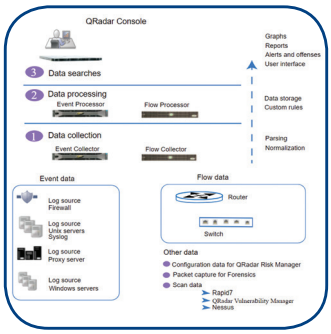
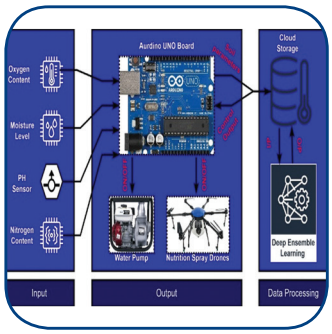
Volume 17, Number 2, 2026 ISSN 1847-6996

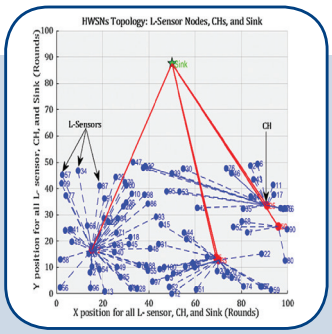
FERIT
FACULTY OF ELECTRICAL ENGINEERING, COMPUTER SCIENCE AND INFORMATION TECHNOLOGY **OSIJEK**

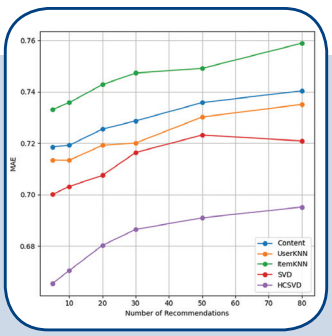
IJECES International Journal of Electrical and Computer Engineering Systems

International Journal of Electrical and Computer Engineering Systems









ISSN: 1847-6996

INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING SYSTEMS

Published by Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Josip Juraj Strossmayer University of Osijek, Croatia

Osijek, Croatia | Volume 17, Number 2, 2026 | Pages 83 - 170

The International Journal of Electrical and Computer Engineering Systems is published with the financial support of the Ministry of Science and Education of the Republic of Croatia

CONTACT
International Journal of Electrical and Computer Engineering Systems (IJECES)

Faculty of Electrical Engineering, Computer Science and Information Technology Osijek, Josip Juraj Strossmayer University of Osijek, Croatia
Kneza Trpimira 2b, 31000 Osijek, Croatia
Phone: +38531224600, Fax: +38531224605
e-mail: ijeces@ferit.hr

Subscription Information
The annual subscription rate is 50€ for individuals, 25€ for students and 150€ for libraries.
Giro account: 2390001 - 1100016777,
Croatian Postal Bank

Journal is referred in:

- Scopus
- Web of Science Core Collection (Emerging Sources Citation Index - ESCI)
- Google Scholar
- CiteFactor
- Genamics
- Hrčak
- Ulrichweb
- Reaxys
- Embase
- Engineering Village

Bibliographic Information
Commenced in 2010.
ISSN: 1847-6996
e-ISSN: 1847-7003
Published: quarterly
Circulation: 300

IJECES online
https://ijeces.ferit.hr

TABLE OF CONTENTS

From Reactive to Proactive: Automating IP Threat Intelligence in SIEM Systems for Cyber Threat Detection

Original Scientific Paper

Abeer Alhuzali *
King Abdulaziz University,
Faculty of Computing and Information Technology, Department of Computer Science
Jeddah, Saudi Arabia
aalhathle@kau.edu.sa

Asrar Alshareef
King Abdulaziz University,
Faculty of Computing and Information Technology, Department of Computer Science
Jeddah, Saudi Arabia
aalshareef0190@stu.kau.edu.sa

*Corresponding author

Abstract *– Digital transformation has provided more opportunities for cybercriminals and exposed organizations to sophisticated threats. Organizations should continuously evaluate their security measures and implement defensive actions to prevent attacks by cybercriminals. Security Information and Event Management (SIEM) systems, deployed within Security Operations Centers (SOCs), allow organizations to identify security risks and vulnerabilities, monitor unusual behavior, and automatically respond to security events. However, SIEM platforms require certain functional enhancements. For instance, security analysts often use external threat intelligence platforms to check suspicious IP addresses manually. This results in longer response times and a greater likelihood of human error. Hence, this paper proposes an integration framework that correlates the functionality of an external threat intelligence platform (AbuseIPDB) with a SIEM system (IBM QRadar) to automatically validate suspicious IP addresses without the need for manual checking. The goal of this integration is to increase the efficiency of threat analysis, incident response, and SIEM-based threat detection. Tests demonstrated that our proposed framework shortens the threat validation time by up to 97.7%, compared to manual processes. Additionally, our system reduces false positives by capitalizing on contextual threat intelligence, thus allowing SOC teams to prioritize critical alerts.*

Keywords: *SIEM, Security Operations Center (SOC), threat intelligence, IP threat, integration*

Received: July 26, 2025; Received in revised form: September 27, 2025; Accepted: September 29, 2025

1. INTRODUCTION

The extensive adoption of technology increases security vulnerabilities because cyberattacks proliferate and organizational IT systems become vulnerable to sophisticated threats. Kuzio *et al.* [1] identified a marked rise in cyberattacks between 2016 and 2023, including ransomware, cyber fraud, and attacks on critical infrastructure. Countries lacking adequate cybersecurity measures were particularly vulnerable. Thus, organizations require advanced security solutions capable of large-scale data analysis and proactive threat detection to prevent data breaches and protect vital assets [2].

Security Information and Event Management (SIEM) systems have emerged as core technologies within Security Operations Centers (SOCs). SIEM platforms are deployed to collect, correlate, and analyze log data from diverse sources. Nonetheless, SIEM platforms face critical limitations, particularly their inability to validate

suspicious IP addresses autonomously. Hence, security analysts have to investigate such indicators manually using external threat intelligence platforms [3]. This reliance on manual processes results in delays and increases the risk of human error.

Several studies [4–7] have sought to solve this problem by introducing automated integration frameworks for SIEM systems and threat intelligence platforms. However, those studies faced several limitations. For example, they overlooked the impact of real-time integration on the performance of SIEM tools. Additionally, only a few studies have addressed the integration of SIEM systems and IP threat intelligence platforms, which are crucial for enhancing the accuracy of the former.

Therefore, this paper addresses the abovementioned limitations by integrating a reliable external threat intelligence platform (AbuseIPDB) with a SIEM system (IBM QRadar). This study demonstrates how such an integration can automate the validation of suspicious IP addresses, improving detection accuracy, alleviating the burden on analysts, and expediting incident response within enterprise environments. Our work makes the following contributions:

- It develops a systematic and modular integration framework combining IBM QRadar and a threat intelligence platform.
- It shows how to automate IP-related threat analysis in QRadar to reduce reliance on manual tools and shorten incident response time.
- It enhances threat detection accuracy using AbuseIPDB to classify malicious IP addresses based on global data.
- It enables the automated generation of reports to improve the speed and precision of decision-making for security analysts.
- It comprehensively evaluates the integration framework to assess its effect on the SIEM.

The remainder of this paper is organized as follows: Section 2 provides background information and Section 3 reviews related work. Section 4 describes the methodology used for the integration. Section 5 details the implementation process and system configuration. Section 6 analyzes the results. Finally, Section 7 concludes the paper and provides recommendations for future research.

2. BACKGROUND

2.1. COMPUTER SECURITY INCIDENT RESPONSE TEAMS

Computer Security Incident Response Teams (CSIRTs) receive, analyze, and respond to security issues affecting data and computer systems. CSIRTs work within organizations, governments, or regions [8]. These teams monitor security events to detect unusual activity that may endanger the information technology (IT) assets of their organizations. They provide reactive services (e.g., incident analysis and response coordination) and proactive services (e.g., vulnerability handling, threat analysis, and cybersecurity information dissemination). They also raise awareness and act as central contact points for incident reporting. Their success relies on four fundamental principles: technical excellence to ensure adequate guidance and solutions, trust to encourage sharing sensitive information, resource efficiency for effective responses, and cooperation with internal and external stakeholders [9]. CSIRTs generally work within SOCs, which are centralized units belonging to IT departments. SOCs continuously monitor, analyze, and respond to security events to detect threats [10].

2.2. SECURITY INFORMATION AND EVENT MANAGEMENT

SIEM systems collect and analyze logs from various sources, including firewalls, intrusion detection and prevention systems (IDS/IPS), and servers [11]. SIEM platforms integrate advanced tools, such as User and Entity Behavior Analytics (UEBA) and machine learning, to improve threat detection and support data-driven decision-making. Using these systems, administrators can define security policies and manage events from multiple sources.

SIEM architecture includes elements for log collection, normalization, analysis, rule-based correlation, storage, and continuous monitoring. Each module can function independently; however, their integration is crucial for optimal system performance [12]. Fig. 1 shows a simplified view of the SIEM log-processing workflow, illustrating how data is normalized and analyzed for monitoring and incident response [10].

Fig. 1. Simplified SIEM log processing workflow: from log collection to monitoring (source: [10]).

2.3. IBM QRADAR

IBM developed QRadar, one of the top SIEM systems, designed to help organizations monitor threats and manage security incidents. QRadar employs machine learning and user behavior analytics to detect unusual activities and enable fast responses [13]. This system gathers and analyzes event data and network flows from various sources, including operating systems, endpoints, and applications. Then, it correlates this information to generate unified alerts that facilitate security investigations. QRadar is a commercial product and thus employs proprietary software and a licensing system that grants IBM full control over the source code. QRadar enhances security and helps organizations fight cyber threats.

Fig. 2 illustrates QRadar's system architecture, highlighting its components for data collection, processing, and analysis [14]. QRadar's first layer collects and normalizes log events and network flows from various sources, converting them into structured data for analysis. The Custom Rule Engine (CRE) layer analyzes event and flow data in real-time; it evaluates rules and building blocks to trigger alerts when security conditions are met. Security analysts use QRadar's GUI to search, filter, and investigate processed data for reporting and offense analysis.

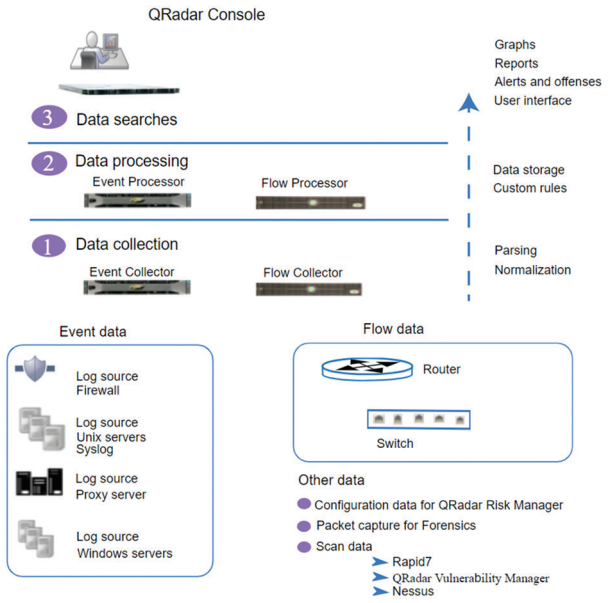


Fig. 2. IBM QRadar system architecture: data collection, processing, and analysis (source: [14]).

2.4. IP THREAT INTELLIGENCE

IP threat intelligence is the use of timely and reliable information on IP addresses associated with malicious activities, such as cyberattacks, system intrusions, or botnet command and control. Thus, IP threat intelligence is key for enhancing the capability of security systems to detect and respond to threats, particularly when integrated with log analysis platforms, such as SIEM systems [15].

2.5. IP-BASED THREAT INTELLIGENCE FEEDS

Threat intelligence feeds are used for collecting Indicators of Compromise (IoCs), including malicious IP addresses. Acquiring accurate and up-to-date information about adversarial behavior enables defenders to refine their security practices and reduce the window of vulnerability, defined as the period during which an organization remains exposed because of a lack of awareness of current attack techniques. Organizations increasingly rely on third-party providers to collect, filter, and curate threat intelligence data, given the complexity of developing such intelligence. The growing market demand in this domain has contributed to notable investments and operational interest [16].

2.6. ABUSEIPDB

AbuseIPDB provides an API that enables users to retrieve detailed information about an IP address, including whether it is blacklisted, its associated geolocation, and the types of threats attached to it. AbuseIPDB is more effective in detecting malicious IP addresses than other public databases; thus, it is recognized as one of the most reliable tools for IP reputation analysis. Lewis *et al.* [17] conducted a comparative evaluation and reported that AbuseIPDB detected 46% of malicious IP addresses, outperforming VirusTotal (13%) and MyIP.ms (16%).

Furthermore, researchers have used AbuseIPDB to investigate IP addresses associated with Advanced Persistent Threat (APT) campaigns, such as *Grizzly Steppe* and *Hidden Cobra*. AbuseIPDB facilitates the collection of rich metadata, such as country codes, activity patterns, and behavioral indicators, and thus enables the precise profiling of malicious infrastructure [18].

These results demonstrate AbuseIPDB's reliability and near real-time capability for assessing IP reputation. The accuracy and accessibility of AbuseIPDB make it a popular choice in academia and enterprise security operations. Fig. 3 shows AbuseIPDB's web interface, which supports IP reporting, historical IP searches, and access to reputation data through the public API.

Fig. 3. AbuseIPDB web interface for IP reputation checking (source: [19])

3. RELATED WORK

Recent studies have emphasized the advantage of integrating SIEM platforms with external threat intelligence sources to improve detection accuracy and reduce false positives [4, 5, 7].

For example, Owen [4] reported that incorporating threat intelligence feeds into SIEM systems augments alert precision and provides security analysts with rich insights through visual dashboards. Owen showed that such a synergistic system enhanced SIEM's capabilities by providing up-to-date information on malicious actors, boosting performance, and addressing data gaps. Similarly, Smeriga [5] explored ways for integrating Cisco Global Threat Alerts [20] with third-party SIEM solutions, emphasizing the need for flexible integra-

tion through efficient APIs. Smeriga also introduced interactive dashboards to help analysts with data interpretation and analysis. Suskalo *et al.* [13] compared IBM QRadar and Wazuh [21], two popular SIEM tools. They found that QRadar exhibits integration flexibility, but requires careful tuning for open-source intelligence feeds, such as AbuseIPDB. By contrast, Wazuh showed robust threat detection and log analysis capabilities, supported by an active user community. The authors presented practical scenarios illustrating how both platforms responded to different attack attempts and evaluated the accuracy of their alerts. Their findings were useful for our selection of IBM QRadar as the SIEM for the current work. Tulcidas [6] proposed using Snort [22], an open-source IDS, and integrating it with Wazuh, an open-source SIEM, in a large-scale academic network. The integration reduced false alarms by enriching events and correlating them with external sources, enabling a faster and more accurate response to internal and external attacks. Tulcidas also compared various SIEM solutions, highlighting their differences in core functionalities, such as aggregation, analysis, and compliance. Sauerwein and Staiger [23] evaluated 13 threat intelligence-sharing platforms, including MISP, OpenCTI, and OTX, using over 50 functional and non-functional criteria that covered aspects including data collection, processing, analysis, dissemination, and integration. However, they noted gaps in data reliability and quality. Although they excluded AbuseIPDB from their analysis, their study provided valuable insights into the factors to consider when selecting a platform. Esseghir *et al.* [7] proposed an open-source platform combining SIEM and IDS functionalities for network monitoring and security alert management. Their proposed system can detect malware in encrypted network traffic using heuristics within a decision tree model. Other studies have investigated the use of machine-learning algorithms to detect cyber-attacks, such as DDoS [24] and other network traffic intrusions [25].

Integrating IBM QRadar with external threat intelligence feeds like AbuseIPDB shows potential for enhancing security monitoring. However, there are no structured methods for combining QRadar with AbuseIPDB. Thus, we propose a novel integration that automates IP reputation analysis to improve response times and accuracy.

4. INTEGRATION METHODOLOGY

Here, the AbuseIPDB reputation platform is integrated with IBM QRadar to automate the analysis of suspicious IP addresses. The proposed system improves threat detection precision and the response to cyber incidents.

4.1. PLATFORM SELECTION RATIONALE

IBM QRadar can collect, analyze, and correlate event logs; thus, it was chosen as the integration platform. Furthermore, QRadar supports the creation of custom correlation rules to generate security alerts. Its flexibility, scalability, and support for threat intelligence make it a popular choice in the cybersecurity industry. AbuseIPDB was selected as the external threat intelligence source because of its extensive database of malicious or suspicious IP addresses and its API, which allows direct reputation searches.

Combining these tools results in an automated workflow that extracts IP addresses from incoming logs, checks their reputation through AbuseIPDB, and returns the enriched data to QRadar. This setup allows QRadar to take suitable security actions based on the classification results.

4.2. SYSTEM ARCHITECTURE

We propose a framework that integrates a malicious IP address reputation service (AbuseIPDB) with a SIEM platform (IBM QRadar). Fig. 4 depicts the framework and its components.

The Integration Control and Management (ICM) module, which oversees the integration process, is at the heart of the system. The ICM handles communication with AbuseIPDB and QRadar. As a result, a final security report is generated for each analyzed IP address. To initiate an analysis, the ICM sends a Syslog message containing the IP address in the Check-IP= xxx.xxx.xxx.xxx format. These messages follow the RFC-5424 standard to ensure QRadar can parse them correctly. The messages are transmitted via the UDP protocol to a predefined port (for example, 5514) on the server running the ICM listening service.

The ICM operates as a background service that continuously monitors the assigned port. When a new message arrives, the ICM uses a regular expression to extract the IP address from the content. The obtained IP is then submitted to AbuseIPDB for a reputation check. AbuseIPDB provides details, such as a confidence score, the country code associated with the IP address, and the date of the most recent reported incident. The confidence score, ranging from 0 to 100, reflects the likelihood that the IP address is linked to malicious activity, based on the number of incident reports.

Then, the ICM uses this data to evaluate the risk level of the IP, which is classified according to its confidence score:

- High-risk: score ≥ 75
- Medium-risk: $20 \leq$ score < 75
- Low-risk: score < 20

This classification is associated with the color-coded ranges in the AbuseIPDB interface. Red indicates high risk, orange denotes medium risk, and yellow represents low risk. This method was validated by testing on real IP data. As a result, IPs with higher confidence scores were consistently associated with multiple abuse reports and malicious behavior. Therefore, we adopted this three-tier classification to enhance the QRadar analysis [19].

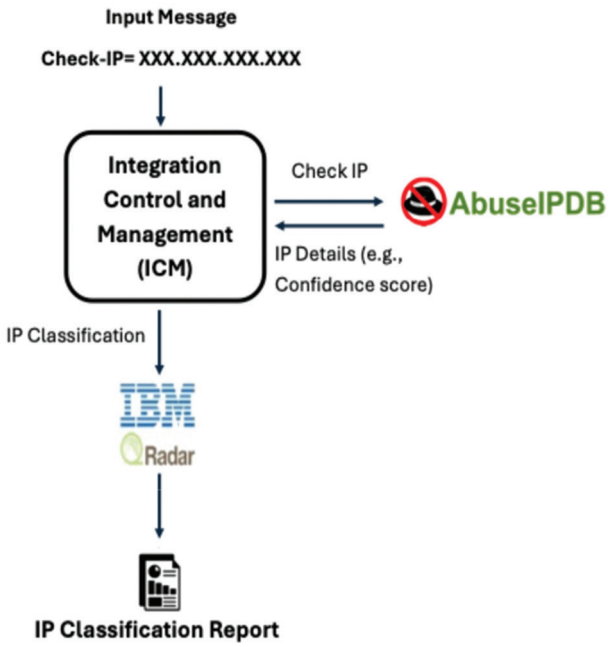


Fig. 4. System architecture for QRadar-AbuseIPDB integration (source: created by the authors)

After the classification process is completed, the ICM re-formats the results into a Syslog-compliant message that includes key data, such as the IP address, country, confidence score, category, and the last reported date. This message is then transmitted to the QRadar platform via UDP, using a custom log source identifier. Next, a custom data source module (DSM) receives and parses the message. If the IP address is classified as high risk or medium risk, the ICM generates a detailed analytical report that lists the IP address, country, confidence score, category, and last reported date, together with analyst-oriented notes tailored to the risk level. The report is automatically stored in a designated network folder, and its file path is inserted into the Syslog message sent to QRadar. This method allows analysts to access the report directly from the event record in *Log Activity*, eliminating the need for manual searching.

5. INTEGRATION IMPLEMENTATION DETAILS

Integration is implemented through the ICM module, which receives request messages, analyzes IP addresses, extracts IP addresses from the messages, assesses their reputation using the AbuseIPDB API, and sends the enriched results back to QRadar through a designated log source. More details are provided in the following sections.

5.1. PHASE ONE: IP MESSAGE CONFIGURATION

The ICM initializes the environment and begins listening for incoming messages. UDP socket port 5514 on the server hosting the ICM's listening service is opened to receive messages from source devices, such as PowerShell scripts running on Windows machines.

5.2. PHASE TWO: PROCESSING SYSLOG MESSAGES AND REPUTATION ASSESSMENT VIA ABUSEIPDB

In this phase, the ICM—a Python code that uses the *socket* library—listens for incoming messages on port 5514. Then, the ICM scans each received message using a regular expression to detect the presence of an IP address; it does this by checking whether the message contains the tag *CHECK-IP*. If the tag is absent, the message is disregarded. Otherwise, the script extracts the IP address. Upon identifying an IP address, the script sends a query to the AbuseIPDB API via an HTTP GET request, using the *requests* library. This request includes a predefined and valid API key. The response, returned in JSON format, provides detailed information about the reputation of the IP address, including the confidence score (a numerical value representing the likelihood that the IP address is malicious), country (a two-letter code indicating the geographic location of the IP address; e.g., US), and the last reported date (the most recent date an abuse report was submitted for this IP). The IP address is classified according to the confidence scores detailed in Section 4.2.

5.3. PHASE THREE: FORMATTING THE ANALYZED DATA AND FORWARDING IT TO QRADAR

Once the IP address is classified according to its risk level, the system generates a message containing the IP address, confidence score, country, last reported date, and final classification (high risk, medium risk, or low risk). This message is built per the structure and formatting requirements of the QRadar platform, following RFC 5424 standards. The message is transmitted to port 514 and received by a designated log source within QRadar for analysis. Then, the message is stored and indexed in the *Log Activity* module.

5.4. PHASE FOUR: CONFIGURING QRADAR TO EXTRACT FIELDS FROM INCOMING MESSAGES

QRadar was configured to extract specific fields from incoming messages by creating custom event properties within the DSM editor. The extraction process relied on precise regular expressions to enable the automatic identification of the following values:

- IP address
- Confidence score
- Country
- Classification
- Last reported date
- Report path

These extracted values were made available within the *Log Activity* module, allowing security analysts to review and analyze each log entry thoroughly.

5.5. PHASE FIVE: ACTIVATING A SECURITY RULE IN QRADAR

After classification, the results are immediately forwarded to QRadar in the form of a Syslog-compliant message. The ICM treats each classification level accordingly. For high-risk IPs, the ICM automatically triggers a security offense via a custom rule. This custom security rule is created within the QRadar platform to automatically trigger an alert upon receiving any message with a *high-risk* classification. This rule is designed to detect high-risk threats without requiring manual intervention. Once triggered, the resulting event is recorded in the *Log Activity* module as a high-priority offense. For *medium-risk* and *low-risk* IPs, the events are logged in the *Log Activity* module without triggering alerts.

5.6. PHASE SIX: FINAL REPORT GENERATION

An analytical PDF report is generated only if the IP address is classified as high risk or medium risk. The report includes key information, such as the IP address, country, confidence score, and last reported date. Furthermore, it provides the security analyst with analytical notes and tailored security recommendations based on the evaluated risk level. Once generated, the report is stored in a predefined shared folder on the network, and the full file path is embedded within the Syslog message sent to QRadar. This mechanism enables analysts to access the report from the event log directly, supports informed decision-making, and contributes to the systematic documentation of analyzed cases. Finally, the ICM returns to its listening state, allowing it to receive and process new messages continuously. The diagram in Fig. 5 depicts the workflow and outlines the steps executed.

6. RESULTS

6.1. EXPERIMENTAL SETUP

The integration framework and experiments were conducted on Oracle VirtualBox with a Red Hat (64-bit) OS, an Intel Core i7-1550H processor of 2.6 GHz, and 8 GB of memory. The integration code was written in Python and used several libraries, including the requests library (detailed in Section 5).

6.2. RESULTS SUMMARY

A series of tests was conducted on 30 randomly selected IP addresses across different confidence score levels from AbusIPDB to evaluate the accuracy of the classification mechanism and the overall effectiveness of the automated response. Table 1 summarizes these selected IPs along with their characteristics. Ten IPs were classified as low risk, 12 as medium risk, and eight as high risk. All of the 30 IPs were correctly classified by our QRadar–AbuseIPDB integration framework. The "Validation Time" column lists the total time required to analyze and categorize each IP automatically. Values ranged from 0.487 seconds (IP# 21) to 18.419 seconds (IP# 17). To compare these results with the manual verification process of each IP, we assumed that manual verification for an IP takes between 30 and 60 seconds. This is reasonable because, in the manual scenario, the analyst must manually verify the IP in AbuseIPDB and then insert the result into the SIEM. For all 30 IPs, the manual process would require between 900 seconds (15 minutes) and 1800 seconds (30 minutes). Hence, our integration framework decreased the threat validation time range by 95%–97.7%.

The integration of QRadar and AbuseIPDB performed well in the real-time analysis of suspicious IP addresses, enabling automated risk-based classification and the generation of security reports to support the incident response team.

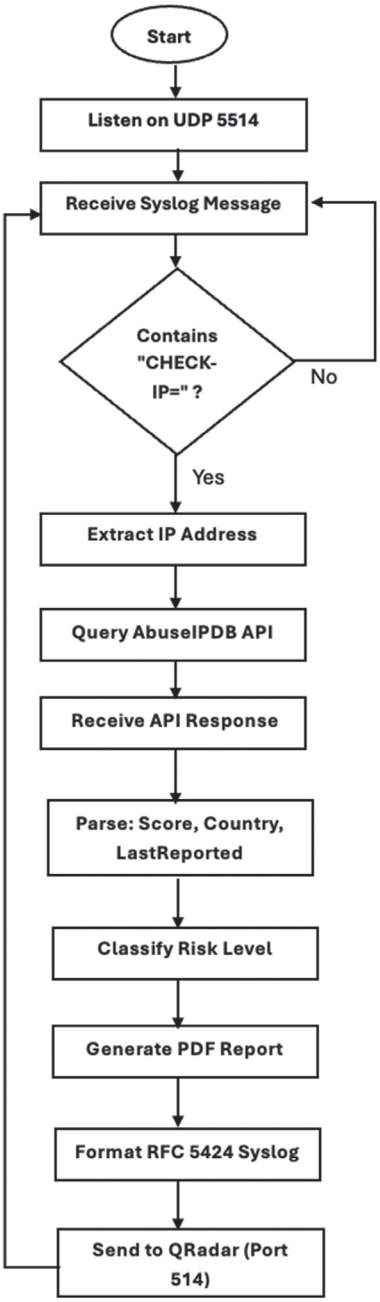


Fig. 5. IP reputation processing workflow (source: created by the authors)

Table 1. IP risk classification results generated by the QRadar–AbuseIPDB integration

#	IP	Score (%)	Country	Classification	Validation Time (sec)
1	209.85.217.42	9	US	Low risk	1.214
2	209.85.167.50	31	FI	Medium risk	0.563
3	216.244.66.245	83	US	High risk	0.690
4	209.85.222.193	64	US	Medium risk	1.615
5	2.57.121.215	54	RO	Medium risk	0.706
6	4.156.21.66	7	US	Low risk	1.145
7	209.85.167.50	31	FI	Medium risk	0.770
8	209.85.222.193	64	US	Medium risk	0.569
9	149.56.160.230	29	CA	Medium risk	0.692
10	209.85.216.66	65	US	Medium risk	0.678
11	185.220.101.26	100	DE	High risk	0.732
12	117.50.47.222	5	CN	Low risk	0.815
13	44.202.169.35	14	US	Low risk	0.568
14	142.4.9.200	18	US	Low risk	0.695
15	185.204.1.182	86	FI	High risk	0.717
16	185.220.101.174	90	DE	High risk	1.268
17	218.92.0.229	100	CN	High risk	18.419
18	77.32.148.7	25	FR	Medium risk	0.637
19	209.85.208.172	26	FI	Medium risk	0.683
20	93.185.162.14	83	ID	High risk	0.601
21	40.107.94.90	2	US	Low risk	0.487
22	103.176.90.16	79	NL	Medium risk	0.723
23	216.239.36.158	1	US	Low risk	0.729
24	209.85.166.230	21	US	Medium risk	0.585
25	88.214.25.62	16	DE	Low risk	1.485
26	146.88.240.123	100	US	High risk	0.593
27	139.59.94.202	15	IN	Low-Risk	0.757
28	216.244.66.236	87	US	High-Risk	0.646
29	209.85.214.200	35	US	Medium-Risk	0.632
30	110.185.37.103	11	CN	Low-Risk	1.065
		Total validation time			**41.479**

6.3. EXAMPLE OF A HIGH-RISK CLASSIFICATION

The IP address *185.204.1.182* was checked by our integration framework, extracted, analyzed, and assigned a high-risk classification (Fig. 6). Therefore, QRadar automatically generated a PDF security report in the specified path (Fig. 7).

The report included the IP address, confidence score, country of origin, last reported date, and tailored mitigation recommendations. All generated reports for high-risk IPs are saved in the "*C:\ThreatReports\High-Risk*" folder, with filenames reflecting the IP address and creation date to facilitate tracking. The full file path of each report was embedded in a Syslog message sent to QRadar via the designated log source. Upon receiving the message, QRadar parsed the content using a custom DSM and extracted the relevant fields. A custom correlation rule within QRadar automatically evaluates the classification field, and if the value is high risk, it triggers an offense of type *Suspicious Activity*.

