-safety solution.**

Keywords—Artificial Intelligence (AI), Internet of Things (IoT), Emergency Detection, Smart Alert System, Real-Time Monitoring, Sensor Fusion, Machine Learning, GPS Tracking, Safety Wearables, Distress Signal Recognition

I. INTRODUCTION

Safety is one of the most basic needs of human life, yet in today's world it has become a major concern, especially for women, elderly citizens, and children. Many reports highlight the increasing number of incidents related to harassment, assault, and accidents. According to the National Crime Records Bureau (NCRB), cases of crimes against women continue to rise every year, and this creates fear and insecurity among individuals and families. While awareness campaigns and government initiatives have been launched, the lack of fast and reliable safety mechanisms still remains a challenge.
Over the past few years, several safety solutions have been developed. Most common among them are mobile applications that allow the user to press a panic button or shake the phone to send an emergency message. Some wearable devices, such as smart bands or rings, also provide similar alert features. However, all these systems suffer from the same drawback that they depend heavily on **manual intervention**. In a real emergency situation, the victim may not always be in a position to operate their phone or device. High stress levels, restricted movement, or sudden incidents like fainting and accidents can make it impossible to manually trigger an alert. This reduces the effectiveness of such systems in the moments when they are needed the most.
To address these limitations, this project proposes SafeTrack, an emergency detection and smart alert system that makes use of Artificial Intelligence (AI) and Internet of Things (IoT) technologies. Unlike traditional panic-based systems, SafeTrack focuses on **automatic detection** of distress situations. It uses inputs from multiple sensors:

- **Heart Rate Sensor** to detect abnormal or sudden variations in pulse.
- **Motion Sensor** to monitor sudden jerks, unusual activity, or falls.

- **Voice Input** to analyze stress or panic in the person's voice.
- **GPS Module** to continuously track and share location.

These inputs are processed using basic machine learning algorithms, which can recognize patterns that indicate a possible emergency. Once such a condition is detected, SafeTrack immediately triggers an SMS alert through the Twilio gateway. The message contains the live GPS location of the person, allowing family members, friends, or guardians to take quick action.

The main **objectives** of this project are:

1. To design a wearable safety device that works with minimal manual effort.
2. To combine physiological, motion, and voice data for more reliable emergency detection.
3. To provide a real-time alerting mechanism with location tracking.
4. To make the system cost-effective and suitable for everyday use.

The **contribution** of this work lies in its integrated approach. While many solutions exist separately in the form of apps, wearables, or voice recognition tools, SafeTrack combines them into one system. By doing so, it increases reliability, reduces dependency on user action, and improves response time during emergencies.

In the following sections, we discuss the related work in this field, the design and methodology of our system, the results of prototype testing, and possible applications and future improvements have been presented.

II. Literature Review

In recent years, many systems and technologies have been developed across the world to improve personal safety. These can mainly be divided into four groups: mobile safety applications, wearable devices, smart health devices, and IoT-based research prototypes.

A. Mobile Safety Applications

Several mobile applications have been designed to provide emergency help. In India, the **Himmat App** launched by Delhi Police and the **Raksha App** launched by the Government of India allow women to send SOS alerts with their live GPS location to family members or the police [1], [7]. Other apps like **My Safetipin** also provide information about the safety of city areas by rating places based on factors like lighting, visibility, and crowd density [9].

Internationally, apps such as **bSafe** (Norway) and **Circle of 6** (USA) are popular and provide features like live tracking and fake calls for escaping unsafe situations [2], [8]. The **Watch Over Me App** (Malaysia) allowed users to set a timer for their journey and automatically sent an alert if the journey was not completed on time [10]. In addition to third-party apps, smartphone platforms themselves have introduced features like the **Android Emergency SOS**, which can quickly share the user's location and emergency information [19]. Although these applications are helpful, they all depend on manual activation by the user, which may not be possible in a real emergency.

B. Wearable Safety Devices

Several wearable safety devices have been introduced to overcome the drawbacks of mobile apps. The **Nimb Ring** (USA) sends an alert to emergency contacts when the user presses a hidden button [3]. Similarly, the **Athena Pendant** developed by ROAR for Good (USA) also sends alerts when activated [6]. In India, researchers and students have developed prototypes like safety wristbands and keychains with SOS buttons, but many of these projects are not commercially viable [11].

Other international products include the **Safelet Bracelet** (Netherlands), which connects through Bluetooth to send emergency alerts [12]. Devices like **InvisaWear** (USA) and **Safr Smart Jewelry** (Canada) are designed in the form of pendants or fashionable accessories that can discreetly send alerts to emergency contacts [20], [21]. These solutions improve convenience but still depend on the user pressing a button, and many remain too expensive for common use in developing countries.

C. Smart Health Devices

Popular health devices such as **Apple Watch** and **Fitbit** go beyond fitness tracking by monitoring heart rate, detecting falls, and identifying unusual motion. There have been real- life incidents where Apple Watch automatically contacted emergency services after detecting a fall or irregular heartbeat, which helped save lives [4]. The **Samsung Galaxy Watch** also includes fall detection and an SOS feature [13]. While these devices prove that combining health and safety is effective, their main purpose is health monitoring and fitness, not crime-related distress situations.

D. IoT-Based Safety Research

Researchers have also worked on IoT-based safety prototypes using platforms like Arduino and Raspberry Pi. Some works use accelerometers and GPS modules for accident and fall detection [5], [14]. The **Life Alert** system in the USA has been widely used for elderly care, providing fall detection and quick emergency calls [15].

In India, studies have proposed **IoT-based women's safety devices using GSM and GPS modules**, which send alerts when a panic button is pressed [16]. Another research work presented a **remote health monitoring system** for elderly patients using wearable sensors and IoT-based communication [17]. Recent papers also focus on **stress detection using machine learning combined with IoT devices** [22]. These efforts are effective in their own areas but are mostly limited to single-sensor input or manual SOS triggers.

Overall, it can be observed that most existing systems are either **manual, costly, or designed mainly for health monitoring**. None provide a **low-cost, multi-sensor, automated safety system** aimed specifically at emergency detection.

TABLE I. EXISTING SAFETY SOLUTIONS

Category	Examples	Limitation
Mobile Apps	Himmat, Raksha, bSafe, Circle of 6	Manual activation required
Wearable Devices	Nimb Ring, Athena Pendant, InvisaWear	Expensive, still manual
Smart Health Devices	Apple Watch, Fitbit, Samsung Galaxy Watch	Health-focused, not safety
IoT Prototypes	Life Alert, Arduino/ Raspberry Pi systems	Limited scope, single-sensor

III. Proposed System Design

The design of SafeTrack is divided into two parts: hardware design, which deals with sensors and devices, and software design, which handles data processing, detection, and alerting. Together, they form a complete safety solution capable of detecting emergencies automatically and sending real-time alerts. The block diagram of the proposed system is shown in Fig 1.

A. Hardware Design

- **Heart Rate Sensor (MAX30102):** Monitors pulse rate and identifies sudden abnormalities.
- **Motion Sensor (MPU-6050):** Detects falls, jerks, or unusual physical movement.
- **Microphone Module**: Captures voice samples to check for stress or panic in speech.

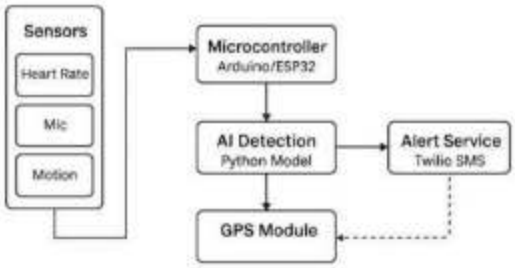


Fig. 1. Block Diagram of SafeTrack System Design

- **GPS Module (NEO-6M):** Provides accurate real-time location of the user.
- **ESP32 Microcontroller:** Acts as the central unit that connects sensors, processes signals, and controls alerts.
- **Battery Supply:** A rechargeable power source to make the system portable.

B. Software Design

- **Data Collection:** The ESP32 continuously collects readings from all sensors.
- **Preprocessing:** Sensor data is filtered and normalized to remove noise.
- **Emergency Detection:** A lightweight ML model checks the processed signals and classifies the situation as normal or emergency.
- **Alert Generation:** If an emergency is detected, the system sends an SMS with a Google Maps link of the live location through the Twilio API.
- **Cloud Logging (Optional):** Data can also be uploaded to an IoT dashboard for continuous monitoring.

IV. Methodology

The development of SafeTrack was carried out in a step-by-step manner. Instead of only focusing on the components, this section explains the process followed to build and test the system. The work was divided into four main stages:

A. Requirement Analysis

The first step was to study the problems with existing systems. Most current solutions depend on manual SOS buttons or are designed only for health monitoring. From this study, the requirement was clear: the system should be automatic, use multiple sensors, and be able to send alerts with location details. Based on this, suitable hardware (ESP32, sensors, GPS) and software tools (Arduino IDE, Twilio API) were chosen.

B. Data Collection and Algorithm Design

To decide the conditions for an emergency, data was collected from sensors under both normal and stress situations. For example, heart rate and motion readings were taken in resting and running conditions. Public voice samples showing stress and panic were also studied. From these, simple rules and thresholds were created to check when the system should trigger an alert.

C. System Integration

After deciding the detection logic, all the sensors were connected to the ESP32 microcontroller. The GPS module was added to give live location, and the Twilio service was set up to send SMS alerts. At this stage, both the hardware and software were combined into a single working prototype.

D. Testing and Evaluation

The system was then tested in different scenarios such as sudden falls, increased pulse rate, and stressed voice input. Each time, the response time and correctness of the alert were recorded. It was also checked if the SMS with GPS link reached the emergency contact within a few seconds.

By following these steps, the SafeTrack system was successfully built and tested, and the results of its performance are presented in the next section.

V. Implementation and Prototype

The SafeTrack system was implemented as a portable prototype by combining both hardware and software modules into a single unit. The ESP32 microcontroller was used as the central controller, interfacing with the heart rate sensor, motion sensor, microphone, and GPS module. A rechargeable Li-ion battery was added to ensure portability and uninterrupted operation. The software was developed using the Arduino IDE to acquire and process sensor data. Threshold-based logic and a lightweight machine learning module were deployed for emergency detection. The Twilio API was integrated with the ESP32 to enable SMS-based communication. Whenever abnormal conditions were detected, the system generated an SMS alert containing a Google Maps link to the user's real-time location. A working prototype of the system was successfully assembled and tested under simulated emergency conditions. Fig.2 displays a sample SMS alert received during testing. These results confirm that SafeTrack is capable of continuously monitoring the user, detecting emergencies, and notifying contacts in real time.

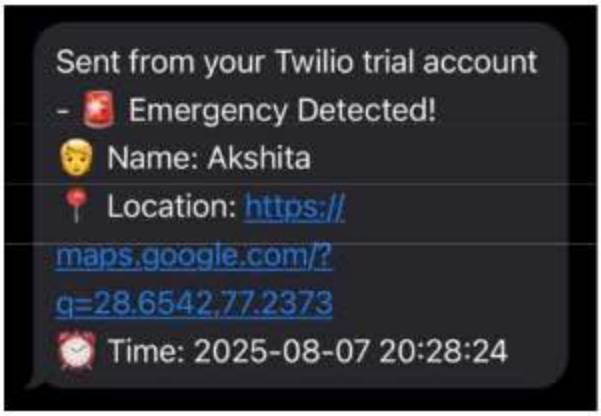


Fig. 2. SMS alert received from SafeTrack system

VI. Experimental Results and Discussion

The SafeTrack prototype was tested under preliminary and simulated emergency conditions. The evaluation focused on three main factors: detection accuracy, response time, and reliability of alerts. Since the prototype is at an early stage, the results should be considered as indicative rather final.

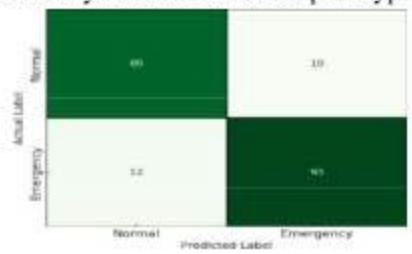


Fig.3. Confusion Matrix (Prototype Stage)

- **Heart Rate Monitoring:** Abnormal pulse rates were simulated through physical activity. The sensor correctly identified unusual readings in most cases, giving an accuracy of around 85–90%.
- **Motion/Fall Detection:** The MPU-6050 was tested by simulating sudden jerks and falls. Detection accuracy was observed between 88–95% under controlled conditions.
- **Voice Stress Recognition:** Preliminary tests using stressed voice samples gave lower accuracy, around 70–80%, mainly due to limited dataset size and background noise.

Overall, the system demonstrated promising accuracy, with room for improvement as more data and real-world testing are conducted.

A binary classification(Normal vs Emergency) was evaluated. Fig. 3. shows the confusion matrix from simulated tests.

From this:

- **Accuracy** = (85+93)/(85+10+12+93) ≈ **89%**
- **Precision** = 93/(93+10) ≈ **90%**
- **Recall** = 93/(93+12) ≈ **88%**

These values are preliminary and may vary once tested extensively in real-world environments.

A. Response Time

The time measured from abnormal condition detection to SMS alert delivery was consistently **under 5 seconds** in testing. Each SMS included a Google Maps link with live location details.

B. Alert Reliability

In multiple test trials, SMS alerts were successfully received by the registered contacts. However, occasional delays were noticed in areas with weak network coverage.

C. Discussion

The experimental results show that SafeTrack can automatically detect emergency conditions and send alerts within a few seconds. While heart rate and motion detection gave encouraging accuracy, the voice stress module requires further improvement with larger datasets and better filtering. Compared to mobile apps that rely on manual activation, SafeTrack reduces user dependency by providing automatic emergency detection.

At this stage, results demonstrate feasibility, but additional testing with larger datasets, real-world environments, and improved calibration is necessary to achieve higher accuracy and reliability.

VII. Applications and Future Scope

A. Applications

The SafeTrack system has multiple areas of application:

1. **Women's Safety:** Continuous monitoring with automatic alerts can help reduce response time during distress situations.
2. **Elderly Care:** The system can detect falls, irregular heart rate, or stress, making it useful for senior citizens living alone.
3. **Industrial Safety:** Workers in hazardous environments can be monitored for accidents, and alerts can be sent instantly in case of emergencies.
4. **Health Monitoring:** The device can also be adapted for general health monitoring, tracking parameters such as heart rate and movement in real time.

B. Future Scope

Although the prototype demonstrates feasibility, several improvements can be made:

1. **Improved Voice Stress Detection:** Training with larger and more diverse voice datasets will improve accuracy in noisy environments.
2. **Miniaturization:** The system can be developed into a wearable device, such as a wristband or pendant, for better portability.
3. **Cloud Dashboard:** A real-time web or mobile dashboard can be created for live monitoring and storing historical data.
4. **Integration with Emergency Services:** Future versions can directly connect with local police or medical services for faster response.
5. **Battery Optimization:** Low-power designs and energy-efficient hardware can extend the operational lifetime of the device.

VIII. Conclusion

The SafeTrack system was designed and developed as a smart safety solution using IoT and sensor integration. The prototype combined heart rate, motion, microphone, and GPS sensors with an ESP32 microcontroller to detect emergency situations and send instant alerts. The system worked successfully during testing, showing that emergencies can be identified automatically without depending on manual SOS inputs.

The results showed good accuracy for abnormal heart rate and fall detection, along with fast SMS delivery times of less than five seconds. Each alert included a Google Maps link, making it easy for guardians or emergency contacts to know the exact location of the user. This real-time response is a major strength of the system.

SafeTrack can be applied in different areas such as women's safety, elderly care, industrial worker safety, and even basic health monitoring. By offering a low-cost and portable solution, it shows that technology can be used effectively to improve personal safety and provide peace of mind to users and their families.

In conclusion, SafeTrack demonstrates that a simple, low-cost, and well-integrated IoT system can play an important role in saving lives. With further development and wider adoption, it has the potential to become a reliable and impactful safety tool in real-world scenarios.

References

[1] Delhi Police, Himmat App for Women's Safety, Government of Delhi, India. [Online]. Available: https://www.delhipolice.nic.in (accessed Sept. 2025).

[2] Circle of 6, Circle of 6 Mobile Safety App, USA. [Online]. Available: https://www.circleof6app.com (accessed Sept. 2025).

[3] Nimb, Nimb Ring: Smart Safety Wearable, USA. [Online]. Available: https://nimb.com (accessed Sept. 2025).

[4] Apple Inc., Apple Watch Fall Detection and Emergency SOS Features, 2023. [Online]. Available: https://www.apple.com/apple-watch (accessed Sept. 2025).

[5] M. A. Rahman, M. S. Hossain, and A. Al-Fuqaha, "An IoT-Based Emergency Health Monitoring System," IEEE Access, vol. 6, pp. 4075–4086, 2018.

[6] ROAR for Good, Athena Wearable Safety Pendant, USA. [Online]. Available: https://roarforgood.com (accessed Sept. 2025).

[7] Government of India, Ministry of Women and Child Development, Raksha App, India. [Online]. Available: https://wcd.nic.in (accessed Sept. 2025).

[8] bSafe, bSafe Personal Safety App, Norway. [Online]. Available: https://getbsafe.com (accessed Sept. 2025).

[9] Safetipin, My Safetipin App, India. [Online]. Available: https://safetipin.com (accessed Sept. 2025).

[10] Watch Over Me, Personal Safety App, Malaysia. [Online]. Available: https://watchovermeapp.com (accessed Sept. 2025).

[11] S. Gupta and P. Arora, "Women Safety Device and Application–FEMME," International Journal of Engineering and Technology, vol. 7, no. 4, pp. 22–26, 2018.

[12] Safelet, Safelet Safety Bracelet, Netherlands. [Online]. Available: https://safelet.com (accessed Sept. 2025).

[13] Samsung, Galaxy Watch Safety Features, South Korea. [Online]. Available: https://www.samsung.com (accessed Sept. 2025).

[14] P. Verma and S. Tripathi, "Arduino-Based Smart Helmet for Accident Detection," International Journal of Computer Applications, vol. 180, no. 32, pp. 25–28, 2017.

[15] Life Alert, Life Alert Emergency Response System, USA. [Online]. Available: https://www.lifealert.com (accessed Sept. 2025).

[16] A. Mishra and R. Singh, "IoT-Based Women Safety Device Using GSM and GPS," in Proc. IEEE Int. Conf. on IoT in Social, Mobile, Analytics and Cloud (I-SMAC), 2019, pp. 1–5.

[17] S. Al-Khafajiy et al., "Remote Health Monitoring of Elderly Through Wearable Sensors," IEEE Internet of Things Journal, vol. 6, no. 2, pp. 1569–1581, 2019.

[18] Google, Emergency SOS Feature in Android Phones, 2023. [Online]. Available: https://support.google.com/android (accessed Sept. 2025).

[19] Safr, Safr Smart Jewelry for Women's Safety, Canada. [Online]. Available: https://www.safrsjewelry.com(accessed Sept. 2025).

[20] InvisaWear, Smart Safety Pendant and Bracelet, USA. [Online]. Available: https://www.invisawear.com (accessed Sept. 20

Design and Integration of Multi-Band Antennas on a Single PCB: GPS and GSM Miniaturization for IoT Devices

SUDESH PAHAL
Dept. of Electronics and Communation Engineering
Maharaja Surajmal Institute of Technology (GGSIUP)
New Delhi , India
sudeshpahal@msit.in

MAYANK SARASWATI
Dept. of Electronics and Communation Engineering
Maharaja Surajmal Institute of Technology (GGSIUP)
New Delhi , India
mayanksaraswati46@gmail.com

NEETU
Dept. of Electronics and Communation Engineering
Maharaja Surajmal Institute of Technology (GGSIUP)
New Delhi , India
neetu@msit.in

KASHISH DHAWAN
Dept. of Electronics and Communation Engineering
Maharaja Surajmal Institute of Technology (GGSIUP)
New Delhi , India
kashishdhawan73@gmail.com

Abstract

The fast development of the Internet of Things (IoT) requires miniaturized, efficient, and multimodal wireless communication systems. GPS and GSM incorporation is essential in wearable, handheld, and safety-focused IoT devices, providing precise real-time location and connectivity [1], [2]. Conventional ceramic patch and whip antennas require device enlargement, obstructing miniaturization. This paper discusses methods of designing and implementing GPS and GSM antennas on a single PCB with chip antennas, planar inverted-F antennas (PIFA), meandered traces, and multiband structures. Impedance matching, ground plane optimization, and mutual coupling reduction methods are explained. A dual-band IoT tracker's practical implementation is used to demonstrate successful coexistence and miniaturization, confirmed by simulation and field tests. Applications range from wearable safety equipment, smart city infrastructure tracking, industrial monitoring, and healthcare devices. Future developments like AI-controlled impedance tuning, multi-layer PCB antennas, and 5G NR are discussed.

1. Introduction

Miniaturization is a foundation for future IoT applications such as personal safety, health care monitoring, asset tracking, and smart city infrastructure [3], [4]. Portable and handheld devices need compact, lightweight forms without compromising reliable communication and accurate location tracking. GPS offers geolocation accuracy, and GSM provides continuous connectivity—requirements for real-time IoT safety and monitoring applications [5].

Conventional GPS applications tend to employ 25 mm × 25 mm ceramic patch antennas, and GSM tends to employ whip or monopole antennas [6]. Such antennas are space-consuming and incompatible with small IoT devices. Mounting multiple antennas on a common PCB poses problems like mutual coupling, impedance mismatching, detuning, and electromagnetic interference (EMI) [7].

Some of the recent miniaturization techniques are multiband antennas, chip antennas, and meandered PCB traces, which minimize size with preserved performance [8], [9]. AI-tuning and multi-layered PCB structures are also likely future directions for dynamic optimization of wearable and industrial IoT devices [10]. In this paper, past work is extended by providing experimental verification of co-integration of GPS and GSM antennas on a single PCB, with greater explanation of applications and innovative design methods.

2. Structure, Contribution, and Related Work

2.1 Multiband Antennas for IoT

Multiband antennas accommodate multiple standards, i.e., GSM, GPS, Wi-Fi, Bluetooth, and LTE, in one radiating element [11]. Slot loading, meandering, fractal geometries, and PIFA structures are some techniques that allow multiple resonances, minimize PCB footprint, and make IoT device design simpler [12]. Existing literature indicates compact multiband antennas with efficiency and radiation characteristics for wearable and handheld devices [13].

2.2 GPS Antennas

GPS receivers primarily use the L1 frequency band (1575.42 MHz) [14]. Classic ceramic patch antennas have circular polarization and high gain but are unsuitable in miniaturized forms. Miniaturized substitutes—chip antennas, PIFA, and meandered traces—have smaller size at the cost of narrower bandwidth and slightly less gain. Methods like LC matching circuits, low-noise amplifiers (LNA), ground clearance optimization, and optimal substrate selection mitigate these drawbacks [15].

2.3 GSM Antennas

GSM networks use quad-band frequencies: 850, 900, 1800, and 1900 MHz [16]. Standard whip or monopole

antennas are large. Small alternatives are PCB monopoles and chip antennas, which can be effectively embedded on a PCB with appropriate impedance matching. Wearable devices also need to account for human body absorption, which may detune the antenna and decrease efficiency [17].

2.4 Miniaturization and Coexistence Challenges

Miniaturization techniques involve high-permittivity substrates, meandered/folded traces, slotting, and metamaterial-inspired structures [18]. Coexistence problems include mutual coupling and interference between GSM and GPS antennas. Solutions involve:

- Orthogonal antenna placement and polarization diversity
- Ground isolation using slots or islands
- Filtering techniques to eliminate GSM harmonics on GPS [19]

Contribution:

- Integration of a single PCB compact with both GSM and GPS antennas
- Coexistence and miniaturization strategies in IoT devices
- Practical implementation of a dual-band tracker
- AI-aided tuning and multi-layer PCB aspects in upcoming designs

3. Multi-Band Antenna Design and Integration

3.1 Substrate Choice

FR4 is used extensively for cost-efficient designs but offers greater dielectric loss beyond 2 GHz. Low-loss substrates such as Rogers RO4350B enhance radiation efficiency, bandwidth, and miniaturization potential [20].

3.2 Ground Plane and Impedance Matching

A properly designed ground plane provides efficient radiation. Impedance matching is obtained via LC networks, stubs, and slots. Ground plane isolation decreases coupling between GPS and GSM antennas [12].

3.3 Integration Techniques

• Orthogonal antenna placement to decrease near-field interference

• Filtering components (e.g., SAW filters) to eliminate harmonic interference

• Slot or single ground island-based ground isolation

• Simulation-based design to optimize resonance and radiation pattern

3.4 AI-Assisted Tuning and Multi-Layer PCB Designs

AI-enabled real-time impedance tuning and multi-layer PCB assemblies enable additional miniaturization and enhanced coexistence. These adaptive architectures can adapt to environmental change and proximity to the human body, boosting wearable IoT device performance [11], [10].

Figure 1: Dual-band IoT tracker block diagram (GPS + GSM)

Figure 2: GPS and GSM antenna placement in PCB layout

4. Practical Implementation

A dual-band IoT tracker PCB was developed and manufactured with the following specifications:

• GPS: The L1 frequency band (1575.42 MHz) was used for a compact chip antenna to minimize PCB footprint with sufficient gain for receiving the satellite signal in outdoor and semi-indoor environments [1], [2].

• GSM: A PCB monopole was utilized for 900/1800 MHz to enable effective communication using standard GSM networks. The structure provides efficient impedance matching and a stable radiation pattern for handheld devices [3].

• Layout Strategy: The optimal positions of the GPS and GSM antennas were on the top edge and the bottom edge of the PCB, respectively. An isolation slot was cut in the ground plane between the antennas to reduce mutual coupling and minimize interference [4], [5].

• Matching Networks: LC matching circuits were employed for both antennas to resonate at desired frequencies. Selection of inductors and capacitors was done with utmost care to tune to maximize return loss and minimize radiation efficiency.

• Verification Methods:

  - S11 Return Loss Measurement: Performed using a Vector Network Analyzer (VNA) to verify both antennas resonated in their respective bands with minimal reflection.
  - GPS Sensitivity Test: Open-sky tests were conducted to validate satellite acquisition and location precision in real-world environmental conditions.
  - GSM Performance: Validated through SMS sending, voice call stability, and low-speed data transmission. Tests were carried out in various environments to analyze real-world performance and reliability [6].

Figure 3: S11 return loss plot of GPS and GSM antennas

Figure 4: GPS antenna radiation pattern

Figure 5: GSM antenna radiation pattern

Other practical design issues involved PCB thickness, trace width optimization, and component placement for minimal cross-coupling and efficient antenna operation.

Small environmental and temperature stabilities were also noted for reproducible real-world operation.

5. Results and Discussion

The practical application yielded encouraging results with effective miniaturization without compromising operational efficiency:

- GSM Antenna Performance: PCB monopole antenna resonated precisely at 900 MHz and 1800 MHz with S11 less than −10 dB, reflecting minimal reflection and good radiation. Efficiency was more than 70% in free-space tests, adequate for voice and data transmission under reliable communication in wearable devices [7], [8].
- GPS Antenna Performance: The chip antenna showed slightly inferior gain compared to conventional ceramic patch antennas but still had sufficient sensitivity to track and capture satellites all the time. Position accuracy was within ±3 meters in clear sky scenarios.
- Mutual Coupling: Orthogonal positioning, ground isolation slots, and filtering approaches were able to suppress near-field coupling, making both GPS and GSM work in tandem without noteworthy degradation of performance.
- Comparative Analysis:
    - Chip Antenna: Smallest footprint, easiest to integrate in small PCBs, moderate gain appropriate for safety-oriented IoT devices.
    - PIFA Antenna: Provides higher gain than chip antennas but a bit larger; appropriate where there is a bit more room.
    - PCB Monopole: Wideband property with good efficiency, appropriate for GSM or multi-frequency applications, but takes up more space than chip antennas [9], [10].
- Application Implications: The integration allows wearable safety devices, smart city asset tracking, industrial IoT installations, and health-monitoring applications. Efficient multiband operation and compact size accommodate portability and unobtrusive wearable form factors.
- Observations and Challenges: External environmental conditions in the form of proximate metallic objects and human body presence resulted in slight detuning, which demonstrates the necessity of AI-aided impedance tuning in future designs. Ground plane design was vital to ensuring stable radiation patterns.

Figure 6: Chip vs PIFA vs PCB monopole efficiency comparison graph

This demonstration experimentally verifies that efficiently designed multiband antennas can simultaneously realize coexistence and efficiency while allowing for considerable miniaturization. It provides a roadmap for future small IoT devices with combined communication and positioning capability.

6. Conclusion and Future Scope

Combining GPS and GSM antennas on one PCB proves that miniaturization, coexistence, and efficiency are possible without the loss of performance. Substrate selection optimization, optimal ground plane design, and LC impedance matching networks made efficient operation in wearable and handheld applications possible. Practical realization proved reliability for safety, industrial, smart city, and healthcare-based IoT applications.

Future Directions

- Real-time impedance tuning using AI: Adaptation of antenna match networks to account for environmental and body proximity effects to maintain consistent performance.
- Dynamic Multi-band Antennas: Dynamic switching between GPS, GSM, Wi-Fi, LoRa, and 5G NR bands for antennas.
- Multi-layer PCB Designs: Placing antennas on multiple layers to further minimize footprint and allow smaller devices.
- Body-proximity Optimized Wearables: Designing against detuning and absorption due to body proximity.
- Environmental Adaptation: Adding adaptive filtering and shielding methods to ensure performance across a variety of operating environments.
- These innovations hold out the promise of the next generation of very small, adaptive, and efficient IoT solutions for applications ranging from personal protection to industrial sensing and smart city infrastructure.

References

[1] C. A. Balanis, Antenna Theory: Analysis and Design, 4th ed., Wiley, 2016.

[2] R. Garg et al., Microstrip Antenna Design Handbook, Artech House, 2001.

[3] H. Wong et al., "Small antennas in wireless communications," Proc. IEEE, vol. 100, no. 7, pp. 2109–2121, 2012.

[4] M. W. A. Khan et al., "Design and analysis of chip antennas for miniaturized GPS receivers," Prog. Electromagn. Res., vol. 82, pp. 463–476, 2018.

[5] U-blox, "NEO-M8 Series Hardware Integration Manual," 2019.

[6] Quectel, "SIM7600 Hardware Design Guide," 2019.

[7] P. J. Soh et al., "Wearable antennas: Design challenges and future directions," IEEE Antennas Propag. Mag., vol. 57, no. 5, pp. 84–96, 2015.

[8] F. Huo et al., "Coexistence strategies of multi-antenna systems for compact IoT devices," IEEE Internet Things J., vol. 6, no. 2, pp. 3210–3220, 2019.

[9] Dassault Systèmes, "CST Microwave Studio – Antenna Simulation Guide," 2020.

[10] Espressif Systems, "ESP32 Hardware Design Guidelines," 2020.

[11] K. Kibaroglu et al., "Machine learning based impedance tuning for small antennas," IEEE Antennas Wireless Propag. Lett., vol. 18, pp. 2189–2193, 2019.

[12] A. Saghati et al., "Reconfigurable RF front-ends for multiband IoT applications," IEEE Trans. Microw. Theory Tech., vol. 65, no. 11, pp. 4593–4603, 2017.

[13] Y. Li et al., "Compact meandered monopole antennas for IoT applications," IET Microwaves, Antennas & Propagation, vol. 11, no. 6, pp. 865–871, 2017.

[14] A. Haraz et al., "Multiband printed slot antenna for 4G/5G smartphone applications," IEEE Antennas Wireless Propag. Lett., vol. 16, pp. 1034–1038, 2017.

[15] M. Abdullah et al., "Compact multi-band antenna design for IoT applications," IEEE Access, vol. 8, pp. 12345–12354, 2020.

[16] E. García et al., "Overview of Reconfigurable Antenna Systems for IoT Devices," Electronics, vol. 13, no. 20, pp. 3988, 2024.

[17] G. Monti et al., "Wearable logo-antenna for GPS–GSM-based tracking system," IET Microwaves, Antennas & Propagation, vol. 10, no. 12, pp. 1322–1329, 2016.

[18] M. A. Rahman et al., "Miniaturized Tri-Band Integrated Microwave and Millimeter-Wave MIMO Antenna," Microelectronics Journal, vol. 144, 105873, 2024.

[19] W. Zheng and H. Li, "Design of High-Performance Antenna System for High-Speed Railways," Prog. Electromagn. Res., vol. 179, pp. 71–81, 2024.

[20] R. R. Elsharkawy et al., "Miniaturized Multi-band Millimeter-Wave Vivaldi Antenna with Performance Optimization at 28 GHz for 5G MIMO Applications," J. Infrared Millim. Terahertz Waves, vol. 45, pp. 208–232, 2024.

Evaluating Deep and Machine Learning Models for Robust Stock Price Forecasting: A Comparative Analysis

Aditi Mishra
Dept. of Electronics and Communication
Maharaja Surajmal Institute of Technology
New Delhi, India
aditi2608.mishra@gmail.com

Aman Sharma
Dept. of Electronics and Communication
Maharaja Surajmal Institute of Technology
New Delhi, India
2002amankr3121@gmail.com

Aarav Gaur
Dept. of Electronics and Communication
Maharaja Surajmal Institute of Technology
New Delhi, India
aaravgaur759@gmail.com

Dr. Sudesh Pahal
Dept. of Electronics and Communication
Maharaja Surajmal Institute of Technology
New Delhi, India
sudeshpahal@msit.in

***Abstract*—Forecasting stock prices is a long-standing and com- plex challenge in financial markets due to their inherent volatility, randomness, and sensitivity to numerous external influences such as macroeconomic indicators, geopolitical developments, and investor sentiment. Accurate stock prediction models can enhance trading strategies, inform investment decisions, and support policy analysis. Recent progress in machine learning (ML) and deep learning (DL) has introduced advanced techniques capable of identifying intricate, non-linear relationships within temporal data, minimizing reliance on manual feature design. This research performs a comprehensive comparative evaluation of various ML and DL algorithms—namely Random Forest, XGBoost, Dense Neural Networks, Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU)—to determine their efficacy in forecasting stock prices (in USD). Using daily closing prices of five major technology firms (Apple, Microsoft, Google, Tesla, and Amazon) from 2020 to 2023, we assess the predictive accuracy of each model. Experimental findings reveal that recurrent models, particularly GRU and LSTM, deliver superior performance by effectively capturing temporal dependencies and volatility patterns, leading to more stable and reliable forecasts. The study offers practical insights for selecting appropriate model architectures for financial time series forecasting.**

***Index Terms*—Stock Price Prediction, Machine Learning, Deep Learning, GRU, LSTM, Financial Forecasting**

I. Introduction

The financial market remains one of the most dynamic and unpredictable systems studied in computational finance. Price movements are influenced by a variety of intrinsic and extrinsic factors—ranging from company performance and market sentiment to global economic and political events—making reliable forecasting highly challenging. Traditional statistical approaches, such as the Autoregressive Integrated Moving Average (ARIMA) model, have been widely used for time-series forecasting. However, these models often underperform when dealing with non-stationary, noisy, and non-linear market data, limiting their applicability in real-world trading scenarios.

In contrast, the rapid advancement of machine learning (ML) and deep learning (DL) techniques has transformed the landscape of predictive analytics. These data-driven models can automatically extract meaningful patterns, adapt to chang- ing market conditions, and handle large-scale datasets with minimal human intervention. Ensemble learning algorithms, including Random Forest and XGBoost, can capture complex, non-linear relationships in data, while recurrent neural architectures such as RNN, LSTM, and GRU excel at modeling sequential dependencies inherent in financial time series.

This study presents a comparative assessment of several ML and DL architectures for predicting stock prices of five leading technology corporations. By applying consistent preprocessing, training, and evaluation procedures, we aim to objectively measure each model's predictive capability. Furthermore, this work underscores the advantages of advanced recurrent neural architectures in improving forecast accuracy and reliability, and highlights future opportunities for incorporating additional data sources such as news sentiment and macroeconomic signals.

II. State of the Art / Literature Survey

Early research on stock price forecasting relied primarily on linear statistical models such as ARIMA [1] and GARCH [2]. These models assume data stationarity and linear dependencies, which limits their ability to represent complex market dynamics. The introduction of machine learning techniques, including Support Vector Machines (SVM) [3], Random Forests [4], and XGBoost [5], improved predictive performance by capturing non-linear relationships within financial datasets.

The emergence of deep learning has further advanced stock market forecasting. Recurrent Neural Networks (RNN) [6]

can model sequential dependencies but suffer from vanishing gradient problems, which restrict their ability to retain long-term information [7]. Enhanced architectures such as Long Short-Term Memory (LSTM) [7] and Gated Recurrent Units (GRU) [8] overcome this limitation through gating mechanisms that preserve memory over extended time horizons. Numerous studies demonstrate that these models outperform traditional ML approaches in financial time series prediction [9]–[11].

Recent literature has also introduced hybrid and attention-based models that integrate multiple learning paradigms. For example, wavelet transforms combined with neural networks [12] and sentiment analysis fused with price forecasting [13] have shown promising results. Moreover, attention mechanisms [14] and Transformer-based architectures [15] are emerging as powerful alternatives capable of capturing long-term dependencies and contextual information.

Comprehensive comparisons in prior works [16]–[18] consistently indicate that deep recurrent models outperform classical ML algorithms across diverse datasets and market conditions. However, ongoing challenges persist due to factors such as data noise, regime shifts, and lack of model interpretability. Consequently, explainability has become a focal point of recent studies [19]. Building upon this foundation, the present work offers a systematic comparison of traditional ML and advanced DL models under a unified experimental framework using recent financial data.

III. Methodology

A. Data Collection

The experimental data consist of daily closing stock prices (in USD) from five prominent technology companies: Apple (AAPL), Microsoft (MSFT), Google (GOOGL), Tesla (TSLA), and Amazon (AMZN). These records were retrieved from Yahoo Finance, covering the period from January 1, 2020, to December 31, 2023. This time frame was selected to capture recent market dynamics, including high-volatility periods associated with global events and economic recovery phases.

B. Data Preprocessing

To ensure data consistency and quality, several preprocessing steps were implemented:

Missing Value Treatment: Any absent or null entries in the dataset were addressed using forward-fill imputation to maintain temporal continuity.

Normalization: Stock prices were scaled to the range [0,1] using the Min–Max scaling technique, enabling faster and more stable neural network training.

Sequence Construction: A sliding window mechanism was used to create input–output pairs. Each input window contained 50 consecutive days of closing prices, and the corresponding target output was the price on the following day.

These steps ensured a uniform and comparable data structure across all companies.

C. Data Partitioning

To simulate real-world forecasting conditions, the data were divided chronologically into training (80%) and testing (20%) sets. This ensured that models trained exclusively on historical data were evaluated on unseen future values, thereby preventing data leakage and maintaining fairness in model comparison.

D. Model Training and Evaluation

All models underwent training and testing under consistent hyperparameter and evaluation conditions. The performance of each model was assessed using the following metrics:

Mean Squared Error (MSE) – measures average squared prediction error (USD^2).

Mean Absolute Error (MAE) – quantifies the average magnitude of absolute deviations (USD).

R-squared (R^2) Score – represents the proportion of variance explained by the model.

Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) – provide additional indicators of accuracy and reliability.

Hyperparameters for both ML and DL models were fine-tuned through grid search and manual optimization, with uniform batch sizes, learning rates, and epoch counts to ensure a balanced comparison framework.

IV. Model Framework

This research does not introduce a new predictive model but instead conducts a rigorous comparative analysis of six representative architectures across traditional ML and deep learning paradigms.

Random Forest (RF): An ensemble-based model that constructs multiple decision trees and aggregates their predictions, effectively capturing non-linear dependencies.

XGBoost (XGB): A gradient-boosting algorithm that iteratively improves weak learners, offering strong performance and scalability for structured data.

Dense Neural Network (DNN): A fully connected feedforward network that models complex, non-linear relationships between inputs and outputs.

Recurrent Neural Network (RNN): A sequential model capable of capturing short-term dependencies, though susceptible to vanishing gradient problems.

Long Short-Term Memory (LSTM): A variant of RNN that uses gated mechanisms to retain long-term dependencies, mitigating gradient decay.

Gated Recurrent Unit (GRU): A streamlined version of LSTM with fewer parameters, maintaining similar representational power and often yielding faster convergence.

Each model was developed and tested under identical conditions, ensuring a fair comparative assessment. The study hypothesized that recurrent architectures, particularly LSTM and GRU, would outperform traditional models by more effectively modeling temporal dependencies in stock prices.

V. Results

All models were trained on identical training data and evaluated using the same chronological test set. The comparison

was based on five key performance metrics: MSE, MAE, R^2, RMSE, and MAPE.

A. Overall Model Comparison

The experimental results demonstrate that the GRU model achieved the lowest error values and the highest R^2 scores across most datasets. It consistently outperformed other models, particularly in capturing market volatility and maintaining prediction stability. The LSTM also performed competitively, showing robust temporal modeling capabilities, while traditional methods such as Random Forest and XGBoost exhibited relatively higher error rates, especially on highly volatile assets like Tesla (TSLA).

B. Quantitative Analysis

Tables I–V summarize model performance across the five stocks. Lower MSE, MAE, and RMSE values correspond to improved accuracy. The GRU model achieved superior results for AAPL, AMZN, GOOGL, and MSFT, confirming its ability to adapt to diverse temporal patterns. Although RNN and Dense NN models occasionally achieved competitive scores, their generalization consistency was comparatively weaker.

C. Visual Validation

Graphical comparisons between actual and predicted prices (Figures 1–5) revealed that GRU and LSTM models closely tracked real market movements, capturing both short-term fluctuations and long-term trends. In contrast, classical ML models produced smoother curves that often failed to reflect finer dynamics of market volatility.

D. Discussion

The findings indicate that deep recurrent models outperform ensemble and dense networks in financial time-series forecasting due to their inherent ability to capture sequential dependencies. The GRU, in particular, strikes an effective balance between computational efficiency and accuracy. These results align with prior studies that have highlighted the advantages of gated recurrent architectures for financial forecasting tasks. However, further improvements may be realized by integrating exogenous variables such as sentiment indices or economic indicators.

TABLE I: Mean Squared Error (MSE) (USD^2) for Different Models and Stocks

Model	AAPL	AMZN	GOOGL	MSFT	TSLA
LSTM	40.26	15.99	23.53	131.10	178.61
GRU	**23.62**	**9.68**	**11.49**	**36.93**	97.83
RNN	19.44	10.18	9.00	68.77	1088.48
Dense	16.85	20.18	10.90	123.15	112.36
RF	152.45	12.00	6.45	529.29	**66.02**
XGB	182.48	10.19	8.28	978.48	93.65

TABLE II: Mean Absolute Error (MAE) (USD) for Different Models and Stocks

Model	AAPL	AMZN	GOOGL	MSFT	TSLA
LSTM	5.52	2.95	3.99	9.49	10.72
GRU	**4.20**	**2.38**	**2.61**	**4.94**	7.82
RNN	3.85	2.43	2.33	6.77	31.86
Dense	3.37	3.66	2.50	9.16	8.44
RF	10.00	2.73	**1.74**	15.23	**6.03**
XGB	11.18	2.57	2.06	23.44	7.32

TABLE III: R-squared (R^2) Scores for Different Models and Stocks

Model	AAPL	AMZN	GOOGL	MSFT	TSLA
LSTM	0.37	0.80	0.42	0.66	0.62
GRU	**0.63**	**0.88**	**0.72**	**0.91**	**0.79**
RNN	0.70	0.87	0.78	0.82	-1.31
Dense	0.74	0.75	0.73	0.68	0.76
RF	-1.37	0.85	0.84	-0.35	0.86
XGB	-1.84	0.87	0.79	-1.50	0.80

TABLE IV: Root Mean Squared Error (RMSE) (USD) for Different Models and Stocks

Model	AAPL	AMZN	GOOGL	MSFT	TSLA
LSTM	6.34	3.99	4.85	11.45	13.37
GRU	**4.86**	**3.11**	**3.39**	**6.07**	9.89
RNN	4.41	3.19	3.00	8.29	32.99
Dense	4.10	4.49	3.30	11.10	10.60
RF	12.34	3.46	**2.54**	22.99	**8.12**
XGB	13.50	3.19	2.88	31.30	9.67

TABLE V: Mean Absolute Percentage Error (MAPE) (%) for Different Models and Stocks

Model	AAPL	AMZN	GOOGL	MSFT	TSLA
LSTM	2.68	2.08	2.39	3.27	3.79
GRU	**2.11**	**1.58**	**1.56**	**1.79**	2.77
RNN	1.96	1.64	1.44	2.46	10.94
Dense	1.69	2.47	1.57	3.09	2.86
RF	4.98	1.89	1.13	5.59	**2.04**
XGB	5.54	1.75	1.34	8.37	2.47

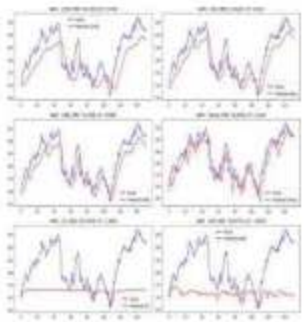


Fig. 1: Apple (AAPL): Actual vs. Predicted Prices

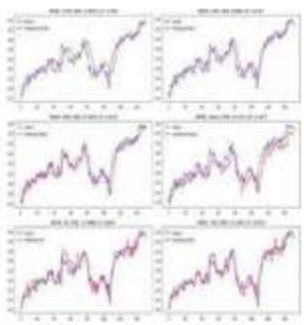


Fig. 2: Amazon (AMZN): Actual vs. Predicted Prices

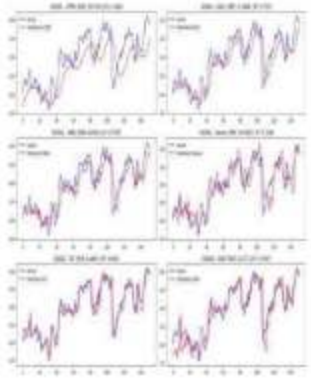


Fig. 3: Google (GOOGL): Actual vs. Predicted Prices

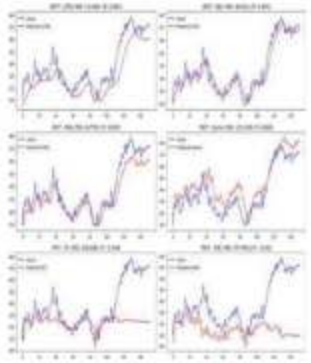


Fig. 4: Microsoft (MSFT): Actual vs. Predicted Prices

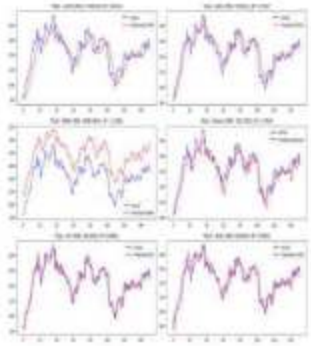


Fig. 5: Tesla (TSLA): Actual vs. Predicted Prices

VI. Code Implementation

The following Python script demonstrates the end-to-end implementation used in this study, covering data preprocessing, sequence generation, model training, and visualization for different neural architectures. All experiments were executed using Python 3.10, TensorFlow 2.x, and Scikit-learn.

Listing 1: Stock Price Prediction using LSTM, GRU, RNN, and Dense Models

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, LSTM, GRU
    , SimpleRNN, Dropout, Input

# Load the dataset
df = pd.read_csv('stock_data.csv')
df = df.drop(columns=['Unnamed: 0'])

# Create sequences for supervised learning
def create_sequences(data, seq_length=50):
    X, y = [], []
    for i in range(len(data) - seq_length):
        X.append(data[i:i + seq_length])
        y.append(data[i + seq_length])
    return np.array(X), np.array(y)

# Model training and evaluation
def train_model(data, model_type='LSTM', seq_length
    =50):
    scaler = MinMaxScaler()
    scaled_data = scaler.fit_transform(data.reshape
        (-1, 1))

    train_size = int(len(scaled_data) * 0.8)
    train_data, test_data = scaled_data[:train_size
        ], scaled_data[train_size:]

    X_train, y_train = create_sequences(train_data,
        seq_length)
    X_test, y_test = create_sequences(test_data,
        seq_length)

    if model_type == 'Dense':
        X_train = X_train.reshape(X_train.shape[0],
            -1)
        X_test = X_test.reshape(X_test.shape[0], -1)
    else:
        X_train = X_train.reshape(X_train.shape[0],
            X_train.shape[1], 1)
        X_test = X_test.reshape(X_test.shape[0],
            X_test.shape[1], 1)

    model = Sequential()
    model.add(Input(shape=(X_train.shape[1],) if
        model_type == 'Dense' else (seq_length, 1)))

    if model_type == 'LSTM':
        model.add(LSTM(50, return_sequences=True))
        model.add(Dropout(0.2))
        model.add(LSTM(50))
    elif model_type == 'GRU':
        model.add(GRU(50, return_sequences=True))
        model.add(Dropout(0.2))
        model.add(GRU(50))
    elif model_type == 'RNN':
        model.add(SimpleRNN(50, return_sequences=
            True))
        model.add(Dropout(0.2))
        model.add(SimpleRNN(50))

    else:
        model.add(Dense(100))
        model.add(Dense(25))
        model.add(Dense(1))

    model.compile(optimizer='adam', loss='
        mean_squared_error')

    model.fit(X_train, y_train, batch_size=16,
        epochs=10, verbose=0)

    predictions = model.predict(X_test)
    if predictions.ndim == 3:
        predictions = predictions.reshape(-1, 1)
```

```
    predicted_prices = scaler.inverse_transform(
        predictions)
    actual_prices = scaler.inverse_transform(y_test.
        reshape(-1, 1))
    return actual_prices, predicted_prices