This offense was logged in the Log Activity module and made visible to analysts, enabling them to access the associated report and take immediate action.

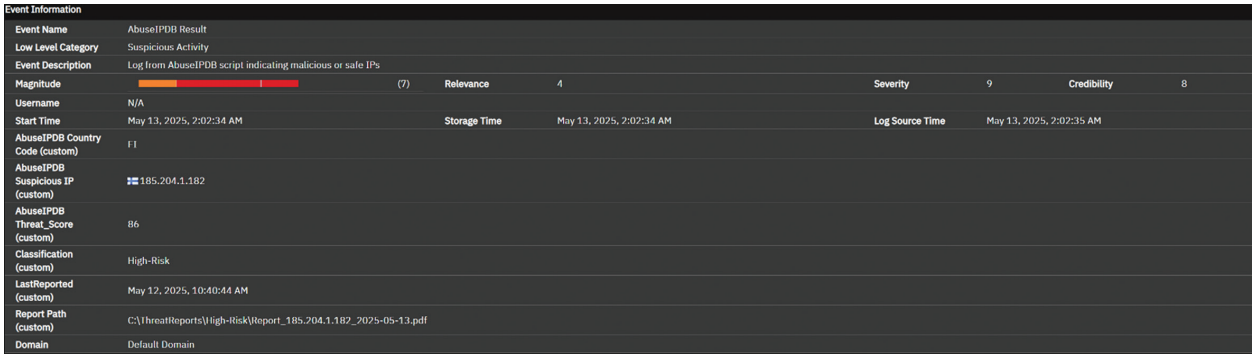


Fig. 6. QRadar's log of high-risk IP 185.204.1.182 (source: created by the authors)

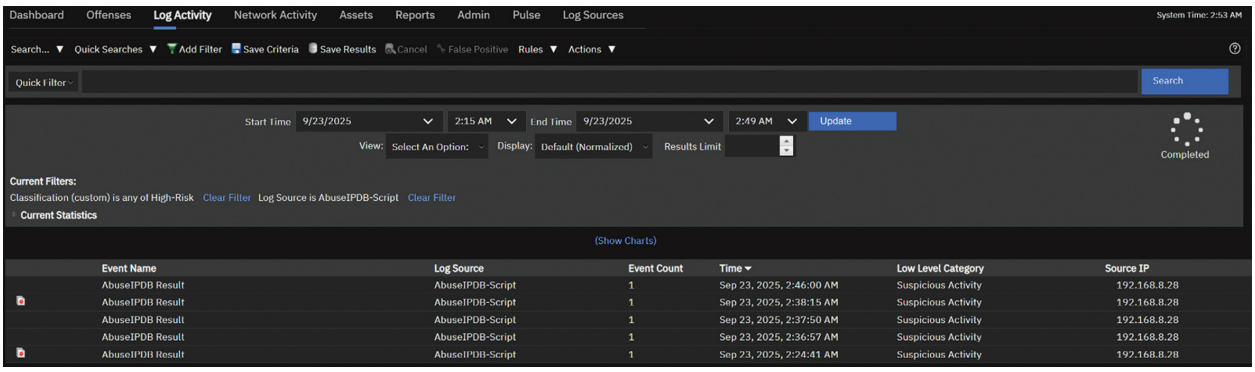


Fig. 7. QRadar's Log Activity module, showing the triggering of a high-risk IP offense (source: created by the authors)

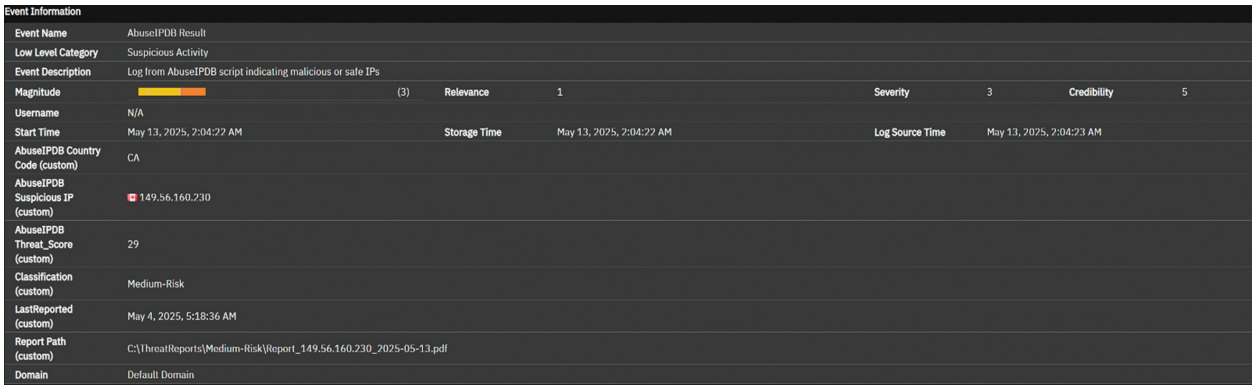


Fig. 8. QRadar's log of medium-risk IP 149.56.160.230 (source: created by the authors)

- **IP address**: *185.204.1.182*
- **Score**: *86%*
- **Country**: *FI*
- **Classification**: *High risk*
- **Last reported**: *May 12, 2025, 10:40:44 AM*
- **Report path**: *C:\ThreatReports\High-Risk\Report_185.204.1.182_2025-05-13.pdf*

6.4. EXAMPLE OF A MEDIUM-RISK CLASSIFICATION

The IP address *149.56.160.230* was classified as medium risk by our tool, based on a confidence score of 29% assigned by AbuseIPDB (Fig. 8). Our system generated a PDF report, which included the IP address, confidence score, country, and the date of the last reported update.

The report was saved in the "*C:\ThreatReports\Medium-Risk*" folder, with a filename including the IP address and creation date. The report also included analytical notes and initial recommendations to help analysts monitor the cases and decide whether escalation is needed if risk levels increase. After the report was generated, our system sent a Syslog message containing the file path and classification result to QRadar. The case was automatically logged in the *Log Activity* module as a medium-severity event without triggering an offense.

- **IP address**: *149.56.160.230*
- **Score**: *29%*
- **Country**: *CA*
- **Classification**: *Medium risk*
- **Last reported**: *May 4, 2025, 5:18:36 AM*
- **Report path**: *C:\ThreatReports\Medium-Risk\Report_149.56.160.230_2025-05-13.pdf*

6.5. EXAMPLE OF A LOW-RISK CLASSIFICATION

IP address *117.50.47.222* was correctly classified as low risk by our system, based on a confidence score of 5% assigned by AbuseIPDB (Fig. 9). The ICM in our tool sent the classification result to QRadar and logged it in the *Log Activity* module as a low-risk event without triggering an offense. Note that no reports are generated for low-risk IPs. Even though this IP address was classified as low risk, it continues to be systematically monitored because of the possibility of score escalation in future reports. This approach ensures timely reclassification and an appropriate response in the event of malicious activity, reflecting a preventive security strategy aimed at minimizing potential threats at an early stage.

- **IP address**: *117.50.47.222*
- **Score**: *5%*
- **Country**: *CN*
- **Classification**: *Low risk*
- **Last reported**: *May 12, 2025, 7:02:18 AM*

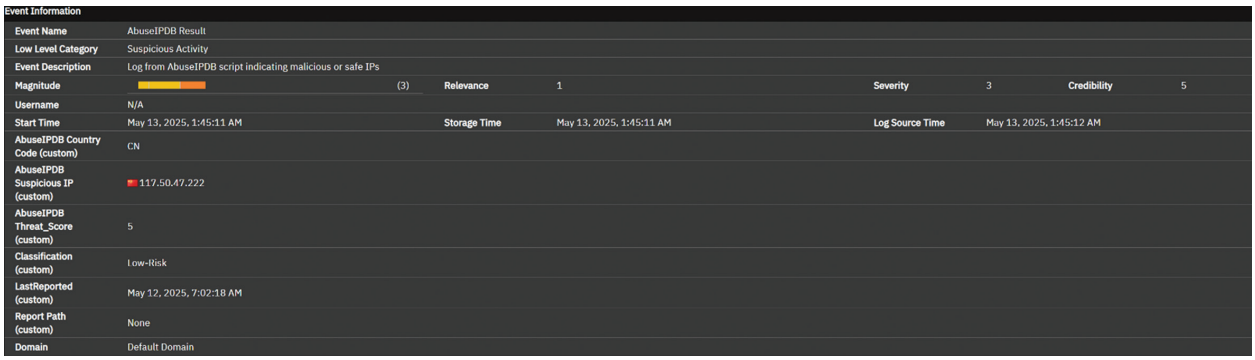


Fig. 9. QRadar's log of low-risk IP 117.50.47.222 (source: created by the authors)

7. CONCLUSION AND FUTURE DIRECTIONS

This paper demonstrates the feasibility and effectiveness of integrating the AbuseIPDB threat intelligence platform with IBM QRadar to enhance the automated analysis of threats and the response to incidents. Our proposed solution embeds a real-time verification mechanism within the SIEM framework, streamlining the detection, classification, and documentation of suspicious IP addresses. The integration is implemented using the ICM, which listens for Syslog messages, extracts IP addresses, and checks their reputation via the AbuseIPDB API. This approach allows the automated classification of threats, avoiding a reliance on manual processing and reducing false positives. Tests confirmed that the integration effectively identifies high-risk IP addresses and enriches event logs with trusted reputation data. The system also generates analytical reports that help security analysts make decisions. Furthermore, this study demonstrates QRadar's flexibility in gathering and analyzing structured data, making it particularly suitable for dynamic security environments that need real-time responsiveness and detailed documentation.

In conclusion, this paper presents a practical model for enhancing SIEM platforms by integrating external threat intelligence. The proposed approach improves organizational readiness against evolving cyber threats. However, our proposed integration approach focuses on analyzing IP addresses and does not incorporate other threat indicators, such as domain names or malware signatures. Additionally, the system relies solely on data from AbuseIPDB, but its effectiveness could be enhanced by integrating additional threat intelligence feeds. Future research directions include (1) expanding threat precision by incorporating threat indicators, such as domain names and file hashes, in addition to IP addresses, to enhance analytical depth, (2) integrating multiple threat intelligence sources, including platforms like VirusTotal and IBM X-Force Exchange, to enable multi-source correlation and improve classification accuracy, (3) incorporating User and Entity Behavior Analytics (UEBA) capabilities into the system to facilitate the detection of anomalous behaviors and advanced persistent threats, (4) enhancing the alerting mechanism within the IBM QRadar platform to support multilevel alert generation, based on the confidence scores and the temporal frequency of the incident reports, (5) assessing integration performance in large-scale environments, such as governmental or financial institutions, to evaluate robustness under high-volume data conditions, and (6) applying machine-learning techniques to improve threat classification accuracy and reduce false positives through advanced behavioral and technical analysis.

8. REFERENCES:

[1] A. Kuzior, I. Tiutiunyk, A. Zielińska, R. Kelemen, "Cybersecurity and Cybercrime: Current Trends and Threats", Journal of International Studies, Vol. 17, No. 2, 2024, pp. 220-239.

[2] A. Tariq, J. Manzoor, M. A. Aziz, Z. U. A. Tariq, A. Masood, "Open Source SIEM Solutions for an Enterprise", Information & Computer Security, Vol. 31, No. 1, 2022, pp. 88-107.

[3] D. Sim, H. Guo, L. Zhou, "A SIEM and Multiple Analysis Software Integrated Malware Detection Approach", Proceedings of the IEEE International Conference on Service Operations and Logistics, and Informatics, Singapore, 11-13 December 2023, pp. 1-7.

[4] T. Owen, "Threat Intelligence & SIEM", Lewis University, Master Thesis, 2014.

[5] J. Smeriga, "Integration of Cisco Global Threat Alert to 3rd Party Product", Masaryk University, Faculty of Informatics, Brno, Master Thesis, 2022.

[6] N. R. Tulcidas, "Event Correlation in Ciências", Universidade de Lisboa, Faculdade de Ciências, Departamento de Informática, Lisbon, Master Thesis, 2024.

[7] A. Esseghir, F. Kamoun, O. Hraiech, "AKER: An Open-Source Security Platform Integrating IDS and SIEM Functions with Encrypted Traffic Analyt-

ic Capability", Journal of Cyber Security Technology, Vol. 6, No. 1-2, 2022, pp. 27-64.

[8] S. K. Bhatt, P. K. Manadhata, L. Zomlot, "The Operational Role of Security Information and Event Management Systems", IEEE Security & Privacy Magazine, Vol. 12, No. 5, 2014, pp. 35-44.

[9] M. Bada, S. Creese, M. Goldsmith, C. Mitchell, E. Phillips, "Computer Security Incident Response Teams (CSIRTs): An Overview", Global Cyber Security Capacity Centre, University of Oxford, Technical Report, 2014.

[10] G. González-Granadillo, S. González-Zarzosa, R. Diaz, "Security Information and Event Management (SIEM): Analysis, Trends, and Usage in Critical Infrastructures", Sensors, Vol. 21, No. 14, 2021, p. 4759.

[11] O. Podzins, A. Romanovs, "Why SIEM is Irreplaceable in a Secure IT Environment?", Proceedings of the IEEE eStream Conference, Riga, Latvia, 25 April 2019, pp. 1-5.

[12] M. Sheeraz, M. A. Paracha, M. U. Haque, M. H. Durad, S. M. Mohsin, S. S. Band, A. Mosavi, "Effective Security Monitoring Using Efficient SIEM Architecture", Human-centric Computing and Information Sciences, Vol. 13, No. 17, 2023, pp. 1-18.

[13] D. Šuškalo, Z. Morić, J. Redžepagić, D. Regvart, "Comparative Analysis of IBM QRadar and Wazuh for Security Information and Event Management", Proceedings of the 34th DAAAM International Symposium on Intelligent Manufacturing and Automation, Vienna, Austria, 2023, pp. 96-102.

[14] M. Seppänen, "Methods for Managed Deployment of User Behavior Analytics to SIEM Product", JAMK University of Applied Sciences, Degree Programme in Information and Communications Technology, Jyväskylä, Finland, Bachelor Thesis, 2021.

[15] H. Griffioen, T. M. Booij, C. Doerr, "Quality Evaluation of Cyber Threat Intelligence Feeds", Proceedings of the 18th International Conference on Applied Cryptography and Network Security, Rome, Italy, 19-22 October 2020, pp. 251-270.

[16] V. G. Li, M. Dunn, P. Pearce, D. McCoy, G. M. Voelker, S. Savage, K. Levchenko, "Reading the Tea Leaves: A Comparative Analysis of Threat Intelligence", Proceedings of the 28th USENIX Security Symposium, Santa Clara, CA, USA, 2019, pp. 739-756.

[17] J. L. Lewis, G. F. Tambaliuc, H. S. Narman, W.-S. Yoo, "IP Reputation Analysis of Public Databases and Machine Learning Techniques", Proceedings of the International Conference on Computing, Networking and Communications, Big Island, HI, USA, 17-20 February 2020, pp. 181-186.

[18] R. Ando, H. Itoh, "Characterizing Combatants of State-Sponsored APT in Digital Warfare by Reported Blocklist Database", International Journal of Computer Science and Network Security, Vol. 22, No. 3, 2022, pp. 541-546.

[19] AbuseIPDB, "IP Address Abuse Checker", www.abuseipdb.com (accessed: 2025)

[20] Cisco Systems, "Cisco Global Threat Alerts", www.cisco.com/security/alerts (accessed: 2025)

[21] Wazuh, "Wazuh: Open Source Security Platform", wazuh.com (accessed: 2025)

[22] M. Roesch and the Snort Team, "Snort: An Open-Source Network Intrusion Detection and Prevention System", www.snort.org (accessed: 2025)

[23] C. Sauerwein, T. Staiger, "Cyber Threat Intelligence Sharing Platforms: A Comprehensive Analysis of Software Vendors and Research Perspectives", University of Innsbruck, Department of Information Systems, Production and Logistics Management & Department of Computer Science, Innsbruck, Austria, Master Thesis, 2021.

[24] T. Hussein, "Deep Learning-based DDoS Detection in Network Traffic Data", International Journal of Electrical and Computer Engineering Systems, Vol. 15, No. 5, 2024, pp. 407-414.

[25] R. Singh, R. Ujjwal, "Intrusion Detection System based on Chaotic Opposition for IoT Network", International Journal of Electrical and Computer Engineering Systems, Vol. 15, No. 2, 2024, pp. 121-136.

Enhanced Crop Yield through IoT-Based Soil Monitoring and Machine Learning Analysis for Rice and Sugarcane Cultivation

Original Scientific Paper

Deepthi Gorijavolu*
Department of Computer Science and Engineering,
Amity School of Engineering and Technology,
Amity University Madhya Pradesh, Gwalior, India
deepthigorijavolu555@gmail.com

Kapil Sharma
Department of Computer Science and Engineering,
Amity School of Engineering and Technology,
Amity University Madhya Pradesh, Gwalior, India
ksharma@gwa.amity.edu

N. Srinivasa Rao
Indian Council of Agricultural Research (ICAR),
Hyderabad, India
ns.rao@icar.gov.in

*Corresponding author

Abstract *– Agriculture, a cornerstone of global economies, faces persistent challenges in efficient crop monitoring. This study introduces a groundbreaking IoT-based framework, integrated with a novel Deep Ensemble Learning (DEL) technique. The cuurent study objective is to enhance rice and sugarcane yield through monitoring soil parameters precisely. The framework employs an array of sensors, including moisture and pH sensors, to determine key soil properties: moisture content, pH level, Nutrient Retention Capability (NRC), and oxygen content. These parameters are crucial in assessing nutrient availability, Organic Carbon Content (OCC), soil texture, and root health. Data captured by sensors is transmitted via an Arduino kit to the cloud, where it undergoes analysis by advanced deep learning models, namely Bidirectional Long Short-Term Memory (Bi-LSTM). The ensemble of models ensures high accuracy in predicting soil parameter. The farmers acquires the processed data through a mobile application that offers actionable insights and facilitating real-time, automated agricultural interventions. Empirical results from field trials demonstrate a significant enhancement in soil parameter detection and monitoring accuracy.The application enables the IoT and DEL-based system in rice and sugarcane fields that enhances the crop yeild by 97% compared to traditional schemes. The study demonstrates the potential of integrating IoT and machine learning in agriculture paradigm shift towards the precision farming, and sets a new standard for sustainable, efficient agricultural practices.*

Keywords*: Precision Agriculture, IoT in Farming, Deep Ensemble Learning, Soil Parameter Monitoring, and Crop Yield Optimization*

Received: July 14, 2025; Received in revised form: September 28, 2025; Accepted: September 29, 2025

AI	Artificial intelligence
Bi-LSTM	Bidirectional Long Short-Term Memory
CSV	Comma-Separated Values
DEL	Deep Ensemble Learning
DL	Deep Learning
GFRC	Global Report on the Food Crisis
GRBF	Generalized Radial Basis Function
GRU	Gated Recurrent Unit
HAR	Human Activity Recognition
HTTP	Hypertext Transfer Protocol
IIS	Intelligent Irrigation System
IoT	Internet of Things
ISNPHC	Integrated Soil Nutrient Prediction and Health Classification
LSTM	Long Short-Term Memory
ML	Machine Learning
NPK	nutrients
NRC	nutrient retention capacity
PPV	Positive Prediction Value
ReLU	Rectified Linear Unit
RNN	Recurrent Neural Network
TAN	Tree Augmented Naïve Bayes
TL	Transfer Learning
TPR	True Positive Rate

1. INTRODUCTION

Agriculture, vital to India's economy, predominantly relying on traditional practices. It serves not only the cornerstone of food production but also as a primary source of income for a large portion of Indian population. The Global Report on the Food Crisis (GFRC) mid-year update of 2023 underscores the alarming state of the global food crisis, emphasising the urgent need for innovation in agricultural practices [1], [2]. Persistent conflicts, economic downturns, and extreme weather events continue to exacerbate global hunger and malnutrition [3]. Conventional agriculture faces several challenges, including heavy reliance on pesticides, high resource consumption, and labor-intensive processes. A significant concern is the inadequate financial returns for farmers. Furthermore, optimal plant growth requires meticulous routines and daily monitoring. Each crop species requires specific management, typically involving manual watering and nutrient application, which is inefficient and laborious [4].

The Technological advancements are gradually addressing limitations through Smart farming, Internet of Things (IoT), artificial intelligence (AI), and robotic machinery to enhance agricultural productivity. However, these innovations can be costly and require specialised knowledge for effective implementation [5, 6]. Precision farming, a critical sector in the agricultural, heavily relies on the integration and transfer of information technology. In this context, IoT plays a vital role by facilitating the transmission of data to farmers[7-9].

The sensor technology has gained prominence in agriculture for real-time data collection, with applications extending to healthcare, military, and telecommunications. In farming, sensors are deployed to monitor soil and environmental conditions that are crucial for crop growth. Soil quality: encompassing soil types and ecosystem characteristics is vital for sustainable plant growth. However, accurately assessing soil quality is complex, and necessitating advanced automatic techniques [10]. The effective farming hinges on a robust system for monitoring soil characteristics. IoT devices are well-suited for this purpose, enhancing various aspects agricultural management [11].

The Key factors for soil condition surveillance include soil temperature and moisture content. Proper balance of these factors aids in determining optimal irrigation schedules. Although watering is not directly correlated with other soil parameters like pH, vitamins, minerals, and salinity levels, these factors remain significant for soil classification [12]. The current research focuses on two major Indian crops, paddy and sugarcane where enhancing yield depends on the effective management of soil parameters. Traditional schemes for monitoring soil parameters have limitations,motivation the adoption of deep ensemble learning for soil parameter classification in this study.

The primary objective is this research is to improve crop growth by optimizing irrigation watering practices and applying nutrients in acordance with real-time field conditions. Thus, to address these challenges, an IoT-based framework is proposed for crop growth monitoring, employing soil parameter analysis. The main contributions of this study are as follows:

- Deployment of soil sensors for crop growth monitoring,and addressing power constraint challenges.
- Extraction of complex soil features using novel Deep Learning (DL) models such as Gated Recurrent Units(GRU) and Bidirectional Long Short Term Memory(Bi-LSTM) networks.
- Adoption of the Bi-LSTM model to effectively capture both past and future temporal dependencies in agricultural time series data.This bidirectional context modelling facilitates accurate interpretation of the dynamic patterns in crop growth, weather varioations, and soil conditions.
- Enhancement of classification performance through an ensemble learning technique that integrating features derived from multiple DL models [13-15].

2. RELATED WORK

The intergation of IoT and Machine Learning (ML) Technologies into agriculture has gained considerable attention in a recent research, with various studies demonstrationg their potential to enhance crop monitoring and yield.

Sharma et al. developed a smart irrigation system for rice cultivation using IoT,and Intelligent Irrigation System (IIS). The system employs soil sensors to continuously monitor soil conditions in rice fields. The sensors data transmitted wirelessly to a web-based database. The databse processes the soil information to determine optimal watering levels and subsequently controls water nozzles through HTTP protocols.The collected data ate stored on a central server and visualized through a interactive dashboard. The key feature of system is an ability to remotely control water pumps based on parameters like soil moisture and flow rate, thereby showcasing the practical applicability of IoT in precision irrigation. [3, 16]

Similarly, Bhushan et al. [16] addressed the challenge of user interfaces in agricultural IoT devices dy designing a low-power remote communication module. This work emphasizes the transition from wired to wireless systems, highlighting to towards more flexible and user-friendly IoT solutions for agriculture. The study furhter anticipates substantial growth in agriculture productivity by 2024 using IoT and wireless sensor networks. This vision is rooted in the ability of IoT to effectively manage soil quality, crop temperature requirements, and irrigation practices [17].

Sahu et al.[3] extended the integrating of IoT in agriculture by incorporating ML for comprehensive crop monitoring. In their approach, wireless sensors collect field data, which is subsequently transformed into CSV format for ML-based processing. The study emphasises the practical deployment of ML modules in agriculture environments and demonstrates how soil parameters and environmental conditions affect plant growth. The application of real-time data within ML models offers valuable insights into climate prediction and the optimization of farming practices, ultimately conserving time, and resources while mitigating crop losses for farmers [18].

Vijayalakshmi et al. explored the application of supervised ML algorithms to classify and map crops based on soil types. The study highlights the efficacy of combining IoT and ML, particularly through ensemble techniques, to achieve precise crop type selection. Thereby, contributing to enhanced agricultural yield [19]. In related study, Afzaal et al. [20] applied various supervised learning methods to improve potato production in Canada's Atlantic region. While Nishant et al. [21] emphasised the importance of improved stacked extrapolation approaches for generating more accurate crop yield forecasts. The collective studies demonstrate the potential growth of deep learning models to furhter improve accuracy and efficiency in agricultural decison making.

Senapaty et al. [22] delved on the analysis of soil nutrients through IoT enabled framework for precision farming. Their framework encompasses multiple stages including data acquision through IoT sensors, preserving real-time data on cloud platforms, accessing data through an Android application, data preparation and subsequent analysis leveraging diverse computational techniques [22]. This approach not only contributes to enhanced crop yeild but also reduces dependency on chemical fertilizers,reinforcing the critical role of IoT in optimising agricultural productivity.

3. METHODOLOGY

This section provides an overview of smart crop growth monitoring with a focus on soil parameters. In this framework a variety of sensors are deployed in paddy and sugarcane fields to measure key soil parameters such as pH, moisture, temperature, dissolved oxygen content, nitrogen content levels, nutrient retention capacity (NRC), and nutrient availability. The collected sensor data is transmitted to the cloud via an Arduino based interface. Once stored in the cloud, the data is processed and analysed using a Deep Ensemble Learning technique as illustrated in the Fig 1.

The Bi-LSTM network is leverged to extract deeper temporal features from the sensor data. An ensemble learning approach is applied by integrating the features derived from the both models to generate a high-quality feature vector. Thector is subsequently employed for classification,enabling the system to determine whether the filed requires irirgation/water or nutrient suppementation. Based on the classification results, the cloud platform communicates with the Arduino kit, which in turn activates the motor for irrigation ON/OFF control and dispatches signal to a drone for precision nutrient application. The architecture of the proposed crop growth monitoring systeme based on soil parameters is depicted in Fig. 2.

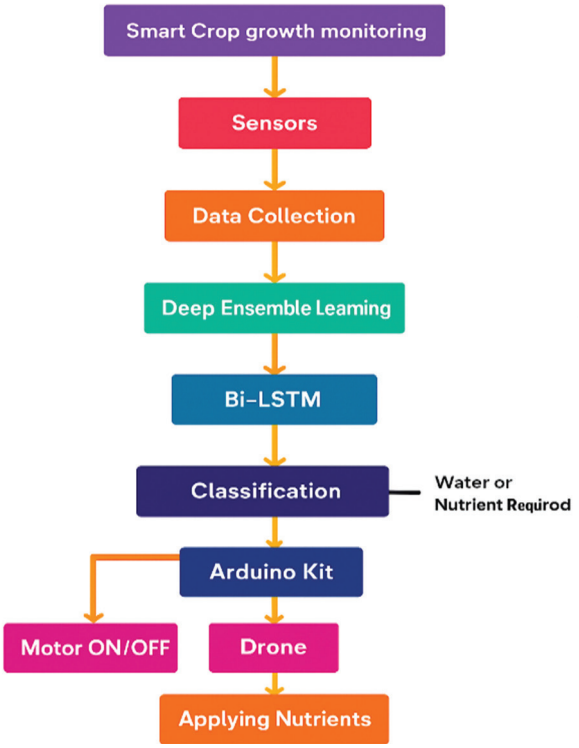


Fig.1. Flow diagram for proposed scheme

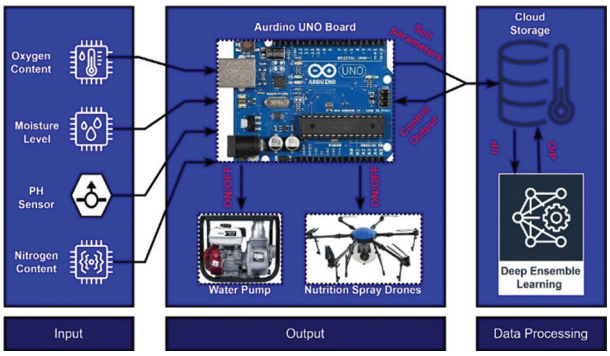


Fig. 2. Overview of the Crop Growth Monitoring System Architecture, Highlighting Soil Parameter-Based Sensing, Data Processing, and Automated Response Mechanisms

A. Data Acquisition Using Soil Sensors

Real-time data are collected from rice and sugarcane fields through the deployemnt of soil sensors. The type of sensors employed for data acquisition are summarized in Table 1 and Table 2. Specifically, an electrochemical sensor is employed to measure the pH value and nutrient content in both paddy and sugarcane fields.

Table 1. Sensors and its Measurements in Paddy Crop Field Monitoring

Soil Parameters	Sensor Used	Ideal Level	Alert Level	Actions/Reasons
PH	PH Sensor	6.0 to 6.7	<6.0 & >6.7	Adjust nutrients if outside ideal range.
Moisture Level	Moisture Sensor	Device reads 270	Device reads 435	Water plants if moisture is high
Oxygen Content	Oxygen Meter	0 to 40%	-	Overwatering lowers soil oxygen, avoid

Table 2. Sensors and its Measurements in Sugarcane Crop Field Monitoring

Soil Parameters	Sensor Used	Ideal Level	Alert Level	Actions/ Reasons
PH	Electrochemical Sensor	5.5 to 6.5	<5.5 & >6.5	Adjust nutrients if outside the ideal range
Moisture Level	Moisture Sensor	80 to 85% (early stage), 50 to 65% (ripening stage)	-	Keep appropriate moisture according to growth stage
Nitrogen Content	NPK Sensor	-	-	Fertilize in a 3:1:2 ratio for healthy crops
Oxygen Content	Soil Oxygen Meter	0.0 to 0.3	0.3 to 0.7	Act if oxygen affects rooting

Soil moisture sensors are employed to measure field humidity and water levels. The oxygen sensors capture dissolved oxygen content in the soil, an essential parameter for effective plant rooting. The soil acidity and alkalinity are determined leveraging a pH sensor, which directly influences microbial activity and micronutrient availability. The pH scale ranges from 0 to 14, with 7 denoting neutrality. The values lower than 5.5 indicate high acidity, values between 5.5 and 6.5 indicte mild acidity, values between 6.5 and 7.5 represent neutrality, values beyond 7.5 reflect slightly alkaline, and values above 8.5 indicate high alkalinity. In addition, the Light Dependent Resistor (LDR) based soil color sensing scheme is leveraged to determine RGB color values which serve as an indirect indicator of soil quality.

The data collection carried out multiple paddy and sugarcane fields in consultation with knowledgeable farmers and agricultural specialists. The topographical map of the region is referenced to account for water availability ensuring somprehensive field coverage. A GPS sensor, integrated with an Arduino UNO board, is employed to determine the latitude and longitude of the field locations enabling spatial tagging of soil data. The informtion combined with sensor measurements is uploaded to cloud storage for further processing.

The communication infrastructure includes a wireless networking module connected to the Arduino board enabling the TCP-enabled internet connectivity facilitating Wi-Fi. The various sensors such as pH,m oisure,electrochemical,temperature and NPK are connected to the Arduino microcontroller board with data aquisituion acheieved through programmed control. The sensor configurations used for data collection are illustrated in the input unit of Fig.1. Once data is acquired and preserved in cloud.Further, the collected data undergoes analysis and classification leveraging an ensemble of Transfer learning (TL) techniquies. This analytical framework facilitates precise assessment of water and nutrient requirements for paddy and sugarcane crops. According to the experiemtnal results actution are triggered automatically wherein drones are deployed for targeted nutrient application and water pumps are controlled for optimized irirgation.

B. Deep Ensemble Learning Method

In this study, a deep ensemble learning method is employed to enhance crop growth monitoring performance. The architecture integrates Bi-LSTM and GRU models. The Bi-LSTM is employed to analyse temporal parameter dependencies in soil data while GRU model focuses on capturing critical soil parameters such as nutrients nutrient levels,pH and moisture content. By combine the predicitive capabilities of both modles the ensemble approach generates high-quality outputs that support precise decision-making for crop growth management.The detailed analysis and netwoek architecture are described in the subsequent subsection.

The soil sensor data uploaded to the cloud undergoes processing throguh a deep learning network. The LSTM based architecture is adopted to address the vanishing gradient commonly encountered in backpropagation. For soil parameter analysis, a multilayer architecture is implemented to ensure robust prediction.The Bi-LSTM model consit of n layers, where the input data are processed sequentially across multiple time steps in both forward and backward directions,thereby capturing past and future contextula dependencies.

The LSTM network operates through three primary gates: the input, forget, and output gate that regulate information flow. Additionally, a candidate gate regulates cell state updates. Their roles are mathematically expressed as in eq. 1 to 4.

$$it=\sigma(W_{t1}ht+W_{t2}ht\text{-}1+B_i) \quad (1)$$

$$ft=\sigma(W_{t1}ht+W_{t2}ht\text{-}1+B_f) \quad (2)$$

$$ot=\sigma(W_{t1}ht+W_{t2}ht\text{-}1+B_0) \quad (3)$$

$$kt=\sigma(W_{t1}ht+W_{t2}ht\text{-}1+B_k) \quad (4)$$

Where, it ,ft , ot , and kt represent the input gate, forget gate, output gate, and candidate gate, respectively. W_{t1} and W_{t2} are the weight parameters of the successive cell in layer n respectively. B_i, B_f, B_0, and B_k are denotes the bias parameters of the input gate, forget gate, output gate, and candidate gate, respectively. The cell state of the LSTM network is defined as follows:

$$Ct=ft.Ct\text{-}1+it.kt \quad (5)$$

In this formulation, the weight parameter and bias parameter of each cell are distributed across all layers of the LSTM network. The Hadamard element-wise operations together with the sigmoid function and the hyperbolic tangent (tanh) activation function regulate information flow and reduce the number of hidden neurons and weight parameters. In this work, the Bi-LSTM processes soil parameters, such as nutrients, pH, moisture, and humidity in both forward and backward directions. The concatenation of forward and backward hidden states are estimated and the outcome is fed into successive layers:

$$h_{t_{fb}}=h_{t_f}.h_{t_b} \quad (6)$$

The Bi-LSTM demonstrates improved capability in modeling sequential dependencies through bidirectional processing and it also requires less momory for problem-solving makes this mechanism more efficient sharing compared to conventional deep learnign approaches. Further, to complement Bi- LSTM,the GRU is adopted as the secondary network within ensemble. While RNNs are effective in handling sequential data through hidden state updates.However, RNN suffering from gradient instability over long sequences.

Therefore, GRU simplify the architecture by introducing two gates: reset adn update that regualte memory flow across time steps,thereby capturing both long and short-term depenencies with reduced complexity. The memory in the RNN network is maintained through a hidden state, calculated using the formula:

$$ht=fl(w_1\, h_{t-1}+w_2\, x_t+B) \quad (7)$$

Where, ht is the hidden state of the RNN, x_t is the input data, w_1 and w_2 are the weight parameters of the hidden nodes, and B is the bias parameter. fl is a nonlinear activation function. The current state pt is estimated as:

$$pt=w_p h_t+B_p \quad (8)$$

Bi-LSTM is adept at modelling complex time patterns but is limited by the problems of vanishing and exploding gradients, and its accuracy decreases over longer time durations. Thus, this network was proposed to address these issues, but its extensive training process can be a limitation for real-time applications. In our research, we employ the GRU, a variation of the recurrent neural network.

Both RNN and GRU feature chain-based self-looping units, but GRU's units are more complex. GRU has two gates: update and reset, which regulate the flow of soil data. These gates map soil parameters in the range [0,1], where the number represents the proportion of memory retained. Thus, GRU can handle both long-term and short-term dependencies in time-series data.

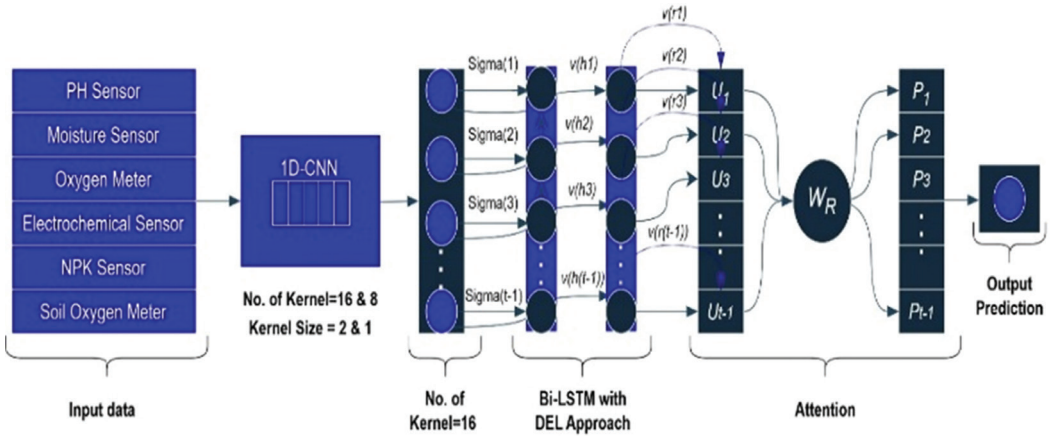


Fig 4. Bi-LSTM structure adapted for DEL method

The GRU's two novel gates, reset and update, are mathematically expressed as follows: The reset gate R_t controls the data transferred from the previous hidden state to the current hidden state:

$$R_t=\sigma(\omega_R\,[h_{t-1}, x_t]+B_R) \quad (9)$$

The memory state m_t is defined using the reset gate and a hyperbolic activation function:

$$mt=tanh(\omega.(R_t.h_{t-1}, x_t)) \quad (10)$$

The update gate U_t localizes which hidden states need updating:

$$U_t=\sigma(\omega_U\,[ht_{-1}, x_t\,]+B_U) \quad (11)$$

Finally, the hidden state link is generated using both the reset and update gates:

$$h_t=(1-U_t)\, h_{t-1}+U_t\, m_t \quad (12)$$

The output layer of the network generates outputs based on the final hidden state. Depending on the specific task, this output could be a single number, an array of values, or a probability distribution among soil parameter classifications.

The prediction classification outputs of the Bi-LSTM technique and the GRU technique are integrated to enhance soil parameter analysis for better crop growth in cultivation fields. The advantage of using the deep ensemble technique lies in its ability to leverage expertise from multiple classification systems, creating a more robust and effective deep learning model. In this research, the outcomes of two distinct DL models are combined to form a multilayered deep ensemble learning model. As the output of each model corresponds

to a single node, these nodes are each connected to a single neuron, activated by the Softmax function using a 3-dimensional vector. A batch normalisation layer is included to improve the precision of the output. Each batch received is processed by this layer, which normalises it using its specific average and standard deviation, then rescales the data using two trainable parameters. Essentially, batch normalisation adjusts its inputs in a coordinated manner. The ReLU activation function is employed in this activation layer to enhance the training speed of the network. The dropout layer serves as a regularisation method that, during training, randomly deactivates a predetermined proportion of neurons within the network. This prevents overfitting and encourages the development of robust and sparse features. It utilises the average and standard deviation of each batch to normalise unit values. This approach can accelerate optimisation by scaling components to a similar scale, irrespective of the network's depth. To further prevent overfitting, it randomly eliminates a set ratio of units from the neural network during training.

4. RESULTS AND DISCUSSION

The soil parameters listed in Tables 1 and 2 are collected from paddy and sugarcane cultivation fields to effectively monitor crop growth. The collected data is pre-processed before being fed into two novel deep learning techniques: the Bi-LSTM network and the GRU network. These dual networks perform a deeper feature analysis of the input data, generating individual classification outputs. To integrate the prediction results of both deep learning networks by employing the Deep Ensemble technique. The classification results of the soil parameter analysis are obtained as single-node outputs from the dropout layer ensuring robustenssnd reduced overfitting.

A. Performance Analysis

Table 3 and Table 4 present the classification outputs of NPK achieved by the proposed DEL technique across various categories. These tables display the maximum prediction accuracy achieved through DEL model. The DEL technique effectively classified soil nutrients into four categories: organic carbon, nitrogen, phosphorus, and potassium for both sugarcane and paddy cultivate fields.

In the paddy field, the classification of the four soil nutrient classes achieved a *Positive Prediction Value (PPV) of 0.8912* and a *True Positive Rate (TPR) of 0.9132*. Additionally, the model attained an overall *accuracy and F1 - of 0.9365 and 0.9736, respectively* that indicates a reliable prediction performance for nutrient classification.

In the sugarcane field, the four nutrient classes are classified with a *PPV of 0.8712 and a TPR of 0.9232*. Furthermore, the accuracy and F1-score for soil nutrient classification in the sugarcane field are approximately *0.9765 and 0.9636, respectively*. The sugarcane cultivate results demonstrates the higher accuracy comapred to paddy field classification with consistently strong precision and recall balance.

Table 3. TPR and PPV content of soil nutrients in the paddy field

Concentration	PPV	TPR	Accuracy	F1 Score
Low	0.865	1	0.986	0.9325
Medium	0.8795	0.832	0.906	0.9625
High	0.8965	0.9012	0.9023	0.9726
Average	0.8912	0.9132	0.9365	0.9736

Table 4. TPR and PPV content of soil nutrients in sugarcane

Concentration	PPV	TPR	Accuracy	F1 Score
Low	0.855	0.5881	0.956	0.9425
Medium	0.8695	0.822	0.966	0.9525
High	0.8765	0.9312	0.9723	0.9626
Average	0.8712	0.9232	0.9765	0.9636

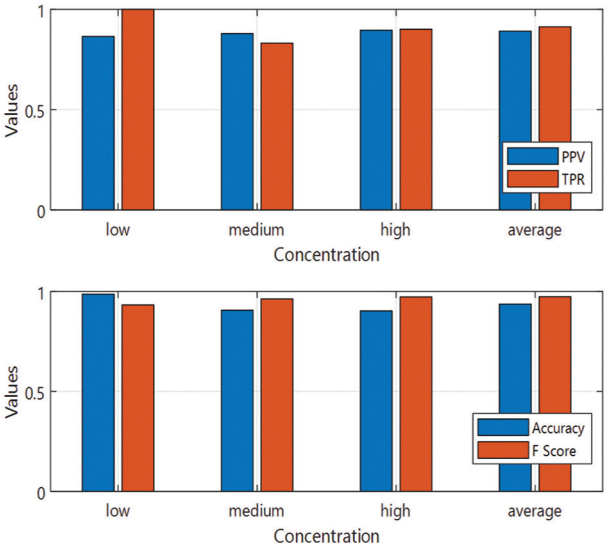


Fig. 4. TPR and PPV content of soil nutrient in paddy field

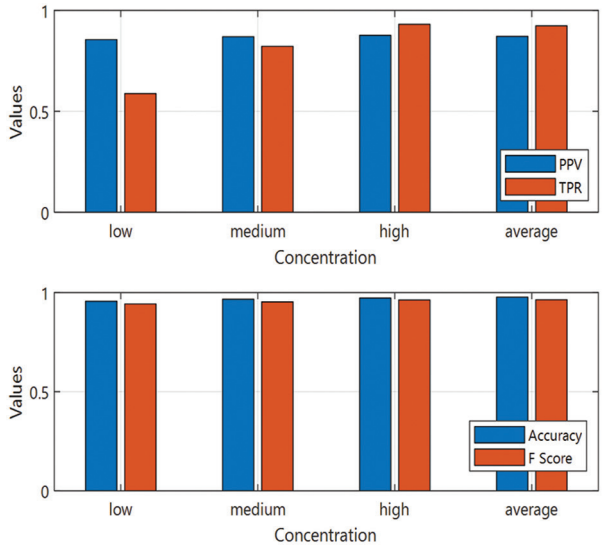


Fig. 5. TPR and PPV content of soil nutrient in sugarcane field

Figs. 4 and 5 illustrate the graphical representations of accuracy and F1-score for soil nutrient classifica-

tion in both paddy and sugarcane fields. The proposed technique effectively classifies soil nutrients, soil pH, and soil moisture across both paddy and sugarcane fields.

In the paddy field, the classification of organic carbon achieved an accuracy and F1-score of approximately 0.9303 and 0.956, respectively. In the sugarcane field, the corresponding values for organic carbon are approximately 0.9405 and 0.9685, respectively. The results are presented in Table 5 which emphasizes the classifiaction outcomes for soil pararmetes across the paddy and sugarcane fields.

Table 5. Soil Parameter Classification Under Different Classes

Soil Parameters	Paddy		Sugarcane	
	Accuracy	F1-Score	Accuracy	F1-Score
Organic Carbon	0.9303	0.956	0.9405	0.9685
Nitrogen	0.9325	0.9611	0.9625	0.9645
Phosphorous	0.9456	0.9405	0.9524	0.9654
Pottasium	0.9085	0.9125	0.9025	0.9125
Soil PH	0.8725	0.8735	0.8751	0.8574
Moisture	0.8565	0.8567	0.8125	0.8525

For nitrogen classification, the F1-score in both paddy and sugarcane fields is approximately 0.9611 and 0.9645, respectively. In the paddy field, the accuracy of DEL for classifying nitrogen, potassium, and phosphorus concentrations is approximately 0.9456, 0.9025, and 0.8725, respectively. In the sugarcane field, the accuracy and F1-score for soil pH classification are approximately 0.8725 and 0.8735, respectively. Additionally, soil moisture classification achieves an accuracy of approximately 0.8565 in the paddy field and approximately 0.8525 in the sugarcane field. Thus, the performance metrices are illustrated in Fig. 5.

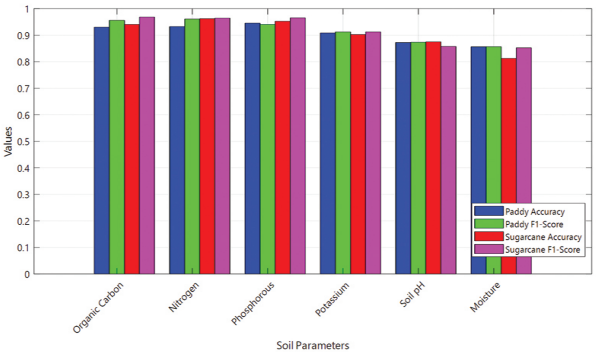


Fig. 6. Soil Parameter Classification under Different Classes

The comparative analysis of the prediction classification results presented in Table 6. The performance of the proposed DEL technique in the context of soil parameter classification. The results build upon the insights discussed earlier in this study. Noteably,the DEL stands out as a remarkable achievement, reaching an impressive 97.5%. which indicates significantly surpasses the performance of traditional DL techniques. This high level of accuracy is critical for precision agriculture where accurate assessment of soil nutrient levels and environmental conditions directly impacts crop health and yeild.

The study also includes four benchmarking techniques for comparison, providing a clear perspective on the state-of-the-art performance of DEL. The results, graphically represented in Fig.6, clearly demonstrate DEL's superiority robustness and reliability over the traditional DL approaches. These findings reinforce the potential of DEL framework as an effective and powerful tool for agricultural research enabling enhanced crop management and decision-making.

Table 6. Comparative Analysis of Soil Parameter Classification Accuracy

Techniques	Nitrogen	Phosphorous	Potassium	Ph	Moisture
ISNPHC	0.9503	0.9412	0.9029	0.8281	0.8752
HAR	0.732	0.8627	0.7584	0.8692	0.8546
GRBF	0.8369	0.9	0.7854	0.8865	0.8185
TAN	0.8125	0.8896	0.7854	0.8695	0.8524
DEL [Proposed]	0.9925	0.9456	0.9085	0.8725	0.8565

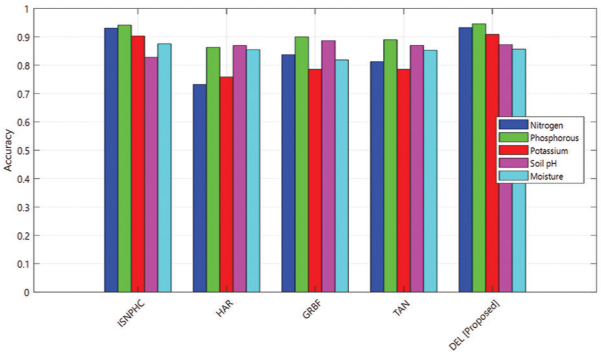


Fig. 7. Comparative Analysis of Soil Parameter Classification Accuracy

A notable aspect of DEL's success is its ability to handle multiple soil parameters including organic carbon, nitrogen, phosphorus, potassium, soil pH, and moisture. The traditional DL models often struggle with multi-parameter classification due to the complexity of the soil data. In contrast, the DEL's deep ensemble approach effectively address these challenges as evidenced by its outstanding performance across all evaluated soil parameters. Furthermore, the results indicate that DEL maintains consistent predictive across different crop fields, including paddy and sugarcane. This versatility suggests that the technique can be applied to a wide range of agricultural contexts, offering valuable insights into soil nutrient levels regardless of the specific crop being cultivated.

5. CONCLUSION

In this research study, a novel DEL technique is designed to classify soil parameters obtained from soil

sensors into six distinct categories: soil pH, moisture, nitrogen, phosphorus, potassium, and organic carbon. The DEL technique employs two advanced deep learning models: Bi-LSTM and GRU, to enhance prediction accuracy. To assess the effectiveness of the DEL method, a comprehensive simulation study is conducted using a standardised dataset. The results demonstrate the superior performance of the DEL technique compared to conventional DL methods. Especially, the accuracy values achieved for soil nutrient classification in sugarcane and paddy fields are remarkable, reaching 0.9765 and 0.9365, respectively. The results signifies a substantial improvement in soil parameter prediction.

The demonstrated capabilities of the DEL technique have the potential to rimplication fro real-time agricultureal operations. By leveraging automation this approach can streamline agricultural practices, optimising crop management and resource allocation. However, certain soil parameters like potassium, pH, and moisture, exhibited moderate accuracy that indicating areas for improvement. Future research can focus on refining feature selection and optimising DL architectures and testing the scalability of the DEL technique on larger adn more diverse datasets. Overall, this study emhasizes the potential of DEL as a cutting edge solution for precise soil parameter classification with strong applicability in modern precision agriculture. The outstanding accuracy achieved in our experiments underscores its relevance and promise for improving decision-making in crop management.

6. REFERENCES

[1] G. Kaur, "Food Security in India: A Comprehensive Examination of Progress, Challenges, and Pathways to Sustainable Development", International Journal For Multidisciplinary Research, Vol. 7, 2025.

[2] B. Philip, G. A. Mathew, R. T. Sebastian, A. A. Thomas, "From Farm to Future: Charting India's Agricultural Path to Global Competitiveness and SDGs Alignment", Current Agriculture Research Journal, Vol. 12, No. 3, 2025, pp. 1239-1248.

[3] C. D. R. Sahu, A. I. Mukadam, S. D. Das, S. Das, "Integration of Machine Learning and IoT System for Monitoring Different Parameters and Optimising farming", Proceedings of the International Conference on Intelligent Technologies, Hubli, India, 25-27 June 2021, pp. 1-5.

[4] D. Rani, N. Kumar, B. Bhushan, "Implementation of an Automated Irrigation System for Agriculture Monitoring using IoT Communication", Proceedings of the 5th International Conference on Signal Processing, Computing and Control, Solan, India, 10-12 October 2019, pp. 138-143.

[5] Rishab R, Prajwal M, Somashekara G, Santhrupth H R, "The Economic Impacts of Ai-driven Agriculture on Small-scale Farmers in Emerging Economics", International Journal For Multidisciplinary Research, Vol. 7, 2025.

[6] K. Elhattab, S. Elatar, "Evaluating low-cost internet of things and artificial intelligence in agriculture", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 37, No. 2, 2025, pp. 968-975.

[7] A. Vangala, A. K. Das, N. Kumar, M. Alazab, "Smart Secure Sensing for IoT-Based Agriculture: Blockchain Perspective", IEEE Sensors Journal, Vol. 21, No. 16, 2021, pp. 17591-17607.

[8] K. Sekaran, M. N. Meqdad, P. Kumar, S. Rajan, S. Kadry, "Smart agriculture management system using internet of things", TELKOMNIKA -Telecommunication, Computing, Electrical & Electronics, and Instrumentation & Control, Vol. 18, No. 3, 2020, pp. 1275-1284.

[9] Dinesh P. M, Sabeenian R. S, Lokeshvar R. G, Paramasivam M. E, Thanish S, Manjunathan A, "IOT Based Smart Farming Application", E3S Web Conferences, Vol. 399, 2023, p. 04012.

[10] D. Saikia, R. Khatoon, "Smart monitoring of soil parameters based on IoT", International Journal of Advanced Technology and Engineering Exploration, Vol. 9, No. 88, 2022.

[11] G. Zhang, X. Li, "Estimate Cotton Water Consumption from Shallow Groundwater under Different Irrigation Schedules", Agronomy, Vol. 12, No. 1, 2022, p. 213.

[12] R. Singh, S. Srivastava, R. Mishra, "AI and IoT-Based Monitoring System for Increasing the Yield in Crop Production", Proceedings of the International Conference on Electrical and Electronics Engineering, Gorakhpur, India, 14-15 February 2020, pp. 301-305.

[13] Md. B. Rahman et al. "Smart Crop Cultivation System Using Automated Agriculture Monitoring Environment in the Context of Bangladesh Agriculture", Sensors, Vol. 23, No. 20, 2023, p. 8472.

[14] G. S. Nagaraja, K. Vanishree, F. Azam, "Novel Framework for Secure Data Aggregation in Precision Agriculture with Extensive Energy Efficiency", Journal of Computer Networks and Communications, Vol. 2023, No. 1, 2023, p. 5926294.

[15] B. B. Sharma, N. Kumar, "IoT-Based Intelligent Irrigation System for Paddy Crop Using an Internet-Controlled Water Pump", https://www.igi-global.com/article/iot-based-intelligent-irrigation-system-for-paddy-crop-using-an-internet-controlled-water-pump/www.igi-global.com/article/iot-based-intelligent-irrigation-system-for-paddy-crop-using-an-internet-controlled-water-pump/273708 (accessed: 2025)

[16] M. W. Rasooli, B. Bhushan, N. Kumar, "Applicability Of Wireless Sensor Networks & Iot In Saffron & Wheat Crops: A Smart Agriculture Perspective", International Journal of Scientific & Technology Research, Vol. 9, No. 2, 2020, pp. 2456-2461.

[17] R. Vijayalakshmi, M. Thangamani, M. Ganthimathi, M. Ranjitha, P. Malarkodi, "An automatic procedure for crop mapping using agricultural monitoring", Journal of Physics: Conference Series, Vol. 1950, No. 1, 2021, p. 012053.

[18] F. Abbas, H. Afzaal, A. A. Farooque, S. Tang, "Crop Yield Prediction through Proximal Sensing and Machine Learning Algorithms", Agronomy, Vol. 10, No. 7, 2020, p. 1046.

[19] P. S. Nishant, P. Sai Venkat, B. L. Avinash, B. Jabber, "Crop Yield Prediction based on Indian Agriculture using Machine Learning", Proceedings of the International Conference for Emerging Technology, Belgaum, India, 5-7 June 2020, pp. 1-4.

[20] M. K. Senapaty, A. Ray, N. Padhy, "IoT-Enabled Soil Nutrient Analysis and Crop Recommendation Model for Precision Agriculture", Computers, Vol. 12, No. 3, 2023, p. 61.

[21] Md. D. Hossain, M. A. Kashem, S. Mustary, "IoT-Based Smart Soil Fertiliser Monitoring And ML-Based Crop Recommendation System", Proceedings of the International Conference on Electrical, Computer and Communication Engineering, Chittagong, Bangladesh, 23-25 February 2023, pp. 1-6.

[22] P. Agarwal, D. Gorijavolu, G. H. Sastry, V. Marriboyina, D. V. Babu, G. K. Kishore, "Real-time crop field monitoring system using agriculture IoT systems", International Journal of Nanotechnology, Vol. 20, 2023, pp. 586-599.

A Novel Approach for Diabetes Mellitus Detection Using a Modified Binary Multi-Neighbourhood Artificial Bee Colony Algorithm with Mahalanobis-Based Feature Selection (MBMNABC-Ma) and an Optimized Decision Forest Framework

Original Scientific Paper

Gaurav Pradhan *
Department of Computer Applications,
Sikkim Manipal Institute of Technology,
Sikkim Manipal University (SMU),
Majitar, India
gaurav.p@smit.smu.edu.in

Gopal Thapa
Department of Computer Applications,
Sikkim Manipal Institute of Technology,
Sikkim Manipal University (SMU),
Majitar, India
gopal.t@smit.smu.edu.in

*Corresponding author

Ratika Pradhan
Department of Computer Applications,
Sikkim University,
Gangtok , India
rpradhan01@cus.ac.in

Bidita Khandelwal
Department of General Medicine,
Sikkim Manipal Institute of Medical Sciences,
Sikkim Manipal University (SMU),
Tadong , India
bidita.k@smims.smu.edu.in

Abstract *– Diabetes is a critical global health issue caused by high blood sugar (hyperglycemia), leading to complications like cardiovascular disease, blindness, neuropathy, and kidney failure. Machine learning (ML) algorithms improve both the accuracy and efficiency of medical diagnoses. This study applies a Modified Binary Multi-Neighbourhood Artificial Bee Colony with Mahalanobis-based (MBMNABC-Ma) for a feature selection algorithm, combined with diverse ML models for diabetes identification. Compared to the conventional Binary Multi-Neighbourhood Artificial Bee Colony (BMNABC), MBMNABC-Ma improves classification accuracy and reduces computational complexity. Five diabetes datasets were analyzed using a 70-30% holdout cross-validation. The MBMNABC-Ma model, trained on Optimal Decision Forest (ODF) with Random Forest Ensemble (RFE), demonstrated high effectiveness. It achieved 97.23% accuracy on the Merged Datasets (comprising 130 US and PIMA datasets), 97.93% on the Iranian Ministry of Health Dataset, 96.05% on the Questionnaire Dataset, 98.39% on the Hospital of Sylhet Dataset, and 80.98% on the PIMA Dataset, with high specificity and sensitivity scores across all cases.*

Keywords: *Machine Learning, Feature Selection, Biomedical Data Analysis, Ensemble Learning, Diabetes Detection, Data Mining*

Received: July 8, 2025; Received in revised form: August 9, 2025; Accepted: August 19, 2025

1. INTRODUCTION

Diabetes is a common endocrine disease, defined by increased blood glucose levels caused by defects in the production or action of insulin or both [1]. The increase in diabetes cases has created a severe risk to the global healthcare system. An individual suffering from diabetes can have consequences like cardiovascular disease, blindness, renal failure, and neuropathy [2]. Hence, diabetes mellitus must be identified and treated as soon as possible to reduce the complications. In the meantime, prominent trends in diabetes-related early death were found through a study. Between 2000 and 2010, the death rate fell in high-income countries [3].