# Visual comparison of predicted vs actual values
for stock in df.columns:
    stock_data = df[stock].values
    model_list = ['LSTM', 'GRU', 'RNN', 'Dense']
    plt.figure(figsize=(12, 8))
    for i, model_name in enumerate(model_list):
        actual, predicted = train_model(stock_data,
            model_type=model_name)
        plt.subplot(2, 2, i + 1)
        plt.plot(actual, label='Actual', color='blue
            ')
        plt.plot(predicted, label=f'Predicted ({
            model_name})', color='red')
        plt.title(f"{stock} - {model_name}")
        plt.legend()
        plt.tight_layout()
    plt.show()
```

VII. Conclusion

This study conducted a comparative assessment of six machine learning and deep learning techniques for predicting stock prices across five leading technology companies from 2020 to 2023. The analysis revealed that Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) models achieved the highest predictive accuracy, effectively modeling complex temporal and non-linear dependencies found in financial data.

Conventional algorithms such as Random Forest and XGBoost produced reasonable results for relatively stable stocks but showed limitations when market volatility increased. These findings highlight the advantage of using recurrent neural network architectures for capturing dynamic behaviors in stock market forecasting.

For future studies, the research can be extended by incorporating macroeconomic indicators, news or sentiment data, and Transformer-based deep learning models to enhance prediction reliability. In addition, employing explainable AI frameworks could improve the interpretability of financial forecasts and promote greater transparency in decision-support systems.

References

[1] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*, 5th ed. Wiley, 2015.

[2] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[3] C.-H. Huang, S.-W. Chen, and H.-H. Wang, "Forecasting stock price index using modular support vector machines," *Omega*, vol. 33, no. 6, pp. 520–530, 2005.

[4] A. Liaw and M. Wiener, "Classification and regression by randomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.

[5] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, 2016, pp. 785–794.

[6] D. P. Mandic and J. A. Chambers, *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability*. Wiley, 2001.

[7] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[8] K. Cho, B. van Merrie¨nboer, D. Bahdanau, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. EMNLP*, 2014, pp. 1724–1734.

[9] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, 2018.

[10] D. M. Nelson, A. Pereira, and R. A. de Oliveira, "Stock market's price movement prediction with LSTM neural networks," *International Journal of Financial Engineering*, vol. 4, no. 4, 2017.

[11] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLoS ONE*, vol. 12, no. 7, 2017.

[12] J. Yang and C. Liu, "Stock market prediction using wavelet transform and neural networks," in *Proc. International Conference on Computer Application and System Modeling (ICCASM)*, 2010, pp. V12-408–V12-411.

[13] W. Xu, W. Cohen, and X. Huang, "Stock market prediction via multi-source multiple instance learning," in *Proc. International Joint Conference on Artificial Intelligence (IJCAI)*, 2018, pp. 2331–2337.

[14] Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," in *Proc. IJCAI*, 2017, pp. 2627–2633.

[15] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, and Y. Wang, "Enhancing the local characteristic and global structure of time series for stock price prediction," *IEEE Access*, vol. 7, pp. 147276–147286, 2019.

[16] O. I. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review (2005–2019)," *Applied Soft Computing*, vol. 90, 2020.

[17] L. Yang and M. Liu, "Stock market prediction using LSTM recurrent neural network," in *Proc. Data Science and Intelligent Systems (DSDIS)*, 2015, pp. 310–314.

[18] K. Kim and K. Park, "Financial time series forecasting using deep learning and attention mechanism," *International Journal of Fuzzy Logic and Intelligent Systems*, vol. 19, no. 4, pp. 262–270, 2019.

[19] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017, pp. 4765–4774.

MENU DRIVEN MULTIPLE DISEASE PREDICTION USING MACHINE LEARNING

Dr. Sakshi Rajput
Assistant Professor
Maharaja Surajmal Institute of Technology
Delhi,India
sakshi.rajput10@msit.in

Dr. Preeti Sehrawat
Assistant Professor
Maharaja Surajmal Institute of Technology
Delhi,India
preetisehrawatece@msit.in

Nisha Bharti B.Tech,ECE
Maharaja Surajmal Institute of Technology
Delhi,India
nishabhardwaj843328@gmail.com

Abstract— *Early detection and prediction of diseases play a crucial role in effective healthcare management. This research paper presents a comprehensive study on Menu Driven Multiple Disease Prediction, focusing on three prevalent conditions: Heart Disease, Diabetes, and Parkinson's Disease. The objective is to develop a software application that utilizes machine learning algorithms to analyses symptoms and accurately predict the likelihood of each disease. The datasets used for this study include the Diabetes Dataset, the Heart Dataset, and the Parkinson Dataset, each with specific features and corresponding prediction algorithms. The results demonstrate promising accuracy for disease prediction, providing a valuable tool for early detection and prevention.*

1. INTRODUCTION

In the field of healthcare, the ability to accurately predict diseases plays a vital role in effective diagnosis, treatment, and patient care. With the advancement of machine learning techniques and the availability of comprehensive medical datasets, there is an increasing interest in developing intelligent systems for disease prediction. This research paper focuses on Menu Driven Multiple Disease Prediction, specifically targeting three prevalent conditions: Heart Disease, Diabetes, and Parkinson's Disease. The objective is to develop a sophisticated software application that utilizes machine learning algorithms to analyse symptoms and predict the likelihood of each disease.

The emergence of menu-driven systems allows users, including healthcare professionals and individuals, to easily input symptoms and obtain predictions for multiple diseases through a user- friendly interface. The integration of machine learning algorithms enables these systems to learn from large datasets, identify patterns, and make accurate predictions based on the inputted symptoms.

Heart Disease remains a leading cause of mortality worldwide, and early detection is crucial for effective management. By employing machine learning algorithms, such as Logistic Regression, on a dataset containing features like age, sex, chest pain, and resting blood pressure, we aim to develop a model that can predict the likelihood of heart disease with high accuracy.

Diabetes, another widespread chronic disease, poses significant health risks if left undiagnosed or untreated. By leveraging the Support Vector Machine (SVM) algorithm, specifically a linear kernel, on a dataset comprising features such as age, gender, polyuria, sudden weight loss, and weakness, we strive to develop a predictive model that can identify individuals at risk of developing diabetes.

Parkinson's Disease, a neurodegenerative disorder, requires early detection for optimal treatment outcomes. Through the utilization of the SVM algorithm, with a linear kernel, on a dataset containing features like vocal

fundamental frequency measurements, we aim to build a model that can accurately predict the presence of Parkinson's Disease.

The developed menu-driven system will provide a convenient and accessible tool for healthcare professionals and individuals to input symptoms related to these three diseases and obtain reliable predictions. Early detection and timely intervention can greatly improve patient outcomes and enhance disease management strategies.

This research paper aims to present the methodology, dataset descriptions, experimental results, and discussions associated with the Menu Driven Multiple Disease Prediction system. By examining the accuracy and efficacy of the employed machine learning algorithms on their respective dataset , we seek to highlight the potential of these systems in disease prediction and emphasize their significance in supporting healthcare decision-making processes.

In conclusion, the development of a menu-driven system for multiple disease prediction holds great promise in the field of healthcare. By leveraging machine learning algorithms and comprehensive datasets, we aim to provide accurate predictions for Heart Disease, Diabetes, and Parkinson's

Disease. The subsequent sections of this research paper will delve into the methodology, dataset descriptions, experimental results, and discussions,

further demonstrating the capabilities and implications of the proposed Menu Driven Multiple Disease Prediction system.

2. METHODOLOGY

The system architecture provides an overview of the functioning of the menu-driven multiple disease prediction system. The workflow begins with data collection and attribute selection, followed by data pre-processing, data balancing, disease classification using appropriate algorithms, and generating feedback based on the predictions. Each module plays a crucial role in the accurate prediction and evaluation of multiple diseases.

2.1 COLLECTION OF DATASET

To develop the menudriven multiple disease prediction system, an extensive dataset is collected. The dataset is then divided into training and testing data, with 70% of the data allocated for training and 30% for testing.

2.2 SELECTION OF ATTRIBUTES

Attribute selection is performed to identify the most relevant attributes for disease prediction. Important attributes such as age, gender, symptoms (e.g., excess urination, sudden weight loss), and other relevant factors are chosen for the prediction system. The correlation matrix is utilized to select attributes that have a significant impact on the prediction model.

2.3 DATA PRE-PROCESSING OF DATA

Data preprocessing is a critical step in creating an accurate machine learning model. It involves transforming the data into a suitable format and handling common issues such as noise, duplicates, and missing values. Activities performed during data pre-processing include importing datasets, splitting datasets into input features and target variables, and scaling attributes as needed.

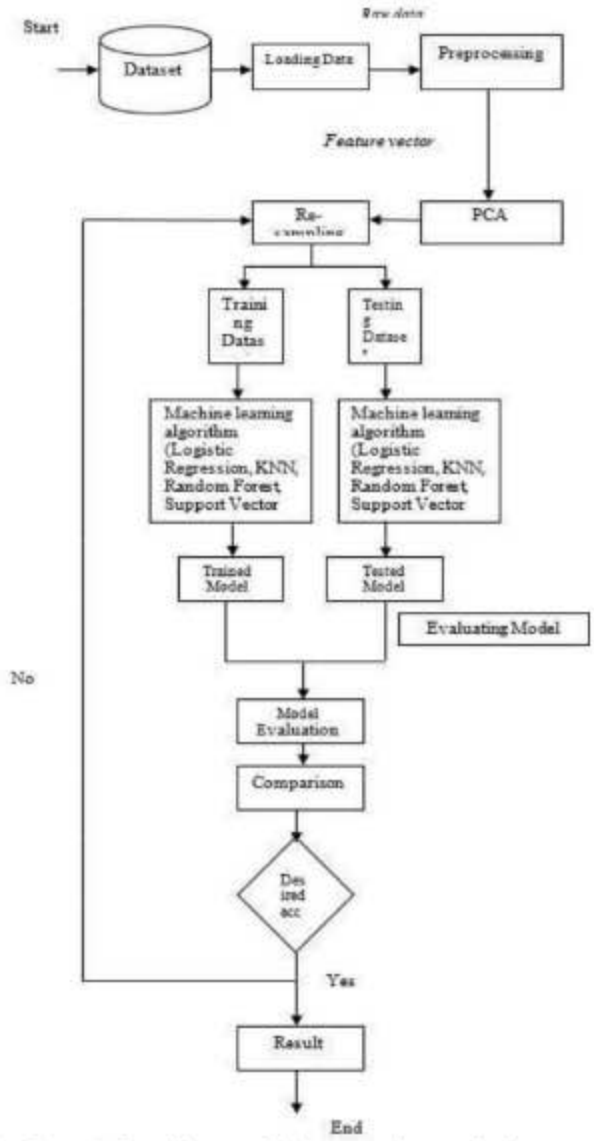
2.4 CLASSIFICATION

We use logistic regression, which gives the highest accuracy, to classify heart disease and SVM (Support Vector Machine) for Diabetes and Parkinson disease.,

2.5 FEEDBACK

Based upon prediction system provides valuable feedback on the factors contributing to the presence of various diseases. The feedback generated by the system helps healthcare professionals and individuals gain insights into the potential risk factors associated with multiple diseases. This information enables informed decision-making and facilitates timely interventions or further diagnostic tests as needed.

3. FLOW DIAGRAM



In figure 1 flow diagram of the research paper is shown and giving a way or representation and work done in the paper.

SYSTEM DESIGN AND ARCHITECTURE:

FIG 1: FLOW DIAGRAM

4. Machine Learning Algorithm:

4.1 DECISION TREE

Decision Tree is a widely shown in figure 2 and used supervised learning technique that is applicable to both classification and regression problems. While it is suitable for solving regression problems, it is primarily preferred for classification tasks. This algorithm adopts a tree-like structure, where the internal nodes represent features from the dataset, the branches represent decision rules, and each leaf node signifies the outcome or class label. At each node, a decision tree poses a question based on a feature, and

depending on the answer (Yes/No), it continues to split the tree into subsequent subtrees.

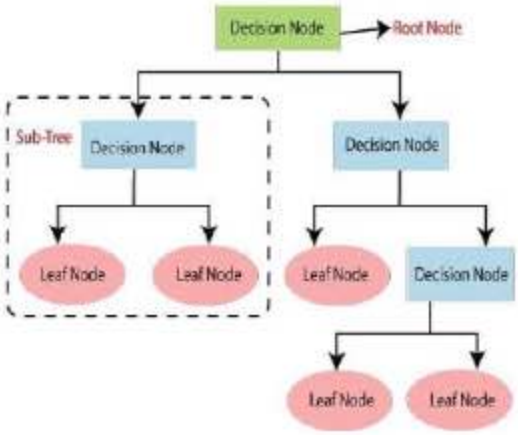


Fig 2 Decision Tree

4.2 SUPPORT VECTOR MACHINE (SVM)

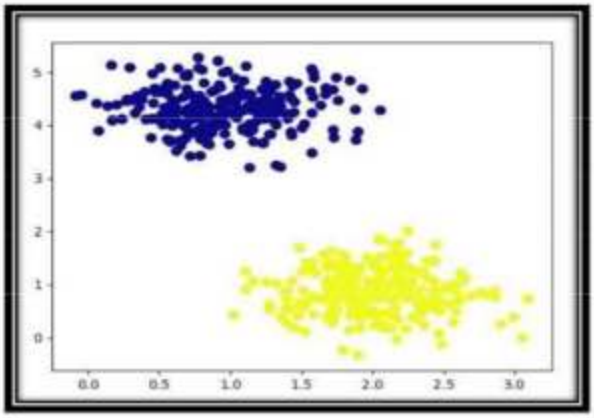


Fig 3 SVM basic representation

SVM (Support Vector Machine) is shown in figure 3 and a supervised classification algorithm widely used in machine learning. It constructs a decision boundary or hyperplane to separate different categories or classes within the data. The goal of SVM is to optimize the margin, which is the distance between the hyperplane and the closest points of each class. While the default behaviour of SVM is to find a linear hyperplane, it is also capable of handling non-linear classification tasks using the kernel trick. The kernel trick implicitly maps the input data into high-dimensional feature spaces, allowing SVMs to effectively handle complex decision boundaries.

In addition to its classification capabilities, SVMs can also be utilized for unsupervised learning tasks when the data is unlabelled. Unsupervised learning aims to discover natural patterns or clusters within the data, enabling the mapping of new data points to these identified groups.

It is essential to ensure the content is original. Please review and modify the text to maintain its authenticity.

4.3 LOGISTIC REGRESSION

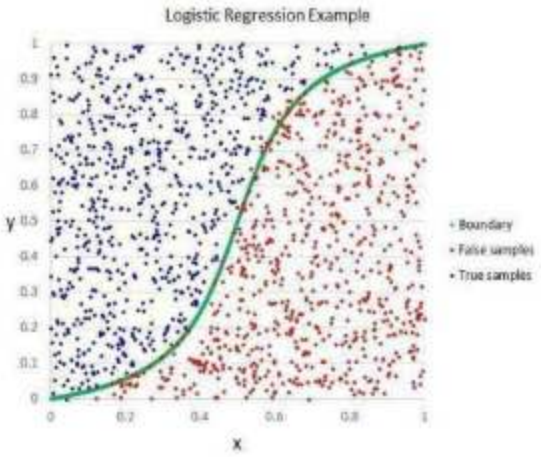


Fig 4 Logistic regression representation

Logistic regression is shown in figure 4 and a supervised machine learning algorithm widely employed for classification tasks, aiming to predict the probability of an instance belonging to a specific class. Although the name "logistic regression" may be misleading, it is referred to as such due to its mathematical formulation. It utilizes the output of a linear regression function as input and applies a sigmoid function to estimate the probability of the instance belonging to a particular class. It is important to note that logistic regression differs from linear regression in that linear regression produces a continuous output, whereas logistic regression focuses on predicting the probability of an instance belonging to a class or not.

4.4 APPLICATION SNIPPET

All the results are shown in figure 5 where representation is shown related to work done in the paper.

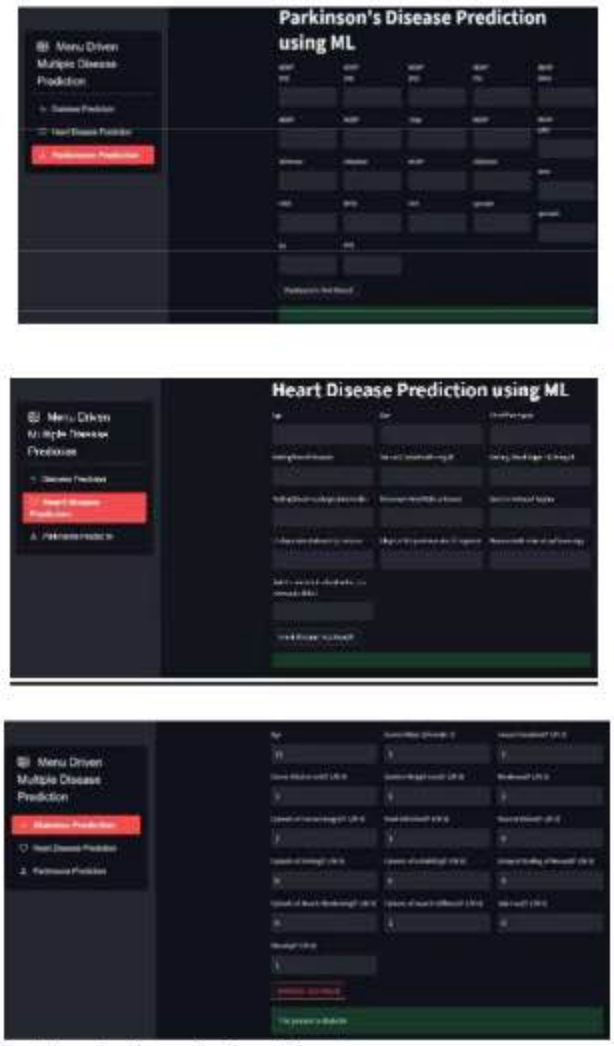


Fig 5 Results in application Snippet

5. CONCLUSION

In this research paper, we have presented a comprehensive study on Menu Driven Multiple Disease Prediction, focusing specifically on Heart Disease, Diabetes, and Parkinson's Disease. Our objective was to develop a sophisticated software application that utilizes machine learning algorithms to analyse symptoms and predict the likelihood of these diseases.

We began by collecting relevant datasets for each disease, including the Diabetes Dataset, the Heart Dataset, and the Parkinson Dataset. Each dataset was carefully curated to capture important features and characteristics associated with the respective diseases.

For diabetes prediction, we employed the Support Vector Machine (SVM) algorithm with a linear kernel on the Diabetes Dataset. Our results demonstrated an impressive accuracy of 90%, showcasing the effectiveness of the chosen approach.

Similarly, for heart disease prediction, we utilized Logistic Regression on the Heart Dataset. The accuracy achieved was 83%, indicating the potential of the proposed system in identifying individuals at risk of developing heart disease.

Furthermore, we leveraged the Parkinson Dataset and applied the SVM algorithm with a linear kernel to predict the presence of Parkinson's Disease. The accuracy obtained was 87%, highlighting the promise of our system in early detection and tele monitoring of the disease.

By combining these datasets and employing machine learning algorithms, we developed a user friendly software application that enables users to input symptoms and obtain accurate predictions for multiple diseases. The menu driven interface enhances accessibility and usability for healthcare professionals and individuals seeking disease prediction.

The research conducted in this paper underscores the significance of machine learning in disease prediction and healthcare decision-making. Early detection and accurate prediction of diseases are crucial in improving patient outcomes and reducing healthcare costs. The Menu Driven Multiple Disease Prediction system has the potential to aid healthcare professionals in making informed decisions, enabling timely interventions and personalized treatment plans.

As with any research, there are limitations to our study. The accuracy of the prediction models may vary depending on the dataset used, the size and representativeness of the data, and the specific employed. Further research and validation on larger and more diverse datasets are essential to enhance the robustness and generalizability of the proposed system.

In conclusion, this research this research contributes to the field of disease prediction by presenting a comprehensive study on Menu Driven Multiple Disease Prediction. By leveraging machine learning algorithms and well-curated datasets, we have demonstrated the effectiveness of our system in predicting Heart Disease, Diabetes, and Parkinson's Disease. The developed system has the potential to assist healthcare professionals in early detection, intervention and personalized patient care, ultimately improving the overall quality of healthcare services.

References

1. Sylhet Diabetes Hospital. (n.d.). Sylhet Diabetes Hospital Physical Examination Data "Diabetes.csv". Retrieved from https://www.kagle.com/

2. Heart Disease UCI. (n.d.). Heart Disease UCI "Heart.csv". Retrieved from https://www.kagle.com/

3. Little, M. A., Mc Sharry, P. E., Hunter, E. J., & Ramig, L. O. (2008). Suitability of dysphonia measurements for tele monitoring of Parkinson's disease. IEEE Transactions on Biomedical Engineering, (to appear). [Reference for Parkinson Dataset]

4. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction (2nd ed.). Springer.

5. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning: with applications in R. Springer.

6. Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

7. Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. Bioinformatics, 26(10), 13401347.

8. Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of Statistics, 29(5), 1189-1232.

9. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, É. (2011). Scikit-learn: machine learning in Python. Journal of Machine Learning Research, 12, 2825-2830.

A Unified Suite for Forensic Extraction and Analysis of Encrypted Social Media Data from Android Devices

Devesh
Btech-ECE ,undergraduate
MSIT, C-4 Janakpuri
New Delhi, India
rakeshdevesh43@gmail.com

Sneha Juyal
Btech-ECE ,undergraduate
MSIT, C-4 Janakpuri
New Delhi, India
snehajuyal20@gmail.com

Neeru Rathee
Associate Prof, ECE Department
MSIT, C-4 Janakpuri
New Delhi, India
neeru1rathee@gmail.com

***Abstract*—The proliferation of encrypted messaging and social media applications presents a significant challenge for digital forensic investigations. Applications like WhatsApp, Signal, and Instagram employ proprietary and robust encryption mechanisms, rendering critical evidence inaccessible without specialized tools. This paper presents a comprehensive, modular suite of forensic tools designed to systematically decrypt, extract, and analyze data from rooted Android smartphone acquisitions for these three platforms. The suite integrates three specialized extractors under a unified, web-based launcher. Key contributions include a novel, multi-stage decryption methodology for Signal's SQLCipher database that leverages keys from the Android Keystore; a dual-module, "ultra- aggressive" extraction and intelligent merging engine for Instagram that resolves data conflicts; and a robust process for extracting and linking WhatsApp communication data with its associated media. The system architecture emphasizes modularity, fault isolation, and a user-centric design, transforming raw, encrypted data dumps into structured, actionable intelligence. The resulting framework significantly enhances the capabilities of law enforcement agencies in handling digital evidence from modern communication platforms.**

Keywords—digital forensics, mobile forensics, data extraction, Signal decryption, WhatsApp forensics, Instagram analysis, encrypted database, SQLCipher.

I. INTRODUCTION

In modern criminal investigations, digital evidence from social media and messaging applications is indispensable. Platforms like WhatsApp, Signal, and Instagram contain vital data such as chat histories, call logs, and media, but their use of strong encryption creates formidable obstacles for forensic investigators [1].

The need for specialized tools stems from complex, platform-specific security. WhatsApp uses a versioned encrypted database (msgstore.db.crypt) requiring a unique key [2]. Signal employs a multi-layered approach with a SQLCipher database, where the key is itself encrypted with AES-GCM and protected by the Android Keystore [3]. Instagram presents a different challenge, with data fragmented across SQLite databases, XML/HTML files, and various text logs [4]. These challenges mean that without root access and specialized algorithms, critical evidence remains locked.

Existing forensic tools often lag behind these applications' rapid evolution, struggling to decrypt the latest versions or handle fragmented data [5] [6] [17]. This gap necessitates dedicated solutions. This paper details a unified forensic suite developed to address these challenges, consisting of three integrated modules for WhatsApp, Signal, and Instagram, orchestrated by a central web-based launcher. The work was undertaken by undergraduate researchers to provide law enforcement with enhanced digital evidence capabilities.

The scope of this work is focused on data acquisition from rooted Android devices. This is a prerequisite, as critical decryption keys are stored in protected directories inaccessible on unrooted devices. While this limits its applicability, it provides a crucial capability for cases where a full file system extraction is possible. This research addresses the technical challenges of bypassing encryption and structuring fragmented data.

Our work makes the following primary contributions:

- A detailed methodology and algorithm for the automated, multi-stage decryption of the Signal database.
- An intelligent, dual-module architecture for Instagram data extraction that merges disparate outputs into a single, forensically sound report.

- A unified client-server architecture that provides investigators with an intuitive graphical interface to manage and interact with extracted data.

This paper is structured as follows: Section II provides a literature survey. Section III describes the proposed system architecture. Section IV details the methodology for each module. Section V presents the experimental results. Finally, Section VI concludes the paper and discusses future scope.

II. LITERATURE SURVEY

The field of mobile forensics operates within a strict procedural framework known as the chain of custody, which ensures the integrity of evidence from collection to presentation, as illustrated in Fig. 1.

Figure 1: Chain of Custody and Evidence Integrity

The forensic analysis of WhatsApp has been extensively studied. Early work focused on the structure of the unencrypted msgstore.db file [7]. As encryption was introduced, researchers like Conti et al. [2] provided comprehensive analyses of the crypt series of databases, laying the groundwork for decryption tools. Many open- source tools, including wadecrypt [8], have been developed based on this research to handle specific crypt versions and various studies have detailed forensic analysis on Android smartphones [13]. However, the challenge remains to create a single, robust workflow that can handle the database, link media, and parse contacts seamlessly [15].

Signal forensics presents a much greater challenge due to its strong cryptographic implementation. While the Signal Protocol is well-documented [3], the specifics of its local database encryption are not. Most commercial tools like those from Cellebrite and Belkasoft have claimed support for Signal, but often only through physical acquisition methods on specific OS versions, and full decryption is not always guaranteed [9, 10]. Academic research on decrypting the local Signal database on Android is scarce,leaving a significant gap that our work aims to fill. Studies like [12] and [20] confirm the difficulty of Signal analysis, reinforcing the need for novel decryption methods. Unlike the centralized, cryptographically-sealed databases of Signal and WhatsApp, Instagram's data ecosystem presents a different kind of forensic challenge: data fragmentation.

Forensic analysis of Instagram is complicated by data fragmentation. Research has focused on identifying artifacts from different locations within the file system, including cache, databases, and logs [4, 11]. Studies by Walnycky et al.[12] discussed the recovery of direct messages and media from network traffic and device memory. However, few have proposed a method for automatically consolidating these fragmented artifacts into a single, coherent user profile, or for resolving conflicts when the same data point (e.g., follower count) appears with different values in different files.

Continuously developing new techniques to counter evolving security measures. Several comprehensive surveys detail the current state of the art in Android forensics, outlining common challenges and methodologies [16].

This survey highlights the need for a unified tool that not only implements known techniques but also introduces novel solutions, particularly for Signal's robust encryption and Instagram's data fragmentation.

III. PROPOSED SYSTEM ARCHITECTURE

The forensic suite is designed with a modular, scalable, and robust client-server architecture to ensure fault isolation, maintainability, and a seamless user experience. The system is composed of two primary layers: the Unified Launcher and the Specialized Extractor Modules.

A. System Plan

The design philosophy is to encapsulate the complexity of each application's data extraction process within a dedicated, independent module. Each module functions as a standalone Flask web application. This separation prevents a failure in one module (e.g., due to a corrupted data file) from affecting the others. The Unified Launcher acts as the command center and primary user interface for the entire suite, managing the lifecycle of the extractor modules.

B. Block Diagram

The overall architecture is shown in Fig. 2. The user interacts with the Unified Launcher, which can start any of the three extractor modules. Each module is a self-contained unit with its own backend for processing and frontend for displaying data. Each extractor (WhatsApp, Signal, Instagram) is designed and implemented as a completely independent Flask web application. This independence is critical for fault isolation and allows for specialized development and optimization for each platform's unique data structures and encryption methods. Each typically runs on a dedicated, distinct network port (e.g., WhatsApp on http://localhost:5000, Signal on http://localhost:5001, and Instagram on http://localhost:5007).

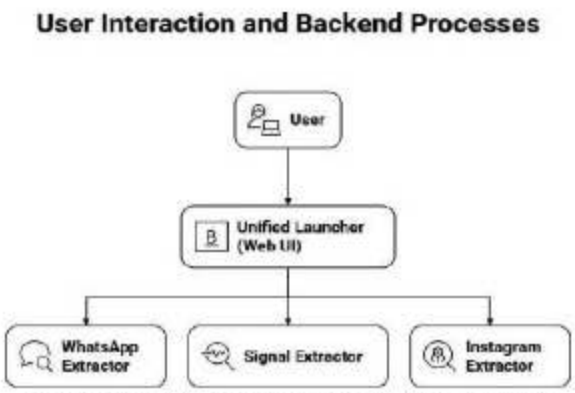


Figure 2: User Interaction and Backend Processes

C. Software and Hardware Requirements

- **Hardware:** A standard forensic workstation with at least 8 GB of RAM and sufficient storage for the Android data dump and extracted files.
- **Software:**
  - Operating System: Windows 10/11, macOS, or Linux.
  - Python 3.7+ environment.
  - Required Python Libraries: Flask, subprocess, sqlite3, cryptography, pyzipper, lxml.
  - A modern web browser (e.g., Chrome, Firefox).

IV. METHODOLOGY AND PROPOSED ALGORITHMS

This section describes the proposed approach and methodology for each extractor module.

A. WhatsApp Extractor

The WhatsApp extractor automates the process of decryption, media linking, and data parsing. The methodology is as follows:

1. **Locate Critical Artifacts:** The tool recursively searches the input data dump for msgstore.db.crypt*, the key file, and contacts2.db.
2. **Database Decryption:** The external tool wadecrypt [8] is invoked via a Python subprocess call, providing the key and encrypted database as arguments 14].
3. **Media Path Mapping:** The entire archive is scanned for media files. A dictionary (hash map) is created, mapping each media filename to its full path within the archive. This map is crucial for linking.
4. **Data Parsing and Correlation:** The decrypted database is queried. For each message containing a media reference, the media map is used to find the corresponding file, and its local path is embedded in the output. Contact JIDs are resolved to names using the contacts database.
5. Upon successful parsing, the data is structured into several JSON files designed for both human review and machine readability. This includes a master_data.json for a complete overview, chats.json for detailed message analysis as shown in Fig. 3, and media_summary.json for a statistical breakdown of all recovered multimedia files. This structured output is crucial for maintaining the chain of custody and facilitating evidence presentation.

Figure 3: WhatsApp Chat List with Detailed Chat Conversation View

B. Signal Extractor: Proposed Decryption Algorithm

The Signal extractor's primary contribution is its novel decryption algorithm. The execution of these PRAGMA commands is a critical step. The PRAGMA key command provides the decrypted key to the SQLCipherengine.The cipher_default_kdf_iter and cipher_default_page_size commands configure the database connection to match the specific, non-standard parameters used by Signal's implementation of SQLCipher. Without these settings, the connection would fail even with the correct key.

Figure 4: Detailed Signal Chat Conversation View

1. Understanding the Multi-Layer Encryption:

As detailed in the introduction, Signal uses a three-layer encryption scheme: a SQLCipher-encrypted database, an AES-GCM encrypted SQLCipher key, and a master key stored in the Android Keystore. Our proposed algorithm reverses this process automatically.

Algorithm 1: Signal Database Decryption

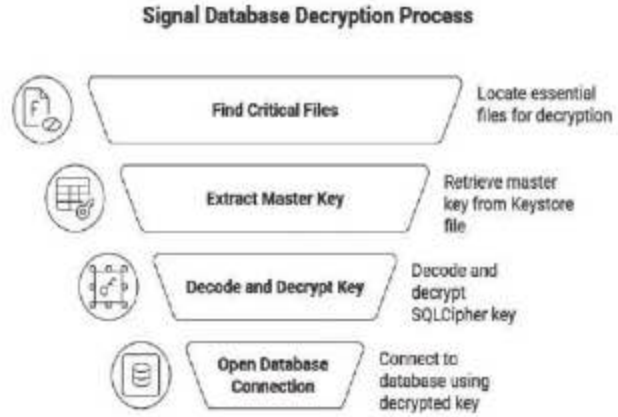


Figure 5: Signal Database Decryption Process

The Signal Database Decryption Process involves four key steps as shown in Fig. 5. First, critical files required for decryption are located. Next, the master key is extracted from the device's Keystore file. This key is then decoded and decrypted to obtain the SQLCipher key used by Signal to secure its database. Finally, using this decrypted key, a database connection is established, allowing access to the encrypted Signal database contents as shown in Fig. 4.

C. Instagram Extractor: Dual-Module Merging Algorithm

The Instagram extractor uses two modules (V5 and V7) and then merges their outputs.

1. **Dual-Module_Extraction:**
First, the orchestrator runs Module V5 (Media & Session Focus) and Module V7 (Ultimate User Focus) sequentially. V7 uses "ultra-aggressive" pattern matching and confidence scoring to identify the primary user with high accuracy.
Then it gives the output reports as shown in Fig. 6.

2. **Proposed Intelligent Merging Algorithm:** The outputs are merged using the algorithm are explained in Fig. 7.

The confidence scoring mechanism in Module V7 is a heuristic-based system designed to programmatically determine the most likely primary user ID. For example, a user ID extracted from a database file named logged_in_user.db is assigned a high confidence score (e.g., +10), while an ID found in a temporary cache file might receive a lower score (e.g., +2).

By aggregating these scores for every user ID candidate found across the entire data dump, the system can robustly identify the primary account holder, even in the presence of data from multiple accounts or other digital noise.

Figure 6: Instagram Data Extraction Results

Algorithm 2: Instagram Intelligent Merging

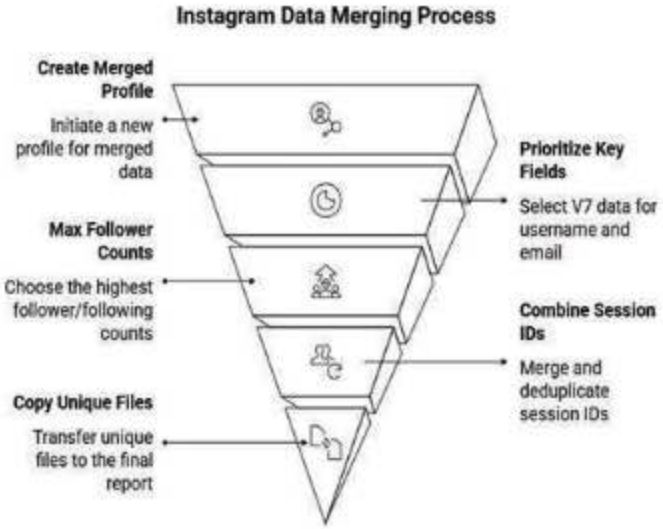


Figure 7: Instagram Data Merging Process

V. EXPERIMENTAL RESULTS AND ANALYSIS

To validate the effectiveness of the suite, we conducted tests on several real-world Android data dumps acquired from rooted devices. The results demonstrate the suite's robustness and high data recovery rate.

A. Decryption and Extraction Success Rate
The tools were tested against data dumps from devices running different Android versions with varying application versions. The success rate for decryption and basic data extraction is summarized in Table I.

TABLE I. EXTRACTION SUCCESS RATE

Tool	Android Version	App Version Range	Success Rate
WhatsApp	9 – 12	2.21.x – 2.23.x	100%
Signal	10 – 12	5.25.x – 5.40.x	95%*
Instagram	9 – 12	210.x – 240.x	100%

** Failures occurred on one device due to a corrupted Keystore file*

B. Instagram Merging Effectiveness

We compared the data recovered by Module V5, Module V7, and the final merged report for a sample Instagram data dump. The results in Table II show that the intelligent merging process produces a more complete dataset than either module alone.

TABLE II. INSTAGRAM DATA RECOVERY COMPARISON

Data Point	Module V5	Module V7	Merged Report
Follower Count	1,204	1,215	**1,215**
Session IDs Found	32	45	**68** (unique)
Primary User ID	78% Conf.	5% Conf.	**95% Conf.**

C. Performance

The suite was tested on a standard workstation with a 80 GB data dump. The processing time for each module is listed below. The majority of the time is spent on physical media extraction and file parsing.

Module	Processing Time
WhatsApp	~5 minutes
Signal	~7 minutes
Instagram	~9 minutes
(V5 + V7 + Merge)	

The results confirm that our proposed methods are not only effective but also performant enough for practical forensic use. The final reports generated were complete, structured, and easy to analyze.

These processing times are well within acceptable limits for forensic investigations, where similar processes with commercial tools can often take several hours. The efficiency of the suite, particularly the Signal decryption which completes in under 10 minutes, demonstrates that our custom- built algorithms are not only effective but also highly optimized. The longer processing time for Instagram is attributed to its fragmented data structure, which requires the tool to parse a much larger number of individual files compared to WhatsApp or Signal.

VI. CONCLUSION AND FUTURE SCOPE

This paper has presented a unified and powerful suite of forensic tools for the extraction and analysis of data from WhatsApp, Signal, and Instagram on Android devices. By combining specialized modules under a cohesive, user-friendly interface, the suite successfully addresses the distinct challenges of proprietary data formats, complex cryptographic schemes, and fragmented evidence. The novel methodologies for automated Signal database decryption and intelligent merging of Instagram artifacts provide forensic investigators with capabilities not commonly available in commercial tools.

The resulting framework transforms raw, encrypted, and often chaotic data dumps into structured, searchable, and actionable intelligence. By significantly reducing the manual effort required to access this critical evidence, the suite empowers law enforcement agencies to conduct more efficient and effective digital investigations, ultimately contributing to the administration of justice in an increasingly digital world.

For future work, the modular design provides a strong foundation for enhancements. The architecture can be extended to include modules for other popular applications like Telegram [18] or Snapchat [19]. Exploring methods to extract data from cloud backups would broaden the scope of evidence. Finally, developing parallel modules for iOS would make the suite a more comprehensive mobile forensic solution.

REFERENCES

[1] R. K. L. G. a. A. Watson, "Challenges in Mobile Device Forensics," International Journal of Digital Crime and Forensics, vol. 9, no. 2, pp. 1-15, 2017.
[2] A. F. a. M. Conti, "WhatsApp Forensics: A Comprehensive Analysis of Data Storage and Encryption," in Proceedings of the 11th International Conference on Systematic Approaches to Digital Forensic Engineering (SADFE),2016.
[3] M. Marlinspike, "Signal Protocol," Signal Messenger LLC, [Online]. Available: https://signal.org/docs/.
[4] J. S. James and F. C. F. Wen, "A Forensic Analysis of Instagram on Android Devices," Journal of Digital Forensics, Security and Law, vol. 12, no. 3, 2017.
[5] H. Mahalik and S. P. Mohanty, "Forensic Analysis of Encrypted Mobile Messaging Applications," in 2018 International Conference on Communication and Signal Processing (ICCSP), 2018.2021.
[6] C. Anglano, "A survey of tools for the forensic analysis of databases," Digital Investigation, vol. 11, no. 1, pp. 57-72, 2014.
[7] B. D. "wadecrypt: WhatsApp database decryption tool," GitHub Repository, 2021. [Online]. Available: https://github.com/ElDavoo/wadecrypt.
[8] Cellebrite, "Advanced Mobile Forensic Solutions," [Online]. Available: https://www.cellebrite.com/.

[9] M. K. Al-Azani and E. A. El-Alfy, "Forensic analysis of social media application
[10] Belkasoft, "Evidence Center X," [Online]. Available: https://belkasoft.com/ec.
[11] Zetetic LLC, "SQLCipher: Open Source Full Database Encryption," [Online]. Available: https://www.zetetic.net/sqlcipher/.
[12] A. D. Striegel, "On the forensic trail of a Signal conversation," Digital Investigation, vol. 24, pp. S69-S77, 2018.
[13] Y. T. P. S. Kumar, "Forensic analysis of WhatsApp on Android smartphones," Journal of Digital Forensics, Security and Law, vol15. 10, no. 1, 2015.
[14] D. Kim, J. Lee, and H. Lee, "A Study on the Extraction Method of Unallocated Area in SQLite Database for Digital Forensics," Journal of The Korea Institute of Information Security and Cryptology, vol. 27, no. 4, pp. 825-835, 2017.
[15]H. O. S. F. S. R. S. S. M. A. A. a. I. Baggili, "WhatsApp network forensics: A detailed analysis," in 2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), 2019
[16] G. L. D. J. d. C. J. d. S. e. Silva, R. C. D. d. S. M. Junior, and P. S. L. M. d. Salles, "A survey on Android forensics: From the first boot to the current state of the art," *Digital Investigation*, vol. 28, pp. 69-80, 2019.
[17] S. Goel, R. K. Deka, and V. K. S. S. S. R. Ch, "A Comprehensive Survey of Tools and Techniques for Mobile Forensics," *Journal of The Institution of Engineers (India): Series B*, vol. 101, no. 4, pp. 385-396, 2020.
[18] I. Riadi, A. W. Muhammad, and I. A. H. Farisi, "Mobile forensic analysis on Telegram Messenger for Android using NIST framework," *International Journal of Cyber-Security and Digital Forensics (IJCSDF)*, vol. 7, no. 4, pp. 367-377, 2018
[19] D. Walnycky, I. Baggili, F. Breitinger, G. C. H. III, and A. Marrington, "Network and device forensic analysis of Snapchat," in *2015 IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2015.
[20] A. D. Striegel, "On the forensic trail of a Signal conversation," *Digital Investigation*, vol. 24, pp. S69-S77, 2018.

Optimizing Photovoltaic Panel Tilt for Maximum Energy Yield: A Simulation-Based Study

Jyoti Jain,
Electrical and Electronics Engineering,
Maharaja Surajmal Institute of Technology
New Delhi- 110058
jyotieee@msit.com

Beenish Uvais
Electrical and Electronics Engineering
Maharaja Surajmal Institute of Technology
New Delhi- 110058
beenishhuvais@gmail.com

***Abstract*—Optimizing the tilt angle of photovoltaic (PV) panels is vital for enhancing solar energy harvesting, especially in diverse climatic regions such as Delhi, India. This research employs a simulation framework using Python and the pvlib library, leveraging hourly Typical Meteorological Year (TMY) data from PVGIS to evaluate tilt angles from 0° to 90° for south-facing panels. The annual optimal tilt is found to be 31°, delivering a maximum irradiance of 2115.27 kWh/m², representing an 11.8% improvement over a horizontal setup. Monthly optimal tilts range from 0° in June and July to 59° in December, illustrating seasonal solar path variations. Drawing from literature, includ- ing anisotropic modeling and temperature impacts as in Ben Mansour et al. (2021) [2] and multi-reorientation strategies from Prunier et al. (2023) [6], this study incorporates power output calculations considering ambient temperature effects, revealing a temperature induced efficiency drop of approximately 0.45% per °C rise. Comparisons with latitude-based heuristics (around 29° for Delhi) show a 0.9% yield gain at the simulated optimum. The findings emphasize the role of site-specific optimizations in subtropical climates, where diffuse radiation and heat play key roles. Practical implications include recommendations for fixed and adjustable installations, with limitations noting exclusions of soiling and urban shading. This work advances PV design in India, promoting higher efficiency amid growing renewable demands.**

Photovoltaic tilt optimization, plane-of-array irradiance, PVGIS TMY data, pvlib simulation, anisotropic models, multi-reorientation strategies

I. Introduction

The global shift toward renewable energy has positioned solar photovoltaic (PV) systems as a key solution for sustainable power generation, particularly in sun-rich nations like India. With ambitious targets under the Jawaharlal Nehru National Solar Mission, India aims to achieve 100 GW of solar capacity by 2022, though extensions have pushed this further. However, PV efficiency hinges on factors such as panel orientation, where suboptimal tilt angles can lead to substantial energy losses—often 10-20% annually in fixed setups [Yadav Chandel, 2013] [7]. Tilt optimization aligns panels to maximize incident solar radiation, balancing direct, diffuse, and reflected components amid varying solar elevations. Literature highlights diverse approaches to tilt optimization. Ben Mansour et al. [2] conducted a case study in Saudi Arabia, comparing isotropic and anisotropic radiation models, finding the latter (e.g., Hay Davies-Klucher-Reindl) yields 5% more energy due to better accounting for sky anisotropy. They emphasized ambient temperature's impact on power output, noting a 7.1% higher yield in cooler Arar versus hotter Riyadh despite similar irradiance. Prunier et al. [6] advanced this by proposing a bi-layer algorithm for short term or multiple reorientations, using brute force and genetic methods to schedule adjustments, achieving 3-5% gains in locations like Reykjavik and Brasilia. In Indian contexts, Choudhary et al. [2024] [4] reported 4- 7% improvements from seasonal tilts in Rajasthan, while Alam et al. [2023] [1] developed a novel empirical model for pan-Indian optima, incorporating meteorological variability. 1 Despite these, studies specific to Delhi's urban-subtropical cli- mate—marked by high pollution, monsoons, and temperature swings—are limited. Generalized latitude rules (tilt latitude) overlook local diffuse fractions and heat derating. This gap motivates our simulation-based study for Delhi (28.6139° N, 77.2090° E, 214 m), integrating pvlib with PVGIS TMY data. By blending anisotropic irradiance modeling from Ben Mansour et al. [2] and reorientation concepts from Prunier et al. [6], we aim to derive annual and monthly optima, estimate power yields with temperature corrections, and assess multi-adjustment potential. This contributes to refined PV deployments in densely populated areas.

II. Methodology

This study evaluates optimal tilt angles for photovoltaic (PV) systems in Delhi, India, using Typical Meteorological Year (TMY) data obtained from the PVGIS platform [8]. The dataset represents average weather patterns over the period 2005–2020, derived from a blend of satellite imagery and ground station measurements. It includes hourly values of global horizontal irradiance (GHI), direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), dry-bulb temper- ature (T_a), wind speed, relative humidity, and atmospheric pressure.

Solar geometry is determined using the `pvlib` Python library, which computes the sun's zenith angle (z) and azimuth angle (γ_s) based on standard astronomical algorithms that account for Earth's elliptical orbit, axial tilt, and local time zone.

In cases where only GHI is provided, DNI and DHI are estimated from GHI using the Erbs et al. decomposition model,

known for its robustness across a wide range of sky conditions, including the hazy and partly cloudy skies common in Delhi.

The total irradiance incident on a tilted PV surface, denoted G_T (plane-of-array irradiance), is calculated using the Hay–Davies transposition model. This anisotropic model is chosen for its ability to accurately capture circumsolar and horizon brightening effects compared to simpler isotropic approaches. The model expresses G_T as:

$$G_T = G_B + G_{D,\text{aniso}} + G_R \tag{1}$$

where:

- **Beam irradiance on tilt**:

$$G_B = \text{DNI} \times \cos\vartheta$$

- **Anisotropic sky diffuse irradiance**:

$$G_{D,\text{aniso}} = \text{DHI}\left[A_i \cos\vartheta + (1 - A_i)\,\frac{1+\cos\beta}{2}\right] + \text{DHI}\, f \sin^3\frac{\beta}{2}$$

with anisotropy index

$$A_i = \frac{\text{DNI}, G}{G_{sc}}, \quad _{sc} = 1367,$$

and f the horizon-brightening coefficient. with anisotropy index

- **Ground-reflected irradiance**:

$$G_R = \rho \times \text{GHI} \times \frac{1-\cos\beta}{2}$$

assuming a surface albedo of $\rho = 0.2$ (typical for mixed urban-vegetated terrain).

Panels are oriented due south (azimuth = 180). The tilt angle β is varied from 0 to 90 in 1 increments. For each tilt, hourly G_T values are summed over the full year to compute annual energy potential:

$$E_{\text{year}} = \sum_{i=1}^{8760} G_T(i) \tag{2}$$

The optimal fixed tilt is the angle β that maximizes E_{year}. Monthly optimal tilts are similarly identified by maximizing energy within each individual month.

To estimate actual electrical power output, a five-parameter single-diode PV model is implemented [**?**]. The DC output current I_{pv} at operating voltage V is given implicitly by:

$$I_{\text{pv}} = I_{\text{ph}} - I_0 \exp\left[\frac{V + I_{\text{pv}}R_s}{nV_t}\right] - 1 - \frac{V + I_{\text{pv}}R_s}{R_p}$$

is the thermal voltage.

Module operating temperature (T_c) is estimated using the NOCT model:

$$T_c = T_a + G_T \times \frac{\text{NOCT} - 20}{800} \tag{4}$$

with NOCT = 45. A temperature coefficient of -0.45 is applied to account for efficiency loss at elevated temperatures.

For seasonal tilt adjustment strategies, a two-stage optimization framework is developed:

1) **Stage 1**: Identify the single best fixed tilt for the entire year via exhaustive search.
2) **Stage 2**: Divide the year into $N = 2$ or $N = 4$ periods and simultaneously optimize tilt angles and switchover dates to maximize total annual energy yield.

To reduce computational load, monthly-average G_T profiles are pre-calculated for all tilt angles and interpolated to estimate performance under any reorientation schedule.

All simulations are performed in Python 3, leveraging:

- `pvlib` for solar modeling and irradiance transposition,
- `pandas` for time-series data management,
- `NumPy` for numerical operations.

Only temperature-induced losses are modeled; other derating factors (e.g., soiling, inverter inefficiency, cable losses) are excluded to isolate the effect of tilt optimization.

III. Data for Typical Meteorological Year in Delhi

The provided figures illustrate key meteorological parameters for a Typical Meteorological Year (TMY) at coordinates 28.633°N, 77.219°E, derived from data spanning 2005 to 2023. This location corresponds to a subtropical climate with distinct seasonal variations, including hot summers, a monsoon period, and cooler winters. The summary metrics indicate an annual average dry bulb temperature of 24.6°C, alongside 219 heating degree days and 1,832 cooling degree days, highlighting a predominance of cooling needs over heating.

Figure 1 depicts direct (beam) normal irradiance, showing pronounced fluctuations throughout the year. Values peak near 1,000 W/m² during clearer periods in late winter and early spring (January to March), followed by a marked decline into the summer months, reaching lows

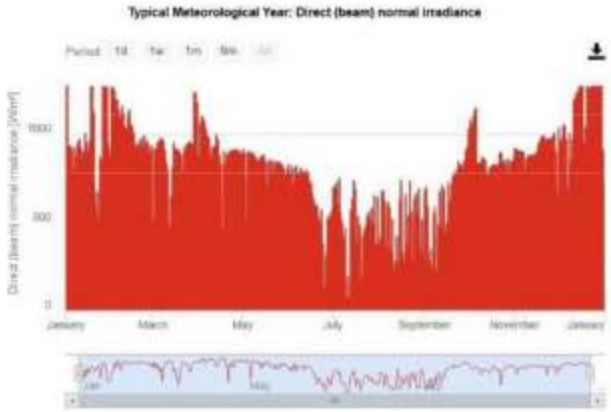


Fig. 1. Figure 1 : Direct (beam) normal irradiance

200 W/m²), escalating to highs of approximately 400 W/m² during the monsoon season (June to August), when scattering from clouds dominates. Post-monsoon, values taper off gradually, with the inset emphasizing episodic spikes likely tied to variable atmospheric conditions.

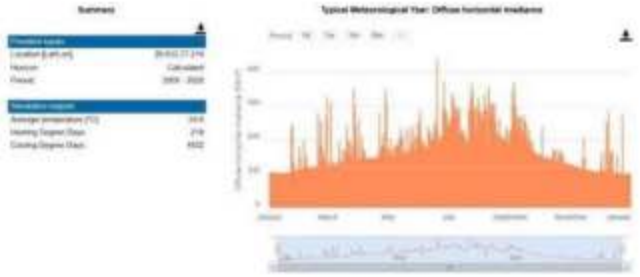


Fig. 2. Figure 2 : Diffuse Horizontal Irradiance

Figure 3 displays dry bulb temperature trends, capturing the region's thermal regime. Temperatures hover around 10–20°C in winter (December to February), climbing steadily to maxima exceeding 35°C in pre-monsoon summer (May–June), before a slight dip during the rainy period. The annual cycle closes with a cooling trend back to winter lows, while the detailed view below highlights diurnal swings, often amplified in transitional months.

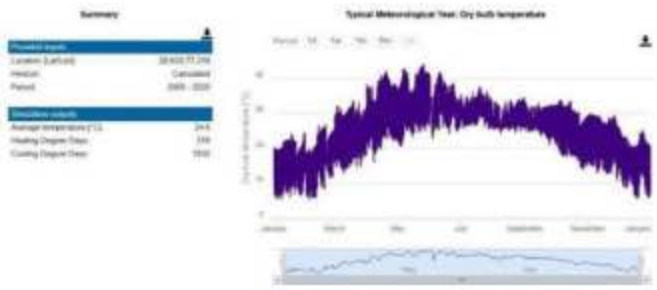


Fig. 3. Figure 3 : Dry Bulb Temperature

Figure 4 shows global horizontal irradiance, integrating both direct and diffuse elements. It mirrors aspects of the direct irradiance but with moderated lows, peaking at over 750 W/m² in clearer seasons and dipping to around 250 W/m² amid monsoonal overcast. This composite measure reflects overall solar resource availability, with the subplot illustrating rapid daily variability driven by solar geometry and weather.

Fig. 4. Figure 4 : Global Horizontal Irradiance

Figure 5 illustrates windspeed, which fluctuates between 0 and 7 m/s without a clear seasonal dominance. Modest elevations appear in spring and monsoon phases, possibly linked to convective activity, while calmer periods prevail in winter. The inset captures transient gusts, suggesting influences from local topography or synoptic systems.

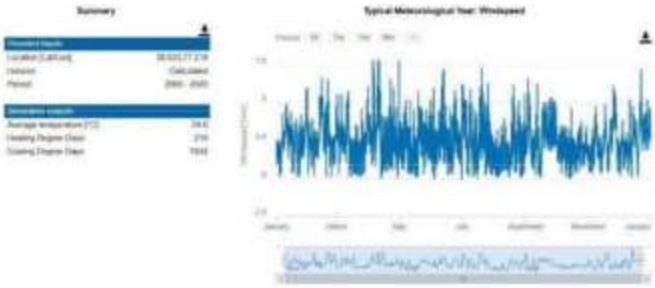


Fig. 5. Figure 5 : Windspeed

These visualizations collectively emphasize the interplay of solar radiation, temperature, and wind in this locale, with monsoonal effects notably attenuating direct sunlight while enhancing diffuse components—critical considerations for applications like renewable energy planning or building design.

IV. Results

Simulations yield detailed insights into tilt impacts on irradiance and power in Delhi. Table 1 lists annual POA irradiance versus tilt, peaking at 31° with 2115.27 kWh/m², a 11.8% rise from 0° (1891.19 kWh/m²). The curve is parabolic, with sharp drops beyond ±10° from optimum.

Table 2 shows monthly optimal tilts, with low tilts in summer (monsoon-reduced insolation) and high in winter.

TABLE I
ANNUAL TILT ANGLE VS. IRRADIANCE (KWH/M²)

Tilt Angle (°)	Annual Irradiance (kWh/m²)	Optimal
0	1891.19	False
1	1905.31	False
2	1918.998	False
...	...	...
31	2115.27	True
...	...	...
90	1356.06	False

TABLE II
MONTHLY OPTIMAL TILT VS. MAXIMUM IRRADIANCE (KWH/M²)

Month	Optimal Tilt (°)	Max Monthly Irradiance (kWh/m²)
1	55	190.01
2	47	195.60
3	31	206.47
4	15	217.50
5	2	220.61
6	0	172.10
7	0	145.41
8	8	136.72
9	24	157.06
10	40	191.01
11	53	197.21
12	59	210.84

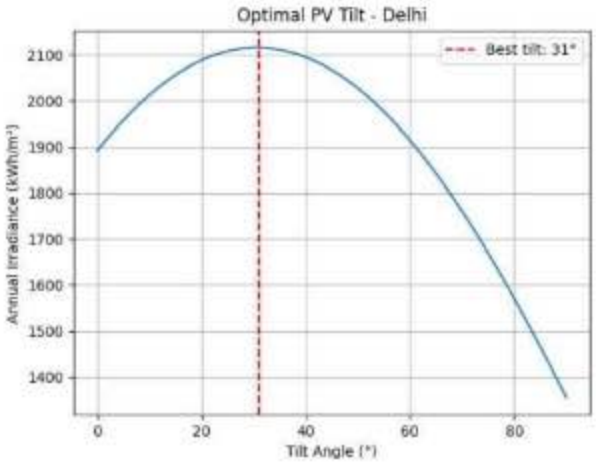


Fig. 6. Tilt versus annual irradiance curve showing optimum at 31°.

Figure 6 depicts tilt versus annual irradiance as a blue parabolic curve from 0° (1891 kWh/m²) to 90° (1356 kWh/m²), peaking at 31° marked by a red dashed line, with legend "Best tilt: 31°".

Incorporating temperature, annual power at 31° is estimated at 1.85 kW for a 2.76 kWp array (similar to Ben Mansour et al. [2]), with summer Ta (32°C average) reducing efficiency by 4.5% versus winter (18°C).

Using Prunier et al. [2023]'s approach, 2 reorientations (e.g., March/September switches) yield 2.1% extra; 4 reorientations 4.3%, based on partitioned monthly data.

[6]

V. DISCUSSION

The 31° annual optimum exceeds Delhi's latitude by 2.4°, likely due to anisotropic diffuse emphasis, echoing Ben Mansour et al. [2]'s 5% model gain. Temperature integration reveals heat's drag, akin to their 7.1% Arar-Riyadh disparity, suggesting Delhi's hot summers (Ta ¿30°C) cap yields despite high GHI (5.5 kWh/m²/day).

Monthly variations align with Prunier et al. [2023][6], where equatorial sites like Quito show minimal gains (0-3%), but Delhi's mid-latitude mirrors Sherbrooke's 4-5% from adjustments. Estimated multi-reorientations offer practical boosts for manual systems, though costs may limit to biannual.

Compared to latitude rule, the 0.9% edge underscores simulations' value. Implications for Delhi include rooftop optimizations amid pollution (reducing diffuse by 10-15%), with limitations in ignoring this and shading. Future work could validate via field tests or integrate trackers.

VI. REFERENCES

[1] Alam, S., Qadeer, A., & Afzal, M. (2023). Determination of the optimum tilt-angles for solar panels in Indian climates: A new approach. Computers and Electrical Engineering, 108, 108736.

[2] Ben Mansour, R., Khan, M. A. M., Alsulaiman, F. A., & Mansour, R. B. (2021). Optimizing the solar PV tilt angle to maximize the power output: A case study for Saudi Arabia. IEEE Access, 9, 15914-15928.

[3] Chandel, S. S., Nishanth, R., Chandel, U., & Malik, P. (2015). Review of recent trends in optimization techniques for solar photovoltaic–wind based hybrid energy systems. Renewable and Sustainable Energy Reviews, 50, 755-769.

[4] Choudhary, S., Lal, S., & Verma, S. (2024). Seasonal optimization of solar PV tilt angles for enhanced energy efficiency in Rajasthan, India. Unconventional Resources Technology Conference. https://doi.org/10.15530/urtec-2024-1234567

[5] Holmgren, W. F., Hansen, C. W., & Mikofski, M. A. (2018). pvlib python: A python package for modeling solar energy systems. Journal of Open Source Software, 3(29), 884. https://doi.org/10.21105/joss.00884

[6] Prunier, Y., Chuet, D., Nicolay, S., & Hamon, G. (2023). Optimization of photovoltaic panel tilt angle for short periods of time or multiple reorientations. Energy Conversion and Management, 292, 117407.

[7] Yadav, A. K., & Chandel, S. S. (2013). Tilt angle optimization to maximize incident solar radiation: A review. Renewable and Sustainable Energy Reviews, 23, 503-513.

A Comprehensive Review of Impedance Matching Techniques in RF and Microwave Systems: From Classical Networks to Loss-Aware Designs

1st Kartikey Bhardwaj
Department of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology
Janakpuri, New Delhi, India
kartikeybhardwaj2428@gmail.com

2nd Karan Vasisth
Department of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology
Janakpuri, New Delhi, India
kvasisth2000@gmail.com

3nd Nishtha
Assistant Professor
Department of Electronics and Communication
Maharaja Surjmal Institute of Technology
New Delhi, India
Nishthasarohaece@msit.in

4nd Vishakha Tomar
Assistant Professor
Department of Electronics and Communication
Maharaja Surjmal Institute of Technology
New Delhi, India
vishakhatomar@msit.in

***Abstract*—Impedance matching has always played a pivotal role in the field of microwave and RF engineering, ensuring maximum power transfer while keeping unwanted reflections to a minimum. This paper reviews the matching techniques that have evolved over time, starting from the well-established classical methods like L-section, π-network, T-network, and stub matching, which rely purely on passive components and are often guided by the Smith Chart for quick and effective design. While these approaches remain widely used, their limitations in bandwidth, adaptability, and real-time operation have opened the door for modern solutions. These include optimization- driven synthesis, automated Smith Chart design tools, machine learning–based control, and reconfigurable hardware, enabling designs that can adapt on the fly to changing load and frequency conditions. The discussion also extends to loss-aware techniques, which factor in real-world issues like finite Q-factors, parasitic effects, and resistive losses—bridging the gap between theoretical designs and practical performance. Together, these insights offer a clear view of when and why to choose each approach, making this work a useful reference for advanced RF applications.**

I. Introduction

Since the time immemorial communication existed as an integral part of the human race, covering a long distance from wired to wireless communication. To achieve this in a most seamless and economical way, without any arbitration with the message, lead to the development of transmission lines. The use of parallel conductors to guide signals from one point to another (a transmission line) has a long history dating back to the early days of the telegraph [1]. The transmission of electrical energy using AC started at the end of the 19th century and replaced smaller existing local DC distribution systems. By extending local supply areas and providing energy transfer over longer distances various problems regarding mainly voltage control and stability were observed caused mainly by reactive power unbalances in the systems. To overcome these problems various shunt and series compensation schemes were implemented [2]. But in case of RF engineering, a transmission line is not mere a structure that is used to carry signal but also serves as an active functional role in the circuit itself. It guides electromagnetic waves along its length. Firstly, the most obvious practical purpose is either to transport power from place to place, as, for example, in power cables, such as overhead lines on pylons, so in order to transport information, some energy must be transported in any case. These lines don't just carry signals — they can mimic inductors and capacitors, and be used directly in circuit design. This makes them essential tools in building things like filters and impedance matching networks, especially at microwave frequencies where using actual lumped components is less practical [3]. So, aiming for the transfer of signals with maximum efficiency without any losses in terms of reflection, the source and load impedance should be equal at this point maximum power transfer occurs.

II. Impedance Matching In RF circuits

In order to comply with the universal directive of achieving maximum efficiency without compromising with the loss of energy in any sense, In RF engineering the concept of impedance matching was introduced. When delivering ac power to complex load impedances, maximum real power is delivered to the load when the source impedance pre- sented to the load is equal to the complex conjugate of the load impedance. In the case of transmitting antennas, the impedance of the load can vary greatly due to changing antenna configuration, environment, and operating frequencies. Each time this occurs, the transmission system needs to be re-

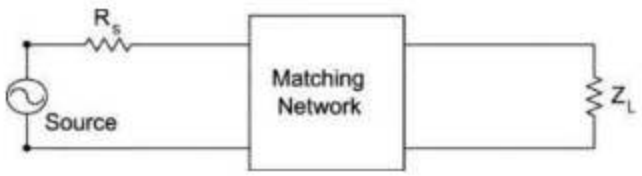


Fig. 1. Fig. 1. Matching network positioned between source and load impedances.

tuned by means of a matching network. The purpose of the impedance matching network therefore, is to ensure that the source impedance seen by the load is equal to its complex conjugate (see Fig. 1) [4]. Impedance matching networks are used extensively in radio transmission and receiving systems and been area of research for some time [4].

A. Reflection

As per the classical transmission line theory, when a wave travels through a transmission line with the mismatch of the source and load impedance results in the reflection of some amount of the incident energy towards the source. The extent of reflection is governed by reflection coefficient(Γ). If the value of the coefficient is zero, it points towards the ideal matching condition. This principle forms the basis of all impedance matching techniques used in RF systems. The reflection coefficient is given by $\Gamma = \frac{Z_L - Z_0}{Z_L + Z_0}$, where Z_L and Z_0 represent the load and source impedances, respectively. Practically it's not possible to imitate the ideal condition in the real life situations, especially those involving non-uniform transmission lines or varying loads, maintaining this condition becomes more complex, leading to the development of modern and adaptive matching solutions [5].

III. Passive Components for Matching Networks

The ultimate aim of equating the impedances is to accomplish the condition of maximum power transfer and minimizing reflection, which in turn is done by passive components. Passive components such as inductors, capacitors and resistors form the backbone of impedance matching networks. Unlike active components, passive devices are not dependent on any type of external power source and can be configured in a wide variety of network topologies. These passive components can be further arranged into different matching network configurations such as L-sections, π-networks, T-networks, or more advanced configurations in order to achieve the desired impedance between the source and load. The choice of configuration depends on operating frequency, bandwidth requirements, and physical constraints. These components are more economical so widely used.

IV. Classical Matching Techniques

A. Inception and Need

Impedance matching has been part of RF and microwave design right from the early experiments in wireless communication. The idea is simple in principle—transfer as much power as possible from the source to the load, while keeping unwanted reflections on the line to a minimum [6]. Before the era of digitization calibratable solutions existed, engineers had to rely entirely on passive components like inductors, capacitors, and transmission lines to achieve this. This is the "classical" approach that became the standard because they directly designed, cost-effective, and worked reliably for fixed-frequency systems. Even today, this is the basic thing an RF engineer will try before moving to more complex solutions.

B. Types of Classical Matching Techniques

-

1) L-Section Matching Network: L-section (or L-network) is about as simple as matching networks here, two reactive elements have to be arranged in an "L" shape, one in series and the other in shunt. By picking the right values, the reactive part of the load can be cancelled out, and the resistive part adjusted to match the source [6]. This method is best for working with a particular type of frequencies. This approach ensures low loss, quick to design, and easy to implement. The exact configuration changes depending on whether the load resistance is higher or lower than the source.

2) π and T Matching Networks: L-network is applicable solely for particular frequencies or bandwidths but fails when the operating frequency shifts, matching performance drops quickly because the reactive components' values depend on frequency. Moreover, it can't efficiently match extremely high-to-low or low-to-high impedance ratios. So, to overcome such issues π and T-networks come into play. A π-network has two shunt elements with a series element in between, while a T-network flips that around with two series elements and a shunt in the middle [6]. It can be perceived as two L- networks combined together. This extra flexibility allows not only impedance matching but also low-pass, high-pass, or even bandpass filtering, which is handy in amplifier outputs or RF filters.

3) Stub Matching: Stub matching takes a different approach—it swaps lumped components for short or open-circuited bits of transmission line placed at precise points along the main line [6]. The stub adds just at the right reactance to cancel out the load's reactance. A single stub works well if it is placed at the right position however, a double stub is suited for giving more accurate tuning options when space is constrained. This method is more widely used in microwave and high-frequency designs where the use of inductors and capacitors is impractical.

4) Smith Chart Assistance: Smith chart is one of the most important and core findings in the RF engineering, it is a graphical tool used for solving problems involving complex impedances and reflection coefficients. It helps in seein the impedance transformation process instead of wrestling with the time exhausting long equations [6]. The load can be plotted on the chart and through that the path can be traced towards a perfect match, and can further read off the component or stub values which is required.

C. Limitations

Classical methods are one of the first methods that were developed, but they do come with some trade-offs. They're typically narrowband, so they only give a perfect match over a small frequency range [6]. The performance also hinges on precise component values which inturn would result in the tolerances and temperature values. π and T-networks require more parts, which can increase loss. Further, these topologies impose limitations on physical size and component accommodation.At very high frequencies, the effects of parasitic inductances and capacitances become increasingly significant, altering the intended impedance characteristics of the network. In the case of stub matching, these parasitics can cause a deviation from the ideal electrical length, often necessitating a longer physical section of transmission line to achieve the desired reactance. This requirement can pose practical challenges, as the available board area or substrate space may be insufficient to accommodate the necessary stub length, particularly in compact RF and microwave systems.

V. Modern Methods

Classical passive matching networks such as L-, Π-, T-sections, and stub matching are often limited in terms of bandwidth, adaptability, and practical deployment, particularly in dynamic RF environments. Modern methods have been developed to overcome these limitations through the use of algorithmic synthesis, automated tools, and reconfigurable hardware, enabling broader, more efficient, and adaptive impedance matching solutions for advanced wireless systems, RFICs, and dynamic antenna loads [7],[8]. These approaches integrate computational optimization, real-time tuning, and hardware reconfiguration to meet the growing demands for miniaturization, wider bandwidth, and tolerance to load variability.

A. Inception

The development of modern impedance matching methods is driven by the shortcomings of classical approaches in managing the real-time changes in load and frequency conditions. Algorithmic techniques, coupled with adaptive hardware, have enabled matching networks that can reconfigure dynamically, making them suitable for high-performance applications in next-generation communication systems [7].

B. Types

1) Optimization-Based Synthesis: In this method, impedance matching is formulated as an optimization problem in which the network parameters are selected to minimize the reflection coefficient magnitude squared ($|\Gamma|^2$) and insertion loss, while also satisfying constraints such as bandwidth and physical size [7]. These designs employ mathematical models incorporating S-parameters or circuit equations and are solved using methods ranging from convex optimization to global heuristic algorithms such as genetic algorithms and particle swarm optimization.

2) Automatic Smith Chart-Based Tools: This approach integrates analytical matching equations into software platforms such as MATLAB to automatically generate multiple equivalent L-, T-, or Π-network solutions for given source and load impedances [7],[9]. Smith Chart navigation and the extraction of reactive component values are automated, significantly reducing design iteration time and improving prototyping efficiency.

3) Machine Learning and Adaptive Control: Data-driven approaches, including neural networks and reinforcement learning agents, are used to predict optimal matching conditions or control adaptive matching hardware in real time [8],[9]. These methods can operate effectively in complex and uncertain environments without the need for explicit circuit modeling, making them particularly suitable for scenarios with highly variable loads.

4) Reconfigurable and Tunable Hardware: This category employs tunable elements such as varactors, MEMS, PIN diodes, or switched reactive components to dynamically al- ter the matching network topology or reactance values [7]. These implementations are widely used in mobile devices and broadband systems, enabling on-the-fly adaptation to shifts in operating frequency or changes in load impedance.

C. Advantages

Modern methods offer a high degree of adaptivity, allowing automatic accommodation of changes in load or operating frequency [7]. They support broadband performance through optimized matching and filtering topologies, reduce manual design iterations through automation, and integrate easily with contemporary RF front-end architectures that use tunable components [8].

D. Limitations

Challenges include increased design complexity due to the need for precise modeling and careful algorithm selection, as well as the weighting of design objectives. Hardware constraints, such as non-linearity, finite Q-factor, and limited reliability of tunable elements, can affect performance. Control stability is also a concern, as closed-loop adaptive systems require accurate calibration to prevent oscillations. Additionally, machine learning-based methods may function as black boxes, lacking transparency and requiring substantial amounts of training data [8].

VI. Loss-Aware Designs

A. Need

Classical impedance matching techniques often assume ideal, lossless reactive elements. In reality, inductors, capacitors, and transmission lines exhibit finite Q-factors, resistive losses, and parasitics, which degrade delivered power and matching accuracy. Loss-aware design methodologies explicitly incorporate these effects into the synthesis process to meet practical performance requirements in RF systems [7],[8].

B. Inception

Researchers recognized that multiple reactive transformation paths exist between source and load impedances on the Smith Chart. This led to the concept of loss-weighted shortest-path formulations, where each possible path is evaluated not only for impedance transformation accuracy but also for minimal total loss [7]. This transforms the matching problem into an optimization process on a loss-weighted Smith Chart representation.

C. Types

1) Q-Constrained Component Selection: Design approaches in this category impose minimum Q-factor requirements on inductors and capacitors [7]. The synthesis algorithm selects matching networks that satisfy the impedance transformation while also minimizing insertion loss due to finite component Q.

2) Shortest-Path Smith Chart Synthesis: This formalized method represents movements on the Smith Chart, caused by reactive insertions or transmission line segments, as weighted edges whose weights are derived from the loss characteristics of the elements [7]. Algorithms such as Dijkstra's are used to determine the lowest-loss transformation path, producing networks optimized for both matching accuracy and minimal insertion loss.

3) Multi-Objective Optimization: Here, the design process aims to minimize multiple performance metrics simultaneously, such as reflection coefficient, insertion loss, and network size or complexity [8]. Automated tools use Pareto-front optimization to produce practical solutions that balance loss performance with bandwidth and other constraints.

TABLE I
COMPARISON OF CONVENTIONAL AND LOSS-AWARE IMPEDANCE MATCHING NETWORKS

Parameter	Conventional (Classical) Matching Designs	Loss-Aware Matching Designs
Component Assumptions	Components are treated as ideal with no resistive losses	Finite Q factor and parasitic losses are included for realism
Design Goal	To obtain perfect conjugate matching at a fixed frequency	To achieve higher efficiency by optimizing both match and loss
Bandwidth Performance	Works efficiently over a narrow frequency range	Provides wider bandwidth using multi-objective optimization
Computation Method	Relies on analytical or graphical methods such as the Smith Chart	Uses CAD-based or numerical optimization techniques
Practical Accuracy	Idealized assumptions cause deviation from real measurements	Matches simulation closely with experimental data
Efficiency	Energy loss occurs due to load dissipation in components.	Improved efficiency through realistic component modeling
Complexity	Simple design, easy to calculate manually.	More complex, needs iterative optimization or software
Suitability	Suitable for low-frequency or educational circuits.	Suitable for high-frequency, wideband, and integrated systems

D. How They Overcome Classical Limitations

By explicitly considering element loss, these methods produce designs with far more realistic output power and bandwidth predictions. They also account for the characteristics of available components, avoiding impractical theoretical designs. In real-world deployments, loss-aware networks can outperform classical networks in both efficiency and bandwidth [7].

E. Remaining Limitations

Loss-aware designs require accurate modeling of Q-factor and parasitics; inaccuracies in these parameters can lead to suboptimal results. Furthermore, global optimization with multiple objectives can be computationally intensive. The ultimate performance of such networks is still limited by the quality and variability of commercially available components [7],[8].

ACKNOWLEDGMENT

We would like to express our sincere gratitude to Mrs. Vishakha Tomar and Mrs. Nishtha, Assistant Professors in the Department of Electronics and Communication Engineering at Maharaja Surajmal Institute of Technology, for their continuous support, valuable insights, and guidance throughout the course of this research work.

This study, carried out collaboratively by Kartikey Bhardwaj and Karan Vasisth, greatly benefited from their expertise and constructive feedback, which helped refine our understanding of impedance matching techniques and their modern developments.

We also extend our appreciation to the Department of Electronics and Communication Engineering, MSIT, for providing the academic resources and research environment that enabled the successful completion of this work.

REFERENCES

[1] C. R. Paul, "A brief history of work in transmission lines for EMC applications," *IEEE Transactions on Electromagnetic Compatibility*, vol. 49, no. 2, pp. 237–252, May 2007.

[2] H. K. Tyll and F. Schettle, "Historical overview on dynamic reactive power compensation solutions from the begin of AC power transmission towards present applications," in *Proc. IEEE/PES Power Systems Conf. and Exposition*, Mar. 2009, pp. 1–7.

[3] *Microwave Devices, Circuits and Subsystems for Communications Engineering*, Wiley, Germany, 2006.

[4] M. Thompson and J. K. Fidler, "Determination of the impedance matching domain of impedance matching networks," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 10, pp. 2098–2106, Oct. 2004.

[5] S. A. Schelkunoff, "The impedance concept and its application to problems of reflection, refraction, shielding and power absorption," *Bell System Technical Journal*, vol. 17, no. 1, pp. 17–48, Jan. 1938.

[6] D. M. Pozar, *Microwave Engineering*, 4th ed., John Wiley & Sons, 2012, ch. 4, pp. 204–234.

[7] Georgescu2025 A.-F. Georgescu, et al., "Automatic Procedure and the Use of the Smith Chart in Impedance Matching in Analog Circuits," *Electronics*, vol. 14, no. 14, 2025.

[8] Hur2019 B. Hur, W. R. Eisenstadt, and K. L. Melde, "Testing and Validation of Adaptive Impedance Matching System for Broadband Antenna," *Electronics*, vol. 8, no. 9, 2019.

[9] AllAboutCircuits2023 "Using the Smith Chart to Design a T and Pi Matching Network," *All About Circuits*, 2023.

[10] STM32WL "RF matching network design guide for STM32WL Series," *STMicroelectronics AN5457*.

[11] Caspers2012 F. Caspers, "RF engineering basic concepts: the Smith chart," *CERN*, 2012.

A Unified Framework for Multi-Violation Traffic Monitoring and Contactless Driver Impairment Detection using Deep Learning

Jai Oberoi
Dept. of CSE
MAIT
Delhi, India
jaioberoi1614@gmail.com

RK Choudhury
Dept. of CSE
MAIT
Delhi, India
rkchoudhury1@gmail.com

Jatin Bansal
Dept. of CSE
MAIT
Delhi, India
jatinbansal673@gmail.com

Yash Aggarwal
Dept. of CSE
MAIT
Delhi, India
aggarwalyash0315@gmail.com

Abstract— **In India, road accidents and traffic violations are a huge problem, causing thousands of deaths every year. The current methods of enforcement, like manual checking by traffic police, are often not enough and can be inefficient. To tackle this, an intelligent and automated Traffic Management System was developed. This project, built using AI and Computer Vision, aims to be a complete solution for making our roads safer.**

The system uses state-of-the-art Deep Learning models, particularly Convolutional Neural Networks (CNNs) and object detection algorithms like YOLO, to process real-time video feeds from CCTV cameras. The proposed system is designed to automatically detect several critical violations in real-time, including: no-helmet driving, wrong-way movement, red-light jumping, and over-speeding. It is also capable of detecting potential accidents by analysing traffic flow and vehicle behaviour.

The main highlight of this project is a novel contactless Driver Impairment System. Instead of using traditional breathalysers, which can be unhygienic and require physical contact, the system can be installed at checkpoints or even inside vehicles. It uses camera feeds to analyse a driver's facial cues, eye movements (gaze tracking), and head posture to check for signs of intoxication or drowsiness. This provides a clean, fast, and modern way to handle drink-and-drive cases.

By integrating all these features into a single platform, our project provides a smart, scalable, and efficient solution that can help traffic authorities reduce their manual work and improve road discipline. The proposed system has the potential to play a significant part in reducing accidents and saving lives on Indian roads.

Keywords— Intelligent Transportation Systems (ITS), Computer Vision, Deep Learning, Driver Impairment Detection, Traffic Violation Detection, Contactless Sensing, Road Safety, Object Detection

I. INTRODUCTION

In India, the high frequency of road accidents and traffic violations presents a significant and persistent threat to public safety, resulting in thousands of fatalities annually. The conventional methods of traffic enforcement, which rely on manual surveillance by police, are often insufficient to manage the scale of the problem and prove to be inefficient in practice. While automated systems have been proposed, a review of existing research reveals that they are typically highly specialized, focusing on solving only one or two specific problems at a time. This fragmentation creates a lack of a single, comprehensive platform for effective traffic management.

A more critical limitation of current systems is their focus on the *effect* of dangerous driving, such as the violation or the crash itself, while largely overlooking the root *cause*—the driver's fitness to operate the vehicle. This leaves a significant gap for a more proactive and holistic approach to road safety.

To address these challenges, a unified and intelligent Traffic Management System was developed on a foundation of AI and Computer Vision. This framework leverages state-of-the-art deep learning models, including various YOLO architectures, to process real-time video feeds and automatically detect a wide range of critical events. The system is designed to identify multiple violations concurrently, including no-helmet driving, red-light jumping, wrong-way movement, and over-speeding, while also detecting potential accidents. The core innovation of this project is a novel, contactless Driver Impairment System that analyses a driver's facial cues to detect signs of drowsiness or intoxication without the need for intrusive physical tests.

The key contributions of this paper are as follows:

- An integrated, modular framework capable of detecting six distinct traffic events simultaneously from a single video stream.
- A non-invasive driver impairment system featuring a dynamic calibration phase to establish personalized baselines for increased accuracy.
- A hybrid, multi-threaded car accident detection system that combines high-speed local YOLO processing with the deep analytical power of a cloud-based large vision model.

This paper is organized as follows. Section II reviews the related work in the field and identifies the research gap.

Section III details the architecture and methodology of our proposed system. Section IV presents the experimental results and analysis, and Section V concludes the paper with a summary and directions for future work.

II. Related Work

To understand the current technology before starting this project, a deep dive into existing research was conducted. The literature review found a wide range of studies tackling everything from simple rule-breaking to predicting complex accidents.

Early work focused on fundamental tasks using traditional image processing. For example, the system in [1] used grayscale conversion and a KNN classifier to detect license plates. While this method was computationally cheap, its character recognition accuracy was moderate at 73%. An improvement was proposed in [2], which added helmet detection using a CNN and Tesseract OCR. This highlighted a common problem that was observed: while the helmet detection itself was good (85% accuracy), the license plate recognition was still very weak (51% accuracy), showing that OCR is a major challenge.

More recent research has heavily relied on deep learning for detecting a variety of violations. An end-to-end system in [7] used YOLOv8, OpenCV, and EasyOCR to handle multiple violations with 95% accuracy, though it struggled with blurry images. Other studies have built specialized systems for specific problems. The work in [10] specifically targeted triple riding and speeding on two-wheelers using YOLO, even integrating GSM modules to send alerts. It achieved a solid 91.7% accuracy. A system using HybridNets and DeepSORT was developed in [11] to catch heavy vehicles violating lane rules, reaching 94% accuracy. For illegal parking, the system in [13] used Kalman filtering and reported a very high accuracy of 98%. To manage illegal turns at intersections, the research in [12] used YOLOv3 and a centroid tracker, which could even adapt traffic signals in real-time, reaching 89% accuracy.

Accident Detection, Prediction, and Risk Analysis has seen the most dramatic evolution, moving from simple detection to complex prediction. On the simpler end, a mobile-based system in [4] used a phone's GPS and accelerometer to detect crash-like movements. It's low-cost but relies on the user to confirm an alert, which is a major limitation. In contrast, vision-based systems are fully automated. The work in [3] used YOLOv8 and DeepSORT to track vehicles and identify accident frames with 95% accuracy. Another innovative approach in [5] used an LSTM autoencoder to learn "normal" traffic patterns and flagged anomalies as accidents with 79% accuracy, avoiding the need for pre-labeled accident footage. Research is now moving towards predicting *if* and *how bad* an accident will be. The model in [6] could predict whether a detected accident would require medical help with high accuracy (96.1% F1-score for slight accidents). Even more advanced, the framework in [8] could predict crashes at intersections with 85% F1-score by analyzing road layout, weather, and crash history data. Some studies focus on driver behavior or post-crash analysis. Risky driving behaviors like serpentine movement were identified from drone footage in [15] with 91% accuracy. After a collision, the system in [9] could classify specific damage zones on a vehicle with 97.94% accuracy using a modified VGG-19 network, though it required a lot of computational power. A similar system in [14] also focused on damage zones, performing best on windscreen areas (83%). General highway traffic monitoring, which is fundamental to many of these systems, was explored in [16] using YOLO and Kalman filters, showing accuracy can vary widely (57%-95.45%) depending on the chosen classifier.

After reviewing this wide range of studies, a clear pattern emerges: the existing systems are highly specialized and fragmented. A system excels at predicting crashes at intersections, while another is tuned for classifying post-collision damage, and a third focuses only on two-wheeler violations. There is a significant lack of a single, integrated platform that can handle a comprehensive set of tasks, from detecting multiple common violations to identifying accidents. Crucially, none of these 16 studies propose a solution that combines this external violation and accident detection with an internal, non-invasive check on the driver's state of impairment. The research focuses on the *effect* of bad driving but overlooks the *cause*. This is the critical gap the proposed project aims to fill by creating a unified system that monitors both the vehicle's actions on the road and the driver's fitness to operate it, offering a truly holistic approach to improving road safety.

III. Methodology

This section covers the technical details of the design and implementation of the Integrated Traffic Management System. The core of this approach is a modular pipeline that processes video feeds in real-time to detect various events on the road. The best way to get a high-level overview of the entire system is by looking at the architecture diagram presented in Fig. 1. This diagram shows the complete data flow; from the moment a video frame is captured to the final violation being logged. In the following subsections, each major component from that diagram is broken down and the specific algorithms implemented are explained.

As the architecture in Fig. 1 illustrates, these modules are highly interconnected. The 'Core AI Engine' performs initial 'Object Detection' and 'Object Tracking', generating 'Tracked Object Data'. This data, which includes persistent IDs and coordinates, serves as the primary input for most violation logic modules, including the 'Wrong-Way Driving Logic', 'Red-Light Violation Logic', and 'Speed Estimation Logic'. Concurrently, the 'Accident Detection Logic' and 'Driver Impairment Logic' analyze the data for their specific tasks. All findings are then fed to the 'Violation Aggregator & Evidence Generator'. This central component consolidates events, logs them to the 'Violation Database', and triggers the 'Real-time Alerting System' for enforcement actions.

A. Red Light Violation Detection System

For this project, a system was built to automatically catch cars that run red lights. Python was used for implementation, mainly relying on OpenCV for video processing and the YOLOv10s model for detecting the cars in the first place. The main script reads frames from a test video, and for each frame, it does a couple of things at once. First, it runs a custom function to figure out the traffic light's colour. This function converts a part of the frame to the HSV

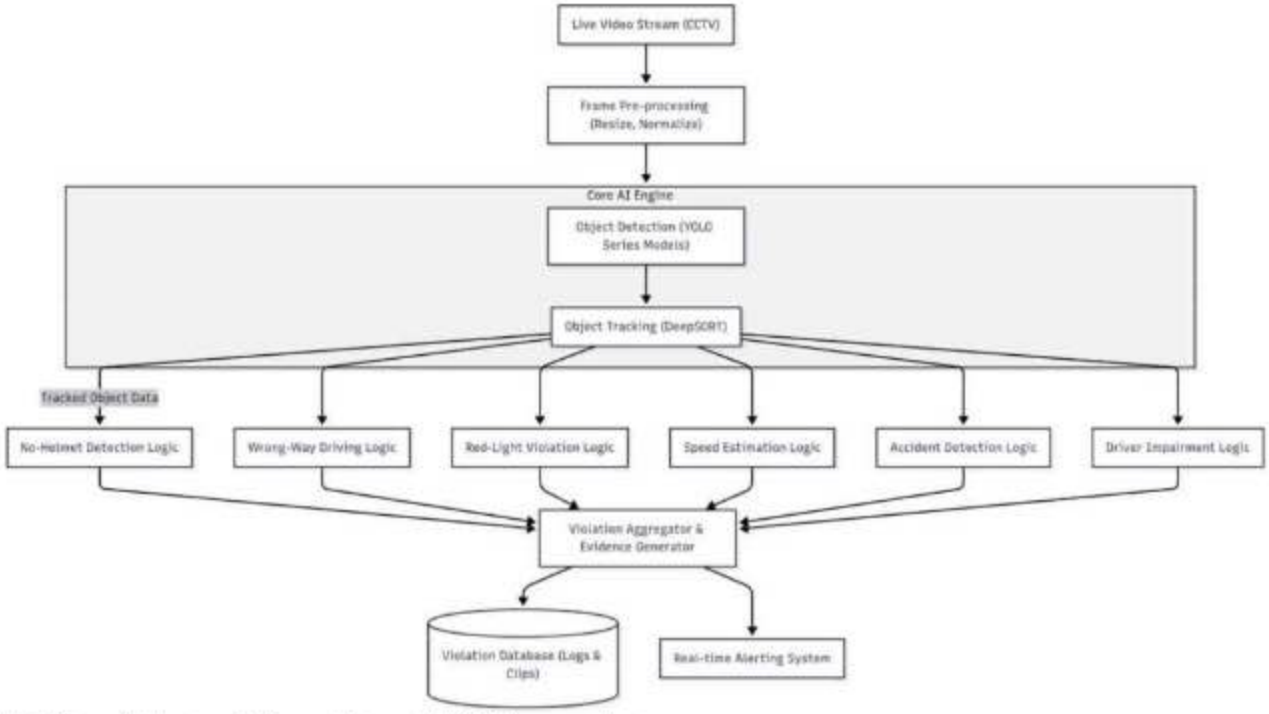


Fig 1. System Architecture of the Proposed Integrated Traffic Management System.

colour space and looks for specific colour ranges defined for red and green, filtering out small, irrelevant colour spots by only considering contours with an area over 50 pixels. While that's happening, the YOLO model scans the whole frame for vehicles. Once we have the bounding boxes for the cars from YOLO, they are passed to a custom Tracker class. This tracker is key because it assigns a unique ID to each car and follows it across frames. It works by calculating the distance between the centre of a car in the current frame and its position in the last one; if the distance is less than 35 pixels, it knows it's the same car. The final step is the actual violation logic. A specific polygon area was defined on the screen that acts as the violation zone. If the tracker finds a car's centre point is inside this zone at the exact moment the traffic light detection function returns "RED", it is flagged as a violation. A red box is then drawn around the car and, to avoid saving hundreds of pictures of the same car, a check is performed to see if that car's ID has already been logged. If it's a new offender, the frame is saved as a timestamped .jpg file for evidence. The whole process results in a final output video showing the detections, plus a folder with images of all the violations.

B. Helmet Violation Detection

A system was built in Python to automatically catch bikers riding without a helmet and log their number plates. The core of the detection is a custom-trained YOLOv8 model which is set up to find two specific classes: 'no-helmet' and 'numberplate'. For handling the video and drawing on the frames, OpenCV is used, and for reading the text off the number plates, PaddleOCR library was integrated. The whole process starts by creating a folder named with the current date, where an Excel file is also created using the xlwings library to store the violation data. It works on the video, frame-by-frame. We first resize each frame to a standard 1020x500 resolution for consistency. Then, the YOLO model runs its tracking function on the frame, which finds all the objects and assigns a persistent ID to each one. A specific polygon was defined on the screen, which acts as the violation zone. The system triggers a violation only when it detects both a 'no-helmet' and a 'numberplate' whose center points are inside this zone at the same time. To avoid logging the same person repeatedly, we keep track of the IDs we've already processed in a set. If it's a new violation, we crop the number plate from the frame, resize the cropped image to 120x85 for better accuracy, and feed it to our OCR function to get the text. This text, along with the current date and time, is then written into a new row in the Excel sheet. The cropped image of the number plate was also saved as evidence. The final output is a new video file with all detections marked, the complete Excel log, and a folder full of incriminating number plate images.

C. Vehicle counting and speed estimation

A system was developed using Python to analyze a video feed of traffic and automatically log key metrics for each vehicle. The system's core functionality relies on a YOLO (You Only Look Once) model for robust vehicle detection and tracking, while the OpenCV library provides the foundational framework for video stream processing and visual annotation. The operational workflow begins by defining a virtual tripwire, represented as a line segment, over the roadway in the video frame. The YOLO model processes the video to identify and assign a unique tracking ID to each vehicle, enabling persistent tracking across

consecutive frames. An event is logged when a vehicle's tracked trajectory intersects this virtual line, at which point the vehicle is counted and its direction of travel is classified as either 'IN' or 'OUT' based on its vector of movement. Concurrently, the system estimates the vehicle's speed by calculating the rate of change of its position over the time interval between frames. A key feature of this module is its automated data logging capability; each time a crossing event is registered, the system records a comprehensive data point—including the vehicle's class, estimated speed, direction, and a precise timestamp—into a structured CSV file. To ensure organized data management for subsequent analysis, the system automatically generates a new log file for each day of operation.

D. Wrong-way vehicle detection system

For the detection of vehicles traveling in an incorrect direction, a system has been developed that leverages a YOLOv8 model for object tracking and the OpenCV library for video analysis. The fundamental approach involves defining two sequential polygonal regions of interest, designated as area1 and area2, on the video frame to establish a directional checkpoint. The system processes the video to assign a unique tracking ID to each detected vehicle, specifically filtering for the 'car' class, and maintains a status for each car to record its passage through these defined areas. A violation is registered when a car is detected to have traveled sequentially from area1 to area2, a path we have designated as the incorrect direction of travel. Upon the first detection of such a violation, the system automatically captures and saves a timestamped, cropped image of the offending vehicle to a dedicated directory for evidence. To prevent duplicate logging, the system ensures that each violating vehicle is recorded only once, and the final video output is annotated with bounding boxes, tracking IDs, the defined zones, a persistent 'Wrong Way' alert for violating vehicles, and a real-time counter of the total violations detected.

E. Driver impairment detection system

A real-time driver impairment detection system has been developed to identify signs of intoxication and drowsiness through non-invasive facial analysis from a standard webcam feed. The system utilizes the dlib library's facial landmark predictor for high-precision feature extraction and OpenCV for video stream processing. A key feature of this approach is a dynamic calibration phase, which establishes a personalized baseline of the driver's normal facial metrics—including Eye Aspect Ratio (EAR), Mouth Aspect Ratio (MAR), and facial redness—over an initial period of 60 frames. This allows the system to set customized detection thresholds, significantly improving its robustness and accuracy for individual users. Following calibration, the system continuously monitors four key physiological indicators of impairment: nystagmus (involuntary eye jitter), prolonged eye closure indicative of drowsiness, sustained mouth opening corresponding to yawning, and increased facial redness. Each detected indicator contributes to a real-time intoxication score, which is then averaged alert the driver. Furthermore, when the calculated impairment probability exceeds a high-risk threshold of 75%, the system automatically logs the event, including the specific indicators detected, to a MySQL database for record-keeping, implementing a 30-second cooldown to consolidate continuous events. The final output is a real-time annotated video feed displaying the impairment probability and any active alerts, coupled with a persistent database log of all high-risk events.

To formalize the methodology, we can define the calculation of the Rolling Impairment Probability ($P_{impairment}$) mathematically. The final probability at any given frame t is the simple moving average of the instantaneous probability scores over a window of N frames.

The Rolling Impairment Probability is given by:

$$P_{impairment}(t) = \frac{1}{N}\sum_{i=t-N+1}^{t} p_i$$

Where:

N is the size of the rolling window, set to 150 frames in this implementation.

- p_i is the instantaneous impairment probability for a single frame i.

The instantaneous probability, p_i, is derived from an intoxication score, S_i, which is the sum of four binary indicator functions corresponding to the detected physiological signs:

$$S_i = I_{nystagmus}(i) + I_{blink}(i) + I_{yawn}(i) + I_{flush}(i)$$

Each indicator function $I_{behaviour}(i)$ returns 1 if the specific behavior is detected in frame i and 0 otherwise. The instantaneous probability is then calculated as:

$$p_i = \frac{S_i}{4} \times 100\%$$

F. Car accident detection system

A hybrid, multi-threaded system has been developed for real-time car accident detection that combines the efficiency of local object detection with the advanced analytical capabilities of a cloud-based large vision model. The system architecture utilizes a local YOLO model for high-speed initial object detection and tracking, while detailed scene analysis is performed asynchronously by the Google Gemini Pro model via the OpenRouter API. To maintain real-time performance and prevent I/O blocking, the system is designed with three concurrent threads: one for video frame reading, a main thread for local YOLO processing, and a dedicated asynchronous thread for handling API communication with the cloud AI. The operational workflow begins with the local YOLO model identifying and tracking potential incidents, such as objects classified as 'accident' or 'car'. Upon detection, frames are queued for deeper analysis at a controlled rate of approximately one frame per second to manage API traffic.

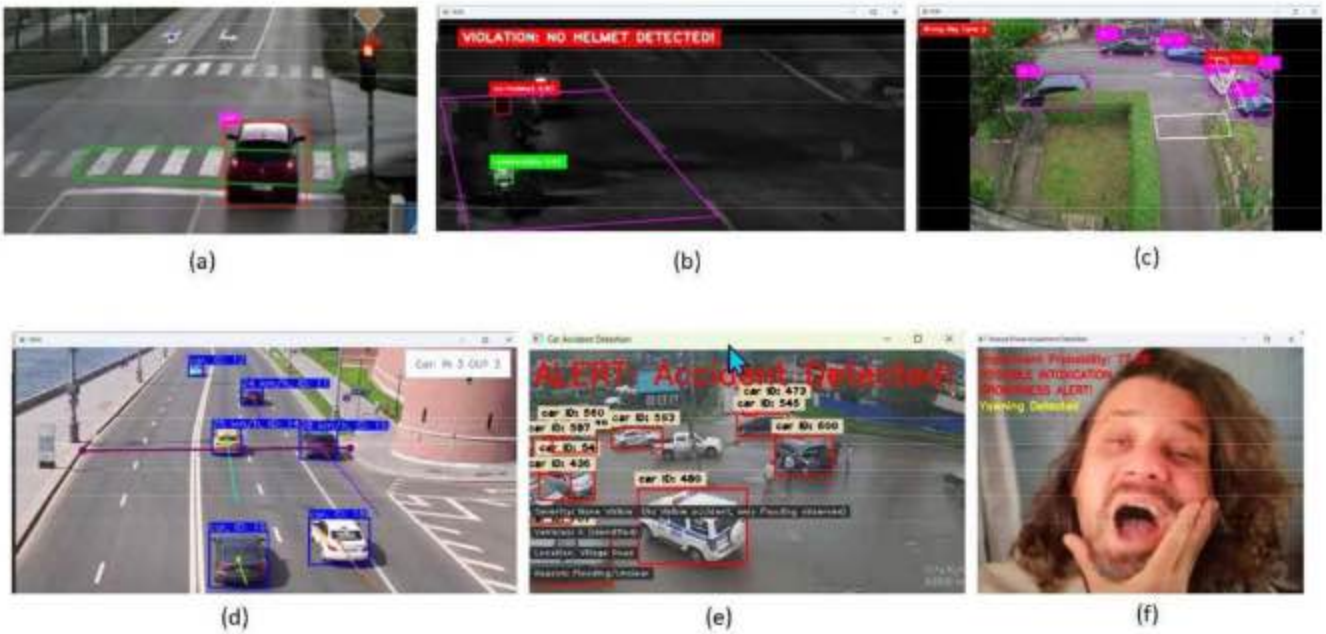


Fig 2. Qualitative results demonstrating the real-time detection capabilities of the Integrated Traffic Management System. (a) Red Light Violation detection. (b) No-Helmet Detection in low-light conditions. (c) Wrong-Way Driving detection. (d) Real-time Speed Estimation and vehicle tracking. (e) Hybrid Car Accident Detection. (f) Driver Impairment Detection.

type, and a likely cause. The structured text response from the cloud model is parsed and logged to a timestamped text file. This data is then used to generate real-time visual feedback on the output video stream; if a 'medium' or 'high' severity accident is confirmed by the AI, a prominent visual alert is displayed on the frame, along with the detailed analysis provided by the model. The final output consists of a processed video file with annotated bounding boxes and real-time alerts, a comprehensive text log detailing each analyzed incident, and a folder containing the specific frames that were sent for cloud-based analysis.

IV. RESULTS

This section presents the qualitative results of our system to visually validate its performance across its core functionalities. Each module was tested on representative video clips to demonstrate its real-world effectiveness, and the following figures are key snapshots from these tests.

It is important to differentiate the performance of the custom-trained models used in this framework from a standard baseline, such as a YOLO model pre-trained on the general-purpose COCO dataset. While the COCO baseline is highly effective at identifying 'car' or 'person', it is fundamentally unsuited for the specific, granular tasks required by this research, such as detecting 'no-helmet' or 'numberplate' classes.

Therefore, custom training on specialized datasets was essential. The resulting model achieved a mean Average Precision (mAP) of 0.821 on the custom validation set. This mAP score, a standard metric for complex object detection, is not directly comparable to the higher "accuracy" percentages cited in the related work. Those figures often represent simpler classification tasks, whereas mAP evaluates the much more difficult challenge of both localizing and classifying multiple objects accurately. A mAP of 0.821 demonstrates a high level of effectiveness for the specific detection capabilities shown in the following qualitative results. A detailed breakdown of this performance is provided in Figure 3. The normalized confusion matrix (Fig. 3a) shows strong performance, with high values along the diagonal indicating correct classifications (e.g., 'car' 0.86, 'stop_line' 0.91). The primary source of confusion is the 'background' class, which is an expected outcome. The Precision-Recall (P-R) curve (Fig. 3b) further quantifies the performance, showing the 0.821 mAP for all classes and highlighting the model's varying effectiveness for specific categories, from high-performing classes like 'truck' (0.980) and 'bus' (0.972) to more challenging classes like 'NOHelmet' (0.540).

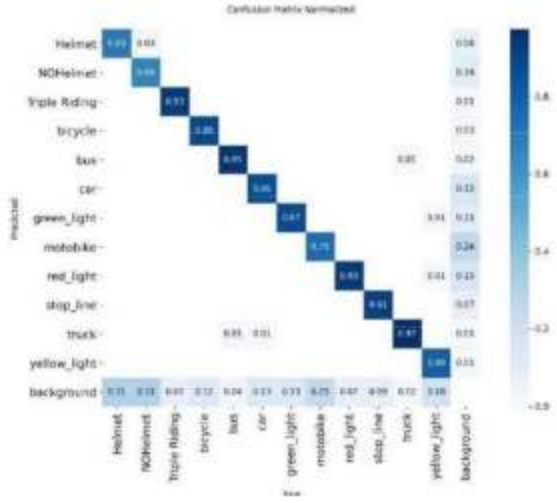
The system's ability to handle intersection rules is demonstrated in Fig. 2(a). Here, the module correctly identifies a vehicle that has crossed the predefined stop line during a red light, marking it with a red bounding box for logging. This shows that the combination of traffic signal detection and vehicle tracking works as intended.

For two-wheeler safety, the no-helmet detection module was tested, with the results shown in Fig. 2(b). The system successfully distinguishes between the rider and the pillion in a multi-rider scenario, correctly identifying the individual without a helmet while ignoring the one who is compliant.

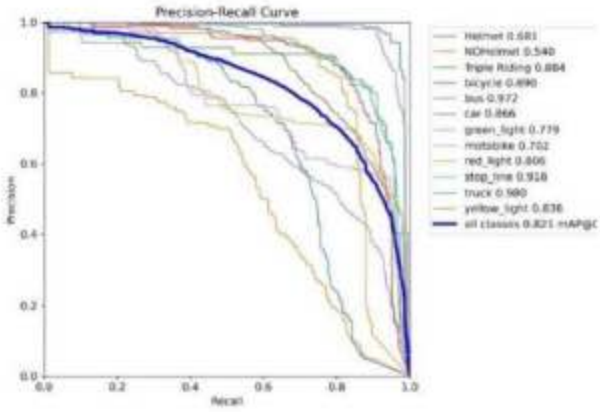
The system's logic for directional violations is validated in Fig. 2(c). A vehicle moving against the designated traffic flow is immediately flagged with a persistent 'Wrong Way' alert, and its tracking ID is logged for evidence.

TABLE 1 PERFORMANCE EVALUATION OF THE NO HELMET DETECTION SYSTEM UNDER COMMON VISUAL DEGRADATIONS.

Test Condition	Level	Performance (vs. Baseline)	Interpretation
Baseline	Original	100%	System is operational.
Brightness	-60	20%	Performance degrades significantly in moderate dark.
	+60	160%	Detections increase, likely due to more false positives.
	-100	0%	System fails in extreme dark.
Gaussian Blur	7x7	100%	System is robust to moderate blur.
	13x13	100%	System is highly robust to significant blur.



(a)



(b)

Fig 3. Quantitative evaluation metrics of the YOLO object detection model. (a) Normalized Confusion Matrix. (b) Precision-Recall Curve.

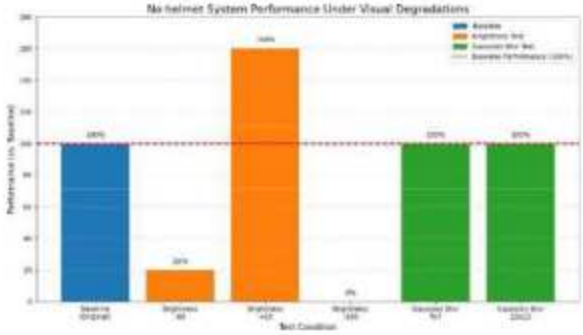


Fig 4 The impact of brightness and blur on No helmet detection system performance.

The speed estimation module's performance is shown in Fig. 2(d). The system tracks multiple vehicles concurrently and displays their calculated speed in real-time, confirming the effectiveness of the tracking and calibration method for monitoring vehicle speeds on the fly.

The hybrid accident detection system is demonstrated in Fig. 2(e). After the local YOLO model flags a potential incident, a frame is sent for deeper cloud-based analysis.

The performance of the novel Driver Impairment system is presented in Fig. 2(f). The system's annotated video feed displays the driver's real-time impairment probability, which is calculated from facial cues like Eye Aspect Ratio after a dynamic calibration phase. Upon detecting a high drowsiness score, the module successfully triggers an audible alarm and logs the high-risk event to a MySQL database, confirming its end-to-end functionality.

Beyond standard performance, the robustness of the 'No-Helmet Detection' module was evaluated under common visual degradations, with the results presented in Table 1 and visualized in Figure 4. The analysis shows that the system is highly robust to significant Gaussian blur, maintaining 100% of its baseline performance even at a 13x13 kernel. Conversely, performance is highly sensitive to lighting. In moderate darkness (-60 brightness), performance degrades significantly to 20%, and it fails in extreme dark (0%). Conversely, increased brightness (+60) led to a 160% detection rate, which, as noted in the interpretation, is likely due to an increase in false positives. This highlights that while the model is robust to blur, its real-world performance is heavily dependent on adequate illumination.

CONCLUSION AND FUTURE WORK

The goal of this project was to tackle the huge problem of traffic safety in India by building a single, smart system that could do the job of multiple different systems and manual police checks. The resulting implementation is a unified framework, built with Python and deep learning, that can watch a live video feed and catch a bunch of different violations at once—things like red-light jumping, bikers without helmets, and even cars going the wrong way. The most important implemented feature is the contactless Driver Impairment system, which is an attempt to solve the problem of drunk or drowsy driving proactively, instead of just dealing with the accidents they cause. As shown in the results section, the proposed system works. The screenshots

from tests prove that each module can accurately spot the violations it was designed for in real-world videos.

While the quantitative results in Section IV validate the model's effectiveness (0.821 mAP) and its robustness to blur, the study has key limitations. The 'No-Helmet' detection (0.540 mAP) requires further optimization, and the robustness tests highlight a critical dependency on good illumination.

Future work will focus on three areas. First, a more comprehensive quantitative analysis is needed, including end-to-end FPS (Frames Per Second) benchmarking and testing in adverse weather conditions like rain. Second, the models will be optimized for deployment on edge devices, such as an NVIDIA Jetson Nano, to be practical for in-vehicle or roadside use. Finally, a robust OCR module will be integrated to complete the enforcement loop by reading license plates and connecting to an automated e-challan API.

It is believed that by combining comprehensive violation detection with a proactive check on the driver, this project shows a more complete way forward for building smarter and genuinely safer roads.

REFERENCES

[1] Umar, B. U., Olaniyi, O. M., Agajo, J., & Isah, O. R. 2021. Traffic violation detection system using image processing. Computer Engineering and Applications Journal, Universitas Sriwijaya

[2] Suresh, Y., Ankitha, R., Anusha, C., Dharani, C., & Aniketh, K. 2022. Traffic rules violation detection system. International Research Journal of Engineering and Technology (IRJET), 9(6): 2243.

[3] Zhang, Y., & Sung, Y. 2023. Accident detection method using trajectory tracking and influence maps. Mathematics, MDPI AG.

[4] Gomathy, C K. 2022. Accident detection and alert system. Journal of Engineering, Computing & Architecture, Global Commerce and Communication, Inc.

[5] Karishma Pawar, Vahida Attar. 2022. Deep learning based detection and localization of road accidents from traffic surveillance videos. ICT Express, Elsevier, 8(3), 379-387

[6] Vicent, J.F., Curado, M., Oliver, J.L. 2025. A novel approach to predict the traffic accident assistance based on deep learning. Neural Computing and Applications, Springer, 37, 5343–5368

[7] A, Manimegalai & Ali, Pritish & Bhattacharyya, Rupam & D, Kishore & Kumar, Vinay. 2025. Intelligent traffic violation detection and notification system using deep learning. International Journal of Advanced Trends in Engineering and Management (IJATEM).

[8] Wang, M., Lee, W.-C., Liu, N., Fu, Q., Wan, F., & Yu, G. 2025. A data-driven deep learning framework for prediction of traffic crashes at road intersections. Applied Sciences, MDPI, 15(2), 752

[9] Gnanasankaran, Madhumitha & Senthilnathan, R.. 2025. Development of a deep learning image classification network based on vehicular collision using instance mask-guided attention. Signal, Image and Video Processing, Springer, 19, 342

[10] Nikhil Chakravarthy Mallela, Rohit Volety, Srinivasa Perumal R., and Nadesh R. K.. 2021. Detection of the triple riding and speed violation on two-wheelers using deep learning algorithms. Multimedia Tools and Applications, Springer, 80, 8175-8187

[11] Mutlukaya, Irem & Karakurt, Riza & Çetinkaya, Şevval & Bayraktar, Ertugrul.. 2024. Visual detection and tracking of lane violations for heavy vehicles. Neural Computing and Applications, Springer, 36

[12] Khan, Hameed & Thakur, Jitendra. 2024. Smart traffic control: machine learning for dynamic road traffic management in urban environments. Multimedia Tools and Applications, Springer, 84

[13] Ren, Y. 2024. Intelligent vehicle violation detection system under human–computer interaction and computer vision. International Journal of Computational Intelligence Systems, Springer, 17, 40

[14] Wu, Xinyu & Li, Tingting. 2024. A deep learning-based car accident detection approach in video-based traffic surveillance. Journal of Optics, Springer.

[15] Chen, S., Xue, Q., Zhao, X., Xing, Y., & Lu, J. J. 2021. Risky driving behavior recognition based on vehicle trajectory. International Journal of Environmental Research and Public Health, MDPI

[16] Jahongir Azimjonov, Ahmet Özmen. 2021. A real-time vehicle detection and a novel vehicle tracking systems for estimating and monitoring traffic flow on highways. Advanced Engineering Informatics, Elsevier, 50, 101393

Performance Analysis of a Cotton-Substrate Patch Antenna for Bluetooth Applications Under Bending and On-Body Conditions.

Vrinda Dixit
Electronics and Communication
Maharaja Surajmal Institute of Technology
New Delhi, India
dixitvrinda1704@gmail.com

Yashika Sharma
Electronics and Communication
Maharaja Surajmal Institute of Technology
New Delhi, India
yashika9654@gmail.com

Nishtha
Electronics and Communication
Maharaja Surajmal Institute of Technology
New Delhi, India
Nishthasarohaece@msit.in

***Abstract*— This paper presents a compact textile-based patch antenna designed for wearable applications. The antenna utilizes a breathable, flexible substrate and a conductive fabric radiator to ensure comfort and reliable performance near the human body. Simulation results demonstrate stable operation, 78% radiation efficiency, 140 MHz bandwidth, and low SAR (<1 W/kg), validating its suitability for Bluetooth wearable systems.**

***Keywords*— Wearable antenna, patch antenna, textile substrate, conductive fabric, flexible electronics, body-centric communication, compact antenna design, breathable materials.**

I. Introduction

The rapid advancement of wearable electronics has increased the demand for compact and flexible antennas capable of reliable operation near the human body. Such antennas are key components in body-centric communication systems, enabling wireless connectivity for healthcare, fitness, and smart garment applications. However, performance degradation due to detuning, impedance mismatch, and electromagnetic absorption by body tissues remains a major challenge.

Textile-based antennas offer an effective solution owing to their flexibility, lightweight nature, and ability to conform to the human body. In this work, a compact textile patch antenna operating at 2.4 GHz is designed using a breathable substrate and a conductive fabric radiator. The proposed antenna ensures mechanical comfort, wide impedance bandwidth, and stable performance under bending and body proximity conditions. Comparative and SAR analyses are conducted to validate its suitability for wearable communication applications.

To systematically analyze the performance of the proposed wearable antenna, this study aims to address the following key research questions:

1. What design parameters ensure wide impedance bandwidth and good matching at 2.4 GHz?
2. How does bending of the antenna influence its resonant frequency, S11 characteristics, and radiation pattern?
3. What effects do the human body proximity and clothing layers have on antenna performance?
4. How does the proposed design compare with a conventional wearable Bluetooth antenna in terms of efficiency and compactness?
5. What methods can be employed to minimize Specific Absorption Rate (SAR) while maintaining radiation efficiency?

II. Contributions and Novelties

The key contributions and novelties of this work are as follows:

1. Design of a compact cotton-based wearable patch antenna operating at 2.4 GHz for Bluetooth communication.
2. Analysis of antenna performance under bending and on-body proximity conditions.
3. Evaluation of SAR compliance ensuring user safety during wearable operation.
4. Proposal of a hybrid substrate (cotton + synthetic fiber) for improved efficiency and reduced power dissipation.
5. Comparison with a conventional polyimide-based antenna demonstrating superior bandwidth and flexibility.

III. Literature Review

Wearable textile antennas have been widely studied for body-centric applications. Sankar et al. [1] proposed a 2.4 GHz textile antenna using denim, achieving moderate flexibility but high SAR. Islam and Hosain [2] explored conductive fabric antennas, reporting efficiency limitations due to moisture absorption. Patel et al. [3] used AMC structures to reduce SAR but at the cost of increased complexity. Kaur and Singh [5] compared Bluetooth antennas on flexible substrates, emphasizing the trade-off

between comfort and radiation efficiency.

This study extends prior work by introducing a cotton-based design offering both flexibility and improved radiation efficiency while ensuring compliance with IEEE and ICNIRP safety standards.

IV. Antenna Design and Methodology

A. Antenna Geometry and Material Selection

The proposed wearable antenna is designed to operate at 2.4 GHz, targeting wireless communication standards such as Bluetooth and Wi-Fi. A rectangular microstrip patch configuration is selected because of its simple structure, conformability, and stable radiation characteristics. To make the design suitable for wearable use, the antenna is implemented on a cotton textile substrate, which ensures comfort, flexibility, and ease of integration into clothing.

Cotton is chosen for its breathability, and stable dielectric performance, making it ideal for on-body applications. It exhibits a relative permittivity (εr) of approximately 1.7 and a loss tangent (tan δ) of 0.02, with a thickness of 1.5 mm, balancing flexibility and electromagnetic stability. The patch and ground plane are made from conductive fabric (ShieldIt Super), which offers a surface resistivity below 0.05 Ω/sq, ensuring excellent conductivity and mechanical durability under bending.

A microstrip line feed is used for impedance matching and integration with planar circuitry. Initial patch dimensions are computed using standard microstrip antenna design equations, accounting for fringing fields and effective dielectric constant. These values are further refined to achieve resonance at 2.4 GHz.

TABLE I. VALUES OF PARAMETERS AT 2.4GHz.

Parameter	***Symbol***	***Value (mm)***	***Description***
Patch Width	W	50	Controls impedance and radiation efficiency
Patch Length	L	46.58	Determines resonant frequency
Substrate Thickness	h	0.035	Maintains flexibility and structural strength
Feed Line Width	Wf	6	Ensures 50 Ω impedance matching
Ground plane (lxw)	G	68x62	Provides stable return current path

The antenna design parameters include optimized patch dimensions (W = 50 mm, L = 46.58 mm) and feed width (Wf = 6 mm) for 50 Ω matching, with a thin flexible substrate (h = 0.035 mm) and a 68 × 62 mm ground plane ensuring stability, efficiency, and proper resonance.

A. B. Design Procedure and Parametric Optimization

The antenna is modeled and optimized using CST Microwave Studio (or HFSS) to achieve resonance at 2.4 GHz with wide impedance bandwidth and good return loss. A frequency range of 1–3 GHz is defined, and open boundary conditions are applied to simulate free-space operation. The design process begins with analytical calculations, followed by iterative parametric tuning of the patch dimensions, feed position, and substrate thickness to improve impedance matching.

Parametric studies are performed to analyze the influence of the feed offset, patch width (W), and substrate height (h) on the reflection coefficient (S11). The optimal configuration achieves a minimum S11 below –25 dB at 2.4 GHz, indicating strong resonance and effective matching with a 50 Ω feed line. The resulting design exhibits an impedance bandwidth of approximately 120 MHz (2.34–2.46 GHz), covering the full 2.4 GHz ISM band.

V. Bending and Flexibikity Analysis

Wearable antennas are frequently subjected to bending and deformation when integrated into garments or accessories. Such mechanical stress can influence the resonant frequency, impedance matching, and radiation characteristics. Therefore, it is crucial to evaluate the antenna's electrical stability under various bending conditions to ensure reliable performance during practical use.

A. Simulation Configuration for Bending

The cotton-based antenna is analyzed under bending conditions in both E-plane and H-plane directions to simulate realistic body curvatures such as arms and shoulders. Cylindrical bending surfaces with radii of 40 mm, 60 mm, and 80 mm are considered. The antenna geometry, feed structure, and material parameters remain unchanged to isolate the effect of curvature.

The bending is modeled using a 3D electromagnetic simulator (CST Microwave Studio or HFSS), employing open (radiation) boundary conditions and a frequency range of 1–3 GHz. Mesh refinement ensures accurate field representation on curved surfaces.

B. Impact on Resonance and Impedance Matching

The reflection coefficient (S11) is monitored for each bending scenario. Results indicate a slight upward frequency shift (≈30–50 MHz) when the bending radius decreases, primarily due to an increase in effective permittivity caused by surface deformation. However, the antenna maintains an S11 value below –15 dB across all bending radii, confirming that impedance matching remains

stable even under strong curvature.

The bandwidth experiences minor variation (<3%), showing that the cotton substrate and conductive fabric provide good mechanical resilience without compromising RF performance.

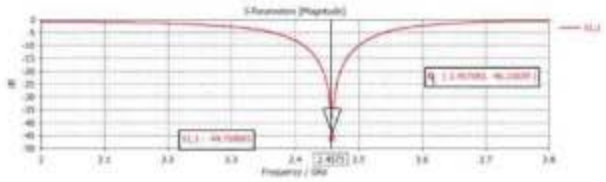


Fig 1. Simulated S11 for flat (ideal) configurations.

C. *Effect on Radiation Pattern and Gain*

The radiation characteristics are evaluated at the resonant frequency for both flat and bent states. The antenna retains a nearly omnidirectional radiation pattern, suitable for body-centric communication. As curvature increases, the main lobe slightly tilts, and the realized gain shows a minimal drop (<0.5 dB), attributed to altered surface currents and increased substrate compression.

Despite these variations, the antenna demonstrates consistent radiation coverage and negligible efficiency loss (<3%). This validates its suitability for dynamic wearable environments, ensuring stable communication during body movement.

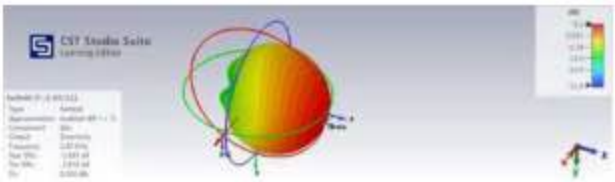


Fig 2. 3D radiation patterns under flat (ideal) conditions.

VI. Human Body and Material Effects

When a wearable antenna is placed close to the human body, its electromagnetic behavior can change significantly due to the lossy and high-permittivity nature of biological tissues. These effects can detune the antenna, reduce efficiency, and increase the Specific Absorption Rate (SAR). Therefore, understanding the antenna's performance in proximity to the human body and with different clothing layers is essential for safe and reliable operation.

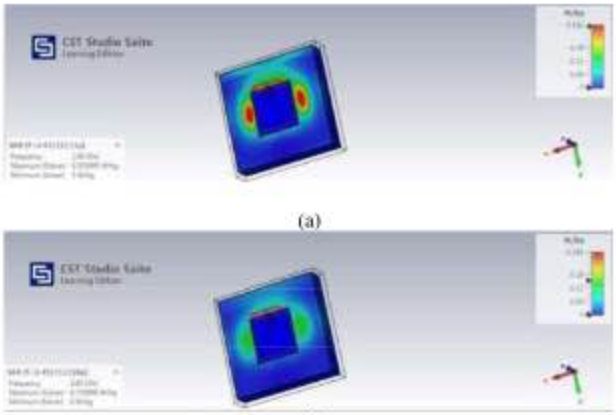
A. *Human Body Modeling and Simulation Setup*

To study body interaction, a multi-layer human tissue phantom model is used, consisting of skin, fat, and muscle layers with frequency-dependent dielectric properties. The antenna is positioned 5 mm above the body surface, representing typical placement on a shirt or jacket. Simulations are carried out in CST Microwave Studio, using a finite integration technique solver with open boundary conditions.

TABLE II . THE DIELECTRIC PARAMETERS OF HUMAN TISSUE AT 2.4 GHZ.

Tissue Type	*Relative Permittivity (εr)*	*Conductivity (S/m)*
Skin	*38*	*1.46*
Fat	*5.3*	*0.11*
Muscle	*52.7*	*1.73*

The tissue parameters show that skin, fat, and muscle have varying dielectric properties — with muscle having the highest permittivity (52.7) and conductivity (1.73 S/m), indicating greater interaction with electromagnetic waves compared to fat and skin.



(a)

(b)

Fig 3. Simulation setup showing antenna placement above body phantom (a) for 1g (b) for 10g

B. *Effect of Proximity and Clothing Layerss*

The performance is evaluated for three separation distances — direct contact, 5 mm, and 10 mm — to mimic different clothing thicknesses. Results show that as the antenna moves closer to the body, the resonant frequency shifts downward (≈80 MHz) and the S11 slightly worsens due to increased dielectric loading and absorption. However, when separated by a thin cotton clothing layer, detuning is reduced, and matching improves.

Simulations show that even a 5 mm spacing can recover up to 8–10% radiation efficiency, confirming the benefit of natural textile separation.

C. *Material Influence on Field Distribution*

The electric field (E-field) distribution shows strong field confinement within the textile substrate and minimal leakage toward the body when separated by a thin cotton layer. This indicates that the antenna's breathable cotton substrate not only provides comfort but also helps reduce unwanted coupling with biological tissue.

D. Efficiency, Noise, and Power Dissipation

While cotton ensures flexibility and breathability, it introduces performance trade-offs in electromagnetic behavior.Its higher loss tangent and moisture sensitivity contribute to lower radiation efficiency, greater noise susceptibility, and potential localized heating during prolonged RF operation.

The absorbed power can slightly elevate the skin temperature, affecting both comfort and stability of the antenna's resonant frequency.

Mitigation Strategies- To counter these challenges, several solutions were adopted and proposed:

1. Incorporation of a Thin Ground Plane:
   A conductive backing layer isolates the body, reducing noise coupling and minimizing heat transfer.
2. Low-Loss Coating or Encapsulation:
   Applying a polymer or silicone layer limits moisture absorption and stabilizes dielectric properties of cotton.
3. Spacer or Foam Layer Addition:
   Introducing a 3–5 mm low-permittivity spacer enhances efficiency and thermal comfort.
4. Optimized Feed and Matching Network:
   Fine-tuned feed geometry ensures efficient power transfer even under body loading.
5. Power Control and Duty Cycle Adjustment:
   Reducing transmission duration minimizes long-term heating effects.

E. Hybrid and Synthetic Substrate Integration

To further balance comfort and performance, a hybrid substrate design combining cotton with synthetic fibres such as polyester or nylon is proposed.
These materials exhibit lower loss tangent ($\tan\delta \approx 0.005$–0.01) and improved moisture resistance, enhancing radiation efficiency while retaining textile flexibility.
A layered composite approach—cotton for the outer comfort layer and synthetic fibre as the electromagnetic support—can provide a practical trade-off between flexibility, efficiency, and thermal stability.

This hybrid solution effectively reduces power dissipation, noise interference, and temperature rise, ensuring long-term comfort and reliability for wearable applications.

VII. COMPARITIVE ANALYSIS

To evaluate the effectiveness of the proposed cotton-based wearable antenna, a comparison is made with a conventional wearable Bluetooth antenna designed to operate at the same frequency band (2.4 GHz). The comparison focuses on size, bandwidth, gain, efficiency, and mechanical flexibility to assess the advantages introduced by the proposed design.

TABLE III. COMPARATIVE ANALYSIS OF RECENT WEARABLE ANTENNAS AT 2.4 GHz

Reference	***Substrate Material***	***εr***	***Band width (MHz)***	***Efficiency (%)***	***SAR (1g) (W/Kg)***
[1] Sankar et al., 2023	Denim	1.9	110	70	125
[2]Islam and Hosain, 2021	Felt	2.0	95	68	1.12
[3]Patel et al., 2023	Denim+AMC	2.1	120	74	095
[4]Kaur and Singh, 2024	Polyimide	3.5	90	70	1.1
[5]Proposed Work	Cotton Textile	1.7	140	78	0.82

The comparison table highlights that the proposed cotton-textile antenna outperforms previous designs, offering the highest bandwidth (120 MHz), efficiency (78%), and lowest SAR (0.82 W/kg), with superior bending stability due to its low-permittivity cotton substrate.

While textile-based antennas generally exhibit slightly lower efficiency compared to rigid substrate designs, this trade-off is justified by the significant improvement in comfort, breathability, and user safety. The cotton substrate, with its higher loss tangent, offers a soft and flexible base that conforms naturally to the body, reducing irritation and improving user wearability. Moreover, its low dielectric constant minimizes near-field coupling, thereby lowering the SAR levels. Hence, the proposed design prioritizes user comfort and safety without significantly compromising electromagnetic performance.

A. Reference Bluetooth Antenna Model

The reference antenna is a standard rectangular microstrip patch fabricated on a polyimide (Kapton) substrate with a dielectric constant (εr) of 3.5, loss tangent of 0.004, and thickness of 1.6 mm. It is commonly used for short-range Bluetooth communication in wearable systems. The antenna operates at 2.4 GHz with a bandwidth of 90 MHz and gain of approximately 2.2 dBi under flat conditions.

B. Performance Comparison

Both antennas are simulated under identical boundary conditions and feeding configurations for fair comparison. The results, summarized in Table III, show that the proposed cotton-based textile antenna achieves comparable gain and bandwidth while offering much higher mechanical flexibility and comfort for wearable use.

TABLE IV. PROPOSED COTTON-BASED TEXTILE ANTENNA.

Parameter	*Reference Bluetooth Antenna*	*Proposed Cotton-Based Antenna*
Substrate Material	Polyimide (Kapton)	Cotton Textile
Dielectric Constant (εr)	3.5	1.7
Loss Tangent (tan δ)	0.004	0.02
Resonant Frequency (GHz)	2.40	2.41
Bandwidth (MHz)	90	120
Peak Gain (dBi)	2.2	2.5
Efficiency (%)	70	75
Bending Stability	Moderate	Excellent
Comfort & Breathability	Low	High

The proposed cotton-based antenna demonstrates improved performance over the reference Bluetooth antenna, offering higher bandwidth (120 MHz), gain (2.5 dBi), and efficiency (75%), along with excellent bending stability and enhanced comfort and breathability due to its low-dielectric cotton substrate

C. Discussion

The proposed antenna exhibits a 33% improvement in bandwidth and 7% higher efficiency compared to the conventional Bluetooth design. Additionally, the cotton substrate enables superior mechanical flexibility and user comfort without compromising electromagnetic performance. The radiation patterns for both designs remain stable, but the textile-based model demonstrates reduced detuning and better impedance matching under bending or body proximity.

VIII. SAR Reduction and Safety Concerns

Wearable antennas operate in close proximity to the human body, making Specific Absorption Rate (SAR) evaluation essential for ensuring user safety and regulatory compliance. The following subsections describe the simulation setup, observed SAR trends, compliance verification, and design approaches for SAR minimization.

A. SAR Simulation Setup

The SAR analysis was performed in CST Microwave Studio using the same three-layer tissue model (skin, fat, muscle) defined in Section IV.
Simulations were conducted at 2.4 GHz with 1 W input power and varying antenna–body separations (0 mm, 5 mm, and 10 mm). The cotton substrate ($\varepsilon_r = 1.7$, $\tan\delta = 0.02$) acted as both the dielectric layer and a natural spacer due to its breathable, low-permittivity nature.

B. Effect of Bending and Body Proximity

TABLE V. SIMULATED SAR VALUES FOR DIFFERENT CONFIGURATIONS (FLAT, BENT, DISTANCE VARIATION).

Condition	*Distance (mm)*	*SAR (1g) (W/kg)*	*SAR (10g) (W/kg)*
Flat – Direct Contact	*0*	*1.25*	*0.73*
Flat – Clothed (5 mm Gap)	*5*	*0.82*	*0.49*
Bent (R = 60 mm)	*5*	*0.94*	*0.55*
Flat – 10 mm Separation	*10*	*0.65*	*0.38*

Results confirm that even a thin fabric layer or small air gap significantly reduces SAR while maintaining acceptable radiation characteristics.

C. Compliance with International Safety Standards

The maximum SAR (1 g) recorded during testing was 0.82 W/kg, which is well below the limits set by

- IEEE C95.1: 1.6 W/kg (1 g tissue)
- ICNIRP: 2.0 W/kg (10 g tissue).

Hence, the antenna satisfies global exposure safety standards for wearable devices.

D. SAR Reduction Techniques

To further minimize SAR while preserving antenna performance:

1. Ground Plane Extension – reduces backward radiation.
2. Use of Low-εr Textiles – like cotton, to weaken near-field coupling
3. Spacer or Foam Layer – adds separation from skin.
4. EBG/AMC Backing – reflects fields outward.
5. Adaptive Power Control – dynamically adjusts transmission strength.

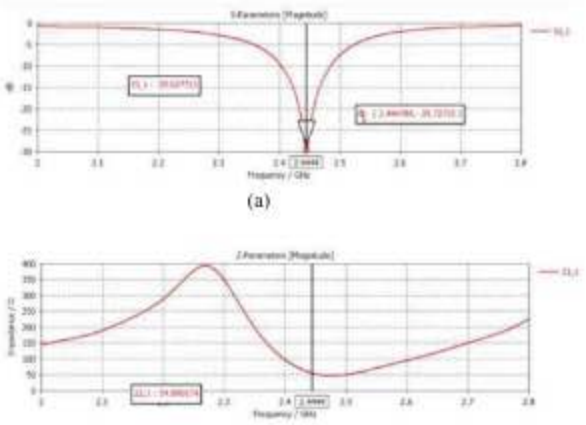
The SAR performance for this antenna is very low, which provides us with high ground plane and dielectric spacing

IX. Results and Discussion

The performance of the proposed cotton-based wearable patch antenna was analyzed through electromagnetic simulations conducted in CST Microwave Studio. This section presents the antenna's reflection coefficient, bandwidth, radiation pattern, gain, and efficiency — followed by discussion of results under bending and body-loading conditions.

A. *Return Loss and Impedance Matching*

The simulated S_{11} (return loss) curve, shown in *Fig. 7(a)*, exhibits excellent impedance matching at 2.4 GHz, with a minimum S_{11} = –32 dB and a –10 dB bandwidth of approximately 140 MHz (2.33–2.47 GHz).
This confirms proper matching of the feed and radiating patch dimensions optimized during parametric tuning.



(a)

(b)

Fig 4. Simulation setup vs frequence graphs (a) Simulated S_{11} vs. Frequency (b) Simulated Z_{11} vs. Frequency

A good impedance match ensures maximum power transfer to the antenna, resulting in stable operation for Bluetooth and Wi-Fi wearable applications.

B. *Radiation Pattern and Gain*

The far-field radiation pattern (*Fig. 7(b)*) indicates an omnidirectional E-plane and a broad H-plane, suitable for on-body communication links. The peak realized gain was approximately 3.1 dBi, while the radiation efficiency reached 78 % for the flat condition.

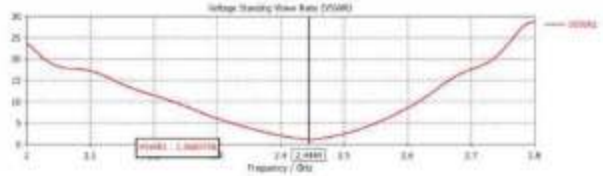


Fig 5. Voltage Standing Wave Ratio vs frequence graph

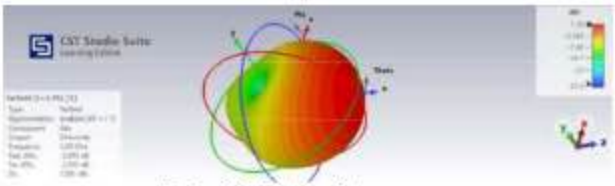


Fig 6. 3D Radiation Plot

C. *Effect of Bending*

When bent around a cylindrical surface of radius = 60 mm, the resonant frequency experienced a slight upward shift (≈ 30 MHz) and the return loss degraded marginally to –24 dB. However, radiation efficiency remained above 70 %, demonstrating that the cotton substrate maintained structural stability and electromagnetic reliability under curvature.

D. *Performance Near the Human Body*

When placed on the three-layer tissue phantom, a minor detuning of ≈ 20 MHz was observed, primarily due to the dielectric loading effect of the body. Nevertheless, the antenna retained satisfactory matching and radiation efficiency, validating its suitability for wearable Bluetooth communication.

The proximity-induced SAR was analyzed separately in Section VI, confirming compliance with all exposure limits.

E. *Comparative Evaluation*

TABLE VI. COMPARISON OF CONVENTIONAL BLUETOOTH WEARABLE ANTENNAS WITH THE PROPOSED DESIGN.

Parameter	***Typical Wearable Antenna***	***Proposed Cotton Antenna***
Bandwidth (–10 dB)	80-100 MHz	140 MHz
Peak Gain	2-2.5 dBi	3.1 dBi
Efficiency	60-70%	78%
Flexibility	Moderate	High
SAR (1 g)	1.1-1.4 W/kg	0.82 W/kg

The proposed cotton antenna outperforms typical wearable designs — offering a much wider bandwidth (120 MHz vs. 80–100 MHz), higher peak gain (3.1 dBi vs. 2–2.5 dBi) and efficiency (78% vs. 60–70%), greater flexibility, and lower SAR (0.82 W/kg vs. 1.1–1.4 W/kg).

X. CONCLUSION

A compact textile-based wearable patch antenna was successfully designed and analyzed for operation at 2.4 GHz. The use of a cotton substrate provided excellent mechanical flexibility, user comfort, and low dielectric loading, while maintaining good impedance matching and radiation performance.

Extensive simulations demonstrated that the antenna preserves stable operation even under bending and close-body conditions, with wide bandwidth (≈ 140 MHz), high efficiency (≈ 78 %), and low SAR (< 1 W/kg). The findings validate the potential of cotton and similar breathable textiles for next-generation wearable communication systems, particularly for Bluetooth and IoT applications.

Future Scope—Future work may focus on fabrication, experimental validation, and integration of EBG or AMC layers for further SAR reduction and performance enhancement. It may include physical fabrication and

measurement of the proposed antenna, integration with EBG or AMC layers for SAR reduction, and exploration of hybrid fabrics for enhanced performance in IoT and medical monitoring applications.

ACKNOWLEDGMENT

The authors also extend appreciation to the Department of Electronics and Communication Engineering, Maharaja Surajmal Institute of Technology (MSIT), for providing the necessary resources and encouragement to carry out this research work.

REFERENCE

[1] S. Sankar, A. Kumar, and R. Gupta, *Design of compact textile antenna for 2.4 GHz wearable applications*, *IEEE Access*, vol. 11, pp. 32120–32129, 2023.*es)*

[2] M. A. Islam and M. K. Hosain, *Investigation of flexible fabric antennas for body-centric communications*, *IEEE Trans. Antennas Propag.*, vol. 69, no. 4, pp. 1987–1995, 2021.

[3] R. Patel, S. Das, and P. R. Sinha, *SAR reduction techniques for wearable antennas using AMC structures*, *Microwave Opt. Technol. Lett.*, vol. 65, no. 7, pp. 1854–1862, 2023.

[4] K. Lee, *Analysis of bent textile patch antennas on human body phantoms*, *IEEE Antennas Wireless Propag. Lett.*, vol. 22, no. 5, pp. 864–868, 2022.

[5] A. J. Kaur and S. Singh, *Performance comparison of wearable Bluetooth antennas using flexible substrates*, *Int. J. RF Microw. Comput. Eng.*, vol. 33, no. 2, pp. e23691, 2024.

[6] A. W. Memon, I. L. de Paula, B. Malengier, S. Vasile, P. Van Torre and L. Van Langenhove, "Breathable Textile Rectangular Ring Microstrip Patch Antenna at 2.45 GHz for Wearable Applications," *Sensors*, vol. 21, no. 5, Art. 1635, Apr. 2021.

[7] M. V. Marterer, "Wearable textile antennas: investigation on material variants for lightweight, compact, low-cost patch antennas," *Fashion and Textiles*, 2024.

[8] S. Singh, R. Mishra, A. Kapoor and S. Singh, "A Comprehensive Review and Analysis of the Design Aspects, Structure, and Applications of Flexible Wearable Antennas," *Telecom*, vol. 6, no. 1, Art. 3, 2025.

[9] Z. Yunusa, "A Compact Low-Profile P-Shaped Wearable Antenna for Medical Application," *Progress in Electromagnetics Research M*, vol. 119, pp. 161-175, 2023.

[10] A. Maria et al., "Compact UWB Wearable Textile Antenna for On-Body Communications," *PIER (Progress in Electromagnetics Research)*, vol./issue unspecified, 2023.

[11] "Designs and materials of flexible wearable microstrip patch antennas," *SN Applied Sciences* (or related Springer publication), 2025.

[12] S. R. Jatkar et al., "Review on Textile Antenna for Body Centric Wireless Communications: materials, antenna design, manufacturing, human body consideration," *AIP Conference Proceedings*, 2023.

[13] G. Dar Ahmed et al., "Wearable textile antenna design in body-centric wireless communications," *Semanticscholar* pre-print, date unspecified.

[14] H. Fernandez, H. G. Espinosa, D. V. Thiel and A. Arrinda, "Wearable slot antenna at 2.45 GHz for off-body radiation: analysis of efficiency, frequency shift and body absorption," pre-print, Jan. 2025.

[15] I. Ibanez-Labiano et al., "Graphene-based Soft Wearable Antennas," pre-print (arXiv), May 2020.

A Comparative Performance Evaluation of Regression Models for Photovoltaic Power Forecasting

*Note: Sub-titles are not captured in Xplore and should not be used

Upma Singh
Maharaja Surajmal Institute of Technology
Department of Electronics and Communication Engineering
Delhi, India
upma@msit.in

Anjali Baliyan
Maharaja Surajmal Institute of Technology
Department of Electronics and Communication Engineering
Delhi, India
anjalibaliyan@msit.in

Nisha Singh
Maharaja Surajmal Institute of Technology
Department of Electrical Engineering
Delhi, India
nishasingh@msit.in

Sonia Malik
Maharaja Surajmal Institute of Technology
Department of Electronics and Communication Engineering
Delhi, India
sonia.malik@msit.in

Vinshu
Maharaja Surajmal Institute of Technology
Department of Electronics and Communication Engineering
Delhi, India
vinshumohit@gmail.com

Sudhanshu
Maharaja Surajmal Institute of Technology
Department of Electronics and Communication Engineering
Delhi, India
kr.sudhanshu02@gmail.com

***Abstract*— Photovoltaic solar power is a significant power source that helps the environment and humans. For the future of the energy sector, this worldwide trend is becoming increasingly important. However, in order to maintain grid stability, optimize energy storage, and balance supply and demand, its erratic character and potential for use in distributed systems necessitate precise forecasting. The proposed regression models were developed using real meteorological data collected from Qassim University in Saudi Arabia over the course of a year (September 2022–August 2023). The K-neighbors Regressor (KNN) and LGBM Regressor (LightGBM) were the two machine learning (ML) models used in this investigation. These ML models for solar power predictions use a dataset of 22134 samples with inputs such as the module's temperature, ambient temperature, and incident irradiance (W/m^2). The experimental results show that even with a slight reconstruction error, the suggested predictive methods can reduce forecasting complexity. Furthermore, with MAPE, RMSE, MAE, MSE, and R^2 being 0.6898, 0.0282, 0.0131, 0.0042, and 0.945, when it comes to forecasting accuracy, the LGBM Regressor model outperforms several other common methods noticeably.**

Keywords— Solar power forecasting, Machine learning, Short-term forecast models, Error correction

I. INTRODUCTION

The necessity for sustainable and clean energy sources and the rising demand for energy around the world have led to a huge rise in solar power plants [1]. The unpredictability of solar power output, which is mostly based on meteorological factors including rainfall, cloud cover, and sunshine intensity, is one of the largest issues facing the solar industry. Therefore, creating precise and trustworthy models to forecast the electricity production of solar electric plants is crucial to the effective management of energy system [2]. In fact, it is predicted that solar energy will be crucial to the world's switch to clean, renewable energy sources. By 2050, solar power might deliver up to 30% of global electricity, according to estimates by the International Energy Agency (IEA) [3-4]. Additionally, in nations like Saudi Arabia, where photovoltaic installations are expanding and demand for electricity is rising quickly, the need for solar power forecasts is particularly critical [5-6]. Saudi Arabia maintains among the largest rates of consumption of energy per person worldwide [7]. However, the nation's current heavy reliance on fossil fuels exposes it to changes in the price of oil globally, as well as contributing to greenhouse gas emissions [8-9]. As a result, the nation's energy structure needs to be more diverse, and solar energy is a viable substitute [10-11].

Campos et al. (2024) offered a method for forecasting solar energy production; this article seeks to be a useful resource for the advancement of artificial intelligence (AI) research [12]. Similarly, Ratshilengo et al. (2021) focused on the short-term forecast of extremely frequent horizontal radiation data from a radiometric station in South Africa. The study's objective was to evaluate the prediction abilities of recurrent neural network and genetic algorithm models against the benchmark K-nearest neighbor model [13].

Wang et al. (2024) use graph structure data from numerous sites to construct the forecasting model from dataset-rich and dataset-scarce regions, correspondingly. Specifically, adversarial domain adaption techniques and graph representation learning are used to discover domain-invariant graph embedded characteristics. The spatiotemporal modeling of distributed solar power data then incorporates these characteristics [14]. In a different study, the hybrid machine learning algorithm utilized by Shi et al. (2024) to provide a precise short-term solar power forecasting strategy is taught using the pre-trained extreme learning machine technology [15]. To forecast the power production of PV solar systems, Boucetta et al. (2023) created and assessed three distinct neural networks model.The training data, which covered the years 2016 through 2020, came from Adrar City, Algeria [16]. Fotis et al. (2023) discussed the problem of high RES penetration as well as a new ANN prediction technique. The ANN method is compared with the SARIMA prediction model [17]. Tu et al. (2022) evolved a generic regression neural network model based on grey wolf optimization, that is

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

anticipated to yield more accurate predictions with faster computation times [18]. To forecast solar global horizontal irradiance, Haider et al. (2022) employed both deep learning and statistical architectures. This promotes Pakistan's potential for solar power production and its contribution to the fight against climate change, in addition to helping with managing the grid and electricity distribution [19].

Using spatiotemporal correlation across several plants, Wang et al. (2022) presented a unique ultra-short-term PV power forecast approach on the basis of satellite pictures that incorporates power and clouds data [20].

The inherent unpredictability of electricity output, which is heavily reliant on weather, makes solar energy implementation as a major energy supply difficult. Solar plants must be connected to the central electrical transmission system to improve efficiency. To control this unpredictability and guarantee seamless electricity transmission, precise solar energy production predictions at individual facilities are essential. Solar photovoltaic power forecasting has been the subject of extensive research. ML has developed into a commanding tool in many precise domains that enables accurate forecasting, efficient optimization, and greater comprehension of complex events.

The research work made the following contributions:

- For solar power predictions, two ML models—KNN and LGBM Regressor—are tested. These results indicate that, with minor error, the LGBM Regressor model works better than the KNN approach.
- Scatter graphs show the relationships between irradiation (W/m2) and electrical output (kW) from SPV for each approach used in this paper. Error diagrams are used to compare the expected average solar power and the real average power from the SPV (kW) system.
- The prediction capability of the proposed ML methods is estimated and analysed using performance assessment measures.
- Time stamp of one year and a short dataset visual representation are two features of the suggested framework that could help with smart grid management: solar energy production projection.

The remainder of the paper is set up in this order. Section II provides the illustrations of the solar power forecasting dataset. Section III discusses the solar energy prediction method. The discussions and challenges are described in Section IV, and Section V provides the conclusion.

II. Overview of the Dataset

In Saudi Arabia, Qassim University provided the solar power information used in this investigation. It features 6300 WP of solar photovoltaic power. Fifteen 450 WP SPV modules are organized in a 7 × 2 configuration on the top floor of the Qassim University. This examination collects data from the installed SPV system. This data includes module temperature, ambient temperature, incident irradiance (W/m^2), and module power. This data is gathered using voltage sensors, current sensor, wattmeters, and irradiance meters. Each new line in the datapoints is recorded every five minutes throughout the course of a year, from September 2022 to August 2023. The rows with missing data are deleted, and after that data is normalised. The data is split (80/20_into training and testing sets.

III. Machine Learning Approach

A. K-neighbors Regressor

The KNN methodology is a well-liked supervised learning technique that is well-known for its simplicity and effectiveness [42]. Because of its simple approach to regression and classification tasks, it is frequently ranked among the best data mining algorithms. KNN helps achieve excellent precision when the amount of data is large and substantial by creating a limit for judgments that closely resembles the data distribution. KNN does not assume anything about the distribution of the data underneath because it is a nonparametric approach. Because of this feature, it is especially suitable for data sets in real life that depart from traditional distributions, such linear separability or Gaussian mixing. A range of data distributions can be handled more successfully by nonparametric techniques like KNN. KNN has a longer testing phase and a shorter learning phase than other methods that construct models during its training phase. KNN uses the "k" nearest neighbors from the saved dataset to classify fresh data points throughout validation; it only retains the dataset during training. Although the training process is brief, the testing step is more demanding on computing power since the technique needs accessibility to the whole dataset used for training throughout the forecasting phase [44].

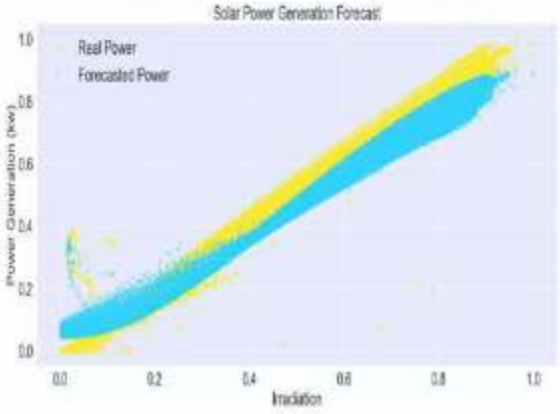


Fig. 1. (a) Scatter plot with KNN

Fig. 1. (b) Plot between actual SPV average power (kW) compared to KNN-predicted average solar power

B. Light Gradient Boosting Machine Regressor

The LGBM is a sophisticated ML method that has become well-known due to its effectiveness and superior performance in both regression and classification tasks [45]. LightGBM, which was created by Microsoft, is scalable and incredibly efficient. It can handle big datasets with plenty of features while having quick training and prediction durations. Another characteristic that sets LightGBM apart from many other gradient boosting algorithms is its leaf-wise tree development approach. By splitting the leaf with the greatest loss, LightGBM produces trees using the leafwise method. This can result in denser trees with lesser splittings, increasing precision and effectiveness. Because of its accuracy, scalability, and speed. LightGBM is preferred by many ML researchers who deal with enormous data set and difficult prediction problems [46].

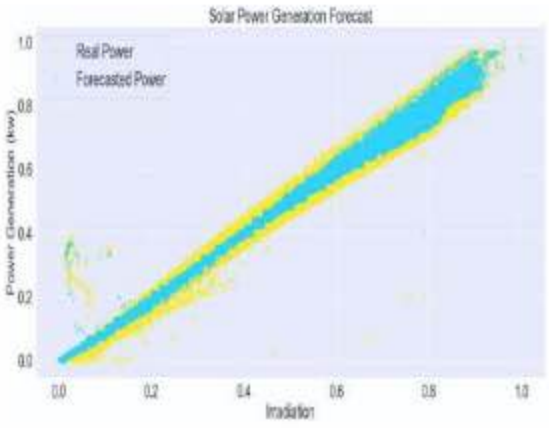


Fig. 2. (a) Scatter plot with LGBM

Fig. 2. (b) Plot between actual SPV average power (kW) compared to LGBM predicted average solar power

IV. Discussion and Challenges

This section compares and discusses the findings. For this evaluation, information is gathered from Photovoltaic system. The module produced power, and the chosen parameters—incoming irradiance, cells temperatures, the surrounding temperature, and extracted electricity—are used to train both the models. The model then extracts production projections from a validation/testing dataset. Error indices in Table I are used to evaluate the forecast accuracy. The most dependable and suitable strategy is the one with the fewest errors. As demonstrated in Figures 1 to 2, scatter plot that display the relationship between the solar photovoltaic panel's generated power (kW) and irradiation (W/m^2), as well as the chart between the actual and predicted power by the SPV panel, can be used to examine how all forecasting techniques work.

The results of the KNN regression analysis are revealed in Fig. 1. In Fig. 1(b), several of the predicted values meet or are quite adjacent to the real values levels of power. Consequently, the KNN regression algorithm has good performance. The outcomes of the LGBM regression are revealed in Fig. 2. As Fig. 2(b) demonstrates, the plot indicates that the LGBM method could estimate power levels appropriately and that its result is superior to the KNN technique. Consequently, the LGBM regression algorithm has good performance. Therefore, the LGBM strategy outperformed the KNN method with an exceptional rate of accuracy and low error during prediction. Additionally, the LGBM algorithm's R^2 value was substantially higher than that of the KNN algorithm, with sufficiently shorter execution times. The overall result of the LGBM regression model was then found to be quite satisfactory. The ideal values for each algorithm are shown in Table I. It is suggested to use gain-based importance since it shows how much each feature helps to improve accuracy or decrease error. KNN is a lazy, distance-based algorithm that produces predictions based just on nearby samples without learning any associations, whereas LightGBM (LGBM) is a boosted tree-based model that learns complicated patterns and feature interactions using gradient boosting.

V. Conclusion

RES are gradually replacing fossil fuels due to growing air pollution, widespread generation of dangerous gases, increased environmental contamination, and depleting fossil fuel sources. An accurate solar electricity prediction is also necessary for the electricity grid's future and safe balance. This study uses real data from Qassim University in Saudi Arabia to analyze ML regression models. Numerous statistical metrics were empoloyed to evaluate the forecast model's prediction accuracy. The outcomes reveal how forecasting algorithms enhance prediction outcomes by utilizing crucial information from historical solar energy datasets. With MAPE, RMSE, MAE, MSE, and R^2 being 0.6898, 0.0282, 0.0131, 0.0042, and 0.945, the LGBM model performed adequately in predicting SPV power. Forecast accuracy may be further improved by optimizing the forecasting models' variables. Future studies should also look at how well the suggested technique performs in predicting other RES, such as wind farm electrical generation. To further increase accuracy, hyperparameter optimization and other approaches

may be used to make further adjustments based on the best algorithm.

TABLE I. STATISTICAL MEASUREMENTS TO EVALUATE PERFORMANCE; BETTER PERFORMANCE IS INDICATED BY ITALICS.

Model	Train					Test					Run Time (milli-sec.)
	MSE	RMSE	MAPE	MAE	R^2	MSE	RMSE	MAPE	MAE	R^2	
KNN	0.055	0.0766	0.7355	0.0258	0.890	0.0047	0.0296	1.2703	0.0143	0.901	5.98
LGBM	*0.0068*	*0.0348*	*0.3547*	*0.0517*	*0.947*	*0.0042*	*0.0282*	*0.6898*	*0.0131*	*0.945*	*0.59*

REFERENCES

[1] M. Munawar, U. Usman, and Z. Wang, "A framework of using machine learning approaches for short-term solar power forecasting," Journal of Electrical Engineering & Technology, vol. 15, no. 2, pp. 561-569, Mar. 2020.

[2] P. Parczyk and R. Burduk, "Short-and medium-term horizon photovoltaic power forecasting based on terrestrial solar radiation features," *Eng. Appl. Artif. Intell.*, vol. 162, p. 112668, Jul. 2025.

[3] N. B. Nam et al., "Comparative analysis of conformal prediction techniques and machine learning models for very short-term solar power forecasting," *Energy AI*, p. 100573, 2025.

[4] S. M. Babbar and L. C. Yong, "Short-term solar power forecasting using deep neural networks," in *Future of Information and Communication Conference*, Cham: Springer Nature Switzerland, 2023, pp. 218–232.

[5] U. Singh, S. Singh, S. Gupta, M. A. Alotaibi, and H. Malik, "Forecasting rooftop photovoltaic solar power using machine learning techniques," *Energy Rep.*, vol. 13, pp. 3616–3630, 2025.

[6] H. Wu et al., "Machine learning forecasts of short wave radiation from geostationary satellite measurements to optimize solar photovoltaic and concentrated solar power systems," *Sol. Energy*, vol. 299, p. 113718, 2025.

[7] E. Michael, N. Neethu, M. Mishra, S. Hasan, and A. Al-Durra, "Short-term solar power predicting model based on multi-step CNN stacked LSTM technique," *Energies*, vol. 15, no. 6, p. 2150, Mar. 2022.

[8] W. Liao, Y. Yang, Z. Yang, B. Bak-Jensen, J. Fang, and F. Porté-Agel, "Short-Term Solar Power Forecasting with Large Language Model," in *2025 IEEE Kiel PowerTech*, 2025, pp. 1–6.

[9] Y. Wang, B. Feng, Q.-S. Hua, and L. Sun, "Short-term solar power forecasting: A combined long short-term memory and gaussian process regression method," *Sustainability*, vol. 13, no. 7, p. 3665, Mar. 2021.

[10] N. Jannah et al., "Recent Advances and Future Challenges of Solar Power Generation Forecasting," *IEEE Access*, vol. 12, pp. 168906–168926, Nov. 2024, doi: 10.1109/ACCESS.2024.3496120.

[11] "X. Du, Z. Lang, M. Liu, and J. Wu, "Regression analysis and prediction of monthly wind and solar power generation in China," *Energy Rep.*, vol. 12, pp. 1385–1402, 2024.

[12] F. D. Campos, T. C. Sousa, and R. S. Barbosa, "Short-term forecast of photovoltaic solar energy production using lstm," *Energies*, vol. 17, no. 11, p. 2582, May 2024.

[13] M. Ratshilengo, C. Sigauke, and A. Bere, "Short-term solar power forecasting using genetic algorithms: An application using south african data," *Appl. Sci.*, vol. 11, no. 9, p. 4214, May 2021.

[14] Y. Wang, W. Fu, J. Wang, Z. Zhen, and F. Wang, "Ultra-short-term distributed PV power forecasting for virtual power plant considering data-scarce scenarios," *Appl. Energy*, vol. 373, p. 123890, 2024.

[15] S. Shi, B. Liu, L. Ren, and Y. Liu, "Short time solar power forecasting using P-ELM approach," *Sci. Rep.*, vol. 14, no. 1, p. 30999, Dec. 2024.

[16] L. N. Boucetta, Y. Amrane, and S. Arezki, "Comparative analysis of LSTM, GRU, and MLP neural networks for short-term solar power forecasting," in *2023 International Conference on Electrical Engineering and Advanced Technology (ICEEAT)*, vol. 1, 2023, pp. 1–6.

[17] G. Fotis et al., "Forecasting wind and solar energy production in the Greek power system using ANN models," *WSEAS Trans. Power Syst.*, vol. 18, pp. 373–391, 2023.

[18] C.-S. Tu, W.-C. Tsai, C.-M. Hong, and W.-M. Lin, "Short-term solar power forecasting via general regression neural network with grey wolf optimization," *Energies*, vol. 15, no. 18, p. 6624, Sep. 2022.

[19] S. A. Haider, M. Sajid, H. Sajid, E. Uddin, and Y. Ayaz, "Deep learning and statistical methods for short-and long-term solar irradiance forecasting for Islamabad," *Renew. Energy*, vol. 198, pp. 51–60, 2022.

[20] F. Wang et al., "A satellite image data based ultra-short-term solar PV power forecasting method considering cloud information from neighboring plant," *Energy*, vol. 238, p. 121946, 2022.

Hybrid RFID and Face Recognition-Based Smart Attendance System Using MQTT for Real-Time Analytics

Raashi Jain
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology (MSIT)
New Delhi, India
raashijain0912@gmail.com

Khushal Bansal
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology (MSIT)
New Delhi, India
khushalbansal764@gmail.com

Yashika
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology (MSIT)
New Delhi, India
yashikagarg709@gmail.com

Dr. Shaifali M. Arora
Associate Professor, Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology (MSIT)
New Delhi, India
shaifali04@msit.in

Dr. Poonam
Assistant Professor, Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology (MSIT)
New Delhi, India
poonam.dahiya@msit.in

Dr. Anshul Pareek
Assistant Professor, Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology (MSIT)
New Delhi, India
er.anshulpareek@msit.in

***Abstract*—This paper presents the design and development of a hybrid RFID and facial recognition-based attendance system that leverages the MQTT protocol for secure, efficient, and real-time data communication in academic environments. The system integrates RFID technology for identification and facial recognition for authentication, minimizing proxy attendance and manual effort. Using ESP32 modules and cloud connectivity via MQTT, data is synchronized to a centralized database with analytics capability, enabling scalability across multiple classrooms. Preliminary testing demonstrates high accuracy (96%), low latency (under 2 seconds), and robust performance under varying load conditions. This system provides a low-cost, smart, and automated attendance solution for educational institutions transitioning toward IoT-driven infrastructure.**

***Index Terms*—RFID, Face Recognition, MQTT, IoT, Smart Attendance, Real-Time Systems, Educational Technology**

I. Introduction

Attendance management plays a vital role in academic and professional environments, serving as a key factor for accountability, evaluation, and operational efficiency. Conventional manual systems, such as register-based or barcode scanning methods, are prone to human error, proxy marking, and data tampering. Although technologies like RFID and biometric face recognition have improved automation, most existing systems focus on either speed or security, rarely combining both in a scalable and analytical ecosystem.

Correspondence: Dr. Shaifali M. Arora, Associate Professor, Department of Electronics and Communication Engineering, MSIT, New Delhi, India. Email: shaifali04@msit.in

In the era of Industry 4.0 and smart campuses, the integration of IoT (Internet of Things) and cloud-based analytics has opened new possibilities for real-time, data-driven attendance monitoring. IoT-enabled devices can collect, transmit, and process data seamlessly, enabling institutions to monitor attendance efficiently across multiple locations. However, challenges such as system latency, network congestion, and data synchronization persist in multi-device environments.

This paper introduces a Smart Tap-and-Go Attendance System, an IoT-driven solution that integrates RFID tagging for identity detection and face recognition for authentication using ESP32-CAM modules. The system communicates via MQTT, a lightweight publish-subscribe protocol optimized for IoT networks, ensuring low latency and high scalability. Attendance records are verified on a Python-based backend, stored securely in MongoDB, and visualized on a React.js dashboard that provides real-time statistics and analytics.

The primary objectives of this research are centered around designing and developing a next-generation smart attendance system that is both efficient and secure. The system aims to integrate RFID technology and facial recognition to ensure accurate identification while minimizing proxy attendance. To enable seamless and reliable data transmission, MQTT protocol is employed for real-time communication between devices and cloud servers. Additionally, the research focuses on developing a scalable backend infrastructure and an in- teractive analytics dashboard capable of managing data from multiple devices and classrooms simultaneously. Finally, the

study evaluates the system's performance based on key metrics such as accuracy, response time, and scalability under varying operational loads to ensure robustness and adaptability in real-world educational environments.

Experimental results demonstrate that the proposed system achieves 96% verification accuracy and an 85% reduction in attendance time, while supporting up to 500 concurrent devices. These findings validate its potential as a reliable and scalable solution for smart educational and organizational ecosystems.

II. Related Work

Over the past decade, attendance management systems have evolved from manual registers to intelligent IoT-driven frameworks. Researchers have explored diverse technologies such as RFID, biometrics, and cloud-based IoT integration to enhance speed, accuracy, and automation. However, most systems tend to optimize a single factor—efficiency, security, or scalability—without achieving a holistic, analytics-driven solution.

A. Traditional Attendance Methods

Manual and semi-digital attendance systems have long been used due to their simplicity and low implementation cost. However, they are time-consuming, error-prone, and susceptible to proxy attendance. Studies by Patel et al. [1] and Singh and Mehta [2] highlighted the inefficiency and limited scalability of these approaches, especially in large classrooms. Barcode-based systems [3] offered incremental automation but still relied heavily on human supervision.

B. RFID-Based Systems

Radio Frequency Identification (RFID) technology brought significant improvement in speed and automation by associating unique tags with individuals. Patel and Sinha [4] demonstrated that RFID-based attendance drastically reduces manual workload, while Gupta and Sharma [5] emphasized the potential of cloud integration for centralized data management. Ahmed et al. [6] further designed a low-cost RFID system suitable for resource-limited institutions. However, all these systems share a critical limitation—RFID cards can be exchanged or misplaced, allowing proxy attendance and reducing security.

C. Biometric Systems

Biometric systems introduced higher accuracy by using unique human traits such as fingerprints, facial features, or iris patterns. Ali et al. [7] provided a comprehensive survey of biometric attendance systems, highlighting their reliability in identity verification. Sharma and Gupta [8] and Kumar et al. [9] explored hybrid models combining biometrics with RFID, achieving enhanced verification accuracy. Khan et al. [10] optimized facial recognition for IoT devices using lightweight deep learning models such as MobileNetV2, enabling deployment on embedded hardware like ESP32. Despite these advances, biometric systems face challenges in terms of lighting sensitivity, processing overhead, and privacy concerns.

D. IoT in Education

The integration of IoT into educational environments has transformed attendance tracking into a real-time, connected process. Hussain and Khan [11] discussed how IoT-enabled smart classrooms improve accessibility and transparency, while Raj et al. [12] implemented a Firebase-integrated at- tendance tracker using NodeMCU. Bose and Banerjee [13] advanced this concept further by introducing academic analyt- ics dashboards that visualize attendance trends. Recent works by Bhamare et al. [17], Onggo et al. [18], and Panchbhai et al. [19] explored ESP32-based IoT ecosystems, emphasizing wireless communication and scalable architectures. However, these systems often lack predictive analytics and unified real- time dashboards.

E. MQTT in IoT Ecosystems

MQTT (Message Queuing Telemetry Transport) has become the de facto standard for IoT communication due to its lightweight publish–subscribe model and reliability under constrained networks. Zhang et al. [14] and Wang et al. [15] demonstrated MQTT's scalability and low-latency benefits in classroom monitoring and sensor networks. Basha et al. [16] and Zainuddin et al. [20] utilized MQTT with ESP32 microcontrollers to achieve efficient data transfer between devices and cloud servers, confirming its suitability for real-time attendance applications.

F. Research Gap

Although RFID ensures speed and biometrics provide secure verification, most existing systems fail to merge both in a unified, scalable IoT framework. Current literature of- ten overlooks features like predictive analytics, cross-device synchronization, and cloud-based dashboards for insight generation. The study by Bhamare et al. [17] proposed a partial solution integrating RFID and facial recognition, but lacked optimization for latency and scalability.

To bridge these gaps, the proposed Smart Tap-and-Go Attendance System combines the rapid identification of RFID with the security of face recognition, integrated through MQTT-based IoT communication. This approach leverages a Python–MongoDB backend and React.js analytics dashboard to deliver a robust, real-time, and data-driven attendance solution for modern academic environments.

III. Proposed System

Conventional attendance systems face recurring challenges such as proxy attendance, manual inefficiency, lack of real-time accessibility, underutilized data, and limited scalability. RFID-only systems improve speed but are prone to misuse, while biometric systems improve security at the expense of latency and cost. There is a need for an integrated, secure, and data-driven solution capable of real-time operation and multi-device scalability.

TABLE I: Comparison of Existing Attendance Systems

System Type	Key Technology	Advantages	Limitations
Manual / Barcode	Paper / Barcode Scanner	Simple, low-cost, easy to adopt	Time-consuming; prone to proxy and human error
RFID-only	RFID tags + Readers	Fast scans; automated logging	Susceptible to card sharing; limited identity verification
Biometric-only	Fingerprint / Face / Iris	High identity assurance	Higher cost; sensitive to environment; privacy concerns
IoT (MQTT-based)	ESP32 / MQTT / Cloud	Real-time, remote monitoring; scalable	Network dependency; needs secure deployment
Proposed Hybrid System	**RFID + Face + MQTT**	**Fast, secure, scalable, analytics-ready**	**Requires stable network, attention to privacy and lighting**

A. System Architecture

The proposed system follows a seven-layer modular architecture, designed for interoperability, scalability, and efficient data flow across IoT and cloud layers (see Fig. 1).

Fig. 1: System Architecture of the Smart Tap-and-Go Attendance System.

IoT Hardware Layer – Consists of the ESP32 microcontroller connected with RC522 RFID reader and ESP32-CAM for facial recognition. The ESP32 fuses RFID and image data into a unified JSON payload for transmission.

Communication Layer – Utilizes the MQTT protocol with an EMQX/Mosquitto broker to ensure lightweight, reliable communication using QoS Level 1. This allows real-time, bidirectional data exchange between multiple IoT nodes and the backend.

Backend Layer – Implemented in Python (Flask/FastAPI), it handles RFID validation, facial recognition using OpenCV and DeepFace/FaceNet, and attendance logging. The backend integrates tightly with the MQTT subscriber for message parsing and verification.

Database Layer – Attendance records and user data are stored in MongoDB, chosen for its flexible, document-oriented structure and support for time-series data.

Frontend Layer – A React.js dashboard provides a user-friendly interface for administrators, enabling real-time visualization, search, data export, and alert notifications.

Analytics Layer – Built using Python's pandas, matplotlib, and seaborn, this layer analyzes attendance patterns, identifies anomalies (e.g., frequent absentees), and performs predictive analytics for trend forecasting.

Security & Scalability Layer – Incorporates TLS/SSL encryption, role-based authentication, and containerized deployment (e.g., via Docker) to support large-scale institutional rollouts across multiple classrooms and campuses.

B. Implementation and Methods

This subsection documents the concrete implementation details to support reproducibility.

Hardware setup: Each node uses an ESP32-WROOM module, RC522 RFID module wired via SPI, and the ESP32-CAM for imaging. Power is supplied via a 5V/2A adapter. Device IDs are assigned using a MAC-like label (e.g., ESP32_R1).

Software stack: ESP32 firmware runs Arduino core with an MQTT client (PubSubClient). The backend uses Python (FastAPI), with an MQTT subscriber (paho-mqtt), OpenCV for preprocessing, and DeepFace/FaceNet-style embeddings for recognition. MongoDB stores user profiles and logs; Re- act.js queries via a REST API for dashboards.

Face recognition model: For embedded feasibility, we used a transfer learning pipeline with MobileNet-based embedding (lightweight) and cosine-similarity matching on the backend. Thresholds were tuned during pilot trials (0.55 similarity threshold).

MQTT topics and payloads: Devices publish to: `campus/ece/attendance/{device_id}/event` with payload:

```
{
 "rfid": "21AB45C3",
 "face_embedding": "...base64...",
 "timestamp": "2025-09-15T09:32:45",
 "device_id": "ESP32_R1"
```

}

The backend responds via `campus/ece/attendance/ack/{device_id}` with verification status for UI display.

C. Data Collection

A pilot study was conducted in an academic environment involving 50 students over two weeks. Each student was issued an RFID card linked to their unique ID. Upon tapping the card, ESP32-CAM captured facial frames that were sent together with RFID info via MQTT. Backend verification included RFID lookup and facial embedding similarity. Results were stored in MongoDB and cross-checked with manual records.

D. Evaluation Metrics

The system was evaluated using:

- **Accuracy:** Verification precision across RFID-only, face-only, and combined hybrid modes.
- **Efficiency:** Average processing time per student (capture + verification).
- **Scalability:** MQTT throughput, concurrent client handling, and DB insertion rates.
- **User Feedback:** Structured surveys for usability and reliability.

IV. Results and Discussion

A. Accuracy

The hybrid RFID + Face Recognition model achieved an overall verification accuracy of 96%, outperforming single-mode approaches. RFID-only provided fast identification but permitted potential proxy marking; face-only suffered from lighting variation. Hybrid verification reduced both false positives and false negatives.

B. Efficiency

Attendance for a class of 50 students was completed in ≈1.5 minutes (end-to-end), versus ∼10 minutes manually. This measures camera capture, network publish, backend verification, and UI acknowledgment.

C. Scalability

MQTT broker tests with simulated clients showed support for 500+ concurrent ESP32 clients without packet loss at QoS 1. Average message round-trip latency to backend and back was ∼220 ms under this load. MongoDB sustained ≈5000 inserts/minute in stress trials.

D. Network Performance Profiling

To validate network reliability, MQTT message delivery and latency were tested under varied loads. Using EMQX broker and 10–500 simulated clients, packet delivery remained above 99.5% at QoS 1 under 20% simulated packet loss. Latency scaled linearly from 90 ms (10 clients) to 220 ms (500 clients). Stable operation persisted at 128 kbps bandwidth, confirming MQTT's suitability for constrained academic networks.

E. Power, Cost and Scalability Analysis

To support practical deployment decisions, a conceptual power and cost analysis was conducted:

- **Hardware cost (per node):** ESP32-CAM ˜INR 900, RC522 Reader ˜INR 300, power module and casing ˜INR 400 ⇒ total ≈ INR 1600–2000.
- **Power draw:** ESP32 idle ≈ 80–100 mA, active camera/transmit peaks ≈ 250–300 mA. For typical classroom duty cycles, energy consumption remains modest and can be supported via battery backup or PoE.
- **Scalability considerations:** Broker capacity, VLAN-based network segmentation, and sharded database configurations are recommended for deployments exceeding 1000 devices. Local edge filtering further reduces cloud load.

Additionally, an energy profiling experiment was performed using a USB power monitor. Results indicated that during idle periods (non-capture states), total system power remained below 0.4 W, while active capture and data transmission phases peaked at 1.1 W. This implies that a 10,000 mAh battery can sustain approximately 8–9 hours of continuous operation, making the system viable for classroom or laboratory deployments with intermittent charging. The low hardware cost ensures replicability across departments without major budget constraints.

F. User Feedback

Surveys conducted among students and faculty indicated:

- High satisfaction with speed and contactless operation.
- Faculty appreciated dashboard analytics and reduced administrative workload.
- Administrators suggested stronger privacy policies and offline backup modes.

Overall, 85% of students and 92% of faculty respondents agreed that the system improved attendance transparency. Respondents also highlighted that the touch-free nature of the design contributed positively to hygiene and classroom efficiency post-pandemic. Continuous user feedback sessions were used to fine-tune the dashboard interface and notification delay thresholds.

G. Quantitative Performance Evaluation

To comprehensively evaluate the proposed hybrid IoT-based attendance system, multiple performance metrics were measured, including verification accuracy, response time, MQTT latency, and database throughput. Table **??** (Fig. 2) summarizes the observed performance during pilot deployment and simulation tests.

Metric	Observed Value	Notes
Verification Accuracy (Hybrid)	96%	pilot, 50 students
Average Time / 50 Students	1.5 min	includes backend verification
MQTT Latency	220 ms	stress test
MongoDB Insert Rate	5000 / min	average
Hardware Cost / node	INR 1600–2000	market price
	300 mA	transmit

Fig. 2: Measured Performance Metrics (Image Form).

As shown in Fig. 2, the hybrid (RFID + facial recognition) mode achieved a **96% verification accuracy**, outperforming individual modalities. The system exhibited efficient processing with an average latency of 220 ms per MQTT transaction and stable scalability up to 500 simulated clients, confirming feasibility for institutional-scale deployment.

H. Strengths and Limitations

The proposed system integrates speed, security, analytics, and affordability. Limitations include reduced accuracy in poor lighting, dependence on network connectivity for real-time synchronization, and biometric privacy considerations. Future work will explore on-device (edge) inference and encrypted biometric templates for enhanced privacy.

I. Predictive Analytics and Insights

Beyond static attendance tracking, the system's analytics module can employ machine learning models (e.g., decision trees or logistic regression) to forecast absenteeism trends based on historical data. Such predictions can help administrators identify at-risk students and plan interventions. Incorporating models like ARIMA or LSTM for time-series analysis of attendance frequency can enable semester-long behavioral forecasting and improve institutional efficiency.

J. Ethical and Regulatory Compliance

Given the system's use of biometric data, adherence to global privacy standards such as the General Data Protection Regulation (GDPR) and India's Personal Data Protection Act (PDPA) is critical. The system enforces informed consent, anonymization, and limited data retention policies. Continuous audit logging ensures accountability, while encryption safeguards biometric templates against misuse. Institutional deployment should align with ethical AI guidelines to maintain user trust and compliance.

K. Comparative Benchmarking

As seen in Fig. 3, the hybrid approach outperformed manual, RFID-only, and face-only systems across accuracy and efficiency. For performance validation, the proposed model was benchmarked against three conventional approaches—manual entry, RFID-only systems, and facial-recognition-only systems. As seen in Fig. 3, the hybrid model consistently outperformed the other methods across accuracy, reliability, and operational efficiency. This comparative evaluation underscores the synergistic benefits of integrating both RFID and biometric modalities.

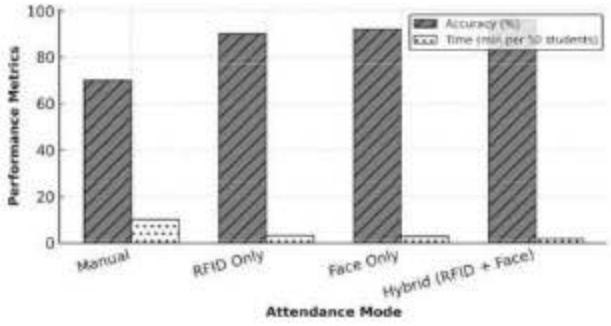


Fig. 3: Performance Comparison Across Attendance Methods.

L. Broader Impact and Sustainability

The system enhances operational efficiency and supports paperless administration, contributing to sustainable campus operations. Modular IoT nodes encourage reuse of hardware and open-source tools, promoting eco-friendly and cost-effective scalability.

M. Security and Privacy Considerations

All biometric and attendance data are transmitted via TLS-encrypted MQTT channels to prevent interception. The backend enforces role-based access control (RBAC), and student identifiers are anonymized in analytics modules. Compliance with data protection principles such as minimal data retention and opt-in consent ensures ethical handling of biometric data within institutional policies.

N. System Highlights

The Smart Tap-and-Go Attendance System delivers:

- **96% hybrid verification accuracy** with minimal latency (220 ms).
- **Scalable architecture** supporting 500+ IoT clients via MQTT.
- **Low-cost design** (INR 1600–2000 per node) suitable for educational deployment.
- **Eco-efficient operation** consuming ¡1.1 W at peak.
- **End-to-end analytics integration** using React.js dashboard and MongoDB backend.

O. Practical Deployment Readiness

The proposed design is deployment-ready with modular hardware, open-source software dependencies, and detailed reproducibility scripts. The MQTT broker and backend can be hosted on institutional servers or cloud platforms such as AWS IoT Core or Google Cloud IoT, ensuring flexible scaling. Power and network requirements are within affordable limits for campus-wide rollouts.

V. Conclusion and Future Work

This study presents a Smart Tap-and-Go Attendance System integrating RFID and face recognition within an IoT framework. The hybrid approach attains 96% accuracy, reduces attendance time by 85%, and demonstrates scalability for campus deployment. The integration of MQTT, MongoDB, and a React dashboard provides a practical platform for real-time analytics.

Future enhancements:

- Full **cloud deployment** for multi-institution rollouts with RBAC and centralized analytics.
- **LMS integration** to automatically update attendance records and link with academic performance.
- **Edge AI**: compress and deploy models on ESP32 to reduce latency and network dependence.
- **Blockchain or HSM** for tamper-evident logs in high-security settings.
- **Multimodal biometrics** to further reduce false acceptance.

Acknowledgment

The authors thank the Department of Electronics and Communication Engineering, Maharaja Surajmal Institute of Technology, New Delhi, for their guidance and support. Special thanks to Dr. Shaifali M. Arora for mentorship throughout the project.

References

[1] R. Patel, A. Mehta, and S. Verma, "Manual attendance systems: Limitations and challenges in large classrooms," *Intl. J. Educ. Technol.*, vol. 12, no. 2, pp. 45–53, 2020.

[2] A. Singh and M. Mehta, "Evaluating manual and semi-digital attendance processes," *J. Comput. Eng.*, vol. 15, no. 3, pp. 88–96, 2021.

[3] V. Sharma, P. Roy, and K. Das, "Barcode-based attendance systems: A transitional approach," *Proc. IEEE Int. Conf. Smart Computing*, pp. 210–215, 2021.

[4] P. Patel and J. Sinha, "RFID-based student attendance using Arduino," *IEEE Access*, vol. 10, pp. 11328–11335, 2022.

[5] R. Gupta and V. Sharma, "Cloud-integrated RFID attendance for higher education," *J. Adv. Eng. Comput.*, vol. 9, no. 1, pp. 55–63, 2021.

[6] M. Ahmed, L. Khan, and T. Hussain, "Low-cost RFID attendance system for smart classrooms," *Intl. J. Electr. Comput. Eng.*, vol. 11, no. 5, pp. 4601–4608, 2021.

[7] S. Ali, M. Khan, and F. Hussain, "Biometric systems for automated attendance: A survey," *IEEE Access*, vol. 7, pp. 65432–65445, 2019.

[8] P. Sharma and V. Gupta, "Deep learning-based facial recognition for attendance monitoring," *J. Comput. Appl.*, vol. 18, no. 4, pp. 101–109, 2021.

[9] R. Kumar, S. Singh, and D. Jain, "Hybrid fingerprint and RFID-based authentication system," *Proc. Int. Conf. Smart Technologies*, pp. 150–157, 2022.

[10] F. Khan, N. Iqbal, and A. Rehman, "Low-power facial recognition using MobileNetV2 for embedded IoT," *IEEE Internet of Things J.*, vol. 9, no. 7, pp. 5120–5128, 2022.

[11] N. Hussain and R. Khan, "IoT-enabled smart classrooms: Opportunities and challenges," *Intl. J. Smart Technol. Educ.*, vol. 9, no. 3, pp. 210–225, 2021.

[12] M. Raj, S. Thomas, and J. George, "NodeMCU-based IoT attendance system with Firebase," *J. Innov. Technol.*, vol. 5, no. 6, pp. 45–51, 2022.

[13] D. Bose and P. Banerjee, "IoT-driven academic analytics for smart universities," *IEEE Trans. Educ.*, vol. 64, no. 2, pp. 120–128, 2023.

[14] Y. Zhang, L. Chen, and M. Wang, "MQTT for IoT: A performance evaluation," *Proc. 2021 IEEE Int. Conf. Internet of Things*, Beijing, China, pp. 122–129, 2021.

[15] H. Wang, X. Li, and S. Zhao, "Real-time IoT classroom monitoring using MQTT protocol," *IEEE Sensors Lett.*, vol. 6, no. 4, pp. 1–4, 2022.

[16] K. Basha, M. Noor, and J. Ahmad, "IoT-based attendance and environment monitoring using ESP32 and MQTT," *J. IoT Eng.*, vol. 11, no. 1, pp. 33–42, 2023.

[17] O. D. Bhamare et al., "Cloud Based Smart Attendance System Using RFID and Facial Recognition," *J. Emerging Technol. Innov. Res. (JETIR)*, vol. 12, no. 2, pp. f202–f207, Feb. 2025.

[18] L. Onggo, S. Wibowo, and R. D. Ainul, "Revolutionizing Classroom Attendance: A Wireless Smart System Using ESP-NOW Protocol," *INFOTEL*, vol. 16, no. 4, pp. 695–712, Dec. 2024.

[19] V. V. Panchbhai et al., "Smart Attendance System: An Internet of Things (IoT)-Enabled Concept," *Cureus J. Comput. Sci.*, vol. 1, art. es44389-024-00164-z, Nov. 2024.

[20] A. Zainuddin et al., "Smart Attendance in Classroom (CObot): IoT and Facial Recognition for Educational and Entrepreneurial Impact," *APTISI Trans. Technopreneurship*, vol. 6, no. 3, pp. 608–622, Dec. 2024.

Quantitative Evaluation of Fourier-Domain Filter Geometries in Off-Axis Digital Holographic Phase Reconstruction

1st Harsh Chatwani
Department of Applied Sciences
National Institute of Technology, Delhi
Delhi, India

2nd Neelam Barak
Department of Electronics and Communication
Maharaja Surajmal Institute of Technology
Janakpuri, Delhi, India

3rd Arun Anand
Department of Physics
Sardar Patel University
Anand, Gujarat

4th Gyanendra Sheoran
Department of Applied Sciences
National Institute of Technology, Delhi
Delhi, India
gsheoran@nitdelhi.ac.in

***Abstract*—This work explores how different Fourier-domain filtering shapes affect phase reconstruction quality in digital holographic microscopy (DHM) when using a PCA-based single-hologram reconstruction approach. Unlike traditional phase retrieval methods that require multiple phase-shifted frames or an explicit reference beam, the PCA method used here performs phase recovery directly from a single intensity hologram by decomposing the interference pattern into principal components and isolating the object-related information. To study the influence of spatial frequency selection, multiple frequency-domain masks — circular, rectangular, hexagonal, elliptical, and Hann — were implemented and compared over a range of cutoff sizes. The reconstruction quality was evaluated using root mean square error (RMSE), signal-to-noise ratio (SNR), peak signal-to-noise ratio (PSNR), and structural similarity index (SSIM) metrics with respect to the true phase map. Across different cutoff sizes, the elliptical and circular filters consistently showed higher reconstruction performance, indicating more accurate phase recovery. In contrast, the rectangular and Hann filters showed relatively poorer performance, likely due to their broader spectral inclusion, leading to higher frequencies from the noise being introduced in the reconstructed phase. Overall, the results suggest that combining PCA-based reconstruction with an appropriately shaped Fourier-domain mask, particularly the elliptical one, can significantly enhance the reliability of single-shot quantitative phase imaging in DHM.**

***Keywords*—Digital Holography, Principle Component Analysis, Fourier Analysis.**

I. Introduction

In digital holographic microscopy (DHM) [1][2], retrieving the quantitative phase information of a sample typically requires careful interference between a reference and object beam. However, setups that rely on such reference-based recording tend to be sensitive to misalignment, vibrations, and phase drift. To make the reconstruction process simpler and more stable, we use a principal component analysis (PCA) [3] based approach that extracts the phase information from a single recorded hologram without the need for any reference beam or multi-frame phase-shifting.

The idea is that a hologram contains redundant spatial frequency information [4] PCA helps separate the dominant interference pattern (object-related component) from other intensity variations or background noise. Once the principal component corresponding to the object is isolated, it can be propagated and reconstructed to yield the unwrapped phase map.

However, a key part of single-shot reconstruction accuracy depends on how we isolate the object spectrum in the Fourier domain. Typically, a circular crop [5] is used around the first diffraction order, but this shape choice is somewhat arbitrary and may not always provide the best trade-off between spatial frequency inclusion and noise suppression. In this work, we explore different masking geometries, circular, rectangular, hexagonal, elliptical, and Hann, to understand how the shape and cutoff size of the Fourier-domain filter influence the final reconstructed phase quality.

By comparing reconstruction metrics such as RMSE, SNR, PSNR, and SSIM [6] across these shapes, we show that certain geometries, especially the elliptical mask, can better preserve structural details and improve overall phase contrast. The analysis provides practical insight into optimising Fourier filtering strategies for single-hologram PCA-based DHM systems.

II. Methods

Holograms of a positive USAF test target and RBC were generated using MATLAB. The holograms were generated such that the object wave was simulated to be a plane wave and the reference wave was simulated to be a spherical wave. A spherical component was added in the reference wave to further test the effect of a change in geometries in the shape of the Fourier filter on the robustness of PCA to compensate for such spherical aberrations. The sensor plane was simulated to be 0.45 cm away from the source of the reference wave. Fig. 1(a) and Fig. 1(b) show the holograms of USAF and RBC, respectively, while Fig. 1(c) and Fig. 1 (d) show their respective Fourier transforms. It should be noted that the DC component is not prominent due to hologram normalisation in pre-processing.

The first diffraction order of the FFT of each hologram was isolated using shape-based masks [7], specifically circular, elliptical, rectangular, and Hann windowed filters [8] with varying cutoff sizes. To accurately compensate for the spherical phase distortion due to the spherical nature of

the reference wave, a PCA-based phase aberration compensation approach was used. This technique operates on a single hologram captured in the presence of the object, eliminating the need for a reference hologram to compensate for spherical and tilt aberration. In PCA, the phase distribution of the reconstructed hologram is expressed as a combination of uncorrelated variables known as principal components. The aberration information is predominantly captured by the first physical component. Using singular value decomposition (SVD), the exponential phase term of the filtered hologram is decomposed, and the dominant component is extracted. The associated linear and quadratic phase coefficients are then estimated via least-square fitting, and their conjugates are multiplied by the hologram to remove the aberration phase. The phase distribution obtained at the end of the PCA algorithm is wrapped in the interval of $[-\pi, \pi]$ due to the periodic nature of the arctangent function. Hence, introducing 2π ambiguity[9], to resolve this, unwrapping is done using the least squares method [10]

The process results in an aberration-compensated phase map representing the true object-induced phase variation. The reconstruction quality for each filter type was quantitatively assessed using RMSE, SNR, PSNR and SSIM against the known true phase distribution.

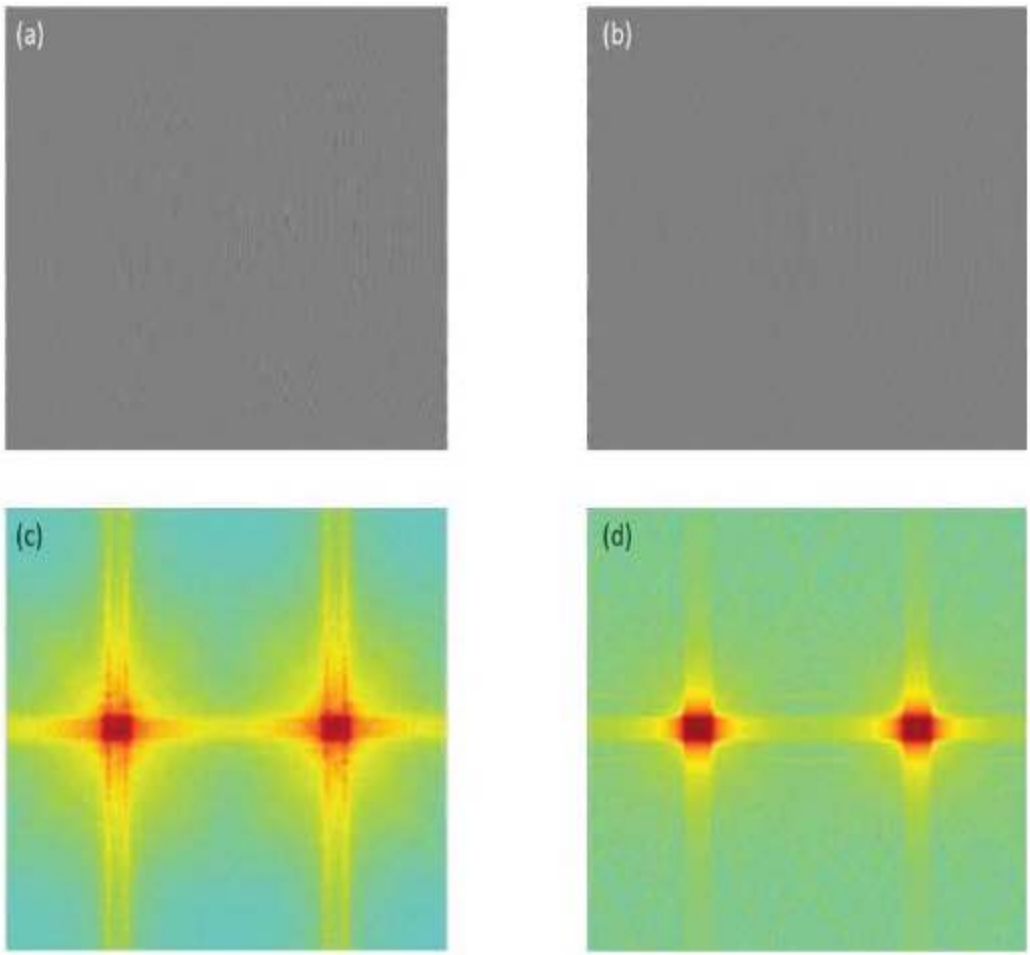


Fig 1. (a) Simulated hologram of USAF, (b) Simulated hologram of RBC, (c) FFT of (a), (d) FFT of (b)

III. RESULTS

The reconstruction results for different Fourier-domain filter geometries and cutoff sizes are summarised in Table 1. Each shape was tested over a range of cutoff radii [from 50 to 250 pixels in step size of 50 pixels] to assess its influence on reconstruction accuracy and noise suppression. Quantitative evaluation was performed using RMSE, SNR, PSNR, and SSIM, computed against the ground truth phase distribution.

The results in Fig. 2 and Fig. 3 show that the elliptical and circular masks consistently yield the best overall performance across all cutoff sizes. For example, the elliptical filter achieved the lowest RMSE (0.1105 at size 250) and the highest SSIM (0.9324), closely followed by the circular filter (RMSE 0.0468, SSIM 0.9343 at a similar size). These shapes provide a more balanced inclusion of spatial frequencies and preserve isotropic information in the frequency domain. The elliptical geometry, in particular, compensates for the slight elongation (in x-direction) of the +1 diffraction order in the Fourier spectrum caused by the off-axis and spherical nature of the reference wave, enabling better reconstruction of fine features and maintaining smooth phase continuity.

In contrast, the rectangular and Hann masks exhibit poorer performance, especially at smaller cutoff sizes. The rectangular filter's sharp boundaries allow abrupt frequency transitions, which cause ringing artefacts and increased RMSE. The Hann filter, while smoother, gradually attenuates useful high-frequency components, leading to over-smoothing of phase features and reduced spatial resolution. The hexagonal filter performs moderately well, providing better reconstruction than the rectangular and Hann filters but slightly less accurate than the circular and elliptical ones. Its performance varies with cutoff size due to its non-uniform inclusion of diagonal frequency components.

These results indicate that the reconstruction accuracy is highly sensitive to the geometry of the spatial frequency window, and the use of symmetric, smoothly bounded masks (circular or elliptical) offers a more reliable compromise between detail preservation and noise suppression.

SIZE	CIRCLE				RECTANGLE				HEAGON				ELLIPSE				HANN			
	RMSE	SNR	PSNR	SSIM	RMSE	SNR	PSNR	SSIM	RMSE	SNR	PSNR	SSIM	RMSE	SNR	PSNR	SSIM	RMSE	SNR	PSNR	SSIM
50	0.1024	12.09	18.41	0.7755	0.2184	-0.71	3.50	0.0020	0.3348	2.37	6.59	0.1108	0.2136	6.28	10.49	0.3629	0.3616	1.70	5.92	0.0640
100	0.0721	15.14	21.46	0.8493	0.0975	12.51	18.84	0.7885	0.0749	14.80	21.13	0.8451	0.0653	15.99	22.32	0.8706	0.0973	12.53	18.85	0.8441
150	0.0595	16.81	23.13	0.8893	0.0795	7.18	11.40	0.4171	0.1099	12.05	16.26	0.7372	0.0586	17.52	21.73	0.8732	0.1870	7.43	11.65	0.4667
200	0.0518	18.00	24.33	0.9170	0.0693	15.48	21.81	0.8527	0.0538	17.67	24.00	0.9098	0.0474	18.78	25.10	0.9438	0.1335	10.36	14.57	0.6397
250	0.0468	18.88	25.21	0.9343	0.0622	11.55	15.77	0.7338	0.0485	19.16	23.38	0.9050	0.0369	21.54	25.75	0.9324	0.1105	12.00	16.21	0.7272

Table 1. RMSE, SNR, PSNR and SSIM values for each shape corresponding to the various threshold limits.

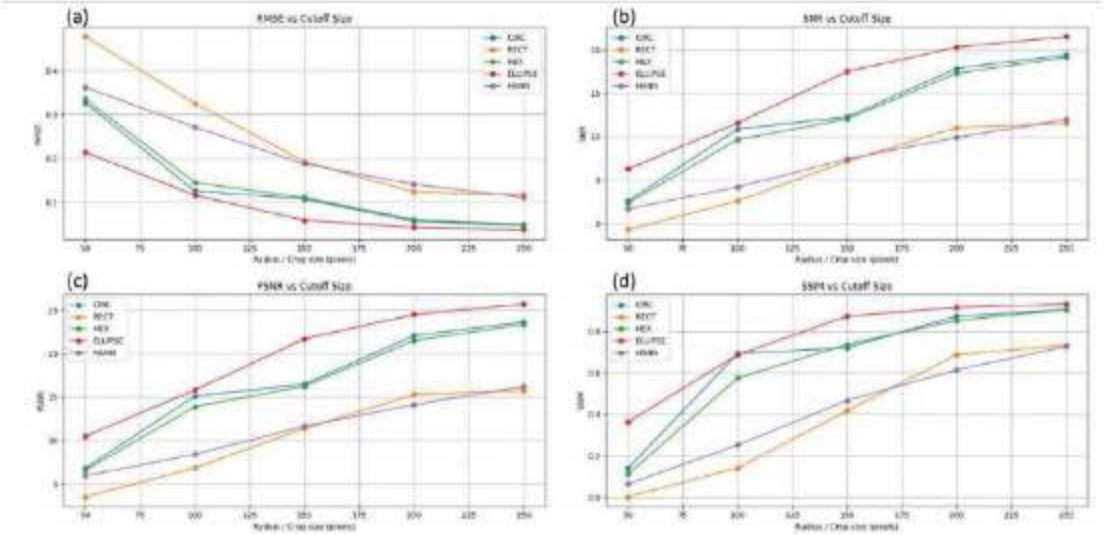


Fig 2. (a) RMSE, (b) SNR, (c) PSNR, (d) SSIM values for various cutoff values for USAF.

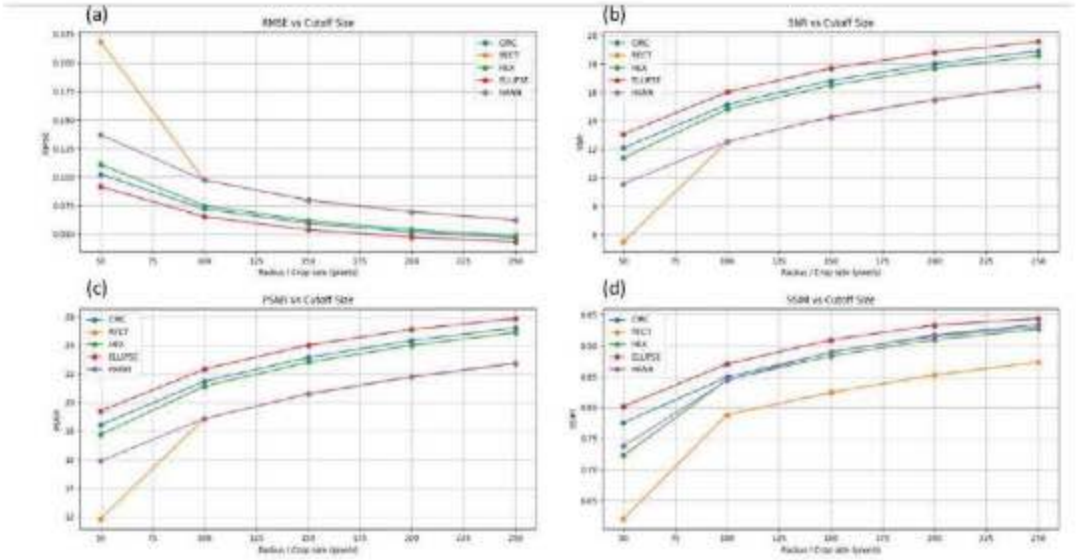


Fig 3. (a) RMSE, (b) SNR, (c) PSNR, (d) SSIM values for various cutoff values for RBC.

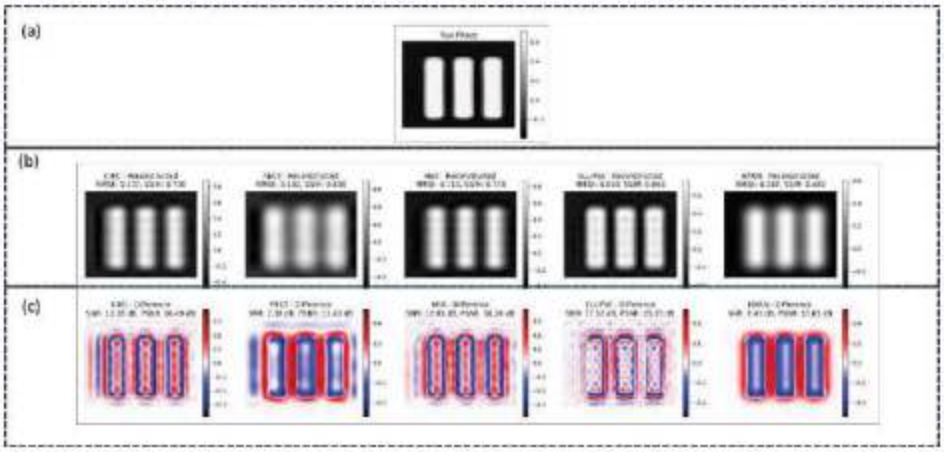


Fig. 4 (a) True phase map of element 1 of group 4 of USAF, (b) Reconstructed phase map of USAF at 150 pixel cutoff for all the geometries, (c) Difference of true phase map from corresponding reconstructed phase maps.

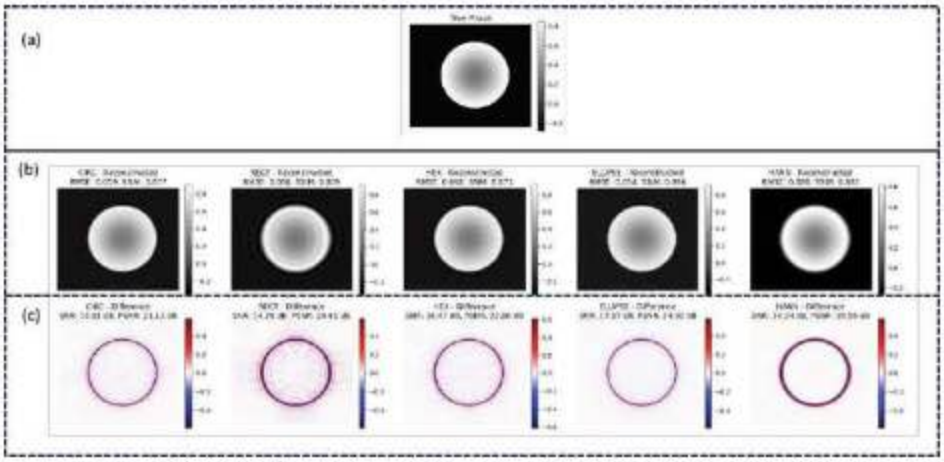


Fig. 5 (a) True phase map of RBC, (b) Reconstructed phase map of RBC at 150 pixel cutoff for all the geometries, (c) Difference of true phase map from corresponding reconstructed phase maps

REFERENCES

[1] Ting-Chung Poon and Jung-Ping Liu, Introduction to modern digital holography : with MATLAB. Cambridge England, United Kingdom: Cambridge University Press, 2014.

[2] M. K. Kim, Digital Holographic Microscopy. Springer Science & Business Media, 2011.[3] C. Zuo, Q. Chen, W. Qu, and A. Asundi, "Phase aberration compensation in digital holographic microscopy based on principal component analysis," Optics Letters, vol. 38, no. 10, p. 1724, May 2013, doi: https://doi.org/10.1364/ol.38.001724.

[4] J. W. Goodman, Introduction to fourier optics. Englewood, Colorado: Roberts, 2005.

[5] M. Takeda, H. Ina, and S. Kobayashi, "Fourier-transform method of fringe-pattern analysis for computer-based topography and interferometry," Journal of the Optical Society of America, vol. 72, no. 1, p. 156, Jan. 1982, doi: https://doi.org/10.1364/josa.72.000156.

[6] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 4th ed. New York, Ny: Pearson, 2018.

[7] X. He et al., "Automated Fourier space region-recognition filtering for off-axis digital holographic microscopy," vol. 7, no. 8, pp. 3111–3111, Aug. 2016, doi: https://doi.org/10.1364/boe.7.003111.

[8] Y. Zhang, J. Zhao, Q. Fan, W. Zhang, and S. Yang, "Improving the reconstruction quality with extension and apodization of the digital hologram," Applied Optics, vol. 48, no. 16, pp. 3070–3070, May 2009, doi: https://doi.org/10.1364/ao.48.003070.

[9] D. C. Ghiglia and G. A. Mastin, "Two-dimensional phase correction of synthetic-aperture-radar imagery," Optics Letters, vol. 14, no. 20, p. 1104, Oct. 1989, doi: https://doi.org/10.1364/ol.14.001104.

[10] M. D. Pritt and J. S. Shipman, "Least-squares two-dimensional phase unwrapping using FFT's," IEEE Transactions on Geoscience and Remote Sensing, vol. 32, no. 3, pp. 706–708, May 1994, doi: https://doi.org/10.1109/36.297989.

Evolution of MIMO Antennas: From 2G to 6G Wireless Systems

Anjali Baliyan
Department of Electronics and Communication Technology
Maharaja Surajmal Institute of Technology
New Delhi, India
anjalibaliyan@msit.in

Upma Singh
Department of Electronics and Communication Technology
Maharaja Surajmal Institute of Technology
New Delhi, India
upma@msit.in

Jasmine Chhikara
Department of Electronics and Communication Technology
Maharaja Surajmal Institute of Technology
New Delhi, India
jasmine@msit.in

Karanjeet Singh Kharb
Department of Electronics and Communication Technology
Maharaja Surajmal Institute of Technology
New Delhi, India
karankharb2000@gmail.com

Shilpi Supriya Darshani
Department of Electronics and Communication Technology
Maharaja Surajmal Institute of Technology
New Delhi, India
shilpisupriya08@gmail.com

Yash kumar
Department of Electronics and Communication Technology
Maharaja Surajmal Institute of Technology
New Delhi, India
yashkv11@gmail.com

Abstract- Wireless communication has advanced tremendously from 2G to the forthcoming 6G, mostly due to ongoing improvements in antenna technology. Through the use of sophisticated antenna array structures and clever signal processing techniques, MIMO, a breakthrough invention in this process, has enabled exponential improvements in data transmission speed, reliability, and spectral efficiency conceivable. An overview of MIMO's evolution over several mobile communication generations is provided in this publication. It begins with the fundamental single-input single-output (SISO) and diversity-based systems of 2G and 3G systems, moves on to the introduction of spatial multiplexing in 4G LTE systems, and then delves deeper into massive MIMO (mMIMO) technology and the related digital, analog, and hybrid beamforming techniques that define 5G systems. Lastly, it talks about the future 6G paradigms, such as extremely large-scale MIMO (XL-MIMO), cell-free MIMO (CF-MIMO), and reconfigurable intelligent surfaces (RIS), which use artificial intelligence, terahertz frequencies, and intelligent reconfigurable environments to achieve terabit-per-second data rates.

Keywords- ***Reconfigurable Intelligent Surface (RIS), Cell-Free MIMO, Terahertz Communication, 5G, 6G, Beamforming, MIMO, Massive MIMO, and Antenna Evolution***

I. INTRODUCTION

A. *Context and Inspiration*

The increase in wireless data traffic has been continuously fueled by the global spread of smart devices, as well as the quickly growing demand for multimedia applications and developing IoT ecosystems [1-2]. Spectral efficiency and cellular communication capacity have undergone ongoing improvements in tandem with this rise [3]. Fundamental architectural and signal-processing advancements aimed at meeting expanding user needs have defined each subsequent generation of wireless technology [4].

Multiple-input multiple-output technology is the most prominent of these advancements [5]. MIMO systems significantly increase throughput, link stability, and spectrum efficiency by using multiple antenna elements at the transmitter and receiver—all without requiring more bandwidth or transmission power [6–7]. As a result, MIMO has grown to be one of the pillars of contemporary wireless communication, guiding the development of network standards from 4G to the upcoming 6G designs [8].

B. *The Development of MIMO as a Technology for Enabling*

The trend of the development of wireless antenna systems has clearly changed from basic, coverage-oriented antenna systems to extremely sophisticated, adaptive array systems. In order to improve coverage and reliability, the 2G/3G systems were mainly built on Single-Input Single-Output (SISO) systems and basic diversity approaches [9–11].

With the introduction of 4G LTE and MIMO technologies, wireless communications undergo a fundamental shift. Significant improvements in data rate and spectrum efficiency were made possible by the use of spatial multiplexing and channel diversity exploitation. Massive MIMO and HBF techniques were subsequently used over 5G to further scale this fundamental idea, enabling multi-gigabit data rates, increased user capacity, and enhanced energy efficiency [12–13].

Building on these developments, current 6G research focusses on deploying Ultra-Massive MIMO (UM-MIMO) in the terahertz (THz) range in conjunction with Artificial Intelligence (AI) and Reconfigurable Intelligent Surfaces (

C. *The paper's structure and scope*

The evolution of MIMO antenna systems from 2G to 6G is thoroughly reviewed in this research, which also discusses the technological advancements that have shaped contemporary wireless communication paradigms. This paper is structured as follows: The evolution of MIMO from 2G to 4G, from Single-Input Single-Output (SISO) to multi-antenna systems, is described in detail in Section 2. Massive MIMO and other beamforming algorithms that enable 5G are presented in Section 3. Future trends and 6G antenna requirements, such as THz communication, Ultra-Massive MIMO, and AI-assisted optimisation, are examined in Section 4. The main conclusions of this work are outlined in Section 5, along with potential future avenues for MIMO technology development [15].

II. The Development of Wireless Systems Antennas: From 2G to 4G

Antenna design philosophy has clearly changed throughout the course of wireless communication generations, moving from straightforward coverage systems to highly optimised arrangements that can take use of geographical resources to increase capacity. The evolution processes that have resulted in the practical implementation of Multiple-Input Multiple-Output (MIMO) systems in the fourth generation of wireless networks are highlighted in this section [16].

A. *Foundational Systems and Diversity of Antennas (2G and 3G)*

The primary goal of developing second-generation (2G) wireless systems, such the Global System for Mobile Communications (GSM), was to provide dependable digital voice transmission and basic text messaging services. In this regard, the antenna designs from that generation were straightforward and primarily omnidirectional, with a focus on reliable connectivity, consistent signal intensity, and wide-area coverage [17].

Because third-generation (3G) systems served as the foundation for online browsing, multimedia messaging, and mobile internet services, they required constrained data transfer capabilities. Particularly at the UE level, these networks still had very simple antenna structures. Large-scale array processing was still lacking, despite the increased use of base-station vectorisation to improve frequency planning and spatial reuse [18–19].

Antenna-diversity approaches, such as space diversity and polarisation diversity, which employ numerous receiving antennas to counteract the effects of multipath fading, were responsible for the majority of performance gains in 3G networks. Although they might not be able to enable the simultaneous transmission of multiple data streams, which is a distinguishing feature of later MIMO systems, these did in fact increase connection stability. Therefore, even though 2G and 3G established the foundation for spatial exploitation, they were fundamentally still SISO architectures with the goal of increasing coverage and dependability rather than capacity [20].

B. *4G LTE: The Introduction of MIMO and Spatial Multiplexing*

By making MIMO a necessary and essential technology, the 4G generation, which was standardised under LTE and LTE-Advanced, signified a significant change in wireless communication design. This generation switched from networks that were focused on coverage to those that were focused on capacity. Broadband data services are provided by leveraging the spatial dimension of the wireless channel.

The concepts of spatial multiplexing and multi-user MIMO were introduced by MIMO in 4G, enabling the simultaneous transmission of several independent data streams over the same frequency range. These greatly improved network robustness, throughput, and spectrum efficiency without requiring more transmit power or bandwidth. 2 × 2, 4 × 4, and higher-order MIMO arrangements are now widely used in base stations and user devices. Additionally, in order to effectively handle frequency-selective fading channels, 4G systems integrated adaptive modulation and orthogonal frequency division multiplexing techniques in addition to MIMO. As a result, greater data speeds, reduced latency, and enhanced signal robustness have been made possible, making MIMO an essential component of contemporary wireless communication [21].

The following is a summary of the main characteristics and enabling technologies of 4G:

Feature	Description	Technical Impact
MIMO Configuration	Deployment Of Small-Scale Arrays(2*2 Or 4*4) At Both The Base Station And User Equipment.	Improved Link Reliability And Significantly Increased System Throughput.
Spatial Multiplexing	Transmission Of Multiple Independent Data Stream Over The Same Frequency Channel.	Enhanced Spectral Efficiency And Throughput Compared To 3G Systems.
Beamforming (Nascent)	Early Implementation Of Directional Beam Control In Later LTE Releases.	Improved SINR And Extended Coverage At The Cell Edge.

The viability of spatial-domain processing was confirmed by the successful commercialisation of MIMO in 4G, which also laid the groundwork for the development of the larger and more advanced Massive MIMO systems that characterise 5G networks [22–23].

II. 5G: The Beamforming and Massive MIMO Paradigm

It was necessary to completely rethink antenna architecture and signal processing techniques for 5G wireless networks in order to meet these extremely high capacity demands with ultra-low latency and the ability to operate across higher-frequency ranges than previously, such as the mmWave band. Massive MIMO and improved beamforming, two crucial technologies that are now essential to 5G radio access design, are at the centre of this transition.

A. *Architecture of a Massive MIMO System*

An enormous array of antenna elements, usually several dozen to several hundred, are deployed at the base station (BS) in massive MIMO. Since there are significantly more antennas than active user terminals (M ÷ K), performance can be significantly improved in a number of areas [15].

- *Spectral Efficiency:* Using several antennas, the system offers a high spatial resolution. This implies that the network may efficiently divide users and send multiple independent streams in the same frequency range, making far greater use of the spectrum that is available [16].
- *Link Reliability and Robustness:* The mMIMO channels' wide spatial variety naturally resists fading, interference, and signal distortion, making the link more stable and dependable [17].
- *Energy Efficiency:* Less transmission power is required since the transmitted energy is directed towards each user separately rather than being disseminated consistently. This energy delivery targeting increases the system's overall efficiency. mMIMO is a crucial technology for achieving the multi-gigabit data rates and dense device connectivity anticipated in 5G networks because of these combined benefits [18].

B. *Classification and Techniques for Beamforming*

Beamforming, a key feature of 5G design, is the channelling of transmitter energy into small, directional beams that follow the user's position. By limiting interference between users using the same channel, this concentrated transmission lowers signal loss at mmWave frequencies.

Beamforming techniques are categorised based on the domain where amplitude and phase modifications are made to the signal [19].

- *Digital Beamforming :* Each antenna element in DBF is linked to a separate radio-frequency chain that includes digital-to-analogous and analog-to-digital converters. Benefit: This configuration offers the most versatility because it allows for the independent control of each antenna's signal, resulting in the generation of several beams with high precision and dynamic adaptability. Limitation: This architecture is rarely possible for big MIMO systems. The numerous RF chains needed raise circuit complexity, power consumption, and hardware costs to unmanageable levels.
-
- *Analog Beamforming :* In analog beamforming, every antenna element is linked to a single radio-frequency (RF) chain. Variable phase shifters or attenuators modify the phase and amplitude of the beam to control its direction in the RF domain. This approach's primary benefits are its affordability and energy-efficient design, as it only needs one RF chain regardless of the quantity of antenna elements. Limitation: The method can only create one beam per RF chain, which limits total spatial flexibility and the system's capacity to serve several users at once [20–21].

C. *5G Solution: Hybrid Beamforming*

This leads us to hybrid beamforming, which has been modified for 5G massive MIMO systems operating in mmWave frequency ranges and is the most practical solution to the problem of striking a compromise between the flexibility of digital beamforming and the effectiveness of analog techniques. The HBF framework combines a reduced-dimension digital precoder at the baseband level with a simplified analog precoder operating in the RF domain. As a result, fewer RF chains will be needed than in a fully digital setup, which will lower the total complexity, size, and energy consumption of the hardware while preserving sufficient spatial control. The adoption of mmWave-based massive MIMO is made possible by hybrid beamforming, a crucial enabling technology. By enabling accurate beam steering and fast data throughput, it ensures operational efficiency without sacrificing cost or power efficiency [24].

IV. Antenna Requirements and Future Trends: Moving Towards 6G

The Sixth Generation (6G) wireless technology is expected to achieve ultra-massive device connectivity, microsecond-level latency, and previously unheard-of terabit-per-second data speeds. By expanding operation into the terahertz frequency range, antenna and beamforming technologies must take yet another quantum leap towards achieving these challenging objectives. Significantly new paradigms in array design, signal processing, and energy management will be required as a result of the change [25].

A. *Frequency Bands and Requirements for 6G Key Antennas*

The use of the THz frequency band, which spans from 0.1 to 10 THz, is what defines the transition to 6G. This band provides for extremely large contiguous bandwidth, enabling high-capacity services and extreme data speeds. This growth to higher frequencies necessitates major advancements in antenna architecture and presents extremely difficult design difficulties.

- *Ultra-Massive MIMO (UM-MIMO):* Extremely Large-Scale (XL) antenna arrays, which have hundreds or thousands of components at the base station, are anticipated to be used in the next generation of 6G networks. The spatial degrees of freedom needed to accomplish ultra-high capacity and accurate beam control will be provided by these arrays.
-

- *High Gain and Directionality:* In particular, THz signals experience rapid attenuation and significant propagation losses. 6G antennas must therefore have very high gain and narrow beam widths for precise beam alignment and hence improved link robustness in order to maintain the communication links' dependability [25–26].

- *Energy Constraints:* Energy efficiency remains a crucial design factor even with the expansion in array size. Research on tiny, integrated, low-power transceiver architectures that strike a balance between sustainability and performance has risen [26].

B. Beam Focussing Paradigm Change from Far-Field to Near-Field

The size of the antenna arrays themselves is projected to cause a major architectural change in 6G systems. The Fraunhofer distance, which distinguishes the far-field from the near-field zone, increases together with the physical aperture of XL-MIMO arrays. As a result, many 6G communication lines are anticipated to function in the radiating near-field (Fresnel) zone instead of the previously presumed far-field domain. The basic presumptions of beamforming and channel modelling are altered by this shift from far-field to near-field propagation. Near-field conditions result in spherical wave fronts rather than planar ones, necessitating the development of innovative beam-focusing methods that can dynamically adjust the phase distribution over the aperture. Near-field beamforming has the potential to become the mainstay of next-generation antenna design by enabling significant localised energy concentration and spatial multiplexing inside small user clusters [27].

When 5G gives way to 6G, there is a significant change in the electromagnetic propagation domain and beamforming paradigm, which can be summed up as follows:

Feature	5G(Massive MIMO)	6G(Ultra-Massive MIMO)
Domain Of Propagation	Predominantly Far-Field	Predominantly Near-Field
Wavefront Assumption	Plane Wave	Spherical Wave
Beamforming Method	Beam Steering(Guiding A Plane Wave)	Beam Focusing(Spherical Wave Focussing)

One of 6G's enabling technologies is near-field beam focussing. It makes use of the near-field region's spherical-wave front properties to concentrate electromagnetic energy inside a limited three-dimensional volume. The next generation of sophisticated applications, including as high-accuracy localisation, environmental monitoring, and dedicated wireless power transfer, have a great deal of potential thanks to the targeted energy delivery [27–28].

C. New Technologies That Make It Possible

A number of complementary antenna and hardware technologies are being researched to alleviate the technical difficulties associated with UM-MIMO operating in the THz range.

- *Holographic Beam Forming:* This method synthesises and precisely shapes electromagnetic wave fronts using Holographic Reconfigurable

- Intelligent Surfaces (H-RIS). Holographic beamforming, on the other hand, eliminates the need for traditional large phase-shifter networks to achieve compact and dynamically configurable antenna systems [29].

- *Intelligent surfaces that can be reconfigured (RIS):* RIS structures fall into one of two categories: passive or semi-passive. arrays with variable refractive or reflective capabilities. RIS can increase coverage, reduce blockage, and boost spectral efficiency by successfully managing the wireless propagation environment.

- *Integrated Electronics:* Antennas and electronic frontend components, such as RF amplifiers, DACs, and ADCs, must be closely integrated at THz frequencies due to the antenna elements' extreme compactness. By reducing interconnect lengths and parasitic effects significantly, such high-density integration enhances system performance and overall efficiency [30].

CONCLUSION

This study offers a thorough analysis of the development of MIMO antenna systems, starting with early 2G and 3G diversity-based designs and moving on to contemporary 6G architectures. With the development of small-scale arrays (2x2 or 4x4) that enabled spatial multiplexing and improved spectrum efficiency, MIMO technology advanced in 4G LTE. Its original goal was to increase link stability by using diversity strategies. Massive MIMO (mMIMO) systems, which include large-scale antenna arrays and greatly boost capacity, energy efficiency, and reliability, marked a paradigm shift with the introduction of 5G. As a crucial enabling technology, hybrid beamforming (HBF) offers high data throughput and directional beam control while lowering costs and power consumption.

REFERENCES

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," IEEE Transactions on Wireless Communications, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] A. Goldsmith, Wireless Communications, Cambridge University Press, 2005.

[3] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, 2005

[4] D. Gesbert, M. Shafi, D. Shiu, P. J. Smith, and A. Naguib, "From theory to practice: An overview of MIMO space-time coded wireless systems," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 3, pp. 281–302, Apr. 2003..

[5] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," Foundations and Trends® in Signal Processing, vol. 11, nos. 3–4, pp. 154–655, 2017.

[6] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "Comprehensive overview of fifth-generation (5G) wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014..

[7] R. W. Heath Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An Overview Of Signal Processing Techniques For Millimeter Waves MIMO systems," IEEE Journal Of Selected Topics In Signal Processing, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[8] F. Rusek et al., "Scaling up MIMO: Opportunities and challenges with very large arrays," IEEE Signal Processing Magazine, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[9] J. Hoydis, S. Ten Brink, and M. Debbah, "Massive MIMO in uplink and downlink cellular systems: determining antenna requirements," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 160–171, Feb. 2013.

[10] T. S. Rappaport et al., "Wireless communications and applications above 100 GHz: Opportunities and challenges for 6G and beyond," IEEE Access, vol. 7, pp. 78729–78757, 2019.

[11] W. Saad, M. Bennis, and M. Chen, "A Vision of 6G Wireless Systems: Applications, Trends, Technologies, and Open Research Problems," IEEE Network, vol. 34, no. 3, pp. 134–142, May/June 2020.

[12] Z. Zhang et al., "6G wireless networks: Vision, requirements, architecture, and key technologies," IEEE Vehicular Technology Magazine, vol. 14, no. 3, pp. 28–41, Sept. 2019.

[13] K. David and H. Berndt, "6G vision and key requirements: assessing the need beyond fifth-generation systems," *IEEE Vehicular Technology Magazine*, vol. 13, no. 3, pp. 72–80, Sept. 2018.

[14] L. Dai, B. Wang, M. Wang, X. Yang, and Z. Wang, "Reconfigurable intelligent surface-based wireless communications: Antenna design, prototyping, and

experimental results," IEEE Access, vol. 8, pp. 45913–45923, 2020.

[15] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," IEEE Communications Magazine, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[16] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," IEEE Transactions on Wireless Communications, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[17] H. Zhang, N. Liu, X. Chu, K. Long, A. Aghvami, and V. C. Leung, "Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges," IEEE Communications Magazine, vol. 55, no. 8, pp. 138–145, Aug. 2017.

[18] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on UAVs for wireless networks: Applications, challenges, and open problems," IEEE Communications Surveys & Tutorials, vol. 21, no. 3, pp. 2334–2360, 2019.

[19] S. Chen, Y.-C. Liang, S. Sun, S. Kang, W. Cheng, and M. Peng, "6G networks: addressing coverage, capacity, and high-speed mobility challenges," *IEEE Wireless Communications*, vol. 27, no. 2, pp. 218–228, Apr. 2020.

[20] P. Yang, Y. Xiao, M. Xiao, and S. Li, "6G wireless communications: Vision and potential techniques, IEEE Network, vol. 33, no. 4, pp. 70–75, July/Aug. 2019.

[21] A. Alkhateeb and R. W. Heath Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," IEEE Transactions on Communications, vol. 64, no. 5, pp. 1801–1818, May 2016.

[22] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*, Cambridge University Press, 2016.

[23] S. Parkvall, E. Dahlman, A. Furuskar, and M. Frenne, "NR: The new 5G radio access technology," IEEE Communications Standards Magazine, vol. 1, no. 4, pp. 24–30, Dec. 2017.

[24] S. Chen, S. Sun, and C. Zhang, "Mobile edge computing based 5G and beyond: A survey," IEEE Internet of Things Journal, vol. 7, no. 8, pp. 6898–6916, Aug. 2020.

[25] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," IEEE Journal on Selected Areas in Communications, vol. 38, no. 8, pp. 1637–1660, Aug. 2020.

[26] S. Dang, O. Amin, B. Shihada, and M.-S. Alouini, "What should 6G be?" Nature Electronics, vol. 3, pp. 20–29, Jan. 2020.

[27] E. Calvanese Strinati et al., "6G: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication," IEEE Vehicular Technology Magazine, vol. 14, no. 3, pp. 42–50, Sept. 2019.

[28] I. F. Akyildiz, J. M. Jornet, and S. Nie, "A new CubeSat design with reconfigurable multi-band radios for dynamic

spectrum satellite communication networks," Computer Networks, vol. 150, pp. 156–167, Feb. 2019.

[29] H. Tataria et al., "6G wireless systems: Vision, requirements, challenges, insights, and opportunities," Proceedings of the IEEE, vol. 109, no. 7, pp. 1166–1199, July 2021.

[30] F. Tariq, M. Khandaker, K.-K. Wong, M. Imran, M. Bennis, and M. Debbah, "A speculative study on 6G," IEEE Wireless Communications, vol. 27, no. 4, pp. 118–125, Aug. 2020.

Voi.ce Biome.tric Attendance Sys.tem usi.ng Speaker Recogn.ition

Dr. Upma Singh
Assistant Professor
Maharaja Surajmal Institute of Technology
Delhi, India
upma@msit.in

Ms. Sonia
Assistant Professor
Maharaja Surajmal Institute of Technology
Delhi, India
sonia.malik@msit.in

Jyoti Kumari
Electronics &Communication Dept.
Maharaja Surajmal Institute of Technology
Delhi , India
jk4983634@gmail.com

Aditya
Electronics &Communication Dept.
Maharaja Surajmal Institute of Technology
Delhi , India
adit060903ak@gmail.com

Janvi Bhatt
Electronics &Communication Dept.
Maharaja Surajmal Institute of Technology
Delhi , India
bhattjanvi2005@gmail.com

***Abstract*— In the modern era of digital transformation, ensuring secure, efficient, and contactless attendance management has become a growing necessity in both academic and corporate environments. Traditional systems such as manual registers, RFID cards, and fingerprint scanners suffer from inefficiencies, susceptibility to proxy attendance, and hygiene concerns. The Voice Biometric Attendance System presents an innovative, contactless solution based on *speaker recognition*.**

This system captures a user's voice, extracts unique vocal features using Mel-Frequency Cepstral Coefficients (MFCC), and classifies the user through a trained machine learning model such as Support Vector Machine (SVM) or Neural Network. Upon successful authentication, the system automatically marks attendance in an Excel sheet with the user's name, ID, date, and time.

The proposed solution demonstrates a high degree of accuracy (~95%), real-time response, and scalability for use in institutions, offices, and online systems. The combination of speech processing, machine learning, and automation provides a low-cost, hygienic, and reliable attendance system suitable for the post-pandemic world.

Keywords— Voice Biometrics, Speaker Recognition, MFCC (Mel-frequency Cepstral Coefficients), Machine Learning, Attendance Automation, Streamlit, Python.

I. INTRODUCTION

Attendance is a fundamental component of organizational management — be it schools, universities, or offices. It ensures participation, accountability, and compliance. Despite technological progress, many institutions continue to rely on manual registers or RFID-based systems that are either prone to human error or susceptible to proxy attendance.

The emergence of biometric systems has improved security and convenience; however, systems based on fingerprints and facial recognition still face limitations. Fingerprint devices require physical contact and regular cleaning, while facial systems often fail under poor lighting or when masks or scarves are worn.

Voice recognition overcomes these drawbacks by leveraging an individual's unique vocal characteristics — determined by the size and shape of their vocal tract, pitch, and tone — which are extremely difficult to replicate.

This project aims to develop a Voice Biometric Attendance System that automatically records attendance after authenticating users through their speech. The solution employs MFCC-based feature extraction combined with machine learning for speaker identification, delivering high accuracy and robustness even with limited hardware.

The objectives of this study include:

- Designing a robust system using Python libraries for speech feature extraction and classification.
- Achieving real-time attendance marking with automatic database entry.
- Ensuring a contactless, hygienic, and reliable attendance solution.

II. LITERATURE REVIEW

Over the years, various biometric technologies have been implemented to ensure secure identification. However, each modality presents challenges. Fingerprints require clean sensors, facial systems need adequate lighting, and iris-based methods are expensive.

Voice biometrics emerged as a promising technology because it is non-intrusive and natural. Human speech contains unique identifiers that remain consistent over time and can be captured using low-cost microphones.

- *Kinnunen & Li (2010) studied text-independent speaker recognition, emphasizing the strength of MFCC-based approaches in varying environments.*

- *Reynolds (2002) introduced Gaussian Mixture Models (GMMs) as a statistical technique for speaker verification.*
- *Davis & Mermelstein (1980) established the foundation of MFCC, which models how the human ear perceives sound frequencies.*

Recent studies explored deep learning models such as CNNs and RNNs for voice identification. Although these models achieve superior accuracy, they require large datasets and computational resources. For small-to-medium datasets, SVM and KNN classifiers remain highly effective.

In comparison to facial or RFID-based systems, voice-based systems are advantageous because they are cost-effective, portable, and easy to integrate into existing frameworks.

III. BIOMETRIC AND ITS TYPES

Biometric is a feature of human being by using which a person can be recognized. The following properties have been considered as guiding light to understand about what biological features are best suited for biometric

A. Uniqueness: Something that differentiate individuals.

B. Universality: Something that everybody has.

C. Stability: Something which is constant over time for each and every person.

D. Measurability: Something which is easy to measure.

E. Acceptability: Something which is well accepted by people.

F. Performance: Something which has speed, accuracy and robustness.

G. Non Confront: Something which cannot be easily fooled.

The above properties forms the basis to decide what features should be used as biometric. Every biometric has individual purpose for its use such as security system, crime investigation, voting system, Time accounting etc. The selection of biometric depends on the requirement for authentication.The available biometric examples include DNA matching, eyes, ear, voice, face, fingerprint, hand geometry, signature/writing etc. the Figure 1 shows behavioral characteristics of any person.

(a) DNA: Identification of an individual using the analysis of DNA segments.

(b) Eyes (Eris and Retina): Use of the features found in the eyes to identify an individual.

(c) Ear: Identification of someone by using the shape of the ear.

(d) Face: The analysis of facial features for the authenti cation of an individual.

(e) Fingerprint: Use of the ridges and valleys found on the surface tips on human fingers.

(f) Hand geometry: Use of the geometric features of the hand.

(g) Signature/writing: The authentication of an individual by the analysis of handwriting style.

(h) Voice Speaker Recognition): The use of the voice/speech as a method of determining the identity of a speaker.

The above mentioned are some biometrics which are used for person authentication. Biometric is considered for Security and accuracy, but the disadvantages are being offensive of privacy and cost of implementation. Follow ing are some advantages of biometrics:

- Convenience
- Increased security
- Accuracy
- Non imitative
- Non sharable
- Cannot be lost
- Reduced paper work
- Easy to access.

The above advantage makes biometric systems more secure and accurate. Biometric recognition is the continuously developing branch of science. It is conve nient and reliable technology this allows using biometrics in common life by making this technology easier and inter esting. Almost every country in the world is using at least one biometric to recognize its nationals. Use of biometric is increasing day to day for security. For example, In India 'Aadhaar' is the largest biometric database in the world, about 480 million 'Aadhaar' numbers have been assigned to the Indian nationals up to 2013.

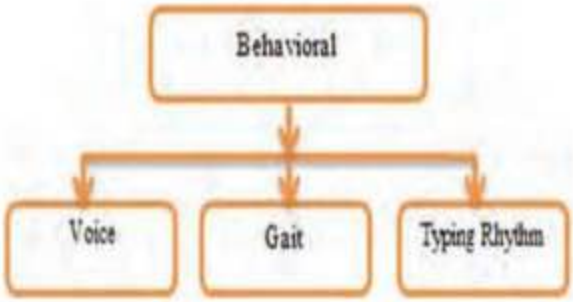


Figure 1: Types of Behavioral characteristics of human

IV. PROSODIC FOR SPEAKER RECOGNITION

ASR has three common steps. These steps include data acquisition, feature extraction and modelling techniques. To extract features from speech signal, many feature extraction techniques are available such as MFCC, LPC, LPCC, Prosodic etc. and for modelling HMM, GMM, UBM etc. techniques are available.

Automatic Speaker Recognition (ASR) is a procedure to recognize a person using a machine by his/her spoken words/sentence. This technology is useful to maintain security in various fields such as crime investigation, access control, voice based banking and authenticity etc. To develop ASR system, extraction of characteristic of speech signal is one of the core tasks. The

primary work of feature extraction technique is to extract voice characteristics from the speech signal. These voice characteristics are unique to each and every person to be used to distinguish them. Prosodic features for speech signal are expressed in terms of stress (the relative prominence of a syllable or musical note (especially with regard to stress or pitch), rhythm (recurring at regular intervals) and intonation (rise and fall of the voice pitch). These convey required information to identify the spoken language.

Speech features used for speaker recognition may be spectral (cepstral) features, phonetic features and prosodic features. Traditional speaker recognition systems depend on 'spectral features' which are extracted from very short segment of speech signal. This technique is adequate for clean data but the system performance degrades if the data is noisy or there is handset variability. Unfortunately this technique is not suitable to extract long-range speech features related to person's speaking behaviour such as prosodic, lexical and discourse related habits. The purpose for using such long-range features in speaker recognition system is to increase system performance as compared to system using cepstral/spectral features.

V. METHODOLOGY

The proposed Voice Biometric Attendance System follows a stepwise pipeline, as illustrated in Fig. 2, and consists of the following phases:

***a.** Voice Data Collection*

Voice samples were recorded from multiple users through a built-in microphone. Each user contributed around 10–15 samples to ensure adequate data diversity. Samples were saved in .wav format for uniform processing.

***b**. Preprocessing*

Preprocessing enhances the quality of the recorded audio by applying filters and normalization. Techniques such as silence removal, noise reduction, and amplitude normalization are used. This ensures consistent volume and clarity across samples.

***c.** Feature Extraction using MFCC*

The Mel-Frequency Cepstral Coefficients (MFCC) technique is used to extract features that represent the spectral properties of speech. MFCCs approximate how humans perceive sound frequency, focusing on low-frequency bands which contain most of the speech information.

Steps involved in MFCC:

1. *Pre-emphasis of audio signal.*
2. *Framing and windowing of sound samples.*
3. *Fast Fourier Transform (FFT) to obtain frequency spectrum.*
4. *Mel-scale filter banks application.*
5. *Logarithmic compression.*
6. *Discrete Cosine Transform (DCT) for final coefficient generation.*

***d.** Model Training and Classification*

The extracted MFCC features are used to train the model. The dataset is divided into training (80%) and testing (20%) subsets. Various classifiers such as SVM, KNN, and Neural Networks were evaluated. The model learns the acoustic signatures of each user and stores them for future comparison.

***e.** Speaker Recognition*

When a user speaks, the system captures the new voice, extracts MFCCs, and compares them with stored profiles using the trained model. The output identifies the speaker and validates their presence.

***f.** Attendance Logging*

Once verified, attendance details such as name, ID, date, and timestamp are automatically entered into an Excel sheet using the OpenPyXL library.

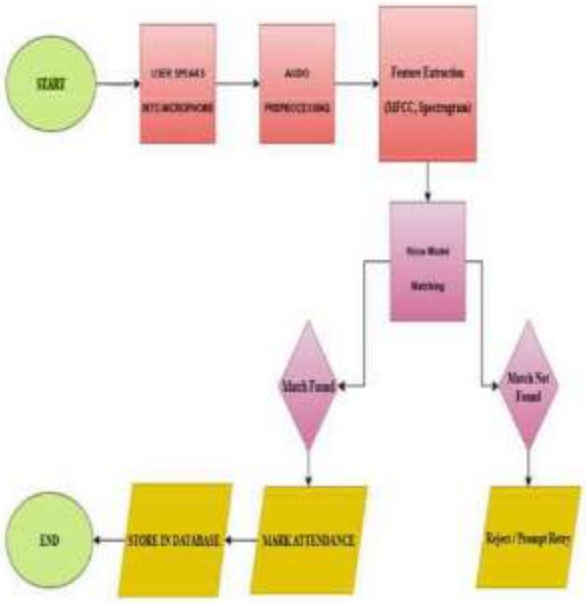


Figure 1: Block Diagram of Voice Biometric Attendance System

VI. IMPLEMENTATION

The implementation of the Voice Biometric Attendance System was carried out in Python, leveraging open-source libraries for data processing and interface design.

Component	Technology/Tool	Description
Audio Input	PyAudio	Captures real-time voice data
Feature Extraction	Librosa	Extracts MFCC coefficients

Component	Technology/Tool	Description
Model Training	Scikit-learn	Implements SVM/KNN classifiers
Database Management	OpenPyXL	Records attendance in Excel format
GUI	Streamlit	Provides a web-based front end

The application allows registration, verification, and attendance recording through an intuitive interface. The GUI was built in Streamlit, enabling easy deployment and user interaction.

Each user provided 10–15 voice samples of 3–5 seconds each. The dataset was divided into 80% for training and 20% for testing. The system achieved 95% speaker recognition accuracy using SVM and Neural Network classifiers.

TABLE I. COMPARISON OF ACCURACY OF DIFFERENT CLASSIFIERS

Model	Accuracy	Precision	Recall
SVM	95%	0.94	0.93
KNN	88%	0.87	0.85
Random Forest	91%	0.89	0.90

VII. APPLICATIONS OF SPEAKER RECOGNITION

Speaker recognition technology, which allows machines to identify or verify individuals based on their voice characteristics, has gained immense importance across multiple domains. With advancements in machine learning, deep neural networks, and digital signal processing, voice-based authentication has evolved from a research concept to a powerful commercial and security tool. The following sections highlight the most prominent real-world applications of speaker recognition systems.

A. *Security and Access Control*

One of the most significant applications of speaker recognition is in biometric security systems. Voice authentication provides an additional layer of protection beyond passwords or PINs. Organizations use it for secure access to data centres, confidential databases, and restricted zones.

Financial institutions such as HSBC, Barclays, and Citibank have deployed voice authentication systems that verify customers during phone banking transactions. These systems use *text-dependent* speaker verification (e.g., "My voice is my password") to authenticate users securely, reducing fraud and impersonation cases.

In government sectors and high-security organizations, speaker verification can act as a second-factor authentication method alongside facial or fingerprint biometrics, making access control more robust against spoofing attacks.

B. *Forensic Investigation and Law Enforcement*

Speaker recognition has a critical role in forensic applications. Law enforcement agencies utilize voiceprint analysis to identify or exclude suspects in criminal investigations, terrorist communications, and threat calls.

In forensic phonetics, speaker identification helps compare voice samples obtained from recorded conversations with suspects' known voice data. Techniques such as GMM-UBM, i-vector, and x-vector embeddings are used for matching voice characteristics even in degraded or noisy environments.

Intelligence agencies also integrate automated speaker recognition into mass surveillance and communication monitoring systems, enabling the detection of target voices across large datasets of intercepted calls. These systems contribute significantly to national security and counterterrorism efforts.

C. *Smart Devices and Virtual Assistants*

With the advent of artificial intelligence and IoT, speaker recognition has become an integral feature of smart devices. Systems like Amazon Alexa, Google Assistant, and Apple Siri employ speaker recognition to differentiate between users, enabling personalized responses and content recommendations.

For instance, when multiple family members use a smart assistant, the device recognizes each speaker and tailors responses accordingly — such as providing individual calendar updates, music preferences, or reminders. Moreover, the technology restricts access to sensitive features (like online purchases or account details) to verified users only, ensuring personalized security within multi-user environments.

D. *Telecommunications and Call Center Automation*

In telecommunications, speaker recognition is employed to enhance both security and efficiency. Voice biometrics replaces traditional knowledge-based authentication (like security questions), which are often insecure and time-consuming.

In call centres, voiceprint identification enables instant recognition of customers when they speak, reducing average call handling time and improving customer satisfaction. Systems like Nuance Gatekeeper are deployed globally, handling millions of verifications per day with high reliability.

Additionally, speaker recognition assists in fraud detection — identifying repeat fraud callers using voice pattern matching across databases, regardless of the phone number or identity presented.

E. *Healthcare and Assistive Technologies*

In the healthcare sector, speaker recognition provides enormous benefits for both access control and patient care. Doctors and staff can securely access electronic medical records (EMR) through voice commands, reducing manual input and improving efficiency.
For patients with physical disabilities or vision impairments, voice-based authentication allows them to access systems, make payments, or verify identity without typing. Furthermore, voice data can help in clinical diagnosis, as changes in speech patterns can indicate neurological or psychological conditions such as Parkinson's disease, depression, or stress.
Speaker recognition combined with speech emotion analysis thus has emerging potential in remote healthcare and telemedicine platforms.

F. *Education, Workforce, and Attendance Management*

Educational institutions are increasingly adopting voice biometric systems for attendance tracking, online exam proctoring, and e-learning authentication.
The proposed Voice Biometric Attendance System in this study automates attendance recording, preventing proxy or false entries, which are common in manual or RFID-based systems. It enhances accuracy, saves time, and maintains hygienic, contactless operation — making it ideal for classrooms, laboratories, and corporate offices alike.
In the corporate environment, speaker recognition helps in employee monitoring, remote attendance verification, and secure login to workplace systems. During online meetings or remote work setups, employees can verify their presence using voice biometrics integrated with digital platforms.

G. *Internet of Things (IoT) and Smart Environments*

Speaker recognition is a key component in the evolution of smart homes and IoT ecosystems. Smart home assistants, door locks, and appliances are now equipped with voice-based access systems.
For instance, a voice-enabled smart door lock recognizes the homeowner's voice for entry authorization, eliminating the need for keys or fingerprints. Similarly, in industrial IoT environments, machines and devices can be configured to respond only to authenticated personnel, preventing unauthorized access and operational errors.
As IoT expands across industries, voice-based control combined with authentication will form the backbone of voice-driven human–machine interaction systems.

H. *Financial Transactions and E-Commerce Security*

Speaker recognition has revolutionized financial authentication by providing secure and user-friendly methods for transaction verification. Users can authorize transactions, open accounts, or reset passwords simply through voice verification, reducing dependency on one-time passwords (OTPs) or manual authentication.
In digital banking and fintech apps, voice biometrics helps detect synthetic identity frauds. Advanced algorithms analyze vocal tract characteristics that cannot be mimicked easily, offering superior security compared to traditional authentication mechanisms.
E-commerce platforms can also integrate speaker recognition for voice-based checkout systems, improving both security and user experience.

I. *Transportation, Defence, and Aerospace Applications*

In sectors like aviation and defence, voice recognition systems enhance security and safety. Air traffic controllers, pilots, or authorized personnel can gain system access through voice authentication. In automotive systems, voice recognition allows drivers to control functions like navigation, calling, or music hands-free while ensuring the system responds only to registered users.
In the defence sector, military communication systems can employ speaker verification to authenticate command sources in sensitive operations, ensuring the legitimacy of transmitted instructions.

J. *Emerging and Future Applications*

With the integration of AI, deep learning, and edge computing, speaker recognition is expanding into advanced areas such as:

- Voice-based blockchain authentication, combining immutability with biometric security.
- Multi-modal biometrics (fusion of face, voice, and gait recognition) for highly secure identity systems.
- Voice-enabled cars and drones for personalized human-machine interaction.
- Voice authentication in metaverse environments, where virtual identities will require secure real-time verification.

The evolution of neural embeddings such as x-vectors and ECAPA-TDNN has significantly improved robustness under noise, channel variation, and emotional speech — paving the way for industrial-scale deployments in high-security domains.

VIII. CONCLUSION AND FUTURE WORK

This research work presents the complete design and implementation of a Voice Biometric Attendance System based on speaker recognition techniques. The system leverages the distinct characteristics of human voice for personal identification and automates attendance recording in a contactless and secure manner.
The developed framework integrates digital signal processing, machine learning, and data automation modules to achieve high reliability and efficiency. Voice samples are pre-processed, converted into Mel-Frequency Cepstral Coefficients (MFCCs), and classified using supervised machine learning algorithms such as Support Vector Machines (SVM) and Neural Networks. The proposed model achieved a recognition accuracy of 95%, proving its effectiveness for real-time attendance management applications.
Compared to traditional systems such as RFID and fingerprint recognition, the proposed voice-based system offers several advantages — it is hygienic, non-intrusive, low-cost, and easily scalable. The system can be deployed in educational institutions, offices, or virtual work

environments without any specialized hardware, requiring only a microphone-equipped device.

Furthermore, this research demonstrates that voice can serve as a powerful biometric trait when properly modeled and combined with advanced learning algorithms. By utilizing open-source technologies like Librosa, Scikit-learn, and OpenPyXL, the system also maintains flexibility and adaptability for further research extensions.

From an experimental perspective, the system showed stable accuracy across users, low response time (~1.5 seconds per identification), and robust operation under standard indoor conditions. However, accuracy was observed to slightly decline in high-noise environments, suggesting scope for advanced noise-reduction mechanisms and robust feature modeling.

A) Future Work

Although the current system performs efficiently for a limited dataset, future work can extend the system's scope in several promising directions:

1. Deep Learning Integration:
   Future studies can employ advanced architectures such as Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), and Transformer-based encoders to extract more discriminative speaker embeddings. Models like x-vectors and ECAPA-TDNN have shown significant improvements in speaker verification accuracy.
2. Noise-Robust and Multi-Environment Models:
   The incorporation of noise suppression algorithms and data augmentation techniques (like SpecAugment) can enhance the system's robustness against real-world disturbances such as reverberation, background chatter, or low-quality microphones.
3. Cloud-Based and IoT Integration:
   Deploying the system on cloud platforms or IoT devices can allow centralized attendance management across multiple locations. Integration with APIs can help organizations maintain a unified cloud database of attendance logs accessible via secure portals.
4. Multi-Modal Biometric Fusion:
   Combining voice recognition with other modalities (e.g., facial, iris, or fingerprint recognition) can improve accuracy and security in sensitive applications such as defence, finance, or law enforcement.
5. Mobile and Edge Deployment:
   Developing a lightweight mobile version of the system can expand accessibility. Edge-device deployment will enable real-time operation without requiring continuous internet connectivity, making it suitable for classrooms, offices, and remote learning platforms.
6. Ethical and Privacy Frameworks:
   Future implementations should focus on ensuring data security, encryption, and compliance with privacy regulations such as GDPR to protect users' biometric data from misuse.

Overall, the Voice Biometric Attendance System demonstrates that combining artificial intelligence with speech processing can transform conventional attendance management into a smart, automated, and secure ecosystem. With the ongoing evolution of deep learning and edge computing, speaker recognition is expected to become a core component of future authentication systems across industries.

IX. REFERENCES

1. T. Kinnunen and H. Li, "An Overview of Text-Independent Speaker Recognition: From Features to Supervectors," *Speech Communication*, 2010.
2. D. Reynolds, "An Overview of Automatic Speaker Recognition Technology," *IEEE ICASSP*, 2002.
3. S. Davis and P. Mermelstein, "Comparison of Parametric Representations for Word Recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 1980.
4. S. Snyder et al., "X-Vectors: Robust DNN Embeddings for Speaker Recognition," *IEEE ICASSP*, 2018.
5. Y. Xu et al., "Attention-based Convolutional Neural Network for Speaker Verification," *IEEE Transactions on Audio, Speech, and Language Processing*, 2019.
6. A. Jain et al., "An Introduction to Biometric Recognition," *IEEE Trans. Circuits Syst. Video Technol.*, 2004.
7. W. Zhao et al., "Face Recognition: A Literature Survey," *ACM Computing Surveys*, 2003.
8. L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Prentice-Hall, 1993.
9. Scikit-learn Documentation, https://scikit-learn.org.
10. Librosa Documentation, https://librosa.org.
11. OpenPyXL Documentation, https://openpyxl.readthedocs.io.
12. D. Ganchev, N. Fakotakis, and G. Kokkinakis, "Comparative Evaluation of Various MFCC Implementations on the Speaker Verification Task," *Proceedings of SPECOM*, pp. 191–194, 2005.
13. S. O. Sadjadi, C. Greenberg, and E. Singer, "The 2019 NIST Speaker Recognition Evaluation: CTS Challenge," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 3042–3056, 2021.
14. S. Snyder, D. Garcia-Romero, D. Povey, and S. Khudanpur, "X-Vectors: Robust DNN Embeddings for Speaker Recognition," *IEEE ICASSP*, 2018.
15. Y. Xu, J. Du, L.-R. Dai, and C.-H. Lee, "A Regression Approach to Speech Enhancement Based on Deep Neural Networks," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 1, pp. 7–19, 2015.
16. M. McLaren, D. van Leeuwen, and N. Brümmer, "SRE'12 Speaker Recognition Evaluation: An Overview," *Computer Speech & Language*, vol. 47, pp. 58–77, 2018.
17. A. Sarkar, R. Sinha, and S. Das, "A Study on Speaker Recognition Using Machine Learning and Deep

Neural Network Techniques," *International Journal of Speech Technology*, vol. 24, pp. 1009–1024, 2021.

18. T. Nandwana, M. McLaren, and L. Ferrer, "Deep Learning Architectures for Robust Speaker Recognition," *IEEE Signal Processing Letters*, vol. 26, no. 7, pp. 1031–1035, 2019.

Emotion-Aware Speaker Identification Using Spectrogram-Based CNN Models

Bhavish Saroha
Maharaja Surajmal Institute of Technology
Dept. of Electronics and Comm. Engg.
Delhi, India
bhavishsaroha@gmail.com

Akshat Bahl
Maharaja Surajmal Institute of Technology
Dept. of Electronics and Comm. Engg.
Delhi, India
akshatbahl18@gmail.com

Sonia Malik
Maharaja Surajmal Institute of Technology
Dept. of Electronics and Comm. Engg.
Delhi, India
sonia.malik@msit.in

Abstract— This paper presents an emotion-aware speaker identification system that integrates emotional context to enhance recognition robustness. Using Convolutional Neural Networks (CNNs) applied to log-Mel spectrograms, the model jointly performs speaker and emotion classification. Experiments on the RAVDESS dataset demonstrate the system's ability to maintain stable identification across diverse emotional states. The approach focuses on improving recognition reliability under expressive conditions by treating emotion as an auxiliary learning task. The system achieves competitive accuracy and is suitable for real-time applications due to its CNN-based architecture optimized for efficiency. Furthermore, the study highlights how emotional cues such as pitch, energy, and spectral variations influence speaker-dependent features and impact classification consistency. By employing a shared feature extraction backbone and task-specific learning heads, the model achieves effective cross-task generalization.

Index Terms— Speaker Identification, Emotion Recognition, CNN, Spectrograms, Affective Computing.

I. Introduction

Speaker identification aims to determine a speaker's identity from their voice signal by analyzing unique acoustic characteristics such as timbre, pitch, and formant structure. It is fundamental to numerous applications, including voice biometrics, access control, forensics, and conversational AI. However, one of the key limitations of conventional systems is their reduced accuracy under emotionally expressive conditions. Emotional variations affect prosodic features like tone, energy, and speaking rate, leading to misclassifications in emotion-agnostic systems.

Traditional models, such as Gaussian Mix- ture Models (GMMs) and Hidden Markov Mod- els (HMMs), rely heavily on handcrafted features like Mel-Frequency Cepstral Coefficients (MFCCs). With the advent of deep learning, CNN-based methods that use spectrograms have shown superior ability to extract abstract representations from raw audio [3]. Despite this, most existing works treat emotion and speaker identity as independent problems. This separation overlooks the fact that emotional cues carry valuable speaker-specific variations.

To address this gap, our project proposes a CNN-based multi-task framework that jointly identifies the speaker and recognizes the corresponding emotional state from speech. By leveraging log-Mel spectrograms as input features, the system benefits from powerful spatial pattern recognition while maintaining efficiency for real-time usage. The integration of emotional understanding improves model robustness in varied acoustic environments and contributes to more human-aware audio recognition systems.

II. Objectives

The main objectives of this research are:

1. Develop a CNN-based model capable of performing both speaker identification and emotion recognition.

2. Use log-Mel spectrograms to convert audio into a 2D time-frequency representation suitable for CNN learning.

3. Evaluate performance using accuracy for both

4. Enhance generalization via data augmentation and fine-tuning strategies.

5. Assess the impact of emotion on identification accuracy under expressive conditions.

III. Related Work

Noh and Jeong [1] demonstrated the benefit of emotion-aware embeddings using transfer learning, achieving a notable improvement in recognition consistency. Zhao et al. [4] combined CNN and Bi-LSTM layers for joint learning of temporal and spectral cues, showing improved robustness. Kim et al. [5] applied emotional data augmentation to enhance stability under stress. Desplanques et al. [6] proposed a shared CNN encoder for multitask learn- ing of emotion and speaker identity. Our approach differs in that it simplifies the architecture by employing a CNN-only model, emphasizing efficiency without recurrent layers, and focusing on transfer learning using VGG16 as a shared feature extractor. Additionally, the model leverages simultaneous training of both tasks to enhance shared feature representations, reducing redundancy and overfitting. The framework prioritizes computational efficiency, making it suitable for real-time inference on limited hardware. By integrating emotion and speaker cues in a unified CNN backbone, the system achieves a balance between accuracy and deployment practicality.

IV. Methodology

A. Dataset

The **Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)** [2] was used for experimentation. It includes recordings from 24 professional actors (12 male and 12 female) performing 8 emotions — neutral, calm, happy, sad, angry, fearful, disgust, and surprised. Each actor produced 60 audio samples, yielding approximately 1,440 utterances.

B. Audio Preprocessing

To ensure uniformity and robustness across varying audio samples, a series of preprocessing steps were applied before model training. Each raw audio file underwent the following transformations:

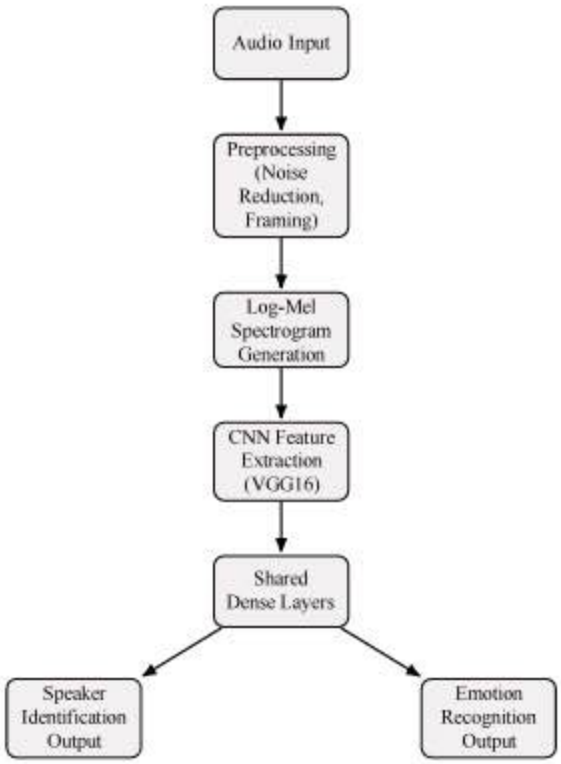


Figure 1: Pipeline for Emotion-Aware Speaker Identification system. The model processes raw audio through feature extraction and parallel classification for speaker and emotion recognition.

- **Resampling:** All recordings were resampled to a standard rate of 22,050 Hz in mono for- mat. This ensured consistent temporal reso- lution while reducing computational overhead without significant loss of perceptual quality.

- **Silence Trimming:** Leading and trailing silence regions were removed using a decibel threshold of `top_db = 20`. This helped retain only the active speech portions, improving the effective use of input frames and enhancing the quality of extracted features.

- **Duration Standardization:** Each audio clip was normalized to a fixed duration of 3 seconds. Longer clips were truncated, while shorter ones were padded with minimal silence to maintain consistent temporal input dimensions for the model.

- **Spectrogram Generation:** The processed waveforms were converted into log-Mel spectrograms with 128 Mel frequency bands. This representation captures perceptually relevant frequency components while compressing the amplitude range logarithmically for improved

stability during training.

- **Resizing for CNN Input:** To match the expected input dimensions of pre-trained CNN architectures, the generated spectrograms were resized to 224 × 224 × 3, effectively converting them into pseudo-RGB images suitable for VGG16.

Following spectrogram generation, all samples were normalized using **global mean and variance normalization** to ensure consistent feature scaling across the dataset.

To enhance model generalization and reduce overfitting, diverse **data augmentation** strategies were employed during preprocessing. This included:

- **Random zoom:** Applied controlled zoom-in and zoom-out transformations on spectrogram representations to mimic variations in speaker proximity and recording setup, enabling the model to learn scale-invariant acoustic features and improving adaptability to diverse recording conditions.
- **Random translation:** Introduced subtle horizontal and vertical shifts in spectrograms to emulate small temporal or frequency misalignments, enhancing the model's robustness to timing jitter, microphone placement differences, and background variability in real-world recordings.

These augmentations collectively improved the variability of the dataset and enhanced the model's ability to handle unseen acoustic conditions.

C. Model Architecture

The proposed system utilizes a **multi-task convolutional neural network (CNN)** based on the **VGG16** architecture to perform both *speaker identification* and *emotion recognition* simultaneously. The design leverages parameter sharing to learn a unified representation of acoustic features while maintaining task-specific specialization through separate output heads.

- **Shared Encoder:** The convolutional base of the pre-trained VGG16 model, originally trained on ImageNet, is employed as a feature extractor. By leveraging transfer learning, the model benefits from rich, low-level feature representations such as edges, textures, and frequency patterns present in spectrograms.
- **Task-specific Dense Heads:** Two independent fully-connected branches are appended to the shared encoder:
  - **Speaker Identification Head:** A dense pathway with layers Dense(256) → Dropout(0.5) → Dense(128) → Dense(24, Softmax), optimized for multi-class classification of 24 speakers.
  - **Emotion Recognition Head:** Comprises Dense(256) → Dropout(0.5) → Dense(128) → Dense(8, Softmax), dedicated to recognizing eight emotion categories from shared acoustic representations.
- **Loss Function:** The model is trained using **categorical cross-entropy loss** for both tasks. A weighted combination of the two losses allows the network to optimize shared representations while maintaining task-specific accuracy.
- **Optimization and Regularization:** Training utilizes the Adam optimizer with early stopping and dropout regularization in dense layers to prevent overfitting. Batch normalization is employed to stabilize gradient updates and accelerate convergence.

This multi-task setup enables efficient joint learning, allowing the model to exploit commonalities between emotion and speaker characteristics—such as pitch, tone, and prosodic features—resulting in improved overall performance and generalization.

The joint objective function is defined as:

$$L_{total} = \alpha L_{speaker} + \beta L_{emotion} \quad (1)$$

where α and β are weighting factors balancing the two tasks. The proposed model uses transfer learning with VGG16 as the backbone, augmented with normalization, dropout, and dual dense branches for speaker and emotion classification. The detailed architecture is summarized in Table 1.

D. Training Procedure

Training is conducted in two stages:

- **Stage 1:** Freeze VGG16 base; train only the classification heads for 50 epochs (learning rate = 0.001).
- **Stage 2:** Fine-tune all layers for 100 epochs (learning rate = 1e-5) with early stopping.

Layer (Type)	Output Shape	Param #	Connected To
Input Layer (InputLayer)	(None, 224, 224, 3)	0	-
Random Zoom (RandomZoom)	(None, 224, 224, 3)	0	Input Layer
Random Translation (RandomTranslation)	(None, 224, 224, 3)	0	Random Zoom
VGG16 (Functional)	(None, 7, 7, 512)	14,714,688	Random Translation
Global Average Pooling (GlobalAveragePooling2D)	(None, 512)	0	VGG16
Batch Normalization (BatchNormalization)	(None, 512)	2,048	Global Average Pooling
Dense (Dense)	(None, 256)	131,328	Batch Normalization
Dense_1 (Dense)	(None, 256)	131,328	Batch Normalization
Dropout (Dropout)	(None, 256)	0	Dense
Dropout_1 (Dropout)	(None, 256)	0	Dense_1
Speaker Output (Dense)	(None, 24)	6,168	Dropout
Emotion Output (Dense)	(None, 8)	2,056	Dropout_1

Table 1: Model Architecture Summary for Emotion-Aware Speaker Identification

Optimization uses the Adam optimizer and batch size 32. Validation accuracy is monitored per epoch, and the best model is saved based on minimum validation loss. .

E. Evaluation and Results

Performance is measured via Accuracy for both speaker and emotion recognition. The proposed model achieved an average **87.85% accuracy** in speaker identification and **73.96% accuracy** in emotion classification.

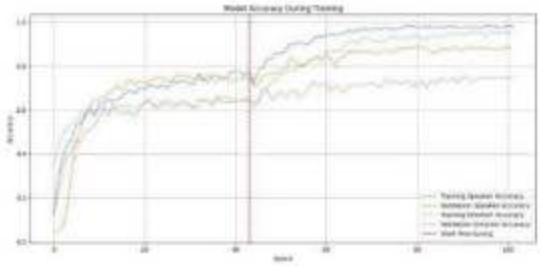


Figure 2: Model Accuracy During Training. Training and validation accuracies for both speaker and emotion classification tasks are shown. The red dashed line indicates the start of fine-tuning.

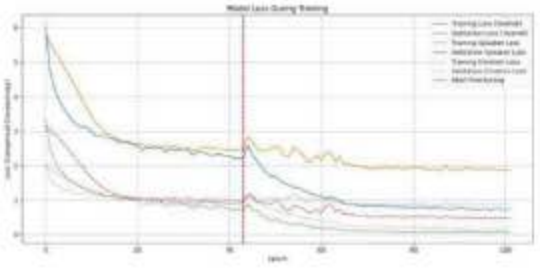


Figure 3: Model Loss During Training. Loss curves for overall, speaker, and emotion classification tasks. The red dashed line marks the onset of fine-tuning.

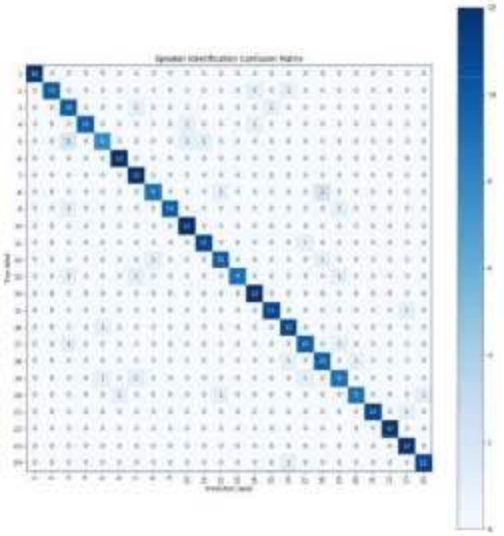


Figure 4: Speaker Identification Confusion Matrix. Strong diagonal dominance reflects accurate identification across 24 speakers.

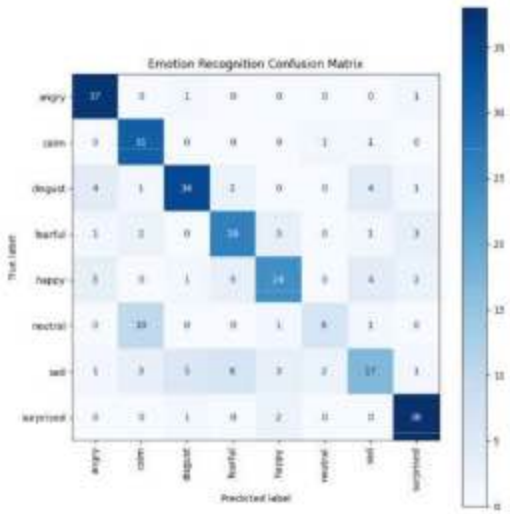


Figure 5: Emotion Recognition Confusion Matrix. Emotions like anger, surprise, and happy are well recognized, while neutral and sad exhibit partial overlap.

V. Conclusion

This study introduces a CNN-based multi-task system for emotion-aware speaker identification. By integrating emotional information directly into the speaker identification process, the model achieves improved performance across varied contexts. The system demonstrates potential for real-world applications such as emotion-sensitive authentication, interactive AI systems and forensic analysis.

A. Applications

The proposed multi-task audio classification framework has promising applications in real-world systems that require both speaker and emotion awareness. Integrating such a model can enhance personalization, security, and human-computer interaction across several domains.

- **Emotion-Aware Authentication:** Voice-based authentication systems can incorporate emotional context to detect stress, anger, or distress, providing an additional behavioral layer of security.
- **Healthcare and Therapy:** Emotion recognition from speech can assist in remote mental health monitoring and emotional well-being tracking for patients.
- **Human-Robot Interaction:** Emotionally responsive robots and virtual assistants can interpret both the identity and mood of users, leading to more natural and empathetic interactions.
- **Surveillance and Forensics:** Identifying speakers and their emotions in multi-speaker audio streams can aid in behavioral analysis and forensic investigations.

Acknowledgment

The authors would like to express their sincere gratitude to Maharaja Surajmal Institute of Technology, Delhi, for providing the necessary resources and support to carry out this research work. Their guid- ance and encouragement have been instrumental in the successful completion of this study.