There was a further rise in death rates between 2010 to 2016 due to diabetes. In low-income countries, the rates of premature death continued to rise [4]. In India, over 77 million people suffer from diabetes, according to forecasts published in 2019 [5, 6]. Likewise, by 2045, there would be close to 134 million cases of diabetes in India, according to these estimates [7]. Diagnosing diabetes and its related diseases is important and cannot be exaggerated at an early age. Therefore, this offers a good opportunity for early intervention and better treatment tactics, which lowers the risk of serious issues, including heart disease and nerve damage [8]. Data Mining techniques and machine learning have become essential tools in detecting various diseases, specifically in the detection of diabetes. It aids with prediction, diagnosis, and complication management [9]. The strength and efficiency of machine learning come from its ability to find patterns that human experts might overlook, improving the diagnostic accuracy [10]. In order to create a successful ML model, feature selection helps in dimensionality reduction, model generalization, and computing efficiency [11, 12]. This study proposes a Modified Binary Multi-Neighbourhood Artificial Bee Colony algorithm with a Mahalanobis-based (MBMNABC-Ma) feature selection algorithm for the detection of diabetes. The study attempted to compare the performance against traditional algorithms by testing it on five diabetes datasets within an Optimal Decision Forest (ODF) framework. We evaluate the proposed model against other methods, including MBMNABC-Ma + k-NN, MBMNABC-Ma NB, MBMNABC-Ma + C4.5, MBMNABC-Ma + RS, and MBMNABC-Ma + SVM, using performance metrics like accuracy, specificity, and sensitivity. Results show that MBMNABC-Ma + ODF(RFE) outperforms most models in terms of accuracy and effectiveness.

1.1. LITERATURE REVIEW

Significant research work has been carried out on choosing features, imputation strategies, and managing values that are missing in the past; this section discusses and examines a few of the appropriate studies [10].

A. Negi et al. [11] used the UCI machine learning library to gather two of the datasets [13][14], those were then combined based on their similarities. Illustrations of unknown or missing data (unidentified data) were replaced with the value of 0. Some of the non-numeric entries were changed into numerical equivalents, and the features that were not relevant to the identification of diabetes were eliminated. A script that was included in LibSVM was employed to prepare the data, and the data points were normalized using a scale ranging from 0 to 1. The mean value was used to replace the amalgamated data points, and numerical values were assigned to symbolic representations. The split of 60%-40% of the dataset were considered for training and testing respectively. In the process of selecting relevant attributes, the Weka software platform was used to create the F-select script of the LibSVM package. A Support Vector Machine (SVM) was employed to create a predictive model. For validation, the training data were used through a 10-fold cross-validatio. Then, Feature selection was performed using both wrapper and ranker methods, resulting in 71% accuracy with the wrapper technique and 72% with the ranker technique. Moreover, the LibSVM F-select script was then employed, yielding 63% accuracy. Finally, by selecting all features, an accuracy of 72.92% was obtained. Heydari et al. [15] utilized a set of data consisting of 2536 occurrences from the University of Medical Sciences, Tabriz, to predict the presence of diabetes. Various methods, such as ANN, SVM, nearest neighbor, Bayesian network, and Decision Tree, were contrasted to identify the most effective procedure for diabetes diagnosis. Among all, the best accuracy was performed by ANN with 97.44%. On the other hand, 5-NN, Decision Tree, SVM, and Bayesian Network reached accuracy percentages of 81.19%, 95.03%, 90.85%, and 91.60%, respectively. Prerna et. al. [12] used online and offline surveys to generate a dataset containing eighteen inquiries about family history, lifestyle, and health. RStudio and the R programming language were implemented for analysis employing classifiers like k-Nearest Neighbor, Support Vector Machine, Decision Tree, Naïve Bayes, Logistic Regression, and Random Forest. The Random Forest produced the maximum accuracy of 94.10% and the dataset was split into 75% for the training and 25% for testing. Islam et al. [16] utilized machine learning methods, including Decision Trees, Naïve Bayes, Logistic Regression, and Random Forest for diabetes detection. 520 instances of the dataset were collected by direct investigations from Sylhet Diabetes Hospital. Using Cross-validation and an 80-20 split, Random Forest has achieved the finest accuracy with 97.4%. Dzulkalnine et al. [17] applied fuzzy principal factor analysis for feature selection from the PIMA dataset. In an 80-20 data split, the maximum accuracy of 72.078% was achieved using FPCA-SVM classification. A better hybrid imputation FPCA-SVM-FCM model was created, and accuracy measures revealed that FCM showed FCM surpassed SVM-FCM. Oladimeji et al. [18] obtained and used a set of data from Sylhet Diabetes Hospital, balancing with the SMOTE oversampling method. Abedini et al. [19] employed a PIMA dataset to propose an ensemble, hierarchical model. The models were individually trained and then integrated at an advanced level. At first, Decision Tree and Logistic Regression model were utilized, and next, feeding their results into a Neural Network for improved accuracy, an accuracy rate of 83% was obtained. Iyer et al. [20] uses Pima Indian Diabetes sets of data to forecast diabetes in women using Decision Tree (C4.5) and Naïve Bayes classifiers. Using the cross-validation and percentage split procedures, the data sets were divided into preparation and test sets. Moreover, 10-fold cross-validation was used to prepare the data with an achieved accuracy of 79.56%. Chang et al. [21] introduced a classification model suitable for elec-

tronic diagnostic systems. The three models: the C4.5 decision tree, Naïve Bayes, and Random Forest classifier were used on the diabetes datasets of Pima Indians. Position ranking, k-means clustering, and principal component analysis (PCA) were employed in the dataset analysis. Several matrices, such as accuracy, sensitivity, precision, F-score, specificity, and AUC (area under the curve), were used to evaluate the model's performance. On the entire dataset, Random Forest outperformed Naive Bayes and C4.5 decision trees, achieving 79.57% accuracy, 89.40% precision, 75.00% specificity, 85.17% f-score, and 86.24% AUC. Out of the three models, C4.5 achieved the highest sensitivity, at 88.43%.

Overall, research shows that ensemble-based classifiers and efficient feature selection methods are essential for accurate diabetes prediction. But the majority of current methods are either less accurate, have a lot of computing overhead, or aren't adequately generalizable to a variety of datasets. To improve the accuracy and efficiency of the detection of diabetes, Modified Binary Multi-Neighbourhood Artificial Bee Colony algorithm with Mahalanobis-based distance (MBMNABC-Ma), with Optimal Decision Forest and Random Forest Ensemble (ODF-RFE) has been proposed.

The paper is planned as follows: *Section 2* details the MBMNABC-Ma feature selection algorithm and the classification techniques used; *Section 3* presents the experimental results and comparisons with existing models; and *Section 4* concludes the study.

2. MATERIALS AND METHODOLOGY

This study aims to build an accurate and efficient system to detect diabetes by using advanced methods like model creation and feature selection. The approach uses the Random Forest Ensemble (RFE) along with the Optimized Decision Forest (ODF) for developing the model, and the MBMNABC-Ma algorithm is used to choose the most relevant features.

2.1. COLLECTION OF DATA

Various datasets were collected for validating the proposed model. The primary dataset, PIMA, and the secondary dataset, Diabetes 130-US hospitals, were provided by Negi et al.[11] from the UCI machine learning repository. These datasets span from 1999 to 2008 and were merged into one, with 102,536 participants—64,419 healthy, 38,115 unhealthy. The dataset's age range was 5 to 95 years, consisting of 47,055 males and 55,480 females. After filtering missing data, 5,000 instances were used. Another dataset from Heydari et al.[15] contained 2,209 individuals tested for type 2 diabetes in Tabriz, Iran, including 698 males and 1,837 females, ages 30 to 90, with 15 missing entries. A third dataset, provided by Neha Prerna et al. [12], included 952 participants aged 40-60, with 580 males and 372 females, of which 266 were diabetic and 685 non-diabetic. They also used the PIMA dataset. M. M. F. Islam et al.[16] introduced a dataset with 502 individuals (186 non-diabetic and 315 diabetic), aged 16 to 90, from the Sylhet Diabetes Hospital in Bangladesh. Finally, Kaggle [13] provided a dataset with 768 records, consisting of 500 non-diabetic and 268 diabetic cases, all female participants, aged 21 to 81. The study utilized binary classification datasets, each labeled with two classes: diabetic (1) and non-diabetic (0). These labels were originally included in the datasets and served as the ground truth during both training and evaluation. Table 1 summarizes these datasets.

Table 1. Datasets Used

Name of Dataset	Number of features	Instances	Classes
Merged Dataset (130 US and PIMA) [11]	46	5000	2
Iranian Ministry of Health [15]	19	2536	2
Questionnaire Dataset [12]	17	952	2
Hospital of Sylhet Dataset [16]	16	502	2
PIMA Dataset [13]	8	768	2

2.2. FEATURE SELECTION USING THE MODIFIED BINARY MULTI-NEIGHBORHOOD ARTIFICIAL BEE COLONY -MAHALANOBIS BASED (MBMNABC-MA) METHOD

The MBMNABC-Ma technique assessed the distance between neighbors *i* and *k* across all datasets using Mahalanobis distance rather than the traditional Euclidean or Hamming distance. The rationale behind adopting MBMNABC-Ma over existing feature selection techniques is its enhanced ability to identify optimal features that maximize classification accuracy while simultaneously minimizing the number of selected features. The classical BMNABC algorithm consists of the following stages: initialization of food sources for the bees; the employed bee phase, where each bee explores a new candidate solution based on neighborhood information; the onlooker bee phase, where bees select promising solutions based on the fitness of neighbors; and the scout bee phase, where new random solutions are introduced when stagnation is detected. In traditional BMNABC, the distance between the *ith* bee and its neighbors was calculated using Hamming distance, which may not adequately capture the relationships in datasets with continuous and correlated features. In contrast, the proposed MBMNABC-Ma algorithm employs Mahalanobis distance, which considers the covariance among features, providing a more reliable measure of dissimilarity between solutions. Moreover, the fitness evaluation process is improved by incorporating a multi-objective function that balances classification accuracy assessed through k-fold cross-validation and the number of selected features. Memetic Algorithms are hybrid optimization techniques that combine global search from algorithms like ABC with local refinement strategies to improve both convergence speed

and solution quality. In the MBMNABC-Ma method, the memetic part improves the top-performing solutions by locally adding or removing features to increase fitness. This hybrid structure helps the algorithm not just explore different regions of the search space globally, but also focus more closely on the best areas through local fine-tuning, resulting in better and smaller feature subsets. Through the use of Mahalanobis distance and a multi-objective fitness, the MBMNABC-Ma algorithm exhibited enhanced performance in both feature selection and classification accuracy, especially in the context of diabetes detection. The detailed working principles of MBMNABC-Ma using Mahalanobis distance measures and multi-objective optimization are outlined in Table 2 to Table 5.

Table 2. The Modified Binary Multi-Neighborhood Artificial Bee Colony with Mahalanobis (MBMNABC-Ma) Feature Selection using Mahalanobis Distance Measure

INPUT		
Dataset:	$X \in R^{N \times D}$	Diabetes mellitus dataset with N samples and D features
Class Labels:	$Y \in \{0,1\}^N$	Binary class labels, where 1 indicates a diabetic and 0 a non-diabetic
Parameters:	N_{pop}	Number of candidate solutions
	T	Maximum number of iterations
	R_{max}	Maximum neighbourhood radius for identifying far neighbours using Mahalanobis distance
	Cls	A supervised classifier for fitness evaluation
	k	Number of folds for cross-validation
	α, β	Weights for multi-objective fitness
OUTPUT		
A reduced subset of features $F_{best} \subseteq F$ that provides the highest classification accuracy and minimizes the number of features for detecting diabetes.		
PROCESS		
Step 1.	**Initialize the Population**	
	Step 1.1.	Generate initial solutions $x_1, x_2, \ldots, xN_{pop}$ where each solution x_i is a binary vector of length D (representing selected features): $x_{ij} = \begin{cases} 1 & if\ rand\ (0,1) \geq 0.5 \\ 0 & otherwise \end{cases}$ for $i = 1,2, \ldots, N_{pop}$ and $j = 1,2, \ldots \ldots, D$ (1)
	Step 1.2.	Evaluate the fitness $f(x_i)$ of each solution using multi-objective fitness: Classifier trained on the selected features corresponding to x_i. Perform k-fold cross-validation and compute the average accuracy $Acc(x_i)$ Compute fitness: $f(x_i) = \alpha \times Acc(x_i) - \beta \times \frac{Number\ of\ Selected\ Features\ in\ x_i}{D}$ (2)
Step 2.	**Far Neighbour Exploration and New Solution Generation**	
for t=1 ***to*** T ***do***		
	Step 2.1.	**Far Neighbour Identification** For each solution x_i compute Mahalanobis distance to every other solution x_k: $MD(x_i, x_k) = \sqrt{(x_i - x_k)^{-T} S^{-1} (x_i - x_k)}$ (3) where S is the covariance matrix of the feature data. Identify x_k as a far neighbor if: $MD(x_i, x_k) \geq max(MDi) \geq R \times mean(MD_i)$, where R is a neighborhood radius that is updated dynamically: $R = R_{max}\left(1 - \frac{t}{T}\right)$ (4)
	Step 2.2.	**Generate a New Candidate Solution** Call the Neighbor-Based Solution Generation Algorithm (Table 3) with the current solution x_i and its far neighbors x_k (identified using Mahalanobis distance) to generate a new solution v_i. $v_i = NeighborBasedSolution\ Generation\ (x_i, x_k)$ (5)
	Step 2.3.	**Selection Based on Fitness** Calculate the selection probability p_i for each solution x_i based on its fitness $p_i = \frac{f(x_i)}{\sum_{k=1}^{N_{pop}} f(x_k)}$ (6) //Use these probabilities to probabilistically select solutions for further exploration
	Step 2.4.	**Local Search on Top Solutions (Memetic Search)** Every M iteration, perform local refinement on top k solutions: Try adding or removing one feature at a time. Accept changes that improve the fitness $f(x)$.
	Step 2.5.	**Explore Near Neighbors** Call the Fitness-Based Neighbor Exploration Algorithm (Table 4) to explore nearby solutions using Mahalanobis distance for neighbor selection and generate a new solution v_i $v_i = FitnessBased\ Neight)$ (7)
	Step 2.6.	**Scout Bee Exploration and Solution Replacement**
	Step 2.6.1.	**Identify Stagnating Solutions** If a solution does not improve after a certain number of trials (stagnation), it is marked for replacement.
	Step 2.6.2.	**Generate a New Random Solution** Call the Random Solution Generation Algorithm (Table 5) to replace stagnating solutions with a new random solution x_i^{new}. $x_i^{new} = RandomSolution\ Generation\ (x_i)$ (8)
	Step 2.7.	**Memorize the Best Solution** After each iteration, keep track of the solution x_{best} with the highest accuracy.
end		

Step 3.	**Best feature subset F_{best} identification**
	Return $F_{best} \subseteq F$ the subset of features for the solution x_{best} for which the highest accuracy was received.

The neighbor-based solution generation of the MBMNABC-Ma feature selection approach is designed to explore the feature space effectively by leveraging the information from far neighbors identified using Mahalanobis distance. After identifying the far neighbors in Step 2.2 of the MBMNABC-Ma algorithm, the Neighbor-Based Solution Generation Algorithm is invoked to generate a new candidate solution by utilizing the structure of the best-performing neighbors. By considering the data's covariance structure through Mahalanobis distance, the algorithm ensures that neighbors are more meaningfully selected based on feature interdependencies. This leads to a more informed and efficient exploration of the search space, increasing the probability of discovering feature subsets that enhance overall fitness. The detailed process flow of neighbor-based solution generation has been outlined in Table 3.

Table 3. The Neighbor-Based Solution Generation Algorithm

INPUT	
• Current solution x_i	
• Set of far neighbors x_k, identified using Mahalanobis distance	
OUTPUT	
New candidate solution v_i	
PROCESS	
Step 1.	**Compute Average Best Neighbor**
	Compute the APB_{ij} of the far neighbors x_k $APB_{ij} = \frac{1}{N_{far}} \sum_{k=1}^{N_{far}} p^{best_{kj}}$ (9) //where $p^{best_{kj}}$ denotes the best solution found so far for the k^{th} far neighbor
Step 2.	**Generate New Solutions**
	Generate a new solution v_i using the best information from the far neighbors: $v_{ij}=x_{ij}+rand(0,1)\times(x_{ij}-APB_{ij})$ (10)
Step 3.	**Convert to Binary**
	Convert v_{ij} into a binary value for feature selection $v_{ij} = \begin{cases} 1 & if\ rand\ (0,1) \geq 0.5 \\ 0 & otherwise \end{cases}$ (11)
Step 4.	**Return New Solution**
	Return the new solution v_i to MBMNABC

Similar to the Neighbor-Based Solution Generation Algorithm, the Fitness-Based Neighbor Exploration Algorithm, as described in Table 4, also aims to enhance the feature space exploration. However, instead of focusing on far neighbors, it concentrates on near neighbors, which are solutions that are close to the current solution x_i based on Mahalanobis distance and have relatively high fitness values. This method refines the search by exploiting the best-performing nearby solutions.

Table 4. The Fitness-Based Neighbor Exploration Algorithm

INPUT	
• Current solution x_i	
• Set of far neighbors x_k, identified using Mahalanobis distance	
OUTPUT	
New candidate solution v_i	
PROCESS	
Step 1.	**Select Best Neighbor**
	Select the best neighbor x_{best_k} from the set of near neighbors based on fitness.
Step 2.	**Generate New Solutions**
	Generate a new solution v_i using the best near neighbor: $v_{ij}=x_{ij}+rand(0,1)\times(x_{ij}-x_{best_{kj}})$ (12)
Step 3.	**Compare and update fitness**
	Compare the fitness of the new solution $f(v_i)$ with the current solution $f(x_i)$ *if* $f(v_i)>f(x_i)$ then $x_i=v_i$ (13) *end*
Step 4.	**Return updated solution x_i**

In contrast to both the Neighbor-Based Solution Generation Algorithm and the Fitness-Based Neighbor Exploration Algorithm, which rely on the information from neighboring solutions, the Random Solution Generation Algorithm is used when a current solution has stagnated, i.e., it fails to improve after several trials. This algorithm seeks to provide variety to the search space and avoid it from getting stuck in local optima.

Table 5. The Random Solution Generation Algorithm

INPUT	
Current solution x_i marked for replacement	
OUTPUT	
New random solution x_i^{new}	
PROCESS	
Step 1.	**Generate Random Solution**
	For each feature *j*, generate a random binary value for the new solution: $x_{ij}^{new} = \begin{cases} 1 & if\ rand\ \ (0,1) \geq 0.5 \\ 0 & otherwise \end{cases}$ (14)
Step 2.	**Return the new random solution x_{ij}^{new}**

The MBMNABC-Ma algorithm for feature selection operates through multiple stages of exploration and exploitation, guided by neighborhood-based solution generation, fitness-based exploration, and random solution generation. Together, these methods ensure that the algorithm effectively navigates the feature space to identify an optimal subset of features that maximizes classification accuracy while minimizing redundancy. At the core of the MBMNABC-Ma algorithm the Mahalanobis distance measure is used to define

the proximity between solutions in the search space. Unlike hamming distance, the Mahalanobis distance adjusts distances according to the covariance structure of the data and considers feature correlations [22]. This provides a more accurate representation of feature relationships in continuous-valued datasets, enabling more precise identification of significant and non-redundant features. Consequently, MBMNABC-Ma is particularly effective in feature selection tasks involving real-world datasets where features exhibit interdependencies, such as in diabetes detection.

To evaluate the effectiveness of the MBMNABC-Ma algorithm, feature selection was independently performed on each of the five diabetes datasets. These datasets contain heterogeneous features ranging from medical test results and medication usage to lifestyle indicators. The algorithm was applied after preprocessing, enabling the identification of feature subsets most relevant to classification while eliminating redundant or irrelevant attributes. Table 6 presents the summary of selected features for each dataset, listing the total number of features, the count of those selected by the algorithm, and the specific feature names retained.

The results indicate that MBMNABC-Ma effectively adapts to varying dataset structures. For instance, the Merged Dataset (130-US + PIMA), originally containing 46 features, was reduced to 21 highly informative variables, such as Repaglinide, Pioglitazone, Change, Readmitted, Race, and Tolbutamide. Similarly, the Iranian Ministry of Health dataset retained 11 out of 19 features, with BMI, Triglyceride, and Cholesterol appearing as the most influential. In the Questionnaire dataset, 15 out of 17 features were preserved, including lifestyle-related variables like Family_diabetes, BMI, Stress, and Sleep. Notably, all 16 features from the Sylhet Hospital dataset were selected, highlighting the clinical relevance of symptoms such as Polydipsia, Polyuria, Alopecia, and visual blurring. Likewise, the PIMA dataset retained all 8 of its original features, suggesting their collective importance in diabetes detection.

This dataset-specific selection underscores the adaptability and robustness of the MBMNABC-Ma algorithm across varying feature spaces. It also demonstrates the algorithm's capacity to uncover both clinical and behavioral indicators that are strongly associated with diabetes, thereby enhancing model interpretability and predictive power.

Table 6. Dataset-wise Feature Selection Results Using the Modified Binary Multi-Neighbourhood Artificial Bee Colony with Mahalanobis-based (MBMNABC-Ma) Algorithm

Dataset	Total Features	No. of Selected Features	Features Name
Merged Dataset (130-US + PIMA) [11]	46	21	Repaglinide, Pioglitazone, Change, Readmitted, Race, Tolbutamide, Gender, Age, A1Cresult, DP_function, Weight, Max_glu_serum, Pregnancy, Troglitazone, Glipizide, Citoglipton, TriFold_Skin Thickness, Acetohexamide, Examide, BMI,
Iranian Ministry of Health [15]	19	11	BMI, Triglyceride, Cholesterol, Weight, HDL, History_of_pregnancy, FBS, Result_of_high_blood_pressure_screening, Age, History_of_diabetes, Family_history_of_diabetes
Questionnaire Dataset [12]	17	15	Family_diabetes, BMI, Age, Stress, Physically_active, Sleep, Soundsleep, Urinationfreq, Regularmedicine, Bplevel, Alcohol, Pregnancies, Gender, Highbp, Junkfood
Hospital of Sylhet Dataset [16]	16	16	Polydipsia, Polyuria, Age, Gender, Sudden_weight_loss, Irritability, Alopecia, Weakness, Itching, Polyphagia, Visual_blurring, Delayed_healing, Genital_thrush, Muscle_stiffness, Obesity, Partial_paresis
PIMA Dataset [13]	8	8	Glucose, Age, Insulin, Pregnancies, BloodPressure, BMI, SkinThickness, DiabetesPedigreeFunction

2.3. PROPOSED DIABETES DETECTION MODEL COMBINING MBMNABC-MA AND OPTIMIZED DECISION FOREST

The proposed method for diabetes detection uses a combination of two powerful techniques: Modified Binary Multi-Neighbourhood Artificial Bee Colony with Mahalanobis Distance (MBMNABC-Ma) for selecting features and an Optimized Decision Forest (ODF) for classification. The process begins by preparing the diabetes dataset through steps like dealing with missing values, normalization, and encoding. MBMNABC-Ma identifies the most important features by using Mahalanobis distance, which takes into account the relationships between features and the dataset's internal structure. This makes it highly effective for continuous datasets where feature interdependence plays a major role in prediction accuracy. The algorithm starts with an initial group of feature subsets and evaluates them based on how well they support classification. It explores both nearby and distant features in the dataset to ensure a thorough search and maintains diversity by introducing new random solutions when progress slows down. This cycle continues until the best feature subset, F_{best} is found.

Once the relevant features are selected, the ODF classifier is applied for predicting diabetes. ODF is an improved version of the Random Forest algorithm. It enhances performance by selecting only the most important decision trees to form a subforest, which helps reduce computation time and improve accuracy. Since ODF focuses on key features identified by MBMNABC-Ma, it provides better results, minimizes redundancy, and improves the model's ability to generalize. Additionally, it makes the prediction process more transparent, helping medical professionals understand how decisions are made, a valuable benefit in healthcare applications.

Overall, combining MBMNABC-Ma for feature selection with ODF for classification results in a highly accurate and efficient system for detecting diabetes. It handles high-dimensional medical data effectively, offers strong prediction performance in terms of accuracy, sensitivity, and specificity, and maintains clarity in decision-making, all of which are essential in clinical settings. The proposed model is illustrated in Fig. 1.

3. RESULTS AND DISCUSSION

This research compares feature selection using the conventional BMNABC algorithm and the Modified BMNABC with Mahalanobis distance (MBMNABC-Ma) across five transcontinental diabetes datasets. All the datasets were first preprocessed using the KNN imputation method and then passed through the Optimized Decision Forest (ODF) framework with the help of the Random Forest Ensemble (RFE) algorithm.

The performance of the proposed MBMNABC-Ma + ODF (RFE) method was compared with other classifiers like MBMNABC-Ma + k-Nearest Neighbors (kNN), MBMNABC-Ma + Support Vector Machine (SVM), MBMNABC-Ma + Naïve Bayes, MBMNABC-Ma + Rough Set (RS), and MBMNABC-Ma + C4.5 decision tree. Comparative results were also analyzed against the conventional BMNABC and previously published research. Accuracy, specificity, and sensitivity were employed as evaluation metrics. The MBMNABC-Ma + ODF (RFE) approach demonstrated superior performance across all metrics and datasets, highlighting its potential for robust, real-world diabetes detection applications. To ensure a strong comparison, key performance metrics were analyzed, including the Receiver Operating Characteristic (ROC) curve. These illustrations make it easier to distinguish the models' capabilities in various contexts. To comprehensively examine the diagnostic capabilities of each model, a detailed evaluation of key metrics, namely accuracy, sensitivity, and specificity, was performed. This rigorous comparison offers a clear understanding of the advantages and distinctive strengths of the proposed MBMNABC-Ma, combined with ODF using RFE about competing approaches. An educated viewpoint on the suggested model's possible real-world applications is facilitated by this thorough assessment, which provides insightful information on the model's efficacy and dependability in the context of diabetes diagnosis.

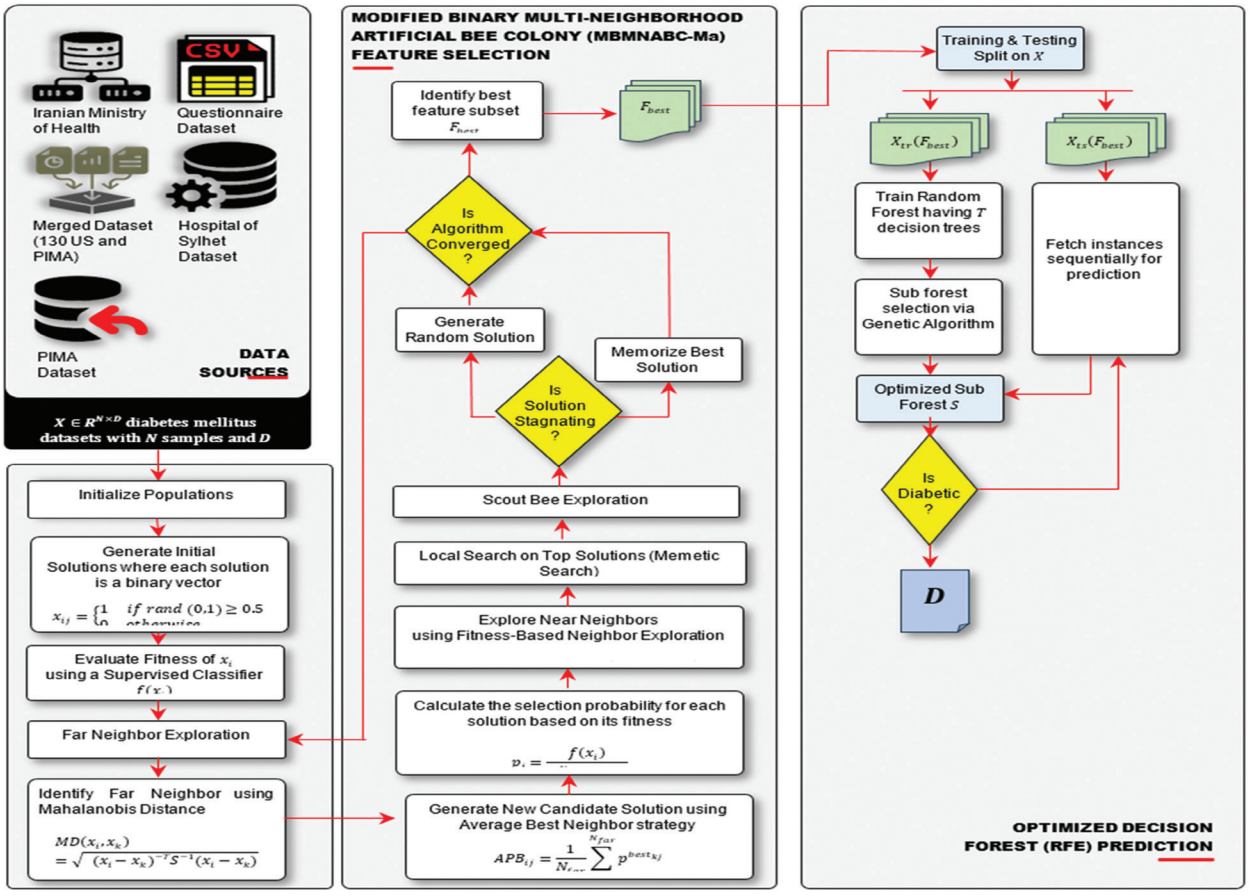


Fig. 1. Block diagram of the proposed diabetes detection model

A thorough analysis of several models, including MBMNABC-Ma + C4.5, MBMNABC-Ma + k-NN, MBMNABC-Ma + NB, MBMNABC-Ma + RS, and MBMNABC-Ma + SVM, in combination with the MBMNABC-Ma + ODF (RFE) model, is shown in Fig. 2. The combined dataset, which incorporates information from both the US and PIMA sources, is used for this evaluation. A useful illustration of the performance evaluation performed on the

Iranian Ministry of Health dataset may be seen in Fig. 3. The assessment findings from the questionnaire dataset are further displayed in Fig. 4, emphasizing the models' comparative capabilities. Additionally, the results of the performance evaluation carried out on the Sylhet Diabetes Hospital dataset are presented in Fig 5.