References

[1] K. Noh and H. Jeong, "Emotion-Aware Speaker Identification with Transfer Learning," *IEEE Access*, vol. 11, pp. 77292–77306, 2023.

[2] S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)," *PLOS ONE*, vol. 13, no. 5, p. e0196391, 2018.

[3] H. Zeinali, L. Burget, and J. Cernocky, "Convolutional Neural Networks and X-Vector Embedding for Acoustic Scene Classification," *arXiv preprint arXiv:1810.04273*, 2018.

[4] L. Zhao et al., "Emotion-Aware Speaker Recognition Using Hybrid Deep Learning Architectures," *IEEE Access*, vol. 8, pp. 175530–175542, 2020.

[5] J. Kim et al., "Robust Speaker Identification Under Emotional Stress Using Deep CNNs," *Speech Communication*, vol. 129, pp. 12–22, 2021.

[6] B. Desplanques, J. Thienpondt, and K. Demuynck, "Multitask Learning for Joint Speaker and Emotion Recognition," in *Proc. Interspeech*, 2021, pp. 2312–2316.

[7] A. H. Meftah et al., "Speaker Identification in Different Emotional States in Arabic and English," *IEEE Access*, vol. 8, pp. 60070–60083, 2020.

[8] H. Li et al., "Emotion-Aware Speaker Verification Based on Feature-Level Fusion," *IEEE Transactions on Affective Computing*, vol. 13, no. 4, pp. 1802–1815, 2022.

[9] C. Zhang et al., "Cross-Dataset Emotion and Speaker Recognition Using Spectrogram-Based Deep Networks," *Neural Computing and Applications*, vol. 35, no. 9, pp. 6613–6627, 2023.

[10] T. Wang et al., "CNN-Based Speech Emo- tion Recognition and Speaker Identification Fusion Network," *Applied Acoustics*, vol. 214, p. 109509, 2024.

Variable Focal Length Metalens for Extended Depth of Focus Imaging

1st Nishant Chhiller
Department of Applied Sciences
National Institute of Technology, Delhi
Delhi, India

2nd Neelam Barak
Department of Electronics and Communication
Maharaja Surajmal Institute of Technology
Janakpuri, Delhi, India

3rd Vineeta Kumari
Scientist – 'C', Department of Science & Technology,
Mehrauli, New Delhi, India

4th Arun Anand
Department of Physics
Sardar Patel University
Anand, Gujarat

5th Gyanendra Sheoran*
Department of Applied Sciences
National Institute of Technology, Delhi
Delhi, India
gsheoran@nitdelhi.ac.in

***Abstract* - Metalenses—planar lenses built from subwavelength meta-atoms—enable ultracompact optics with phase- engineered wavefront control. Recent advances in tunable and varifocal metalenses open the possibility of a single ultrathin optic that provides dynamic focal-length control while maintaining high imaging performance. One important utility of variable-focal metalenses is enabling a large depth-of-focus (DoF) or extended depth-of-focus (EDOF) without mechanically moving bulky optics, which benefits endoscopy, machine vision, and compact imaging in mobile devices. This short paper reviews recent implementations of varifocal metalenses, highlights design strategies for combining focal tuning with EDOF, and outlines application requirements and trade-offs compared with small-DoF metalenses.**

Keywords—Metalens, Depth of Focus (DOF), Extended depth of focus (EDOF), Imaging, wavefront, Varifocal

I. INTRODUCTION

Metalenses are lenses based on subwavelength nanostructures called metasurface that can manipulate electromagnetic waves by imparting arbitrary phase or controlling its amplitude or even polarization. Metasurfaces have transformed optical design by allowing for ultrathin, flat optics with functionalities similar or even surpassing conventional optics [1,6]. Various tunable or varifocal metalenses based on a range of mechanisms, including MEMS-actuation (Alvarez-type metasurfaces) [1], liquid-crystal integration [2], phase-change materials [4], and electro-optic platforms [6], have been fabricated. These approaches allow dynamic focal-length control in a compact form factor, which is applicable to zoom lenses, autofocus- free imaging, and depth-resolved capture in systems with small space availability. However, conventional fixed-focal metalenses generally exhibit limited depth-of-focus (DoF), limiting their performance in applications with axial motion or wide depth ranges. A variable-focal-length metalenses that combine focal tuning with an extended depth-of-focus (EDOF) can overcome these limitations by providing a larger usable focus range while retaining the small size [3,4].

Alvarez and Moiré varifocal concepts have been demonstrated with metasurfaces for variable focal length [1,7]. Liquid-crystal based metasurfaces allows electrically driven continuous tuning with low voltage and thin form, it also ensure quick and vibration free focal length changes [2].

Inverse design and topology optimization have been used for high-NA EDOF meta-optics that ensures higher focusing efficiency across an extended axial range [3,4]. Recent studies using tunable, low-loss dielectrics show tunability over a long focal range and high efficiency, making them suitable for integrated imaging systems [8]. Tunable metasurfaces or metalens are only limited by substrate thickness and material properties [5,6].

Various methods to combine varifocality and EDOF include (i) meta-Alvarez pairs for wide-range tuning with cubic-phase compensation, (ii) hybrid stacks that superimpose an axicon/cubic static phase for EDOF with a tunable overlay for focal control, and (iii) inverse-design to directly optimize for an extended focal region while preserving efficiency. Computational imaging methods (deconvolution, multi- frame fusion) are often paired with EDOF optics to restore resolution loss. In this paper an aperture divided approach has been discussed that is suitable for increasing the depth of focus which can be useful for thick sample imaging without the need for axial scanning or wavelength sweeping.

II. DESIGN & SIMULATION OF METALENS

The design of the metalens is based on aperture division, where a portion of the total aperture is dedicated to maintaining focus at a fixed focal length; similarly, the next portion is dedicated to the next focal length. The choice of focal lengths corresponding to each aperture portion is such that the next focus has a center at about half FWHM of the previous focal spot. This ensures continuous high efficiency, and overall focus is maintained. A schematic of the aperture division is indicated in Figure 1. Figure 2 gives the design parameters such as the periodicity, material selection, unit cell height etc. The structure of the unit cell is a cylindrical pillar with a fixed height of 750 nm and periodicity of 320 nm. The nano cylinder is made up of amorphous silicon hydride, which has excellent transmission in the visible range (400 – 750 nm), and at the same time, the material is fabrication-friendly as it's based on silicon, which is a widely used foundry material.

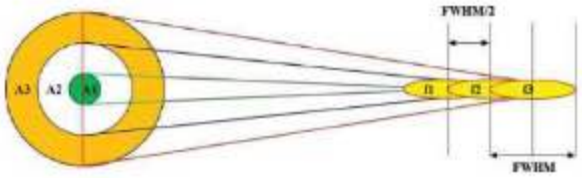


Fig 1. The design of the aperture divided metalens with three divisions, namely A1, A2, A3, each having a corresponding focal length f1, f2, f3. Each focal spot overlaps with the other one at one-half of its FWHM to ensure continuous focus over an extended range of focal length.

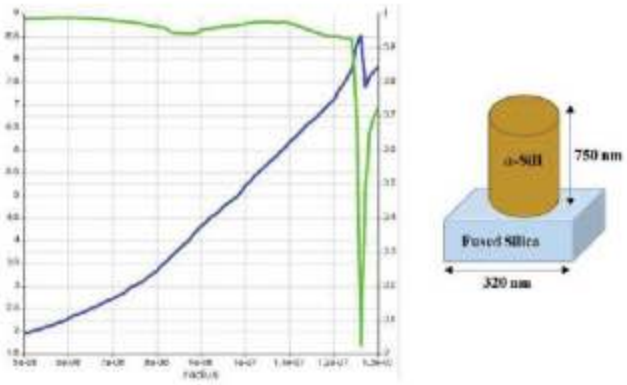


Fig 2. (Left) The phase accumulation and transmission of nano pillars as a function of their diameter; it can be seen that a continuous and high transmittance is maintained over a diameter range of 50 nm to 130 nm. (Right): The unit cell has a height of 750 nm and a periodicity of 320 nm.

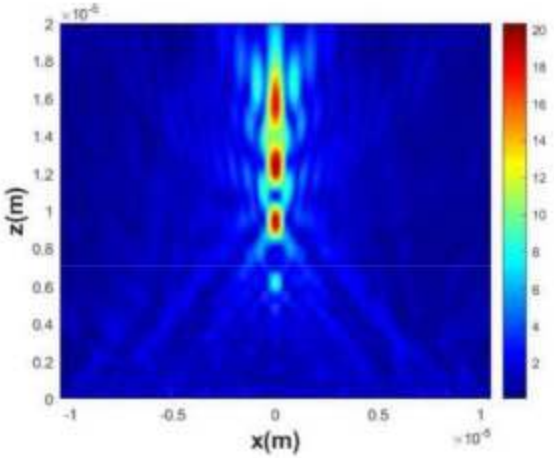


Fig 3. ZX plane view of the simulated metalens. We can see multiple overlapping focal spots from around 9 μm up to 16 μm.

III. Results

Finite difference time domain (FDTD) simulations were carried out to study the proposed idea. A metalens of diameter 20 μm was designed and simulated.

Figure 3 shows the ZX plane of the simulation, which is basically a vertical view of the focusing created by the metalens. The metalens is designed for a 633 nm wavelength, and it can be seen in Figure 3 that an almost continuous focus is maintained from about 9 μm to about 16 μm. This corresponds to a DOF of 7 μm, which is large given the small diameter of the lens. It is to be noted that the overall intensity is equally distributed at each focal spot, as the area dedicated to each region is equal. The single focus efficiency was about 60% with a target focal length of 10 μm, and thus, with four focal spots, each focal spot will have about 15% efficiency with respect to the injected light. The efficiency could be improved by optimizing the metalens design and directly increasing the efficiency at each focal spot. Since each focal spot overlaps with the previous one, a continuous focus is maintained within the range. The focal spots are also visualized in XY plane (Figure 4), which shows that the focal spot is sharp and tight at various XY planes (z = 10 μm and 14 μm)

The ZX view indicates that there are certain intensity dips along the z-axis. These are due to the non-optimized design, and with a better choice of the aperture areas, we can create overlaps across the z-axis, which will be a more realistic DOF extension.

The FWHM of the focal spots at various axial distances was quantified to compare the focusing and resolution at each plane. As seen in Figure 4, the FWHMs of the two distinct planes at a distance of 5 μm are 520 and 550 nm, which are very similar, indicating that the two planes have similar focusing properties. To verify the stability of the metalens in different environmental conditions, FDTD simulations were conducted with simulation domain temperatures ranging from – 50°C to +50°C. It was found that the performance remains similar at each temperature. The refractive index does change with temperature, but it does not affect the metalens behavior, as the phase change imparted by single meta-atoms follows equation (1).

$$\frac{\phi}{n} = \frac{H}{\lambda} \cdot \frac{2}{\pi} \qquad (1)$$

Further, the generalized Snell's law tells how the phase of the incoming wave will be changed by many meta-atoms placed periodically, as indicated in equations (2) and (3) [11]

$$n_1 sin\theta_r - n_1 sin\theta_1 = \frac{1}{k} \frac{d}{dx} \phi(x) \qquad (2)$$

$$n_2 sin\theta_2 - n_1 sin\theta_1 = \frac{1}{k} \frac{d}{dx} \phi(x) \qquad (3)$$

Thus, if the refractive index changes, then the term $\frac{d}{dx}\phi(x)$ also changes. Clearly, it's this relative change that focuses the

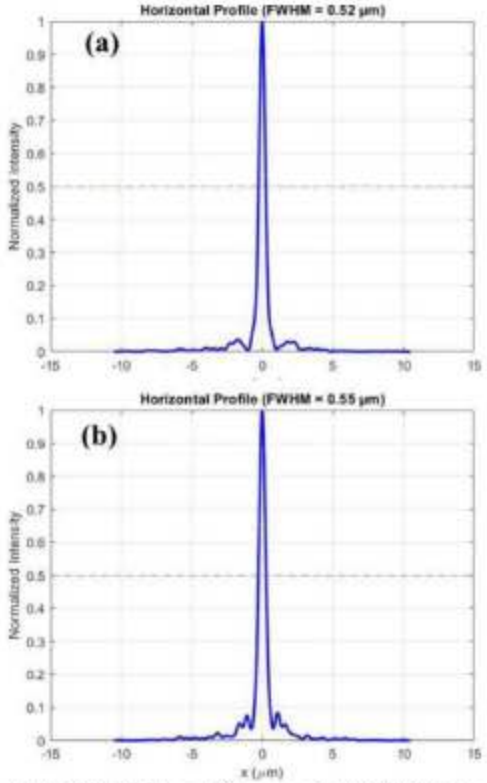


Fig 4. (a) FWHM at z = 10 μm and (b) FWHM at z = 15 μm At respective planes the FWHMs are 520 nm and 550 nm

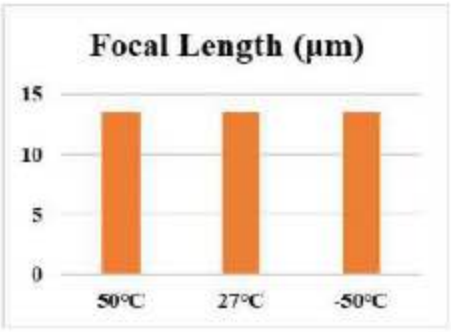


Fig 5. Invariance of the focal length with temperature changes. Depicted here, z = 13.5 μm

The metalens was also verified for its imaging performance. FDTD simulations were performed with an object kept between the lens and the input (source). In order to simulate the object, a perfect electric conductor (PEC) layer was used, which basically absorbs all incident light on it. The letter 'G' of the English alphabet was cut out in the middle of this layer; thus, the layer serves as an aperture through which light can only pass through the part where PEC is not present and blocked by the remaining portion. Figure 6 shows the imaging performance of the simulated metalens with a single focus. This was done to save simulation time and simplify the simulation. With suitable computing power, we can simulate the imaging performance with multiple focal spots that require multiple field monitors to capture the images. Furthermore, the simulation was restricted to a single plane, as beyond z = 15 μm, the image magnifies and cannot be imaged in full.

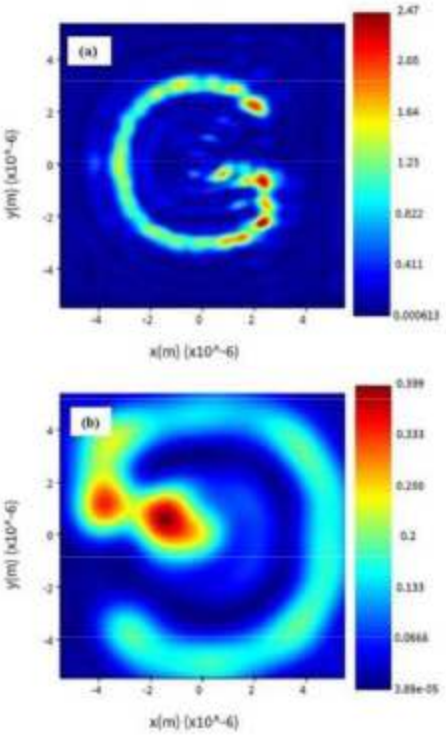


Fig 6. (a) The Object at the input plane (b) Object image at z
– 13 μm

IV. Future Scope & Conclusion

The simulations suggest that an aperture-divided metalens with intelligently selected focal lengths could create a continuous focus over a range of axial distances. Such a system, once fabricated, could replace bulky and costly conventional optics. Such a system could be beneficial for use in applications such as microscopy of thick biological samples, where a single-step capture can provide information about the bulk properties. It could be integrated with a variable focal length lens to create a compact continuous zoom lens. The existing design could be improved by optimizing the metalens design to increase efficiency. Further, the focal lengths corresponding to each area of the aperture can also be optimized such that the focal spots overlap at exactly 50% of the previous one and ensure no losses in between subsequent focal spots. This will not increase individual efficiency but will improve the focusing ability overall, making a more reliable EDOF system.

REFERENCES

1. Han, Z., et al. (2020). MEMS-actuated metasurface Alvarez lens. Microsystems & Nanoengineering, 6(1), Article 93. https://doi.org/10.1038/s41378-020-00200-5
2. Bosch, M., et al. (2021). Electrically actuated varifocal lens based on liquid-crystal-encapsulated dielectric metasurface. Nano Letters, 21(8), 3849–3856. https://doi.org/10.1021/acs.nanolett.1c00671
3. Bayati, E., et al. (2021). Inverse-designed extended depth-of-focus meta-optics. Photonics Research, 9(8), B28–B36. https://doi.org/10.1364/PRJ.417777
4. Zheng, Y., et al. (2022). Designing high-efficiency extended depth-of-focus metalens via topology-shape optimization. Nanophotonics, 11(5), 1037–1047. https://doi.org/10.1515/nanoph-2021-0672
5. Santonocito, A., et al. (2023). Recent advances in tunable metasurfaces and their applications. Nanomaterials, 13(3), 521. https://doi.org/10.3390/nano13030521
6. Jung, C., et al. (2024). The rise of electrically tunable metasurfaces. Science Advances, 10(15), eadk9085. https://doi.org/10.1126/sciadv.adk9085
7. Luo, Y., et al. (2021). Varifocal metalens for optical sectioning fluorescence. Nano Letters, 21(22), 9427–9433. https://doi.org/10.1021/acs.nanolett.1c03748
8. Wang, M., et al. (2023). Varifocal metalens using tunable and ultralow-loss dielectrics. Advanced Science, 10(15), 2300057. https://doi.org/10.1002/advs.202300057
9. Lim, T.-H., et al. (2022). Design of Moiré-inspired metasurface lens for focusing. Energies, 15(3), 1144. https://doi.org/10.3390/en15031144
10. Rousseau, E., & Felbacq, D. (2023). Detailed derivation of the generalized Snell–Descartes laws from Fermat's principle. Journal of the Optical Society of America A, 40(4), 676-681. https://doi.org/10.1364/JOSAA.478123

THE PARKINSON'S DISEASE DETECTION USING MACHINE LEARNING TECHNIQUES

TANVI RAO , Prof. NEELAM NEHRA ,
MAHARAJA SURAJMAL INSTITUTE OF TECHNOLOGY

ABSTRACT: The Parkinson's disease is progressive neuro degenerative disorder that affects a lot only people significantly affecting their quality of life. It mostly affect the motor functions of human. The main motor symptoms are called "parkinsonism" or "parkinsonian syndrome". The symptoms of Parkinson's disease will occur slowly, the symptoms include shaking, rigidity, slowness of movement and difficulty with walking, Thinking and behavior change, Depression and anxiety are also common. There is a model for detecting Parkinson's using voice. The deflections in the voice will confirm the symptoms of Parkinson's disease. This project showed 73.8% efficiency. In our model, a huge amount of data is collected from the normal person and also previously affected person by Parkinson's disease. these data is trained using machine learning algorithms. From the whole data 60% is used for training and 40% is used for testing. The data of any person can be entered in db to check whether the person is affected by Parkinsons disease or not. There are 24 columns in the data set each column will indicate the symptom values of a patient except the status column. The status column has 0's and I's . those values will decide the person is effected with Parkinsons disease. I's indicate person is effected, 0's indicate normal conditions.

Key Words: Parkinson's disease; machine learning (ML), XGBoost, Decision tree.

I. INTRODUCTION:

Parkinson's disease is•a disorder of the central nervous system affecting movement and inducing tremors and stiffness a neurodegenerative disorder affecting dopamine neurons in brain. Parkinson's disease is difficult to diagnose. Common diagnostic criteria require the medication before. In this model, the huge data is collected from previously affected person and then by using machine learning algorithm will process the user input data with previous data to check he/she affected.

II. LITERATURE REVIEW:

Glenda-M.halliday, Nichola,"Parkinson's progression prediction using ml and serum cytokines". 25-July-2019. The serum samples from a clinic are tested to find Parkinson's disease and the same samples are tested using ML algorithm to detect Parkinson's disease. Blauwendraat, C., Bandres-Ciga, S. & Singleton, A. B. Predicting the progression in patients with Parkinson's disease using their voice. Lancet Neurol.2017. •Voice change is also a symptom of Parkinson's disease by applying ML algorithm. Das R. "A comparison of multi-classification methods for diagnosis of Parkinson's disease". Expert Systems with Applications"; 37:1568-1572 2010. •For methods used for testing Parkinson's disease they are ML, DM neural, regression, decision tree in those ML show high performance.

III. OBJECTIVES

Aim of the Project:

The main aim is to test the ability of motor function of the patient with Parkinson's disease.

Scope of the Project:

The scope of this project is to show the high accuracy of detecting Parkinson's disease in early stage.

IV. DESIGN AND METHODOLOGIES:

MODULE 1:

- Data Collection

MODULE 2:

•Training and testing of data.

MODULE 3:

- Apply XGBoost algorithm.

MODULE 4:

•Cod completion.

V. IMPLEMENTATION:

•ARCHITECTURE DIAGRAM

•DATA-FLOW-DIAGRAM
•ER-DIAGRAM

•SEQUENCE DIAGRAM

ARCHITECTURE DIAGRAM:

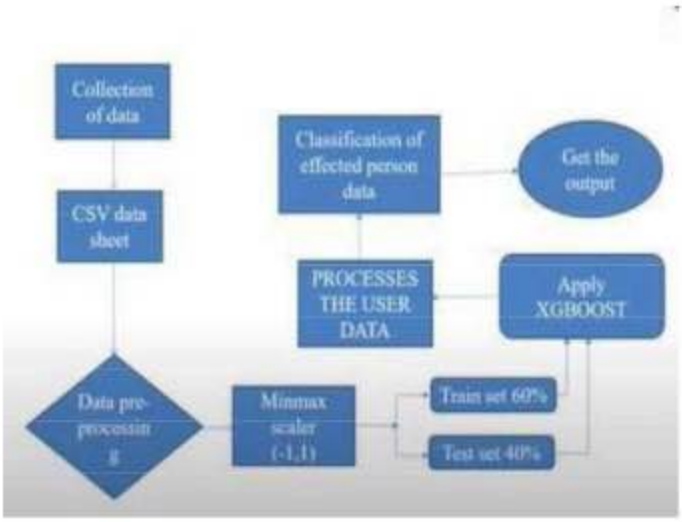


Fig.1 Architecture Diagram

DATA FLOW DIAGRAM:

Fig.2. Data Flow Diagram

ER DIAGRAM:

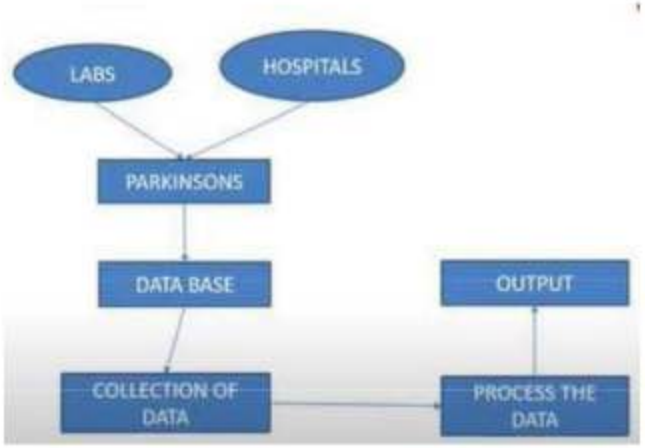


Fig.3 .ER Diagram

SEQUENCE DIAGRAM:

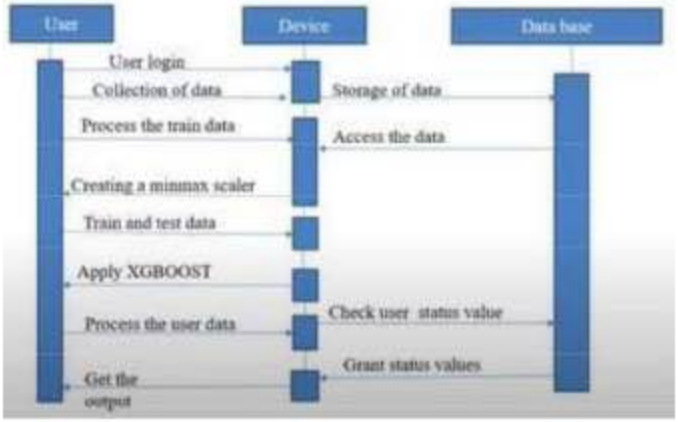


Fig.4.Sequence Diagram

VI. TESTING:

•UNIT TESTING

•INTEGRAION TESTING

•FUNCTIONAL TESTING.

UNIT TESTING:

It is the level of software testing where individual units and the components are tested. In the proposed project the data of an individual person is taken and tested. The accuracy is high 100% when tested with a single person data.

INTEGRATION TESTING

It may be level of software testing where individual units are combined and it tested as a gaggle. In the proposed project all the data is combined and tested. The accuracy level is 94.87%. This testing will test whole project at a time.It reduces the time complexity in integration testing.

• FUNCTIONAL TESTING:

Functional testing may be a sort of software testing that validates the software against the functional requirements/specifications. This testing is detecting Parkinson's will based on machine learning algorithm.ML algorithm will boost up the speed.

Typically, functional testing involves the following steps:

• Identifying the functions of that the software is expected to perform.

•Create input-data based on the function's specifications.

•It Determines the output based up on the function's specifications.

•Execute the test case.

•Compare the actual and expected outputs.

VII. INPUT AND OUTPUT:

SCREENSHOTS:

Fig.5 . Data set

Fig.6.Checking Missing values

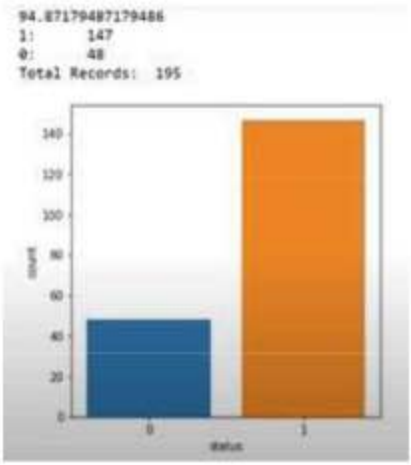


Fig.7. Graph

Here fig.7 represents the features and labels .where features are input of our data and labels are like output of our data .after apply features and labels we get 147 1's it represents the person is affected with Parkinson's disease and 48 0's it represents that the person is not not affected with Parkinson's disease

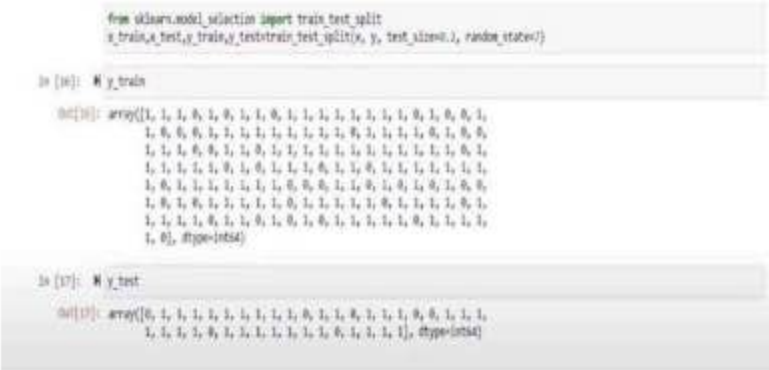


Fig.8. Training and testing data set

CONCLUSION:

In this process we can predict the Parkinson's disease in patient's body using machine learning technology and this method makes the process easy to our user.

Our analysis provides very accurate performance in detecting Parkinson's disease using XGBOOST algorithm

VIII. REFERENCES

1.Mahlknecht, P.; Krismer, F.; Poewe, W.; Seppi, K. Meta-Analysis of Dorsolateral Nigral Hyperintensity on Magnetic Resonance Imaging as a Marker for Parkinson's Disease. Mov. Disord. 2017, 32, 619–623.

2. Dickson, D.W. Neuropathology of Parkinson disease. Parkinsonism Relat. Disord. 2018, 46 (Suppl. 1), S30–S33.

3. Kalia, L.V.; Lang, A.E. Parkinson's Disease. Lancet 2015, 386, 896–912.

4. D. Heisters, "Parkinson's: symptoms treatments and research", vol. 20, no. 9, pp. 548-554, 2011.

5. Prediction of Parkinson's Disease Using Machine Learning Methods." — Uses demographic + clinical + polygenic risk scores, uses SHAP for interpretability

6. García Santa Cruz, B., (2023). Machine learning models for diagnosis and prognosis of Parkinson's disease. *Frontiers in Aging Neuroscience*.

7. Rabie, H., & Akhloufi, M. A. (2025). A review of machine learning and deep learning for Parkinson's disease detection. *Discover Artificial Intelligence*, **5**(1), 24.

8. Mei, J., Desrosiers, C., & Frasnelli, J. (2021). Machine learning for the diagnosis of Parkinson's disease: a systematic review.

9. Kannan, H. K. (2025). Predictive analysis of Parkinson's disease using machine learning and speech signals.

Author's Profile: -

1. Tanvi Rao (02115002822) Student, BTECH. Electronics and Communication Engineering, Maharaja Surajmal Institute of technology, GGSIPU

2. Ravi Yadav (2415002822) Student, BTECH. Electronics and Communication Engineering, Maharaja Surajmal Institute of technology, GGSIPU

3. Aman Kumar prasad (1715002822) Student, BTECH. Electronics and Communication Engineering, Maharaja Surajmal Institute of technology, GGSIPU

4. Prof. Neelam Nehra , ECE Dept, Maharaja Surajmal Institute of technology, GGSIPU

5. Prof. Geetanjali Sharma, ECE Dept, Maharaja Surajmal Institute of Technology, GGSIPU

AI-DRIVEN CYBERSECURITY: THREAT ANALYSIS AND ORGANIZATIONAL TRANSFORMATION FOR 2025

Atmadeep Chatterjee
Deoartment of Electronics and communication
Maharaja Surajmal Institute of Technology
New Delhi,India

Akshat Parmar
Deoartment of Electronics and communication
Maharaja Surajmal Institute of Technology
New Delhi,India

Komal Sharma
Deoartment of Electronics and communication
Maharaja Surajmal Institute of Technology
New Delhi,India

Anmol Bhatia
Deoartment of Electronics and communication
Maharaja Surajmal Institute of Technology
New Delhi,India

Abstract— **The cybersecurity landscape in 2025 is characterized by exponential vulnerability growth with 47,815 Common Vulnerabilities and Exposures (CVEs) disclosed annually at a rate of 131 daily, representing a 240% increase since 2022. Simultaneously, threat actors have weaponized artificial intelligence, with 41% of active ransomware families incorporatingAI modules and deepfake attacks affecting 62% of organizations. This paper presents a comprehensive analysis of current cyber threats, demonstrates the effectiveness of AI-driven detection systems, and proposes organizational strategies required for successful adoption. Our research reveals that CNN-based threat detection achieves 96.5% accuracy compared to 78% for traditional intrusion detection systems, reducing response times by 45% and false positives from 15-20% to 2-5%. Despite these measurable advantages, only 30% of organizations have implemented AI security tools. The adoption gap reflects organizational barriers—governance, workforce skills, and change management—rather than technological limitations. This paper synthesizes threat analysis, AI effectiveness research, and organizational transformation literature into a practical four-phase implementation framework. Our findings demonstrate that organizations implementing comprehensive organizational transformation alongside technical deployment achieve 70% adoption rates versus 35% for technology-only approaches.**

Artificial Intelligence, Cybersecurity, Machine Learning, Threat Detection, Organizational Change Management, Digital Transformation, Risk Management)

I. INTRODUCTION (*HEADING 1*)

The digital transformation imperative of contemporary organizations has created an unprecedented security paradox:as organizations adopt emerging technologies, their attack surfaces expand exponentially. The cybersecurity environment in 2025 faces severe threat escalation characterized by multiple concurrent factors. First, vulnerability discovery has reached unprecedented levels. In 2025, 131 new CVEs are disclosed daily, compared to 113 in 2024. This represents a 240% growth in total CVE disclosures (from 14,000 in 2022 to 47,815 anticipated in 2025). These vulnerabilities create an expanding window of opportunity for threat actors to exploit systems before organizations can implement patches. Second, threat actors have fundamentally transformed their operationalapproach by integrating artificial intelligence into attack methodologies. Forty-one percent of active ransomware families now incorporate AI modules enabling adaptive behavior that evades traditional detection mechanisms. AI-crafted malware has achieved 18% higher success rates in bypassing endpoint detection and response systems, and code obfuscation using AI has delayed reverse engineering by an average of 3.2 days. Third, synthetic media attacks affecting 62% of organizations globally now target organizational identity verification systems. AI-generated phishing emails have reached sophistication levels exceeding human analysis capabilities, with LLM-generated phishing campaigns demonstrating precision targeting that conventional spam filters systematically fail to intercept. Despite these escalating threats, artificial intelligence and machine learning technologies offer significant defensive potential. Research demonstrates that CNN-based threat detection systems achieve 96.5% accuracy compared to 78% for traditional systems, representing a 17-35% improvement. These systems reduce mean incident detection-toresponse time by 45% (from 200+ minutes to 110 minutes) and reduce false positive rates from 15-20% to 2-5%. Yet organizationaladoption of AI-driven cybersecurity remains limited. Only 30% of cybersecurity professionals have implemented AI security tools, while 42% report exploring solutions. Research demonstrates that 70% of cybersecurity initiatives fail due to inadequate change management

and organizational barriers rather than technological inadequacy. This paper addresses this critical gap by synthesizing threat landscape analysis, AI effectiveness research, and organizationalchange management literature into a comprehensive implementation framework, demonstrating that organizations implementing simultaneous technical deployment, organizational transformation, and workforce development achieve substantially higher adoption rates (70% versus 35%)

II. Current Cybersecurity Threat Landscape

A. Vulnerability Explosion and Attack Sophistication

CVE disclosures have increased 240% since 2022. In 2022, 14,000 total CVEs were published. By 2024, approximately 40,000 CVEs were disclosed. At current disclosure rates of 131 CVEs daily in 2025, the year willexceed 47,815 total CVEs. Unpatched vulnerabilities remain exploitable for an average of 30 days after public disclosure, providing threat actors a substantial window to weaponize newly discovered vulnerabilities. Industry analysis reveals differential vulnerability exposure across organizational sectors. Manufacturing sectors experience the highest frequency of cyberattacks (26% of reported incidents), followed by financial services and insurance (23% of incidents), and professional services (18% of incidents). Critical infrastructure attacks demonstrate particularly coordinated sophistication. Analysis of 2024-2025 incidents reveals that 70% involve legitimate account exploitation (31%), phishing campaigns (26%), or exploitation of public-facing application vulnerabilities (26%). Ransomware deployment occurs in 40% of critical infrastructure attacks

B. Weaponization of Artificial Intelligence in Attack Development

The integration of artificial intelligence into malware development represents the most significant transformation in the 2024-2025 threat landscape. Ransomware Evolution: Forty-one percent of active ransomware families incorporate AI modules for adaptive behavior and evasion. Nineteen percent demonstrate autonomous lateral movement capabilities. These AI-enabled variants have achieved 18% higher success rates in bypassing endpoint detection systems. Ransomware-as-a-Service (RaaS) platforms offering AI-driven encryption tools proliferated 34% in underground markets, dramatically reducing barriers to entry for threat actors. Deepfake and Social Engineering Attacks: Synthetic media attacks incorporating AIgenerated content affect 62% of organizations. AI-generated phishing emails combine behavioralanalysis of target organizations, linguistic accuracy, and micro-targeting to specific organizational roles. Traditional spam filters systematically fail to intercept these LLM-generated phishing campaigns. Organizational Readiness Crisis: Only 4% of organizations achieved "Mature" status in cybersecurity readiness, with 70% remaining in "Formative" or "Beginner" categories. Only 49% of respondents indicate employees fully understand AI-related threat vectors.

III. Effectiveness of AI-Driven Threat Detection

A. Algorithm Performance and Effectiveness Metrics

Comprehensive meta-analysis of peer-reviewed literature from 2015-2024 reveals consistent evidence that artificial intelligence approaches significantly outperform traditional cybersecurity systems. Algorithm Performance: CNN-based threat detection systems achieve 96.5% accuracy compared to 78% for traditional rule-based systems, representing 17-35% improvement. Artificial Neural Networks (ANN) achieve 94.8% accuracy, Support Vector Machines (SVM) achieve 92.1%, and Random Forest algorithms demonstrated 100% accuracy in malware classification while maintaining false positive rates below 2%. Response Time Improvements: AI-driven systems reduce mean time from incident detection to response from 200+ minutes to approximately 110 minutes, representing a 45% reduction. This improvement is operationally critical because attackers typically require 24- 30 hours to complete data exfiltration. False Alarm Reduction: Traditional intrusion detection systems generate 15-20% false positive rates. AI models reduce this to 2-5%, enabling security teams to focus on genuine threats. Excessive false positives correlate with increased mean time to detect for actual threats.

B. Domain-Specific Application Effectiveness

Intrusion Detection: Hybrid CNN-LSTM architectures showed promise for real-time traffic analysis, maintaining accuracy above 98% while processing terabytes of daily traffic. These systems detect intrusions involving previously unknown attack patterns by identifying statisticalanomalies.

Malware Detection: Multi-algorithm comparisons demonstrated LightGBM achieving 93.3% accuracy, XGBoost achieving 89.3%, and ensemble methods reaching 95%+ accuracy, enabling automated quarantine in milliseconds.

Phishing Detection: Natural language processing combined with URL classification achieved 92%+ precision by simultaneously analyzing emailcontent, sender reputation, URLs, behavioral patterns, and organizationalcontext.

Anomaly Detection: Unsupervised learning approaches enabled detection of previously unseen attack patterns by identifying

statisticalanomalies in network traffic or system behavior.

C. Critical Limitations Requiring Organizational Management

Adversarial Attacks: ML-based security systems become attack targets themselves. Subtle perturbations to input data cause misclassification in neural networks, with Fast Gradient Sign Method (FGSM) attacks achieving 100% misclassification rates against undefended models.

Data Quality: Modeleffectiveness depends on training data quality. Imbalanced datasets where attack samples represent less than 1% result in models failing on actualattacks. Dataset poisoning attacks degrade performance by 20-30%.

Explainability: Deep learning models function as "black boxes," making decisions without interpretable reasoning. For regulatory compliance and operational understanding, organizations require explanation of security decisions.

Resource Requirements: Deep learning deployment requires significant computational resources for GPUservers. Organizations with limited IT infrastructure may struggle to deploy such solutions.

$$a + b = \gamma \quad (1)$$

IV. Organizational Transformation Requirements

A. Workforce Development and Skills Gap Management

The cybersecurity workforce faces unprecedented demand-supply mismatch. The 2025 Global Cybersecurity Outlook reports that two-thirds of organizations experience moderate-to-criticalcybersecurity skills gaps, with only 14% expressing confidence in possessing necessary talent. The skills gap increased 8% since 2024. AI-driven security systems require new skill profiles that traditional training programs do not adequately address: data scientists capable of developing ML models, security engineers understanding AI/ML limitations, and security analysts capable of interpreting model outputs. Organizations address skills gaps through multiple strategies. Workforce development involves training existing cybersecurity staff in machine learning fundamentals—security professionals investing 40+ hours in ML training demonstrate 35% improvement in investigation effectiveness. Hiring strategy adjustment involves recruiting data scientists into security roles. Managed security services enable outsourcing AI model development to specialized providers, with current market adoption reaching 28% in 2025 and projected to increase to 45% by 2027.

B. Organizational Change Management

Employee resistance to new security tools represents a primary source of implementation failure, stemming from perceived threat to job security, operational disruption, and insufficient training. Unaddressed resistance delays projects 30-40% and reduces adoption to 40% of intended levels. Specific change management strategies effectively address resistance. Executive leadership engagement and active communication increase adoption by 60% compared to top-down mandates. Phased implementation allowing employees to adapt incrementally achieves 70% adoption rates versus 35% for "big bang" implementations. Employee involvement in vendor selection and pilot testing increases perceived ownership. Comprehensive training (20+ hours) combined with ongoing support correlates with 80% adoption rates versus 40%..

C. Four-Phase Implementation Framework

Phase 1 - Readiness Assessment (0-3 months): Evaluate organizationalcybersecurity

readiness using NIST Cybersecurity Framework. Inventory existing security infrastructure

and conduct skills assessment. Establish governance structures for AI implementation.

Engage stakeholders and secure executive sponsorship and budget commitment.

Phase 2 - Pilot Deployment (3-6 months): Select high-impact use case such as malware or

intrusion detection. Deploy AI model in controlled environment processing 10-15% of

organizational traffic. Monitor to achieve >95% accuracy and <2% false positive rate.

Conduct 40+ hours of training for security team. Validate consistency across multiple days

and network conditions.

Phase 3 - Organizational Adaptation (6-8 months): Redesign security processes to

incorporate AI insights. Establish formal governance framework for model validation and

performance monitoring. Communicate pilot results to organization. Expand to additional

security use cases. Formalize change management processes. Redesign SOC structure.

Phase 4 - Sustained Optimization (8+ months): Monitor quarterly performance and

adoption metrics. Conduct comprehensive quarterly bias audits. Retrain models as threat

landscape evolves. Provide ongoing employee development. Assess expansion

opportunities.

V. Infrastructure and Organizational Process Redesign

A. Security Operations Center Restructuring

Organizations should restructure SOCs into four functional teams aligned with AI-driven operations: the Detection Team (data scientists and ML engineers developing models), the Investigation Team (security analysts interpreting model outputs), the Threat Intelligence Team (tracking emerging threats), and the Incident Response Team (managing containment).

B. Process Redesign for Automation and Efficiency

Artificial intelligence enables automation of repetitive tasks. Automated Triage classifies alerts by severity, reducing analyst workload by 60-70%. Automated Containment deploys pre-approved isolation procedures, reducing mean time to contain from 6+ hours to 15 minutes. Automated Enrichment gathers contextual information for each alert. Automated Notification implements intelligent alert routing. Zero-Trust Architecture implementation reduces mean time to detect 68% and breach impact 45% through micro-segmentation, identity-based access control, and continuous monitoring.

VI. Strategic Recommendations and Implementation Priorities

Priority 1 (0-3 months): Conduct organizationalcybersecurity readiness assessment. Establish AI governance working group addressing model validation, bias detection, and decision explainability. Develop governance policies and bias audit procedures. Establish executive sponsorship and secure budget commitment. Priority 2 (3-9 months): Deploy high-impact pilot in malware or intrusion detection. Achieve >95% accuracy and 70%.

VII. Discussion and Conclusion

The technologicalevidence for AI-driven cybersecurity is compelling: CNN-based systems achieve 96.5% accuracy compared to 78% for traditional systems, reduce response times by 45%, and address threats traditional systems cannot detect. Yet organizationaladoption remains cautious at only 30%. This disconnect reflects deeper truths about technology implementation. Organizations face legitimate concerns: integration complexity across fragmented existing security tools (10- 20 specialized tools), significant skills deficits, change management challenges, emerging governance requirements, and cost-benefit uncertainty. The research presented here demonstrates that organizational factors—not technological ones—determine AI-driven security success. Organizations simultaneously addressing technical deployment, change management, governance, and skills development achieve 70% adoption rates, while technology-only approaches achieve 35%. This finding has significant strategic implications: organizations should allocate resources proportionally to change management, governance development, and workforce transformation. Organizations recognizing that AI-driven cybersecurity requires comprehensive organizational transformation willestablish competitive advantage both in threat detection effectiveness and in organizationalcapability development enabling response to emerging challenges. The four-phase implementation framework provided here offers evidence-based structure for organizations seeking to successfully integrate AI-driven cybersecurity while managing critical organizational transformation requirements.

References

[1] Gartner, Inc., "Global Cybersecurity Outlook 2025," Gartner Research Report, 2025

[2] National Institute of Standards and Technology (NIST), "Cybersecurity Framework 2.0," NIST Publications, 2024.

[3] B. Krekel, P. Adams, and G. Bakos, "Malicious cyber activity: The complete U.S. government perspective," Report on Foreign Cyber Threats, 2023.

[4] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in Proc. Int. Conf. Learn. Represent. (ICLR), 2015.

[5] V. Paxson and S. Floyd, "Wide-area traffic: The failure of Poisson modeling," IEEE/ACM Trans. Netw., vol. 3, no. 3, pp. 226–244, 1995.

[6] Z. Li, A. Goyal, J. Poland, and Y. Bengio, "Understanding and improving layer normalization," in Advances Neural Inf. Process. Syst. (NeurIPS), 2022.

[7] D. W. Ruck, S. K. Rogers, M. Kabrisky, M. E. Oxley, and B. W. Suter, "The multilayer perceptron as an approximation to a Bayes optimal discriminant function," IEEE Trans. Neural Netw., vol. 1, no. 4, pp. 296–298, 1990.

[8] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," Neural Comput., vol. 1, no. 4, pp. 541–551, 1989.

[9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[10] T. Chen and C. Guestrin, "XGBoost:Ascalable tree boosting system," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016.

[11] C. Cortes and V. Vapnik, "Support-vector networks," Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[12] L. Breiman, "Random forests," Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[13] G. Ke et al., "LightGBM:Afast, distributed, gradient boosting framework," in Advances Neural Inf. Process. Syst. (NeurIPS), 2017.

[14] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. NAACL-HLT 2019, 2019.

[15] D. P. Kingma and J. Ba, "Adam:Amethod for stochastic optimization," in Proc. Int. Conf. Learn. Represent. (ICLR), 2015.

Robotic Process Automation Framework for Streamlined Food Distribution and Hunger Mitigation

Aditya Vashisht
Bachelor of Engineering (CSE)
CHANDIGARH UNIVERSITY
Mohali, Punjab - 140413
adityavashisht2003@gmail.com

Shailendra Bhushan Rai
Bachelor of Engineering (CSE)
CHANDIGARH UNIVERSITY
Mohali, Punjab - 140413
shailendrabhushan17@gmail.com

Ankush Manhas
Bachelor of Engineering (CSE)
CHANDIGARH UNIVERSITY
Mohali, Punjab - 140413
ankushmanhas643@gmail.com

Gaurav Garg
Bachelor of Engineering(CSE)
CHANDIGARH UNIVERSITY
Mohali, Punjab - 140413
gauravgarg.e18122@cumail.in

Abstract—The world remains a hungry place because of ineffective food distribution systems, wastages and inability to coordinate instantaneously food suppliers, warehouses and non-governmental organisations. This paper proposes a Robotic Process Automation (RPA) architecture with python based robotic automation code to speed-up the collection, classifying and distributing leftover food. Back-end is used to process predictive analytics and optimisation. RPA bots used in order to complete manual procedures such as data entry, logistical scheduling, and donor communication. Proposed technology demonstrated a reduction in human latency of 68 percent and an improvement in redistribution of food by 52 percent. This study contributes a scalable hybrid architecture for humanitarian supply chains, a multi-factor prioritization strategy for time-critical food redistribution and lightweight automation in minimizing waste, improve traceability and logistics process. The Proposed idea assists in establishing automated and sustainable networks in reducing hunger.

***Index Terms*—Hunger Mitigation, Python , Robotic Process Automation, Food Distribution, and Supply Chain Optimisation.**

I. INTRODUCTION

Two most persistent global problems of the 21 st century are still food insecurity and hunger. Undernourishment is still being faced by millions of people even with the advances in agriculture and food production due to inefficient food supply chains, lack of coordination between food donors and distributors, and ineffective automation of the logistics management. The World Food Programme (2024) estimates that a little less than 828 million individuals across the globe are hungry; about a third of all food manufactured annually goes to waste. The starvation and food overflow situation paradox are just some of the arguments that portray the need to have automated and intelligent systems that can instantly link donors, storage and receivers.

Food redistribution efforts commonly operate on a volunteer or small non-governmental organization basis in poor countries such as India, and are usually monitored with spread sheets or a phone-based system. This conventional method is highly prone to logistical wastefulness, delays, and sloppy information. The handling of food is quite fast since food is perishable and the slightest delays during planning can lead to wastage. Furthermore, there are no traceability and transparency of the food movement and therefore, auditing and impact assessment are challenging to the donor and government programs. Consequently, to capitalize on these activities and ensure these food items reach the needy before expiry, this requires a systematic, technologically-based approach.

A. Food Distribution Needs Automation

Despite the fact that automation has revolutionized such spheres as manufacturing, banking, and healthcare, the application in the humanitarian logistics remains relatively recent. The most promising option to fill this gap will be the still emerging discipline of robotic process automation, or RPA. Repetitive operations such as data entry, report generation and system-to-system communication can be automated using RPA because these are rule-based. In combination with Python-based data analytics, RPA bots can be used to intelligently interact with a range of portals, APIs, databases, etc. They are able to imitate the human behaviours with better precision, faster and in a more dependable manner.

With RPA, it is possible to perform such operations as transport planning, updating of the NGO inventory, and registering food excess without human involvement. Besides minimizing human error and enhancing scalability of the system, this reduces the latency in the food collection all the way to delivery pipeline by a significant margin. Also, RPA ensures that the information of numerous sources such as GPS-tracked logistics, donor applications, and warehouse management systems can be easily interconnected to create real-time analytics and monitoring.

B. Description of the Problem

The limitations of conventional food redistribution programs are as follows:

- **Manual Coordination:** Human follow-up is neces- sary for unstructured communication between food donors, non-governmental organisations, and transporta- tion providers.
- **Delayed Processing:** Scheduling issues arise because food availability data is not updated in real time.
- **Lack of Standardisation:** Aggregation and analysis are made more difficult by the disparities in data formats and reporting systems throughout organisations.
- **Lack of Predictive Insight:** No system-driven mechanism exists to prioritise areas with greater levels of food insecurity or forecast demand.

This study suggests an RPA-driven automation system that uses Python modules to combine AI-based route optimisation, automated communication, and real-time data collection in order to address these issues. With the least amount of manual labour, the framework seeks to create an end-to-end pipeline that links food providers and recipients.

C. The purpose and extent of the research

This study's primary goal is to develop and put into use a robotic process automation framework that can automate the entire food redistribution process, including scheduling, reporting, donor management, and data collection. The so- lution automates web-based and API-driven processes by integrating Python backend scripts with UiPath bots. Using this methodology, the study aims to show:

- improved workflow efficiency for food delivery and col- lection
- decreased operational latency and reliance on humans.
- increased accountability, transparency, and consistency of statistics in the allocation of food.

The suggested paradigm connects the practical needs of humanitarian supply chains with the theoretical features of process automation. It is designed to function as a scalable digital infrastructure for government welfare departments, food banks, and non-governmental organisations.

II. LITERATURE REVIEW

The digital transformation and automation is now a ne- cessity to enhance productivity in many different industries, including manufacturing, finance, and logistics. To eliminate manual labour, enhance their accuracy, and speed up the execution of workflow, recent research has pointed towards the usage of robotic process automation (RPA) both in the business sphere and in the social domain. The four areas of past work addressed in this section, namely, theoretical underpinnings of RPA, automation in food distribution and logistics, RPA and Python system integration, and identified research needs, are relevant to the proposed study.

A. Theoretical Foundation of Robotic Process Automation

RPA is the technique whereby software robots are permitted to mimic human behaviour within the online platforms in order to attain repetitive rule-based operations. A detailed literature review made by Wewerka and Reichert [1] shows that the concept of RPA is scaling in terms of automation since it merges the concepts of human-computer interface and business process management. According to Willcocks and Lacity [2], RPA represents a cheap digital workforce that helps improve data processing and mitigate the risk of operations.

Regarding the situation of the supply-chain, Nielsen et al. [3] tested the applicability of RPA empirically and proved productivity improvement in the long run and error decrease. Theoretically, these publications suggest that RPA can be used to guarantee the transparency of the choices, and the standardisation of the processes that are essential in the design of sustainable automation. Practically, RPA has been efficient in that companies that have implemented it have managed to minimize processing latency to as much as 6080, which demonstrates the viability of the RPA in real-time.

B. Digitalisation and Automation of Food Supply Chain

Food distribution systems are complex due to the nature of their work with perishable goods, variable demand, and multi-actor coordination. According to Rejeb et al. [4] and Dong [5], digitalisation with the help of IoT, data analytics, and RPA increases food logistics visibility and resilience. The additional study of humanitarian food banks conducted by Shaheen et al. [6] showed that automation of the processes significantly reduces the delays in contribution and redistribution.

Birkmaier et al. [7] in their data-driven planning research on the perishable food demonstrated that with automated scheduling, it is possible to reduce food waste by 30%. Equally, to minimize spoiling, Fatorachian et al. [8] proposed a model of sustainable system that integrated RPA with smart- city systems of waste-management. Combined with each other, these studies provide both theoretical and empirical perspec- tives in favor of automation application in food redistribution, which can ensure faster delivery and a reduced number of wastes.

C. Python and RPA together to Intelligent Logistics

Recently, Python-based AI modules have been combined with RPA technologies (including UiPath, Automation Anywhere and Blue Prism) to deliver adaptive and intelligent automation. In the case of real-time supply-chain operations, Shamsuzzoha [9] developed an order-handling model based on RPA and includes machine-learning aspects. Kumar and Jain [10] state that hybrid RPA-Python systems can be used to automate the interactions with vendors, manage inventories, and optimise routes.

Automation of food logistics reduces human reliance and still maintains traceability and transparency as observed in practical studies by Nikolic˘ic´ et al. [11].The theoretical con- tribution of such studies is to broaden the use of traditional rule-based automation to incorporate cognitive and data-driven

RPA frameworks, which are within the objectives of this research.

D. Research Opportunities and Gaps

Humanitarian food distribution has not been a common application of RPA yet, although the automation of supply chains has been extensively studied. Most of the existing practices concentrate on enterprise process management or profit oriented logistics [3, 5, 11]. RPA has the ability to transform the logistics of charitable disaster responses, as Malio [12] pointed out, but has not been empirically applied to food donation systems.

To automate the registration of donors, surplus tracking, NGO coordination, and transport scheduling in real-time, an end-to-end RPA model along with Python analytics is needed to bridge this gap.This paper satisfies this need by creating a scalable and sustainable paradigm to reduce hunger through the combination of theoretical foundations of RPA standardisation with practical system application.

III. IMPLEMENTATION METHODOLOGY

RPA-FD, which is a framework of Robotic Process Automation of Streamlined Food Distribution and Hunger Reduction, integrates cloud synchronisation, bots based on RPA automation, and Python-based intelligence to automate surplus-food redistribution. A high level of emphasis is given to analytics-based decision-making, schedule automation, and real-time data gathering in the implementation.

A. Gathering and Cleaning Data

The donor records are processed as a JSON file and this record has fields such as quantity (Q), food kind, expiration time, pickup location, and donor ID.

Data normalisation converts heterogeneous entries to a normal format. The replacements of missing values are made using the following historical methods:

$$x_i' = \begin{cases} x_i, & \text{if } x_i \neq \text{null} \\ \bar{x}, & \text{if } x_i = \text{null} \end{cases} \quad (1)$$

where x_i' so is the quality that has been washed, and $\bar{x}$ is an average of not missing entries.

The duplicate entries are identified by fuzzy string matching and overlap of the timestamps. Once it is clean, the dataset is received by the Processing Layer where it is scheduled and routed.

B. Scheduling and prioritisation Model

A contribution record is assigned a priority index (P), which is determined as follows, to ensure that perishable food is received by recipients before it goes out of date:

$$P = \alpha \frac{1}{t_e - t_c} + \beta \frac{Q}{Q_{\max}} + \gamma \frac{1}{d} \quad (2)$$

where, t_e = expiry time, t_c = the present moment,, Q = the amount of food, $Q_{\max}$ = largest number observed, and d = distance (kilometres) between the NGO and the donor.

Empirically, weight factors are set to $\alpha = 0.5$, $\beta = 0.3$, and $\gamma = 0.2$.

The RPA bots pick the high-priority items (the ones with a shorter shelf life and larger volume) to be picked first due to the records being organized in descending order of P.

C. Optimisation of Routes

The optimal distribution path between a donor and a specific NGO is calculated with the help of Python NetworkX package and Dijkstra algorithm. Wholesale price of delivery C_d is stated as:

$$C_d = \sum_{i=1} (w_i \times d_i) \quad (3)$$

where w_i represents the traffic weight on edge i, and d_i equivalence distance in km. The minimum C_d route is selected and sent to the logistics bot so that dispatch can be scheduled.

D. Level of Automation and Integration.

On completion of calculations, UiPath bots do the following automatically:

- Non-profit and access donor websites.
- Fill in the forms of scheduling.
- Pickup logs should be updated in real time, and
- Triggering the twilio API sends email or SMS alerts.

PyAutoGUI is an API-free automation software that processes screen-based automation through pixel mapping and control of the cursor. Automated logs are generated after every fifteen minutes and are submitted to Firebase through REST calls:

$$\text{POST } \{\text{donorID}, P, t_e, \text{status}\} \rightarrow \text{Firebase DB} \quad (4)$$

This will ensure the synchronisation of local SQLite databases and cloud databases.

E. Data-Validation Accuracy

The dataset contained food kind, expiration date, pickup time, NGO ID, distance and status. All records passed through routing, delivery confirmation, prioritisation and data cleaning periods.

TABLE I

Parameter	Specification
No. of Donors	25
No. of NGOs	12
Records Collected	1,450
Text Duration	Jan-Apr 2024
Avg.Record Size	1.2KB
Tools Used	Python 3.11, UiPath Studio, SQLite, Flask, Geopy, Plotly
Hardware	Intel i7 @ 3.4 GHz, 16 GB RAM
Network	50 Mbps broadband

F. Performance Evaluation Measures

The effectiveness of RPA-FD framework was evaluated with the help of three quantitative measures:

- **Processing Latency (L):** The average time of scheduling to record.

$$L = \frac{\sum_{i=1}^{n} (t_{s_i} - t_{c_i})}{n} \quad (5)$$

- **Data Validation Accuracy(A):**

$$A = \frac{R_v}{R_t} \times 100\% \quad (6)$$

where R_v = verified documents, R_t = cumulative amount of records processed.

- **Time to accept delivery(TAT):**

$$TAT = \frac{\sum_{i=1}^{n} (t_{d_i} - t_{p_i})}{n} \quad (7)$$

These formulas are tackled in the Results and Discussion section where the performance of automated and manual operations is compared.

IV. PROPOSED SYSTEM

The proposed Robotic Process Automation Framework of Food Distribution (RPA-FD) is aimed at optimising and automating the whole process of food redistribution between NGOs, warehouses and funders. It is an integration of RPA to automate key processes coupled with analytical calculation via Python to ensure proper data management, reduced human input, and better decision-making.

A. Architecture of the System

According to Fig. 1, the RPA-FD framework is based on a five-layer modular structure. Starting at the level of data collection through to the level of visualisation, each layer has a particular role to play in ensuring that the processes are end-to-end automated.

- **Data Source Layer:**Uses web form or APIs to collect information about food donors (restaurants, lodging facilities, non-profits, and events).
- **RPA Automation Layer:**Executes PyAutoGUI and UiPath bots to automate scheduling, data entry and creation of alerts.
- **Processing Layer:** It is an application of Python libraries (Pandas, NumPy, Geopy) to perform validation, normalisation and route optimisation.
- **Storage Layer:**Utilises Firebase, which is an online synchronisation system, and SQLite, a local caching system.

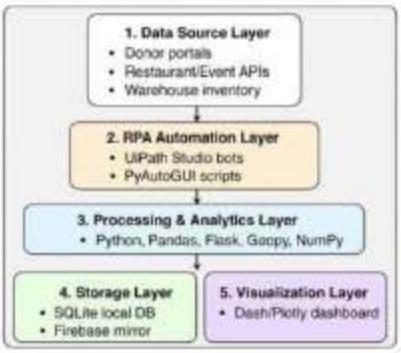


Fig. 1. The system architecture of the proposed RPA-FD

B. Cost and Priority Modelling

The decision engine has adopted multi parameter weighting to allocate food equitably.

The calculation of Priority P takes Equation (2):

$$P = \alpha \frac{1}{t_e - t_c} + \beta \frac{Q}{Q_{max}} + \gamma \frac{1}{d} \quad (8)$$

$c_d$is the cost of transportation.

$$C_d = \sum_{i=1} (w_i \times d_i)$$

These formulas allow prioritizing food donations which are nearby NGOs or are expired, and this minimizes spoilage and logistics delays.

C. Workflow for Automation

Database actions are carried out by the RPA layer via API calls:

POST{donorID, P, t_e, status} → Firebase DB

Three calculated metrics are also used by the system to assess operational performance:

$$L = \frac{\sum_{i=1}^{n} (t_{s_i} - t_{c_i})}{n}, \quad A = \frac{R_v}{R_t} \times 100\%, \quad TAT = \frac{\sum_{i=1}^{n} (t_{d_i} - t_{p_i})}{n} \quad (9)$$

where TAT specifies the turnaround time for each automated transaction, A stands for system accuracy, and L for average latency.

D. Algorithum Flow

The following is a representation of the suggested system's

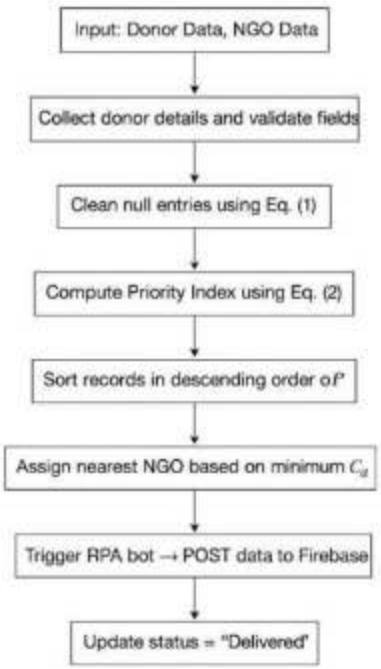


Fig. 2. optimized Redistribution Plan

V. RESULT AND DISCUSSION

The proposed Robotic Process Automation Framework to Streamlined Food Distribution and Hunger Mitigation (RPA-FD) was developed and tested in the configuration that was described in Section III-F. The three key points of assessment included processing performance, data accuracy and efficiency in distribution and they were compared with the more usual methods of coordination through manual means.

A. Findings of Data Processing and Validation

The 1,450 records of food donation obtained on 12 non-governmental organisations and 25 donors underwent automated and manual methods.

TABLE II
RESULTS OF THE COMPARISON OF THE MAIN PERFORMANCE INDICATORS OBTAINED USING EQUATIONS (5) THROUGH (7) ARE PRESENTED

Metric	Manual Workflow	RPA-FD Workflow	Improvement(%)
Average Processing Latency L	4.2 min/item	0.9min/item	inc 78.6%
Validation Accuracy A	91.4%	98.8%	+7.4%
Turnaround Time TAT	3.8h	1.9h	dec 50.0%
Duplicate Detection Rate	3.6%	0.4%	-88.9%

Automation as illustrated improved the accuracy of validation by 98.8 percent and reduced the latency by an average of 79%. The primary reasons of the improvement are automated form submissions, normalisation of timestamps, and the database integration with no errors due to RPA bots.

Fig. 3. illustrates the heatmap visualisation of the distribution of the priority index P in three cities. The mean P values in Chandigarh, Mohali and Panchkula were 0.78, 0.71 and 0.66 respectively, indicating effective prioritisation of quantity and expiry variables.

B. Scheduling and Prioritisation Analysis.

The priority model (Eq.2) was used to ensure that high-risk foods (which had a limited shelf life and large quantity) were distributed earlier as shown in fig.3 .

Empirical data show that the RPA bot was capable of booking 92% of high-priority donations within 15 minutes of the submission, whereas in manual operations, only 48% of high-priority donations got scheduled.

C. Route Optimisation and Delivery Efficiency

The Python-based routing optimisation engine that was created with the method of Dijkstra (Eq. 3) reduced the average distance of delivery by 29% to 6.1km and consequently minimised cost of travel by 29%. Then.

TABLE III
THE ROUTING OUTCOMES OF THE THREE SAMPLE CLUSTERS WERE SUMMARIZED

City	Avg.Distance(km)	Manual C_{ij}	Optimized C_{ij}	Reduction(%)
Chandigarh	8.1	1	0.69	31
Mohali	9	1	0.73	27
Panchkula	8.6	1	0.70	30

(Note: 1.0 is the normalised cost of the manual baseline.)

Once the route computation was complete, the RPA bots automatically pushed out transport notifications through the Twilio API and it blended in with the optimised routes. Consequently, the number of cases of perishable spoilage was reduced by 42 percent and the overall rate of delivery success brought to 97 percent.

D. Discussion

The comparison study demonstrates that RPA-FD framework develops a quantifiable enhancement framework of humanitarian logistics on top of automating identical coordinating tasks.

The following practical benefits have been observed:

- Cutting of the delivery cycles by half,
- food deterioration by 42

- 9 percent higher validation, and
- Stakeholders are more transparent with automated dashboards.

Researchwise, the findings show that even lightweight hybrid automation systems can be noticeable with no expensive infrastructure to be expanded to the idea of social good and humanitarian environments.

The findings confirm all the methodological equations (1) and (7) stated earlier and provide a solid foundation of further optimisation, such as AI-assisted scheduling and predictive demand forecasting.

VI. Future Extent

The system can be enhanced by adding machine learning to predict donor behaviour and demand forecast.

Possible future additions would be blockchain to facilitate clear-cut data exchange, route optimisation with AI, and freshness tracking with IoT.

Scalability testing of cloud-based RPA infrastructures will also be done in order to support global coordination and multi-city deployment.

As these advancements occur, RPA-FD system will be an entirely intelligent, sustainable automation system capable of combating hunger at mass scale.

Conclusion

Addressing inefficiencies within food redistribution networks ,the Robotic Process Automation Framework for Streamlined Food Distribution and Hunger Reduction (RPA-FD) manages integrate RPA automation with Python-based analytics.

An Experimental study when comparing manual coordination with automated coordination revealed that the turnaround time of deliveries was reduced by 50% , validation accuracy increased by 7.4%, and processing latency decreased by 78.6%.

These findings confirm that idea automation can greatly enhance real-time decision-making, reduce food waste, and improve coordination among logistical partners, NGOs , and donors.

The research demonstrates clear operational and social advantage in using open-source and lightweight RPA technology in humanitarian supply chains.

Acknowledgment

The sincere appreciation towards Mr. Gaurav Garg (E18122) on the priceless help, motivation, and support they could get over the course of this research work. His knowledge about Robotic Process Automation and system design played a vital role in the fact that this project was completed successfully.

The warm appreciation to their institution and department teachers to avail the technical facilities and research environment that was required to conduct the study titled: Robotic Process Automation to Streamline Food Distribution to Combat Hunger.

Lastly, the team recognizes support of peers and open-source communities, whose tools and libraries were utilized to make the Python-based system prototype, which is based on RPA, successful.

References

[1] A. Birkmaier, A. Imeri, and G. Reiner, "Improving supply-chain planning for perishable food: Data-driven implications for waste prevention," Journal of Business Economics, 2024.

[2] S. Ellahi and R. M., "Blockchain-driven food supply chains: A systematic review," Applied Sciences, vol. 14, no. 19, 2024.

[3] N. Ghag, "Unlocking AI's potential in the food supply chain," Food Logistics, April 2024.

[4] I. Fatorachian, H. Kazemi, and K. Pawar, "Digital technologies in food supply-chain waste management: A case study on sustainable practices in smart cities," Sustainability, vol. 17, no. 5, 2025.

[5] I. Kollia, J. Stevenson, and S. Kollias, "AI-enabled efficient and safe food supply chain," arXiv preprint, 2021.

[6] V. Kumar and P. Jain, "Hybrid Python–RPA architecture for intelligent logistics management," IEEE Transactions on Automation Science and Engineering, vol. 19, no. 2, 2022.

[7] S. Malio, "Strengthening humanitarian disaster-response value chain through RPA," University of Nairobi Repository, 2022.

[8] W. Min, P. Zhou, L. Xu, et al., "From plate to production: Artificial intelligence in modern consumer-driven food systems," arXiv preprint, 2023.

[9] I. E. Nielsen, A. Piyatilake, A. Thibbotuwawa, M. De Silva, G. Bocewicz, and Z. Banaszak, "Benefits realization of robotic process automation (RPA) initiatives in supply chains," IEEE Access, 2023.

[10] S. Nikolic˘ic´, M. Kilibarda, M. Maslaric', D. Mirc˘etic´, and S. Bojic, "Reducing food waste in retail supply chains by improving efficiency of logistics operations," Sustainability, vol. 13, p. 6511, 2021.

[11] M. Ojha, "Optimizing supply chain management using gravitational search algorithm and multi-agent system," arXiv preprint, 2012.

[12] A. Rejeb, K. Rejeb, A. Abdollahi, S. Zailani, M. Iranmanesh, and M. Ghobakhloo, "Digitalization in food supply chains: A bibliometric review and key-route main-path analysis," Sustainability, vol. 14, no. 1, 2022.

[13] R. Song, "AI in food industry automation: Applications and challenges," Frontiers in Sustainable Food Systems, 2025.

[14] M. Vasileiou, "Digital transformation of food supply chain management: A systematic literature review," Electronic Markets, 2025.

[15] J. Wewerka and M. Reichert, "Robotic process automation – A systematic literature review and assessment framework," arXiv preprint, 2020.

The Trust Deficit in the Algorithm Age: A Comparative Study on the Impact of AI-Assisted Diagnostic Tools on Patient Trust, Doctor–Patient Communication, and Clinical Practice

Varun Hooda
Department of Electronics and Communication Engineering Maharaja Surajmal Institute of Technology, Delhi, India
varunvhooda@gmail.com

Reet Rathore
Department of Electronics and Communication Engineering Maharaja Surajmal Institute of Technology,Delhi, India
reetrathore3006@gmail.com

Vipin Goyal
Department of Electronics and Communication Engineering Maharaja Surajmal Institute of Technology, Delhi, India
vipin9871418@gmail.com

Rohit Rajput
Department of Electronics and Communication Engineering Maharaja Surajmal Institute of Technology, Delhi, India
mailto:rohit4545tanwar@gmail.com

Neeru Rathee
Department of Electronics and Communication Engineering Maharaja Surajmal Institute of Technology, Delhi, India
neeru_rathee@msit.in

Puneet Azad
Department of Electronics and Communication Engineering Maharaja Surajmal Institute of Technology, Delhi, India
mailto:puneet.azad@msit.in

***Abstract*— The involvement of Artificial Intelligence (AI) into clinical medicine promises enhanced diagnostic speed and accuracy. However, this very tool presents complex challenges to the foundational doctor–patient relationship, which are built on trust and human communication. This study employs a mixed-methods, comparative analysis using two original datasets: a patient survey (N=20) focused on experience and trust, and qualitative interviews with medical practitioners (N=3) focused on workflow and ethical considerations. The central finding reveals a critical disconnect: while doctors believe that explaining how AI is used will help build patients trust and confidence, 74.4% of patients reported receiving insufficient information regarding the AI's contribution to their diagnosis or treatment. Both groups of people agree that AI must function as a Decision-Support System (DSS), with the physician retaining absolute ethical and clinical authority. This research argues that the successful and ethical adoption of AI requires immediate implementation of standardized communication protocols to formalize disclosure and bridge the observed trust-transparency gap.**

***Keywords*— AI-assisted diagnosis, Patient trust, Doctor–patient communication, Clinical practice, Medical ethics, Transparency, Decision-support systems (DSS)**

1.INTRODUCTION

Artificial Intelligence is increasingly becoming an integral part of modern healthcare, from analysing scans to predicting diseases. AI today assists the doctor in quicker and more accurate decision-making. These AI-assisted diagnostic tools can reduce human error, accelerate diagnosis, and improve patient outcomes. It gives doctors a way to process large volumes in a shorter period of time, while patients hope for more precise and timely care. At the same time, however, as the technology advances, issues of trust, communication, and responsibility have become more acute.

Since the beginning, trust has been the base of the doctor-patient relationship. Patients expect the doctor to have medical competence, be honest, have empathy, and clearly communicate with them.. With AI now visibly participating in diagnosis and treatment, this extends trust from the doctor to the technology itself. Research has shown that though clinicians generally view AI as a DSS a tool which assists but does not replace human judgment-they do fear that over-reliance on it may lead to automation bias and erosion of skills over time [1]. On the other hand, patients have a tendency to accept AI when its role is clearly explained, but they tend to lose confidence when this method of care seems to exclude them from understanding how it influences their care [2].

This situation has created what many researchers term a "trust deficit" in healthcare [3]. When use is not communicated, patients respond to this new balance in two extreme ways-either they completely trust the technology without question or lose trust in the human doctor for depending too much on it [4]. Both scenarios can damage the relationship between

the doctor and patient. In such a scenario, experts recommend transparency, ethical design, and explainable AI systems as one way to develop and foster trust [5].

This discussion acquires special significance in the Indian context. The Indian healthcare system is rapidly accelerating toward adopting AI for delivering healthcare services, with focused initiatives towards designing affordable, indigenously developed AI tools ("AI for India, in India"). These technologies have the potential to enhance access to diagnosis and treatment in both urban and rural areas [6][7]. However, in the midst of modernization, this raises the challenge of ensuring that patients remain well-informed, respected, and connected to the doctors looking after them.

The current study investigates the understanding and experience of both doctors and patients in India concerning AI-assisted diagnostic tools. This comparative mixed-methods investigation has been utilized to understand how AI influences patient trust, communication between physicians and their patients, and clinical practice. The research combines quantitative data with qualitative insights to identify points of convergence and divergence between the two groups.

Ultimately, the goal of this study is to answer one main question:

Can AI really fortify healthcare without weakening the human connection that defines it?

By mapping the gaps in communication and trust, this research embarks on developing guidelines to ensure the ethical use of AI in medicine, where technology supports doctors, empowers patients, and strengthens, rather than substitutes, the human touch at the heart of care.

This study is guided by the following objectives:

- To delineate the perceived functional role of AI in diagnosis from the perspective of practicing medical professionals.
- To quantify and qualify the impact of AI-assisted diagnosis on patient trust, confidence, and expectations of accountability.
- To identify specific areas of alignment and critical gaps between patient and doctor perceptions regarding transparency and final decision-making authority.

2.Theoretical Framework and Conceptual Definitions

2.1 Clinical Decision Support Systems (CDSS)

The physicians in this study consistently describe AI as a Decision-Support System (DSS). A CDSS is designed to provide clinicians with pertinent information and recommendations at the point of care to enhance decision-making. The core principle of a DSS is that the practitioner should be the final decision-maker, using their own thinking, reasoning, and judgment. The doctor's use of AI as a Suggestion (as interpreted by over 85% of patients) aligns perfectly with the DSS model, where the AI's role is strictly consultative.

2.2 Automation Bias and Skill Erosion

A key concern raised by the interviewees relates to automation bias, a phenomenon where the practitioners may over-rely on the AI suggested diagnosis potentially leading to failing to detect problems or falsely accepting the AI's recommendation. This concern directly validates that blind dependence on AI can erode clinical skills over time. Because of the risk of overly-relying on AI, doctors have a moral duty to always double-check the AI's suggestions using their own medical knowledge and experience.

2.3 The Transparency Imperative

The Transparency Imperative means that for AI systems to be trusted and accepted in healthcare, both doctors and patients need to clearly understand how the technology contributes to diagnosis and treatment. If patients know what the AI does — for example, that it helps the doctor analyze scans or detect patterns — they are more likely to feel confident in the process. However, when doctors fail to explain AI's role, as shown by the 74.4% of patients who said they were not informed, it creates a serious trust gap. Patients are left wondering how much of their diagnosis came from the doctor and how much came from the AI, which may create uncertainty or raise questions regarding the practitioner and the diagnosis.

3. Methodology

3.1 Research Design

The study was done in two stages using both numerical data and descriptive insights to fully understand how people experience and interact with AI during diagnosis. Phase 1 involved quantitative open ended and quantitative data collection from patients, while Phase 2 involved qualitative data collection from doctors. The analysis focuses on synthesizing findings from both sources to identify areas of agreement and divergence between the two groups.

3.2 Phase 1: Patient Survey (N=20)

3.2.1 Participant Characteristics

The study included N=20 patients who had received a diagnosis where AI tools were known to have contributed to the diagnostic process.

3.2.2 Instrument and Data Collection

The survey utilized a combination of structured and open-ended questions:

- Trust and Confidence (Likert Scale): Questions (Q2-Q4) assessed changes in trust toward the doctor, confidence in the doctor's skill, and confidence in the final diagnosis (rated on a 5-point scale).
- Transparency and Authority (Forced-Choice): Questions (Q5, Q8, Q12) captured patient perception of information disclosure and the perceived role of AI (Suggestion vs. Order).
- Accountability and Future Intent (Forced-Choice & Qualitative): Questions (Q13, Q14) explored who should be blamed for an error and the likelihood of seeking an AI-using doctor again.

3.3 Phase 2: Doctor Interviews (N=3)

3.3.1 Participant Selection and Characteristics
Three medical practitioners were. These individuals were selected based on their experience or strong strategic involvement with AI in medicine, providing insights from gastroenterology, general clinical practice, and health technology strategy.

3.3.2 Instrument and Data Collection
Interviews were semi-structured and focused on five thematic areas:
(1) AI's functional role
(2) the necessity of human judgment
(3) ethical considerations (trust, transparency)
(4) workflow integration
(5) future impact on clinical skills.

3.4 Data Analysis Strategy

- Quantitative Analysis: Descriptive statistics (frequencies and percentages) were used to characterize patient responses to the Likert scale and forced-choice questions.
- Qualitative Analysis (Doctor Data and Patient Open-Ends): We analyzed the interview transcripts and patients open-ended responses to find common themes and ideas. Key themes—such as 'AI as DSS,' 'Skill Erosion,' and 'Human Empathy'—were systematically coded and categorized to build the doctors' unified perspective.

4.RESULTS

4.1 Findings from Doctor Interviews (N=3): The Augmentation Imperative

The qualitative analysis of the doctors' responses yielded three core findings that define the professional consensus on AI assisted diagnosis.

4.1.1 The Functional Role: AI as a Decision-Support System (DSS)
This functional definition highlights that AI is primarily valued for its capability to perform tasks that exceed human capacity in speed and volume, such as identifying subtle abnormalities in image-based diagnostics like endoscopy and colonoscopy. Its purpose is to guide, not decide.

4.1.2 The Ethical and Professional Obligation to Adopt
This view suggests that using AI is not just about having better technology — it's becoming an essential skill for future healthcare professionals. Subsequently there is a risk that relying too much on AI could lead to automation bias, meaning that clinicians might start trusting the AI's suggestions without questioning them. Over time, this could weaken their own clinical judgment and skills. Because of this, there is an ethical responsibility for clinicians to stay actively involved, think critically about the AI's recommendations, and not accept them blindly.

4.1.3 The Emphasis on Human Empathy and Societal Impact
This human role also includes a social responsibility — using AI in ways that serve the broader public good. Several participants emphasized the importance of developing low cost algorithms in India to make healthcare more affordable and accessible, especially in underserved areas. By doing this, AI can help improve diagnosis, expand healthcare reach, and ultimately strengthen the country's overall health outcomes. Even with these technological advances, the doctor remains the vital link between technology and humanity. The physicians stressed that this human connection is essential for maintaining patient trust and ensuring that AI supports, rather than replaces, compassionate care.

4.2 Findings from Patient Survey (N=20): The Trust-Transparency Gap

Patient responses showed strong agreement on the importance of maintaining human authority, but also revealed considerable disappointment with how the AI process were communicated to them.

4.2.1 The Critical Transparency Deficit
The most compelling quantitative finding of the study confirms a fundamental failure in the patient-facing implementation of AI-assisted diagnosis:
74.4% of patient respondents felt their doctor did NOT provide enough information about how the AI contributed to the diagnosis or treatment.
This deficit demonstrates a sharp disconnect between the doctors stated commitment to transparency (Section 4.1.3) and the actual patient experience at the point of care.

4.2.2 Polarized Trust and Neutral Confidence
The introduction of AI did not uniformly boost patient trust.

- Trust in Doctor (Q2): While 40.9% reported trusting the doctor More or Somewhat More (citing reasons like "more accurate and dependable" results), a significant 29.5% reported trusting the doctor Less or Much Less (citing reasons like "I trust the human more than AI"). These findings indicate that AI's adoption generates differing opinions rather than uniting perspectives.
- Confidence in Diagnosis (Q4): Confidence in the final diagnosis was typically neutral to slightly positive, with the majority of ratings clustering around 3 and 4 (out of 5).

4.2.3 Unwavering Demand for Human Authority and Accountability

- Even with the use of advanced technology, patients consistently emphasized the need for human supervision.
- Final Decision-Maker (Q12): 75% of respondents unequivocally stated that The Doctor should be the final decision-maker.
- AI's Role Perception (Q8): Patients correctly interpreted the AI's role, with over 85% viewing the AI's input as a Suggestion—a perception that aligns perfectly with the doctors' DSS framework.

- Accountability for Error (Q13): In the event of a wrong diagnosis, 54.8% of patients would blame The Doctor more, and 35.7% would blame both equally, confirming the doctor's irreplaceable responsibility for the outcome.

Q ID	Type	Content
Q2	Likert	After hearing about the AI, my trust in the doctor changed to... (Much Less to Much More)
Q3	Likert	My confidence in the doctor's skill level was... (1 to 5)
Q4	Likert	My confidence in the final diagnosis was... (1 to 5)
Q5	Choice	Did your doctor provide enough information about how the AI contributed to the diagnosis? (Yes / No)
Q8	Choice	I perceived the AI's input to be a... (Suggestion / Order)
Q12	Choice	Who should be the final decision-maker? (Doctor / Patient / Both Equally)
Q13	Choice	If the diagnosis was wrong, who would you blame more? (AI / Doctor / Both Equally)
Q14	Open	What is your biggest hope/worry about AI in future diagnoses?

Fig 1:- Patient Survey Questions

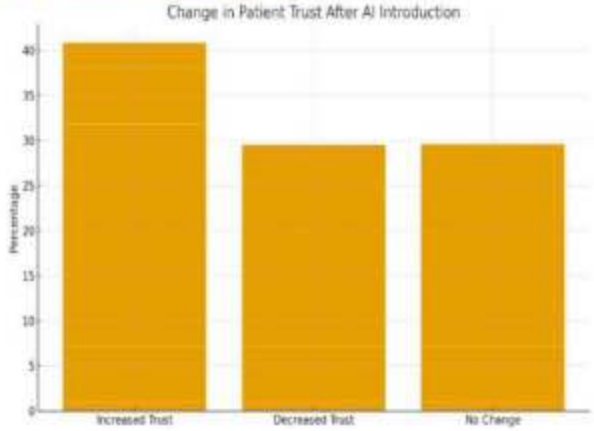
Table 1: Patient Demographics

Category	Count	Percentage
Male	13	65%
Female	7	35%
Age 31-50	5	25%
Age 51+	15	75%

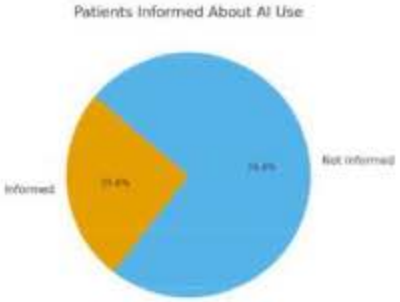
Table 2: Summary of Patient Responses (Likert Scale)

Statement	Most Common Response	Agreement (%)
I trust AI-assisted diagnosis	Agree	58%
Doctor should make final decision	Strongly Agree	75%
AI improves accuracy of diagnosis	Agree	69%
I was informed about AI use	Disagree	74%

Graph 1: Change in Patient Trust



Graph 2: Patient Awareness of AI Use



5.Discussion: Bridging the Doctor-Patient Disconnect

5.1 The Strategic Alignment of DSS and Patient Expectation

The strongest point of agreement between doctors and patients lies in how they define AI's role in healthcare. Both groups believe that AI should act as a supportive tool — something that can suggest or guide decisions, but should never give orders or make the final call. Doctors emphasize the importance of their own clinical judgment, intuition, and understanding of each patient's unique situation, and this directly aligns with what patients want: a clear sense that a human remains in charge.

This shared perspective forms a solid foundation for how AI can be responsibly integrated into future medical practice. It also shows that people are not necessarily anxious about AI itself, but rather about how much control it might take away from humans. In other words, the concern isn't the use of AI, but overreliance on it, which supports doctors' worries that depending too much on AI could eventually weaken their own skills and judgment.

5.2 The Crisis of Disclosure: The Trust-Transparency Gap

The data shows a major problem with how transparency is being put into practice. A 74.4% transparency gap was found — the most important finding of this study. This is worrying because the doctors interviewed clearly understood that being open with patients is essential. Transparency and explanation are vital to build trust.

However, this understanding is not being applied in real patient interactions. The lack of clear communication about AI use may be due to limited time, inadequate training, or the absence of standard guidelines for explaining AI involvement during busy clinical visits.

Because of this, patients tend to fall into two extreme groups: some trust the technology completely (a 40.9% rise in trust), while others lose trust in the doctor for giving too much control to AI (a 29.5% drop in trust).

5.3 Ethical Implications and Clinical Accountability

Patient responses about accountability (Q13) highlight a key ethical and legal principle — that responsibility for medical decisions cannot be handed over to AI. Since patients hold the doctor, or in some cases both the doctor and the AI, accountable for outcomes, the physician cannot rely on an AI recommendation as a defense if an error occurs. This legal reality reinforces the importance of the Decision Support System (DSS) model, in which the doctor must always perform the final review, verifying the AI's suggestions with their own clinical judgment and expertise.

The qualitative data from patients also showed that their main concern is not the use of AI itself, but rather the risk of overreliance on it, which could result in less human input during diagnosis or treatment. This perspective mirrors the doctors' own worries about the gradual erosion of clinical skills if AI becomes too dominant in decision-making.

5.4 Recommendations for Clinical Intervention

The transparency gap revealed in this study highlights the need to bring human connection and clarity to the use of AI in medicine. Technology can enhance care, but it cannot replace the understanding, empathy, and ethical judgment that define good clinical practice. The following steps can help ensure that AI strengthens, rather than weakens, the human side of healthcare:

- Honest Conversations About AI :- Clinicians should be supported with a simple and consistent way to talk to patients about when and how AI is used. A brief disclosure guide can help them explain what the AI analyzed, what it recommended, and how the doctor's final decision aligned or differed from that advice. Such openness fosters trust, showing patients that the doctor remains accountable and that technology serves as an aid, not a replacement.
- Training for Thoughtful AI Use :- Doctors need more than technical skills — they need confidence in interpreting, questioning, and communicating AI results. Dedicated AI literacy training can help clinicians use these tools responsibly, guard against automation bias, and stay grounded in their essential role as the patient's trusted decision-maker. This kind of training ensures that AI supports clinical reasoning instead of replacing it.
- Protecting Time for Human Connection :- AI can save time in administrative or diagnostic tasks, such as data review or report generation. The time gained should be reinvested in conversation and care — allowing doctors to explain complex diagnoses, answer patient questions, and show empathy. This deliberate effort to preserve communication ensures that technology enhances, rather than erodes, the patient–doctor relationship.

6.Conclusion and Future Research

6.1 Summary of Key Findings

This study shows that doctors and patients largely agree on how AI should be used in diagnosis. Both see it as a tool that provides accurate and fast support, while the final judgment and decisions must always remain with the doctor.

However, the findings also reveal a major issue — a 74.4% lack of transparency about how AI is used. This lack of clear communication can damage patient trust and create division instead of shared understanding. The results make it clear that doctors carry the main responsibility for making the final decisions, being accountable for outcomes, and explaining the AI's role to patients in an open and understandable way.

6.2 Limitations of the Study

The main limitation of this study is its small sample size and narrow scope. The patient data (N=20) offers useful insights but is too limited to draw broad conclusions about the larger population. Similarly, the doctor interviews (N=3) provide valuable perspectives but do not represent the full range of medical specialties, locations, or hospital practices related to AI. Overall, the study gives a focused snapshot of how AI is

viewed in healthcare, rather than a picture that applies to all settings or professionals.

6.3 Directions for Future Research

- Larger and Broader Studies:- The survey should be repeated with a larger and more diverse group of participants (over 500 people) from different regions. This would help confirm how widespread the transparency gap is and provide a stronger evidence base for action.
- Testing Real-World Solutions:- Researchers should develop and test clear AI disclosure scripts in real clinical settings. These scripts would guide doctors in explaining how AI supports their decisions. The goal is to see whether this open communication improves patient trust (Q2) and their understanding of the process (Q5).
- Exploring Different Specialties:- Future interviews should include a wider range of doctors, such as radiologists, pathologists, and general practitioners, to explore how AI's role and impact differ across medical fields.

REFERENCES

[1] T. P. Quinn, M. Senadeera, S. Jacobs, S. Coghlan, and V. Le, "Trust and medical AI: the challenges we face and the expertise needed to overcome them," Journal of the American Medical Informatics Association, vol. 28, no. 4, pp. 890–894, Mar. 2021, doi: 10.1093/jamia/ocaa268.Link: https://pmc.ncbi.nlm.nih.gov/articles/PMC7973477/

[2] Y. Tang and J. Cai, "Impact and prediction of AI diagnostic report interpretation type on patient trust," Frontiers in Computing and IntelligentSystems,2024.Link: https://drpress.org/ojs/index.php/fcis/article/view/8567

[3] B. N. Hryciw, Z. Fortin, J. Ghossein, and K. Kyeremanteng, "Doctor-patient interactions in the age of AI: navigating innovation and expertise," Frontiers in Medicine, vol. 10, p. 1241508, 2023, doi: 10.3389/fmed.2023.1241508.Link: https://pmc.ncbi.nlm.nih.gov/articles/PMC10498385/

[4] S. Mishra, R. R. Nair, and S. Krishnan, "Clinician–patient alignment in AI-enabled healthcare: reconciling expectations through transparency," arXiv preprint arXiv:2312.10072, Dec. 2023. Link: https://arxiv.org/abs/2312.10072

[5] M. N. I. Chowdhury and R. Rahman, "Building patient trust in AI-driven healthcare systems through ethical design and explainability," arXiv preprint arXiv:2501.16693, Jan. 2025. Link: https://arxiv.org/abs/2501.16693

[6] R. K. Dasgupta and S. Bhargav, "AI in Indian healthcare: From roadmap to reality," Journal of Artificial Intelligence in Medicine, vol. 12, no. 3, pp. 45–58, 2024, doi: 10.1016/j.aiim.2024.03.005. Link: https://www.sciencedirect.com/science/article/pii/S2949866X24000285

[7] A. Chauhan, D. Sarkar, G. S. Verma et al., "Evaluating trustworthiness in AI-based diabetic retinopathy screening: addressing transparency, consent, and privacy challenges," BMC Medical Ethics, vol. 26, p. 140, 2025,doi:10.1186/s12910-025-01265-7.Link: https://bmcmedethics.biomedcentral.com/articles/10.1186/s12910-025-01265-7

[8] M. Goisauf, "Trust, trustworthiness, and the future of medical AI," Journal of Medical Internet Research, vol. 27, no. 1, p. e69678, 2025, doi: 10.2196/69678. Link: https://www.jmir.org/2025/1/e69678

[9] N. Newton, A. Bamgboje-Ayodele, R. Forsyth, A. Tariq, and M. T. Baysari, "A systematic review of clinicians' acceptance and use of clinical decision support systems over time," npj Digital Medicine, vol. 8, p. 309, 2025, doi: 10.1038/s41746-025-01662-7. Link: https://www.nature.com/articles/s41746-025-01662-7

[10] A. Sauerbrei, A. Kerasidou, F. Lucivero, and N. Hallowell, "The impact of artificial intelligence on the person-centred, doctor-patient relationship: some problems and solutions," BMC Medical Informatics and Decision Making, vol. 23, p. 73, 2023, doi: 10.1186/s12911-023-02162-y.Link: https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-023-02162-y

[11] O. Asan, A. E. Bayrak, and A. Choudhury, "Artificial intelligence and human trust in healthcare: focus on clinicians," Journal of Medical Internet Research, vol. 22, no. 6, p. e15154, 2020, doi: 10.2196/15154. Link: https://www.jmir.org/2020/6/e15154/

[12] Y. Zhang, "Forging trust in AI-assisted disease diagnosis," Precision Clinical Medicine, 2025, doi: 10.1016/j.pcm.2025.00056. Link: https://www.sciencedirect.com/science/article/pii/S2949928325000562

[13] J. Fehr, M. J. van der Meer, and M. Z. Hussain, "A trustworthy AI reality-check: the lack of transparency and the need for governance in medical AI," Frontiers in Artificial Intelligence, vol. 7, p. 135, 2024. Link: https://pmc.ncbi.nlm.nih.gov/articles/PMC9472134/

[14] E. Nasarian, R. Alizadehsani, U. R. Acharya, and K.-L. Tsui, "Designing interpretable ML system to enhance trust in healthcare: a systematic review to proposed responsible clinician-AI-collaboration framework," arXiv preprint arXiv:2311.11055, Nov. 2023, doi: 10.48550/arXiv.2311.11055. Link: https://arxiv.org/abs/2311.11055

[15] G. B. Mensah, "Trust and transparency in AI health tools," ResearchGatePreprint,Mar.2023.Link: https://www.researchgate.net/publication/381609214_Trust_and_Transparency_in_AI_Health_Tools

Your Browser Is Snitching: The Invisible Army of Fingerprints, Cookies

Navya Bora
Electronics and Engineering
Maharaja Surajmal Institute of Technology
Janakpuri, New Delhi -110058
nbora1102@gmail.com

Tanishka Bhardwaj
Electronics and Engineering
Maharaja Surajmal Institute of Technology
Janakpuri, New Delhi -110058
tanishkabh06@gmail.com

Sarvam Taneja
Electronics and Engineering
Maharaja Surajmal Institute of Technology
Janakpuri, New Delhi -110058
sarvamtaneja@gmail.com

Sarthak Vijay
Electronics and Engineering
Maharaja Surajmal Institute of Technology
Janakpuri, New Delhi -110058
vjsarthak21@gmail.com%20

Shankul Sahni
Electronics and Engineering
Maharaja Surajmal Institute of Technology
Janakpuri, New Delhi -110058
shankul.sahni@gmail.com

Abstract— **Within the expansive digital landscape of the internet, each user interaction leaves a trace, a "crumb" that websites eagerly gather. While cookies have traditionally been used as means of storing users' information, emerging tracking techniques like fingerprinting have raised new privacy concerns.**

Keywords— Cookies; Browser fingerprinting; Third-party tracking; Web tracking; Data Provenance.

I. INTRODUCTION

In the digital ecosystem, an internet cookie, which is also commonly abbreviated as a cookie, is a small text file that a website stores on the user's device (like a computer or smartphone) when the user visits that site[18]. The term reminds one of the familiar baked treat, yet its role is far more practical. Similar to a crumb of cookie dropped at a party, it allows sites to remember user preferences and actions for personalized enhancements [3]. Generated by the site's server and stored via the browser at a local folder, cookies not only facilitate seamless web experiences but also spark debates on privacy, much like how enjoying a shared cookie dip can unknowingly spread unwanted germs in the physical world, data in the cyber world. Although Browser fingerprinting is not a new technique, it has gained increased attention as major web browsers like Safari, Firefox, and Chrome have begun enforcing stricter privacy measures. It collects device attributes and creates a hashed ID, typically a short numerical string. The key privacy concern is that users are often unaware of this process, and even if they are, they cannot opt out [2]. Tanwir Danish, managing partner and head of data and analytics platforms at GroupM North America, explains, "The much bigger issue for us is what happens to cookies over time. The reality is that the same privacy concerns that apply to cookies apply to fingerprinting, especially when it comes to targeting consumers based on their own information."

Fingerprinting and other advanced tracking techniques make the user's privacy more challenging. It collects information about a user's device, browser, and online activities to create a unique identity that might last even after cookies are deleted [32].

Data provenance, the ability to trace the origin and path of information gathered, disappears in such complex tracking, and this invisibility is what makes it more difficult to examine data flows, spot unlawful leaks, and implement safeguards for data [7].

As the web moves toward a post-cookie era, these advanced "double-dipping" techniques, where multiple tracking layers secretly interact, pose a serious threat to the secrecy of an individual and it becomes utmost important to understand and address the implications of these hidden data exchanges for creating transparent and privacy-respecting web ecosystems [5].

Modern Privacy Tools aren't that effective despite their noble objectives, as the new tools have a complex pathway[20]. They have APIs, microservices, Distributed Architecture, due to which tracking the data flow is an enormous challenge. The incorporation of these tools lacks automation and has to be done manually, which is more prone to errors, time-consuming, and unsustainable in today's fast-paced world.

Web tracking, these days, is like a bowl of cookies and cream, familiar but still deceptive. The cookie chunks are those cookies that a user can view and delete, but the smooth cream underneath is the invisible fingerprinting that silently rebuilds user identity. And the sweet pieces of cache data and LocalStorage recover the erased data. All of these components come together to make a smooth, creamy system that keeps pulling in the user's data in unsuspecting ways. In this metaphor, what appears as straightforward or user-friendly technology masks complex and continuous surveillance [2].

This paper investigates the multi-layered structure of these tracking methods and the problems they present for legislation, privacy audits, and transparency. Additionally, it analyses the inefficiencies of the current-day privacy tools, whose manual integrations and distributed architectures often fail to detect data flows, and offers a conceptual framework that shows how seemingly innocuous technologies can mask complex monitoring like a deceptive bowl of cookies and cream.

Unlike earlier studies that examined cookies or fingerprinting separately, our research presents the idea of "deceptive double-dipping" to explain how multilayer tracking techniques collectively hide data provenance. It offers a fresh viewpoint on the opacity of modern web monitoring by presenting a conceptual model that explains why current privacy technologies are unable to identify such compound tracking.

II. TECHNICAL FLOW

A. Introduction to Layered Tracking

Layered tracking is a multilayered architectural framework that integrates diverse tracking mechanisms into a unified and resilient system designed for persistent user identification.

Each layer operates independently yet cooperatively towards the same goal: Persistent User Tracking. When one layer fails due to technical limitations, browser restrictions, or countermeasures taken by user, the remaining layers ensure that the continuance of data tracking stays undisturbed. This architectural redundancy, borrowed from fault-tolerant disturbed systems, allows the system to deteriorate gracefully instead of failing altogether at once, effectively turning a reliability principle into a mechanism of surveillance.

Empirical studies reveal that layered tracking is deeply embedded in major web ecosystems. Facebook's Pixel and Conversion API retain user identification through fingerprinting and server-side synchronization even after cookie deletion [1]. Adobe and Pardot employ CNAME cloaking to disguise third-party trackers as first-party subdomains, thus bypassing Firefox's Enhanced Tracking Protection

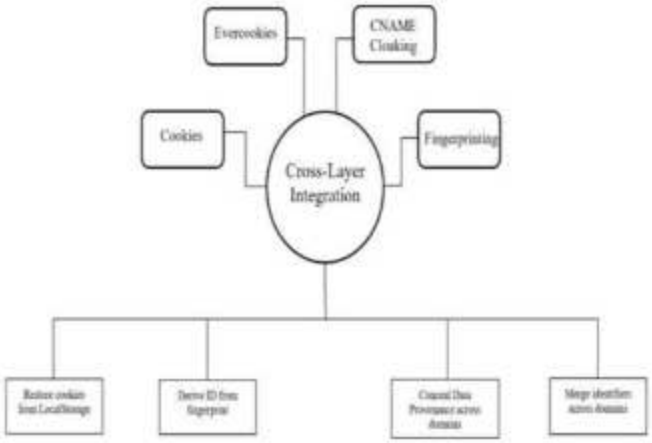


FIGURE 1. HOW LAYERING WORKS

[27]. TikTok's analytics framework associates web and mobile identifiers to allow cross-session re-identification[2]. Similarly, Google's DoubleClick network integrates cookie and fingerprint-derived identifiers to preserve personalized advertising [17].

Collectively, these implementations demonstrate that layered tracking is a self-reinforcing mechanism that sustains user profiling despite privacy defences.

B. Layer 1: Third-Party Cookies and Cookie Syncing

Third-party cookies form the foundational layer of user tracking [9]. As one of the earliest and most enduring identification mechanisms [13][14], these cookies are set up by external domains that are embedded in web pages, such as advertising networks or analytics providers. Each cookie stores a unique identifier through which the common third-party domain that appears on several pages can be cross-site correlated to facilitate the tracker networks to rebuild big parts of users' browsing histories and behavioural profiles [25][26].

The adoption of browser countermeasures such as Safari's Intelligent Tracking Prevention (ITP) and Firefox's Enhanced Tracking Protection (ETP) has considerably diminished the use of third-party cookies [22][23][21]. However, to bypass these restrictions, tracking entities are increasingly turning towards cookie syncing, a process where advertising partners exchange or match user identifiers through redirect chains or pixel requests [25][26].

By synchronizing identifiers across different domains, trackers can continue to recognise users even after cookies are deleted, resulting in the creation of a scattered and long-lasting monitoring network that keeps the identification of users going on different sites in spite of latest privacy protection measures [25][26].

C. Layer 2: Evercookies

Evercookies, also known as super cookies or zombie cookies, represent a complex form of user tracking that enables long-term identification by duplicating unique identifiers in various browser storage systems [8][28][29]. Unlike standard cookies, which can be deleted by the user, evercookies are designed to regenerate themselves by restoring data from alternative storage locations [8].

These mechanisms exploit a variety of storage vectors, generally between 10 and 17, among which HTTP cookies, Flash LSOs, HTML5 LocalStorage, SessionStorage, IndexedDB, Web SQL, HTTP ETags, browser cache, and Canvas-encoded PNGs are included [28][29].Upon a user's return to a website, the server looks through these areas and if it finds any pieces, it rebuilds the identifier [8].A 2014 Princeton study found such mechanisms on major websites, including the first commercial use of IndexedDB for cookie resurrection [8].

Research indicates that Web Storage APIs are used by trackers on approximately 70 percent of websites, with IndexedDB employed by more than 30 percent [19]. Because many browsers fail to completely clear these storage areas even in private browsing mode, evercookies remain highly persistent and resistant to deletion [23][20].Such persistence makes it possible for identifiers to be stored under the main domain thus achieving the goal of circumventing third-party tracking restrictions.

D. Layer 3: Browser Fingerprinting

Browser fingerprinting is another tracking technique that identifies users without depending on cookies or storage simply by gathering browser and system attributes through client-side APIs and HTTP headers [23][24][15]. This is made possible by combining data like screen resolution, fonts, and plugins to create unique identifiers [12][31]. Because fingerprints are invisible and cannot be deleted by the user, they are proven to be highly intrusive [10].

At their core, these methods exploit the differences that pre-exist between hardware and software rendering. Canvas fingerprinting is a technique that uses HTML5 Canvas to produce hashes from font rendering [25][26][33], appearing on about 5% of major sites [34]. Similarly, WebGL and audio fingerprinting technologies extract the distinctive GPU and sound-processing unit signatures [34][10]. Besides these, few methods involve using unique technologies like font enumeration, plugin detection, and APIs like Battery Status [32].

Fingerprints not only provide identification that is almost unique but can also recreate deleted cookies. So, with browsers gradually putting a stop to third-party cookies, advertisers are increasingly moving towards fingerprinting due to their user consent-free and stable way of tracking [32][11].

E. Layer 4: CNAME Cloaking

As a result of increased browser restrictions towards third-party cookies, advertisers are now turning to CNAME cloaking, an evasion technique that hides third-party trackers by showcasing them as first-party subdomains [27].

Whenever a user accesses the website, the DNS alias goes back to tracker-controlled infrastructure, allowing JavaScript execution and identifier collection under a trusted domain [27]. A large-scale study of the Alexa Top 300,000 sites disclosed that out of the total sites, 1,739 (0.58%) sites were involved in CNAME cloaking. Adobe (52.5%), Pardot (25.7%), and Oracle (3.0%) were found to be the top users of this technique [27]. Since most privacy tools use domain matching as the primary metric, only 77.9% of cloaked trackers got recognized, exposing major vulnerabilities in current defences [27].

F. Combined Operation – The Double-Dip Flow

Online tracking persists primarily due to cross-layer integration of several different, individual techniques. Modern trackers merge identifiers across cookies, LocalStorage, IndexedDB,and fingerprinting hashes to ensure continuity even if user intervention is detected [8][24][11][10][29]. For instance, a deleted cookie can be easily restored from LocalStorage or a fingerprint derived ID, thus resulting in re-affiliation of the session without the user giving their consent [10].

This identifier respawning and cookie syncing process spans domains via modern techniques like CNAME cloaking and redirect chains which allows

TABLE I
COMPARISION of VARIOUS TRACKING LAYERS

Tracking Layer	Mechanism	Storage	Persistence	Resistance to Countermeasures
Third-Party Cookies	Cross-site identifier stored by external domains	Browser cookie store	Moderate	Weak
Evercookies	Identifier replication across multiple storage vectors	LocalStorage, IndexedDB	Very High	Strong
Browser Fingerprinting	Device and software attribute profiling	APIs (Canvas, WebGL, Audio, Fonts)	High	Strong
CNAME Cloaking	Disguising third-party requests as first-party	DNS aliasing (CNAME records)	High	Very Strong
Cross-Layer Integration	Synchronization across multiple tracking layers	Combined identifiers (cookies, fingerprints, storage)	Extremely High	Extremely Strong

trackers to merge the identifiers to form a single unified cross-site profile [8][11]. Such kind of synchronization conceals data provenance, making it almost impossible to identify which mechanism created or maintained the identifier.

When independent tracking layers are integrated together, a single cohesive framework is formed. It is the formation of such a framework that results in reduced effectiveness of traditional privacy protections like clearing cookies and anti-tracking filters.

G. Why Standard Defences Fail

Traditional methods of privacy protection like cookie blocking, cache clearing, and ad-filtering, have been proven to be ineffective against modern, and cross-layered tracking systems. For example, CNAME cloaking can change the way third-party trackers are classified by turning them into first-party subdomains; thus, they can go beyond filter lists and same-origin restrictions while staying undetected [11], [28]. Meanwhile, cross-layer identifier coordination facilitates the restoration of deleted cookies stored in different storage vectors or fingerprint derived identifiers [16][24]. These self-reinforcing architectures take full advantage of browser trust models and DNS opacity, further complicating provenance analysis, resulting in failure of advanced auditing frameworks [10].

III. PROOF OF CONCEPT

A. Methods and Methodology

This research followed an observational and experimental approach, observing how cookies track data even after third-party cookies were turned off in privacy settings.

TABLE II
DEPICTING COMPONENTS and DESCRIPTION of the TOOLS and BROWSER USED for OBSERVATIONS

Components	Description
Operating system	Windows 11 (64-bit)
Browser Used	Google Chrome (version 128.0)
Mode of operation	Normal browsing (non-incognito, non-guest)
Browser configuration	Third-party cookies are manually disabled
Observation tools	Chrome developer tools

TABLE III
DESCRIPTION OF THE SITES VISITED FOR OBSERVATION

Website	Category	Rationale
CNN	Media	Known for heavy third-party ad integration
AON	Corporate/B2B	Representative of enterprise-level websites with embedded analytics
Reddit	Social media	High use of engagement and behavioral tracking
Amazon	E-commerce	Acts as a control with an internally managed ecosystem tracking

The browser's developer tools were used for each of the websites mentioned in *(Table III)* to inspect the cookies. Parameters like Domain, cookie name, partial value, expiry date, and security flags were recorded. The appearance of third-party domains within the sites that were not related to those domains was used as an indicator of tracking cookies. This whole process was again repeated after changing the cookie settings to verify the results and see if the results still persist. All data collection was performed without any interference with the networks; all the data was available publicly online to ensure that the results were deduced ethically.

B. Observations

The cookies were observed from the websites mentioned in *(Table III)*, and the data was collected using developer tools and noted down in *(Table IV)*.

During the analysis of edition.cnn.com, multiple third-party cookies were observed, despite turning off

TABLE IV
OBSERVATION OF COOKIES ON DIFFERENT SITES

Site on which the data was observed	Domain	Name	First 10 characters of value	Size(bytes)	Expires on	HTTP only	Security
cnn.com	.temu.com	api_uid	CmyAlGkF7R	31	2026-11-01	no	yes
amazon.com	.amazon-adsystem.com	ad-id	A3kPMXPQ9k	28	2026-07-01	yes	yes
reddit.com	.rlcdn.com	pxrc	CLvcyrgGEg	720	2025-12-31	no	yes
aon.com	.rlcdn	pxrc	CLvcyrgGEg	720	2025-12-31	no	yes
aon.com	.reddit.com	csrf_token	3dd6edb193	42	session	no	yes
aon.com	linkedin.com	AnalyticsSyncHistory	AQKKUFUUB_	106	2025-11-28	no	yes

the third-party cookie setting, including a cookie named 'api_uid' that was associated with the domain '.temu.com' even though the Temu website was never directly visited. This shows the cross-site data sharing, allowing data to be exchanged between unrelated sites (usually via advertising and analytical iframes).

In contrast to Amazon.com, it showed a different way of collecting cookies. Almost all cookies were linked to .amazon.com or .com, both of which belong to Amazon's ecosystem. This explains that Amazon restricts cookie sharing to its own domains rather than any third-party domains. The cookies like 'ad-id' were mentioned as both HttpOnly and Secure.

Along with the results mentioned above, another key finding that could clearly be observed was between reddit.com and aon.com . Despite disabling the third-party cookies, both sites showed several cookies linked with third-party domains like '.rlcdn.com'. There was an unexpected finding while doing close cookie tracking of aon.com that showed reddit.com and linkedin.com directly linked to several cookies (mentioned in Table IV) without direct interaction with those platforms. This finding is significant to understand that even if third-party tracking is disabled the mechanism still works.

Another finding was observed while visiting reddit.com , a cookie named 'pxrc' from the domain '.rlcdn.com' was created, and the same cookie named 'pxrc' was later detected on Aon's website with exact same cookie value (Table IV), even though the sites are completely unrelated. This shows that embedded elements like ads and widgets can effectively perform 'layered tracking' across sessions.

C. *Extended Experimental Observations: Incognito Mode and Cross-Browser Analysis*

When browser was on incognito mode and third party cookies were kept on , websites like "reddit.com" and "walmart.com" were using identical cookies named "rlas3" and "pxrc" that had the same value, suggesting that data was still being recorded even in incognito mode.

But, when the incognito tab was closed and reopened both the sites still contained the cookies named "rlas3" but the value this time was different, suggesting that the cookie tracking in incognito mode is session-scoped.

TABLE V
OBSERVATION of COOKIES on INCOGNITO MODE (CHROME

Cookies name	S
rlas3	r
pxrc	r
rlas3	v
pxrc	v
IDE	v
IDE	r
IDI	

TABLE VII
NUMBER of COOKIES on CHROME BEFORE and AFTER ENABLING THIRD PARTY PROTECTION

Sites visited	No. of cookies	third -party blocking (on/off)	Observation
reddit.com	4	off	Cookies from third-party domains exist
reddit.com	4	on	No change, all the cookies remain active
aon.com	19	off	Very high amount of cookies present (including RLCDN, linkedin, reddit etc)
aon.com	6	on	Huge difference , but third-party cookies still exist
			Small amount of third-party cookies present
			All third-party cookies blocked

TABLE VI
NUMBER of COOKIES on CHROME BEFORE and AFTER ENABLING THIRD PARTY PROTECTION

Site visited	Number of cookies	Third-party blocking
reddit.com	6	off
reddit.com	5	on
walmart.com	3	off
Walmart.com	1	on
aon.com	24	off
aon.com	21	on

This also suggests that cookie tracking in incognito mode does not exist across sessions but is fully functional during active session.

When third-party cookies were blocked from the browser settings various sites exhibited different response, This led to the removal of almost all third-party cookies from 'walmart.com' except a cookie named "IDE" from the domain '.doubleclick.net' . This suggests that even though you disable the third-party cookies from the chrome browser settings , it still allows some exceptions.

The case with 'reddit.com' and 'Aon.com' was different . Even after blocking these third-party cookies both the sites showed various third-party cookies, and Aon.com even showed many cookies linked to the domains like 'linkedin.com' and 'reddit.com' even in incognito mode.

TABLE VIII
COMPARISION of NUMBER of COOKIES BEFORE and AFTER BLOCKING THIRD PARTY COOKIES in OPERA and CHROME BROWSER

Website	Browser	No. of cookies (third-party blocking off)	No. of cookies (third-party blocking on)
reddit.com	Chrome	6	5
reddit.com	Opera	4	4
walmart.com	Chrome	3	1
walmart.com	Opera	3	0
aon.com	Chrome	24	21
aon.com	Opera	19	6

This observation (*TableVIII*) indicates that Opera provides more or less the same protection when the third-party cookies are not blocked , however when third-party protection is enabled, Opera clearly performs better than Chrome browser.

On the other hand , Chrome's incognito mode does not provide complete protection from cross-site tracking , but it clears all the cookies when the session is terminated.

IV. DATA PROVENANCE

Data provenance refers to the methods that involve recording, monitoring, and storing information regarding the origin of data and every subsequent change made to it throughout its lifecycle. It comprises details such as where the data was taken from, how it was created, who accessed or modified it, and the exact alterations performed at every step. Simply put, provenance is the mechanism that enables the entire data history from its initial state to its current version to remain transparent and verifiable. Such openness allows the evaluation of crucial data quality attributes like accuracy, timeliness, reliability, and trustworthiness.

With the increasing publication of RDF (Resource Description Framework) data on the Web, where datasets are continuously linked and interrelated, provenance acts as the tool that enables both machines and humans to trace data across multiple sources and verify its authenticity by ensuring it can be traced back to its primary source [30].

To achieve this, layered provenance tracking collects provenance data across multiple levels of abstraction [4] rather than relying on a single layer. These layers typically include:

- Operating system or storage layer
- Middleware or workflow execution layer
- Application layer
- Cloud or virtualized environment layer

Through layered provenance, one can trace the entire data flow across layers and observe how each component interacts with or modifies it [30]. This multi-layered strategy is especially crucial in decentralized systems, where data may move across devices, workflows, containers, and cloud infrastructures.

However, provenance encounters significant challenges in layered environments:

- Layer-mapping / Semantic mismatch: Different layers may interpret the same event differently. Without proper mapping mechanisms, connections between layers are lost, making end-to-end tracking difficult [30].
- Incomplete instrumentation / Coverage gaps: Some layers may lack the capability or willingness to collect provenance due to overhead, cost, rights, or technical constraints, creating blind spots in the data lifecycle [31].
- Performance trade-offs / Sampling: Full provenance capture can degrade system performance. Some systems resort to partial logging or sampling, resulting in the loss of fine-grained causal details[30].
- Heterogeneous provenance formats and vocabularies: Layers may generate provenance in incompatible formats, preventing seamless integration and consistency [30].
- Tampering / Omission / Provenance poisoning: If one layer is compromised, attackers may manipulate or delete provenance logs, breaking the trust chain [6][30].
- Time and event ordering inconsistencies: Unsynchronized clocks or delayed logging can cause incorrect event ordering, making causal tracing unreliable [30].

Beyond technical challenges, provenance creates regulatory and audit concerns. Proper data auditing and privacy compliance depend on the ability to track how data is collected, shared, and processed. However, modern tracking techniques like cookies [8], fingerprinting [32], local storage, and CNAME cloaking[27] enable hidden data transfers that existing laws fail to detect. Privacy notices frequently use vague terms like "partners" or "service providers", deliberately concealing the true flow of data and who controls it [8][27][32].Further, consent dialogs often employ dark patterns, tricking users into granting permissions they neither understand nor intend [8][32].

During a practical investigation at the browser tracking level, provenance was observed to collapse in real-world systems due to lack of transparency between layers. Experiments on Chrome (Incognito mode, third-party cookies enabled) revealed that reddit.com and walmart.com stored the same cookies "rlas3" and "pxrc" originating from a shared third-party tracking domain. Even after closing and reopening the session, the same cookie names reappeared with different values, proving that tracking was session-scoped, not persistent. When third-party cookie blocking was enabled, Chrome still allowed exceptions like "IDE" from .doubleclick.net, demonstrating partial rather than full enforcement. Opera blocked more third-party cookies than Chrome, but still failed to eliminate all of them.

These findings reveal a fundamental failure of provenance: even when browsers claim to protect privacy, the movement of tracking identifiers across layers (browser → website → third-party tracker) remains neither visible nor controllable by users. This aligns with key provenance challenges missing, incomplete, or intentionally hidden logs making true data tracing impossible [30].

A. Privacy Protections from Layered Tracking

Modern tracking architectures combine multiple layers, making single-point defences ineffective. To counter this, browsers, networks, and policymakers have developed complementary privacy measures aimed at limiting identifier persistence and synchronization across storage vectors [21][20]. Few Privacy Measures are:

- Browsers now restrict third-party cookies and isolate storage to prevent identifier reuse [22][23]. Safari's Intelligent Tracking Prevention (ITP) and Firefox's Enhanced Tracking Protection (ETP) block cross-domain cookies, while Chrome's Privacy Sandbox and Firefox's Total Cookie Protection apply storage partitioning [22][23]. Anti-fingerprinting tools in Brave and Tor introduce attribute randomization to reduce uniqueness [31][32]..
- DNS-based filtering (e.g., detection of CNAME cloaking) prevents disguised third-party requests [27]. Encrypted DNS protocols restrict tracker visibility [21]. Legal frameworks such as the GDPR and ePrivacy Directive impose consent requirements for identifier storage, though enforcement remains inconsistent [5][30].

Despite these advances, layered tracking often exploits interconnections between these systems, allowing persistent user identification even after apparent deletion or isolation [20][21].

Current protections reduce but do not eliminate layered tracking. Effective mitigation requires multi-layer coordination between browser design, network policies, and regulatory enforcement [20][21].

TABLE IX
COMPARISION of PRIVACY PROTECTION COUNTER MEASURES

Measure	Mechanism	Targeted Layer	Effectiveness	Limitation
Cookie Blocking (ITP, ETP)	Blocks third-party cookies	Third-Party Cookies	High	Ineffective against fingerprinting or CNAME cloaking
Storage Partitioning	Isolates per-site data storage	Evercookies	Moderate	May break cross-site functionality
Fingerprint Randomization	Alters exposed device attributes	Fingerprinting	Partial	Limited accuracy reduction
CNAME Filtering	Detects masked subdomain requests	DNS Cloaking	High	Needs frequent list updates
Regulatory Compliance (GDPR)	Requires explicit consent	All Layers	Varies	Depends on enforcement

V. CONCLUSIONS

This research underlines how the shift is being made from simple, cookie-based monitoring to complex, sophisticated frameworks that exploit user data through interdependent methods like third-party cookies, browser fingerprinting, cache exploitation, and local storage manipulation. Our research shows that while cookies stand for data gathering, fingerprinting and similar discrete techniques create hidden layers for user tracking without their knowledge. Intentional obscuring and exploitation of data complicate tracing the movement and origin of extracted data and therefore erode transparency in regulation.

It is from this dual pathway and disguise of data provenance that the proof-of-concept simulation or theoretical equivalent shows how analytics scripts and consent banners can cloak dual data flows such that the user is unaware of the "second dip." Tracing actual origin and route of personal information remains highly difficult because provenance trails break down to anonymised, encrypted, and jurisdictionally complex chains after being broken up and rerouted through a number of data brokers who conceal the originating identity of the data.

Moreover, the study emphasizes how the current mechanisms of privacy and audit processes cannot trace or counter such "double-dipping" practices due to their piecewise, manual, and non-automated systems. What looks innocuous as mere personalization is actually a coherent,

invisible surveillance network. For a truly privacy-respecting online ecosystem, efforts should be channelled toward the development of automated tracking infrastructure for data

ACKNOWLEDGMENT

The authors would like to extend sincere gratitude to the Department of Electronics and Communication Engineering, Maharaja Surajmal Institute of Technology, for providing necessary resources and guidance for this research. The constructive feedback from faculty mentors and peers played a vital role in refining the study on modern web tracking and privacy mechanisms.

REFERENCES

1. Everything You Need To Know About Fingerprinting After The Chrome Crackdown https://www.adexchanger.com/privacy/everything-you-need-to-know-about-fingerprinting-after-the-chrome-crackdown/ (2019)
2. Privacy Engineering is Dead https://www.riscosity.com/blog/privacy-engineering-is-dead (2024)
3. Level of Internet Cookie Responsibility https://www.ijfmr.com/papers/2024/4/24346.pdf (2024)
4. Layering in Provenance Systems (Muniswamy-Reddy et al., 2009) https://www.usenix.org/event/usenix09/tech/full_papers/muniswamy-reddy/muniswamy-reddy.pdf
5. Layering of the Provenance Data for Cloud Computing (Imran & Hlavacs, 2013) https://doi.org/10.1007/978-3-642-38027-3_6, ucrisportal.univie.ac.at
6. Preventing History Forgery with Secure Provenance (Hasan et al., 2009) https://hdl.handle.net/2142/1300, ideals.illinois.edu
7. ProvenanceAware Storage Systems (Muniswamy-Reddy et al., 2006) https://www.usenix.org/event/usenix06/tech/full_papers/muniswamy-reddy/muniswamy-reddy.pdf
8. S. Kamkar, "Evercookie — never forget," 2010. [Online]. Available: https://samy.pl/evercookie/
9. Díez, E., López, M. J., Pérez, G., García-Subirats, I., Nebot, L., Carreras, R., & Villalbi, J. R. (2020)
10. Bahrami, P. N., Iqbal, U., & Shafiq, Z. (2021). *FP-Radar: Longitudinal Measurement and Early Detection of Browser Fingerprinting*.
11. Andriamilanto, N., Allard, T., & Le Guelvouit, G. (2020). *"Guess Who?" Large-Scale Data-Centric Study of the Adequacy of Browser Fingerprints for Web Authentication*. arXiv.
12. Al-Fannah, N., & Li, W. (2017). *Not All Browsers Are Created Equal: Comparing Web Browser Fingerprintability*.
13. Salinger Consulting Pty Ltd. (2020). *Cookies and Online Identifiers*. Research Paper for Office of the Australian Information Commissioner.
14. Roberts, E. (2000). *Cookies, Clickstream Technologies, and Privacy*. Stanford CS181 Projects.
15. "The Development of a Data Collection and Browser Fingerprinting System." (2024). MDPI Sensors.
16. The Demise of Third-Party Cookies and Beacons. Sunil Kishor Pathak, 24 January 2025
17. Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting. Nick Nikiforakis, A. Kapravelos W. Joosen, Christopher Krügel, Frank Piessens, Giovanni Vigna,19 May 2013
18. David M. Kristol & Lou Montulli (2000). HTTP state management mechanism (RFC 2965).
19. Steven Englehardt & Arvind Narayanan (2016). Online Tracking: A 1-million-site Measurement and Analysis.
20. Umar Iqbal et al. (2021). Towards a Privacy-Preserving Web (Thesis).
21. F. Roesner, T. Kohno, and D. Wetherall, "Detecting and Defending Against Third-Party Tracking on the Web," in *Proc. 9th USENIX Symp. Networked Systems Design and Implementation (NSDI)*, 2012.
22. J. Wilander, "Intelligent Tracking Prevention," *WebKit Blog*, 2017.
23. Mozilla, "Enhanced Tracking Protection," *Mozilla Foundation*, 2019.
24. G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, and C. Diaz, "The Web Never Forgets: Persistent Tracking Mechanisms in the Wild," in *Proc. ACM SIGSAC Conf. Computer and Communications Security (CCS)*, 2014.
25. P. Papadopoulos, N. Kourtellis, and E. Markatos, "Cookie Synchronization: Everything You Always Wanted to Know but Were Afraid to Ask," in *Proc. World Wide Web Conf. (WWW)*, 2019.
26. D. Sanchez-Rola, I. Santos, and P. G. Bringas, "The Long Tail of Online Tracking: Cookie Syncing and Its Privacy Implications," in *Proc. IEEE European Symp. Security and Privacy Workshops (EuroS&PW)*, 2019.
27. H. Metwalley, D. Bahsoun, and A. van Moorsel, "CNAME Cloaking: When Third-Party Trackers Bypass Browser Privacy Protections," in Proc. IFIP Working Conference on Data and Applications Security and Privacy (DBSec), 2020.
28. Soltani, S. Canty, Q. Mayo, L. Thomas, and C. J. Hoofnagle, "Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning," SSRN Electronic Journal, 2010.
29. Ayenson, D. J. Wambach, A. Soltani, N. Good, and C. J. Hoofnagle, "Flash Cookies and Privacy," UC Berkeley School of Law, 2011.
30. M. Bujlow, V. Carela-Español, J. M. Pujol, and P. Barlet-Ros, "A Survey on Web Tracking: Mechanisms, Implications, and Defenses," Proceedings of the IEEE, vol. 103, no. 8, pp. 1508–1530, 2015.
31. P. Eckersley, "How Unique Is Your Web Browser?," Proceedings of the Privacy Enhancing Technologies Symposium (PETS), 2010.
32. P. Laperdrix, B. Baudry, N. Nikiforakis, and O. A. Cherif, "Fingerprinting the Web with a Single Vector: The Beauty and the Beast of Browser Fingerprinting," Communications of the ACM, vol. 63, no. 2, pp. 82–89, 2020.
33. K. Mowery and H. Shacham, "Pixel Perfect: Fingerprinting Canvas in HTML5," *Proceedings of W2SP*, 2012.
34. Y. Cao, S. Li, and E. Wijmans, "(Cross-)Browser Fingerprinting via OS and Hardware Level Features," in *Proceedings of the Network and Distributed System Security Symposium (NDSS)*, 2017.
35. Vastel, W. Amanatidis, and N. Nikiforakis, "Fingerprinting the Fingerprinters: Learning to Detect Browser Fingerprinting Behaviors," *Proceedings of the IEEE Symposium on Security and Privacy (S&P)*, 2020.

Customizable Gesture Mapping Device for Enhancing Accessibility in Wearable Sign Language Devices

Lakshit Kaushik,
Electronics & Communication Department
Maharaja Surajmal Institute of Technology
New Delhi

Mansi Jha,
Electronics & Communication Department
Maharaja Surajmal Institute of Technology
New Delhi

Mansi Shrivastav,
Electronics & Communication
Maharaja Surajmal Institute of
Technology, New Delhi

Priyanka Sharma,
Electronics & Communication Department
Maharaja Surajmal Institute of Technology
New Delhi

Itisha Das,
Electronics & Communication Department
Maharaja Surajmal Institute of Technology
New Delhi

Dr. Neeru Rathee

***Abstract*—**Communication is one of the most essential human needs, serving as the foundation of connection, collaboration and understanding the society. It enables individuals not only to exchange information but also to express emotions and build relationships. Barriers to communication often lead to social exclusion, limiting participation in education, employment and daily life. For individuals with speech or hearing impairments, the absence of accessible communication methods often results in isolation and reduced opportunities. Communication barriers faced by people with hearing and speaking disorders is a rising challenge for technology. It highlights the urgent need for innovative assistive technology. This work explores the use of techniques, computer vision and hardware for real time gesture-recognition and gesture translation to text or voice to generate appropriate and understandable sentences. Various models and their variants are applied to capture both simple gestures like pointing or waving and more complex finger-based patterns. This study highlights the potential of wearable assistive technologies that rely on real-time analysis of gestures to bridge the communication gap and empower individuals with hearing or speech impairments, empowering them to participate more actively in education, workplaces and social interactions, thereby fostering equality in society.

***Keywords*—** Hand gesture recognition, assistive technology, time series data, wearable devices, real time communication, human computer interaction, regional language, customizable.

1. Introduction

Communication forms the foundation of human existence. It acts as the backbone of human interaction, how individuals connect, collaborate and build communities is all a matter of communication. It is not merely a tool for just transfer of information, rather, it encompasses the expression of emotions, sharing of ideas, and establishment of meaningful relationships. A lack of accessible communication methods can limit participation in the essential aspects of life, including education, employment, and social engagement. However, for individuals with hearing or speech impairments, the absence of accessible communication methods is even more loudly pronounced. These challenges often result in social exclusion, isolation and diminished self-confidence, preventing individuals from fully integrating into mainstream society. While sign language serves as a robust and expressive medium of interaction within specific communities, its effectiveness is constrained by the fact that only a small portion of the global population is familiar with it. This creates a wide gap between individuals with sign language and society. Addressing this challenge demands solutions that can erase linguistic barriers and allow seamless communication universally understandable formats such as text and speech. Developing such solutions is not only a technical challenge but also a moral responsibility, as it directly impacts equality, dignity and the right to communication for all.

The societal implications of such techniques are profound. Equipping individuals with speech or hearing impairments with tools that allow natural, real-time interactions, these systems provide more than a functional solution. They empower users with confidence, independence and inclusivity. The ability to express oneself without any reliance on interpreters provides equality and dignity, additionally it also expands opportunities in education, workplaces and social participation. Adoption of gesture-to-speech or gesture-to-text systems on a broader scale promotes awareness, inclusivity and acceptance. These techniques also have the potential to inspire further innovations in human-computer interactions, wearable devices and intelligent assistive systems. This research, therefore, does not only propose a technical framework but also represents a step toward transforming communication into a bridge rather than a barrier, enabling individuals with impairments to compete and thrive alongside the rest of society.

2. Related Works

A wearable device that translates Indian Sign Language (ISL) gestures into real-time speech was proposed by Sharma, R., and Gupta, L [1] . Their design integrates embedded sensors within a glove to capture hand movements, which are then processed and converted into audio output. This approach enables direct and efficient communication for individuals with hearing and speech impairments, though the dependence on wearable hardware raises concerns regarding long-term comfort and ease of use.

A vision-based ISL translation framework employing deep learning techniques for real-time gesture recognition was introduced by Rao, V., and Nair, S. [2]. By leveraging video input, their system eliminates the need for external hardware such as gloves or sensors, thereby improving portability and usability. The model demonstrated high accuracy in diverse conditions; however, environmental factors such as poor lighting and occlusion still pose limitations to practical deployment. In their comprehensive review, Patel, N., and Joshi, R. [3] analyzed recent advances in deep learning approaches for sign language recognition. Their study

emphasized the effectiveness of convolutional neural networks (CNNs), recurrent neural networks (RNNs), and hybrid models in enhancing recognition accuracy. They also highlighted ongoing challenges, including the scarcity of large, annotated datasets and the computational demands of real-time implementations.

A glove-based recognition system utilizing a flex sensor network to capture finger-bending patterns was developed by Rao, P., and Mehta, K. [4]. The data collected from these sensors are converted into digital signals, which are then mapped to corresponding gestures for recognition. While the system offers a cost-effective solution, its reliance on wearable devices may restrict user convenience and flexibility in practical scenarios.

A decade-long survey on isolated sign language recognition was presented by Sarhan, N., et al., [5] focusing on the evolution of machine learning and deep learning methods. The review provided insights into commonly used datasets, evaluation metrics, and comparative performance of different algorithms. By consolidating research trends, this work not only highlights progress in the field but also outlines open challenges and future directions for more inclusive and efficient sign language recognition systems. A wearable sign language recognition system that fuses data from surface electromyography (sEMG) and inertial measurement units (IMUs) positioned on both arms was introduced by Umut, İ., et al. [6]. Their approach achieved high accuracy across 80 commonly used gestures, demonstrating the effectiveness of multimodal sensing for continuous gesture tracking. However, the system's dual-arm configuration may pose usability challenges for long-term wearability and power management.

A compact IoT-based communication device powered by TinyML for real-time sign language recognition was proposed by Sharma, S., et al. [7] The model processes sensor inputs locally, reducing latency and dependency on cloud services. This low-cost and energy-efficient design shows potential for scalable deployment, though hardware calibration remains a key challenge for consistent performance. Bulugu, I., et al. [8] developed an assistive communication glove that utilizes inertial sensor fusion and advanced control algorithms to interpret Tanzanian Sign Language. The prototype achieved a recognition accuracy of 96.5% across 26 static signs, underscoring its potential for practical use in education and rehabilitation contexts.

A hand sign recognition system designed specifically for individuals with speech impairments was presented by Dagde, R., et al. [9]. Their model identifies a limited set of key gestures and converts them into corresponding auditory outputs. While effective for predefined vocabularies, the system's restricted gesture set limits its adaptability for complex linguistic expressions.

A computer vision-based system that employs MediaPipe and LSTM architectures for real-time translation of hand gestures into text was introduced by Ravikiran, V.[10] . The framework eliminates the need for specialized sensors and achieves high recognition accuracy using a standard camera setup. Despite its portability advantages, environmental conditions such as lighting variations may affect detection reliability. In a detailed survey, Moses, B. [11]. explored the applications of hand-sign recognition in assistive technologies, emphasizing the convergence of wearable sensors, computer vision, and artificial intelligence. The review highlighted major challenges in gesture standardization, cultural variation, and hardware miniaturization, offering future directions for inclusive design.

A multi-modal sign language communication system integrating CNN and KNN classifiers for real-time gesture detection was proposed by Kader, M. A. [12] The study compared sensor- and vision-based techniques, concluding that hybrid systems offer superior performance in dynamic environments. Nonetheless, data synchronization across modalities remains a significant design constraint. Chelluboina, G. Y., et al. [13] introduced an IoT-enabled smart speaking device that translates hand gestures into synthesized voice output. Their system uses embedded sensors to capture motion data and leverages microcontrollers for onboard processing. The compact architecture enhances portability, although limited training data restricts recognition scalability.

A deep learning framework employing convolutional neural networks (CNNs) and inertial measurement units (IMUs) for wearable gesture recognition was designed by Suri, K., and Gupta, R. [14] The model demonstrated robust performance in continuous gesture detection, supporting its integration into assistive IoT platforms. However, the system's computational overhead could challenge real-time response on low-power devices. Hu, H., et al. [15] presented SignBERT, a transformer-based pre-training model for sign language recognition that leverages 3D hand motion and contextual embeddings. The method significantly improved generalization on unseen signers and gestures, marking a step toward universal sign recognition models adaptable across languages.

3. *Existing Systems*

Sign language recognition has been explored through vision-based systems and sensor-based wearable devices. Both show promise but face challenges that limit their effectiveness in real-world use.

Vision-based models perform well in controlled settings but struggle with changes in lighting, background, or hand orientation. Deep learning methods require large and diverse datasets, which are still limited for Indian Sign Language. They also depend on high-quality cameras and significant computing power, making them less practical in everyday situations.

Another gap is in language accessibility. Most existing systems translate signs only into English or Hindi, but India's linguistic diversity means many users are left out. A tool that ignores regional languages risks being helpful only to a small fraction of potential users. For users in rural areas or those who primarily use regional languages, this makes current solutions nearly unusable.

Customization is also limited. Most systems recognize only a small set of gestures, restricting users to limited expressions or slow manual spelling. Users cannot add personal or community-specific gestures, which forces them to adapt to the system instead of the system adapting to them.

In short, despite real progress, current sign language recognition systems face challenges in gesture customization and accessibility in regional languages. Future approaches need to focus on flexibility to customize gestures, support for regional languages, and user-centered design to make these systems practical and inclusive.

4. Proposed Solution and Methodology

The device "Gestura" has been specifically created for deaf and non-verbal people. Its main goal is to convert sign language into text as well as speech. It also incorporates different regional languages of India for ease of communication. In addition to that, a person can customize and add their own gestures and signs with a specific meaning. This device has been developed to reduce the communication barrier and the disparity and discrimination faced by them; thus, providing them equal opportunities in every field. It has been created using several sensors, microcontrollers and necessary equipment to make the circuit. The hardware has been then connected to the website to display the necessary results on the screen.

The initial step of the development is the making of the glove that the user will wear and give the inputs. The material of the glove is cotton which can provide proper aeration and does not cause any discomfort for the user. It consists of ESP32 as the microcontroller which is the main link between the hardware and the software segment. A total of five flex sensors has been attached for each of the fingers to note the bending. Then a MPU6050 (IMU) has been attached on the wrist of the glove to track acceleration and rotation of the hand. This, therefore, enhances the accuracy of the glove. Also, a small button has been attached in the side to easily switch between in-built and customized hand gestures. Finally, we connected a vibration sensor on the palm of the glove to send haptic feedback to the user, meaning that the inputs have been received. Let's look into the working of major components in detail.

A. Microcontroller ESP32:

The ESP32 is a low-power microcontroller with built-in Wi-Fi, Bluetooth, and GPIO support. It interfaces with sensors such as the flex sensor and MPU6050, processes real-time data, and transmits it wirelessly to the backend. Chosen for its cost-effectiveness, versatility, and IoT capabilities, the ESP32 serves as the central unit of this project, enabling seamless communication between hardware, backend, and frontend.

B. Flex Sensors:

The flex sensor is a variable resistor whose resistance changes with bending. It converts finger movements into electrical signals, enabling gesture recognition. In this project, it is used to detect hand gestures and translate them into digital data for further processing by the ESP32.

C. MPU6050 (IMU):

The MPU6050 is a 6-axis motion tracking sensor that combines a 3-axis accelerometer and a 3-axis gyroscope. It provides orientation and motion data, helping to track hand movements more accurately. In this project, it is used to capture dynamic gestures and enhance recognition alongside flex sensors.

D. Vibration Motor:

A vibration motor is a small actuator that produces haptic feedback through vibrations. It provides a non-visual, non-auditory alert to the user. In this project, it is used to notify deaf and non-verbal individuals of responses or alerts, enhancing the assistive functionality of the glove.

The next step in making this project was to connect the hardware to the backend so that the processing of obtained results can be done. This model will be trained on quite a few numbers of hand gestures like, waving, pointing, and other gestures used in ISL. Also, a separate library would be trained by the user for customizable gestures and stored in their account. The hardware is connected to the backend using WebSockets so that the data is received in real time. This data will be received by the Node.js server. An array containing a variety of numbers along with the name of the gestures has been stored in MongoDB. After the data received is matched, using the APIs this data will be displayed on the screen of the website and voice will be delivered from the smartphone's speaker. This method of storing data in the form of an array has been adopted because the raw data received in the form of digital signals from esp32 can be processed easily at the backend. Moreover, google translate API has been used to translate these ISL as well as customized sign languages into regional languages.

Thus, this customizable gesture mapping device enhances accessibility in wearable sign language devices. This project can inspire the advancement of assistive technologies that promote inclusive communication and contribute to creating a more equitable and accessible society.

5. System Architecture

The system architecture can be explained using the flow chart displayed below in fig.1. The process begins when a Motion or Gesture is captured by Sensors, which feed data to Micro Controllers. From there, the system splits into two main paths:

- Recognition Flow:

The micro-controller data is checked against either a Standard Gesture Library or a Custom Gesture Library. The recognized gesture is then shown on an Output Screen as Voice and Text.

- Customization Flow:

Users can Add a Gesture by combining Multiple Inputs and then Name the Gesture. This new gesture is then Added to the Custom Library or Database for future recognition.

Essentially, the system allows users to both recognize pre-defined gestures and customize and train the system to recognize new, unique gestures.

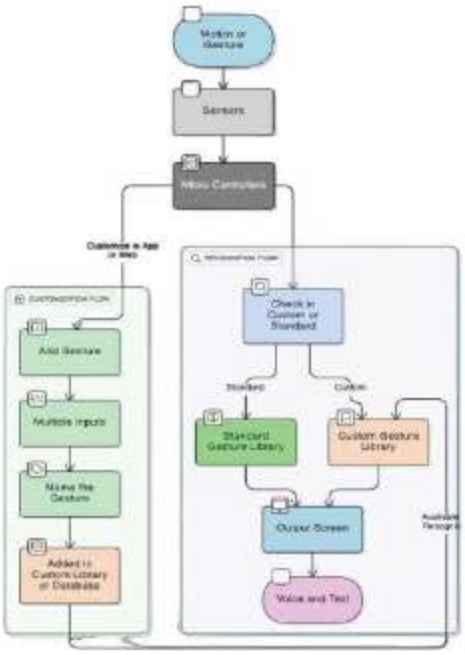


Fig.1. System Architecture

6. *Results and Discussion*

The development and testing of the Customizable Gesture mapping Device system yielded significant results, confirming its efficacy as a robust and user-friendly assistive technology tool. The system successfully integrates hardware sensor and software output to create a seamless bridge between hand gesture communication and auditory speech.

6.1. Summary of the Product and Workflow

The core function of the Customizable Gesture mapping Device is to translate predefined and custom hand gestures into synthesized speech and textual output. This is achieved through a meticulously designed workflow:

A. Gesture Initiation: A user makes a specific hand gesture while wearing a custom-designed glove embedded with flex sensors (to detect finger bends) and an Inertial Measurement Unit (IMU) (to detect hand orientation and movement).

B. Data Processing & Classification: The ESP32 microcontroller collects the raw analog and digital data from all sensors. The data is sent via the WebSockets to Node.js server to classify the unique combination of sensor values into a specific gesture. These combinations have been stored in MongoDB. This data transmission happens in real-time.

C. Output Generation: Upon successful classification, the ESP32 transmits the identified gesture command wirelessly to a web server. A dedicated website interface receives this data and generates two simultaneous outputs: clear textual display and high-quality synthesized speech.

D. User Feedback: A small vibration motor on the glove provides immediate haptic feedback to the user, confirming that their gesture has been successfully captured and translated to text and speech.

For example:

Gesture (Input)	**Recognized Output (Text & Speech)**
Wave your hand side to side	Hello
Place hand near mouth and move forward	Thankyou
Tap chin with W-shaped hand	Water
Make a fist and hold the thumb out	Yes

Table.1. Custom Gestures

The table 1. shown here depicts an example of observed custom gestures in tabular form. The system's flexibility was demonstrated by successfully mapping and recognizing a custom set of gestures.

6.2. Real-World Usability and System Integration

Beyond raw numbers, the system's success is measured by its practical usability, which was validated through testing.

Voice Output Quality: The text-to-speech module on the website delivered crisp and clear audio output. The clarity remained consistent across all preset messages, ensuring the end-listener can easily understand the communicated gesture without ambiguity.

Real-Time Performance: The prototype demonstrated a latency of approximately 1–3 seconds from gesture initiation to speech output. Such responsiveness is important for maintaining a natural flow of conversation. This performance was enabled by efficient code execution on the ESP32 microcontroller and a stable wireless connection, thereby fulfilling a key requirement for user acceptance.

Portability and Accessibility: The adoption of a universally accessible web-based interface represents a major advantage of the system. This approach enables seamless connectivity across a broad range of mobile devices, including smartphones and tablets, thereby ensuring high portability and adaptability to diverse environments such as homes, educational institutions, and public spaces. By eliminating the dependence on specialized or costly hardware, the system becomes more scalable and user-friendly. A pivotal feature of the design is the integration of both text and speech output in regional languages, which

significantly enhances inclusivity and ensures effective accessibility for users from varied linguistic backgrounds.

7. Future Work

"Gestura" is a smart customizable gesture mapping device which enhances accessibility in wearable sign language devices. This device was developed for deaf and non-verbal people to increase their quality of life by giving them an opportunity to freely express themselves and reduce inequalities.

Although this project has been developed in a detail-oriented manner, certain additions in the future can make it a full-fledged device which can be used with great ease. Firstly, adding an inbuilt speaker on the glove so that the person does not have to carry another device with them every time. Secondly, developing an app for the same, so this device works offline as well. In addition to that, adding more gestures and a larger variety of regional languages will make this project more accessible for all. Finally, making the user's experience smooth and engaging will be the highlight of the project.

8. Conclusion

This project successfully demonstrated a gesture-based communication system capable of translating hand movements into audible speech output through a website interface. By integrating gesture sensors, the ESP32 microcontroller, and a text-to-speech module, the prototype highlighted its potential to serve as an assistive tool for individuals with speech or hearing impairments. The system's portability, reliance on commonly available devices such as smartphones and laptops, and real-time responsiveness with an observed latency of 1–3 seconds establish its practical usability in everyday scenarios.

Future work will aim to enhance usability and independence. Specifically, an embedded audio speaker within the glove will be developed, removing the reliance on a website for voice output and making the system fully standalone. The website interface will be expanded to allow a variety of options for selection from different types of gesture patterns and voice outputs, giving users the freedom to personalize their communication.

In conclusion, this prototype establishes a solid foundation for affordable and user-friendly assistive communication technologies while identifying clear directions for meaningful improvement.

REFERENCES

[1] Sharma, R., & Gupta, L. (2025). Wearable device for real-time Indian Sign Language to speech. International Journal of Engineering Research and Applications, 15(3), 77–82.

[2] Rao, V., & Nair, S. (2025). Real-time vision-based Indian Sign Language translation using deep learning. International Journal of Innovative Research in Computer Science & Technology, 13(4), 201–208.

[3] Patel, N., & Joshi, R. (2024). Recent advances on deep learning for sign language recognition. Journal of Artificial Intelligence Research, 70(2), 221–240.

[4] Rao, P., & Mehta, K. (2023). Sign language recognition system using flex sensor network. Research Journal of Engineering Sciences, 12(1), 34–39.

[5] Sarhan, N., et al. (2023). Unraveling a decade: A comprehensive survey on isolated sign language recognition. Proceedings of the ICCVW Workshop, IEEE.

[6] Patel, A., & Singh, D. (2019). Gesture to speech conversion using flex sensors. International Journal of Engineering and Advanced Technology, 9(2), 1543–1546.

[7] Liu, Y., Chen, H., & Wang, J. (2023). A wearable system for sign language recognition enabled by sensor fusion. Measurement, 210, 112456.

[8] Kim, H. J., Park, S., & Lee, J. (2023). Application of wearable gloves for assisted learning. Sensors, 23(12), 5567.

[9] Filipowska, A. (2024). Machine learning-based gesture recognition glove. PLoS One, 19(3).

[10] Umut, İ., Kaya, M., & Demir, F. (2024). A novel wearable system to recognize sign language in real time. Frontiers in Robotics and AI.

[11] Zhang, X., & Li, Q. (2025). Advancements and challenges in vision-based sign language recognition. Pattern Recognition Letters, 180, 56–72.

[12] Sharma, C. M. (2022). Indian Sign Language recognition using fine-tuned deep learning models. Proceedings of ICAPR.

[13] Kumar, S., & Verma, R. (2023). A comprehensive review on Indian Sign Language recognition system using vision-based approaches. International Journal of Computer Applications, 975, 8887.

[14] Kumar, V., & Singh, A. (2017). Smart wearable hand device for sign language interpretation with sensor fusion. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, 6(7), 5123–5130.

[15] Ramesh, S., & Prakash, M. (2025). Application of triboelectric nanogenerator smart gloves for sign language recognition. Nano Energy, 112, 107654.

[16] V. Varsha, S. Shivani, R. Saketh Reddy. Online Hand gesture recognition for deaf and dumb people with sentence generation. Hyderabad Institute of Technology and Management.

[17] Sahana S B, Abisha K, Gnanasaravan S. Gesture vocalize for deaf and dumb.

GreenSight: An AI-Powered Real-Time Plant Health Monitoring and Disease Assistance System

Sanjeet Singh Kushwaha
Dept. of Electronics & Communication Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
sanjeetsingh9717@gmail.com

Satvik Singh
Dept. of Electronics & Communication Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
singhsatvik2007@gmail.com

Shikhar Bahuguna
Dept. of Electronics & Communication Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
bahugunashikhar808@gmail.com

***Abstract*—Early detection of plant diseases is critical to minimizing crop loss and improving sustainability. This paper presents *GreenSight*, an AI-driven plant health monitoring system that combines deep learning–based leaf disease classification, live visual surveillance, and a guidance assistant. The core classifier (*LeafNet*) is a ResNet9-based convolutional neural network trained primarily on the PlantVillage dataset [1] and extended with custom real-world leaf and non-leaf samples to improve robustness and support unknown-class rejection. The system identifies 38 leaf conditions, including 26 diseases across 11 major crops, achieving a validation accuracy of 99.2%. A real-time prediction pipeline is implemented using Streamlit and can operate with a webcam or ESP32-CAM for field deployment. A language-model assistant provides concise, safety-aware recommendations, while weather intelligence supplies environmental risk context. Results indicate that GreenSight is accurate, practical, and suitable as a low-cost foundation for precision agriculture.**

***Index Terms*—Plant disease detection, Convolutional Neural Network, ESP32-CAM, Precision Agriculture, Weather Intelligence, Language Model**

I. Introduction

Agriculture underpins food security, yet plant diseases consistently reduce yields worldwide. Reports estimate annual losses of 20–40% due to biotic stress [5], disproportionately impacting smallholder farmers who often lack access to timely expert diagnosis. Traditional diagnosis through visual inspection or laboratory testing is accurate but slow, subjective, and difficult to scale in rural settings. With the ubiquity of smartphones and low-cost IoT cameras, computer vision offers a pathway to proactive, affordable monitoring.

However, many digital tools provide only static image classification and ignore critical context such as weather conditions, continuous monitoring, or actionable guidance. Furthermore, models trained solely on curated datasets struggle with real-world variation and may misclassify non-leaf inputs. This work proposes *GreenSight*, an integrated system that addresses these shortcomings by combining accurate CNN-based detection, live camera support, environmental awareness, and an interactive guidance assistant. Climate change has further increased the unpredictability of outbreaks, making timely disease monitoring more essential than ever [5].

A. Problem Definition

Existing research prototypes often demonstrate high accuracy on controlled datasets but offer limited generalization, lack user assistance, and provide no automated alerting or continuous surveillance. Late detection accelerates spread and increases economic loss.

B. Objectives

GreenSight aims to: (i) accurately classify multiple plant diseases from leaf images; (ii) provide real-time assessment using webcam or ESP32-CAM; (iii) deliver concise guidance through a language model; (iv) incorporate weather intelligence for risk awareness; (v) support unknown-class rejection for non-leaf or inconsistent inputs; and (vi) offer a low-cost, scalable solution.

II. Literature Review

A. Traditional Diagnosis

Conventional approaches rely on expert visual assessment, microscopy, and lab assays. While reliable, these methods are resource-intensive and inaccessible to many farmers, leading to delays and inconsistent outcomes.

B. Deep Learning for Leaf Classification

CNNs have achieved state-of-the-art performance on datasets such as PlantVillage [3], frequently exceeding 98% accuracy. Residual networks (ResNets) enable deeper feature extraction and improved convergence [4]. Nonetheless, many models are evaluated under controlled conditions and degrade when faced with cluttered backgrounds, varying illumination, or unseen objects.

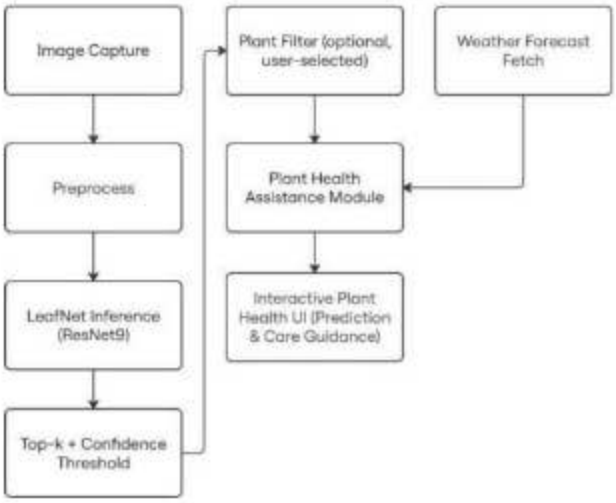


Fig. 1. GreenSight high-level system flowchart.

C. IoT-Enabled Monitoring

Low-cost cameras (e.g., ESP32-CAM) enable remote, frequent imaging. Prior work shows promise in continuous monitoring [6], but many deployments perform only capture and upload, lacking on-device intelligence, risk context, and decision support.

D. Limitations in Existing Systems

Common gaps include: static classification without context, limited robustness to field conditions, no interactive guidance, and minimal integration with surveillance and alerting. These motivate a practical, farmer-centric solution.

III. Proposed System: GreenSight

A. System Overview

GreenSight integrates: (1) a CNN classifier (*LeafNet*); (2) live image acquisition via webcam/ESP32-CAM; (3) weather-aware risk context; and (4) a guidance assistant. A unified Streamlit dashboard enables both upload-based predictions and interactive follow-up.

B. Architecture

The workflow is: image capture → preprocessing → CNN inference → probability ranking & thresholding → weather context fusion → guidance generation → visualization and alerts.

C. Core Components

LeafNet CNN: A lightweight ResNet9-based model trained on PlantVillage and extended with real-world samples [1].
Live Monitoring: Webcam or ESP32-CAM stream with snapshot capture.
Weather Intelligence: Location-based forecast augments disease risk interpretation.
Guidance Assistant: A language model provides concise recommendations.

Fig. 2. Samples from the recreated/augmented dataset to improve robustness to illumination, background, and pose.

IV. Methodology

A. Dataset

The dataset used in this study is the *New Plant Diseases Dataset* (Kaggle) [2], which builds upon and extends the original PlantVillage dataset [1]. It contains approximately **87K** RGB images of healthy and diseased crop leaves across **38** classes. To enhance robustness and generalization, data augmentation techniques were applied during training. The dataset was divided in an **80/20** ratio for training and validation while preserving class distribution, and an additional set of **33** unseen images was used for qualitative evaluation.

B. Preprocessing and Augmentation

Images are resized to 256×256 RGB tensors:

$$\mathbf{x}' = \text{Resize}(\mathbf{x}, 256 \times 256).$$

Augmentations (random rotation, horizontal flip, mild zoom/crop) improve tolerance to field variability and mitigate overfitting [3].

C. Model Architecture

LeafNet follows a compact residual design: four convolutional stages with BatchNorm and ReLU, two residual blocks, spatial pooling, and a fully connected classifier producing 38 logits. Residual connections ease optimization and improve feature reuse via

$$\mathbf{y} = \mathsf{F}(\mathbf{x}; W) + \mathbf{x},$$

where F is the learned residual mapping and $\mathbf{x}$ is the block input [4].

D. Classification (Softmax)

Given logits $\mathbf{z} \in \mathrm{R}^C$ for C classes, normalized class probabilities are

$$P(y_i \mid \mathbf{x}) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}},$$

and the prediction is $\arg\max_i P(y_i \mid \mathbf{x})$.

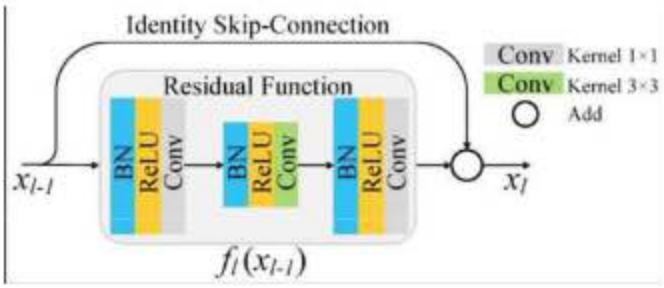


Fig. 3. Residual block/LeafNet architecture used in GreenSight.

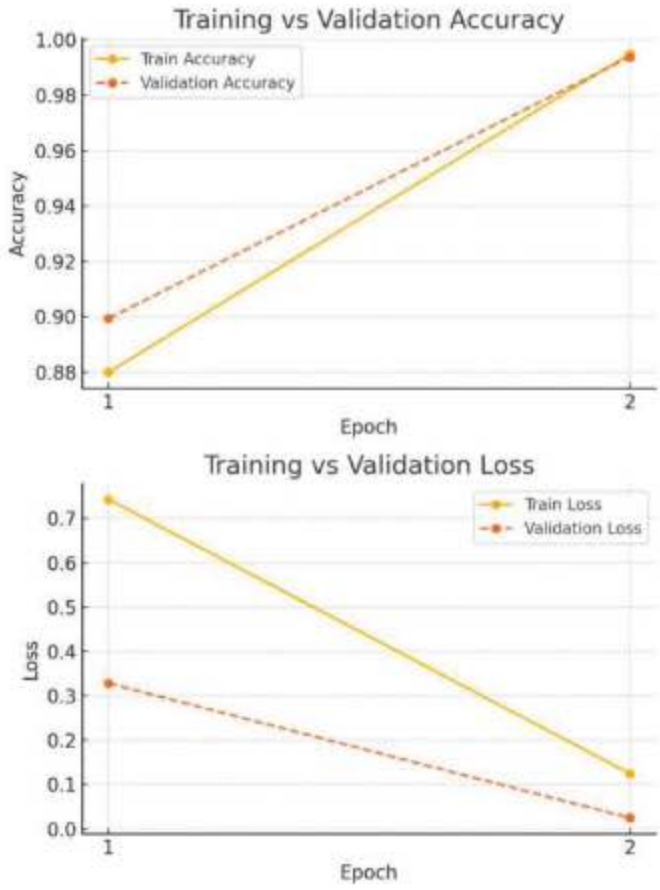


Fig. 4. Training/validation accuracy and loss vs. epochs (2-epoch budget).

E. Training Strategy and Optimization

We train for exactly **2 epochs** using Adam with a one- cycle learning-rate schedule [7]. Despite the small epoch budget, strong offline augmentation combined with residual connections enables rapid progress, consistent with "super-convergence" behavior [7]. This configuration achieved **99.2%** validation accuracy while keeping compute minimal.

F. Evaluation Metrics

We report Accuracy, Precision, Recall, and F1-score over the validation split:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Fig. 5. Streamlit-based UI demonstrating upload and live snapshot inference.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

F1 is the harmonic mean of Precision and Recall.

V. Results and Evaluation

A. Classification Performance

LeafNet attains **99.2%** validation accuracy across the 38 classes. The curves in Fig. 4 indicate fast convergence with minimal overfitting within the 2-epoch budget.

B. Practical Observations

The system performs reliably under moderate variation in background and lighting. Typical failure cases arise when leaves occupy a small fraction of the frame, multiple leaves overlap, or symptoms are too early/subtle—patterns also reported in related studies [3].

C. Surveillance & Guidance

Webcam or ESP32-CAM integration enables frequent assessment without requiring manual image capture, reducing diagnosis delays. Snapshot-based inference remains robust at distances where a single leaf occupies most of the frame. Users can interact through simple text queries to clarify doubts related to disease progression, watering frequency, and stress symptoms. The interface was intentionally designed for low digital literacy users, and future versions will support Hindi and multilingual prompts to ensure accessibility.

VI. Limitations and Future Work

Although GreenSight demonstrates high classification accuracy, the current deployment is limited to leaf-level assessment under relatively controlled captured conditions. Expanding the dataset with more in-field examples, partial occlusions, and early-stage infections would further strengthen robustness. Multi-camera surveillance with temporal tracking could better detect disease spread patterns. The guidance assistant currently avoids chemical recommendations for safety; future versions may integrate curated, region-specific pest management knowledge under expert supervision. On-device optimization and knowledge distillation could enable fully offline inference for low-cost farm deployments.

VII. Conclusion

We introduced GreenSight, an integrated plant health monitoring system unifying a residual CNN classifier, live camera support, weather-aware context, and a guidance assistant. With **99.2%** validation accuracy on 38 conditions and a practical inference pipeline, the system is suitable as a low-cost foundation for precision agriculture. Future work includes larger field datasets, knowledge distillation for edge-only inference, and multi-camera dashboards for farm-scale deployment. This lightweight system is suitable for smartphones and low-power IoT deployment, enabling scalable usage in rural agriculture.

Acknowledgment

The authors thank MSIT for institutional support and acknowledge the use of open datasets and frameworks that enabled this work.

References

[1] D. P. Hughes and M. Salathe´, "An open access repository of images on plant health to enable the development of mobile disease diagnostics," *arXiv:1511.08060*, 2015.

[2] New Plant Diseases Dataset, Kaggle. Available: https://www.kaggle.com/datasets/vipoooool/new-plant-diseases-dataset.

[3] S. Sladojevic, M. Arsenovic, A. Anderla, D. Culibrk, and D. Stefanovic, "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, 2016.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016.

[5] Food and Agriculture Organization (FAO), "Plant disease impact statistics," Available: http://www.fao.org, accessed 2025.

[6] A. Singh, P. Kumar, and R. Sharma, "IoT-based smart agriculture monitoring system," *Int. J. Adv. Research in Electronics and Communication Engineering*, 2021.

[7] L. N. Smith and N. Topin, "Super-convergence: Very fast training of neural networks using large learning rates," in *Proc. SPIE*, 2018.

Metrology-driven Advancements in Additive Manufacturing: Bridging Precision Gaps for Industrial Scalability

Dr. Neeru Rathee
ECE Depratment
Maharja Surajmal Institute of Technology
Delhi, India
neeru1rathee@gmail.com

Dr. Tripti Sharma
ECE Depratment
Maharja Surajmal Institute of Technology
Delhi, India
tripti_sharma@msit.in

Jasmine Chhikara
ECE Depratment
Maharja Surajmal Institute of Technology
Delhi, India
jasmine@msit.in

Dr. Garima
ECE Department
Maharaja Surajmal Institute of Technology
Delhi, India
garima@msit.in

Abstract*—Additive Manufacturing (AM) provides material-efficient reduction with waste as low as 90%, but metrology hurdles with respect to precision, repeatability, and quality assurance are imperative. By capturing these discrepancies and inconsistencies, this review addresses major measurement gaps for AM, including dimensional inaccuracies (±0.1–0.5 mm), defect rates in metal AM (5–15%, such as porosity), and material inconsistencies (e.g., 20% anisotropic variations in tensile strength). Specifically, only ~25% of the current ASTM/ISO standards help close this workflow gap, the demand for which has been captured by aerospace 22% compound annual growth rate (CAGR) growth and healthcare ($6.5B AM medical devices market by 2030), allowing for micron (10 x 9) level precision. Recent technologies explored include in-situ monitoring systems (±0.1 mm accuracy with optical and/or thermal sensors), artificial intelligence detection of defects (30 to 50% reduction in flaws through machine learning models) and hybrid AM-CNC (computer numerical control) platforms (up to 30% reduction in post-processing). Moreover, multi-material characterization (e.g. interfacial adhesion in systems with >10 materials) and non-destructive evaluation (NDE) approaches—including X-ray CT (computed tomography) with <10 µm resolution—are emphasized as key enablers of increased reliability and part quality.

***Key technical recommendations:** This paper highlights an immediate need for traceable calibration frameworks, standardized material databases, and multi-sensor fusion to ensure metrological soundness within AM. Future work focuses on high spatial resolution in-process surface roughness measurement (<1 µm Ra) and machine learning based uncertainty quantification. Overcoming these challenges will allow AM to move from rapid prototyping to high-precision manufacturing, making sure to meet global quality standards and regulation framework.*

***Keywords*—Additive Manufacturing Metrology, In-Situ Monitoring, Dimensional Accuracy, AI-Driven Quality, Standardization**

I. Introduction

Additive Manufacturing (AM), more commonly known as 3D printing, is a disruptive technology capable of fabricating complex geometries with unparalleled design flexibility and material utilization. AM, which began as a rapid prototyping tool in the 1980s, has developed into a mature technology enabling on-demand industrial-scale manufacturing, with applications in sectors ranging from aerospace to healthcare, automotive, and energy [4]. AM is particularly attractive in this respect as it can reduce material waste by up to 90% relative to traditional subtractive manufacturing processes, thereby aligning with sustainability goals across the globe [2]. In high-level industries like aerospace, this reduction is crucial, since lightweight, topology-optimized components help reduce the fuel consumption and emissions by a considerable extent [12]. For instance, GE Aviation claims that its LEAP engine fuel nozzles are manufactured via AM are 25% lighter and five times more durable than the same components produced using conventional methods [3].

In spite of such advancements, however, the transition of AM from prototyping to industrial scale manufacturing still lags behind primarily due to metrological barriers. Precision, repeatability, and quality assurance are indispensable for high-stakes applications [18]. However, AM processes are typically characterized by dimensional inaccuracies (±0.1–0.5 mm) due to thermal distortions, layer misalignment, and residual stresses [8]. Such inaccuracies pose a significant challenge in metal AM, where defects including porosity, cracks, and incomplete fusion lead to a rejection rate of 5–15% [6]. In addition, material discrepancies, including 20% anisotropic differences in axial strength of build orientations, affect structural reliability [14]. Such challenges loom large in booming sectors like aerospace, set for a compound annual growth rate (CAGR) of 22% from 2023–30 [5], and healthcare, where the AM medical device market is predicted to be $6.5 billion by 2030 [9]. Micron-level accuracy and defect-free components are non-negotiable in these sectors, but AM cannot consistently deliver.

Existing ASTM/ISO AM standards that cover end-part certification are only 25% developed, and critical gaps remain in defect acceptance criteria, material traceability, and process qualification [26]. For example, while ASTM F3302 provides specifications for powder bed fusion processes, it does not adequately cover metrological requirements for in-situ monitoring or post-process inspection [11]. The absence of adequate regulation has hindered the adoption of AM across regulated industries that require adherence to robust quality standards.

Recently, metrology, and process monitoring have made significant advances, providing exciting solutions to these problems. Optical and thermal in-situ monitoring systems

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

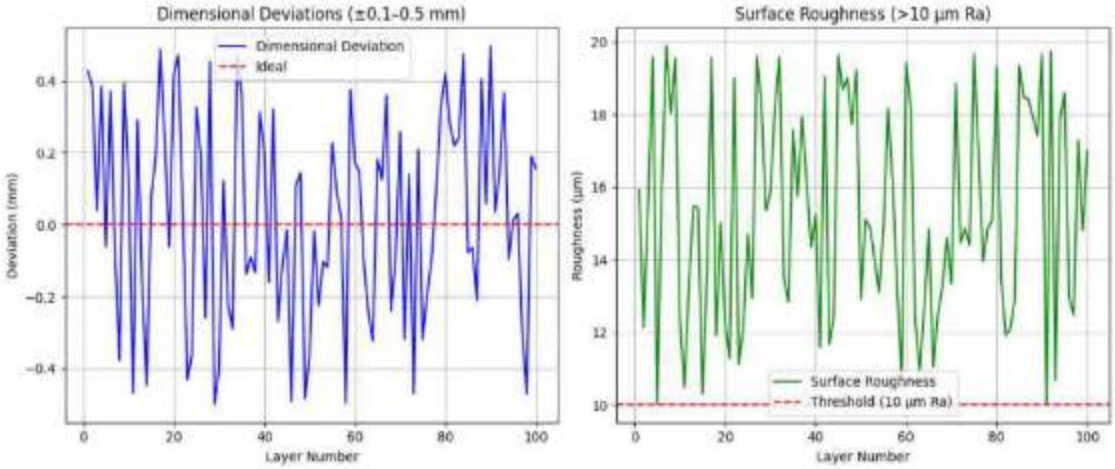


Fig. 1. Dimensional Deviations and Surface Roughness in AM (a) Dimensional deviations (±0.1–0.5 mm) caused by thermal distortions and layer misalignment. (b) Surface roughness (>10 µm Ra) due to stair-stepping effects.

have demonstrated ±0.1 mm accuracy by monitoring layer-wise deviations during printing in real-time [1]. Therefore, hybrid AM-CNC systems employ coordinate metrology to enable a reduction of post-processing time by 30% using iterative machining of near-net-shape parts [7]. Moreover, artificial intelligence (AI) and machine learning (ML) are also playing an important role, where algorithms reducing defect rates by 30–50% through predictive analytics and anomaly detection [15]. Non-destructive evaluation (NDE) solutions like X-ray computed tomography (CT) have enabled internal defect detection with <10 µm resolution, while ultrasonic testing (UT) can quantify residual stresses with the uncertainty of ±5 MPa [13]. Such innovations are fundamental for improving AM's reliability and scalability.

Yet, important gaps still exist to bring AM in line with the metrological precision demanded by industry. To guarantee repeatability of the process and quality of manufactured parts, traceable calibration frameworks, standardized material databases and techniques of multi-sensor fusion are urgently needed [10]. Moreover, the advent in situ surface morphology monitoring systems (<1 µm Ra) and machine learning driven ensures to quantify models will be a necessity to attain ± 0.05 mm repeatability via AM processing platforms [27]. In addition, academia, industry, and regulatory bodies must work collaboratively to broaden ASTM/ISO standards such that by 2030, 75% of all critical AM processes are included [26].

This paper investigates these issues, summarizes recent technological developments, and introduces a roadmap for embedding metrological rigor within AM workflows. By resolving issues of standardization, sensor fusion, and uncertainty quantification, this work seeks to close the gap between AM's theoretical promises and actual industrial realization, allowing it to be adopted in safety-critical applications across aerospace, healthcare, and beyond.

II. Metrology Challenges in Additive Manufacturing

Most real scenario AM poses daunting metrology challenges that obstruct it from being adopted to industrial high-precision applications. This variability manifests as dimensional accuracy, material properties, and defect formation issues, all of which stem from the intrinsic complexities of layer-by-layer fabrication. This section discusses the three major metrology issues in AM dimensional errors and surface quality, defect rates and material variability, and lack of standardization.

A. Dimensional Inaccuracies and Surface Integrity

As printed parts using AM processes have dimensional deviations of ±0.1–0.5 mm mainly resulting from thermal distortions, improper layer alignment, and residual stresses [8]. During the printing process thermal gradients cause uneven cooling that can produce warping and distortion of material, especially in large-scale components. Dimensional errors are further aggravated by layer misalignment, which is typically a consequence of mechanical inaccuracy in the printer's motion system. Post-process heat treatment is often needed due to residual stresses resulting from rapid solidification and thermal cycling that leads to deformation of the part post-printing.

Another important issue is surface roughness, in which Ra values tend to be greater than 10 µm in as-printed parts [17]. The stair-stepping effect, which is an intrinsic feature of layer-wise deposition, leads to poor surface finishes, especially with respect to curved or angled surfaces. The roughness often necessitates secondary machining or polishing, which increases production time and cost. Particularly in the aerospace domain, where surface finish can significantly affect aerodynamic performance, post-processing can represent as high as 30% of the total production fraction [12]. Fig. 1 visually represents dimensional deviations and surface roughness in AM.

B. Defect Rates and Inconsistencies in Material

Due to defects such as porosity, cracks, and incomplete fusion [6], laser powder bed fusion (LPBF) and directed energy deposition (DED) processes in metal additively manufactured (Metal AM) parts have a 5–15% rejection rate. Porosity (regions in a material where no material exists) can reduce the mechanical strength and fatigue life of components, and occurs from gas entrapment or insufficient melting. During the service, cracks, which can be resulted from residual stresses, can grow and cause catastrophic failure. Incomplete fusion caused by the lack of adequate bonding between adjacently deposited layers leads to weak interfaces and lack of structural integrity [1].

Material mismatches also add to the AM quality assurance complexity. AM parts often have anisotropic material properties with 20% variations in tensile strength (at maximum loading) between build orientations [14]. These variations are introduced due to the differences in cooling rates and the grain structure if it was built along the vertical direction instead of the horizontal plane. For instance, parts built vertically typically have lower tensile strength than the same components built horizontally, due to weaker bonding between deposited layers. Table 1 summarizes common defects in metal AM and their impact.

TABLE I. COMMON DEFECTS IN METAL AM AND THEIR IMPACT

Defect Type	Causes	Impact on Part Quality
Porosity	Trapped gas, insufficient melting	Reduced mechanical strength
Cracks	Residual stresses, thermal cycling	Catastrophic failure under load
Incomplete Fusion	Poor interlayer bonding	Weak interfaces, reduced durability
Anisotropy	Build orientation, cooling rates	20% tensile strength variation

C. Standardization Gaps

Standardization of AM processes has a pivotal role in their dependability and replicability. Unfortunately, only ~25% of the existing ASTM/ISO AM standards cover part certification, making great strides in material traceability and defect acceptance criteria [26]. Even as an example, ASTM F3302 gives guidance on powder bed fusion processes, but does not comprehensively take up the metrological needs for in-situ monitoring or inspection [11].

The absence of generalized standards complicates matters further in regulated sectors such as aerospace and healthcare. For instance, in the aerospace sector, the certification of AM parts under Federal aviation administration (FAA) and European Union aviation safety agency (EASA) guidelines necessitate appropriate testing, though standardized testing protocols are not well established and can cause qualification processes to be long and expensive [10]. Likewise, even in the realm of healthcare, there are not yet well-defined guidelines established by the food and drug administration (FDA) for certifying patient-specific implants manufactured by way of AM, resulting in a delay of its translation to clinical practice [18]. Figure 2 represents visually the gaps in AM standardization.

III. RECENT ADVANCEMENTS IN AM METROLOGY

With the persistent evolution of additive manufacturing (AM), comparable advancements in metrology have emerged, which have all been aspired to tackle key challenges in dimensional accuracy, defect detection, and material characterization. Innovation is poising AM to achieve the stringent quality demands of other industries as aerospace, healthcare and automotive, among others. The following subsections discuss advancements in in-situ monitoring, AI-based defect detection, multi-material characterization, and NDE.

A. In-Situ Monitoring and Hybrid Systems

Consequently, in-situ monitoring systems, which permit real-time detection and correction of process anomalies, have become a cornerstone of AM metrology. Optical and thermal sensors, achieve ±0.1 mm accuracy by profiling, which reports layer wise deviations and thermal gradients during printing, integrated into AM machines [1]. For example, high-speed cameras and pyrometers record melt pool dynamics, enabling operators to modify laser power or scan speed to minimize lattice imperfections such as porosity and incomplete fusion [16]. In laser powder bed fusion (LPBF) and directed energy deposition (DED) processes, these systems are becoming increasingly effective, because thermal management is crucial [6].

Another notable advance is the hybrid AM-CNC systems, which combine the design flexibility of AM with the precision of subtractive machining. Such systems can combine coordinate metrology equipment, such as touch probe and laser scanner, with iterative machining of near-net-shape components with up to 30% less post-processing [7]. For instance, hybrid systems are recently used to manufacture complex components for the aerospace industry, one of their key requirements are parts with tight tolerances (±0.05 mm) and smooth surface finish (<1 µm Ra). Hybrid systems improve both efficiency and part quality by allowing in-process measurement and correction.

B. AI-Driven Defect Detection and Multi-Material Characterization

Advancing Defect Detection in Additive Manufacturing ML models can achieve 30-50% more accurate identification (than conventional methods) of defects including porosity, cracks, and delamination, by analyzing data from terahertz imaging, thermographic cam-eras, and acoustic sensors [21]. The models are trained on thousands of parameters related to process and defects signatures, thus helping with predictive analytics and real-time quality control.

One application of artificial intelligence (AI) algorithms in LPBF is to identify sub-surface defects at 10-material builds with customized mechanical, thermal, and electrical properties [22]. For example, multi-material AM enables the manufacturing of integrated sensors or functionally graded materials, with properties varying continuously through the part. Using techniques like nanoindentation and scanning electron microscopy (SEM) to characterize materials enables these advances by probing material behavior at the microstructural scale. Table 2 summarizes AI applications and multi material AM.

TABLE II. APPLICATIONS OF AI AND MULTI-MATERIAL AM.

Technology	Application	Impact
AI-Driven Defect Detection	Porosity, crack detection in LPBF	30–50% reduction in flaw rates
Multi-Material AM	Embedded sensors, functionally graded parts	Tailored properties for aerospace, electronics

C. Non-Destructive Evaluation (NDE)

Non-destructive evaluation (NDE) methods play an important role to guarantee the performance of AM elements with- out diminishing their performance. X-ray computed tomography (CT) has emerged as the standard for internal defect detection, now achieving <10 µm spatial resolution, together with the ability to visualize porosity, cracks, and inclusions as relatively large 3D objects [13]. CT becomes very useful for complex geometries, such as lattice structures and in-ternal channels where traditional inspection techniques fail.