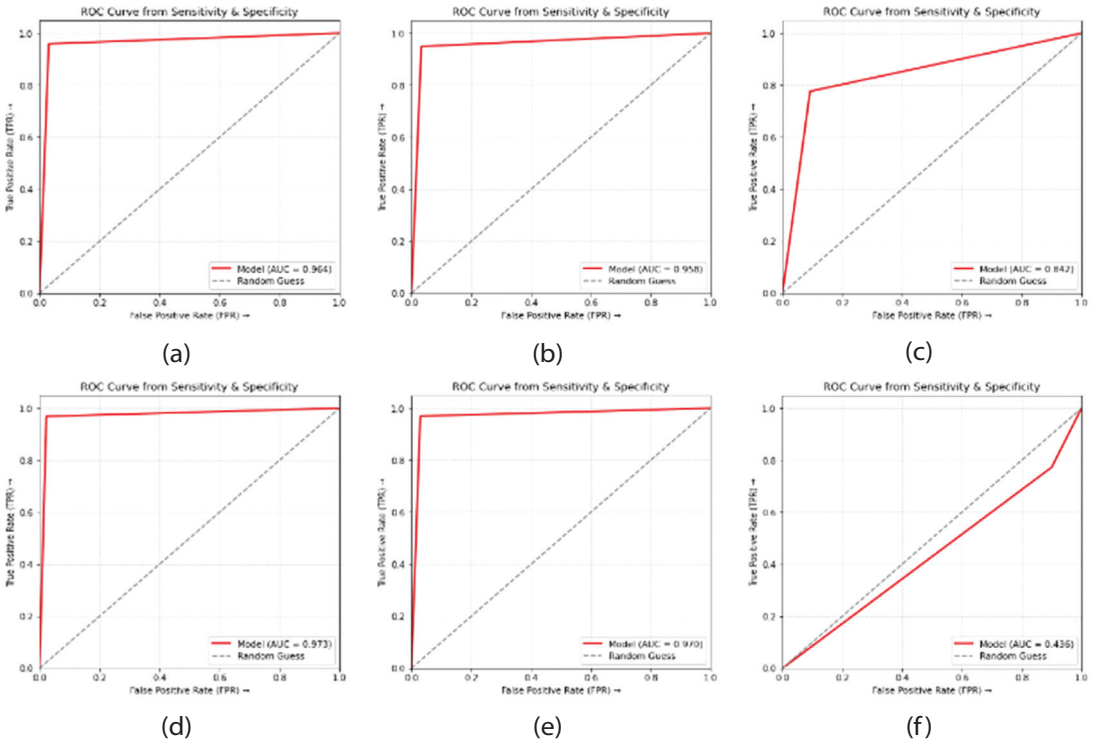


Fig 2. Comparative performance analysis of the models (a) MBMNABC-Ma + C4.5, (b) MBMNABC-Ma + k-NN, (c) MBMNABC-Ma + NB, (d) MBMNABC-Ma + ODF(RFE), (e) MBMNABC-Ma + RS, and (f) MBMNABC-Ma + SVM evaluated on the merged dataset (130 US and PIMA samples)

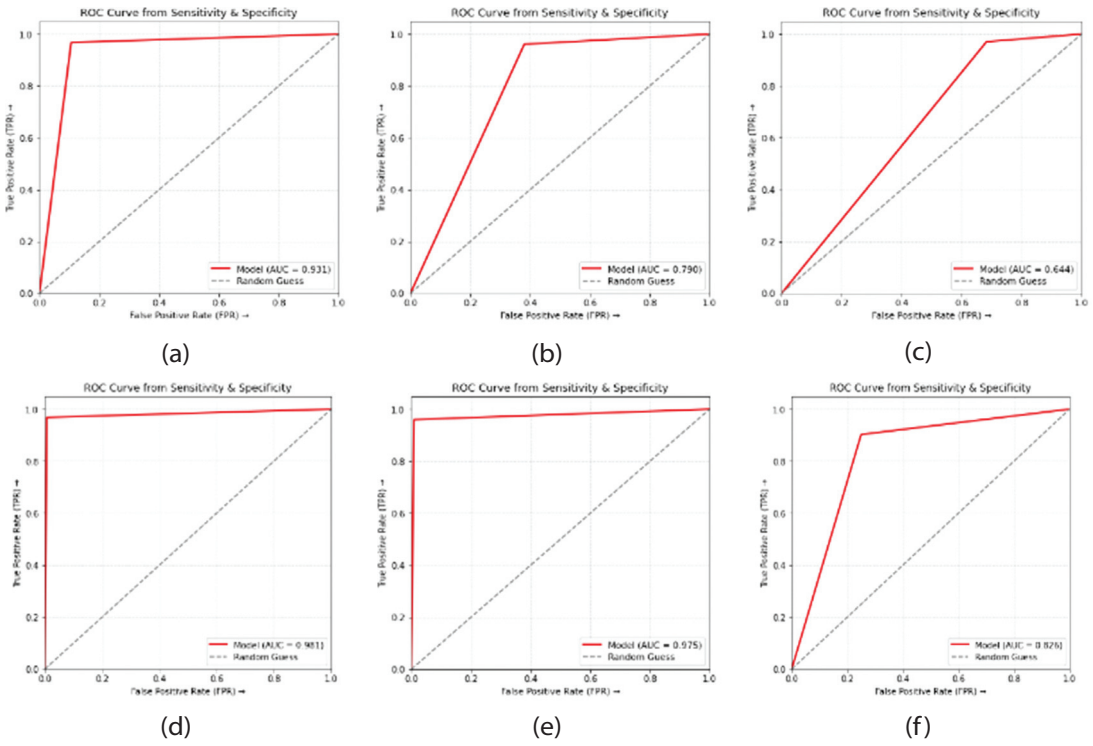


Fig 3. Comparative performance analysis of the models (a) MBMNABC-Ma + C4.5, (b) MBMNABC-Ma + k-NN, (c) MBMNABC-Ma + NB, (d) MBMNABC-Ma + ODF(RFE), (e) MBMNABC-Ma + RS, and (f) MBMNABC-Ma + SVM evaluated on the Iranian Ministry of Health dataset

Lastly, Fig 6 presents a comprehensive visualization of the performance evaluation conducted on the PIMA dataset, offering a clear and comparative perspective on the effectiveness of the models across different datasets.

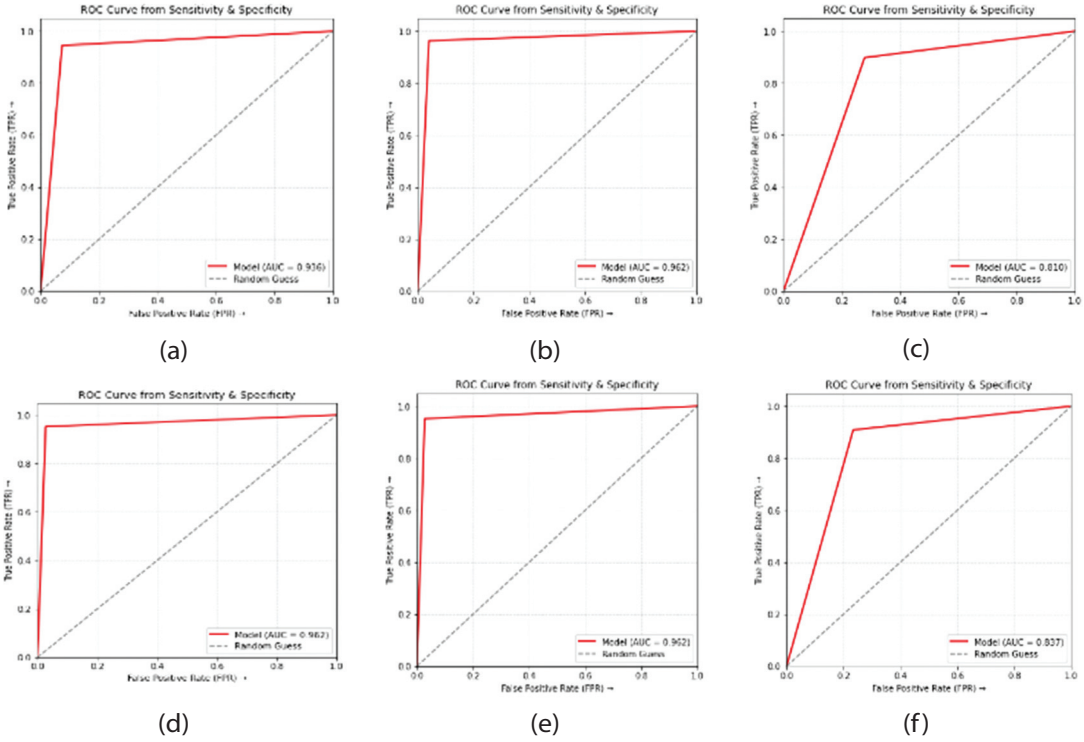


Fig 4. Comparative performance analysis of the models (a) MBMNABC-Ma + C4.5, (b) MBMNABC-Ma + k-NN, (c) MBMNABC-Ma + NB, (d) MBMNABC-Ma + ODF(RFE), (e) MBMNABC-Ma + RS, and (f) MBMNABC-Ma + SVM evaluated on the Questionnaire Dataset

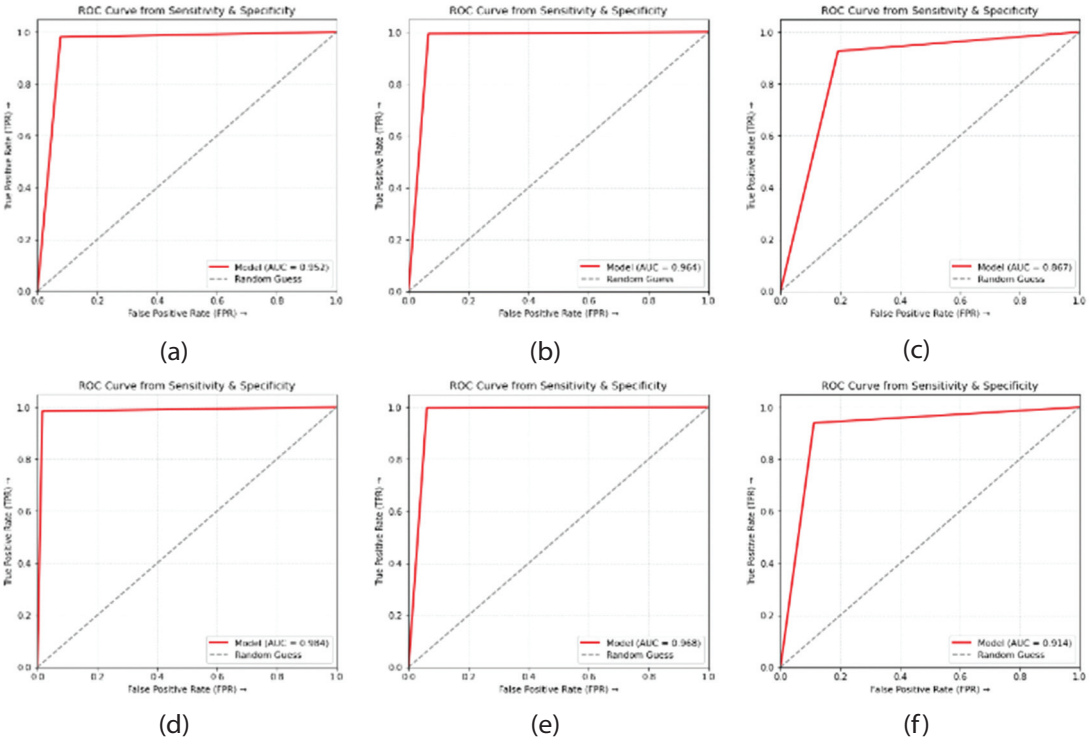


Fig 5. Comparative performance analysis of the models (a) MBMNABC-Ma + C4.5, (b) MBMNABC-Ma + k-NN, (c) MBMNABC-Ma + NB, (d) MBMNABC-Ma + ODF(RFE), (e) MBMNABC-Ma + RS, and (f) MBMNABC-Ma + SVM evaluated on the Sylhet Diabetes Hospital of Sylhet Dataset

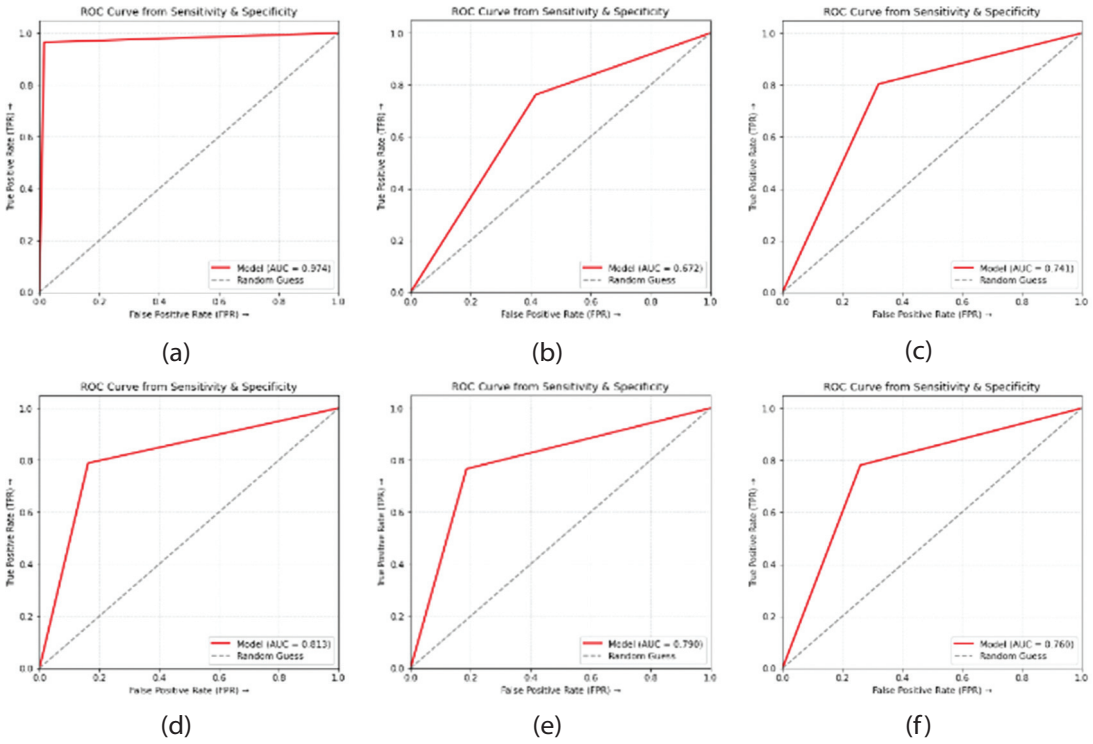


Fig 6. Comparative performance analysis of the models (a) MBMNABC-Ma + C4.5, (b)MBMNABC-Ma + k-NN, (c) MBMNABC-Ma + NB, (d) MBMNABC-Ma + ODF(RFE), (e) MBMNABC-Ma + RS, and (f) MBMNABC-Ma + SVM evaluated on the PIMA dataset

The above figures comprehends the subtle differences in performance of each model across various datasets, which helps to provide a thorough grasp of their prospective applications and probable ramifications in the field of diabetes diagnosis.

3.1. PERFORMANCE OUTCOMES WITH VARIOUS METHODS

The comparative analysis evaluated models based on accuracy, specificity, and sensitivity. The models assessed include MBMNABC-Ma + Random Forest (RF), MBMNABC-Ma + k-Nearest Neighbors (k-NN), MBMNABC-Ma + Naïve Bayes (NB), MBMNABC-Ma + C4.5, MBMNABC-Ma + Rough Set (RS), and MBMNABC-Ma + Optimized Decision Forest (ODF) using Random Forest Ensemble (RFE). The results for the Merged Dataset (130 US and PIMA records) are presented in Table 7, while the performance outcomes for the Iranian Ministry of Health dataset, the Questionnaire Dataset, the Hospital of Sylhet Dataset, and the PIMA dataset are detailed in Tables 8, 9, 10, and 11, respectively.

In Table 7, the MBMNABC-Ma + ODF (RFE) algorithm achieved the highest accuracy (97.23%) for diabetes detection, outperforming all other compared methods. MBMNABC-Ma + Naïve Bayes achieved an accuracy of 82.06%, MBMNABC-Ma + SVM achieved 83.94%, MBMNABC-Ma + k-NN reached 95.68%, MBMNABC-Ma + C4.5 attained 96.37%, and MBMNABC-Ma + Rough Set (RS) scored 96.98%. For specificity, MBMNABC-Ma + SVM achieved the highest value at 100%, followed by MBMNABC-Ma + ODF (RFE) at 97.75%, MBMNABC-Ma + C4.5 at 96.95%, and MBMNABC-Ma + RS at 97.06%. High specificity indicates the ability of the model to accurately identify healthy individuals, thereby reducing false positives. In terms of sensitivity, MBMNABC-Ma + RS (96.93%) and MBMNABC-Ma + ODF (RFE) (96.82%) performed the best, followed by MBMNABC-Ma + C4.5 (95.91%) and MBMNABC-Ma + k-NN (94.89%). Meanwhile, MBMNABC-Ma + SVM (77.27%) and MBMNABC-Ma + Naïve Bayes (77.64%) exhibited the lowest sensitivity, suggesting their potential challenges in accurately detecting all diabetes cases. High sensitivity is crucial to ensure diabetic individuals are correctly identified, minimizing the occurrence of false negatives.

Table 7. Performance of Proposed Detection Methods on the Merged Dataset (130 US and PIMA records) (10-Fold Cross Validation)

Detection Method	Accuracy (%)	Specificity (%)	Sensitivity (%)
MBMNABC-Ma + C4.5	96.37	96.95	95.91
MBMNABC-Ma + kNN	95.68	96.7	94.89
MBMNABC-Ma + NB	82.06	90.78	77.64
MBMNABC-Ma + ODF(RFE)	**97.23**	97.75	96.82
MBMNABC-Ma + RS	96.98	97.06	96.93
MBMNABC-Ma + SVM	83.94	100.00	77.27

Table 8. Performance of Proposed Detection Methods on the Iranian Ministry of Health dataset (10-Fold Cross Validation)

Detection Method	Accuracy (%)	Specificity (%)	Sensitivity (%)
MBMNABC-Ma + C4.5	96.19	89.41	96.76
MBMNABC-Ma + kNN	92.57	61.95	96.07
MBMNABC-Ma + NB	81.34	31.7	97.02
MBMNABC-Ma + ODF(RFE)	**97.93**	99.29	96.82
MBMNABC-Ma + RS	97.37	99.11	95.97
MBMNABC-Ma + SVM	90.22	75	90.25

The MBMNABC Ma combined with ODF using RFE achieved the highest accuracy of 97.93 percent, specificity of 99.29 percent, and sensitivity of 96.82 percent on the Iranian Ministry of Health dataset, demonstrating its strong capability for making precise predictions, as presented in Table 8. The MBMNABC-Ma + RS method closely shadowed by achieving 97.37% accuracy, with the specificity of 99.11% and 95.97% sensitivity. Comparing other methods, such as MBMNABC-Ma + k-NN and MBMNABC-Ma + Naïve Bayes, showed lesser accuracy and specificity, with MBMNABC-Ma + Naïve Bayes particularly displaying a very low specificity (31.7%). Hence, the result emphasizes that the MBMNABC-Ma + ODF(REF) and MBMNABC-Ma + RS performed extremely well in this dataset, showing significant potential for the precise and reliable detection of diabetes on the Iranian Ministry of Health dataset.

Table 9. Performance of Proposed Detection Methods on the Questionary dataset (10-Fold Cross Validation)

Detection Method	Accuracy (%)	Specificity (%)	Sensitivity (%)
MBMNABC-Ma + C4.5	94.01	92.65	94.48
MBMNABC-Ma + kNN	**96.21**	96	96.3
MBMNABC-Ma + NB	84.76	72.16	89.84
MBMNABC-Ma + ODF(RFE)	96.05	97.27	95.18
MBMNABC-Ma + RS	96.05	97.27	95.18
MBMNABC-Ma + SVM	86.86	76.6	90.83

Table 9, with the Questionary dataset, specifies significant insights into the efficiency of various methods for diabetes detection. In this dataset, the MBMNABC-Ma + kNN method gave the admirable accuracy of 96.21%, with strong specificity (96%) and sensitivity (96.3%). The methods, MBMNABC-Ma + ODF(REF) and MBMNABC-Ma + RS, gave the same results in terms of accuracy (96.05%), specificity (97.27%), and sensitivity (95.18%). The MBMNABC-Ma + C4.5 also gave a good accuracy of 94.01%, with specificity (92.65%) and sensitivity (94.48%). The other method, MBMNABC-Ma + NB and MBMNABC-Ma + SVM, provided lower accuracy, specificity, and sensitivity. This highlights that the MBMNABC-Ma + ODF(REF) and MBMNABC-Ma + RS performed best in the detection of diabetes in the Questionary dataset.

Table 10. Performance of Proposed Detection Methods on the Hospital of Sylhet dataset (10-Fold Cross Validation)

Detection Method	Accuracy (%)	Specificity (%)	Sensitivity (%)
MBMNABC-Ma + C4.5	95.8	92.31	98.04
MBMNABC-Ma + kNN	97	93.4	99.34
MBMNABC-Ma + NB	87.82	80.79	92.62
MBMNABC-Ma + ODF(RFE)	**98.39**	98.39	98.41
MBMNABC-Ma + RS	96.36	93.92	99.66
MBMNABC-Ma + SVM	92.01	88.83	93.93

Using the Hospital of Sylhet dataset, Table 10 highlights the performance of various algorithms for diabetes detection. In this dataset, the MBMNABC-Ma + ODF(REF) methods gave the highest accuracy (98.39%) as compared to other methods, with the specificity and sensitivity of 98.39% and 98.41% respectively. Similarly, the MBMNABC-Ma + kNN achieved an accuracy of 97% with a specificity of 93.4% and a sensitivity of 99.39%. Although the MBMNABC-Ma + RS did not provide the best accuracy and specificity as compared to MBMNABC-Ma + ODF(REF) and MBMNABC-Ma + kNN but it achieved the highest sensitivity of 99.66%. The MBMNABC-Ma + NB method gave the lowest accuracy, specificity, and sensitivity of 87.82%, 80.70% and 92.62% respectively, demonstrating that MBMNABC-Ma + ODF(REF) and MBMNABC-Ma + kNN emerged as the top methodologies for diabetes detection in the Hospital of Sylhet dataset.

In Table 11, using the PIMA dataset, the MBMNABC-Ma + ODF (RFE) method achieved the highest accuracy of 80.98% specificity of 83.74% and sensitivity of 78.88% indicating that it is efficient in accurately detecting diabetes in this dataset, The MBMNABC-Ma + RS achieved the second highest accuracy with 78.66%, reasonable specificity of 81.35% and notable sensitivity of 76.65%. The MBMNABC-Ma + c4.5 method gave the

highest sensitivity of 96.44% and specificity of 98.45% but notably lower accuracy of 76.17%. Hence, these results suggest that MBMNABC-Ma + ODF (RFE) and MBMNABC-Ma + RS performed best in the detection of diabetes in the PIMA dataset.

Table 11. Performance of Proposed Detection Methods on the PIMA Dataset (10-Fold Cross Validation)

Detection Method	Accuracy (%)	Specificity (%)	Sensitivity (%)
MBMNABC-Ma + C4.5	76.17	98.45	96.44
MBMNABC-Ma + kNN	70.44	58.30	76.20
MBMNABC-Ma + NB	76.43	67.90	80.38
MBMNABC-Ma + ODF(RFE)	**80.98**	83.74	78.88
MBMNABC-Ma + RS	78.66	81.35	76.65
MBMNABC-Ma + SVM	77.08	73.96	78.13

3.2. COMPARATIVE ANALYSIS WITH EXISTING TECHNIQUES

In Table 12, the proposed MBMNABC-Ma + ODF(RFE) method achieves a remarkable accuracy of 97.23%. The proposed methodology considerably gave better results as compared to the conventional BMNABC + ODF(RFE) reported by Pradhan et al. [23] on the Merged Dataset (130 US and PIMA), which had an accuracy of 96.36%. The substantial improvement of 0.87% emphasizes the effectiveness of MBMNABC-Ma in refining feature selection processes. Other methods, like SMOTE + Random Forest by Pradhan et al.[24] with the accuracy of 84.60% and LIBSVM by Negi et al. [11] with an accuracy of 73.00%, further illustrates the strength of the proposed method. This enhancement not only reinforces its potential as a leading method in the field but also highlights its ability to deliver superior classification outcomes compared to the traditional approach.

Table 12. Comparative result analysis between Conventional BMNABC and MBMNABC-Ma for the Merged Dataset (130 US and PIMA)

Dataset	Authors	Methods	Accuracy (%)
Merged Dataset (130 US and PIMA) [11]	Negi et al. [11]	SVM (Classification) + LIBSVM (Feature Selection)	73.00
	Pradhan et al. [23]	BMNABC + ODF(RFE)	96.36
	Pradhan et al. [24]	Random Forest + SMOTE	84.60
	Proposed	MBMNABC-Ma + ODF(RFE)	**97.23**

In Table 13, the proposed method MBMNABC-Ma + ODF(RFE) performed better than the BMNABC + ODF(RFE) by Pradhan et al. [23], which stated an accuracy of 97.93% accuracy in the Iranian Ministry of Health Dataset. Even though the proposed method performs better than several advanced methods, it's important to identify that MBMNABC-Ma's benefits are obtained from its flexibility and creative feature selection skills. For instance, Heydari et al. [15] achieved an accuracy of 97.44% with expert feature selection combined with ANN, Pradhan et al. [24] with SMOTE combined with Random Forest, achieved 96.80% accuracy, and Habibi et al. [25] reached 97.60% using expert feature selection with C4.5. The little discrepancy in accuracy shows that the proposed approach not only beats the competition but also provides substantial value in terms of the interpretability and pertinence of characteristics. The little discrepancy in accuracy shows that the proposed approach not only beats the competition but also provides substantial value in terms of the interpretability and pertinence of characteristics.

Table 13. Comparative result analysis between Conventional BMNABC and MBMNABC-Ma for the Iranian Ministry of Health

Dataset	Authors	Methods	Accuracy (%)
Iranian Ministry of Health [15]	Heydari et al. [15]	ANN + Expert Feature selection (Manual)	97.44
	Habibi et al.[25]	C4.5 + Expert Feature selection (Manual)	97.60
	Pradhan et al. [24]	Random Forest + SMOTE	96.80
	Pradhan et al.[23]	BMNABC + ODF(RFE)	97.28
	Proposed	MBMNABC-Ma + ODF(RFE)	**97.93**

The MBMNABC-Ma + kNN methodology acquires an accuracy of 96.21% for the Questionnaire Dataset in Table 14, falling only 0.22% short of the top-performing technique, BMNABC + ODF(RFE) by Pradhan et al. [23], which recorded 96.43%.

Table 14. Comparative result analysis between Conventional BMNABC and MBMNABC-Ma for the Questionnaire Dataset

Dataset	Authors	Methods	Accuracy (%)
Questionnaire Dataset [12]	Tigga et al. [12]	Random Forest	94.10
	Pradhan et al. [24]	Random Forest + SMOTE	93.70
	Pradhan et al. [23]	BMNABC + ODF(RFE)	96.43
	Proposed	MBMNABC-Ma+ kNN	**96.21**

However, it does not reach this benchmark, the proposed technique still displays improved performance compared to the BMNABC + ODF (RFE), which obtains 96.43% accuracy. In comparison, Tigga et al [12] achieved 94.10% accuracy with Random Forest and Pradhan et al. [24] reported 93.70% with SMOTE combined with Random Forest. This tiny difference exhibits how competitive MBMNABC-Ma is at achieving appropriate characteristics and classifying data.

The close performance emphasizes MBMNABC-Ma's possibility for additional evolution and offers positive options for improvement.

Table 15 shows the impressive accuracy of 98.39% achieved by the MBMNABC-Ma + ODF(RFE) in the Hospital of Sylhet Dataset. This is still not as accurate as the greatest recorded accuracy of 99.23% by Gündoğdu et al. [26], but it is significantly more accurate than the BMNABC + kNN technique of 97.01% by Pradhan et al. [23]. This improvement of 1.38% underlines the efficacy of the proposed strategy in addressing the challenges of this dataset. Other methodologies, like SMOTE with Random Forest by Pradhan et al.[24] which achieved an accuracy of 98.10% and SFS + ANN by Buyrukoğlu et al. [27] with an accuracy of 99.10% further emphasis the strength of the proposed method. This improvement of 1.38% underlines the efficacy of the proposed strategy in addressing the challenges of this dataset. MBMNABC-Ma's dynamic performance demonstrates how it may improve feature selection and emphasizes its usefulness in intricate real-word situations.

Table 15. Comparative result analysis between Conventional BMNABC and MBMNABC-Ma for the Hospital of Sylhet Dataset

Dataset	Authors	Methods	Accuracy (%)
Hospital of Sylhet Dataset [16]	Islam et al. [16]	Random Forest	97.4
	Pradhan et al. [24]	Random Forest + SMOTE	98.10
	Buyrukoğlu et al. [27]	ANN + SFS	99.10
	Nipa et al. [28]	APGWO-based MLP	97.00
	Gündoğdu et al. [26]	XGBoost + Random forest	99.23
	Prasanth [29]	XG Boost + SelectKBest	98.00
	Yasar [30]	FFNN + CSA	99.04
	Ma [31]	Neural Network + Min–Max	96.20
	Saboor et al. [32]	Optimize Selection + kNN + SMOTE	93.66
	Elsadek et al.[33]	Random Forest + Supervised Attribute Filter	97.88
	Rony et al. [34]	Random Forest + CFS	97.50
	Hasan et al. [35]	Extra Trees + PCC	99.06
	Pradhan et al. [23]	BMNABC + kNN	97.01
	Proposed	MBMNABC-Ma + ODF(RFE)	**98.39**

Finally, in Table 16, the proposed methodology (MBMNABC-Ma + ODF(RFE) achieved an accuracy of 80.98% which is more as compared to the BMNABC + ODF(RFE) result of 77.21% published by Pradhan et al. [23]. For instance, methodologies employing mean imputation and Naïve Bayes reported by Mousa et al. [36] with an accuracy of 85.00% and Chang et al. [21] with an accuracy of 79.13% it yields lower accuracies. In the PIMA dataset, an existing technique varying from 72.90% to 79.13%, the MBMNABC-Ma + ODF(RFE) method shows a competitive advantage. The results show that the proposed methodology, MBMNABC-Ma, is flexible and robust in different datasets: the Merged Dataset (130 US and PIMA), the Iranian Ministry of Health, the Questionnaire dataset, the Hospital of Sylhet, and the PIMA datasets. Although the proposed methodology falls short of the existing benchmarks, its strength lies in its flexibility, capability, and robustness in detecting diabetes.