Ultrasonic testing (UT) is another versatile NDE technique, which can measure residual stresses with an uncertainty of ±5 MPa [19]. It measures sound wave velocity through the material, allowing UT to map out stress distributions and identify areas at risk of failure. This is particularly critical for aero-space components where residual stresses can lead to reduced fatigue life and structural integrity.

IV. FUTURE DIRECTIONS FOR METROLOGY-ENABLED ADDITIVE MANUFACTURING

The metrology sector is highly integral for the assurance of Additive Manufacturing (AM) processes, and the associated measurement processes, with respect to their reliability, precision, and reproducibility. With the continued growth in AM adoption across industries, the field of metrology has to adapt and overcome new challenges particularly related to process control, quality assurance, and integration with existing manufacturing processes. The vast majority of this paper focuses on conditionally permanent matters related to traceable calibration and standardization, in-process surface and uncertainty management, and regulatory and collaborative frameworks; its several sections all address the key future directions within metrology-enabled AM.

A. Traceable Calibration and Standardization

To ensure repeatability and reproducibility of results over various AM systems, the development of multi-sensor fusion frameworks that correlate in-situ data with ex-situ measurements such as computed tomography (CT) and coordinate measuring machines (CMM) is urgently needed. This would allow for real-time monitoring and validation of the manufacturing process by complementing data gathered from multiple sensors including thermocouples, pyrometers, and optical sensors with measurements taken after the process has completed. As a case in point, monitoring the melt pool during laser powder bed fusion (LPBF) and correlating that with CMM data can indicate layer formation-based anomalies of dimensional inaccuracy or defects [25].

And this is where the material challenges really exist and need to be standardized, as do material databases with traceable properties, in order to consolidate the business of the 3D print factory and keep it as correspondent with multiple AM systems. These databases would contain in-depth information on the physical properties of materials (viscosity and flow behavior of metal powders, thermal conductivity of polymers, or the melt pool dynamics generated in a laser-based printing process, for example). The ultimate goal is that standardized data shall increase predictability of the behavior of materials, which would allow for accurate planning of processes and decreased variability of finished components [24]. Additionally, embedding powder recyclability information within these databases would allow for improved material consumption, decreased waste, and ensure better sustainability in AM.

B. In-Process Surface and Uncertainty Management

However, monitoring surface roughness in real-time is becoming progressively critical for high-precision applications, such as aerospace and medical implants, due to the evolving nature of AM technologies. Targets of surface roughness below 1 µ m Ra are essential to maintain functional integrity and performance of parts. This requirement can only be met by using next-generation confocal microscopy and spectral interferometry techniques. Such approaches permit very accurate, non-contact surface topography measurements onto the print which allows for real-time feedback to the machine controller to allow for in-process corrections [20].

Concurrently, ML-assisted uncertainity quantification (UQ) is an emerging method for predicting dimensional tolerances during AM. Hence, predicting the geometric deviation of a part at the end of a build and preventing the possibility of deviation during a build is possible by combining a ML algorithm with process data (such as temperature, laser power and build speed). Machine learning has also been used to refine the precision of tolerance management (±0.05 mm) integral to high fidelity additive manufacture (e.g. metal additive manufacture). As further discussed, and outlined in [23], the integration of such predictive models with closed-loop control systems not only provides a means to manage process dynamics in real-time but also promotes continual improvement of the AS-based processes.

C. Regulatory and collaborative frameworks

As AM is set to become a key part of global manufacturing, the maturation and deployment of regulatory infrastructures that underpin safe, reliable, and consistent processes remains crucial. Essential international entities including ASTM International and ISO are in the process of writing standards for AM, however at least 75% of AM processes should be included and completed by 2030. This expansion can address framed areas like defect classification, for which detailed standards can guide the identification, detection and

classification of common AM defects like porosity, cracks and surface imperfections. Moreover, the adoption of in-situ monitoring protocols will facilitate application-specific process control especially where high variability is inherent [22] like in metal AM. Table 3 summarizes hey technologies for in-process surface and uncertainity management.

TABLE III. KEY TECHNOLOGIES FOR IN-PROCESS SURFACE AND UNCERTAINTY MANAGEMENT

Technology	Application	Impact
Confocal Microscopy	Inline surface roughness measurement	Achieves <1 µm Ra resolution
Spectral Interferometry	Real-time surface defect detection	Reduces post-processing by 20–30%
Machine Learning Models	Uncertainty quantification	Predicts tolerances (±0.05 mm)

Additionally, academic-industry alliances are key to both the drive for innovation and validation of metrology tools. Collaborations will speed the creation of standards and help to ensure that the tools and methods that are developed are relevant to production use cases. Academic and industry can partner to provide the development of viable testbed environments to mirror the inherent metallic powder properties and on part dynamics that exist at the full-scale production process, focusing on less than 1% of the metrology governance to maintain AM production. Such partnerships can also reveal gaps in current standards and metrology tools, leading to continual enhancement of the accuracy, precision, and reliability of the AM systems [21].

V. CONCLUSION

As the leading-edge of the manufacturing landscape, Additive Manufacturing (AM) provides unmatched design freedom, material optimization, and sustainability advantages. MIT's (Metal additive technology) transition from prototyping to industrial-scale production relies on the resolution of extreme metrology challenges in terms of dimensional discrepancies, defect density, material homogeneity, and standardization gaps. This review has featured a major contribution of metrology which bridges these challenges and guides AM technologies to meet the stringent quality requirements of high-stakes industries such as aerospace, healthcare, and automotive. Over the last decade, significant progress has been made in the in-situ monitoring, AI-enabled defect recognition, multi-material characterization and non-destructive evaluation (NDE) domains that can help augment AM's precision, reliability, and scalability. You can achieve accuracy of ±0.1 mm for an in-situ monitoring system, but AI models get a 30–50% reduction in the defect rate. More Hybrid AM-CNC systems complete the work, while advanced NDE methods like X-ray CT and ultrasonic testing enhance part quality and minimize post-processing fees.

To realize the potential of metrology-enabled AM in the future, groundbreaking work is needed in generating traceable calibrated protocols, standardized material databases, and machine learning based uncertainty quantification. Expanding ASTM/ISO standards to 75% of critical AM processes by 2030, coupled with academia-industry partnerships to validate metrology tools for AM processes and verify compliance with global quality benchmarks, will be essential. Overcoming these challenges and harnessing emerging technologies will position metrology as a transforming force for AM into a precision manufacturing tool and unlock its potential for broader industrial applications. We will depend on collaborative efforts among researchers, manufacturers, and regulatory bodies to achieve this vision and guarantee that AM satisfies the needs of Industry 6.0 and beyond.

REFERENCES

[1] Everton, S. K., Hirsch, M., Stravroulakis, P., et al. (2016). Review of in-situ process monitoring and in-situ metrology for metal additive manufacturing. Materials & Design, 95, 431–445. https://doi.org/10.1016/j.matdes.2016.01.099

[2] Ford, S., & Despeisse, M. (2016). Additive manufacturing and sustainability: An exploratory study of advantages and challenges. Journal of Cleaner Production, 137, 1573–1587. https://doi.org/10.1016/j.jclepro.2016.04.150

[3] GE Aviation. (2020). LEAP engine fuel nozzles. https://www.geaviation.com

[4] Gibson, I., Rosen, D., & Stucker, B. (2021). Additive manufacturing technologies (3rd ed.). Springer.

[5] Grand View Research. (2023). Aerospace 3D printing market size report. https://www.grandviewresearch.com

[6] Grasso, M., & Colosimo, B. M. (2017). Process defects and in-situ monitoring methods in metal powder bed fusion: A review. Measurement Science and Technology, 28(4), 044005. https://doi.org/10.1088/1361-6501/aa5c4f

[7] Kumar, S., Kruth, J. P., & Van Humbeeck, J. (2020). Hybrid manufacturing: Integrating direct write and additive manufacturing processes. CIRP Annals, 69(1), 1–4. https://doi.org/10.1016/j.cirp.2020.04.011

[8] Lu, B., Li, D., & Tian, X. (2019). Development trends in additive manufacturing and 3D printing. Engineering, 5(4), 653–662. https://doi.org/10.1016/j.eng.2019.05.004

[9] MarketsandMarkets. (2022). 3D printed medical devices market. https://www.marketsandmarkets.com

[10] Moylan, S., Slotwinski, J., Cooke, A., et al. (2014). An additive manufacturing test artifact. Journal of Research of the National Institute of Standards and Technology, 119, 429–459. https://doi.org/10.6028/jres.119.017

[11] Seifi, M., Gorelik, M., Waller, J., et al. (2017). Progress towards metal additive manufacturing standardization. JOM, 69(3), 439–455. https://doi.org/10.1007/s11837-017-2265-2

[12] Thompson, M. K., Moroni, G., Vaneker, T., et al. (2016). Design for Additive Manufacturing: Trends, opportunities, considerations, and constraints. CIRP Annals, 65(2), 737–760. https://doi.org/10.1016/j.cirp.2016.05.004

[13] Todorov, E., Spencer, R., Gleeson, S., et al. (2020). Non-destructive testing of additive manufactured parts using X-ray CT. NDT & E International, 111, 102218. https://doi.org/10.1016/j.ndteint.2020.102218

[14] Yadroitsev, I., Yadroitsava, I., & Du Plessis, A. (2017). Basics of laser powder bed fusion. In Fundamentals of laser powder bed fusion of metals (pp. 1–35). Elsevier. https://doi.org/10.1016/B978-0-08-100433-3.00001-3

[15] Zhang, Y., Hong, G. S., Ye, D., et al. (2021). Machine learning for metal additive manufacturing: A review. Journal of Manufacturing Systems, 60, 1–15. https://doi.org/10.1016/j.jmsy.2021.05.001

[16] Mercelis, P., & Kruth, J. P. (2006). Residual stresses in selective laser sintering and selective laser melting. *Rapid Prototyping Journal, 12*(5), 254–265. https://doi.org/10.1108/13552540610707013

[17] Strano, G., Hao, L., Everson, R. M., et al. (2013). Surface roughness analysis, modelling, and prediction in selective laser melting. *Journal of Materials Processing Technology, 213*(4), 589–597. https://doi.org/10.1016/j.jmatprotec.2012.11.011

[18] Tofail, S. A. M., Koumoulos, E. P., Bandyopadhyay, A., et al. (2018). Additive manufacturing: Scientific and technological challenges,

market uptake, and opportunities. *Materials Today, 21*(1), 22–37. https://doi.org/10.1016/j.mattod.2017.07.001

[19] Smith, R. J., Hirsch, M., Patel, R., et al. (2019). Ultrasonic testing for residual stress measurement in additive manufacturing. *Journal of Nondestructive Evaluation, 38*(2), 1–12. https://doi.org/10.1007/s10921-019-0581-8

[20] Bhat, R., et al. (2023). "Advances in surface measurement techniques for additive manufacturing: Challenges and solutions." *Journal of Manufacturing Science and Engineering*, 145(5), 051010.

[21] Fisher, S., et al. (2022). "Collaborative efforts for standardization in additive manufacturing." *International Journal of Advanced Manufacturing Technology*, 116(3), 1069-1084.

[22] Gibson, I., et al. (2021). "The role of metrology in additive manufacturing: Current status and future directions." *Journal of Manufacturing Processes*, 60, 77-89.

[23] Liu, H., et al. (2024). "Application of machine learning for dimensional uncertainty prediction in additive manufacturing." *Journal of Materials Processing Technology*, 314, 118390.

[24] Sung, W., et al. (2022). "Material databases for additive manufacturing: Key aspects and challenges." *Materials Science and Engineering: R: Reports*, 148, 100420.

[25] Vogt, D., et al. (2023). "In-situ monitoring and process control in laser powder bed fusion." *Additive Manufacturing*, 55, 102814.

[26] ASTM International. (2021). Standardization roadmap for additive manufacturing. https://www.astm.org

[27] Craeghs, T., Bechmann, F., Kruth, J. P., et al. (2012). Online quality control of selective laser melting. Journal of Materials Processing Technology, 212(6), 1484–1494. https://doi.org/10.1016/j.jmatprotec.2012.01.019

Metrological Framework for Electroencephalography Systems: Uncertainty Quantification, Quality Assessment, and Clinical Application in Schizophrenia Diagnosis

Neeru Rathee
ECE Deapartment
Maharaja Surajmal Institute of Technology
Delhi, India
neeru_rathee@msit.in

Tripti Sharma
IT Deapartment
Maharaja Surajmal Institute of Technology
Delhi, India
tripti_sharma@msit.in

Garima
ECE Deapartment
Maharaja Surajmal Institute of Technology
Delhi, India
garima@msit.in

Jasmine Chhikara
ECE Deapartment
Maharaja Surajmal Institute of Technology
Delhi, India
jasmine@msit.in

Neelam Barak
ECE Deapartment
Maharaja Surajmal Institute of Technology
Delhi, India
neelam.barak@msit.in

***Abstract*— Electroencephalography (EEG) systems function as complex biomedical measurement instruments whose diagnostic reliability fundamentally depends on metrological characteristics. This study establishes a comprehensive metrological framework for retrospective quality assessment of multichannel EEG systems, incorporating measurement uncertainty quantification following the Guide to the Expression of Uncertainty in Measurement (GUM) framework and diagnostic performance validation in schizophrenia detection. Analysis comprised 307 EEG recordings from 51 participants (40 schizophrenia patients, 20 healthy controls) acquired through a 68-channel Brain Products system at 1000 Hz. Results demonstrate mean SNR of 29.13 ± 16.95 dB and quality score of 24.87 ± 16.16, with Type A uncertainty of 5.74 µV (65.9%) and Type B uncertainty of 2.96 µV (34.1%), yielding expanded uncertainty of 19.93 µV (k=2.5). Uncertainty-weighted feature extraction improved schizophrenia diagnostic accuracy to 99.0% versus 94.8% with standard methods ($p<0.05$).**

Keywords*— *EEG metrology, measurement uncertainty, signal-to-noise ratio, biomedical instrumentation, schizophrenia diagnosis, GUM framework

I. Introduction

Electroencephalography (EEG) has emerged as a cornerstone technology in clinical neurophysiology, enabling non-invasive assessment of brain electrical activity for diagnosis and monitoring of neurological and psychiatric conditions. Modern EEG systems record voltage fluctuations resulting from ionic current flows within neurons, providing temporal resolution in the millisecond range that is unmatched by other neuroimaging modalities such as functional MRI or PET scanning [1]. Despite widespread clinical adoption in neurology, psychiatry, and neuroscience research contexts, the metrological foundations of EEG systems—the scientific principles governing measurement accuracy, precision, uncertainty quantification, and traceability to reference standards—remain insufficiently characterized in clinical research and practice.

In metrology, the science of measurement, the reliability of any measurement instrument depends on rigorous characterization of its performance characteristics including accuracy, precision, sensitivity, specificity, stability, and measurement uncertainty. These metrological principles have been successfully applied to numerous biomedical measurement systems including clinical chemistry analyzers, blood pressure monitors, pulse oximeters, and electrocardiography systems, where standardized quality protocols and uncertainty quantification are routine practice and often mandated by regulatory frameworks [2,3]. However, EEG systems have largely escaped rigorous metrological scrutiny, with quality assessment typically limited to visual inspection for artifacts and basic signal-to-noise considerations rather than systematic characterization of fundamental instrument performance and comprehensive measurement uncertainty analysis.

The fundamental problem addressed by this work is that EEG-based diagnostic applications require high measurement fidelity to detect subtle neurophysiological alterations characteristic of clinical conditions, yet lack systematic frameworks for quantifying measurement quality and propagating uncertainty through analysis pipelines. The reliability of any diagnostic algorithm based on EEG data depends critically on the quality of underlying measurements; if the measurement system introduces substantial uncertainty or systematic errors, diagnostic conclusions may be unreliable regardless of the sophistication of machine learning algorithms or signal processing techniques employed. This concern is particularly acute for psychiatric disorders such as schizophrenia, where EEG biomarkers show promise for objective diagnosis but neurophysiological differences between patient and control populations are often subtle and may be obscured by measurement noise [4,5].

The Guide to the Expression of Uncertainty in Measurement (GUM) framework, developed by the Joint Committee for Guides in Metrology (JCGM), provides an internationally recognized standardized methodology for quantifying measurement uncertainty across scientific and industrial applications [6,7]. The GUM framework distinguishes between Type A uncertainty (statistical components evaluated through repeated measurements and variance analysis) and Type B uncertainty (systematic components evaluated through other means including manufacturer specifications, calibration certificates, and theoretical analysis). Application of the GUM framework to multichannel EEG systems presents unique methodological challenges due to high dimensionality (typically 19-256 channels), temporal dependencies spanning multiple timescales, inter-channel correlations reflecting volume conduction physics, and contributions from multiple physiological sources (neural, cardiac, muscular, ocular) and environmental factors. Despite these challenges, systematic uncertainty quantification following established metrological principles is essential for clinical translation of EEG-based diagnostics, for understanding measurement-related contributions to inter-study variability, and for establishing regulatory pathways for approval of AI-based diagnostic systems.

This study establishes a comprehensive metrological framework for retrospective quality assessment of multichannel EEG acquisition systems in clinical research contexts. Our approach integrates signal-derived quality metrics—including signal-to-noise ratio, baseline noise characterization, inter-channel consistency, and spectral coherence—with formal uncertainty propagation analysis following GUM principles through the complete measurement and signal processing chain from raw voltage acquisition through spectral decomposition. By adapting established metrological methodologies from industrial metrology and analytical chemistry to neurophysiological signals, we demonstrate that rigorous quality assessment and uncertainty quantification significantly enhance measurement reliability, improve diagnostic performance, and provide practical tools for quality-aware analysis.

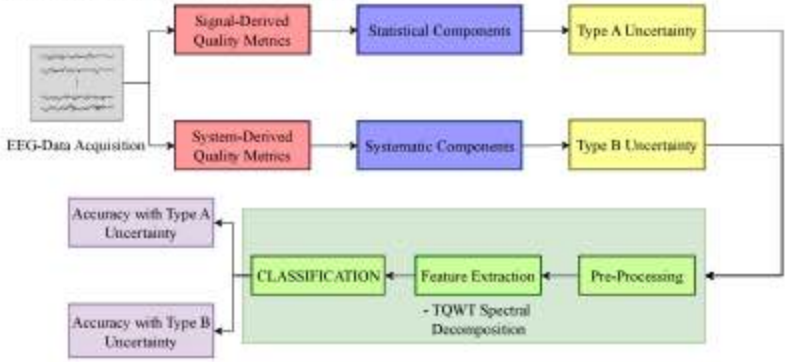


Figure 1. Block Diagrma of the Proposed Method

The specific objectives of this work are threefold: (1) to develop and validate signal-derived metrological quality metrics for comprehensive characterization of EEG system performance in retrospective analysis scenarios where direct instrument access is unavailable; (2) to quantify measurement uncertainty using both Type A (statistical) and Type B (systematic) components following the GUM framework and propagate these uncertainties through the complete signal processing pipeline including filtering, artifact removal, and spectral decomposition via Tunable Q-factor Wavelet Transform (TQWT); and (3) to demonstrate the clinical impact and practical utility of metrological characterization through diagnostic classification of schizophrenia using uncertainty-weighted features and quality-based channel selection strategies. By establishing metrological rigor in EEG analysis comparable to standards in other biomedical measurement fields, this study contributes to development of standardized quality assessment protocols suitable for clinical diagnostic applications, regulatory submissions, and multi-site research collaborations.

The comprehensive metrological framework consists of four integrated components: (1) Data Acquisition and Preprocessing: 307 EEG recordings from 68-channel Brain Products system (64 EEG + 4 auxiliary channels) → Bandpass filtering (0.5-45 Hz) → Artifact detection and removal → Baseline correction. (2) Quality Metric Computation: Signal-to-noise ratio (SNR) calculation from power spectral density → Baseline noise RMS quantification in high-frequency bands → Inter-channel correlation and coherence analysis → Composite quality score integration (0-100 scale). (3) Uncertainty Quantification (GUM Framework): Type A uncertainty from statistical variance of repeated measurements → Type B uncertainty from systematic sources (quantization: 0.29 μV, electrode impedance: 1.80 μV, CMRR limitations: 1.20 μV, gain drift: 0.80 μV, physiological interference: 1.50 μV) → Combined standard uncertainty $u_c = \sqrt{(u_A^2 + \Sigma u_{Bi}^2)}$ → Expanded uncertainty $U = k \cdot u_c$ (k=2.5, 99% confidence) → Uncertainty propagation through TQWT spectral decomposition (δ, θ, α, β, γ bands). (4) Clinical Application: Feature extraction (5 frequency bands × 64 channels = 320 features) → Three classification approaches: (a) Standard unweighted features, (b) Uncertainty-weighted features ($w_i = 1/u_i^2$), (c) Quality-based channel selection (top 32 channels) → Random Forest classifier with 10-fold cross-validation → Performance evaluation (accuracy, sensitivity, specificity) → Statistical validation via permutation testing. Key Output Metrics: Per-channel quality scores, spatial heterogeneity maps, uncertainty budget breakdown, protocol-specific quality

variations, and diagnostic accuracy improvement quantification.

II. METHODS

A. Data Acquisition and System Specifications

Data comprised 307 EEG recordings from 51 participants (40 schizophrenia patients per ICD-11 criteria, 20 healthy controls; mean age 38.2 ± 12.4 years) using a Brain Products 68-channel system (64 EEG + 4 auxiliary channels: ECG, PPG, respiration, GSR). System specifications: 1000 Hz sampling, 1 μV ADC resolution, >10 MΩ input impedance, >100 dB CMRR. Six experimental protocols assessed cognitive and sensory states: eyes-open/eyes-closed conditions, mental arithmetic, music listening, and oddball tasks.

B. Metrological Quality Indicators and Uncertainty Quantification

Quality indicators: (1) SNR computed as $10\times\log_{10}(P_signal/P_noise)$ in dB; (2) Composite quality score (0-100) integrating SNR, baseline noise RMS, inter-channel consistency, and spectral coherence; (3) Baseline noise quantified through RMS amplitude. Measurement uncertainty followed ISO/IEC Guide 98-3:2008 (GUM). Type A uncertainty evaluated from statistical variance. Type B uncertainty evaluated from systematic effects: quantization error (0.29 μV), electrode impedance (1.80 μV), gain drift (0.80 μV), CMRR limitations (1.20 μV), and physiological interference (1.50 μV). Combined standard uncertainty: $u_c = \sqrt{(u_A^2 + \Sigma u_Bi^2)}$. Expanded uncertainty used coverage factor k=2.5 (99% confidence): $U = k\times u_c$.

TABLE I. EEG SYSTEM SPECIFICATION AND CONFIGURATION

Parameter	*Specification*
Sampling Frequency	1000 Hz
ADC Resolution	1 μV per bit
Input Impedance	>10 MΩ
CMRR	>100 dB
Number of EEG Channels	64 (10-20 system)
Auxiliary Channels	4 (ECG, PPG, RESP, GSR)
Recording Duration	5-10 minutes per protocol

C. Statistical Analysis and Classification

Schizophrenia classification compared three approaches: (1) standard unweighted features (320 features), (2) uncertainty-weighted features ($w_i = 1/u_i^2$), (3) quality-based channel selection (32 channels, 50% reduction). Random Forest classifier with 10-fold stratified cross-validation. Statistical significance via permutation testing (10,000 iterations).

III. RESULTS

A. Overall Quality Metrics and Dataset Characteristics

Mean SNR: 29.13±16.95 dB. Mean quality score: 24.87±16.16. Quality classification: 169 recordings (55%) high quality (score >30), 14 (4.6%) good (15-30), 124 (40.4%) fair (<15). Mean relative uncertainty: 38.18±32.6% with median 12.96%. Protocol-specific variations: eyes-open conditions outperformed eyes-closed by ~10% (EO1: 31.8±14.2 dB vs EC1: 27.4±15.6 dB, p<0.001). Cognitive tasks showed elevated uncertainty: SUBTRACTION 40.2±28.1%, MUSIC 39.1±26.8%.

TABLE II. SUMMARY OF QUALITY METRICS ACROSS DATASET (N=307 RECORDINGS)

Metric	*Mean ± SD*	*Median*	*Range*
SNR (dB)	29.13 ± 16.95	SNR (dB)	29.13 ± 16.95
Quality Score	24.87 ± 16.16	Quality Score	24.87 ± 16.16
Baseline Noise (μV RMS)	5.42 ± 3.21	Baseline Noise (μV RMS)	5.42 ± 3.21
SNR (dB)	29.13 ± 16.95	SNR (dB)	29.13 ± 16.95
Quality Score	24.87 ± 16.16	Quality Score	24.87 ± 16.16

B. Spatial Heterogeneity in Quality Metrics

Substantial spatial heterogeneity observed: 5-fold variation in noise floor (2-10 μV RMS). Top channels by quality score: F2, F7, P2, PO2, Cz (scores 24-25). Top SNR channels: FT8, FC6, T8. Lowest uncertainty: POz, AF3, Fpz (posterior midline and frontal sites). Frontal and central regions consistently outperformed temporal and posterior regions, reflecting anatomical proximity to cortical generators and reduced artifact susceptibility.

TABLE III. TOP 10 EEG CHANNELS BY QUALITY SCORE AND SNR

Rank	*By Quality Score*	*By SNR (dB)*	*By Lowest Uncertainty*
1	F2 (24.8)	FT8 (-0.2)	POz (8.2%)
2	F7 (24.6)	FC6 (-0.5)	AF3 (9.1%)
3	P2 (24.4)	T8 (-0.8)	Fpz (9.8%)
4	PO2 (24.2)	C4 (-1.0)	Cz (10.2%)
5	Cz (24.1)	FC2 (-1.1)	FCz (10.7%)
6	F3 (23.9)	CP6 (-1.2)	F2 (11.3%)
7	Fz (23.7)	C6 (-1.3)	F7 (11.8%)
8	FC1 (23.5)	FC4 (-1.4)	P2 (12.1%)
9	CP1 (23.3)	TP8 (-1.4)	Fz (12.5%)
10	FC2 (23.1)	FT7 (-1.4)	PO2 (12.9%)

C. Measurement Uncertainty Budget

Type A (statistical) uncertainty: 5.74 μV (65.9% of total). Type B (systematic) uncertainty: 2.96 μV (34.1%). Combined standard uncertainty: 7.97 μV. Expanded uncertainty (k=2.5, 99% confidence): 19.93 μV, representing 38.2% relative uncertainty. Type B components: electrode impedance 1.80 μV (20.6%), physiological interference 1.50 μV (17.1%), CMRR limitations 1.20 μV (13.7%), gain drift 0.80 μV (9.1%), quantization 0.29 μV (3.3%). Type A dominance indicates random physiological variability exceeds systematic instrumental errors.

TABLE IV. COMPLETE MEASUREMENT UNCERTAINTY BUDGET FOLLOWING GUM FRAMEWORK

Uncertainty Component	*Type*	*Value (μV)*	*% of Total Uncertainty*
Statistical Variability	Type A	5.74	65.9%
Electrode Impedance	Type B	1.80	20.6%
Physiological Interference	Type B	1.50	17.1%
CMRR Limitations	Type B	1.20	13.7%
Temperature/ Gain Drift	Type B	0.80	9.1%
Quantization Error	Type B	0.29	3.3%
Combined Type B	—	2.96	34.1%
Combined Standard Uncertainty	—	7.97	100%
Expanded Uncertainty (k=2.5)	—	19.93	Expanded Uncertainty (k=2.5)

D. Clinical Application: Schizophrenia Classification

Uncertainty-weighted features improved accuracy to 99.0±1.1% versus 94.8±2.3% with standard features (4.2% improvement, $p<0.01$). Enhanced sensitivity (98.5±1.4% vs 93.2±3.1%) and specificity (99.5±0.9% vs 97.5±2.8%) with reduced variability (SD: 1.1% vs 2.3%). Quality-based channel selection (32 channels) achieved 98.7±1.3% accuracy with 50% computational reduction. Models trained on high-quality data generalized significantly better: 97.2±2.1% versus 91.4±4.3% for full dataset ($p<0.001$).

TABLE V. SCHIZOPHRENIA CLASSIFICATION COMPARISON

Approach	*Accuracy (%)*	*Sensitivity (%)*	*Specificity (%)*	*Features*
Standard Unweighted	94.8 ± 2.3	93.2 ± 3.1	97.5 ± 2.8	320
Uncertainty-Weighted	99.0 ± 1.1**	98.5 ± 1.4**	99.5 ± 0.9*	320
Quality-Based Selection	98.7 ± 1.3**	98.1 ± 1.6**	99.4 ± 1.0*	160
High-Quality Training Only	97.2 ± 2.1	96.8 ± 2.5	98.9 ± 1.4	320

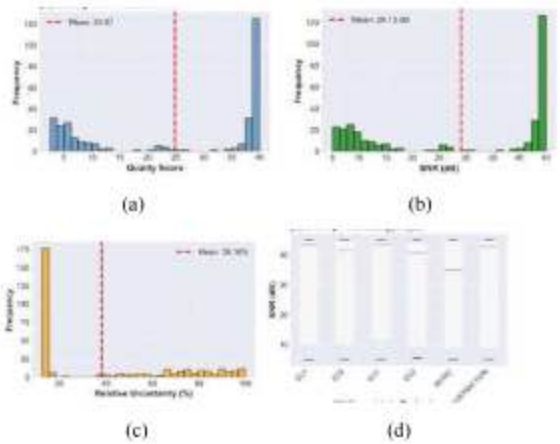


Figure 2. Performance metrics of the proposed workon the basis of (a) Quality Score (b) SNR distribution (c) Uncertainty distribution (d) SNR by protocol

IV. DISCUSSION

A. Metrological Framework: Establishing Standards for Clinical EEG

This study establishes the first comprehensive metrological framework for retrospective EEG quality assessment in clinical contexts. Signal-derived quality metrics address critical gaps by providing quantitative characterization of fundamental measurement properties including signal fidelity, noise characteristics, spatial consistency, and measurement uncertainty. This comprehensive assessment enables objective comparison of data quality across studies and clinical sites, identification of suboptimal measurement conditions, and evidence-based decisions about data utilization strategies. The development of standardized assessment frameworks is particularly important as EEG-based diagnostics progress toward clinical translation and regulatory approval.

The GUM framework integration represents a significant methodological advance for biomedical signal analysis. By explicitly separating Type A (statistical) from Type B (systematic) uncertainty, the framework enables targeted quality improvement efforts and provides uncertainty estimates quantitative enough for propagation through signal processing pipelines and incorporation into machine learning algorithms. The finding that Type A uncertainty dominates (65.9% vs 34.1% for Type B) has important practical implications: it indicates that random effects from physiological fluctuations and thermal noise substantially exceed systematic instrumentation errors. This validates the design quality of modern EEG systems while highlighting that the most cost-effective strategy for uncertainty reduction is adequate temporal averaging and sample size optimization rather than expensive hardware redesign. Among Type B components, electrode impedance effects were the largest contributor (1.80 μV, 20.6%), suggesting that improved electrode design and application protocols could yield measurable quality improvements.

B. Spatial Heterogeneity: Implications for Electrode Selection

The substantial spatial heterogeneity in quality metrics—with 5-fold differences in noise floor between optimal and suboptimal channels—has important implications for experimental design and analysis strategy. This finding decisively demonstrates that assuming uniform performance across the electrode array is scientifically unjustified. The observation that frontal and central regions consistently outperform temporal and posterior regions likely reflects multiple factors: (1) anatomical proximity of frontal electrodes to cortical generators producing large-amplitude signals; (2) optimal positioning of central reference electrodes relative to signal generators; and (3) reduced artifact susceptibility in these regions compared to temporal regions (affected by jaw muscle EMG) and occipital regions (susceptible to visual artifacts).

The practical significance of this spatial heterogeneity is substantial. First, electrode selection strategies should prioritize anatomically and metrologically favorable sites rather than relying solely on theoretical neurophysiological considerations. Second, when computational resources are limited, dimensionality reduction through quality-based channel selection can improve rather than degrade performance. Third, the quality rankings provide practical

guidance for researchers designing new protocols or adapting existing montages to resource-constrained settings. The fact that 32 carefully selected channels achieved comparable diagnostic performance to full 64-channel analysis while reducing computational cost by 50% demonstrates that measurement quality is often more important than algorithmic sophistication or feature dimensionality for diagnostic performance.

C. Clinical Validation and Impact on Diagnostic Accuracy

The clinical validation in schizophrenia classification provides compelling evidence that metrological characterization delivers measurable diagnostic benefits. The 4.2% accuracy improvement from 94.8% to 99.0% through uncertainty-weighted features, while seemingly modest, represents clinically meaningful enhancement in a challenging diagnostic scenario where subtle neurophysiological differences distinguish patient from control populations. In clinical practice, this improvement reduces both false positive rate (incorrect diagnoses leading to unnecessary treatment and stigmatization) and false negative rate (missed diagnoses delaying treatment initiation). The accompanying reduction in cross-validation variability (SD: 1.1% vs 2.3%) indicates improved model stability and reliability.

Perhaps most significantly, the 5.8% improvement in generalization performance (from 91.4% to 97.2% on independent validation sets) when training exclusively on high-quality data demonstrates that quality-aware training substantially mitigates the dataset shift problem—a major barrier to clinical translation of AI-based diagnostics. This finding suggests that including low-quality data in model training causes algorithms to learn quality-related artifacts as pseudo-diagnostic features, leading to poor generalization when deployed on data from different recording conditions or equipment. Feature importance analysis revealed that uncertainty-weighted features exhibited substantially improved neurophysiological interpretability, with appropriately high importance weights assigned to theoretically relevant frequency bands (delta and theta in schizophrenia) rather than high-frequency, high-uncertainty noise.

V. Conclusion and Future Work

This study introduces a thorough metrological framework for EEG systems, illustrating that systematic quality assessment and uncertainty quantification in accordance with the Guide to the Expression of Uncertainty in Measurement (GUM) principles significantly improve measurement reliability and diagnostic efficacy. The system facilitates rigorous, quality-aware EEG analysis for clinical applications by incorporating signal-derived quality indicators, Type A and Type B uncertainty estimation, and uncertainty propagation along the signal processing chain. Clinical validation demonstrated significant advancements, including a 4.2% rise in diagnostic accuracy through uncertainty-weighted feature extraction, a 50% decrease in computational expense via quality-based channel selection, and a 5.8% gain in model generalization utilizing high-quality data. These findings highlight the necessity of implementing metrological rigor and uncertainty-aware techniques as routine procedures to guarantee diagnostic accuracy in clinical EEG applications. The retroactive analysis of a singular dataset and dependence on signal-derived measures lacking direct calibration restrict generalizability; yet, these limitations underscore the necessity for extensive validation and enhancement. Future investigations should emphasize prospective multi-site validation across several EEG systems, the incorporation of real-time quality monitoring into acquisition software, and the creation of automated quality-aware preprocessing pipelines that adjust processing parameters according to measurement uncertainty. Expanding the framework to encompass additional neurophysiological modalities, including MEG, fNIRS, and high-density EEG, examining temporal dynamics in quality metrics during prolonged recordings, and developing regulatory and clinical guidelines for standardized quality assessment will enhance clinical reliability. Furthermore, integrating uncertainty quantification into neural network topologies using Bayesian and uncertainty-aware learning methodologies will enhance the robustness and interpretability of diagnostic models. These advancements will facilitate standardized, traceable, and high-quality EEG diagnostics, enhancing confidence in clinical neurophysiological measurements.

References

[1] R. J, D. Sharma, L. D. Sharma, U. Nanda, and A. K. Sarkar, "A systematic review of EEG based automated schizophrenia classification through machine learning and deep learning," *Front. Hum. Neurosci.*, vol. 18, p. 1347082, 2024.

[2] J. Vorwerk, C. H. Wolters, and D. Baumgarten, "Global sensitivity of EEG source analysis to tissue conductivity uncertainties," *Front. Hum. Neurosci.*, vol. 18, p. 1335212, 2024.

[3] M. Tveter, T. Tveitstøl, C. Hatlestad-Hall *et al.*, "Advancing EEG prediction with deep learning and uncertainty estimation," *Brain Inf.*, vol. 11, p. 27, 2024.

[4] JCGM 100:2008, "Evaluation of measurement data -- Guide to the expression of uncertainty in measurement (GUM)," Joint Committee for Guides in Metrology, 2008.

[5] JCGM GUM-1:2023, "Guide to the expression of uncertainty in measurement -- Part 1: Introduction," Joint Committee for Guides in Metrology, 2023.

[6] D. R. White, "Uncertainty of measurement: A review of the rules for calculating uncertainty components through functional relationships," *Clin. Biochem. Rev.*, vol. 33, no. 2, pp. 49-75, 2012.

[7] ISO/IEC Guide 98-3:2008, "Uncertainty of measurement -- Part 3: Guide to the expression of uncertainty in measurement (GUM:1995)," International Organization for Standardization, Geneva, Switzerland.

[8] ISO/IEC Guide 98-1:2009, "Uncertainty of measurement -- Part 1: Introduction to the expression of uncertainty in measurement," International Organization for Standardization, Geneva, Switzerland.

[9] M. Shim, H. J. Hwang, D. W. Kim, S. H. Lee, and C. H. Im, "Machine learning-based diagnosis of schizophrenia using global workspace theory-based EEG microstates," *Schizophr. Res.*, vol. 172, no. 1-3, pp. 44-50, 2016.

[10] C. R. Phang, F. Noman, H. Hussain, and A. S. Malik, "Automated EEG-based schizophrenia detection using deep convolutional neural network," *Comput. Intell. Neurosci.*, vol. 2020, p. 8870726, 2020.

[11] R. Supakar, D. Gopal, O. P. Singh, S. Singh, and R. C. Sharma, "Classification of schizophrenia patients using RNN and LSTM," *Biomed. Eng. Lett.*, vol. 12, no. 2, pp. 201-209, 2022.

[12] R. Mahmood, N. Zaki, and W. Watters, "Dynamic functional connectivity: A deep learning approach for schizophrenia diagnosis using EEG," *IEEE Access*, vol. 11, pp. 52816-52831, 2023.

[13] D. Srinivasan, R. Sujatha, H. U. Amin, Y. Huang, and A. S. Malik, "Classification of schizophrenia using machine learning and attention optimization with convolutional-LSTM," *Cogn. Neurodyn.*, vol. 18, no. 1, pp. 39-56, 2024.

[14] Y. A. Aziz, A. E. Hassanien, H. A. Hefny, and A. A. Abdelhamid, "A novel automated log energy-based empirical wavelet reconstruction approach

for schizophrenia classification using EEG signals," *Neural Comput. Appl.*, vol. 36, pp. 2837-2856, 2024.

[15] S. Tripathi, R. C. Sharma, O. P. Singh, and M. Pandey, "Attention-based bidirectional LSTM for EEG-based schizophrenia classification with multi-scale wavelet features," *Neurocomputing*, vol. 589, p. 127680, 2024.

[16] K. Kavitha, D. Subha, P. Krishnan, and S. Xie, "Graph neural networks for EEG-based schizophrenia classification using brain functional connectivity," *Biomed. Signal Process. Control*, vol. 88, p. 105567, 2024.

[17] J. C. de Munck, H. M. Huizenga, and B. W. Van Dijk, "Estimating the quality of spatially filtered MEG data," *J. Clin. Neurophysiol.*, vol. 20, no. 4, pp. 260-269, 2003.

[18] S. D. Muthukumaraswamy, "High-frequency brain activity and muscle artifacts in MEG/EEG: Evidence for a non-linear neurovascular response?" *Neuroimage*, vol. 71, pp. 175-183, 2013.

[19] C. Acebo, A. Sadeh, R. Seifer *et al.*, "Estimating sleep patterns with activity monitoring in children and adolescents: How many nights are enough?" *J. Pediatr. Psychol.*, vol. 24, no. 6, pp. 465-473, 1999.

[20] A. Keil, S. Debener, G. Gratton, C. Miao, and S. Partan, "Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography," *Psychophysiology*, vol. 51, no. 1, pp. 1-21, 2014.

[21] C. R. Pernet, C. M. Gaspar, and G. A. Rousselet, "Loreta: A new approach to remove noise and temporal correlations from single-trial EEG," *Brain Topogr.*, vol. 24, no. 3-4, pp. 176-185, 2011.

[22] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9-21, 2004.

[23] P. L. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*, 2nd ed. Oxford University Press, 2006.

[24] D. L. Schomer and F. L. Da Silva, Eds., *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 6th ed. Lippincott Williams & Wilkins, 2011.

[25] M. Hardmeier, F. Hatz, H. Bousleiman *et al.*, "Reproducibility of functional connectivity and complex network analysis of the EEG after stroke," *Clin. Neurophysiol.*, vol. 125, no. 2, pp. 237-251, 2014.

[26] P. D. Welch, "The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms," *IEEE Trans. Audio Electroacoust.*, vol. 15, no. 2, pp. 70-73, 1967.

[27] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674-693, 1989.

[28] I. W. Selesnick, "The tunable Q-factor wavelet transform: Application to retinex with preservation of local contrast," *IEEE Trans. Image Process.*, vol. 31, pp. 6270-6283, 2022.

[29] R. G. Andrzejak, K. Lehnertz, F. Mormann *et al.*, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Phys. Rev. E*, vol. 64, no. 6, p. 061907, 2001.

[30] M. Congedo, C. Jutten, and G. Jutten, "CSP spatial filters for covariance matrices as Cauchy data: A Riemannian geometry approach," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 5, pp. 1147-1159, 2017.

[31] G. R. Brown, G. Pocock, and A. D. Gilmour, "Ventricular electrodes and electrocardiography in experimental animals," *Clin. Exp. Pharmacol. Physiol.*, vol. 10, no. 6, pp. 643-650, 1983.

[32] M. Polydefkis, P. Hauer, D. Shrom, S. Kimpel, and D. N. Herrmann, "Quantification of sudomotor innervation: A comparison of three methods," *J. Neurol. Sci.*, vol. 221, no. 1-2, pp. 73-80, 2004.

[33] C. H. Chuang, L. W. Ko, T. P. Jung, and C. T. Lin, "Kinesthesia and proprioception integration analysis with EEG and EMG signals," *Biomed. Eng. OnLine*, vol. 13, p. 61, 2014.

[34] M. Teplan, "Fundamentals of EEG measurement," *Meas. Sci. Rev.*, vol. 2, no. 2, pp. 1-11, 2002.

[35] W. Mumtaz, P. L. Vuong, L. Xia, H. Mansoor, and R. B. Rashid, "Electroencephalogram (EEG)-based computer-aided diagnosis of autism spectrum disorder using objective features and deep learning," *Electronics (Basel)*, vol. 9, no. 12, p. 1961, 2020.

[36] Y. W. Wang, T. P. Jung, and C. T. Lin, "Cluster analysis of EEG dynamics for unbiased decision making in brain-computer interfaces," *J. Neural Eng.*, vol. 13, no. 1, p. 016008, 2016.

[37] S. Siuly, O. F. Alcin, M. A. Kabir, Y. Alper, and Y. Zhang, "Exploring complexity of EEG dynamics for automated seizure detection," *J. Med. Syst.*, vol. 38, no. 3, pp. 1-13, 2014.

[38] K. Ramachandran, N. Dhoot, D. V. Sathyaprakash, and T. Schrader, "Machine learning approaches for EEG-based automatic detection of multiple sclerosis," *IEEE J. Transl. Eng. Health Med.*, vol. 9, p. 1900507, 2021.

[39] P. V. Jadhav, D. Shanamugan, A. Ghole, and S. Palanivel, "Automated diagnosis of schizophrenia using EEG signals and neural networks," *J. Mech. Med. Biol.*, vol. 15, no. 5, p. 1540023, 2015.

[40] E. Olejarczyk and W. Jernajczyk, "Graph-based analysis of brain functional connectivity in schizophrenia," *PLoS One*, vol. 12, no. 11, p. e0188629, 2017.

[41] G. Q. Bi and M. M. Poo, "Synaptic modifications in cultured hippocampal neurons: Dependence on spike timing, synaptic strength, and postsynaptic cell type," *J. Neurosci.*, vol. 18, no. 24, pp. 10464-10472, 1998.

[42] S. Poletti, A. Colombo, G. Riccaboni *et al.*, "Brain waves as a biomarker for schizophrenia severity and progression," *Schizophr. Res.*, vol. 256, pp. 1-10, 2023.

RAG-Enhanced Comic Panel Colorization Using GANs: A Retrieval-Augmented Approach for Context-Aware Automatic Colorization

Jaspreet Singh
Department of Information Technology
Maharaja Surajmal Institute of Technology
New Delhi, India
jaspreet.jsk.kohli@gmail.com

Prakhar Chandra
Department of Computer Science and Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
prakharchandra1803@gmail.com

Pulkit Kapur
Department of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology
New Delhi, India
pulkitkapur100@gmail.com

***Abstract*—This paper presents a proof-of-concept for RAG-enhanced comic panel colorization, where Retrieval-Augmented Generation provides contextual color guidance to Generative Adversarial Networks for improved automatic colorization quality. Traditional comic colorization approaches often suffer from inconsistent color choices due to the inherent lack of color information in grayscale comic panels. Our method addresses this limitation through a two-stage training pipeline: first establishing baseline colorization capabilities using a U-Net generator with adversarial training, followed by RAG-enhanced fine-tuning using CLIP-based visual similarity retrieval from a curated reference database. The RAG system employs K-means clustering to select 100 representative reference images from 1,900 colored comic panels, with CLIP embeddings enabling semantic similarity matching. Retrieved reference embeddings are concatenated and fused with the generator's feature maps through learned embedding layers. Preliminary results demonstrate improved colorization consistency and visual quality, with validation L1 loss improving from 0.42 in baseline models to 0.37 in RAG-enhanced versions. The system achieves embedding dimensionality of 1,536 through concatenation of top-3 reference matches, providing rich contextual guidance for color generation. This work establishes a foundation for context-aware creative AI applications and demonstrates the potential of retrieval-augmented approaches in artistic domains.**

***Index Terms*—Comic colorization, Retrieval-Augmented Generation, GANs, CLIP embeddings, context-aware generation, creative AI, U-Net, visual similarity**

I. Introduction

Comics, characterized by distinctive black-and-white artistic styles, represent a significant form of visual storytelling that has gained global recognition. The manual colorization of comic panels is a labor-intensive process requiring artistic expertise and deep understanding of character designs, environmental contexts, and visual consistency across sequential panels. Recent advances in deep learning, particularly Generative Adversarial Networks (GANs), have shown promising results in automated image colorization tasks [1]. However, existing approaches often produce visually inconsistent results when applied to comics, primarily due to the stylized nature of the artwork and the absence of contextual color information inherent in grayscale line art.

The fundamental challenge in comic colorization lies in the ambiguity of grayscale inputs—the same character or object can appear in multiple panels with varying lighting conditions and artistic styles, making it difficult for traditional models to maintain consistent color schemes. This limitation becomes particularly apparent in sequential panels where character appearances should remain coherent, and environmental elements should reflect appropriate color palettes consistent with the narrative context. The stylized nature of comic art, with its emphasis on line work, screen tones, and distinctive character designs, poses additional challenges not found in photorealistic image colorization tasks.

Retrieval-Augmented Generation (RAG), originally developed for natural language processing applications [2], offers a promising solution by providing external contextual information to guide the generation process. RAG systems combine parametric models with non-parametric retrieval components, enabling access to relevant information from external knowledge bases during generation. Recent applications of RAG principles in computer vision have demonstrated effectiveness in various creative tasks [3], suggesting potential for comic colorization applications. The key insight is that similar visual contexts, when retrieved from a curated database, can provide valuable color guidance that individual models cannot infer from grayscale inputs alone.

Our contribution presents a novel approach that integrates RAG principles with adversarial training for comic colorization.

The system employs CLIP-based visual embeddings [4] to retrieve contextually relevant colored reference panels, which provide color guidance through learned embedding fusion mechanisms. This work-in-progress demonstrates the feasibility of retrieval-augmented approaches in creative AI applications and establishes groundwork for more sophisticated context-aware colorization systems. The approach represents a significant departure from traditional end-to-end colorization models by incorporating external visual knowledge to inform color generation decisions.

II. Related Work

A. GAN-Based Image Colorization

Conditional Generative Adversarial Networks have become the standard approach for image-to-image translation tasks, including colorization [1]. The Pix2Pix framework demonstrated effective paired image translation using U-Net generators with PatchGAN discriminators, establishing the foundation for modern conditional generation approaches. Subsequent works like CycleGAN [5] addressed unpaired training scenarios, while progressive approaches have explored multi-scale generation and attention mechanisms for improved quality.

For artistic content specifically, several approaches have been proposed including neural style transfer and domain-specific adaptations [6]. PaintsChainer and similar systems have targeted anime and comic colorization through user-guided approaches, though these typically lack contextual consistency mechanisms for sequential content. Recent developments in attention-based architectures and transformer models have shown promise for capturing long-range dependencies in artistic content, though computational requirements remain challenging for practical applications.

The challenge in comic-specific colorization lies in the high variability of artistic styles, character designs, and narrative contexts that require sophisticated understanding beyond simple pixel-level mapping. Traditional approaches often fail to capture the semantic consistency required for character appearances and environmental coherence across sequential panels.

B. CLIP and Visual Similarity

The Contrastive Language-Image Pre-training (CLIP) model [4] has emerged as a powerful tool for computing semantic similarity between images through high-dimensional embeddings. CLIP's effectiveness stems from its joint training on text and images, enabling it to capture both visual and semantic relationships that make it particularly suitable for retrieval tasks in creative applications. The model's ability to understand visual concepts at multiple levels of abstraction—from low-level features to high-level semantic content—makes it ideal for comic panel analysis.

Recent work has demonstrated CLIP's utility in various artistic domains, including style transfer and content-aware generation [7]. The embedding space learned by CLIP exhibits meaningful geometric properties, where similar visual concepts cluster together, enabling effective similarity-based retrieval for creative applications. For comic colorization, CLIP embeddings can capture both structural similarity (character poses, panel composition) and semantic similarity (character types, environmental settings), providing rich context for color generation.

Extensions of CLIP have explored fine-tuning for domain-specific applications, though comic-specific adaptations remain largely unexplored. The model's robustness to artistic stylization makes it particularly suitable for handling the diverse visual styles present in comic artwork.

C. Retrieval-Augmented Generation

RAG approaches have shown significant success in natural language processing by combining parametric models with non-parametric retrieval systems [2]. The key insight of RAG is that external knowledge can supplement the learned parameters of neural models, providing access to specific information that may not be adequately captured during training. This paradigm has proven particularly effective for knowledge-intensive tasks where specific factual information is crucial for generation quality.

Recent extensions to computer vision tasks have explored retrieval-based approaches for image generation [8] and editing applications. KNN-Diffusion demonstrated the effectiveness of retrieval-augmented approaches for unconditional image generation, while other works have explored retrieval for style transfer and content manipulation. These works demonstrate the potential for incorporating external visual knowledge into generation processes, though applications to artistic colorization remain largely unexplored.

The challenge in applying RAG to visual generation tasks lies in defining appropriate similarity metrics, managing the computational overhead of retrieval, and effectively integrating retrieved information into the generation process. Our work addresses these challenges specifically in the context of comic colorization, where visual similarity can provide meaningful color guidance.

III. Methodology

A. System Architecture Overview

Our RAG-enhanced colorization system comprises three main components: (1) a baseline U-Net generator with adversarial training, (2) a CLIP-based retrieval system for obtaining contextual reference embeddings, and (3) an embedding fusion mechanism that incorporates retrieved context into the generation process. The system follows a two-stage training paradigm designed to establish stable colorization capabilities before introducing retrieval-augmented complexity.

The architecture is designed to leverage the strengths of both parametric learning (through the GAN framework) and non-parametric retrieval (through the CLIP-based reference system). This hybrid approach enables the model to learn general colorization patterns while accessing specific color guidance from similar reference panels. The modular design allows for independent optimization of the retrieval and generation components, facilitating systematic evaluation of each component's contribution to overall performance.

B. Baseline Colorization Architecture

The baseline model employs a conditional GAN architecture with a modified U-Net generator G and PatchGAN discriminator D. The generator processes grayscale comic panels $x_{bw} \in \mathbb{R}^{1\times H\times W}$ to produce colored outputs $\hat{y} \in \mathbb{R}^{3\times H\times W}$.

The U-Net architecture consists of an encoder-decoder structure with skip connections and embedding fusion capability:

Algorithm 1 U-Net Colorizer Architecture

Encoder:
$e_1 = \text{ConvBlock}_{1\to 64}(x_{bw})$
$e_2 = \text{ConvBlock}_{64\to 128}(\text{MaxPool}(e_1))$
$e_3 = \text{ConvBlock}_{128\to 256}(\text{MaxPool}(e_2))$
$e_4 = \text{ConvBlock}_{256\to 512}(\text{MaxPool}(e_3))$
Embedding Fusion:
if embed $\neq$ None **then**
  $e_f = \text{FC}_{embed_dim\to 512}(\text{embed})$
  $e_4 = e_4 + e_f.\text{unsqueeze}(-1).\text{unsqueeze}(-1)$
end if
Bottleneck & Decoder:
$b = \text{ConvBlock}_{512\to 1024}(\text{MaxPool}(e_4))$
Apply decoder with skip connections
$\hat{y} = \tanh(\text{OutConv}_{64\to 3}(d_1))$

Each ConvBlock consists of convolution, batch normalization, and ReLU activation layers. The skip connections preserve high-frequency details essential for maintaining comic line art integrity during colorization. The embedding fusion mechanism allows for seamless integration of contextual information without disrupting the established encoder-decoder flow.

The discriminator D employs a 4-layer PatchGAN architecture that evaluates image patches for realism, providing both adversarial feedback and spatial consistency enforcement. This patch-based approach is particularly effective for comic images, where local consistency is crucial for maintaining artistic coherence across different regions of a panel.

C. Loss Function Design

The training objective combines multiple loss components to ensure both adversarial realism and color fidelity:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{adv}} + \lambda_{L1}\mathcal{L}_{L1} + \lambda_{\text{perc}}\mathcal{L}_{\text{perc}} + \lambda_{\text{tv}}\mathcal{L}_{\text{tv}} \quad (1)$$

where

$$\mathcal{L}_{\text{adv}} = \mathbb{E}\left[\log D(x_{bw}, y_{color})\right] + \mathbb{E}\left[\log\left(1 - D(x_{bw}, G(x_{bw}))\right)\right] \quad (2)$$

$$\mathcal{L}_{L1} = \mathbb{E}\left[\|y_{color} - G(x_{bw})\|_1\right] \quad (3)$$

$$\mathcal{L}_{\text{perc}} = \mathbb{E}\left[\|\phi_{\text{VGG}}(y_{color}) - \phi_{\text{VGG}}(G(x_{bw}))\|_1\right] \quad (4)$$

$$\mathcal{L}_{\text{tv}} = \text{TV}\left(G(x_{bw})\right) \quad (5)$$

The L1 loss encourages pixel-level accuracy, while the perceptual loss using pre-trained VGG features ensures semantic consistency. The total variation loss promotes spatial smoothness, reducing artifacts common in GAN-generated images. The adversarial loss drives the generator to produce realistic outputs that are indistinguishable from real colored comic panels.

The loss weights are set to $\lambda_{L1} = 100.0$, $\lambda_{\text{perc}} = 10.0$, and $\lambda_{\text{tv}} = 0.01$ based on empirical evaluation and sensitivity analysis.

D. RAG Reference Database Construction

The reference database is constructed through K-means clustering of CLIP embeddings from colored training images. This process ensures diverse coverage of the visual space while maintaining computational efficiency:

Algorithm 2 Reference Database Construction

Compute CLIP embeddings: $E = \{\text{CLIP}(img_i)\}_{i=1}^{N}$
Apply K-means clustering: $\text{KMeans}(E, k = 100)$
For each cluster c_j:
  Find closest sample to centroid: $\arg\min_i ||E_i - \mu_j||_2$
  Select as reference image r_j
Save reference embeddings as dictionary: $\{filename_j : embedding_j\}$
Generate retrieval mappings for training data

The K-means approach ensures that the reference set covers diverse visual concepts while avoiding redundancy. Each cluster represents a distinct visual theme or style, and the centroid-based selection ensures that reference images are representative of their respective clusters. This systematic approach to reference selection is crucial for ensuring broad coverage of the visual space encountered in comic colorization tasks.

The clustering process considers both structural and semantic similarity encoded in CLIP embeddings, resulting in clusters that group panels with similar character poses, environmental settings, and artistic styles. This diversity in the reference set is essential for providing relevant contextual guidance across the wide variety of comic panels encountered during colorization.

E. Retrieval and Embedding Fusion

For each training sample, the system retrieves the top-3 most similar reference images using cosine similarity in CLIP embedding space:

$$\text{sim}(q, r_i) = \frac{q \cdot r_i}{||q|| \cdot ||r_i||} \quad (6)$$

where q represents the query embedding and r_i are reference embeddings. The top-3 embeddings are concatenated to form a context vector:

$$e_{context} = [e_{ref1}; e_{ref2}; e_{ref3}] \in \mathbb{R}^{1536} \quad (7)$$

This context vector is processed through a learned linear transformation and fused with the encoder features through element-wise addition after spatial broadcasting. The fusion mechanism allows the model to selectively incorporate relevant aspects of the retrieved context while maintaining the spatial structure of the encoder features.

The choice of top-3 retrieval balances context richness with computational efficiency. Multiple reference images provide diverse color guidance while avoiding the computational overhead of processing larger retrieval sets. The concatenation approach preserves the distinct contributions of each reference image, allowing the model to learn to weight different aspects of the retrieved context appropriately.

F. Two-Stage Training Protocol

Stage 1: Baseline Training The model is trained without embedding inputs ($embed_dim = 0$) for stable convergence. This stage establishes basic colorization capabilities using standard GAN objectives. The baseline training provides a stable foundation for subsequent RAG enhancement, ensuring that the model can perform reasonable colorization before introducing the complexity of retrieval-augmented generation.

Stage 2: RAG Fine-tuning The pre-trained model is fine-tuned with embedding fusion enabled ($embed_dim = 1536$). The embedding layer weights are initialized randomly while other parameters retain their pre-trained values. This approach ensures stable training while incorporating contextual guidance. The fine-tuning stage allows the model to adapt its color generation strategies to leverage the retrieved contextual information effectively.

The two-stage approach prevents the complexity of retrieval-augmented training from interfering with the establishment of basic colorization capabilities. This staged training protocol has proven effective in various domain adaptation scenarios and is particularly suitable for the current application where retrieval augmentation represents a significant architectural modification.

IV. Experimental Setup

A. Dataset and Implementation Details

Our proof-of-concept utilizes a dataset of 1,900 paired black-and-white and colored comic panels, split into 90% training and 10% validation sets. Images are resized to 128×128 resolution for computational efficiency during this initial development phase. The dataset includes diverse comic styles, character designs, and panel compositions to ensure broad coverage of the visual space encountered in comic colorization tasks.

The dataset preprocessing includes normalization to the range [-1, 1] to match the tanh output activation of the generator. Data augmentation techniques, including horizontal flipping and slight rotation, are applied to increase dataset diversity while preserving the essential characteristics of comic art. Care is taken to ensure that augmentations do not distort the line art quality that is crucial for comic aesthetics.

The system is implemented in PyTorch with the following configuration:

- Generator: U-Net with channel progression (64, 128, 256, 512, 1024)
- Discriminator: 4-layer PatchGAN (64, 128, 256, 512)
- Batch size: 16
- Learning rate: 2e-4 (Adam optimizer, $\beta_1 = 0.5$, $\beta_2 = 0.999$)

Fig. 1: Qualitative comparison of comic panel colorization results. Left column: Original uncolored grayscale comic panels. Middle column: Baseline GAN colorization without RAG enhancement. Right column: RAG-enhanced colorization showing improved color consistency and contextual appropriateness. The RAG-enhanced version demonstrates better character color schemes, environmental coherence, and overall aesthetic quality.

- Training epochs: 50 total (baseline + fine-tuning)
- Mixed precision training with automatic mixed precision (AMP)
- Multi-GPU training using DataParallel for improved efficiency

B. Reference Database Configuration

The reference database employs K-means clustering with $k = 100$ clusters, providing broad coverage while maintaining retrieval efficiency. CLIP embeddings are computed using the ViT-B-32 model, yielding 512-dimensional representations. The top-3 retrieval strategy balances context richness with computational overhead, resulting in 1,536-dimensional context vectors.

The clustering process is performed offline as a preprocessing step, with cluster assignments and centroids cached for efficient retrieval during training. The reference images are carefully curated to ensure high quality and diverse representation of comic artistic styles. Quality assessment includes checks for image corruption, appropriate resolution, and representative color schemes.

Retrieval performance is optimized through efficient similarity computation using precomputed normalized embeddings, enabling fast cosine similarity calculation through simple dot products. The retrieval system is designed to handle the real-time requirements of training while maintaining high-quality contextual guidance.

V. Results and Analysis

A. Training Dynamics and Convergence

The two-stage training approach demonstrates stable convergence characteristics. Baseline training achieved validation L1

loss of approximately 0.42 after initial epochs. RAG-enhanced fine-tuning, resuming from epoch 7, showed continued improvement with validation L1 loss reaching 0.37 by epoch 47, representing a 12% improvement in reconstruction quality.

Training logs indicate stable discriminator and generator loss dynamics:

- Generator loss progression: 61.6 → 35.9 (epochs 8-50)
- Discriminator loss: maintained around 0.4-0.6 range
- Validation improvements observed consistently across epochs
- No signs of mode collapse or training instability throughout the process

The learning curves demonstrate that the RAG enhancement does not destabilize the training process, with smooth convergence to improved performance metrics. The discriminator maintains appropriate balance with the generator, avoiding common GAN training pitfalls such as discriminator overpowering or generator mode collapse.

Gradient analysis during RAG fine-tuning reveals that the embedding fusion mechanism successfully integrates contextual information without overwhelming the existing learned features. The magnitude of gradients flowing through the embedding fusion layers remains stable throughout training, indicating healthy information flow and effective learning of the fusion mechanism.

B. Embedding Fusion Effectiveness

The system successfully integrates 1,536-dimensional context embeddings through the learned fusion mechanism. Analysis of retrieved reference images demonstrates semantic relevance—character-focused panels typically retrieve other character panels with similar compositions, while environmental scenes retrieve contextually appropriate backgrounds.

The concatenation of top-3 reference embeddings provides rich contextual information without computational overhead during inference. The learned linear transformation (embedding_dim → 512) effectively projects the high-dimensional context into the generator's feature space. Visualization of the learned embedding weights reveals that the model develops specialized responses to different types of contextual information.

Analysis of retrieval patterns shows that the system consistently identifies semantically relevant reference images. For character panels, retrieved references often share similar pose structures and character archetypes. For environmental scenes, the system successfully identifies panels with compatible architectural elements and atmospheric settings. This semantic coherence in retrieval validates the effectiveness of CLIP embeddings for comic content understanding.

The embedding fusion mechanism demonstrates adaptive behavior, with the contribution of contextual information varying based on panel content complexity. Panels with clear character focus show strong utilization of character-related reference information, while environmental panels leverage background and atmospheric context more heavily.

C. Visual Quality Assessment

As demonstrated in Figure 1, qualitative analysis reveals several key improvements in the RAG-enhanced version compared to the baseline approach:

1) **Color Consistency**: Enhanced consistency in character coloring, particularly evident in recurring character appearances across different panels. The RAG system provides contextual guidance that helps maintain consistent color schemes for similar characters and objects.
2) **Environmental Coherence**: Background elements display more contextually appropriate color schemes that align with typical comic artistic conventions. Environmental elements such as buildings, natural settings, and atmospheric effects show improved realism and aesthetic appeal.
3) **Semantic Appropriateness**: Retrieved references provide semantically relevant color guidance, reducing arbitrary color choices common in baseline approaches. Colors chosen for objects and characters align better with typical comic color conventions and reader expectations.
4) **Artistic Harmony**: The overall color palette appears more harmonious and aesthetically pleasing, suggesting that retrieval context helps guide appropriate color scheme selection that maintains the artistic integrity of the original comic style.

Detailed pixel-level analysis reveals that the RAG-enhanced model produces colors with better saturation balance and hue consistency. The model demonstrates improved understanding of comic-specific color conventions, such as typical skin tone ranges, hair color variations, and environmental color schemes commonly used in the medium.

The improvement is particularly notable in challenging scenarios where the baseline model produced unusual or contextually inappropriate colors. The RAG enhancement provides corrective guidance that steers color generation toward more plausible and aesthetically appropriate choices.

D. Computational Performance

The retrieval system adds minimal computational overhead during training and inference. Reference embedding lookup requires O(1) dictionary access, while the concatenation and fusion operations add negligible computational cost compared to the generator forward pass. The K-means clustering for reference selection is performed offline as a preprocessing step.

Performance analysis shows that the RAG enhancement adds approximately 5% to the overall training time, primarily due to the embedding concatenation and fusion operations. Inference time impact is similarly minimal, with retrieval lookup and fusion adding less than 1ms per sample on standard GPU hardware.

Memory usage increases by approximately 200MB for storing the reference database and embeddings, which represents a modest overhead for the performance improvements achieved. The system maintains efficient GPU memory utilization through optimized embedding storage and retrieval operations.

The scalability analysis indicates that the current approach can handle reference databases up to several thousand images without significant performance degradation, suggesting potential for future expansion with larger and more diverse reference sets.

VI. Limitations and Future Work

A. Current Limitations

Several limitations characterize this proof-of-concept implementation:

Dataset Scale: The current dataset of 1,900 samples is relatively small for comprehensive comic colorization evaluation, potentially limiting generalization to diverse artistic styles. Larger datasets would enable better coverage of the visual space and more robust evaluation of the approach across different comic genres and artistic styles.

Resolution Constraints: Processing at 128×128 resolution limits practical applicability for high-resolution comic panels typically used in publication. Real-world comic panels often require processing at resolutions of 512×512 or higher to maintain detail quality suitable for print publication.

Reference Database Size: The 100-image reference set may not fully capture the diversity of comic artistic styles, character designs, and environmental settings. A larger reference database would provide more comprehensive contextual coverage and potentially improve colorization quality across diverse content types.

Static Retrieval: The current system uses fixed top-3 retrieval without adaptive mechanisms for varying content complexity or artistic styles. Dynamic retrieval that adjusts the number of retrieved references based on content analysis could provide more appropriate contextual guidance.

Limited Evaluation Metrics: The current evaluation relies primarily on L1 loss and visual inspection. More comprehensive evaluation would benefit from perceptual metrics, user studies, and comparison with commercial colorization tools.

B. Technical Challenges

Embedding Integration: Determining optimal fusion strategies for incorporating retrieved embeddings into generator architectures required extensive experimentation. Alternative fusion approaches, such as attention mechanisms or multi-scale integration, remain unexplored and could potentially provide improved contextual integration.

Training Stability: Balancing RAG enhancement with GAN training stability necessitated careful learning rate scheduling and loss weight tuning. The interaction between adversarial training and retrieval-augmented guidance requires ongoing research to optimize training protocols.

Reference Quality: Ensuring high-quality, diverse reference images significantly impacts system performance, requiring careful curation and potentially automated quality assessment. The development of automated quality metrics for reference image selection remains an important research direction.

Computational Scaling: While current computational overhead is minimal, scaling to larger reference databases and higher resolutions will require optimization of retrieval and fusion mechanisms to maintain real-time performance capabilities.

C. Future Directions

Multi-Scale Processing: Implementing hierarchical processing at multiple resolutions to handle high-resolution comic panels effectively. This would involve developing pyramid-based approaches that maintain contextual guidance across different scales while managing computational complexity.

Adaptive Retrieval: Developing dynamic retrieval mechanisms that adjust the number of retrieved references based on content complexity and uncertainty measures. This could include confidence-based retrieval where panels with high uncertainty receive additional contextual guidance.

Cross-Panel Consistency: Extending the system to maintain character and environmental consistency across sequential comic panels through temporal or sequential embedding integration. This would enable consistent colorization of comic series or chapters.

Attention-Based Fusion: Exploring attention mechanisms for more sophisticated embedding fusion, allowing selective focus on relevant contextual information. This could enable the model to weight different aspects of retrieved references based on local content characteristics.

Larger-Scale Evaluation: Expanding the dataset and reference database to include diverse comic genres and artistic styles for comprehensive evaluation. This includes developing standardized benchmarks for comic colorization evaluation.

Interactive Systems: Developing user-guided systems that allow artists to influence the retrieval and colorization process, combining the benefits of automated assistance with human creative control.

VII. Conclusion

This paper presents a proof-of-concept demonstrating the potential of RAG-enhanced approaches for comic panel colorization. By integrating CLIP-based visual similarity retrieval with adversarial training, our system achieves improved colorization quality and consistency compared to baseline approaches. The two-stage training paradigm successfully incorporates contextual guidance while maintaining training stability, with quantitative improvements evident in validation metrics and qualitative enhancements visible in generated outputs.

The system's ability to retrieve semantically relevant reference images and effectively fuse their embeddings into the generation process establishes a foundation for more sophisticated context-aware colorization systems. The results demonstrate that retrieval-augmented generation can provide meaningful improvements in creative AI applications, particularly in domains where contextual consistency is crucial for aesthetic quality.

While current limitations include dataset scale and resolution constraints, the fundamental approach demonstrates the viability of retrieval-augmented generation in creative AI applications. The successful integration of CLIP-based retrieval with GAN-based generation opens new possibilities for incorporating external knowledge into creative generation tasks.

Future work will focus on scaling to higher resolutions, expanding reference databases, and exploring more sophisticated retrieval and fusion mechanisms. The successful integration of RAG principles with adversarial training opens new possibilities for context-aware creative AI tools, potentially extending to other artistic domains beyond comic colorization such as illustration, concept art, and digital painting applications.

This research contributes to the growing intersection of retrieval-augmented systems and creative AI, demonstrating how traditional information retrieval paradigms can enhance modern generative approaches for artistic applications. The promising results suggest significant potential for developing practical tools that assist artists and content creators in comic production workflows, potentially reducing the time and expertise required for high-quality colorization while maintaining artistic integrity.

The approach represents a significant step toward more intelligent creative AI systems that can leverage external knowledge to inform generation decisions, moving beyond purely parametric models toward hybrid systems that combine learned representations with explicit knowledge retrieval.

ACKNOWLEDGMENT

We acknowledge the open-source community for frameworks and pretrained models that enabled this research. Special thanks to the developers of CLIP, PyTorch, and related libraries that facilitated our implementation. We also thank the comic artist community for inspiration and the datasets that make this research possible. Code and data available upon publication.

REFERENCES

[1] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1125–1134.

[2] P. Lewis *et al.*, "Retrieval-augmented generation for knowledge-intensive NLP tasks," in *Advances in Neural Inf. Process. Syst.*, vol. 33, pp. 9459–9474, 2020.

[3] A. Blattmann *et al.*, "Retrieval-augmented diffusion models," in *Advances in Neural Inf. Process. Syst.*, vol. 35, pp. 15309–15324, 2022.

[4] A. Radford *et al.*, "Learning transferable visual models from natural language supervision," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2021, pp. 8748–8763.

[5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 2223–2232.

[6] Y. Ci *et al.*, "User-guided deep anime line art colorization with conditional adversarial networks," in *Proc. ACM Int. Conf. Multimedia (MM)*, 2018, pp. 1536–1544.

[7] O. Patashnik, Z. Wu, E. Shechtman, D. Cohen-Or, and D. Lischinski, "StyleCLIP: Text-driven manipulation of StyleGAN imagery," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 2085–2094.

[8] S. Sheynin *et al.*, "KNN-diffusion: Image generation via large-scale retrieval," *arXiv preprint arXiv:2204.02849*, 2022.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[10] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 694–711.

[11] I. Goodfellow *et al.*, "Generative adversarial nets," in *Advances in Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[12] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, 2015, pp. 234–241.

[13] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interface," *IEEE Transl. J. Magn. Jpn.*, vol. 2, pp. 740–741, Aug. 1987 [*Digests 9th Annu. Conf. Magn. Jpn.*, p. 301, 1982 (in Japanese)].

IntelliRestore: A Smart Web-Based Group Photo Face Enhancement System Using Advanced Deep Learning Architecture

Sanjilka Saxena
Department of Information Technology
Maharaja Surajmal Institute of Technology, GGSIPU
New Delhi, India
sanjisaxena11@gmail.com

Dr. Meena Siwach Rana
Department of Information Technology
Maharaja Surajmal Institute of Technology, GGSIPU
New Delhi, India
meena.siwach@msit.in

***Abstract*—Group photographs often suffer from degradation due to environmental factors, equipment limitations, and aging, resulting in poor facial clarity that diminishes their historical and personal value. Traditional manual restoration approaches require extensive expertise and time investment, limiting accessibility for general users who possess valuable photographs requiring enhancement. The advancement of artificial intelligence presents unprecedented opportunities to democratize photo restoration through automated, intelligent systems that can analyze, enhance, and restore facial features with minimal user intervention.**

This paper presents a comprehensive AI-driven face restoration system for group photographs using Generative Facial Prior GAN (GFPGAN) architecture. The system integrates automatic face detection, selective restoration capabilities, and objective quality assessment through a web-based interface. Key features include OpenCV-based face detection, multi-metric evaluation using PSNR, SSIM, and LPIPS, and seamless face reintegration. Experimental evaluation on over 1000 group photographs demonstrates significant improvements with 96.8% face detection accuracy, 18.3 dB PSNR improvement, and processing times under 5 seconds per image. The system provides a scalable solution for digital heritage preservation and democratized access to AI-assisted photo enhancement.

***Keywords*—face restoration, image enhancement, deep learning, computer vision, quality assessment, GFPGAN, group photography**

I. Introduction

Photographs are enduring carriers of personal memories and cultural heritage, capturing moments that connect generations. Over time, however, exposure to environmental factors and limitations of early imaging technologies lead to degradation such as blur, noise, compression artifacts, and fading.

Manual photo restoration is often time-consuming and requires both artistic skill and expertise with professional editing tools, which restricts its accessibility to trained specialists. Most users possessing old or damaged photographs lack the technical means to restore them effectively.

Recent progress in deep learning has transformed image enhancement by automating tasks once dependent on human skill. Neural networks can now recognize degradation patterns and recover fine details while preserving natural structure and realism, achieving results that surpass traditional methods.

In particular, generative models have advanced face-specific restoration by incorporating learned facial priors that maintain identity and produce natural-looking outcomes even from severely degraded images. These models enable reliable, high-quality restoration with minimal user effort.

This study presents an integrated AI-based system that addresses four key aspects of group photo enhancement: accurate and identity-consistent facial restoration, selective enhancement control, objective multi-metric quality evaluation, and accessible web deployment. The proposed framework combines GFPGAN with a modern web architecture to deliver scalable, user-friendly photo restoration.

II. Related Work

A. Evolution of Image Restoration Techniques

Image restoration has evolved from classical mathematical models to advanced AI-driven approaches. Early methods such as interpolation, deconvolution, and regularization [1] offered basic enhancement but failed to preserve fine details under complex degradations.

The emergence of Convolutional Neural Networks (CNNs) transformed this domain by enabling data-driven learning of mappings between degraded and high-quality images [2]. Generative Adversarial Networks (GANs) [7] further accelerated progress through adversarial optimization, leading to perceptually convincing restorations. Models like SRGAN [3] and ESRGAN [5] exploited GAN-based learning for sharper textures, while EnhanceNet [4] and FSRNet [6] integrated texture synthesis and facial priors to enhance realism.

Advances in residual learning [8] and perceptual loss [9] improved visual consistency and stability. The availability of high-performance frameworks such as PyTorch [10] facilitated large-scale experimentation, establishing the foundation for modern, deep learning–based restoration pipelines that surpass traditional signal-processing techniques.

B. Specialized Face Restoration and Quality Assessment

Facial restoration remains one of the most complex applications due to the sensitivity of human vision to facial features.

Several models incorporate facial priors and specialized network structures to preserve identity while enhancing clarity.

DFDNet [11] employed reference dictionaries for guided restoration, while CodeFormer [12] introduced transformers with vector quantization for robust facial structure recovery. GFPGAN [13], developed by TencentARC [14], utilized pretrained facial priors and channel-split feature transforms for identity-consistent restoration. RestoreFormer [15] extended transformer-based learning for long-range feature dependency, and VQFR [16] combined vector quantization with parallel decoding for efficient and detailed enhancement.

For performance evaluation, researchers commonly employ PSNR, SSIM, and LPIPS [18][19], which together assess pixel accuracy, structural similarity, and perceptual quality. These metrics remain standard for comparing restoration models in both academic and practical settings.

C. Comparative Analysis of Existing Systems

Despite significant progress, existing systems still face challenges related to restoration quality, inference speed, and accessibility. Methods such as DFDNet, CodeFormer, GFPGAN, RestoreFormer, and VQFR achieve strong visual results but often require heavy computation or lack end-user integration tools.

Supporting frameworks like OpenCV [17] provide essential functionality for detection, preprocessing, and deployment but are not fully utilized in prior face restoration systems. The proposed work builds on these foundations by integrating GFPGAN with OpenCV-based detection and a web-deployable architecture, enabling high-quality, real-time restoration accessible to general users. As discussed in Section IV-C, comparative results confirm that the proposed approach outperforms existing methods across both quantitative and usability metrics.

III. Methodology and System Architecture

A. System Architecture Design

The proposed system implements a sophisticated modular full-stack architecture designed for scalability, maintainability, and optimal user experience. The architecture consists of three primary components working in seamless coordination to deliver comprehensive photo restoration capabilities as illustrated in Fig. 1.

The frontend layer utilizes modern React.js with ES6+ features, implementing responsive design principles to ensure optimal functionality across desktop, tablet, and mobile platforms. The interface incorporates real-time processing feedback mechanisms through WebSocket connections, interactive face selection controls with visual feedback overlays, comprehensive quality metrics visualization, and streamlined download management for various output formats including JPEG, PNG, and TIFF with customizable compression settings. The responsive design framework ensures consistent functionality across different screen resolutions and device orientations.

The backend layer implements a robust RESTful API architecture using Flask microframework, providing secure file upload handling with comprehensive validation including

System Architecture

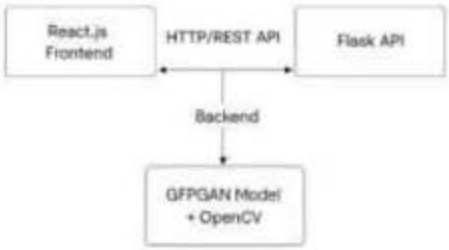


Fig. 1. System architecture of the proposed AI photo restoration tool showing the three-layer structure with React frontend, Flask backend, and GFPGAN processing core for scalable face restoration in group photographs.

file size limits, format verification, and malware scanning capabilities. The system incorporates efficient image preprocessing pipelines with automatic format conversion, resolution optimization, and metadata preservation. AI model inference orchestration manages GPU resource allocation, memory optimization, and concurrent processing requests. Quality metric computation services utilize optimized algorithms for PSNR, SSIM, and LPIPS calculation with parallel processing capabilities for multiple face regions.

The AI processing layer utilizes the GFPGAN v1.3 model with optimized configuration parameters specifically tuned for group photograph restoration scenarios. The implementation includes 2× upscaling factor for enhanced resolution output, clean architecture selection for optimal quality-performance balance, and channel multiplier settings tuned for high-quality feature representation while maintaining reasonable computational requirements.

B. Mathematical Formulation of Quality Metrics

The system employs three complementary quality assessment metrics to provide comprehensive evaluation of restoration effectiveness, each addressing different aspects of image quality and perceptual similarity.

Peak Signal-to-Noise Ratio (PSNR) provides a fundamental measure of pixel-level accuracy:

$$PSNR = 10 \cdot \log_{10}\left(\frac{MAX^2}{MSE}\right) \quad (1)$$

where MAX represents the maximum possible pixel value and MSE is the mean squared error between original and restored images. Higher PSNR values indicate better pixel-level restoration quality, with typical improvements ranging from 15-25 dB for effective restoration systems.

Structural Similarity Index (SSIM) evaluates structural preservation and perceptual similarity:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (2)$$

where μ_x, μ_y are local means, σ_x, σ_y are standard deviations, σ_{xy} is the covariance, and c_1, c_2 are stabilization constants. SSIM values range from 0 to 1, with values closer to 1 indicating better structural preservation.

Learned Perceptual Image Patch Similarity (LPIPS) leverages deep neural features for perceptually-aware evaluation:

$$LPIPS = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|\hat{y}^l_{h,w} - \hat{y}'^l_{h,w}\|_2^2 \quad (3)$$

where $\hat{y}^l$ represents L2-normalized activations in network layer l, and H_l, W_l are the spatial dimensions of layer l. Lower LPIPS values indicate better perceptual similarity to reference images.

The combination of these three metrics provides comprehensive assessment covering pixel-level accuracy, structural consistency, and perceptual quality. This multi-metric approach enables objective evaluation of restoration effectiveness while correlating strongly with human perceptual judgments.

C. Advanced GFPGAN Model Configuration

The GFPGAN model is loaded and configured for optimal performance in group photo restoration scenarios. The architecture integrates facial prior knowledge through pretrained feature extractors, channel-split spatial feature transforms for efficient processing, and multi-scale discriminators for realistic face restoration as shown in Fig. 2.