Table 16. Comparative result analysis between Conventional BMNABC and MBMNABC-Ma for the PIMA Dataset

Dataset	Authors	Methods	Accuracy (%)
PIMA Dataset [13]	Mousa et al. [36]	LSTM + Mean imputation	85.00
	Rajni et al. [37]	RB-Bayes + Mean imputation	72.90
	Chang et al. [21]	Naïve Bayes + PCA	79.13
	Sisodia et al.[38]	Naïve Bayes + Mean imputation	76.30
	Pradhan et al. [23]	BMNABC + ODF(RFE)	77.21
	Proposed	MBMNABC-Ma + ODF(RFE)	**80.98**

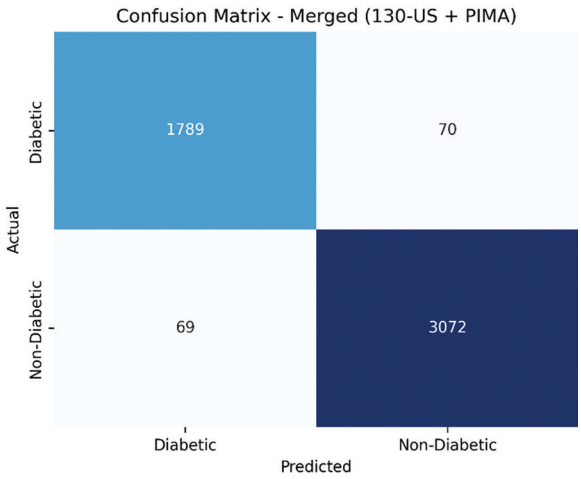


Fig 7. Confusion matrix for the Merged Dataset showing the classification performance of the proposed MBMNABC-Ma+ ODF(RFE) model

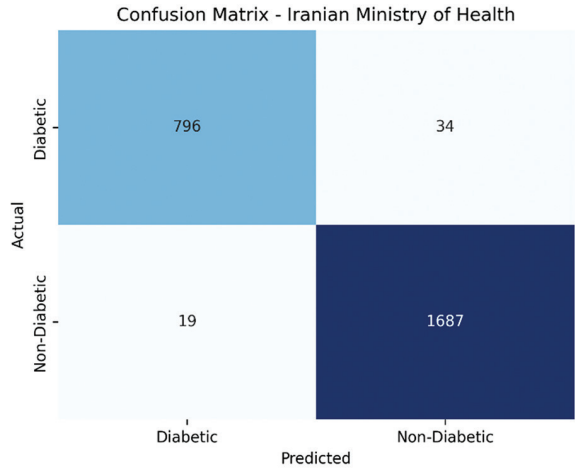


Fig 8. Confusion matrix for the Iranian Ministry of Health Dataset showing the classification performance of the proposed MBMNABC-Ma+ ODF(RFE) model

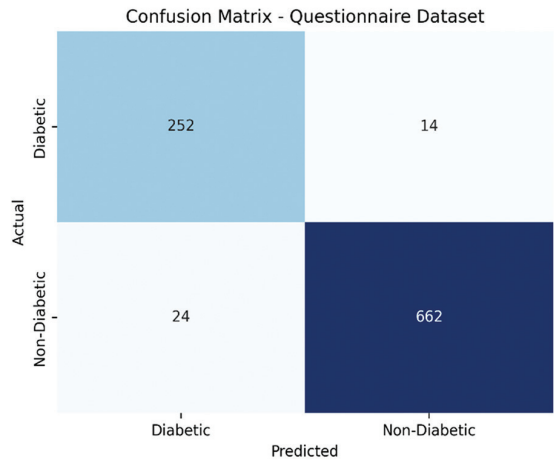


Fig 9. Confusion matrix model for the Questionnaire Dataset showing the classification performance of the proposed MBMNABC-Ma + ODF(RFE) model

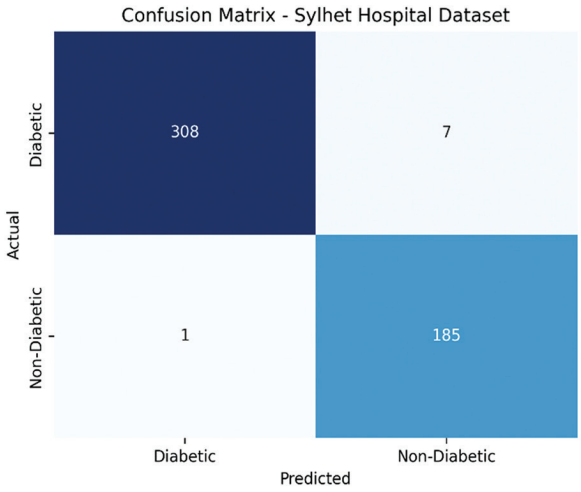


Fig 10. Confusion matrix for the Sylhet Diabetes Hospital Dataset showing the classification performance of the proposed MBMNABC-Ma + ODF(RFE) model

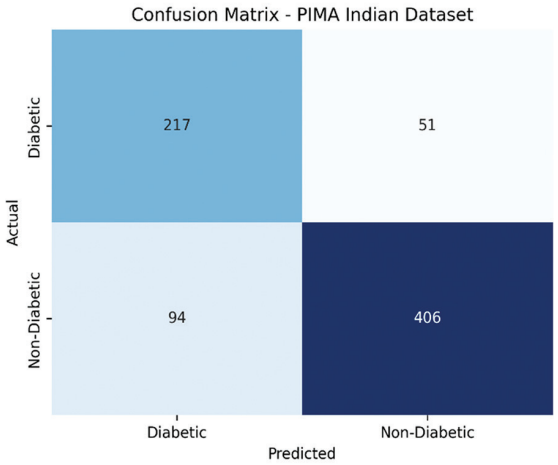


Fig 11. Confusion matrix model for the PIMA Dataset showing the classification performance of the proposed MBMNABC-Ma + ODF(RFE) model

4. CONCLUSION

Diabetes remains a significant public health concern, particularly among adults and elderly individuals, where early detection plays a vital role in reducing the risk of severe complications. This study explored the effectiveness of a novel meta-heuristic feature selection approach for diabetes detection by leveraging five diverse datasets containing a rich set of clinical and demographic variables. Through comprehensive experimentation, the proposed MBMNABC-Ma combined with the Optimized Decision Forest (RFE) framework demonstrated superior performance in terms of accuracy, sensitivity, and specificity compared to traditional methods. These findings not only confirm the robustness of the proposed approach but also underscore its potential for practical implementation in real-world clinical settings.

Moreover, this research emphasizes the importance of advanced feature selection techniques in improving

model precision and reducing redundancy. Looking ahead, future work may incorporate explainability techniques such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) to better understand the predictions made by complex models like ODF, thereby enhancing interpretability and trust in healthcare applications. Additionally, integrating deep learning architectures such as Convolutional Neural Networks (CNNs) may further refine detection capabilities by capturing deeper and more abstract patterns in high-dimensional medical data.

5. REFERENCE

[1] A. Ramachandran, C. Snehalatha, A. Raghavan, A. Nanditha, "Classification and Diagnosis of Diabetes", Textbook of Diabetes, Wiley, 2024, pp. 22-27.

[2] Md. A. Uddin et al. "Machine Learning Based Diabetes Detection Model for False Negative Reduction", Biomedical Materials & Devices, Vol. 2, No. 1, 2024, pp. 427-443.

[3] B. F. Wee, S. Sivakumar, K. H. Lim, W. K. Wong, F. H. Juwono, "Diabetes Detection Based on Machine Learning and Deep Learning Approaches", Multimedia Tools and Applications, Vol. 83, No. 8, 2023, pp. 24153-24185.

[4] L. Al Rayes, M. Haggag, I. Afyouni, "Predicting Pre-Diabetic and Diabetes in Adults and Elderlies Using Machine Learning", Proceedings of Advances in Science and Engineering Technology International Conferences, Abu Dhabi, UAE, 3-5 June 2024, pp. 1-8.

[5] S. Luhar et al. "Lifetime risk of diabetes in metropolitan cities in India", Diabetologia, Vol. 64, No. 3, 2021, pp. 521-529.

[6] G. Pradhan, R. Pradhan, B. Khandelwal, "A Study on Various Machine Learning Algorithms Used for Prediction of Diabetes Mellitus", Proceeding of the International Conference on Computing and Communication, 2020, pp. 553-561.

[7] R. R. A. Bourne et al. "Causes of vision loss worldwide, 1990-2010: A systematic analysis", Lancet Glob Health, Vol. 1, No. 6, 2013, pp. e339-e349.

[8] M. T. Moghaddam et al. "Predicting Diabetes in Adults: Identifying Important Features in Unbalanced Data Over a 5-year Cohort Study Using Machine Learning Algorithm", BMC Medical Research Methodology, Vol. 24, No. 1, 2024, p. 220.

[9] I. Guyon, A. Elisseeff, "An Introduction to Variable and Feature Selection", Journal of Machine Learning Research, Vol. 3, 2003, pp. 1157-1182.

[10] Y. Li, T. Li, H. Liu, "Recent Advances in Feature Selection and Its Applications", Knowledge and Information Systems, Vol. 53, No. 3, 2017, pp. 551-577.

[11] A. Negi, V. Jaiswal, "A First Attempt to Develop a Diabetes Prediction Method Based on Different Global Datasets", Proceedings of the Fourth International Conference on Parallel, Distributed and Grid Computing, Waknaghat, India, 22-24 December 2016, pp. 237-241.

[12] N. P. Tigga, S. Garg, "Prediction of Type 2 Diabetes using Machine Learning Classification Methods", Procedia Computer Science, Vol. 167, 2020, pp. 706-716.

[13] UCI Machine Learning, "Pima Indians Diabetes Database", https://www.kaggle.com/uciml/pima-indians-diabetes-database (accessed: 2025)

[14] B. Strack et al. "Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records", BioMed Research International, Vol. 2014, No. 1, 2014, pp. 1-11.

[15] M. Heydari, M. Teimouri, Z. Heshmati, S. M. Alavinia, "Comparison of Various Classification Algorithms in The Diagnosis of Type 2 Diabetes in Iran", International Journal of Diabetes in Developing Countries, Vol. 36, No. 2, 2016, pp. 167-173.

[16] M. M. F. Islam, R. Ferdousi, S. Rahman, H. Y. Bushra, "Likelihood Prediction of Diabetes at Early Stage Using Data Mining Techniques", Proceedings of the International Symposium on Computer Vision and Machine Intelligence in Medical Image Analysis, 2019, pp. 113-125.

[17] M. F. Dzulkalnine, R. Sallehuddin, "Missing Data Imputation with Fuzzy Feature Selection for Diabetes Dataset", SN Applied Sciences, Vol. 1, No. 4, 2019, p. 362.

[18] O. O. Oladimeji, A. Oladimeji, O. Oladimeji, "Classification Models for Likelihood Prediction of Diabetes at Early Stage Using Feature Selection", Applied Computing and Informatics, Vol. 20, No. 3/4, 2021, pp. 279-286.

[19] M. Abedini, A. Bijari, T. Banirostam, "Classification of Pima Indian Diabetes Dataset using Ensemble of Decision Tree, Logistic Regression and Neural Network", International Journal of Advanced Research in Computer and Communication Engineering, Vol. 9, No. 7, 2020, pp. 1-4.

[20] A. Iyer, S. Jeyalatha, R. Sumbaly, "Diagnosis of Diabetes Using Classification Mining Techniques", International Journal of Data Mining & Knowledge Management Process, Vol. 5, No. 1, 2015, pp. 01-14.

[21] V. Chang, J. Bailey, Q. A. Xu, Z. Sun, "Pima Indians Diabetes Mellitus Classification Based on Machine Learning (ML) Algorithms", Neural Computing and Applications, Vol. 35, No. 22, 2023, pp. 16157-16173.

[22] K. Dashdondov, S. Lee, M.-U. Erdenebat, "Enhancing Diabetes Prediction and Prevention through Mahalanobis Distance and Machine Learning Integration", Applied Sciences, Vol. 14, No. 17, 2024, p. 7480.

[23] G. Pradhan et al. "Optimized Forest Framework with A Binary Multineighborhood Artificial Bee Colony for Enhanced Diabetes Mellitus Detection", International Journal of Computational Intelligence Systems, Vol. 17, No. 1, 2024, p. 194.

[24] G. Pradhan, G. Thapa, R. Pradhan, B. Khandelwal, S. Visalakshi, "A Study on Transcontinental Diabetes Datasets Using a Soft-Voting Ensemble Learning Approach", Proceedings of Advances in Communication, Devices and Networking, 2023, pp. 87-99.

[25] S. Habibi, M. Ahmadi, S. Alizadeh, "Type 2 Diabetes Mellitus Screening and Risk Factors Using Decision Tree: Results of Data Mining", Global Journal of Health Science, Vol. 7, No. 5, 2015, pp. 304-310.

[26] S. Gündoğdu, "Efficient Prediction of Early-Stage Diabetes Using XGBoost Classifier with Random Forest Feature Selection Technique", Multimedia Tools and Applications, Vol. 82, No. 22, 2023, pp. 34163-34181.

[27] S. Buyrukoğlu, A. Akbaş, "Machine Learning based Early Prediction of Type 2 Diabetes: A New Hybrid Feature Selection Approach using Correlation Matrix with Heatmap and SFS", Balkan Journal of Electrical and Computer Engineering, Vol. 10, No. 2, 2022, pp. 110-117.

[28] N. Nipa, M. H. Riyad, S. Satu, Walliullah, K. C. Howlader, M. A. Moni, "Clinically adaptable machine learning model to identify early appreciable features of diabetes", Intelligent Medicine, Vol. 4, No. 1, 2024, pp. 22-32.

[29] B. P. Kumar, "Diabetes Prediction and Comparative Analysis Using Machine Learning Algorithms", International Research Journal of Modernization in Engineering Technology and Science, Vol. 04, No. 05, 2022, pp. 1-9.

[30] A. Yasar, "Data Classification of Early-Stage Diabetes Risk Prediction Datasets and Analysis of Algorithm Performance Using Feature Extraction Methods and Machine Learning Techniques", International Journal of Intelligent Systems and Applications in Engineering, Vol. 9, No. 4, 2021, pp. 273-281.

[31] J. Ma, "Machine Learning in Predicting Diabetes in the Early Stage", Proceedings of the 2nd International Conference on Machine Learning, Big Data and Business Intelligence, Taiyuan, China, 23-25 October 2020, pp. 167-172.

[32] A. Saboor, A. U. Rehman, T. M. Ali, S. Javaid, A. Nawaz, "An Applied Artificial Intelligence Technique For Early Prediction of Diabetes Disease", International Conference on Latest Trends in Electrical Engineering and Computing Technologies, 2022, pp. 1-6.

[33] S. N. Elsadek, L. S. Alshehri, R. A. Alqhatani, Z. A. Algarni, L. O. Elbadry, E. A. Alyahyan, "Early Prediction of Diabetes Disease Based on Data Mining Techniques", International Conference on Computational Intelligence in Data Science, Vol. 611, 2021, pp. 40-51.

[34] Nurjahan, M. A. T. Rony, Md. S. Satu, M. Whaiduzzaman, "Mining Significant Features of Diabetes through Employing Various Classification Methods", Proceedings of the International Conference on Information and Communication Technology for Sustainable Development, Dhaka, Bangladesh, 27-28 February 2021, pp. 240-244.

[35] S. M. M. Hasan, Md. F. Rabbi, A. I. Champa, Md. A. Zaman, "A Machine Learning-Based Model for Early Stage Detection of Diabetes", Proceedings of the International Conference on Computer and

Information Technology, Dhaka, Bangladesh, 19-21 December 2020, pp. 1-6.

[36] A. Mousa, W. Mustafa, R. B. Marqas, S. H. M. Mohammed, "A Comparative Study of Diabetes Detection Using The Pima Indian Diabetes Database", The Journal of University of Duhok, Vol. 26, No. 2, 2023, pp. 277-288.

[37] R. Rajni, A. Amandeep, "RB-Bayes Algorithm for The Prediction of Diabetic in Pima Indian Dataset", International Journal of Electrical and Computer Engineering, Vol. 9, No. 6, 2019, p. 4866.

[38] D. Sisodia, D. S. Sisodia, "Prediction of Diabetes using Classification Algorithms", Procedia Computer Science, Vol. 132, 2018, pp. 1578-1585.

Enhancing Cold-Start Recommendations with Content-Based Profiles and Latent Factor Models

Original Scientific Paper

Amritha P*
Kannur University,
Department of Information Technology,
Kannur University, Kannur, Kerala, India
amritha@chintech.ac.in

Rajkumar K K
Kannur University,
Department of Information Technology,
Kannur University, Kannur, Kerala, India
rajkumarkk@kannuruniv.ac.in

*Corresponding author

Abstract *– Recommendation systems have become an important tool for enhancing personalized recommendations across various domains. However, these systems face challenges, including the cold start problem, data sparsity, etc. In this paper, we present a novel recommendation model that integrates content-based and collaborative approaches to overcome these challenges. The proposed model uses TF-IDF vectorization over multiple item attributes to compute content similarity scores, and the SVD collaborative model captures latent user-item interactions. To further strengthen user preferences, a time-aware exponential decay function is used to acquire the most recent user preferences during the construction of user profiles for content-based prediction. Finally, the rating prediction is generated through a weighted fusion of content and collaborative models. Compared to benchmark models, our approach reduces RMSE by 3.06% and MAE by 3.23%, demonstrating an improvement in prediction accuracy. Furthermore, our method shows stable performance, with only a slight increase in prediction error (MAE with 8% and RMSE with 1.5% with a hybrid weight of 0.5) under cold-start conditions, indicating that the proposed method maintains strong stability and robustness even in data sparsity scenarios.*

Keywords: *recommendation, collaborative, hybrid, content-based*

Received: June 18, 2025; Received in revised form: October 1, 2025; Accepted: October 2, 2025

1. INTRODUCTION

The amount of data over the Internet has increased dramatically in the past few years due to the rapid advancement of information technology. Although the Internet offers more accessibility to users, it also creates the issue of "information overload" [1]. This is a big challenge for consumers to easily and precisely identify the necessary information among the enormous volume of data. Recommendation systems (RS), or recommender systems, are essential tools for consumers to find required personalized details from the internet. In recent years, academics and industry have focused on recommendation systems, which effectively help alleviate the problem of information overload [1, 2].

User-specific collections generated by recommendation systems make exploring the internet a satisfying experience for customers. Recommendation systems consist of three primary categories: content-based methods, collaborative filtering approaches, and hybrid models that incorporate both approaches [3]. Content-based filtering creates recommendations by analyzing the features or metadata of items, such as genre, keywords, or descriptions, and aligning them with the known preferences of users [3]. By identifying patterns and relationships among users and items, collaborative filtering can recommend items. Collaborative filtering is the most commonly employed algorithm in recommendation systems. Based on past behaviour, it predicts user preferences and generates customized rec-

ommendations [1]. Hybrid recommendation systems merge collaborative and content-based systems to provide better suggestions [4]. Recommendation systems typically produce two kinds of outputs: (a) a list of the top N recommended things, and (b) a numerical prediction for a user or set of users.

Recommendation systems help to reduce information overload by giving personalized suggestions, but they still face many challenges [1, 5]. One of the major issues among them is the data sparsity problem, which is a situation where the available user-item interaction data is sparse. This can occur when there are many users and items in the system, but each user has not interacted or given feedback. Another challenge to be addressed is the cold-start problem, which arises when newly introduced items are not getting listed in the recommended item list. Therefore, researchers have been trying to improve these algorithms and explore other techniques and methods to solve these problems and improve the effectiveness, accuracy, and user satisfaction of recommendation systems. The dynamic interests of customers also affect the efficiency of recommendations. For example, people may have long-term interests and short-term interests based on different contexts. Traditional recommendation methods cannot address these kinds of problems. Addressing these challenges requires innovative approaches, including hybrid models that combine different recommendation techniques, the incorporation of contextual and real-time data, and the development of mechanisms to enhance privacy, fairness, and transparency in recommendation processes [2, 3]. Content-based techniques focus on analyzing the item attributes that a user has interacted with and provide suggestions based on similarities between the user's preferences and item features. However, content-based methods do not incorporate user behaviour data, which means they cannot adapt to users' changing preferences. At the same time, collaborative filtering methods use user behaviour data, such as ratings, clicks, and other types of interactions [5].

So incorporating auxiliary data, contextual data, temporal features and user preferences is essential for enhancing the content-based methods, while collaborative approaches are vulnerable to data sparsity since many users do not consistently provide ratings. Traditional methods have their advantages and limitations, which vary depending on the application context. To exploit the strengths of both content-based and collaborative filtering while addressing these shortcomings, hybrid strategies with additional attributes are essential for tailoring recommendations to customers [4, 5].

The main contributions of this study are summarized as follows:

1. This study introduces a hybrid recommendation model that integrates content-based and collaborative filtering techniques to effectively address cold-start challenges, including user cold-start and item cold-start.
2. Our method uses the vectorization of multiple item features in the content-based component to make meaningful predictions, even in the absence of historical user-item interactions. The collaborative part uses Singular Value Decomposition (SVD) to uncover hidden patterns in the user-item rating matrix.
3. These methods incorporate an additional time-aware exponential decay function to capture the item's timestamp feature, which is used for preparing the user profile. This allowed the system to accord greater emphasis to more recently rated items, enhancing the relevance of the user profile preferences.

The following sections of this article are organized as follows: Section 2 provides an introduction about recommendation systems and their types. In Section 3, we discuss the related works on recommendation systems. In Section 4, we outline our proposed recommendation model and methodology. Finally, in the last part, we present the findings of our experimental results and the implications of our study.

2. RECOMMENDATION ALGORITHMS

The two main sections of recommendation systems are content filtering and collaborative filtering. Each uses a different set of techniques to offer items to the user. Collaborative filtering uses user-item interactions or ratings to discover correlations between customers and items. Content-based filtering, in contrast, looks at the characteristics of the items and recommends items similar to what the user has liked before, relying mainly on the description and features of the items [3].

2.1. COLLABORATIVE-BASED RECOMMENDATION

The fundamental input for the collaborative techniques is a user-item rating matrix. The recommendation system assumes that users will have the same preferences in the future if they liked or interacted with similar items. Two major categories of collaborative filtering are memory-based and model-based.

The three basic phases of memory-based collaborative recommendation systems are: 1) calculating similarity; 2) identifying nearest neighbours among similar users/items; and 3) making predictions. Two important subfields of memory-based collaborative filtering recommendation systems are the item-based and user-based approaches. Model-based techniques use mathematical models to learn hidden features of the user-item rating matrix that represent users and items in a lower-dimensional space. These models extract latent features from the user interaction matrix to identify underlying patterns in the data [2, 3]. One of the popular models based on the collaborative filtering method is singular value decomposition [6].

The SVD is a widely recognized matrix factorization (MF) technique in recommendation systems. Its main purpose is to reduce the dimensionality of the user-item rating matrix while preserving important relationships. The notion of SVD involves reducing the dimensionality of the original matrix through its factorization into smaller, low-rank matrices, thereby capturing latent relationships. Consider a matrix A of size m x n is transformed into: U with size m x f, $\sum$ with size f x f, and V with size f x n, as shown in Fig. 1 [7, 8].

In recommendation systems, SVD approximates the rating matrix by decomposing it into two lower-rank matrices, P and Q. The matrix P corresponds to user-feature interactions derived by U x $\sum$, where $\sum$ is treated as a scalar value to preserve the dimensionality. The matrix Q represents item-feature interactions and is equivalent to V. The dot product of P and Q estimates the rating a user might assign to an unseen item [9]. Therefore, the SVD-based matrix factorization formulation can be expressed using the user-item rating matrix $\mathbf{R}$ as follows:

$$\mathbf{R}_{m\times n} \approx \mathbf{P}_{m\times f}(\mathbf{Q}_{n\times f})^T \quad (1)$$

Where $R_{m\,x\,m}$ is the user-item rating matrix. $P_{m\,x\,f}$ denote user latent matrix (users × latent factors). And $Q_{f\,x\,n}$ denote the Item latent matrix (items × latent factors). Fig. 2. depicted SVD-based decomposition of a user-item interaction/rating matrix. In this rating matrix, the blank cells represent missing ratings or values. A higher number of such blank cells indicates greater data sparsity [9]. Matrices P and Q are derived by factorizing the user-item rating matrix R. The matrix P denote a user-latent matrix, and matrix Q depicts an item-latent matrix. These two matrices are of m x f and f x n respectively. Here m denotes the number of users, n refers to the number of items, and f represents the number of latent factors obtained during matrix decomposition.

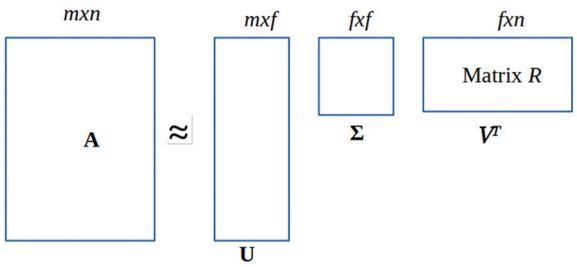


Fig. 1. SVD decomposition

The number of latent factors can be selected based on the required model complexity, as they help to uncover hidden patterns and interactions between users and items [8, 9]. The main advantage of the SVD method in recommendation systems is that it overcomes data sparsity and scalability problems. However, applying SVD directly to collaborative filtering can be problematic due to the presence of many missing entries in the user-item rating matrix. To perform matrix factorization, these missing values are often filled with some default values. Regularized SVD that works through iteration is called matrix factorization [9].

2.2. CONTENT-BASED (CB) RECOMMENDATION

The content-based approach includes a metadata of item characteristics and a user profile which contains the user's historical interests. The central task of this recommendation system is identifying items that closely match with the user's individual preferences [5]. Unlike collaborative methods, content-based recommendation depends on the inherent qualities of items and the user's preferences [5]. The Term Frequency-Inverse Document Frequency (TF-IDF), binary encoding, categorical encoding method, or the frequency encoding method are some of the techniques used to process the textual data in item descriptions in content-based recommendation systems [10, 11]. This model considers only the information provided by the target user and the features of the rated items for predicting the recommendation. Content-based algorithms use user preferences for items and suggest similar ones based on a domain-specific understanding of the item's content [12]. The initial step for a content-based model is textual data from item descriptions processed with a Vector Space Model (VSM) [12]. Below is a list of all the steps involved in the CB recommendation model process:

1. Initially each item's textual feature string is vectorized using TF-IDF to produce a sparse, high-dimensional feature representation.
2. In the second step, these vectors are used to compute *cosine similarity*, which measures the closeness between items. To personalize recommendations, construct a *user profile vector* by taking a weighted average of the TF-IDF vectors of the item the user has rated, where the weights are the actual rating values [8].
3. There are two ways to compute the predicted rating. a) Item-based similarity approach: For a given user and a target object/item, the system takes into account the items that the user has previously rated. It calculates the weighted average of those ratings, where each weight is the cosine similarity across the target item and a previously rated item. b) User profile method: The predicted rating for a new item is computed as the cosine similarity between the user's profile vector, which is constructed from the TF-IDF vectors of items they have rated, and the TF-IDF vector of the target item [8].

Let us take an example. Each movie's content description is formed by concatenating its title, genres, and release year (binned by decade). That is, the movie Toy Story (1995), with the genres Animation and Comedy, becomes the string "toy story animation comedy 90s". This is depicted in Table 1. These textual feature strings are transformed into numerical vectors using the TF-IDF method, which captures the importance of each

term within the corpus. Once vectorized, the cosine similarity is computed between every pair of movies, resulting in a similarity matrix as in Table 2. This matrix reflects how semantically close each pair of movies is based on their textual features. These cosine similarity values are later used for generating content-based recommendations either by comparing item-item relationships or by a user profile vector method.

Table 1. Content Representation Used for TF-IDF Encoding

Movie	Title	Genres	Year	TF-IDF Content String
A	"Toy Story"	"Animation, Comedy"	1995	"toy story animation comedy 90s"
B	"The Lion King"	"Animation, Adventure"	1994	"lion king animation adventure 90s"
C	"Aladdin"	"Animation, Fantasy"	1992	Aladdin animation fantasy 90s"

Table 2. Cosine Similarity Matrix Based on TF-IDF Vectors

	Toy Story	The Lion King	Aladdin
Toy Story	1.000	0.189	0.221
The Lion King	0.189	1.000	0.221
Aladdin	0.221	0.221	1.000

Hybrid recommendation merges various strategies such as content-based and collaborative filtering. This approach allows for more personalized and accurate recommendations for users with diverse preferences and behaviours. The main advantage of hybrid recommendation systems is their ability to overcome the weaknesses of individual recommendation methods [3].

3. RELATED WORK

The development of hybrid recommendation systems has improved in recent years as researchers attempt to overcome the challenges of sparsity, cold start, popularity bias, and evolving user preferences. Traditional collaborative filtering methods are good at capturing latent user–item interactions but get worse in sparse conditions, while content-based filtering can handle new items but struggles with limited attributes and user personalization. To address these shortcomings, numerous hybridization strategies have been proposed, each introducing novel mechanisms but also new trade-offs in complexity, scalability, and effectiveness.