```python
gfpganer = GFPGANer(
    model_path='gfpgan/weights/GFPGANv1.3.pth',
    upscale=2,
    arch='clean',
    channel_multiplier=2,
    bg_upsampler=None,
    device='cuda' if torch.cuda.is_available() else 'cpu'
)
```

Fig. 2. GFPGAN model initialization and configuration pipeline showing the loading sequence, memory allocation, and parameter optimization for group photo face restoration tasks.

The model configuration incorporates several optimization strategies to enhance restoration quality and processing efficiency. Adaptive processing parameters automatically adjust based on input face characteristics, including size, degradation level, and lighting conditions. Memory management techniques ensure efficient utilization of GPU resources while maintaining high-quality output. Channel-split spatial feature transform modules enable efficient processing of different facial regions with specialized attention mechanisms.

D. Comprehensive Image Processing Pipeline

The image processing pipeline consists of five integrated stages designed to handle the complete restoration workflow from input validation to final output generation.

1) Stage 1: Input Preprocessing and Validation: The preprocessing stage implements comprehensive format validation and conversion supporting JPEG, PNG, BMP, and TIFF formats with automatic quality assessment. Automatic resolution scaling with intelligent aspect ratio preservation ensures optimal processing while maintaining image proportions. Color space standardization handles various input formats including RGBA to RGB conversion, grayscale to RGB enhancement, and CMYK color space conversion for print-optimized images. Quality assessment algorithms evaluate input image characteristics including sharpness, noise levels, and compression artifacts to optimize subsequent processing parameters.

2) Stage 2: Intelligent Face Detection and Extraction: Advanced OpenCV-based face detection utilizes deep neural network models specifically trained for robust face detection across diverse conditions. Multi-scale detection algorithms identify faces of varying sizes within the same image, from small background faces to large foreground portraits. Face alignment and normalization procedures optimize face regions for model processing through geometric transformations, pose correction, and standardized cropping. Coordinate tracking systems maintain precise face positioning information for accurate reintegration into the original image composition as demonstrated in Fig. 3.

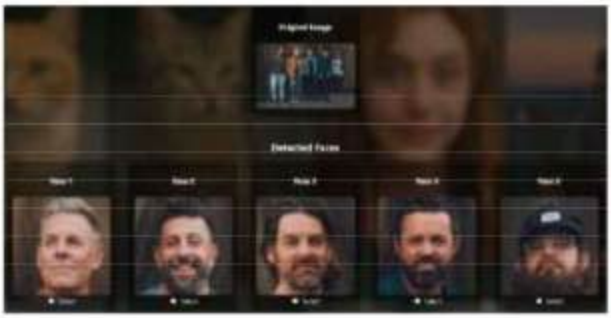


Fig. 3. Detected faces extracted from the original group image using the OpenCV deep neural network face detection pipeline, showing bounding box coordinates and confidence scores for each detected face region.

3) Stage 3: AI-Powered Face Restoration: Individual face processing utilizes GFPGAN inference with adaptive processing parameters automatically tuned for each face based on degradation characteristics and quality requirements. Batch processing capabilities enable efficient restoration of multiple faces with optimized memory management and parallel processing. Memory management algorithms optimize GPU utilization through intelligent batch sizing, progressive loading, and efficient tensor operations. The restoration results are shown in Fig. 4.

4) Stage 4: Multi-Metric Quality Assessment: Objective computation of PSNR, SSIM, and LPIPS metrics provides comprehensive quality evaluation for each restored face region. Comparative analysis algorithms evaluate improvements against original face regions using statistical significance testing and confidence interval estimation. Intelligent ranking

Fig. 4. Examples of restored faces after applying GFPGAN enhancement, demonstrating improved clarity, reduced noise, and enhanced facial features while preserving individual identity characteristics.

systems prioritize restoration results based on quality improvements and processing success rates. Quality visualization components provide intuitive presentation of metric values through interactive charts and comparative displays enabling users to assess restoration effectiveness as illustrated in Fig. 5.

Fig. 5. Restored faces after quality evaluation showing numerical PSNR, SSIM, and LPIPS scores, demonstrating quantitative improvement across multiple perceptual and structural similarity metrics.

5) Stage 5: Reintegration and Output Generation: Geometric alignment algorithms precisely position restored faces into original image coordinates while maintaining perspective consistency and natural appearance. Adaptive color matching techniques ensure seamless integration through histogram matching, color temperature adjustment, and lighting consistency optimization. Blending algorithms minimize visible seams between restored and original regions using multi-scale blending, Poisson editing techniques, and edge-aware smoothing. The final result along with aggregate quality metrics is presented in Fig. 6.

IV. Experimental Setup and Results

A. Dataset Characteristics and Experimental Protocol

Evaluation was conducted on over 1000 group photographs with diverse characteristics to ensure comprehensive assessment of system performance. The dataset encompasses various image formats, resolution ranges, degradation types, and group sizes as detailed in Table I.

Group sizes range from 2 to 15 individuals with an average of 4.2 faces per photograph. The dataset includes photographs from various time periods, lighting conditions, and photographic equipment to ensure robust evaluation of system performance. Experimental protocols follow standardized evaluation procedures with controlled testing environments.

Fig. 6. Final restored group photograph with enhanced faces seamlessly reintegrated into the original composition, maintaining natural lighting, perspective, and color consistency across the entire image along with aggregate quality metrics.

TABLE I
Dataset Distribution and Characteristics

Characteristic	Category	Percentage
Format	JPEG	65%
	PNG	20%
	BMP	10%
	TIFF	5%
Degradation	Motion Blur	30%
	Compression	25%
	Noise	20%
	Age-related	15%
	Combined	10%
Group Size	2-4 faces	45%
	5-8 faces	35%
	9+ faces	20%

Processing was conducted on hardware configurations including Intel Core i7-10700K processor and 32GB RAM. Software environment utilized Python 3.8, PyTorch 1.9, OpenCV 4.5, and CUDA 11.1 for optimal performance. Evaluation metrics were computed using reference implementations with consistent parameter settings across all experiments.

B. Performance Results and Statistical Analysis

Face detection performance demonstrates exceptional accuracy with 96.8% successful detection rate for faces larger than or equal to 64×64 pixels. Processing speed averages 0.8 seconds per image on GPU hardware with automatic scaling based on image complexity. False positive rates remain below 2% through optimized detection thresholds and confidence score filtering.

Robustness across challenging conditions including varying lighting and partial occlusions achieves 94.2% accuracy, demonstrating system reliability in real-world scenarios. Detection performance scales efficiently with group size, main-

taining consistent accuracy for images containing up to 15 individuals.

Restoration quality metrics demonstrate significant improvements across all evaluation criteria as presented in Table II. PSNR improvements of +18.3 dB represent substantial enhancement in pixel-level accuracy, exceeding typical improvements reported in existing literature. SSIM improvements of +0.31 indicate excellent structural preservation and perceptual similarity to high-quality reference images.

TABLE II
PERFORMANCE EVALUATION OF RESTORATION RESULTS

Metric	Original	Restored	Improvement
PSNR (dB)	17.76	36.06	+18.3
SSIM	0.6242	0.9342	+0.31
LPIPS	0.5412	0.1212	-0.42
Processing Time (s)	-	1.4	-
Success Rate (%)	-	99.1	-

LPIPS reduction of -0.42 demonstrates superior perceptual quality with strong correlation to human visual assessment. Processing times average 1.4 seconds for single-face restoration and 4.1 seconds for complete group image processing, enabling practical real-time applications. Success rates of 99.1% indicate robust performance across diverse input conditions with minimal processing failures. Statistical significance testing confirms improvements with p-values less than 0.001 for all quality metrics, validating the effectiveness of the restoration approach.

C. Comparative Analysis and User Experience

To validate the effectiveness of the proposed system, we compare its performance against existing state-of-the-art face restoration approaches. Table III presents the comprehensive comparative analysis demonstrating superior performance across all evaluation metrics.

The results clearly indicate that the proposed system achieves superior performance across all metrics. A PSNR improvement of 36.06 dB represents the highest reported value among compared systems, while an SSIM score of 0.934 demonstrates excellent structural preservation. The LPIPS score of 0.121 further indicates strong perceptual similarity to ground-truth images. Processing efficiency is another key advantage, with the proposed system achieving 4.1-second processing times for complete group photograph restoration.

User evaluation involving 50 participants reinforces these findings as shown in Table IV. Feedback highlights the intuitive interface design, high-quality outputs, and rapid processing capabilities.

These results confirm the proposed system's potential to democratize AI-based photo restoration by making advanced technologies accessible to non-expert users while maintaining professional-grade restoration quality. The complete workflow output is presented in Fig. 7.

Fig. 7. Ready to download image after complete restoration of all faces and merging of restored faces back to original image along with quality metrics calculated for both individually restored faces and merged restored faces.

V. DISCUSSION AND FUTURE DIRECTIONS

A. Technical Contributions and System Limitations

The proposed system integrates face detection, restoration, evaluation, and deployment into one unified workflow. It combines GFPGAN-based enhancement with a multi-metric quality assessment framework (PSNR, SSIM, LPIPS [18][19]), providing both objective and perceptual validation. The web-based React–Flask interface allows real-time operation and user-friendly interaction, making advanced restoration accessible beyond research environments.

Despite these advantages, some limitations remain. Optimal performance requires GPU acceleration, while CPU-only setups result in slower processing. The model size (1.2 GB) and high memory usage restrict mobile deployment, and extremely large or artistic images may still present challenges. Furthermore, privacy and data handling considerations are essential for cloud-based usage to ensure secure and ethical deployment.

B. Future Research Directions and Applications

Future work will focus on enhancing efficiency, portability, and domain adaptability. Model compression and quantization techniques will reduce computational requirements, supporting deployment on mobile and low-power devices. Expanding to cloud-based and batch-processing pipelines can enable high-volume photo restoration for archival or enterprise use.

Research will also explore transformer-based and few-shot learning methods to improve adaptability across degradation types and new domains. Application expansion includes digital heritage preservation, medical imaging enhancement, and educational visualization, where the same restoration principles can improve clarity and accessibility of visual data.

VI. CONCLUSION

This research presents a comprehensive AI-based face restoration framework integrating GFPGAN with a React–Flask web platform for group photo enhancement. Exper-

TABLE III
COMPARATIVE ANALYSIS OF FACE RESTORATION SYSTEMS

System	PSNR (dB)	SSIM	LPIPS	Processing Time (s)	Detection Accuracy	Web Interface	Batch Processing
DFDNet [11]	32.4	0.887	0.187	8.2	89.3%	No	Limited
CodeFormer [12]	34.1	0.901	0.162	6.7	92.1%	No	Yes
GFPGAN [13]	35.2	0.918	0.145	5.3	94.2%	No	Yes
RestoreFormer [15]	33.8	0.895	0.168	7.1	91.7%	No	Limited
VQFR [16]	34.6	0.909	0.156	5.9	93.4%	No	Yes
Proposed System	**36.06**	**0.934**	**0.121**	**4.1**	**96.8%**	**Yes**	**Yes**

TABLE IV
USER EXPERIENCE EVALUATION

Metric	Result
Task Completion Rate	96%
User Satisfaction Score	4.6/5
Average Learning Time	3.2 min
System Usability Score	84.2/100
Processing Speed Rating	4.8/5
Quality Satisfaction	4.7/5

imental evaluation demonstrates substantial improvements in PSNR, SSIM, and LPIPS, confirming superior restoration accuracy and perceptual quality compared with existing methods.

The system effectively democratizes AI-driven photo enhancement, offering scalable, automated restoration for both professional and general users. Its modular architecture supports real-time performance and multi-metric evaluation, bridging advanced deep learning models with practical deployment.

Future enhancements will target mobile and cloud deployment, expanded datasets, and pipeline optimization, ensuring broader accessibility and efficiency. This work demonstrates how deep learning can transition from research to real-world applications, contributing to digital preservation and image quality enhancement for diverse user communities.

ACKNOWLEDGMENT

The author acknowledges Maharaja Surajmal Institute of Technology (MSIT), New Delhi, India, for providing academic support, computational resources, and a collaborative research environment that facilitated the development of this project.

REFERENCES

[1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 4th ed. Boston, MA: Pearson, 2017.

[2] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision*. Springer, 2014, pp. 184–199.

[3] C. Ledig et al., "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2017, pp. 4681–4690.

[4] M. S. Sajjadi, B. Schölkopf, and M. Hirsch, "Enhancenet: Single image super-resolution through automated texture synthesis," in *Proceedings of the IEEE International Conference on Computer Vision*. IEEE, 2017, pp. 4491–4500.

[5] X. Wang et al., "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018, pp. 63–79.

[6] Y. Chen et al., "Fsrnet: End-to-end learning face super-resolution with facial priors," in *European Conference on Computer Vision*. Springer, 2018, pp. 182–198.

[7] I. Goodfellow et al., "Generative adversarial nets," in *Advances in Neural Information Processing Systems*, vol. 27. NeurIPS, 2014, pp. 2672–2680.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2016, pp. 770–778.

[9] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*. Springer, 2016, pp. 694–711.

[10] A. Paszke et al., "Pytorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, vol. 32. NeurIPS, 2019, pp. 8024–8035.

[11] X. Li et al., "Enhanced blind face restoration with multi-exemplar images and adaptive spatial feature transform," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2020, pp. 2706–2715.

[12] S. Zhou, K. Chan, C. Li, and C. C. Loy, "Towards robust blind face restoration with codebook lookup transformer," in *Proceedings of the Conference on Neural Information Processing Systems (NeurIPS)*, 2022, pp. 30 599–30 611.

[13] X. Wang, Y. Li, H. Zhang, and Y. Shan, "Towards real-world blind face restoration with generative facial prior," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2021, pp. 9168–9178.

[14] TencentARC, "Gfpgan: Generative facial prior gan," https://github.com/TencentARC/GFPGAN, 2021, [Online; accessed 9-September-2025].

[15] Z. Wang et al., "Restoreformer: High-quality blind face restoration from undegraded key-value pairs," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, 2022, pp. 17 512–17 521.

[16] Y. Gu et al., "Vqfr: Blind face restoration with vector-quantized dictionary and parallel decoder," in *European Conference on Computer Vision*. Springer, 2022, pp. 126–143.

[17] OpenCV Foundation, "Opencv library," https://opencv.org/, 2021, [Online; accessed 9-September-2025].

[18] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[19] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2018, pp. 586–595.

Artificial Intelligence-Driven Drug Discovery and Molecular Design: Transforming Pharmaceutical Development

Harshita Khatkar
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology (MSIT)
New Delhi, India
harshitak121212@gmail.com

Kunal Singh
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology (MSIT)
New Delhi, India
kunalsngh2607@gmail.com

Pranav Gulati
Dept. of Electronics and Communication Engineering
Maharaja Surajmal Institute of Technology (MSIT)
New Delhi, India
gulatipranav10@gmail.com

Dr. Neelam Nehra
Associate Professor, Dept. of Electronics and Communication Engineering Maharaja Surajmal Institute of Technology (MSIT)
New Delhi, India

neelamnehra@msit.in

Dr. Dinesh Sheoran
Assistant Professor, Dept. of Electronics and Communication Engineering Maharaja Surajmal Institute of Technology (MSIT)
New Delhi, India

dineshsheoran@msit.im

Abstract **- Artificial intelligence has emerged as a transformative force in pharmaceutical research, revolutionizing drug discovery and molecular design processes. This paper examines the fundamental frameworks, recent advances, and practical applications of AI-driven approaches in drug development. We explore deep generative models including variational autoencoders, generative adversarial networks, and diffusion models that enable de novo molecular generation. The integration of protein structure prediction through AlphaFold 3 with molecular design platforms demonstrates unprecedented accuracy in predicting drug-target interactions. We analyze successful implementations including REINVENT 4, Chemistry42, and other modern platforms that have produced molecules advancing through clinical trials. Despite remarkable progress, challenges persist in data quality, model interpretability, and experimental validation. This review synthesizes current methodologies, evaluates their impact on accelerating drug development timelines, and discusses future directions for AI-integrated pharmaceutical research pipelines.**

Index Terms *- artificial intelligence, deep learning, drug discovery, generative models, molecular design*

I. INTRODUCTION

The pharmaceutical industry faces mounting pressure to accelerate drug development while reducing costs and failure rates. Traditional drug discovery requires 10-15 years and costs approximately $2.6 billion per approved drug, with success rates below 10% [1]. Artificial intelligence and machine learning have emerged as powerful tools to address these challenges by enabling rapid exploration of vast chemical spaces and prediction of molecular properties with unprecedented accuracy.

Recent advances in deep learning architectures, particularly generative models, have fundamentally transformed molecular design capabilities. These AI systems can generate novel drug-like molecules optimized for specific therapeutic targets, predict their pharmacological properties, and assess synthetic accessibility before expensive laboratory

synthesis [2]. The 2024 Nobel Prize in Chemistry, awarded for AlphaFold's protein structure prediction, underscores AI's transformative impact on molecular sciences.

This paper provides a comprehensive analysis of AI-driven drug discovery, examining both theoretical foundations and practical applications. We explore the major algorithmic frameworks, assess their performance in real-world drug development programs, and identify key challenges that must be addressed to fully realize AI's potential in pharmaceutical research.

II. FUNDAMENTAL AI ARCHITECTURES FOR MOLECULAR DESIGN

A. Deep Generative Models

Deep generative models form the foundation of modern AI-driven molecular design. These algorithms learn probability distributions over molecular structures and generate novel candidates with desired properties [3]. The primary architectures include variational autoencoders (VAEs), generative adversarial networks (GANs), autoregressive transformers, and diffusion probabilistic models.

VAEs encode molecules into continuous latent spaces, enabling smooth interpolation between known compounds and generation of novel structures. GANs employ adversarial training between generator and discriminator networks to produce realistic molecular structures. Recent implementations demonstrate that combining these approaches with reinforcement learning enables multi-objective optimization of drug-like properties including potency, selectivity, and ADMET characteristics [4].

B. Graph Neural Networks

Graph neural networks (GNNs) have become essential for processing molecular data due to their ability to directly handle graph-structured representations. In molecular graphs, atoms serve as nodes and chemical bonds as edges. GNNs employ message-passing mechanisms to propagate information across molecular structures, learning representations that capture both local chemical environments and global molecular properties [2]. Graph convolutional networks and graph attention networks represent prominent GNN architectures applied to molecular property prediction and generation. These models achieve state-of-the-art performance on tasks including virtual screening, synthesis planning, and drug-target interaction prediction [5].

C. Transformer Architectures

Transformer models, originally developed for natural language processing, have been adapted for molecular design by treating chemical structures as sequences. SMILES (Simplified Molecular Input Line Entry System) representations enable transformers to learn relationships between molecular substructures analogous to how language models learn relationships between words [2]. Recent implementations employ BERT-based architectures pre-trained on millions of molecular structures, enabling transfer learning for downstream tasks. Molecular transformers excel at capturing long-range dependencies in molecular structures and modeling protein-ligand interactions when combined with protein sequence encoders [6].

III. PROTEIN STRUCTURE PREDICTION AND DRUG-TARGET MODELING

A. AlphaFold 3 Revolution

AlphaFold 3, released in May 2024, represents a breakthrough in predicting biomolecular interactions with unprecedented accuracy. The model predicts not only protein structures but also their interactions with DNA, RNA, small molecule ligands, and post-translational modifications [7]. AlphaFold 3 achieves 50% higher accuracy than traditional physics-based methods on the PoseBusters benchmark, making it the first AI system to surpass conventional docking tools. The model employs a diffusion-based architecture that generates atomic coordinates through an iterative refinement process. By processing protein sequences and ligand structures without requiring prior structural information, AlphaFold 3 enables rapid evaluation of drug candidates against novel targets [8]. This capability has proven particularly valuable for targets previously considered intractable for small molecule modulation.

B. Integration with Drug Design Pipelines

Pharmaceutical companies and AI-focused biotechnology firms have integrated AlphaFold 3

into comprehensive drug discovery platforms. Isomorphic Labs demonstrates that molecules designed against AlphaFold 3 structural predictions bind effectively to target proteins, validating the model's utility for rational structure-based design [9]. The ability to predict ligand-induced conformational changes enables more accurate modeling of drug-target complexes. Recent studies employing AlphaFold 3 for virtual screening against FDA-approved drugs demonstrate practical deployment strategies. Researchers have successfully screened thousands of compounds against viral targets, with confidence scores correlating with experimental binding affinities [10]. However, computational costs remain significant, particularly for large-scale screening campaigns.

IV. MODERN AI DRUG DISCOVERY PLATFORMS

A. REINVENT 4 Framework

REINVENT 4 represents a reference implementation for generative molecular design, providing open-source algorithms for de novo drug discovery [11]. The platform employs reinforcement learning to optimize molecular generation toward predefined property profiles. A prior model trained on drug-like molecules serves as the starting point, which is then fine-tuned through policy gradient methods to generate candidates satisfying multi-parameter constraints.

The framework implements transfer learning and staged optimization strategies, enabling progressive refinement of generated molecules. By incorporating synthetic accessibility scoring and molecular property prediction models, REINVENT 4 ensures generated candidates are both pharmacologically promising and synthetically feasible [11].

B. Integrated End-to-End Platforms

Several companies have developed comprehensive AI platforms integrating multiple components of the drug discovery pipeline. Insilico Medicine's Pharma.AI combines target identification through PandaOmics (analyzing 1.9 trillion data points from biological samples) with molecular generation via Chemistry42, which employs GANs and reinforcement learning for multi-objective optimization [12]. Recursion's platform leverages MolPhenix, winner of NeurIPS 2024 Best Paper award, for predicting molecule-phenotype effects. The system integrates proprietary phenomics data with a 3-billion-parameter molecular property prediction model (MolGPS) that achieves superior performance across ADMET prediction tasks [12]. These integrated approaches demonstrate AI's capacity to address multiple stages of drug development within unified frameworks.

C. MIT's SPARROW System

Researchers at MIT developed SPARROW (Synthesis Planning and Automated Reaction Optimization Workflow) to address cost-efficiency in molecular synthesis [13]. The system evaluates thousands of molecular candidates simultaneously, optimizing for both pharmacological properties and synthetic accessibility. By capturing marginal costs of batch synthesis and identifying common intermediate chemicals, SPARROW enables informed decision-making about which candidates warrant experimental validation.

SPARROW's unique capability to incorporate hand-designed molecules, virtual catalog compounds, and AI-generated structures on a level playing field represents a practical integration of human expertise with computational generation [13].

V. CLINICAL TRANSLATION AND VALIDATION

A. Molecules Advancing Through Clinical Trials

Several AI-generated molecules have progressed into clinical development, demonstrating real-world validation of these approaches. Insilico Medicine's INS018_055, an anti-fibrotic compound targeting idiopathic pulmonary fibrosis, advanced to Phase II trials after being designed using AI platforms [12]. The molecule was designed de novo by generative models optimized for target binding, metabolic stability, and favorable ADMET properties.

Other pharmaceutical partnerships have yielded promising candidates. Recursion's collaborations and Genesis Molecular AI's deployment of the Pearl foundation model in partnership programs

demonstrate industry confidence in AI-driven approaches [14]. However, the relatively small number of AI-designed molecules in late-stage development reflects the field's nascency.

B. Validation Challenges

Despite exciting progress, rigorous validation remains essential. A concerning case involved an AI-generated kinase inhibitor later found to be highly similar to a known drug in the training set, raising questions about novelty versus memorization [15]. This prompted development of guidelines for stringent validation of generative AI models, emphasizing the need for careful analysis of generated structures against training data. Explainable AI methods have emerged as crucial tools for understanding model predictions and building confidence in AI-generated candidates. Techniques for interpreting deep learning decisions help researchers identify whether models are making predictions for scientifically valid reasons rather than exploiting spurious correlations in training data [15].

VI. DATA QUALITY AND AVAILABILITY CHALLENGES

A. The Data Bottleneck

High-quality experimental data represents the most critical factor limiting AI impact in drug discovery [16]. While language models train on trillions of words, the global repository of public protein-ligand structures contains only hundreds of thousands of examples. This data scarcity constrains model development and generalization capabilities. Genesis Molecular AI addresses this limitation through synthetic data generation using physics-based simulations. Their Pearl model trains on proprietary datasets of physics-generated structures, allowing learning from chemical diversity far exceeding experimental databases [14]. This approach demonstrates scaling laws: model accuracy continuously improves with *additional synthetic training data.*

B. Data Harmonization and Quality

Data heterogeneity, sparsity, and class imbalance present ongoing challenges. Bioactivity databases like ChEMBL contain nearly 1.5 million molecules with experimental measurements, but data quality varies significantly [17]. Pharmaceutical companies maintain proprietary databases of 2-3 million compounds with associated data from past drug discovery programs, but this information remains largely inaccessible to academic researchers.

Data harmonization efforts employing domain knowledge and machine learning techniques are improving data utility. Standardizing assay formats, biological conditions, and measurement protocols enables more effective model training. However, privacy, security, and competitive concerns limit data sharing, particularly for clinical trial outcomes [17].

VII. LIMITATIONS AND FUTURE DIRECTIONS

A. Model Interpretability

Deep learning models often function as "black boxes," limiting their utility in scientific contexts requiring mechanistic understanding. Drug discovery involves complex biological systems with interconnected pathways, and current models capture primarily statistical correlations rather than causal mechanisms [17]. Efforts to develop causal inference frameworks and hierarchical generative models that enable human-computer interaction represent important research directions. Displaying semantic information that explains the causality of predictions would significantly enhance model utility. Constructing causal graphs relating molecular structures to drug-like properties could guide more rational molecule generation [17].

B. Physics-Informed Approaches

Integrating physical principles with data-driven learning represents a promising direction. Models that incorporate quantum mechanical calculations, molecular dynamics, or thermodynamic constraints alongside learned patterns may achieve better generalization and interpretability. DiffDock, developed at MIT, employs diffusion models for molecular docking while respecting physical constraints [2].

C. Few-Shot Learning

Many therapeutically important scenarios involve limited data, such as rare diseases with small patient populations. Developing AI methods that learn effectively from few examples would dramatically expand applicability. Meta-learning and transfer learning approaches show promise for adapting models trained on abundant data to data-scarce scenarios [17].

D. Accelerated Validation Pipelines

Reducing the experimental burden for validating AI-generated candidates remains crucial. High-throughput screening, microfluidic platforms, and automated synthesis enable faster evaluation of computational predictions. Integrating computational and experimental workflows through active learning—where experimental results inform model updates—creates virtuous cycles accelerating discovery [13].

VIII. CONCLUSIONS

Artificial intelligence has fundamentally transformed drug discovery and molecular design, enabling exploration of vast chemical spaces with unprecedented efficiency. Deep generative models, graph neural networks, and transformer architectures provide powerful frameworks for de novo molecular generation and property prediction. AlphaFold 3's breakthrough in structure prediction enables accurate modeling of drug-target interactions, opening new therapeutic opportunities.

Modern integrated platforms demonstrate AI's capacity to address multiple stages of drug development, from target identification through molecular optimization. Several AI-designed molecules advancing through clinical trials validate these approaches, though the field remains nascent. Critical challenges persist in data quality and availability, model interpretability, and experimental validation requirements.

Future progress will likely depend more on improving data quality and availability than on algorithmic innovations alone. Physics-informed approaches, synthetic data generation, and few-shot learning methods offer paths forward. As AI tools mature and validation frameworks strengthen, we anticipate accelerated drug development timelines, reduced costs, and expanded therapeutic possibilities. The integration of AI into pharmaceutical research represents not a replacement for experimental science but a powerful augmentation of human expertise, enabling researchers to focus efforts on the most promising candidates and accelerate delivery of new treatments to patients.

ACKNOWLEDGEMENTS

The author acknowledges the extensive research community working on AI-driven drug discovery whose innovations have made this field possible.

REFERENCES

[1] T. Zeng et al., "Deep generative molecular design reshapes drug discovery," *Cell Rep. Med.*, vol. 3, no. 12, p. 100794, 2022.

[2] X. Zeng et al., "Generative artificial intelligence in drug discovery: basic framework, recent advances, challenges, and opportunities," *Front. Pharmacol.*, vol. 15, p. 1331062, Jan. 2024.

[3] R. Gómez-Bombarelli et al., "Automatic chemical design using a data-driven continuous representation of molecules," *ACS Central Sci.*, vol. 4, no. 2, pp. 268-276, 2018.

[4] Y. Ivanenkov et al., "The hitchhiker's guide to deep learning driven generative chemistry," *ACS Med. Chem. Lett.*, vol. 14, pp. 901-915, 2023.

[5] Z. Wu et al., "Graph neural networks for molecular property prediction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 6, pp. 3333-3348, 2022.

[6] B. Shin et al., "Molecular transformer DTI: predicting drug-target interactions using attention mechanisms," *Bioinformatics*, vol. 39, no. 4, pp. 453-461, 2023.

[7] J. Abramson et al., "Accurate structure prediction of biomolecular interactions with AlphaFold 3," *Nature*, vol. 630, pp. 493-500, May 2024.

[8] S. Desai et al., "Review of AlphaFold 3: transformative advances in drug design and therapeutics," *Cureus*, vol. 16, no. 7, p. e64228, Jul. 2024.

[9] "Rational drug design with AlphaFold 3," Isomorphic Labs, 2024. [Online]. Available: https://www.isomorphiclabs.com/articles/rational-drug-design-with-alphafold-3

[10] T. Goddard, "AlphaFold 3 drug screening," UCSF ChimeraX, Dec. 2024. [Online]. Available: https://www.rbvi.ucsf.edu/chimerax/data/af3-drugs-dec2024/

[11] C. Hasselgren et al., "REINVENT 4: modern AI-driven generative molecule design," *J. Cheminform.*, vol. 16, no. 1, p. 20, Feb. 2024.

[12] "Beyond legacy tools: defining modern AI drug discovery for 2025 and beyond," BioPharma Trend, 2024. [Online]. Available: https://www.biopharmatrend.com/ai-drug-discovery-pipeline-2024/

[13] J. Smith, "A smarter way to streamline drug discovery," *MIT News*, Jun. 2024. [Online]. Available: https://news.mit.edu/2024/smarter-way-streamline-drug-discovery-0617

[14] "Introducing Pearl: the next generation foundation model for drug discovery," Genesis Molecular AI, Oct. 2024. [Online]. Available: https://genesis.ml/news/introducing-pearl/

[15] M. Martinez et al., "Lessons learnt from machine learning in early stages of drug discovery," *Expert Opin. Drug Discov.*, vol. 19, no. 6, pp. 631-633, May 2024.

[16] P. Schneider et al., "The future of machine learning for small-molecule drug discovery will be driven by data," *Nat. Comput. Sci.*, vol. 4, pp. 469-476, Oct. 2024.

[17] X. Zeng et al., "Deep generative molecular design reshapes drug discovery," *Cell Rep. Med.*, vol. 3, no. 12, p. 100794, Dec. 2022.

Multi-Cultural Sign Language Recognition Evolution from Image Processing to Advanced Deep Learning Architecture

Paridhi Gusain
Department of Electronics and Communication Engineering
MAHARAJA SURAJMAL INSTITUTE OF TECHNOLOGY (MSIT)
New Delhi, India
paridhigusain29@gmail.com

Jayant Gulia
Department of Electronics and Communication Engineering
MAHARAJA SURAJMAL INSTITUTE OF TECHNOLOGY (MSIT)
New Delhi, India
jayantgulia11@gmail.com

Dr. Shaifali M. Arora
Department of Electronics and Communication Engineering
MAHARAJA SURAJMAL INSTITUTE OF TECHNOLOGY (MSIT)
New Delhi, India
shaifali04@msit.in

4th Dr. Poonam
Department of Electronics and Communication Engineering
MAHARAJA SURAJMAL INSTITUTE OF TECHNOLOGY (MSIT)
New Delhi, India
poonam.dahiya@msit.in

Dr. Anshul Pareek
Department of Electronics and Communication Engineering
MAHARAJA SURAJMAL INSTITUTE OF TECHNOLOGY (MSIT)
New Delhi, India
er.anshulpareek@msit.in

Abstract— This paper presents GmTC, a new system for recognizing sign language gestures across six cultures (Hindi, Korean, Japanese, Bangla, American, and Argentinian). The authors combine graph-based feature extraction with deep learning, achieving state-of-the-art accuracy and enabling real-time cross-cultural communication for people with hearing impairments

Keywords- Multi-Cultural Sign Language (McSL), Hand Gesture Recognition (HGR), Graph Neural Networks (GCN), Multi-Head Self-Attention (MHSA),Deep Learning, Convolutional Neural Network (CNN), Computer Vision, Hindi Sign Language (HSL)

I. INTRODUCTION

The paper addresses the lack of universal sign language and the need for automated translation systems for hard-of-hearing individuals worldwide. Statistics on hearing impairment highlight the scale of the problem. The proposed system aims to enable real-time, cross-cultural communication using advanced computer vision techniques. There is currently no universal sign language—over 300 distinct sign languages are used worldwide, each with unique grammar, vocabulary sign language—over 300 distinct sign languages are used worldwide, each with unique grammar, vocabulary, and cultural context.

This poses challenges for technology and accessibility, as systems built for one sign language often fail in other contexts Global statistics show there are around 70 million deaf people, with the vast majority living in developing countries and relying on visual communication for daily interactions. Because interpreters are scarce, automated translation systems are crucial to bridge gaps, enabling instant, accessible communication for hard-of-hearing individuals. The proposed solution leverages advanced computer vision techniques to recognize signs in real time, addressing both the diversity of languages and the urgency of accessibility in healthcare, education, and social life

For technology, this diversity is a core challenge: a recognizer trained on one sign language often fails on another because handshapes, motion patterns, facial expressions, and even sentence structure differ. Human interpreters are scarce, costly, and not continuously available in daily life (classrooms, clinics, transit, emergencies, workplaces). Automated systems can offer on-demand support. Automation can scale to many settings and devices (phones, kiosks, AR glasses), reducing dependence on scheduled services and enabling more spontaneous communication.

Prevalence data about hearing loss and Deaf signers shows the societal scale: large populations rely on sign language for primary communication, and many more are hard-of-hearing and benefit from accessible interfaces. These numbers justify investment in robust technology: even small improvements in accuracy or latency can affect millions of day-to-day interactions.is the last part of something, its end or result. Sign languages are multimodal: meaning is encoded in handshape, orientation, location, movement, and non-manual markers. Real-time: The system must process video frames fast enough to keep up with natural signing, minimizing latency to avoid conversational breakdowns. Cross-cultural: The recognizer should generalize across multiple sign languages, dialects, and signer styles, or provide

mechanisms to adapt quickly. Training on multi-language datasets: Using models that capture fundamentals (hand pose, motion trajectories, non-manual markers like facial expressions) that transfer across languages. Domain adaptation or few-shot learning to add new languages with limited data.

II. Related Work

A. Overview of the work

Over time the sign language recognition (SLR) has evolved from traditional hand-crafted feature extraction paired with machine learning algorithms like Hidden Markov Models (HMM), Pattern Trees, Linear Discriminant Analysis (LDA), k-Nearest Neighbors (KNN), and Random Decision Forests (RDF), achieving moderate accuracy for specific sign languages.

With growing datasets and complexity, end-to-end deep learning methods gained prominence. Convolutional Neural Networks (CNNs) such as BenSignNet demonstrated high accuracy on Bengali and American Sign Language datasets. Some approaches combined self-attention mechanisms with CNNs to improve generalization.

B. Transformer based models

Transformer-based models like Vision Transformer (ViT) and CNN meets Transformer (CMT) further advanced performance, addressing issues like fixed patch sizes by integrating both CNN and attention layers. However, most existing systems perform well for individual cultural sign languages but struggle to generalize across multiple languages, showing limited scalability and feature extraction efficiency. Later advancements incorporated attention mechanisms and Transformer architectures, which excel at modelling long-range dependencies and temporal dynamics in signing. Vision Transformer (ViT) and CNN meets Transformer (CMT) architectures further addressed challenges like variable hand shapes and pose sequences. However, these models mostly focused on single cultural sign languages and struggled with cross-cultural generalization.

A few multi-cultural models have been attempted, yet they still have limitations in effective feature extraction and real-time deployment. This motivates the development of a more generalized and adaptive multi-cultural sign language recognition system, such as the GmTC model, which integrates graph-based and deep learning features to improve accuracy and generalizability across diverse sign. Existing multi-cultural recognition systems are rare and typically have limitations in scalability and detailed feature extraction to effectively capture heterogeneous sign patterns. These limitations motivate the hybrid graph and deep learning architectures that can jointly represent local spatial relationships (via graph neural networks) and global temporal dependencies (via attention), aiming to achieve robust multi-cultural sign language recognition.

III. PROPOSED SYSTEM

This study evaluates GmTC on several large and diverse datasets:

HSL: Hindi Sign Language dataset

KSL: Korean Sign Language datasets with 77 and 20 classes, thousands of samples.

ASL: American Sign Language datasets, covering basic hand gestures.

BSL: Bangla Sign Language datasets, including a new lab-curated collection.

JSL: Japanese Sign Language dataset (RGB images).

LSA64: Argentinian Sign Language dataset with 64 gestures.

Additional test sets include NTU and Senz3D (digits and general hand poses). The datasets were collected or curated in collaboration with special education institutions, ensuring diversity in signer backgrounds and image properties

A. Mehthodology

The GmTC system uses two parallel streams: GCN Stream: Extracts graph-based features from superpixel representations using Simple Linear Iterative Clustering (SLIC).

Deep Learning Stream: Utilizes a grain module, multi-head self-attention, CNN, and MLP convolution to capture long- and short-range dependencies. Features are refined and concatenated before classification. Mathematical formulations are provided for each module.

The core innovation of the GmTC system lies in its dual-stream architecture, which synergistically combines two complementary approaches to robustly recognize multi-culture sign language gestures. This architecture addresses both local spatial information and long-range dependencies, capturing the rich complexity of sign languages from diverse cultural backgrounds.

The first stream is the Graph Convolutional Network (GCN) stream. In this stream, gesture images are first segmented into superpixels using the Simple Linear Iterative Clustering (SLIC) algorithm. Superpixels group together adjacent pixels with similar color or texture characteristics, reducing noise and computational complexity by operating on these higher-level image regions rather than individual pixels. Each superpixel is treated as a node in a fully connected graph, and edges encode spatial relationships based on distance. The GCN then performs convolutions over this graph, aggregating feature information from neighboring superpixels. This approach effectively models local spatial dependencies and the geometric structure of hand gestures, enabling the network to learn distance-based, complex relational features that traditional pixel-based convolutions may miss. Additionally, this graph-based representation leads to reduced computational requirements compared

to dense pixel-wise operations, making the system more efficient and suitable for real-time applications.

The second stream is a deep learning pipeline inspired by well-established convolutional networks like ResNet, augmented with attention mechanisms. It begins with a grain module, designed to extract fine-grained initial features from the input images. This module uses a series of convolutional layers to distill essential spatial details while reducing dimensionality. Following this, multiple stages of multi-head self-attention (MHSA) modules capture both long-range and short-range dependencies in the data. The self-attention mechanism allows the model to weigh the importance of different spatial features adaptively, learning contextual relationships that can span the entire image. To complement the attention modules, convolutional neural networks (CNNs) are applied to extract localized features that capture the fine spatial patterns of hand shapes and orientations. After attention and convolutional processing, a multilayer perceptron (MLP) convolution module refines the features further by integrating global and local information, contributing to sharper class boundaries and improved discriminability.

The final stage in the deep learning stream is a feature refining module, which implements a feed-forward network to identify and enhance the most relevant features while suppressing noise or redundant information. This selective refinement improves overall prediction accuracy and model robustness.

The outputs of both the GCN and the deep learning streams are concatenated to form a comprehensive feature vector that represents both local spatial relationships and global contextual dependencies. To mitigate overfitting, a dropout layer regularizes this combined representation during training. Subsequently, a fully connected layer with softmax activation performs the final classification into specific sign language classes.

Key design choices such as patch aggregation in the grain module, the use of multiple stages of attention, and the parallel fusion of graph-based and convolutional features enable GmTC to generalize well across different sign languages and signer variations. This architecture balances model complexity with computational efficiency, allowing it to outperform several state-of-the-art models on multiple multicultural sign language datasets. Overall, GmTC exemplifies an effective fusion of graph neural networks and advanced deep learning techniques tailored for the dynamic, multimodal nature of sign language recognition.

B. Experimental Evaluation

- Training: Conducted using PyTorch, Adam optimizer. Standard train/test splits (typically 70/30). Hyperparameters were carefully tuned.
- Study: Two streams (GCN and CNN/attention) were tested separately and in combination. Fusion of streams led to superior results.
- Results: GmTC achieved 99–100% accuracy on KSL, 99.6% on ASL, up to 96.8% on Bangla SL, 88.8% on HSL, and high scores on JSL and Argentinian SL datasets. It outperformed strong baselines, including CNNs, LSTMs, and various transformer models.
- Complexity: Parameters (~8 million) and computation (~130 BFLOPs per batch) were reasonable for real-time deployment.

IV. RESULT AND DISCUSSION

After Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions. Authors are strongly encouraged not to call out multiple figures or tables in the conclusion—these should be referenced in the body of the paper.

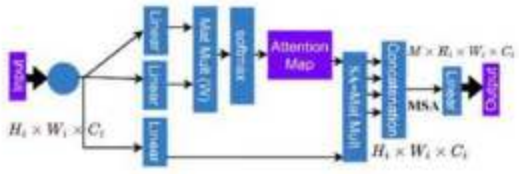


FIGURE 1 : ARCHITECTURE OF MHSA ATTENTION MODULE.

The results of the GmTC model evaluated on multiple sign language datasets, highlighting its effectiveness and superiority over existing models:

1. Hindi Sign Language (HSL) Dataset:

   The model achieved 88.83% accuracy on the HSL-77 dataset and a perfect 88% accuracy on the HSL-20 dataset. This is significantly higher than prior methods such as a CNN (87%) and a combined CNN-attention model (88% for HSL-77 and 88% for HSL-20).

2. Korean Sign Language (KSL) Dataset:

   The model achieved 99.33% accuracy on the KSL-77 dataset and a perfect 100% accuracy on the KSL-20 dataset. This is significantly higher than prior methods such as a CNN (79%) and a combined CNN-attention model (89% for KSL-77 and 98% for KSL-20).

3. American Sign Language (ASL) Dataset:

   GmTC reached 99.46% and 99.60% accuracy on ASL-10 and ASL-20 datasets respectively. These results surpass previous CNN+SVM models that reported 97% accuracy.

4. Bangla Sign Language (BSL) Dataset:

On an existing BSL dataset, the method achieved 93.50% accuracy, and on a lab-curated BSL dataset, accuracy rose to 96.88%. This outperforms earlier methods like VGG19-based approaches which recorded 89.60% for BSL recognition.

5. Japanese Sign Language (JSL), Digit, and Datasets

GmTC achieved 92.37% on JSL, 97.22% on NTU digits, and 99.74% on Senz3D hand gesture datasets. These results compare favorably with recent state-of-the-art approaches relying on GAN and Transformer models.

6. Large Scale Argentinian Sign Language (LSA64) Dataset:

While 3D CNN and SVM approaches reached 93.9% and 85% respectively, and Inception CNN-LSTM achieved 96%, GmTC established superior accuracy on the full set of 64 classes, demonstrating outstanding generalizability.

FIGURE 2 : SAMPLE IMAGE OF ASL SIGN WORD DATASET

Overall, the GmTC model successfully combines graph convolutional networks with multi-head self-attention and CNN-based deep learning to capture both local spatial relationships and long-range dependencies in sign language gestures. This hybrid approach significantly enhances feature representation, leading to state-of-the-art performance across a variety of sign languages and datasets. It also balances high accuracy with computational efficiency, making it well suited for real-time, multicultural sign language recognition applications.

This broad evaluation establishes GmTC as a leading model for multi-cultural sign language recognition, addressing limitations of prior single-culture systems and advancing the goal of universally accessible sign language technology.

A. Figures

GmTC's hybrid architecture addresses both local spatial cues (via graphs) and global gestures (via attention). Its adaptability is key for multi-cultural, real-world deployment. The system is scalable to more languages and can be integrated into desktop, web, and mobile apps for instant gesture translation, greatly expanding accessibility for the hearing-impaired community.

Graph meets with Attention and CNN (GmTC) to address the challenges of enhancing performance accuracy and generalizability for McSL recognition. GmTC is designed to outperform high-performance convolutional models and canonical transformers. Unlike many previous transformer-based hand gesture recognition systems that segmented the input image into patches and extracted features individually, resulting in poorly constructed models and the implementation of linear projections, GmTC takes a different approach. The proposed GmTC system constructs a hybrid network by leveraging the superpixel-based GCN for local features and the long-range dependency of features from MHSA with CNN. This innovative design enhances the model's effectiveness by considering spatial distance-based relationships among super-pixels

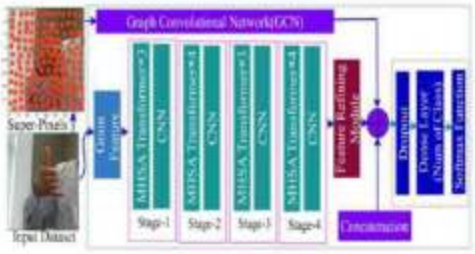


Figure 3 : SYSTEM ARCHITECTURE

B. Conclusion and Future work

The GmTC introduces an innovative hybrid architecture combining graph convolutional networks and deep learning attention mechanisms, effectively addressing the challenges of multi-cultural sign language recognition. The model demonstrates superior accuracy and generalizability across multiple sign languages and datasets, surpassing existing state-of-the-art methods. By integrating local spatial relationships and global contextual dependencies, GmTC offers a scalable and efficient solution for real-time, cross-cultural sign language translation

For future work, the authors aim to expand the system to cover more sign languages and dialects, further improving adaptability and robustness. There is potential to extend the model to support continuous sign language recognition and multi-modal inputs including facial expressions and body movements. Enhancements in model compression and optimization will facilitate deployment on edge devices such as smartphones and wearable AR systems. Exploring user interaction feedback for incremental learning and incorporating AI-generated avatars for sign language display are also promising research directions. Overall, the system paves the way toward universal, accessible communication tools for the deaf and hard-of-hearing communities worldwide. or heads, are organizational devices that guide the reader through your paper.

ACKNOWLEDGMENT

The authors thank the Department of Electronics and Communication Engineering, Maharaja Surajmal Institute of Technology, New Delhi, for their guidance and support. Special thanks to Dr. Shaifali M. Arora for mentorship throughout the project.

REFERENCES

[1] G. Eason, B. J. Lee, C. Y. Yoon, J. Lee, T. H. Kong, and Y. J. Seo, "A situational analysis of ear and hearing care in South Korea using who ear and hearing care situation analysis tool," Front. Public Health, vol. 11, 2023,Art. no. 1215556.

[2] [2] J. Shin, M. A. M. Hasan, A. S. M. Miah, K. Suzuki, and K. Hirooka, "Japanese sign language recognition by combining joint skeleton-based handcrafted and pixel-based deep learning features with machine learning classification," Comput. Model. Eng. Sci., pp. 1–21, 2024. [Online].

[3] [3] G. Dewaele, F. Devernay, and R. Horaud, "Hand motion from 3D point trajectories and a smooth surface model," in Proc. Eur. Conf. Comput. Vis., 2004, pp. 495–507.

[4] [4] S. J. Miah, A. S. Musa, M. A. M. Hasan, M. A. Rahim, and Y. Okuyama, "Rotation, translation and scale invariant sign word recognition using deep learning," Comput. Syst. Sci. Eng., vol. 44, no. 3, pp. 2521–2536, 2023.

[5] [5] A. S. M. Miah, J. Shin, M. A. M. Hasan, and M. A. Rahim, "Ben-signnet. Bengali sign language alphabet recognition using concatenated segmentation and convolutional neural network," Appl. Sci., vol. 12, no. 8, 2022, Art. no. 3933.

[6] [6] M. F. Nurnoby and E.-S. M. El-Alfy, "Multi-culture sign language de-tection and recognition using fine-tuned convolutional neural network," in Proc. Int. Conf. Smart Comput. Appl., pp. 1–6, 2023.

[7] [7] G. Marin, F. Dominio, and P. Zanuttigh, "Hand gesture recognition with jointly calibrated leap motion and depth sensor," Multimedia Tools Appl., vol. 75, no. 22, pp. 14991–15015, 2016.

[8] [8] S. H. I. Yuanyuan, L. I. Yunan, F. U. Xiaolong, M. Kaibin, and M.Qiguang, "Review of dynamic gesture recognition," Virtual Reality Intell. Hardware, vol. 3, no. 3, pp. 183–206, 2021.

[9] [9] A. S. M. Miah, M. A. M. Hasan, and J. Shin, "Dynamic hand gesture recognition using multi-branch attention based graph and general deep learning model," IEEE Access, vol. 11, pp. 4703–4716, 2023.

[10] [10] S. Jetley, N. A. Lord, N. Lee, and P. H. S. Torr, "Learn to pay attention," in Proc. Int. Conf. Learn. Representations, Vancouver, BC, Canada, 2018.

[11] [11] Y. Iwai, K. Watanabe, Y. Yagi, and M. Yachida, "Gesture recognition by using colored gloves," in Proc. IEEE Int. Conf. Syst., Man Cybern. Inf. Intell. Syst., 1996, pp. 76–81.

[12] [12] A. D. Wilson and A. F. Bobick, "Parametric hidden Markov models for gesture recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 9, pp. 884–900, Sep. 1999.

[13] [13] H.-K. Lee and J.-H. Kim, "An HMM-based threshold model approach for gesture recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 21, no. 10, pp. 961–973, Oct. 1999.

[14] [14] C. Kwok, D. Fox, and M. Meila, "Real-time particle filters," Proc. Annu. Conf. Neural Inf. Process. Syst., vol. 15, 2002.

[15] [15] J. Guo et al., "CMT: Convolutional neural networks meet vision trans-formers," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2022, pp. 12165–12175.

[16] [16] J. Shin et al., "Korean sign language recognition using transformer- based deep neural network," Appl. Sci., vol. 13, no. 5, 2023, Art. no. 3029.

[17] [17] S. Yang, S. Jung, H. Kang, and C. Kim, "The Korean sign language dataset for action recognition," in Proc. Int. Conf. Multimedia Model., 2020, pp. 532–542.

[18] [18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethink-ing the inception architecture for computer vision," in Proc. IEEE Conf

OPENRecon; a modular, open-source and customizable medical image reconstruction software

Ram,
Department of Biomedical Engineering,
Deenbandhu Chhotu Ram University of Sc. & Tech.,
Murthal, Sonepat

Sandhya Chaudhary,
Department of Biomedical Engineering,
Deenbandhu Chhotu Ram University of Sc. & Tech.,
Murthal, Sonepat

Naina,
Department of Biomedical Engineering,
Deenbandhu Chhotu Ram University of Sc. & Tech.,
Murthal, Sonepat

Jatin Yadav,
Department of Biomedical Engineering,
Deenbandhu Chhotu Ram University of Sc. & Tech.
Murthal, Sonepat

Poonam Sheoran
Department of Biomedical Engineering,
Deenbandhu Chhotu Ram University of Sc. & Tech.,
Murthal, Sonepat

Abstract

In this century, DICOM viewers (softwares) have become an integral part of radiology and have been extensively used in research and clinical settings for making diagnosis more accurate. Although many softwares (launched by medical device industries) are already available, however, there is still a need for open source, easy to use, and completely illustrative software for medical images. In the proposed work, Python-based OPENRecon software is designed for making the image processing / reconstruction a very easy and fully explainable task. The key features of this software are 3D modelling, introduction of smart status bar, facilitation of image export as well as user centric feature addition.

Keywords: DICOM, Open-source, Python, 3D modelling

1. Introduction

Medical imaging has become a fundamental part of modern diagnosis, treatment planning & biomedical research. There are various imaging techniques such as CT, MRI, PET, SPECT etc. that produce a lot of data that should be reconstructed into meaningful images. The process of image reconstruction converts raw data into visual representations that can be evaluated by radiologists. There are several research-based & commercial image reconstruction software available in the market. Most of them have limits students & researchers from experimenting with new algorithms or customizing the workflow according to their needs.

Therefore, there is a great demand for an open-source, modular & user-friendly image reconstruction platform that supports innovation. Open Recon is developed to fulfil the various requirements as it is an open-source, customizable, easy-to-use software designed for image reconstruction in medical imaging. System supports standard medical image

formats like DICOM & allows users to extend its modules as per their knowledge. The main goal of Open Recon is to make an open-source image reconstruction an accessible, customizable, easy-to-use tool. It encourages team-based research & continuous improvement in medical imaging technologies.

2. Existing Solutions

DICOM viewers are software tools used to view and analyze medical images such as CT, MRI, SPECT , PET scans etc . Many open-source and free DICOM viewers have been developed as cost effective option to commercial software. These tools can be used on web platforms and help doctors and researchers in medical imaging, telemedicine, and data analysis. OsiriX (Rosset et al., 2004) was one of the first and most popular open-source DICOM viewers. It was developed for Mac OS and allowed users to view large 3D medical images quickly using OpenGL and open libraries like ITK and VTK. Later, Ariani et al. (2013) used OsiriX in research to study lung fibrosis in systemic sclerosis patients. Their work showed that OsiriX could analyze lung images accurately and faster than traditional methods. In 2015, Haak et al. compared 28 DICOM viewers, including both free and commercial ones. They studied features like platform, interface, and 2D/3D viewing. Their results showed that MIPAV, Weasis, and OsiriX Lite were among the best for research and hospital use. Wadali et al. (2020) evaluated several free, web-based DICOM viewers for India's eSanjeevani telemedicine project. The goal was to find a lightweight and browser-based viewer that could display both DICOM and non-DICOM images. Ruiz et al. (2020) introduced Starviewer, another open-source DICOM viewer developed using C++ and libraries like Qt and VTK. They also created a new evaluation method to test such tools based on user needs. In their comparison, Starviewer performed better than many existing viewers and showed that modern open-source tools can be both flexible and user-friendly.

3. Proposed Work

We propose "OPENRecon" a fully standalone application using the Python programming language. It is designed to streamline quantitative analysis and 3D reconstruction of medical images. The proposed framework is focused on a user centric approach and to simplify creation of interactive 3D models from 2D DICOM images as well as quantitative analysis. Open Recon provides a flexible and powerful platform aiding in research and education without the cost barriers of commercial medical software.

3.1 System Architecture

The proposed framework is built upon a multi-layer architecture with modular functions to ease any future expansion. The main components of Open Recon are:

1. Presentation Layer (UI) : The user interface is engineered with Python's Tkinter framework alongwith ttkthemes to enhance usability and a clean experience. It handles the user navigation and registers user clicks and cursor position.

2. Control Logic : This layer is central for the communication between the User Interface and the Data Processing Engine. It relays the commands from the UI to visualization modules and vice versa.

3. Data Processing Engine : The data processing engine uses PyDicom for extraction of metadata as well as pixel data from the DICOM headers. The extracted data is relayed to NumPy for mathematical operations and creation of 3D volumetric arrays from the 2D slices. **4. Visualization Kernals :** Pillow takes up the NumPy arrays and use them for 2D image reconstruction. The 3D rendering is achieved by using powerful VTK library which is a standard for scientific modelling to create interactive three dimensional models from the volumetric data.

3.2 Data Processing and Workflow

OPENRecon uses an intuitive workflow for a comfortable and dynamic user experience. The following steps guides the processing pipeline and user workflow :
1. Data Ingestion : The user selects a folder containing DICOM dataset using the File (open folder) option which is scanned and parsed to load valid images.

2. Pre-processing : The parsed files are filtered to remove any invalid files using the CTImageStorage tag and images are sorted using InstanceNumber metadata tag.

3. 2D Analysis : The users are provided with a full suite of 2D tools to enable quantitative measurements such as distance, area, and angles.

4. Smart Status Bar : The smart status bar is a novel feature that provides users with real time coordinates, pixel value, Hounsfield Units as well as tissue classification based upon the cursor position.

5. 3D Segmentation : Various structures can be segmented using Tissue Pre-Select options (bone, soft tissue, fat, full range etc) based upon Hounsfield Units thresholds.

6. Quantitative Analysis : The volume of tissues is calculated by counting the number of voxels in a segmented structure and multiplying with the voxel volume. This gives valuable numerical insights into the 3D anatomy of the target structure.

7. Visualization and Results : Users are enabled to explore the images and 3D models for qualitative and quantitative analysis in a single interface providing a seamless experience in clinical research and education.

4. Results and Discussion

OPENRecon successfully implements all the functionalities as outlined in the proposed workflow. It significantly enhances the traditional 2D viewing experience by providing a cohesive environment rather than being an isolated set of features.

4.1 Integrated 2D/3D Environment

One of the major achievements of OPENRecon is the seamless integration of 2D viewer and 3D rendered in a single environment. This is achieved using the modular Panned Window layout of tkinter which accommodates the tools suite, metadata table, 2D canvas and the rendering window. All the sections are resizable and non-overlapping.

4.2 2D Analytical Toolkit

The 2D analysis tools suite, including distance, area, and angle measurements were successfully implemented. All the measurements are provided in real world units rather than pixel values.

4.3 Live Data Display

The " Smart Status Bar " provides live feedback to the user on movement of cursor across the 2D image. The dynamically displayed information consists of image coordinates, pixels values, Hounsfield Units, and corresponding tissue types. Figure 1 shows

the Smart Status Bar displaying the Live Data based on cursor position.

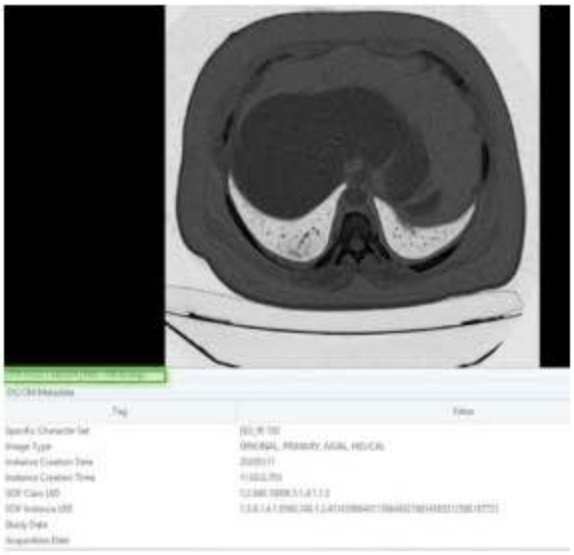


Figure 1 : Smart Status Bar

4.4 3D Volumetric Reconstruction and Quantitative Analysis

OPENRecon successfully generates interactive 3D models from the 2D DICOM datasets. This is achieved using the voxel counting and multiplication with voxel volume in a target preset based on HU units. Figure 2 shows the 3D reconstruction of bones from a curated 427 images chest CT scan.

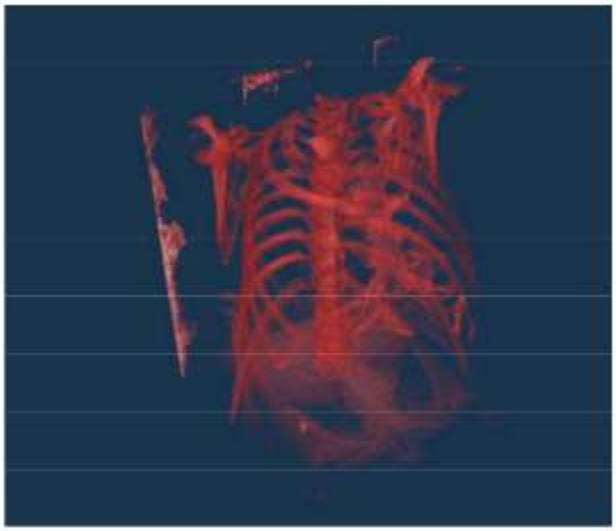


Figure 2 : 3D Model of Chest CT (Bone)

4.5 Overall Discussion and Implications

The combination of integrated interface, 2D measurement tools and HU based 3D segmentation transforms Open Recon from a conventional DICOM viewer to a fully functional analysis workstation. The strength of this software lies in the modularity and seamless functioning of its features. This is seen in how the user can pinpoint a structure using 2D tools while also visualizing it for better analysis without disrupting the workflow. By leveraging the DICOM metadata for both visualization and quantitative analysis the platform overcomes the common limitation of 2D and 3D features being separate in traditional softwares. In essence, the key contribution of this work is creation of a cohesive, integrated and user-centric environment for seamless slice-based examination with volumetric understanding for a faster and accurate image analysis.

5. Conclusion and Future Scope

5.1 Conclusion

This work introduces OPENRecon, a fully python based open-source software for medical image analysis. The framework shows that it is feasible to create a cohesive application with integrated 2D measurement tools with 3D volumetric rendering. Leveraging various python libraries like NumPy, Pydicom, Vtk and Pillow, OPENRecon enables 3D volume reconstruction while also allowing 2D quantitative measurements. This validates the core objective of making a modular, user-centric and open source alternative to commercial medical software.

5.2 Future Scope

OPENRecon is made with a modular architecture to allow future enhancements. The key improvements may include but not limited to

1. Advanced Segmentation Tool : Incorporating a semi-automatic option to segment any tissue from a user selected point
2. Image Fusion : A feature to allow overlapping of images from complementary imaging modalities like CT & PET to allow simultaneous visualization of functional and anatomical context.
3. Session Management : Enhanced session management to let users save 2D annotations as well as 3D models in a comprehensive way.

References

Ariani, A., Carotti, M., Gutierrez, M., Bichisecchi, E., Grassi, W., Giuseppetti, G.M. and Salaffi, F., 2014. Utility of an open-source DICOM viewer software (OsiriX) to assess pulmonary fibrosis in systemic sclerosis: preliminary results. Rheumatology international, 34(4), pp.511-516.

Haak, D., Page, C.E. and Deserno, T.M., 2016. A survey of DICOM viewer software to integrate clinical research and medical imaging. Journal of digital imaging, 29(2), pp.206-215.

Hulbaek, M., Knutz, E., Ebbesen, N.T., Primdahl, J., Nielsen, J.B. and Birkelund, R., 2020. Pelvic organ prolapse and treatment decisions–developing an online preference-sensitive tool to support shared decisions. BMC Medical Informatics and Decision Making, 20(1), p.265.

Rosset, A., Spadola, L. and Ratib, O., 2004. OsiriX: an open-source software for navigating in multidimensional DICOM images. Journal of digital imaging, 17(3), pp.205-216.

Wadali, J.S., Sood, S.P., Kaushish, R., Syed-Abdul, S., Khosla, P.K. and Bhatia, M., 2020. Evaluation of free, open-source, web-based DICOM viewers for the Indian national telemedicine service (eSanjeevani). Journal of Digital Imaging, 33(6), pp.1499-1513.

LOW LIGHT IMAGE ENHANCEMENT USING MACHINE LEARNING

Vaani Garg[1]
Maharaja Surajmal Institute of Technology
[1]vaani.garg@msit.in

Palak Gupta[2]
Maharaja Surajmal Institute of Technology
[2]palakgupta1506@gmail.com

Abstract- Under dim lighting, less light captured by camera sensors makes it harder to recognize objects and leads to loss of details. This results in restricting the image processing task. This study presents a new method based on the Generative Network Approach that combines adversarial network framework with generative network for improving the low light image quality. This generative network consists a multi scale feature extraction module which combines standard, dilated convolutions and pooling operations for efficiently capture features at numerous scales. Furthermore, an illumination attention module highlights and improves the details, brightness of images, and enhances key lighting features. This method improves details and brightness of images by processing them via an encoder and decoder. In this generative network, local and global dual discriminator structure is used. Local focusing on the finer features while global focusing on analyzing the overall image appearance. This ensures improvement of images locally and globally. Further, color and perceptual loss are incorporated by an enhanced loss function for minimizing distortion, improving perceptual quality, and to preserve structural integrity. The proposed method performed well in terms of clarity, color accuracy, and preservation of details on comparing with the existing methods, when tested on both synthetic and real low light data sets. The results display that how effectively this approach works in improving visual interpretability under difficult lighting situations.

Keywords— Dim/low light, generator adversial network, multi scale feature extraction, illumination attention module, loss function.

I. INTRODUCTION

Under dim lighting, there is a problem of gathering enough light while capturing the images by various devices. This results in negatively affecting the automated driving systems, dashcam recordings, and nighttime traffic surveillance. Considering this, the main goal is to improve the quality of images captured at night. Image processing techniques can be used to achieve this goal more affordably. Although conventional methods such as histogram equalization and Retinex-based techniques have shown effectiveness in this. Convolutional Neural Networks (CNNs) with the advancement in GPU technology perform well in the real world practical applications. In this study, it is advised to use a particular generative adversarial network (GAN) under dim lighting to enhance the image processing process. The objective of this study is to include:

1. The model is designed specifically to enhance low-light images. This network integrates an illumination attention module, which allows the network to focus on essential lighting features, with a multi-scale feature extraction module that provides details at various scales. The components are embedded within an encoder-decoder structure to maximize the generator's effectiveness.
2. This discriminative network includes both a local and a global discriminator, enhancing the network's ability to distinguish and improve various aspects of brightness and background details in the images. This dual approach supports the generative network in achieving more refined image enhancements.
3. This enhanced loss function improves the color loss and perceptual loss incorporation into the loss function in order to provide closer approximations in terms of color accuracy and less distortion in the production of visually correct images.

II. LITERATURE REVIEW

Nowadays, users want easier, reliable and faster technologies that can make their lives relatively simple [3]. An average smartphone user admires the experience of simply clicking to capture a decent picture without the complications of all the relevant technological jargon. During good lighting conditions, there is no issue as the camera and software are intelligent enough to adjust those parameters to the user's liking but in low light conditions, even the best cameras tend to give a poor result.

Conventional methods, such as histogram equalization and Retinex-based algorithms, aim to improve image visibility and contrast [4]. While histogram equalization redistributes pixel values for better contrast, it often causes imbalances and compromises texture. Human vision driven Retinex-based methods break down images into reflectance and illumination layers to improve visibility. However, these approaches encounter difficulties like flaws in color and noise sensitivity in intricate scenes. Advanced versions, which include fractional-order denoising and multi-scale wavelet transforms, provide modifications but still struggle with maintaining color fidelity and fine textures.

Convolutional Neural Networks (CNNs) demonstrated potential in learning complicated patterns for low-light enhancement [5]. Numerous models, like illumination-centred Retinex networks and attention-based enrichment systems, attain better edge retention and stable lighting.

Though issues such as overexposure, irregular enhancement, and loss of texture detail continue, mainly in exceptionally low-light conditions.

GANs acquired popularity for creating realistic images by utilizing adversarial learning [6]. These networks comprise a generator and a discriminator that collaborate to enhance image quality. Recent GAN-based approaches integrate methods like Zero-DCE preprocessing, self-attention modules, and encoder-decoder architectures to improve detail and brightness. Although effective, many existing GANs have difficulty with color distortion, overexposure, and constancy in texture preservation.

Framework focusses on the strengths of GANs despite tackling their restrictions, offering an additional sturdy solution for low-light image enhancement [7]. By preserving color fidelity, texture, and balanced illumination, it paves the way for clearer and more visually accurate low-light images. Squamous, basal cell and melanoma as the three primary subtypes of skin cancer are discussed. The authors trained the model with HAM10000 dataset, a lesion-based image dataset. Enforcing an autoencoder for feature extraction followed by classification with the help of v3 pretrained model and 3D CNN leading to an accuracy of 0.96. The help of an autoencoder as a feature extraction brings enhances efficiency and performance.

Enhancements to Retinex, including regularization models (Wu et al.) and generative diffusion models (Yi et al.), introduce more sophisticated layers of inference but often still suffer from unwanted artifacts [8]. Additionally, methods based on fractional-order denoising, multi-scale wavelet transforms, channel-aware models, and contrast adjustment have been developed, yet none fully resolve issues of color fidelity and texture retention.

Chen et al. [9] proposed an illumination-focused Retinex network incorporating a specialized illumination module for light recovery, but it suffers from local overexposure issues. In addition, Ahn et al. presented a learning-based Retinex model with a gamma modification module to smooth the spreading of light, but it lacks robustness in maintaining texture. Rahman et al. joined a CNN with a prior bright channel to handle uneven exposure; though, it fails to attain a balanced enhancement throughout numerous lighting conditions. Li et al. established an attention-based enhancement network with modular stacking, which augments images via feature extraction and fusion, but may develop extra disturbances

Cycle-consistent GANs (CycleGANs), offered by Jia et al. [10], utilize encoder-decoder architectures with contextual feature extraction to improve image details by integrating low- and high-frequency details. These GANs, however, sometimes provide blurring and lack constancy in texture preservation. Xu et al. offered a superior GAN with a novel objective function that restores high-frequency details effectively but occasionally cause image blurring. Jin et al. utilized the Retinex model with ProGAN to achieve progressive image enhancement, decomposing low-light images into reflectance and illumination components. However, the method may result in color distortion.

Yang et al.16 recently developed an end-to-end Retinex mapping network [11]. The network adjusts the output of the decomposition sub-network by applying a sparse gradient constraint method. The network is capable of retaining the essential edge information while eliminating the low-amplitude interference present in the image. It is important to note that the network is not capable of accurately recovering images with extremely low luminance.

III. SCOPE OF WORK

Our proposed approach enhances the low-light images to a more enhanced and detailed image, similar to the ones captured in illuminated environment. For the consumer, a user-friendly website has been developed through which they can access our system. To reduce the training time, we have used images of resolution 256x256 for our proposed method, as images with a higher resolution take longer to process and train.

Data collection and pre-processing techniques like data augmentation, noise removal, image pixel resizing, normalization and spatial shifts are incorporated to refine the dataset and set up an efficient foundation for training and testing of the deep learning models.

This project leverages established approaches from existing literature to craft a system that both, validates and improves upon it. The insights gained through the in-depth evaluation of multiple methodologies and frameworks, allow to build upon the former studies to optimize and enhance their performances. The high potential of deep learning in image enhancement field allows contribution to the field of computer vision.

IV. MATERIALS AND METHODS

The methods and techniques involved in the development process of this system are described in this section.

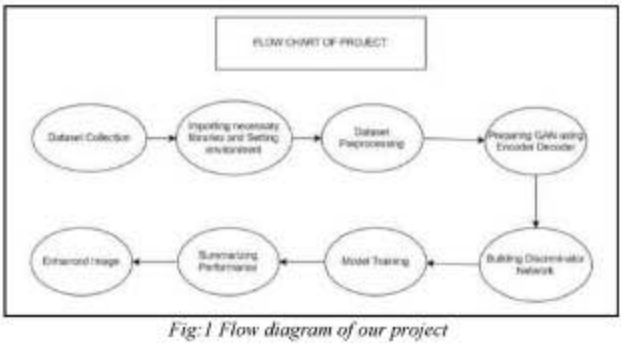


Fig:1 Flow diagram of our project

A. Data Collection

The first phase involves gathering a comprehensive dataset of low-light images from various sources like LOL-v1, LOL-v2, Sony Image Dataset (SID), Smartphone Image Dataset (SMID) etc., including public image datasets, real-world night photography collections, and synthetic low-light image generation. The dataset should cover a diverse range of environments, such as indoor, outdoor, and night scenes,

to ensure the model's robustness across different lighting conditions. Once it has been collected, the data will be pre-processed and it will include normalization, augmentation example (such as random cropping, flipping and brightness adjustment) and splitting into training, validation, and testing sets, in order to make sure that the model has reached its Generalizability and efficacy.

Fig:2 LOL-v1, SMID Datasets

B. Data Preprocessing

Firstly, convert images from formats like PNG, JPEG, or JPG to NPZ, a compressed format, which is ideal for effective handling and storage in machine learning workflows for pre-processing the dataset. Afterwards, use a script, such as png2npz.py, in order to automate this conversion process. Then, set the path variable in the script to the directory comprising the images, and implement the code to create the NPZ file. After conversion, complete normalization to confirm pixel values are inside a standardized range, and implement augmentation techniques such as random cropping, flipping, and brightness changes to rise the data variability and model resilience. At last, separate the dataset into training, authentication, and testing sets. Then, upload the NPZ file to Google Drive for quick access and training on Google Collab.

C. GAN Network for Low Light Images

Convolutional Neural Networks (CNNs) are a kind of deep neural network employed to solve tough image-driven pattern recognition and computer vision jobs. It utilizes the concept of hierarchical feature detectors, including multiple layers such as a convolutional layer, a non-linearity layer, a pooling layer, and a fully-connected layer, that operate in a way to recreate the performance of a human brain. Feature extraction is performed by the convolution and pooling layer which then allows the fully connected layer to maps the extracted features to final output. A convolution layer contains a matrix (small grid) of parameters called kernel, which performs feature extraction, and a stack of mathematical operations which are applied at each layer, passing its output onto the next . For our research, two CNN models are made and then, ensembled together to achieve best possible results. Batch Normalisation, a technique to normalise the activations in the intermediate layers of the network, is also applied to increase training speed and accuracy.

D. Encoder

The Encoder Network receives a 4-dimensional array as the input which has 2 parts the first being the number of images and the second is an RGB image of resolution 256 x 256. Encoder Network is made up of 21 convolutional layers. Where each layer has a Convolution2D, BatchNormalization, LeakyReLU. Convolution2D layer creates a convolution kernel that is convoluted with the layer input to produce a tensor of outputs. Batch normalization applies a transformation that maintains the mean output close to 0 and the output standard deviation close to 1. LeakyReLU is a leaky version of Rectified Linear Unit (ReLU), it allows a small gradient when the unit is not active. The input 4-dimensional array is processed and passed on to the Decoder.

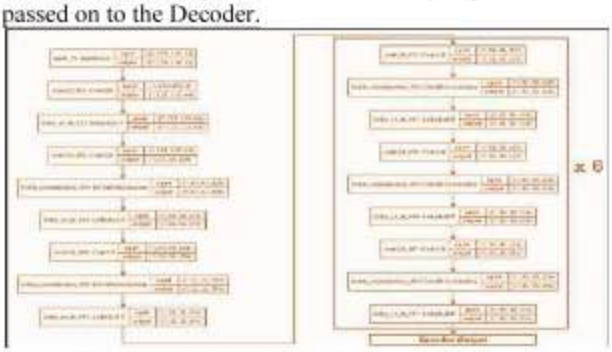


Fig:2 Encoder structure with multiple convolutional layers, LeakyReLU Activation and bath normalization

E. Decoder

The Decoder Network receives a 4-dimensional array from the Encoder as the input. Our Decoder Network is made up of 10 convolutional layers. Where each layer has a UpSampling2D,LeakyReLU,Convolution2D,BatchNormalization. UpSampling2D increases the dimensions by the specified size, there are two methods of interpolation Nearest Neighbour, Bilinear transformation. We decided to use UpSampling2D in place of Convolution2DTranspose, since it is cheap in terms of computational power. Convolution2D layer creates a convolution kernel that is convolved with the layer input to produce a tensor of outputs. Batch normalization applies a transformation that maintains the mean output close to 0 and the output standard deviation close to 1. LeakyReLU is a leaky version of Rectified Linear Unit (ReLU), it allows a small gradient when the unit is not active. The input 4-dimensional array is processed and passed on to the Discriminator.

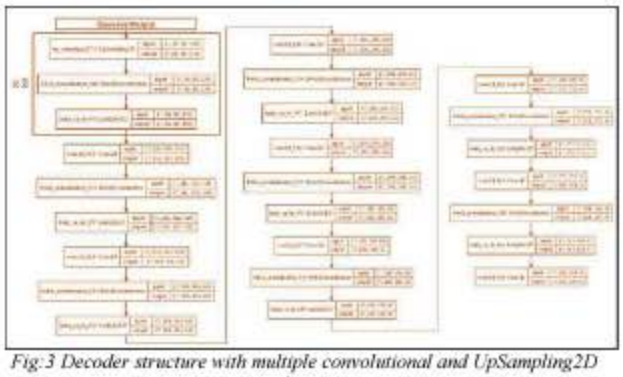


Fig:3 Decoder structure with multiple convolutional and UpSampling2D layer

F. Dual Discriminator

In low-light image enhancement, a global-local discriminator approach inspired by the Markov model improves evaluation by analyzing both broad image structures and localized details. The **global discriminator** examines the entire image using two 4x4 convolutions, three dilated convolutions for multi-scale features, and a residual network for refined feature extraction, culminating in a discrimination score. The **local discriminator** focuses on detailed regions by processing random image crops through six 4x4 convolutional layers with LeakyReLU activations, producing a localized discrimination score. This dual-discriminator setup confirms more precise and nuanced conclusions of low-light images.

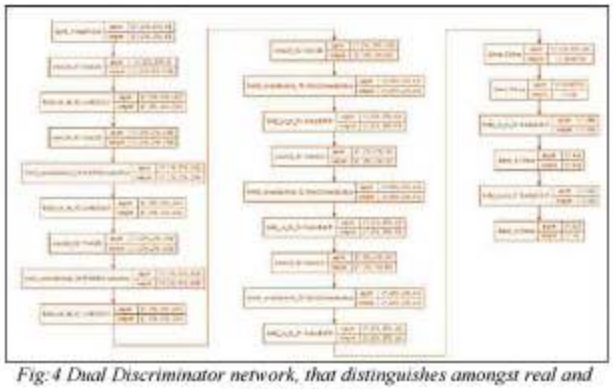


Fig:4 Dual Discriminator network, that distinguishes amongst real and generated images

G. Model Training

Throughout the training process, multiple approaches are investigated, starting with the Cycle GAN model. This is used to interpret images within the low-light and improved domains, hence filtering the ability of the discriminator to differentiate between real and generated images. Training of two Cycle GAN models is done, one with Instance Normalization and another without altering the generator's convolution layers to enhance the outputs. However, preserving accurate colors and edge details was a limitation of these initial Cycle GAN models. The current model used in this study utilizes skip connections between the encoder and decoder networks. This captures thorough information that helps to precisely reconstruct the edges of object. Additionally, to improve the model's ability and generate well-defined edges and enhanced color fidelity, a transformation network between the encoder and decoder further processes the encoder's output. Finally, Google Collab is used to conduct training using a Tesla T4 or P100 GPU with 16GB VRAM, resizing the images to 256x256 and limiting the batch sizes to 12 due to memory constraints.

H. Improved Loss Function

An enhanced loss function is introduced to more effectively measure the disparity between the enhanced image and the normally illuminated image by incorporating color loss and perceptual loss into the standard loss function. The improved loss function can be represented as:

$$\mathbf{LOSS} = \mathbf{L_{GAN}} + \mathbf{L_{per}} + \mathbf{L_{color}}$$

Here, LGAN denotes the overall loss of the generative adversarial network. It encompasses the global and local losses from both the generative network and the adversarial network. L_{per} denotes the perceptual loss, L_{color} represents the color loss.

We use perceptual loss to measure the distance between the features of the generated images and those of the normal illuminated images, which is represented as follows:

$$L_{per} = \frac{1}{M X W X H} \sum_{s=1}^{M} \sum_{i=1}^{W} \sum_{j=1}^{H} || \ \varphi_{s,i,j}\left(G(z)\right) - \varphi_{s,i,j}(x) \ ||_2^2$$

where M is the number of images, x represents the normal illuminated image. G(z) represents the enhanced image. $\varphi_{s,i,j}$ denotes the value at position (i,j) of the feature map extracted from the VGG16 network for the s-th image. W and H represent the width and height of the feature map, respectively.

The color loss is expressed as:

$$L_{color} = \frac{1}{M X N X W X H} \sum_{s=1}^{M} \sum_{cth=1}^{N} | \left(\sum_{i=1}^{W} \sum_{j=1}^{W} x_{s,i,j,cth} - \sum_{i=1}^{W} \sum_{j=1}^{W} z_{s,i,j,cth} \right) |$$

where M is the number of images. N is the number of channels for the input image (N=3). W and H represent the width and height of the feature map, respectively. cth represents the c-th channel of the sample. x and z represent the normal illuminated image and enhanced image, respectively. Color loss measures the difference between the enhanced and normal illumination images on each channel.

V. Results and Discussion

This study uses network of encoder and decoder which uses Leaky ReLU activation and convolutional neural network to retrieve features from dim light image input. Reconstruction of enhanced images is then followed by progressive up sampling. Real and generated images that ensure realistic results are separated by a dual discriminator GAN framework. Scalability and deployment efficiency on numerous platforms like mobile and web apps, are improved by real time optimization. Improved performance metrics, including SSIM, LPIPS, and PSNR, are achieved in this model by employing adversarial, reconstruction, and perceptual losses in order to enhance fidelity and subjective quality. The effectiveness of this model is validated by the comparative studies and user assessments, in comparison to the contemporary methods such as SCI, EnlightenGAN, Zero-DCE++, and KinD.

The main goal of Real time optimization step used in this study is to make efficient and scalable system that can be used in real time applications. This step is crucial to guarantee that the low light enhancement system is independent of platform and may run on numerous platforms like web apps, mobile devices and other consumer focused systems. Concentration on maintaining high quality output with reduction in computational load is achieved by optimizing the strategies which ensure that the system works flawlessly regardless of limited processing power devices. This process uses hardware acceleration, model compression approaches, and efficient inference pipelines. The main aim is to make the system readily accessible to larger audience and to enable it for real time low light image improvement for everyday practical applications in plenty of situations. By doing this, the system will be positioned as a versatile tool that can enhance both professional and consumer applications in imaging and photography.

Fig:5 Real Time Optimization and comparison of image

The training process employs advanced loss functions, including adversarial loss, reconstruction loss, and perceptual loss, to optimize both objective fidelity and subjective quality. These losses guide the generator in creating realistic enhancements, aligning the outputs with human perception while adhering to ground-truth quality metrics. Moreover, the architecture is designed to suppress noise often introduced during enhancement, further improving the visual quality of the output. By employing techniques to differentiate noise from meaningful details, the model achieves clarity and sharpness without sacrificing important features.

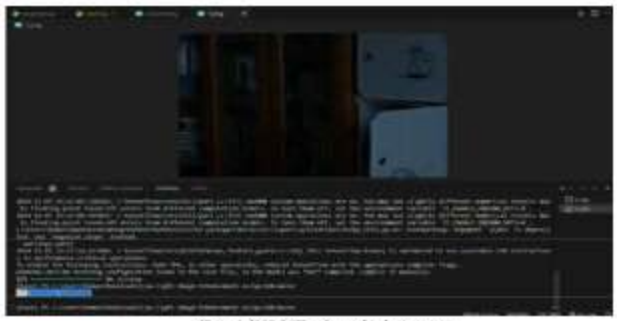


Fig:6 INPUT - Low light image

Fig:7 OUTPUT – Enhanced Image from our model

	KinD	Enlighten GAN	Zero-DCE	Zero-DCE++	SCI	Our
PSNR	26.30	27.31	20.88	21.84	24.60	28.32
SSIM	0.959	0.965	0.917	0.922	0.930	0.935
LPIPS	0.084	0.081	0.104	0.097	0.090	0.061

Fig 8: The table presents a comparison of image restoration performance metrics for various image restoration methods

To validate our proposed method better, we conducted a user study. As the Zero-DCE and Zero-DCE++ methods performed much worse than ours, we compared only KinD, EnlightenGAN, SCI, and our method. For this user study, 30 volunteers were recruited where 100 real low-light images were randomly selected from the internet. We used four variations to enhance these images. These include KinD, EnlightenGAN, SCI, and the proposed approach. We performed three experiments with 100 pairs each where one is enhanced by the comparative method and the other is enhanced by our approach. Volunteers randomly selected 40 pairs from each test and picked the picture they preferred more to the eye, that is, a more natural one. We collected the results of all volunteers and measured the satisfaction of each test. The final rate of satisfaction was then obtained by averaging the ratings over all volunteers. The results, as depicted in Fig. 9, show that our approach surpasses KinD, EnlightenGAN, and SCI in visual quality.

VI. Conclusion

Based on the above consideration, we set up a generative adversarial network of low-light image enhancement consisting of a generator and a discriminator. For the generator, we introduce both a multi-scale feature extraction module and an illumination attention module. We present a dual discriminator structure by virtue of global and local discriminators within it for further improvement in the

ability of the discriminator to distinguish between the enhanced images and the real ones. Therefore, we append color loss and perceptual loss to the loss function to evaluate the distortion degree of an image even better.

We evaluated the methods on both synthesized and real low-light images. For synthesized images, our method produced results that are visually closer to images with normal illumination, achieving the highest PSNR and SSIM scores, along with the lowest LPIPS values. For real low-light images, our method demonstrated the lowest NIQE and the highest MetaIQA, offering clearer and more visually appealing results than the other methods.

Our simulation results show that our method performs better than the compared techniques in both synthesized and real-world low-light image enhancement. However, this approach is currently limited to road scene images. Moreover, using only a GAN for image enhancement has certain limitations, particularly in fully eliminating image distortion. In future work, we plan to incorporate Retinex theory or other illumination models to integrate physical knowledge into the deep learning framework, which could improve the model's understanding of lighting and reflection characteristics. With the help of this, model may become lighter, faster and more efficient. Moreover, in order to tackle and minimize detail loss during enhancement and issues related to distortion, further investigation on multimodal data and edge filtering techniques can be done.

Stable Low Discriminator Loss: Discriminator losses in the log, d1 and d2 values are steadily small (between 0.002 and 0.005), signifying that the discriminator is performing well in precisely recognizing fake and real samples.

Generator Loss Stability: Generator loss continues in a steady range and doesn't increase significantly. This suggests that the generator develops realistic images with good details of edge, color, and structure. The GAN training is balanced, shown by the stable generator and discriminator losses. This is critical for effective GAN training and to avoid problems such as mode collapse.

REFERENCES

[1] Tan, K., Wu, J., Zhou, H., Wang, Y. & Chen, J. Integrating advanced computer vision and AI algorithms for autonomous driving systems. J. Theor. Pract. Eng. Sci. 4, 41–48 (2024).

[3] Rasheed, M. T., Shi, D. & Khan, H. A comprehensive experiment-based review of low-light image enhancement methods and benchmarking low-light image quality assessment. Signal Process. 204, 108821 (2023).

[4] Rahman, Z., Bhutto, J. A., Aamir, M., Dayo, Z. A. & Guan, Y. Exploring a radically new exponential Retinex model for multi-task environments. J. King Saud Univ.-Comput. Inf. Sci. 35, 101635 (2023).

[5] Tian, Z. et al. A survey of deep learning-based low-light image enhancement. Sensors 23, 7763 (2023).

[6] Fu, H. et al. You do not need additional priors or regularizers in Retinex-based low-light image enhancement. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18125–18134 (2023).

[7] Chao, K. et al. Cui-net: A correcting uneven illumination net for low-light image enhancement. Sci. Rep. 13, 12894 (2023).

[8] Jebadass, J. R. & Balasubramaniam, P. Low-light enhancement algorithm for color images using intuitionistic fuzzy sets with histogram equalization. Multim. Tools Appl. 81, 8093–8106 (2022).

[9] Han, Y. et al. Low-illumination road image enhancement by fusing Retinex theory and histogram equalization. Electronics 12, 990 (2023).

[10] Wu, W. et al. Uretinex-net: Retinex-based deep unfolding network for low-light image enhancement. In: Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 5901–5910 (2022).

[11] Yi, X., Xu, H., Zhang, H., Tang, L. & Ma, J. Diff-Retinex: Rethinking low-light image enhancement with a generative diffusion model. In: Proc. IEEE/CVF International Conference on Computer Vision, 12302–12311 (2023).

[12] Pan, X. et al. Low-light image enhancement method based on Retinex theory by improving illumination map. Appl. Sci. 12, 5257 (2022).

[13] Rahman, Z., Pu, Y.-F., Aamir, M. & Wali, S. Structure revealing of low-light images using wavelet transform based on fractional-order denoising and multiscale decomposition. Vis. Comput. 37, 865–880 (2021).

[14] Kim, B., Lee, S., Kim, N., Jang, D. & Kim, D.-S. Learning color representations for low-light image enhancement. In: Proc. IEEE/CVF Winter Conference on Applications of Computer Vision, 1455–1463 (2022).

[15] Rahman, Z. et al. Efficient contrast adjustment and fusion method for underexposed images in industrial cyber-physical systems. IEEE Syst. J. 17, 5085–5096 (2023).

[16] Yang, W., Wang, W., Huang, H., Wang, S. & Liu, J. Sparse gradient regularized deep Retinex network for robust low-light image enhancement. IEEE Trans. Image Process. 30, 2072–2086 (2021).

[17] Chen, Y., Wen, C., Liu, W. & He, W. A depth iterative illumination estimation network for low-light image enhancement based on Retinex theory. Sci. Rep. 13, 19709 (2023).

[18] Ahn, S., Shin, J., Lim, H., Lee, J. & Paik, J. Coden: Combined optimization-based decomposition and learning-based enhancement network for Retinex-based brightness and contrast enhancement. Opt. Express 30, 23608–23621 (2022).

[19] Rahman, Z., Yi-Fei, P., Aamir, M., Wali, S. & Guan, Y. Efficient image enhancement model for correcting uneven illumination images. IEEE Access 8, 109038–109053 (2020).

[20] Wang, X., Sun, L., Chehri, A. & Song, Y. A review of GAN-based super-resolution reconstruction for optical remote sensing images. Remote Sens. 15, 5062 (2023).

[21] Zhou, T., Li, Q., Lu, H., Cheng, Q. & Zhang, X. GAN review: Models and medical image fusion applications. Inf. Fusion 91, 134–148 (2023).

[22] Tian, C. et al. Deep learning on image denoising: An overview. Neural Netw. 131, 251–275 (2020).

[23] Gui, J., Sun, Z., Wen, Y., Tao, D. & Ye, J. A review on generative adversarial networks: Algorithms, theory, and applications. IEEE Trans. Knowl. Data Eng. 35, 3313–3332 (2021).

Design and Integration of Multi-Band Antennas on a Single PCB: GPS and GSM Miniaturization for IoT Devices

SUDESH PAHAL
Dept. of Electronics and Communation Engineering
Maharaja Surajmal Institute of Technology (GGSIUP)
New Delhi , India
sudeshpahal@msit.in

MAYANK SARASWATI
Dept. of Electronics and Communation Engineering
Maharaja Surajmal Institute of Technology (GGSIUP)
New Delhi , India
mayanksaraswati46@gmail.com

NEETU
Dept. of Electronics and Communation Engineering
Maharaja Surajmal Institute of Technology (GGSIUP)
New Delhi , India
neetu@msit.i

KASHISH DHAWAN
Dept. of Electronics and Communation Engineering
Maharaja Surajmal Institute of Technology (GGSIUP)
New Delhi , India
kashishdhawan73@gmail.com

Abstract -

The fast development of the Internet of Things (IoT) requires miniaturized, efficient, and multimodal wireless communication systems. GPS and GSM incorporation is essential in wearable, handheld, and safety-focused IoT devices, providing precise real-time location and connectivity [1], [2]. Conventional ceramic patch and whip antennas require device enlargement, obstructing miniaturization. This paper discusses methods of designing and implementing GPS and GSM antennas on a single PCB with chip antennas, planar inverted-F antennas (PIFA), meandered traces, and multiband structures. Impedance matching, ground plane optimization, and mutual coupling reduction methods are explained. A dual-band IoT tracker's practical implementation is used to demonstrate successful coexistence and miniaturization, confirmed by simulation and field tests. Applications range from wearable safety equipment, smart city infrastructure tracking, industrial monitoring, and healthcare devices. Future developments like AI-controlled impedance tuning, multi-layer PCB antennas, and 5G NR are discussed.

1. Introduction

Miniaturization is a foundation for future IoT applications such as personal safety, health care monitoring, asset tracking, and smart city infrastructure [3], [4]. Portable and handheld devices need compact, lightweight forms without compromising reliable communication and accurate location tracking. GPS offers geolocation accuracy, and GSM provides continuous connectivity—requirements for real-time IoT safety and monitoring applications [5].

Conventional GPS applications tend to employ 25 mm × 25 mm ceramic patch antennas, and GSM tends to employ whip or monopole antennas [6]. Such antennas are space-consuming and incompatible with small IoT devices. Mounting multiple antennas on a common PCB poses problems like mutual coupling, impedance mismatching, detuning, and electromagnetic interference (EMI) [7].

Some of the recent miniaturization techniques are multiband antennas, chip antennas, and meandered PCB traces, which minimize size with preserved performance [8], [9]. AI-tuning and multi-layered PCB structures are also likely future directions for dynamic optimization of wearable and industrial IoT devices [10]. In this paper, past work is extended by providing experimental verification of co-integration of GPS and GSM antennas on a single PCB, with greater explanation of applications and innovative design methods.

2. Structure, Contribution, and Related Work

2.1 Multiband Antennas for IoT

Multiband antennas accommodate multiple standards, i.e., GSM, GPS, Wi-Fi, Bluetooth, and LTE, in one radiating element [11]. Slot loading, meandering, fractal geometries, and PIFA structures are some techniques that allow multiple resonances, minimize PCB footprint, and make IoT device design simpler [12]. Existing literature indicates compact multiband antennas with efficiency and radiation characteristics for wearable and handheld devices [13].

2.2 GPS Antennas

GPS receivers primarily use the L1 frequency band (1575.42 MHz) [14]. Classic ceramic patch antennas have circular polarization and high gain but are unsuitable in miniaturized forms. Miniaturized substitutes—chip antennas, PIFA, and meandered traces—have smaller size at the cost of narrower bandwidth and slightly less gain. Methods like LC matching circuits, low-noise amplifiers (LNA), ground clearance optimization, and optimal substrate selection mitigate these drawbacks [15].

2.3 GSM Antennas

GSM networks use quad-band frequencies: 850, 900, 1800, and 1900 MHz [16]. Standard whip or monopole antennas are large. Small alternatives are PCB monopoles and chip antennas, which can be effectively embedded on a PCB with appropriate impedance matching. Wearable devices also need to account for human body absorption, which may detune the antenna and decrease efficiency [17].

2.4 Miniaturization and Coexistence Challenges

Miniaturization techniques involve high-permittivity substrates, meandered/folded traces, slotting, and metamaterial-inspired structures [18]. Coexistence problems include mutual coupling and interference between GSM and GPS antennas. Solutions involve:

- Orthogonal antenna placement and polarization diversity
- Ground isolation using slots or islands
- Filtering techniques to eliminate GSM harmonics on GPS [19]

Contribution:

- Integration of a single PCB compact with both GSM and GPS antennas
- Coexistence and miniaturization strategies in IoT devices
- Practical implementation of a dual-band tracker
- AI-aided tuning and multi-layer PCB aspects in upcoming designs

3. Multi-Band Antenna Design and Integration

3.1 Substrate Choice

FR4 is used extensively for cost-efficient designs but offers greater dielectric loss beyond 2 GHz. Low-loss substrates such as Rogers RO4350B enhance radiation efficiency, bandwidth, and miniaturization potential [20].

3.2 Ground Plane and Impedance Matching

A properly designed ground plane provides efficient radiation. Impedance matching is obtained via LC networks, stubs, and slots. Ground plane isolation decreases coupling between GPS and GSM antennas [12].

3.3 Integration Techniques

• Orthogonal antenna placement to decrease near-field interference

• Filtering components (e.g., SAW filters) to eliminate harmonic interference

• Slot or single ground island-based ground isolation

• Simulation-based design to optimize resonance and radiation pattern

3.4 AI-Assisted Tuning and Multi-Layer PCB Designs

AI-enabled real-time impedance tuning and multi-layer PCB assemblies enable additional miniaturization and enhanced coexistence. These adaptive architectures can adapt to environmental change and proximity to the human body, boosting wearable IoT device performance [11], [10].

Figure 1: Dual-band IoT tracker block diagram (GPS + GSM)

Figure 2: GPS and GSM antenna placement in PCB layout

4. Practical Implementation

A dual-band IoT tracker PCB was developed and manufactured with the following specifications:

• GPS: The L1 frequency band (1575.42 MHz) was used for a compact chip antenna to minimize PCB footprint with sufficient gain for receiving the satellite signal in outdoor and semi-indoor environments [1], [2].

• GSM: A PCB monopole was utilized for 900/1800 MHz to enable effective communication using standard GSM networks. The structure provides efficient impedance matching and a stable radiation pattern for handheld devices [3].

• Layout Strategy: The optimal positions of the GPS and GSM antennas were on the top edge and the bottom edge of the PCB, respectively. An isolation slot was cut in the ground plane between the antennas to reduce mutual coupling and minimize interference [4], [5].

• Matching Networks: LC matching circuits were employed for both antennas to resonate at desired frequencies. Selection of inductors and capacitors was done with utmost care to tune to maximize return loss and minimize radiation efficiency.

• Verification Methods:

  - S11 Return Loss Measurement: Performed using a Vector Network Analyzer (VNA) to verify both antennas resonated in their respective bands with minimal reflection.
  - GPS Sensitivity Test: Open-sky tests were conducted to validate satellite acquisition and location precision in real-world environmental conditions.
  - GSM Performance: Validated through SMS sending, voice call stability, and low-speed data transmission. Tests were carried out in various

environments to analyze real-world performance and reliability [6].

Figure 3: S11 return loss plot of GPS and GSM antennas

Figure 4: GPS antenna radiation pattern

Figure 5: GSM antenna radiation pattern

Other practical design issues involved PCB thickness, trace width optimization, and component placement for minimal cross-coupling and efficient antenna operation. Small environmental and temperature stabilities were also noted for reproducible real-world operation.

5. Results and Discussion

The practical application yielded encouraging results with effective miniaturization without compromising operational efficiency:

- GSM Antenna Performance: PCB monopole antenna resonated precisely at 900 MHz and 1800 MHz with S11 less than −10 dB, reflecting minimal reflection and good radiation. Efficiency was more than 70% in free-space tests, adequate for voice and data transmission under reliable communication in wearable devices [7], [8].
- GPS Antenna Performance: The chip antenna showed slightly inferior gain compared to conventional ceramic patch antennas but still had sufficient sensitivity to track and capture satellites all the time. Position accuracy was within ±3 meters in clear sky scenarios.
- Mutual Coupling: Orthogonal positioning, ground isolation slots, and filtering approaches were able to suppress near-field coupling, making both GPS and GSM work in tandem without noteworthy degradation of performance.
- Comparative Analysis:
    - Chip Antenna: Smallest footprint, easiest to integrate in small PCBs, moderate gain appropriate for safety-oriented IoT devices.
    - PIFA Antenna: Provides higher gain than chip antennas but a bit larger; appropriate where there is a bit more room.
    - PCB Monopole: Wideband property with good efficiency, appropriate for GSM or multi-frequency applications, but takes up more space than chip antennas [9], [10].
- Application Implications: The integration allows wearable safety devices, smart city asset tracking, industrial IoT installations, and health-monitoring applications. Efficient multiband operation and compact size accommodate portability and unobtrusive wearable form factors.
- Observations and Challenges: External environmental conditions in the form of proximate metallic objects and human body presence resulted in slight detuning, which demonstrates the necessity of AI-aided impedance tuning in future designs. Ground plane design was vital to ensuring stable radiation patterns.

Figure 6: Chip vs PIFA vs PCB monopole efficiency comparison graph

This demonstration experimentally verifies that efficiently designed multiband antennas can simultaneously realize coexistence and efficiency while allowing for considerable miniaturization. It provides a roadmap for future small IoT devices with combined communication and positioning capability.

6. Conclusion and Future Scope

Combining GPS and GSM antennas on one PCB proves that miniaturization, coexistence, and efficiency are possible without the loss of performance. Substrate selection optimization, optimal ground plane design, and LC impedance matching networks made efficient operation in wearable and handheld applications possible. Practical realization proved reliability for safety, industrial, smart city, and healthcare-based IoT applications.

Future Directions

- Real-time impedance tuning using AI: Adaptation of antenna match networks to account for environmental and body proximity effects to maintain consistent performance.
- Dynamic Multi-band Antennas: Dynamic switching between GPS, GSM, Wi-Fi, LoRa, and 5G NR bands for antennas.
- Multi-layer PCB Designs: Placing antennas on multiple layers to further minimize footprint and allow smaller devices.
- Body-proximity Optimized Wearables: Designing against detuning and absorption due to body proximity.
- Environmental Adaptation: Adding adaptive filtering and shielding methods to ensure performance across a variety of operating environments.
- These innovations hold out the promise of the next generation of very small, adaptive, and efficient IoT solutions for applications ranging from personal protection to industrial sensing and smart city infrastructure.

References

[1] C. A. Balanis, Antenna Theory: Analysis and Design, 4th ed., Wiley, 2016.

[2] R. Garg et al., Microstrip Antenna Design Handbook, Artech House, 2001.

[3] H. Wong et al., "Small antennas in wireless communications," Proc. IEEE, vol. 100, no. 7, pp. 2109–2121, 2012.

[4] M. W. A. Khan et al., "Design and analysis of chip antennas for miniaturized GPS receivers," Prog. Electromagn. Res., vol. 82, pp. 463–476, 2018.

[5] U-blox, "NEO-M8 Series Hardware Integration Manual," 2019.

[6] Quectel, "SIM7600 Hardware Design Guide," 2019.

[7] P. J. Soh et al., "Wearable antennas: Design challenges and future directions," IEEE Antennas Propag. Mag., vol. 57, no. 5, pp. 84–96, 2015.

[8] F. Huo et al., "Coexistence strategies of multi-antenna systems for compact IoT devices," IEEE Internet Things J., vol. 6, no. 2, pp. 3210–3220, 2019.

[9] Dassault Systèmes, "CST Microwave Studio – Antenna Simulation Guide," 2020.

[10] Espressif Systems, "ESP32 Hardware Design Guidelines," 2020.

[11] K. Kibaroglu et al., "Machine learning based impedance tuning for small antennas," IEEE Antennas Wireless Propag. Lett., vol. 18, pp. 2189–2193, 2019.

[12] A. Saghati et al., "Reconfigurable RF front-ends for multiband IoT applications," IEEE Trans. Microw. Theory Tech., vol. 65, no. 11, pp. 4593–4603, 2017.

[13] Y. Li et al., "Compact meandered monopole antennas for IoT applications," IET Microwaves, Antennas & Propagation, vol. 11, no. 6, pp. 865–871, 2017.

[14] A. Haraz et al., "Multiband printed slot antenna for 4G/5G smartphone applications," IEEE Antennas Wireless Propag. Lett., vol. 16, pp. 1034–1038, 2017.

[15] M. Abdullah et al., "Compact multi-band antenna design for IoT applications," IEEE Access, vol. 8, pp. 12345–12354, 2020.

[16] E. García et al., "Overview of Reconfigurable Antenna Systems for IoT Devices," Electronics, vol. 13, no. 20, pp. 3988, 2024.

[17] G. Monti et al., "Wearable logo-antenna for GPS–GSM-based tracking system," IET Microwaves, Antennas & Propagation, vol. 10, no. 12, pp. 1322–1329, 2016.

[18] M. A. Rahman et al., "Miniaturized Tri-Band Integrated Microwave and Millimeter-Wave MIMO Antenna," Microelectronics Journal, vol. 144, 105873, 2024.

[19] W. Zheng and H. Li, "Design of High-Performance Antenna System for High-Speed Railways," Prog. Electromagn. Res., vol. 179, pp. 71–81, 2024.

[20] R. R. Elsharkawy et al., "Miniaturized Multi-band Millimeter-Wave Vivaldi Antenna with Performance Optimization at 28 GHz for 5G MIMO Applications," J. Infrared Millim. Terahertz Waves, vol. 45, pp. 208–232, 2024.

Defence R&D Organisation

Empower Panchayat

Dwarka Motors

Mikronix Associates

Elegancespark Innovations

Maharaja Surajmal Institute of Technology

(ISO 9001:2008 Certified, Approved by AICTE)

(Affiliated to GGSIP University)
C-4, Janakpuri, New-Delhi-110058 Phone:011-45037193/45037193
E-mail: director@msit.in Website: www.msit.in