One way to strengthen CF under sparsity is by integrating memory based Nearest-Neighbour model with model based collaborative filtering. Lv et al. [13] proposed a hybrid recommendation algorithm that integrates a User-Nearest-Neighbour (UNN) model with collaborative filtering techniques. The novelty of this approach lies in using the UNN model to fill missing user–item interactions with a weighted similarity metric. After this step, the collaborative filtering methods called ALS, MLP, and NCF are applied on the optimized matrix. The key advantage of this method is that it can reduce sparsity and enhance accuracy in sparse situations. It uses the Spark distributed platform to make it scalable. However, the model is less effective when users have few co-interactions and does not address the item cold-start since it ignores content features [13]. Similarly, Guan et al. [14] came up with an advanced similarity computation with a Wasserstein-distance-based CF, integrating anti-popularity and anti-prominence terms to reduce bias. The main advantage of this work is in its ability to handle sparse datasets and its evaluation across different metrics. However, the method incurs higher computational cost due to the Wasserstein distance calculation and similarity-based CF without explicit incorporation of temporal or content information [14].

Another important direction is adaptive segmentation and neighborhood personalization. Liang et al. [15] proposed a behavior-aware hybrid recommendation framework that separates users into two groups: active groups and inactive groups. For the inactive users, the method designed a fusion algorithm that integrates SVD with content-based filtering, which improved the accuracy measures on the MovieLens dataset. For the active users, the method applied a diversity-enhanced KNN algorithm, which reduced accuracy but increased item coverage, thereby enhancing diversity. The positive aspects of this work is its explicit user-group differentiation, ensuring a solution for sparsity. While it balances accuracy and diversity, it does not explicitly address the cold-start problem, since new users and items are not the primary focus of the framework [15].

Roy et al. [7] took an alternate approach, proposing a weighted hybrid model that combined Adaptive KNN (AKNN) and SVD. AKNN used a hybrid similarity measure that integrates cosine similarity, Pearson correlation, and Variance Mean Difference (VMD). The SVD model is used to capture latent user and item factors through matrix factorization. This hybrid similarity metric captures user–item relationships more effectively than single-measure approaches. The final prediction is generated by optimally weighting the outputs of the AKNN and SVD components, creating a weighted hybrid model. The challenge here is that the dynamic adjustment of the number of neighbours may lead to inconsistent model behaviours for users with sparse or dense user-item interactions [7].

Researchers have also turned toward clustering and multi-stage learning to capture richer similarity patterns. Sourabh et al. [16] presented a hybrid recommendation approach which uses an improved singular value decomposition applied to perform matrix factorization and a content-driven k-Nearest Neighbours model that utilized cosine similarity to identify similarities between movies based on their descriptions, year of release, and user ratings. To find the neighbours, the model used the improved kernel self-organization map with the EISEN cosine correlation distance, which helps reduce cluster

overlap. Additionally, K-means clustering is used to categorize movies, where the silhouette method determines the optimal number of clusters. However, this multi-stage method increases system complexity and computational overhead during both training and prediction phases [16].

Ensemble learning has also been introduced to balance weaknesses in individual recommendation models. Ensemble learning combines multiple models either homogeneous or heterogeneous. Singh et al. [17] integrated content-based filtering, collaborative filtering, and supervised learning models with boosting algorithms. One of the best things about this work is its use of boosting to reduce individual model weaknesses. However, the system introduces added complexity due to multiple model training stages, and it does not explicitly address issues such as cold-start scenarios [17]. In a similar way Behera et al. [18] combined matrix factorization with XGBoost, feeding latent factors and contextual attributes into the boosting model. The technique captured nonlinear relationships effectively, though the computational cost remained high and cold-start challenges were not explicitly been solved [18].

Zhi-Toung et al. [19] came up with domain-specific applications by designing a hybrid recommender integrated with KNN and SVD for food recommendation, successfully uniting memory-based and model-based filtering. However, the absence of temporal and contextual features such as dietary preferences, location, or time-of-day limited its personalization capacity [19]. Explicit cold-start mitigation was focused by Juliet et al. [20] who proposed a hybrid recommendation approach to address the cold-start problem. The method integrates collaborative filtering and content-based filtering through an adaptive weighting scheme. So when rating data is sparse, the algorithm relies more heavily on content similarity; when richer in contexts, collaborative information dominates. This fusion approach outperformed traditional collaborative filtering and content filtering methods. The strength of this work lies in its explicit focus on cold-start mitigation and the use of an adaptive hybridization mechanism rather than fixed weights. The limitation is that the content-based component is simpler and does not incorporate multiple item attributes [20].

Recent advances used deep and meta-learning to improve recommendation. Liu et al. [21] proposed a hybrid model that combines a meta-learning module with an attention module to address the cold-start challenge in recommendation systems. The attention module focuses on learning personalized user interests by assigning weights to different user–item interactions. This ensures that only informative preferences have a greater contribution to the recommendation process. The meta-learning module uses Model-Agnostic Meta-Learning (MAML) to train the recommendation model in tasks, where each task corresponds to a user's preference estimation. This helps the recommendation system adapt quickly to cold-start situations. The strengths of this approach are its ability to model personalized user interests and to generalise effectively in cold-start scenarios. The disadvantages of this approach is that, the combination of attention and meta-learning increases model complexity in sparse datasets [21].

The above studies indicate that hybrid recommendation systems outperform traditional approaches by addressing sparsity and cold-start challenges. But every approach has its advantages and limitations. One research gap across the literature is the limited integration of temporal dynamics and multi-attribute content modeling, both of which are critical for capturing changing user preferences and supporting new items.

4. PROPOSED METHOD

In this section, we will first discuss our proposed model Hybrid Content and Singular Value Decomposition (HCSVD) in detail. The proposed hybrid model integrates collaborative filtering and content-based filtering to capture user preferences and item semantics effectively. This approach is better at handling cold-start problems and data sparsity by utilizing both user-item interactions and multiple content attributes. Our proposed model consists of four stages: the data pre-processing layer, the item similarity prediction layer with multiple attributes, the SVD prediction layer, and finally the hybrid prediction layer. The architecture diagram of the proposed method is shown in Fig. 3.

4.1. PRELIMINARIES

Let $U = \{u_1, u_2, ..., u_m\}$ represent the set of users, $I = \{i_1, i_2, ..., i_n\}$ represent the set of items, R be the user-item rating matrix, r_{ui} is the rating of user u for item i, and $\hat{r}_{ui}$ denote the predicted rating. Table 3. shows the notation used in our proposed approach. The goal of the proposed approach is to predict $\hat{r}_{ui}$ as accurately as possible, especially in both the normal and the cold-start settings, using the proposed hybrid model.

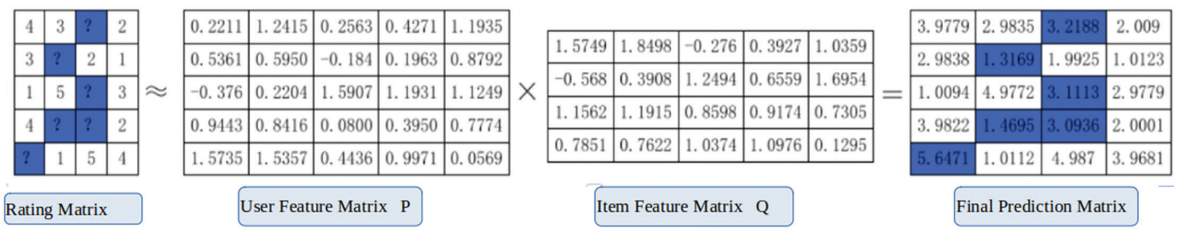


Fig. 2. SVD-based decomposition of Rating matrix R

Table 3. Notations used in the proposed hybrid system

Notation	Description
$U = \{u1, u2, \ldots, um\}$	Set of users
$I = \{i1, i2, \ldots, in\}$	Set of items
r_{ui}	Actual rating
$\mathbf{p}u, \mathbf{q}i$	Latent factor vectors
x_j	TF-IDF vector of item j (content)
$\mathbf{v}_u$	Content-based user profile vector
$\beta \in [0, 1]$	Hybrid Weight
w_j	Weight based on recency
e_{ui}	Rating prediction error
μ	Training set average rating
b_u	Bias value for user u
b_i	Bias value for item i
w_j	Weight based on recency
α	Learning rate

4.2. DATA PRE-PROCESSING

The first stage of this proposed method is data pre-processing, where the user rating matrix and item metadata are collected and extracted from the dataset. To prepare item metadata for content-based recommendation, a structured pre-processing step is applied to transform raw categorical and textual features into a format suitable for vectorization. For example, the attributes, such as item titles, category labels, and timestamp information, were extracted and cleaned. The temporal features were separated into categorical bins to capture historical trends and cold start information. Textual features such as titles were normalized by removing non-informative patterns. Categorical features, including multi-label attributes, were converted into descriptive string formats. These processed components were then concatenated to form a unified content string for each item. This text representation captures semantic, categorical, and temporal characteristics of the items. So here the metadata of the item is represented as a vector from the item metadata using an appropriate vector space model.

4.3. ITEM SIMILARITY COMPUTATION AND PREDICTION

Here the combined strings are vectorized using the TF-IDF method, resulting in a high-dimensional sparse vector. The next step is to construct a personalized user profile vector. The user profile vector is constructed by aggregating the TF-IDF vectors of the item a user has rated, weighted by the corresponding rating values. The user profile vector captures the user's preferences across multiple content dimensions.

The following equation represents the personalized user profile vector:

$$\mathbf{v}_u = \frac{1}{\sum_{j \in I_u} r_{uj}} \sum_{j \in I_u} r_{uj} \cdot \mathbf{x}_j \tag{2}$$

where V_u represent the content-based user profile vector, r_{uj} is the rating given by user u to item j. x_j is the content item vector of item j. I_u denote the set of items rated by user u.

Apart from additional user attributes, we have also included a time decay function to handle users' evolving interests as well as to get preferences for cold start items. We know that the user preference may change over time. To enhance the adaptability of the content-based recommendation system and better reflect users' evolving interests, a time-decay function is integrated into the process of constructing user profiles. Each item rated by a user contributes to the construction of their user profile based on both how much they liked it and how recent it is.

A time-aware exponential decay weight is applied to each item, defined as w_j. Hence updated user profile vector is represented as:

$$\mathbf{v}'_u = \frac{\sum_{j \in I_u} w_j \cdot r_{uj} \cdot \mathbf{x}_j}{\sum_{j \in I_u} w_j \cdot r_{uj}} \tag{3}$$

Where v'_u is the updated content-based user profile vector, r_{uj} is the rating given by user u to item j. x_j is the content item vector of item j. I_u denote the set of items rated by user u.

The time-aware weight based on recency w_j is calculated as:

$$w_j = \exp\left(-\lambda(t_{\text{now}} - t_j)\right) \tag{4}$$

Where λ denote decay rate, t_{now} represent the current time and t_j express the timestamp when item j interacted with the user. The exponential function exp(x) represent e^x where $e \approx 2.718$.

The content-based prediction is computed as the cosine similarity between the user's profile vector and the movie's TF-IDF vector, scaled to the original rating range. The content-based prediction equation is represented as below:

$$\hat{r}_{ui} = R_{\max} \cdot \cos\left(\mathbf{v}'_u, \mathbf{x}_i\right) \tag{5}$$

Where $\hat{r}_{ui}$ denote the predicted rating, R_{max} is the the maximum possible rating range. $\cos(v'_u, x_i)$ denote cosine similarity between the user profile vector and item feature vector.

$$\hat{r}_{ui} = R_{\max} \cdot \frac{\mathbf{v}'^{\top}_u \mathbf{x}_i}{\|\mathbf{v}'_u\| \cdot \|\mathbf{x}_i\|} \tag{6}$$

Where $\hat{r}_{ui}$ is the rating prediction, v'_u is the updated content-based user profile vector, x_i indicate the content item feature vector and R_{max} is the maximum possible rating range.

4.4. SVD BASED PREDICTION

For collaborative filtering, we are using the SVD-based MF algorithm, which models latent user and

item factors learnt from historical rating data. The SVD model is used to capture latent user and item factors through matrix factorization. The changes in the user-item ratings are often influenced by user and item specific biases. We can see that a few users always give better ratings to all items, while another group may give average ratings to items. At the same time, a majority of the users provide true ratings too. To take care of these deviations, the SVD-based rating prediction formula always incorporates bias terms that represent the global average rating, individual user bias, and item bias. This adjustment helps improve the accuracy of predictions by normalizing user behaviour and item popularity.

By considering this, SVD predictive formula for ratings as:

$$\hat{r}_{ui} = \mu + b_u + b_i + \mathbf{p}_u^\top \mathbf{q}_i \tag{7}$$

Where $\hat{r}_{ui}$ is the rating prediction, μ denote the average rating, b_u and b_i implies user bias and item bias values. $\boldsymbol{p}_u$ and $\boldsymbol{q}_i$ represent the latent factor vector for user and item.

The SVD objective function for the rating prediction is represented as follows:

$$\min_{\mathbf{u},\mathbf{i}\in D} \sum (r_{ui} - \hat{r}_{ui})^2 + \gamma \left(\|\mathbf{p}_u\|^2 + \|\mathbf{q}_i\|^2 + b_u^2 + b_i^2 \right) \tag{8}$$

Where set D indicate the users and items set, γ depict regularization parameter to prevent over-fitting. $\boldsymbol{p}_u$ and $\boldsymbol{q}_i$ represent the latent factor vector for user and item, b_i and b_u implies user bias and item bias values.

The error formula for updating the bias value is as below:

$$e_{ui} = r_{ui} - \mu - b_u - b_i - \mathbf{p}_u^\top \mathbf{q}_i \tag{9}$$

Here e_{ui} indicate a difference between the expected and actual values, r_{ui} indicate the actual rating. $\boldsymbol{p}_u$ and $\boldsymbol{q}_i$ represent the latent factor vector for user and item, b_u and b_i implies user bias and item bias values.

We have used stochastic gradient descent for optimizing the result. The parameters have been updated using the stochastic gradient descent approach, as shown in equations 10 to 13.

The user bias and item bias values are updated by the equation 10 and 11 as follows:

$$\mathrm{b}_u \leftarrow b_u + \alpha(e_{ui} - \gamma b_u) \tag{10}$$

$$\mathrm{b}_i \leftarrow b_i + \alpha(e_{ui} - \gamma b_i) \tag{11}$$

The latent vectors are updated by the following equations:

$$\mathrm{p}_u \leftarrow \mathbf{p}_u + \alpha(e_{ui} \cdot \mathbf{q}_i - \gamma \mathbf{p}_u) \tag{12}$$

$$\mathrm{q}_i \leftarrow \mathbf{q}_i + \alpha(e_{ui} \cdot \mathbf{p}_u - \gamma \mathbf{q}_i) \tag{13}$$

Where α denotes the learning rate, e_{ui} indicate a difference between the expected and actual values, $\boldsymbol{p}_u$ and $\boldsymbol{q}_i$ represent the latent factor vector for user and item, b_u and b_i implies user bias and item bias values.

4.5. HYBRID PREDICTION & RECOMMENDATION

In the last stage, the final hybrid prediction is computed as a linear combination of the SVD and content-based predictions, controlled by a weighting parameter β. This fusion approach uses the strengths of content-based filtering with multiple attributes based on metadata, while SVD models latent user-item interactions from historical rating predictions. The content filtering also uses a time decay function to handle users' changing interests. Here we are using a weighting parameter $\beta \in [0,1]$ to control the contribution of each component. The final predicted rating is computed using the following equation:

$$\hat{r}_{ui}^{\text{Hybrid}} = \beta \cdot \hat{r}_{ui}^{\text{SVD}} + (1 - \beta) \cdot \hat{r}_{ui}^{\text{CE}} \tag{14}$$

where $\hat{r}_{ui}^{\text{SVD}}$ is the collaborative filtering predicted score and $\hat{r}_{ui}^{\text{CE}}$ is the content-based predicted score. This hybrid formulation allows the system to balance the two approaches, improving accuracy and cold-start scenarios.

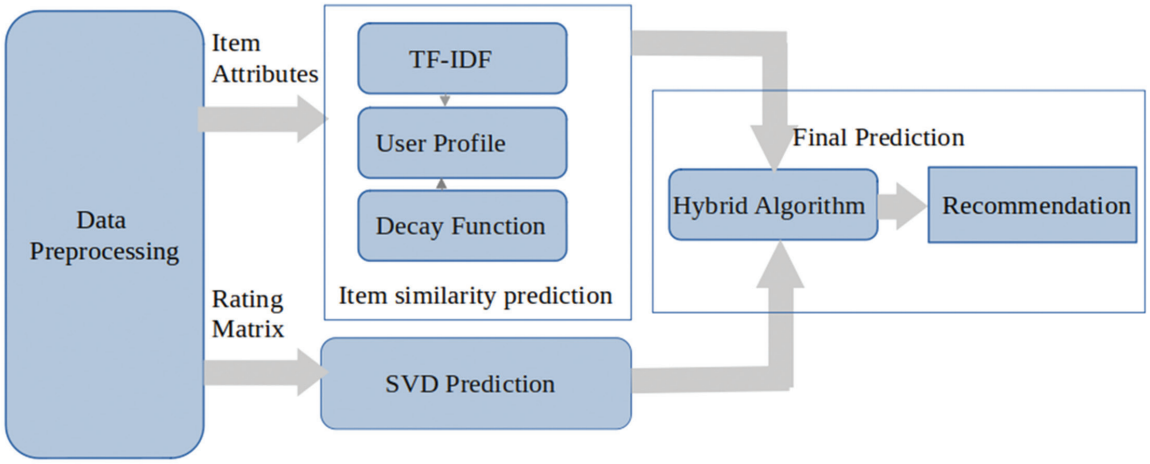


Fig. 3. Architecture diagram of HCSVD

5. EXPERIMENTS AND RESULTS

All the experiments were performed on a machine with Ubuntu 22.04 LTS powered by an 11th Gen Intel® Core™ i5-11260H CPU @ 2.60GHz×12. Python 3.9.15 is the programming language for the experimentation, and the programming environment was a Jupyter Notebook environment through Anaconda Navigator 2.4.3. The proposed method was implemented using the Scikit-Surprise library in Python [22]. Python tools like NumPy 1.23.4, Pandas 1.4.3, and Scikit-learn 1.1.1 were used for pre-processing data, evaluating metrics, and doing essential tasks. We used 75% of the dataset for training and the remaining 25% for testing. To evaluate the model's effectiveness under cold-start conditions, we conducted two controlled experiments: one for user cold-start and another for item cold-start. To test user cold start, we used users who had very few ratings, while for the item cold start case, we utilized items that had no ratings. In SVD prediction, the learning rate used is 0.01, and the number of epochs is 50. The computed time-aware weights lie in the continuous range of 0 to 1. The hybrid weight values range from 0.1 to 0.9.

5.1. DATASET

To implement the recommendation algorithms, we use two Movie-Lens datasets: ML-100K and ML-1M [23]. The Movie-Lens 100K consists of 100,000 records, where 944 users have rated 1683 items. The maximum rating given is 5, and each user has rated a minimum of 20 movies. The Movie-Lens 1M dataset consists of 1,000,209 ratings from 6,040 users on 3,952 movies. Additional metadata about the movies is available in the dataset fields, such as movie titles, release dates, and 19 types of genre vectors, with each genre represented as a binary indicator. This enables content-based methods to exploit rich categorical data. Table 4 provides a detailed description of the dataset.

5.2. BASELINE MODELS

The following baseline recommendation methods are used to compare the performance of our proposed model. They are content-based (CB) filtering, user-based KNN (UKNN) [4], item-based KNN (IKNN) [4], SVD [25], non-negative matrix factorization (NMF) [26], probabilistic matrix factorization (PMF) [27], HCFMR [18] and AMeLU [21].

CB [5, 8]: In the content-based recommendation approach, the recommendations are generated by the correlation between item attributes and the target user's profile. Each item is represented using a Vector Space Model (VSM), where features are transformed into high-dimensional vectors.

UKNN [4, 24]: This memory-based model represents users and items in a user-item rating matrix. It uses correlation-based similarity computation models, like Pearson correlation, cosine similarity, and adjusted cosine similarity, to calculate user-to-user correlations. A prediction function is then applied to generate recommendations based on these computed similarities [24].

IKNN [4, 24]: This memory-based model treats users and items as vectors in a user-item rating as a user interaction matrix. It employs correlation-based similarity computation models such as Pearson correlation, cosine similarity, and adjusted cosine similarity to determine item-to-item correlations. Recommendations are generated using a prediction function based on these calculated similarities [25].

SVD [24, 25]: Within the framework of recommendation engines, SVD can be applied to decompose the user-item interaction matrix, where users and items are represented as vectors. The resulting decomposition captures the underlying structure of the data, allowing for more accurate prediction of user preferences and generating recommendations [25].

NMF [24, 26]: Non-negative Matrix Factorization is a dimensionality reduction method that decomposes a non-negative matrix into two lower-rank matrices. NMF ensures the condition that every element within the matrices must be non-negative. In the domain of recommendation systems, NMF can be utilized to determine latent factors that represent user preferences and item traits, thereby allowing accurate rating estimates [24].

PMF [24, 27]: Recommendation employ probabilistic matrix factorization, to look for latent components that explain observed ratings by treating them as samples from a Gaussian distribution. The user-rating matrix between the user and the item in PMF is split into two lower-dimensional matrices, one for the user factors and one for the item factors [24].

HCFMR [18]: This study employs a Hybrid Collaborative Filtering with a Multi-Relation Reasoning Movie Recommendation Approach, which integrates collaborative filtering with content-based techniques. In this work, two integrated modules are used: one module learns latent user-item factors using a matrix factorization method, while another leverages item content to compute content-based similarities.

AMeLU [21]: Attentional Meta-Learned User Preference Estimator is a recommendation model for cold-start scenarios that fuses meta-learning with an attention mechanism to capture various types of user interests.

5.3. EVALUATION PARAMETERS

In our work, we utilized performance metrics mean absolute error (MAE) and root mean squared error (RMSE) to evaluate the prediction accuracy of the recommendation system [24, 28]. MAE is a popular metric for calculating the recommendation prediction. The following equation is used to compute MAE:

$$\mathrm{MAE} = \frac{\sum_{(u,i)\in T} |r_{ui} - \hat{r}_{ui}|}{|T|} \qquad (15)$$

Where r_{ui} is the actual rating and $\hat{r}_{ui}$ is the predicted rating. T is the set of all user-item pairs in the test set. RMSE can be computed using the following equation:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{(u,i)\in T} \left(r_{ui} - \hat{r}_{ui}\right)^2}{|T|}} \qquad (16)$$

Where r_{ui} is the actual rating and $\hat{r}_{ui}$ is the predicted rating. T is the set of all user-item pairs in the test set.

5.4. RESULTS AND DISCUSSIONS

We have chosen the following benchmark models for our hybrid approach to evaluate performance: the content-based model CB, memory-based collaborative models UKNN and IKNN, the model-based approaches NMF, PMF and SVD, and the hybrid models HCFMR and AMeLU. Table 5 shows a performance comparison of our proposed model against the baseline models. This table shows that HCSVD achieves the lowest error rate with RMSE and MAE values of 0.8552 & 0.6745 on MovieLens 100K and 0.8451 & 0.6686 on MovieLens 1M, outperforming all baseline methods. While hybrid models such as HCFMR and AMeLU have shown better results than standalone methods, they still fall short of HCSVD. It is evident that the hybrid models outperform individual recommendation methods. We observed that the models NMF and PMF show significantly higher error rates, with NMF reaching an RMSE of 0.9671 and MAE of 0.8110 on the Movie-Lens 100K dataset. KNN-based methods, such as IKNN and UKNN, perform moderately better but still fall short of the proposed hybrid method. During the evaluation process, we observed that CB and IKNN individually display nearly identical error values. The SVD model demonstrated superior performance in capturing latent user–item interactions, achieving lower prediction errors compared to other benchmark models. However, hybrid models gave better results by combining both collaborative and content-based features.

We have conducted additional experiments by varying the number of recommendations to evaluate how our method performs compared to other approaches in terms of prediction accuracy. As shown in Fig. 4 and Fig. 5, our approach HCSVD performed better than all other methods across different values of number of N={10,20,30,50,80}, where N is the number of recommendations. However, we observed that prediction accuracy gradually dropped as the number of recommendations increased.

Table 4. Movie-Lens Dataset Details

Dataset	No of Users	No of Items	Total no of Ratings	Density(%)
ML-IOOK	944	1683	100000	6.37
ML-IM	6040	3706	1000209	4.47

Table 5. Analysis of the Proposed method with Baseline models

	Move-Lens 100K		Move-Lens 1M	
Model	**RMSE**	**MAE**	**RMSE**	**MAE**
NMF [27, 25]	0.9671	0.8110	0.9213	0.7986
PMF [28, 25]	0.9590	0.7980	0.9108	0.7656
CB [5, 8]	0.9571	0.7610	0.9375	0.7263
IKNN [4, 25]	0.9500	0.7631	0.9118	0.7385
UKNN [4, 25]	0.9454	0.7435	0.9107	0.7185
SVD [23, 24]	0.9076	0.7146	0.9013	0.7087
HCFMR [19]	0.8850	0.6970	0.8210	0.6291
AmeLU [21]	0.8822	0.7277	0.8756	0.7068
HC SVD	**0.8552**	**0.6745**	**0.8451**	**0.6686**

We have also tested the performance of our proposed recommendation algorithm for cold-start users/items. Here we have conducted it in two ways: 1. Measure how well the model predicts for users with few or no historical ratings; 2. Measure performance on items with no prior ratings in training. In the first case, evaluation is performed, focusing on users with limited interaction history. In this setting, cold-start users are selected by identifying those with very few ratings for training. The remaining ratings for these users are held out for testing. In the second scenario, we have evaluated the model's ability to handle new items. In this evaluation, a subset of items around 5% is randomly selected without rating and removed from the training set. These items are then included only in the test set.

To evaluate the performance of the proposed hybrid model HCSVD under item cold-start conditions, we conducted a series of experiments by adjusting the hybrid prediction parameter β. The impact of hybridization parameter β is directly applied to SVD prediction and 1-β applied to the content prediction. Tables 6 and 7 present the results for the cold-start user scenario, while Tables 8 and 9 present the results for the cold-start item scenario.

Table 6. HCSVD (Cold-Start Users) on Movie Lens 100K

Beta (β)	0.9	0.7	0.5	0.3
MAE	0.7953	0.7498	0.7212	0.6987
RMSE	0.9948	0.9412	0.8911	0.8742

Table 7. HCSVD (Cold-Start Users) on Movie Lens 1M

Beta (β)	0.9	0.7	0.5	0.3
MAE	0.7845	0.7552	0.7208	0.6948
RMSE	0.9875	0.9644	0.8669	0.8711

Table 8. HCSVD (Cold-Start Items) on Movie Lens 100K

Beta (β)	0.9	0.7	0.5	0.3
MAE	0.7934	0.7326	0.7154	0.6939
RMSE	0.9820	0.9027	0.8842	0.8625

Table 9. HCSVD (Cold-Start Items) on Movie Lens 1M

Beta (β)	0.9	0.7	0.5	0.3
MAE	0.7710	0.7578	0.7275	0.6892
RMSE	0.9175	0.8641	0.8577	0.8469

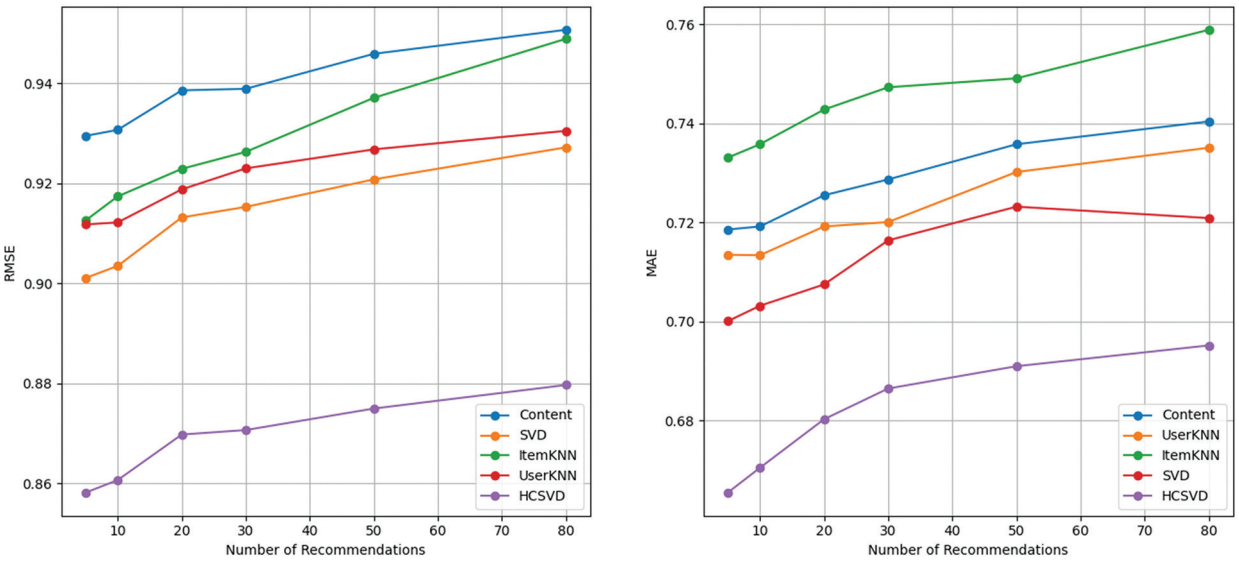


Fig. 4. Evaluation of RMSE and MAE with Varying Number of Recommendations Across Models on Movie Lens 100K

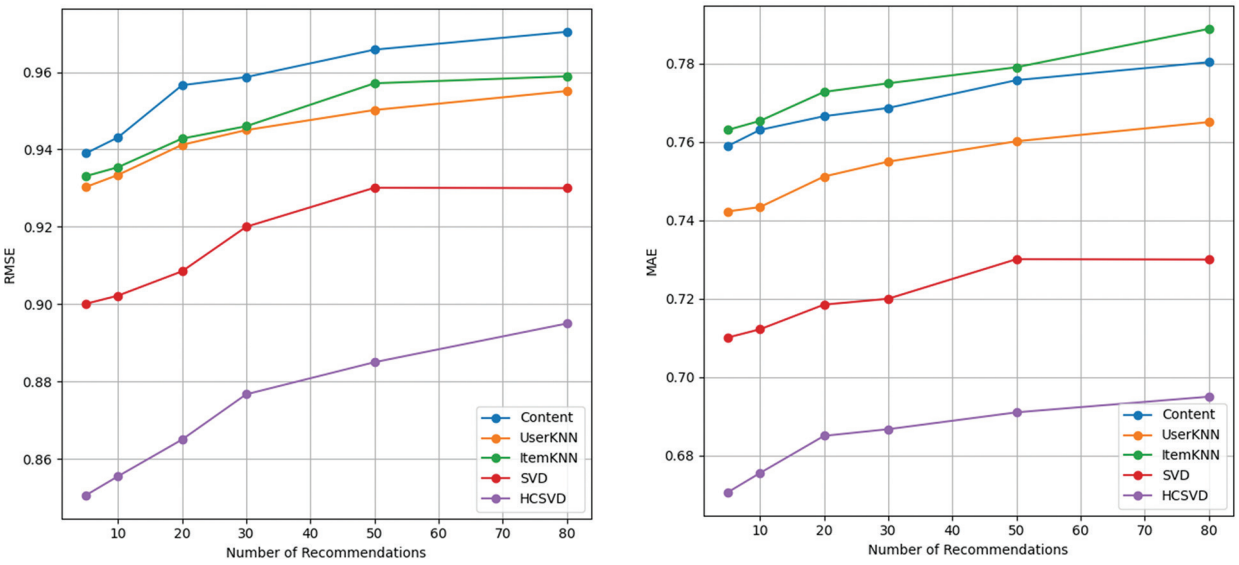


Fig. 5. Evaluation of RMSE and MAE with Varying Number of Recommendations Across Models on Movie Lens 1M

Considering the cold-start item performance, when $\beta = 0.5$ where content-based filtering and the SVD model contribute equally the system achieves an MAE of 0.7275 and RMSE of 0.8577. This demonstrates that the model maintains stable and balanced performance when both components are equally weighted. When $\beta = 0.3$, where content-based filtering contributes 70%, the model produces lowest MAE (0.6892) and RMSE (0.8469). This performance is remarkably close to the normal (non-cold-start) case, where the MAE and RMSE were 0.6686 and 0.8469, respectively. These findings are summarized in Table 9.

The fusion of singular value decomposition enables the model to capture unknown patterns in the user-item matrix. At the same time, incorporating multiple item attributes through content-based filtering enhances prediction accuracy. The hybrid model HCSVD outperforms other baselines in this setting due to its ability to depend on the content-based.

6. CONCLUSION

Currently, the recommendation of a cold-start issue remains an open subject matter, and the recommendation system continues to face a significant challenge when formulating this recommendation. In this paper, we propose a hybrid recommendation model HCSVD that combines content-based filtering and collaborative filtering to address challenges in recommendation systems, particularly cold-start scenarios. Our method utilizes the vectorization of multiple item features, as the content-based component was able to make meaningful predictions even in the scarcity of historical

user-item interactions. The method uses a time-aware exponential decay function derived from the item's timestamp feature to construct the user profile. This approach places greater emphasis on more recently rated items, thereby enhancing the relevance of the user's preference context.

Compared to benchmark models, our proposed method achieves a 3.06% reduction in RMSE and a 3.23% reduction in MAE, demonstrating its superiority in prediction accuracy. Experimental results indicate HCSVD has better performance in prediction accuracy over other benchmark models in normal and cold-start situations. In the future, we are planning to enhance our methods by integrating deep learning techniques and other innovative data representations like knowledge graphs for better recommendations. In future work, we also plan to extend the system for both rating prediction and ranking recommendations. Using both prediction and ranking evaluations will help the system to measure user satisfaction and usefulness more effectively.

7. REFERENCES:

[1] Y. Sun, Q. Liu, "Collaborative filtering recommendation based on k-nearest neighbor and non-negative matrix factorization algorithm", The Journal of Supercomputing, Vol. 81, 2024, p. 79.

[2] L. Wu, P. Sun, R. Hong, Y. Ge, M. Wang, "Collaborative neural social recommendation", IEEE Transactions on Systems, Man, and Cybernetics: Systems, Vol. 51, 2021, pp. 464-476.

[3] R. Duan, C. Jiang, H. K. Jain, "Combining review-based collaborative filtering and matrix factorization: A solution to rating's sparsity problem", Decision Support Systems, Vol. 156, 2022, p. 113748.

[4] R. Chen, Q. Hua, Y. S. Chang, B. Wang, L. Zhang, X. Kong, "A survey of collaborative filtering-based recommender systems: From traditional methods to hybrid methods based on social networks", IEEE Access, Vol. 6, 2018, pp. 76292-76326.

[5] P. Lops, M. de Gemmis, G. Semeraro, "Content-based recommender systems: State of the art and trends", Recommender Systems Handbook, Springer, 2011.

[6] Y. Koren, R. Bell, C. Volinsky, "Matrix factorization techniques for recommender systems", Computer, Vol. 42, No. 8, 2009, pp. 30-37.

[7] T. Roy, P. Shetty, "A Hybrid Approach to Predict Ratings for Book Recommendation System using Machine Learning Techniques", Proceedings of the IEEE Region 10 Symposium, New Delhi, India, 27-29 September 2024, pp. 1-6.

[8] F. Ricci, L. Rokach, B. Shapira, "Rating singular value decomposition: An enhanced matrix factorization technique for recommender systems", Recommender Systems Handbook, Springer, 2015, pp. 291-324.

[9] T. Widiyaningtyas, M. I. Ardiansyah, T. B. Adji, "Recommendation algorithm using SVD and weight point rank (SVD-WPR)", Procedia Computer Science, Vol. 161, 2019, pp. 849-856.

[10] M. J. Pazzani, D. Billsus, "Content-based recommendation systems", The Adaptive Web, Lecture Notes in Computer Science, Vol. 4321, Springer, 2007.

[11] G. Salton, M. J. McGill, "Introduction to Modern Information Retrieval", McGraw-Hill, 1983.

[12] F. Ricci, L. Rokach, B. Shapira, P. B. Kantor, "Recommender systems handbook", Springer, 2015.

[13] S. Lv, J. Wang, F. Deng, Y. Li, Y. Zhang, "A hybrid recommendation algorithm based on user nearest neighbor model", Scientific Reports, Vol. 14, 2024, p. 17119.

[14] J. Guan, B. Chen, S. Yu, "A hybrid similarity model for mitigating the cold-start problem of collaborative filtering in sparse data", Expert Systems with Applications, Vol. 242, 2024, p. 123700.

[15] A. Liang, Y. Bai, M. Wu, J. Wu, G. Wu, "Research on personalized recommendation algorithms for different user behaviors", Proceedings of the 4th International Conference on Big Data, Artificial Intelligence and Risk Management, Guangdong China, 15-17 November 2024, pp. 163-169.

[16] S. Sharma, H. K. Shakya, "Hybrid recommendation system for movies using artificial neural network", Expert Systems with Applications, Vol. 258, 2024, p. 125194.

[17] K. Singh, S. Dhawan, N. Bali, "An ensemble learning hybrid recommendation system using content-based, collaborative filtering, supervised learning and boosting algorithms", International Journal of Electrical and Computer Engineering, Vol. 13, No. 5, 2023, pp. 5599-5608.

[18] G. Behera, S. K. Panda, M.-Y. Hsieh, K.-C. Li, "Hybrid collaborative filtering using matrix factorization and XGBoost for movie recommendation", Electronics, Vol. 13, No. 8, 2024, p. 1490.

[19] Z.-T. Yap, S.-C. Haw, N. Ruslan, "Hybrid-based food recommender system utilizing KNN and SVD approaches", Cogent Engineering, Vol. 11, No. 1, 2024, p. 2436125.

[20] A. N. M. Juliet, "An improved hybrid recommendation system algorithm for resolving the cold-start issues", Journal of Information Systems Engineering & Management, Vol. 10, No. 2, 2025, pp. 243-250.

[21] S. Liu, Y. Liu, X. Zhang, C. Xu, J. He, Y. Qi, "Improving the performance of cold-start recommendation by fusion of attention network and meta-learning", Applied Sciences, Vol. 13, No. 2, 2023, p. 1120.

[22] N. Hug, "Surprise: A Python library for recommender systems", Journal of Open Source Software, Vol. 5, 2020, p. 2174.

[23] F. M. Harper, J. A. Konstan, "The MovieLens Datasets: History and Context", ACM Transactions on Interactive Intelligent Systems, Vol. 5, No. 4, 2015, pp. 1-19.

[24] C. Wenga, M. Fansi, S. Chabrier, J.-M. Mari, A. Gabillon, "A comprehensive review on non-neural networks collaborative filtering recommendation systems", Journal of Machine Learning Theory, Applications and Practice, Vol. 1, No. 1, 2023, pp. 1-44.

[25] X. Zhou, J. He, G. Huang, Y. Zhang, "SVD-based incremental approaches for recommender systems", Journal of Computer and System Sciences, Vol. 81, 2015, pp. 717-733.

[26] X. Zhang, X. Zhou, L. Chen, "Explainable recommendations with nonnegative matrix factorization", Artificial Intelligence Review, Vol. 56, 2023, pp. 3927-3955.

[27] A. Mnih, R. Salakhutdinov, "Probabilistic matrix factorization", Advances in Neural Information Processing Systems, Curran Associates, Inc., 2008, pp. 1257-1264.

[28] W. Zhu, "Statistical parameters for assessing environmental model performance related to sample size: Case study in ocean color remote sensing", Remote Sensing of Environment, Vol. 280, 2022, p. 13179.

Impact of Ammonia (NH_3) on the Energy Production in Photovoltaic Panels

Original Scientific Paper

Diego Rigoberto Aguiar
Instituto Superior Tecnológico Rumiñahui,
Postgraduate School
Sangolquí, Ecuador
diego.aguiar@ister.edu.ec

Luis Daniel Andagoya-Alba*
Instituto Superior Tecnológico Rumiñahui,
Postgraduate School
Sangolquí, Ecuador
luis.andagoya@ister.edu.ec

*Corresponding author

Abstract *– The increase in energy demand, the fossil energy crisis, and the trend towards using renewable energies from sources such as the sun or wind have led to the rise in photovoltaic installations. Some of these installations are being installed on farms. While it is true that irradiation levels, location, and inclination of the panels are considered, the influence of certain gases such as ammonia (NH_3), which is present in poultry, pig and dairy farms, is not considered. The present study is carried out in a poultry farm, through the implementation of two data acquisition devices that will be located in two scenarios, the first one exposed to NH_3 levels and the other one free of the influence of this gas, the prototypes are equipped with a 100W panel to measure the power generated and determine if there is a difference in the energy production produced by the influence of ammonia. Data was obtained for ten consecutive days, in which it was determined that the power generated by the panel decreased in the scenario with ammonia compared to the prototype without of influence of this gas, proving that NH_3 influences the decrease in power generated in the solar photovoltaic panel, obtaining average losses of 5%. It is concluded that ammonia (NH_3) influences the efficiency of energy conversion in photovoltaic solar panels.*

Keywords: *Photovoltaic Panels, Ammonia Exposure, Energy Conversion Efficiency, Environmental Impact, Solar Energy*

Received: April 13, 2025; Received in revised form: August 7, 2025; Accepted: September 16, 2025

1. INTRODUCTION

The use of solar energy has increased in recent years due to its multiple benefits and applications. Photovoltaic solar energy is not only used to provide electricity to areas where access to conventional electricity is challenging and to reduce the carbon footprint, but it also serves as an energy matrix to reduce dependence on electricity generated from other sources [1].

As the adoption of photovoltaic systems grows, it is also essential to understand the factors that influence the efficiency of solar panels. While tilt and orientation are inherent to the design, other factors can reduce the efficiency or performance of the panels, such as dust, shadows, bird droppings, raindrops, and environmental pollution, among others.

The study [2] proposes that the efficiency of a photovoltaic system is reduced due to the accumulation of dust on the surface of solar panels. The power supplied by a solar panel depends on the level of irradiation reaching its solar cells, which can be affected by light shadows caused by atmospheric pollution. In more

severe cases, efficiency is impacted by hard shadows formed by the accumulation of dust particles, which interfere with the absorption of solar radiation by the photovoltaic panel.

According to [3], dust can affect power generation systems composed of photovoltaic panels as it accumulates on their surface and remains suspended in the air, preventing the panels from receiving direct solar radiation [4]. When the amount of radiation reaching the solar cells of a photovoltaic panel decreases, power loss occurs.

Addressing issues related to contamination on the surface of photovoltaic panels and how this factor reduces their energy production efficiency is carried out in the study [5]. They consider the increase in temperature of solar panels, which not only decreases the amount of energy generated but also leads to efficiency losses. They conclude that the accumulation of dirt and other contaminants leads to a decrease in power output and an increase in the operating temperature of the panel [6].

Another study was conducted on four factors affecting the efficiency of photovoltaic panels, addressing elements such as dust, water droplets, bird droppings, and partial shading [7]. These factors are analyzed both separately and together. Although air quality is initially considered in determining panel efficiency, the study focuses on performance reduction due to dust accumulation, without considering other elements of environmental pollution resulting from the use of fossil fuels and industrial waste. The study showed the levels of impact of the aforementioned factors and how each of them reduces the efficiency of photovoltaic panels by a certain percentage. However, it does not consider any type of pollutant gases.

Dust accumulation combined with other external factors is the main cause of decreased energy efficiency in photovoltaic solar panels [8]. Cement dust often compacts upon contact with water, forming a layer that completely blocks the photovoltaic solar panel from utilizing solar radiation, even when radiation levels are ideal for energy production. Therefore, dust accumulation, along with factors such as climate, the location of the photovoltaic system, the type of dust, and other parameters, contributes to the reduction in the performance or efficiency of the solar panel or photovoltaic array.

Other studies also showed that the contamination on the photovoltaic panel's surface reduces the amount of radiation received and simultaneously increases the surface temperature, leading to a decrease in energy production efficiency [9]. Depending on the level and type of contamination, the panel's performance varies. In the study, it is determined that, in addition to solar radiation and temperature, other factors reduce the performance of a photovoltaic panel, including dust, dispersed air molecules, water vapor, and other air pollutants. These elements can cause a refraction effect on sunlight, preventing the panel from receiving direct solar radiation, thereby decreasing solar irradiance on the photovoltaic panel [10]. The same study mentions that these conditions worsen when air pollutants, suspended particles, and air humidity are present, as these factors significantly contribute to the decrease in the performance or energy conversion efficiency of photovoltaic panels.

Dust is considered the primary cause of the reduction in energy captured by solar panels, resulting in a significant decrease in their energy efficiency [11]. This is partly due to suspended particles in the air preventing the solar panel cells from directly receiving solar energy. Additionally, the study demonstrates the presence of industrial dust resulting from the accumulation of environmental pollutants and fertilizers [12]. The study [13] determines that the formation of dirt on the exterior of the panel significantly reduces the electricity generated by the panels and also increases their degradation. It is estimated that there are losses between 5% and 30% in a solar generation system annually due to dirt. In the same context, [14] concludes that, in addition to the impact of temperature and wind—meteorological variables that can affect panel performance—the accumulation of dust and other factors, such as rain, lead to the formation of mud or dirt, resulting in a decrease in panel efficiency by 3.95% and 4.03%. Concludes that dust accumulation in a short period can cause a reduction of up to 6.5%, leading to a decrease in voltage and output power [15]. This is due to periodic dust accumulation. Panels exposed to dirt caused by dust tend to reduce efficiency by up to 50% [16].

Stated that when panels are exposed to the outdoors, they are susceptible to environmental factors such as dust, bird droppings, temperature, precipitation, wind speed, among others, which can vary the energy efficiency of photovoltaic panels [17]. It is indicated that the decrease in efficiency can reach up to 30% per hour. Additionally, it was discovered that humidity levels with relative values of 76% and 86% reduce the power and output efficiency of the system [18].

It was demonstrated through experimentation that the photovoltaic efficiency of the panel tends to decrease gradually as the temperature increases, even though the solar radiation level also rises [19]. This means that even with optimal solar radiation, if the temperature increases, the efficiency of the solar panel tends to decrease. Indicates that temperature and radiation directly influence the panel's efficiency, with current, voltage, and output power decreasing as temperature increases[20]. The output power of a photovoltaic panel tends to decrease by 60% to 70% when temperature and humidity combine with dust [21].

All the research and scientific articles reviewed focus on the decrease in energy efficiency and performance in photovoltaic panels due to factors such as temperature, dust, dirt, shading, and bird droppings, among others. However, not much importance is given to gases like ammonia. This study will examine the presence of ammonia effects on photovoltaic generation systems. Due to the lack of research, it is necessary to de-

termine the level of impact, especially since there are many poultry and livestock areas, as well as industries with significant potential for solar energy utilization in their locations. However, due to the insufficient information regarding the direct or indirect problems that ammonia may cause, there is no guarantee or certainty that the implementation of these renewable energy generation systems will not be affected.

2. METHODOLOGY

The methodology is based on the comparative analysis of variables such as irradiation, temperature, and NH_3 to determine whether there are alterations in the voltage, current, and power measurements of the installed prototypes.

For the acquisition and storage of data, a system based on Arduino, humidity sensors, and temperature sensors is used. The measurement of NH_3 levels and irradiation was conducted independently using separate devices. For ammonia levels, the MQ-137 sensor, an LCD display, and an Arduino UNO board were employed. For measuring irradiation levels, a data acquisition device for solar irradiance was used.

The stage of obtaining and storing data can be visualized in the graph of Fig. 1. For the processing of data from voltage, current, solar irradiation, temperature and NH_3 level sensors provided by each of the prototypes located in areas with and without influence of the NH_3, these values are obtained by a data acquisition system based on Arduino and then stored in SD memory device, to finally process the collected information using processing package such as Python.

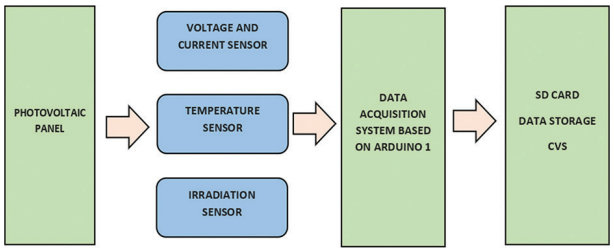


Fig. 1. Information Processing

2.1. OBTAINING AND SAVING DATA FROM THE PHOTOVOLTAIC SYSTEM.

The first stage of the project consists of obtaining experimental data in the two proposed scenarios. The data from the photovoltaic system will be subsequently analyzed using the methodology developed to determine the possible differences in the two study scenarios. This stage consists of the following processes:

Measurement of variables: By placing sensors at specific points on each of the prototypes, values of relative humidity, ammonia (NH_3), panel voltage, and panel current are obtained, and irradiance measurement is carried out separately.

Data acquisition and storage: Data acquisition is performed by means of an Arduino Uno board that processes the analog signals coming from each of the sensors and stores them on an SD card for later analysis.

2.2. PROTOTYPE OF THE DATA ACQUISITION SYSTEM.

The experimental methodology was applied in two scenarios, exposed to high levels of NH_3 and another without NH_3, at the poultry farm "El Belén," located in Pillaro-Ecuador.

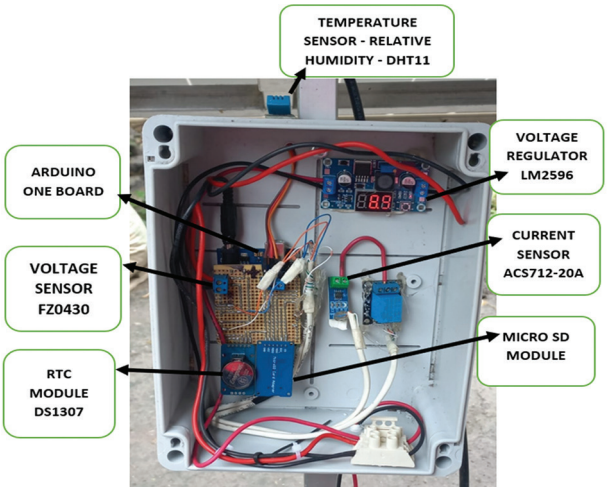


Fig. 2. Data Acquisition System Components

The data acquisition system has temperature, voltage and current sensors that allow us to measure variables such as the voltage and current consumption, in addition to the ambient temperature in the photovoltaic panel, each of the sensors send the information to an Arduino Uno card to convert the analog signals from the sensors to be subsequently analyzed in the micro-SD module. An RTC (real-time clock) module is available to take voltage, current, and temperature measurements.

These will later be analyzed with the NH_3 measurement systems and a solar irradiation measurement device to determine the differences of the mentioned variables in each analyzed scenario.

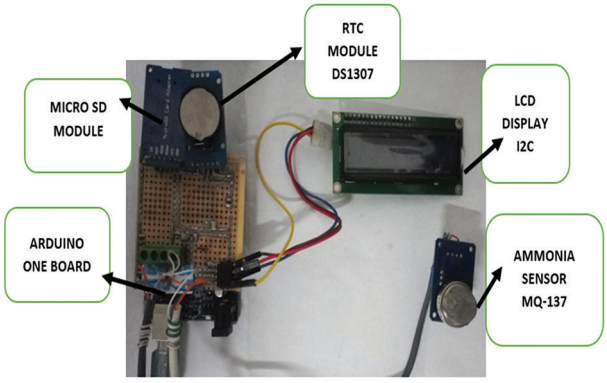


Fig. 3. Ammonia Measurement System

The ammonia NH_3 meter during the pre-stage of data acquisition allows us to determine where the highest NH_3 concentration scenarios exist. In this way, it was determined that shed #6 presents considerable and constant levels of the mentioned gas. It should be emphasized that although there were scenarios in which measurements between 5-15 ppm (parts per million) were obtained, they were not constant due to the air currents used for the ventilation of these sheds.

Using the same operating principle, the ammonia sensor was modified to store gas level readings, which are synchronized with the readings and measurements from the previous device (voltage, current, temperature). This was achieved by adding the RTC module and the Micro SD module for data storage. The collected data was later analyzed using tools such as Excel, considering that each data acquisition device generates a CSV file (comma-separated values). These files will be merged for joint analysis and used to create graphs with the Python tool in order to carry out a comparative analysis based on the generated graphs.

The system used to analyze the power generated in both scenarios consists of the following components:

100 W Photovoltaic Panel with an open-circuit voltage (Voc) of 22.28 VDC and a short-circuit current (Isc) of 6.05 A.

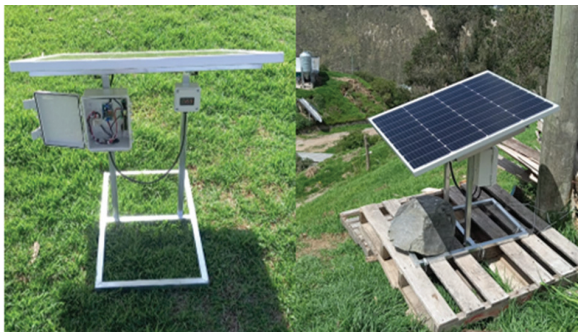


Fig. 4. Experimental Prototype

Fig. 5. Panel cleaned (left) and panel without cleaning (right)

2.3. RESULTS VISUALIZATION.

During the first stage, display elements are available to visualize values of voltage generated by the panel and percentage of ammonia in parts per million (ppm), where a voltmeter and LCD display are available to visualize the percentage of ammonia, respectively.

2.4. PHOTOVOLTAIC SYSTEM TEST PROTOTYPE.

The data acquisition prototype was tested in stages. In the first stage, measurements were made of the panels in Voc to determine through the measurements whether the solar panels generate the same amount of open-circuit voltage.

Subsequently, NH_3 levels were measured in several scenarios to determine where there was a permanent concentration of the gas and whether it was not affected by airflow or forced ventilation systems.

Fig. 6. Prototype under the influence of NH_3

Prototype 1 was located near the source of ammonia. After the measurement to determine in which area or shed there was a higher concentration of this gas, it should be noted that it was chosen considering the concentration and constant levels for experimentation and comparison for the case study.

Fig. 7. Prototype free of NH_3 exposure

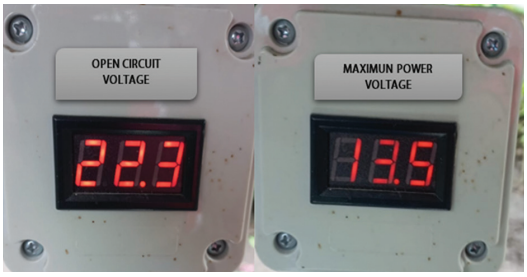
Prototype 2 was implemented away from the influence of ammonia to obtain current and voltage values, allowing for the calculation of the power generated. Then, a comparative analysis was performed to determine the decrease in power due to exposure to NH_3 in prototype 1.

Finally, the prototypes are placed in the area of influence of the mentioned gas to start data acquisition. Once corrected and the necessary modifications have

been made, a period of continuous data acquisition begins for 10 days (07-02-2025 to 16-02-2025) during which data on variables such as temperature, power generated, and irradiation are obtained.

3. RESULTS

During the first testing phase, a voltmeter was implemented to determine whether the system had a voltage signal, that is, whether it was operational.



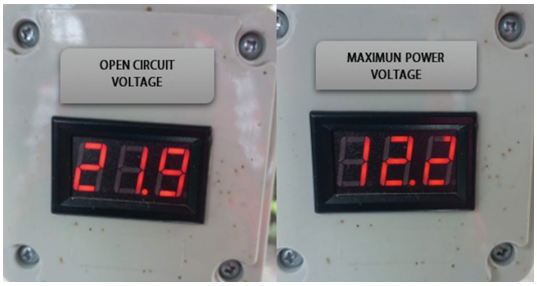
(a) (b)

Fig. 8. Voltage in the scenario without NH_3. (a) Open circuit and (b) with load

In the scenario without exposure to ammonia, a slightly higher open-circuit voltage was measured compared to the voltage recorded in the scenario with NH_3 exposure. It is important to note that initial measurements were taken under open-circuit conditions, that is, without any load connected to the photovoltaic panel.

Subsequently, measurements were repeated with a load connected. The data acquisition was carried out in a synchronized manner under identical irradiation levels, in order to determine whether the presence of ammonia had any effect on the energy conversion process. As observed, there is a voltage drop of 8.8 volts when the load is connected.

In subsequent measurements, the results obtained from the three implemented data acquisition devices were compared to correlate temperature levels (°C), ammonia concentration levels (ppm), and, based on voltage and current readings, determine the power generated by the photovoltaic panel. This data was then used to perform a comparative analysis of the results obtained from the prototype exposed to NH_3 levels.



(a) (b)

Fig. 9. Voltage in the scenario with NH_3. (a) Open circuit and (b) with load

In the scenario exposed to NH_3 levels, measurements of variables such as temperature, voltage, and current were considered under the same solar irradiation conditions as in the ammonia-free scenario. However, an increase in temperature of 1.5°C was observed compared to the 12°C recorded in the unexposed scenario. This suggests that the panel's voltage generation efficiency also decreases with temperature.

As in the previous case, open-circuit voltage measurements were first conducted to assess whether the temperature increase had any effect on the voltage without a connected load. Subsequently, a load was connected to measure the operating voltage. A drop of 0.4 volts was recorded under open-circuit conditions, and a 1.3-volt drop was observed under load conditions.

After analyzing the various parameters used to determine the impact of ammonia on photovoltaic generation systems, the correlation between NH_3 concentration levels and temperature, as well as the relationship between temperature and generated power, is considered significant. The generated power is calculated by multiplying the voltage and current values obtained from each prototype.

This study is based on experimentation using a data acquisition system to compare two scenarios under the influence of ammonia (NH_3). The project focuses on data acquisition from two scenarios: in Scenario 1, the system is exposed to average ammonia levels ranging from 3 to 5 ppm, while in Scenario 2, the prototype is either free from exposure or situated in an environment with low NH_3 levels. The objective is to determine whether the mentioned gas influences energy conversion efficiency.

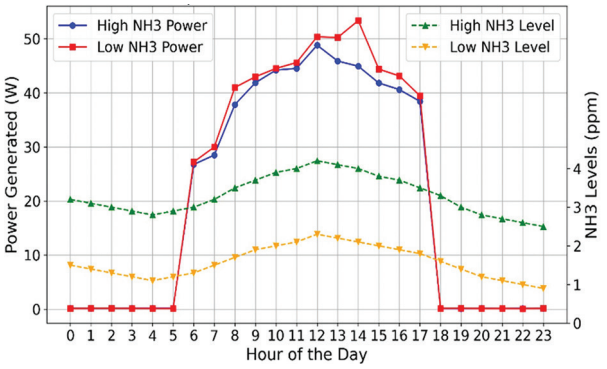


Fig.10. Generated Power vs. NH_3 Levels

According to the results, more power is generated in the scenario without exposure to NH_3, while in the scenario with exposure, there is a slight decrease in the power produced, this decrease is inversely proportional of concerning the levels of ammonia, i.e., if there is a higher concentration of ammonia level, the power produced by the solar panels is lesser.

Based on the values obtained, it is possible to determine the relationship between the power generated and the high and low NH_3 levels as shown in Fig. 6,

which, in the first instance, can be attributed to the fact that NH_3 influences the amount of power generated by the solar panels.

In a subsequent analysis, it was determined that ammonia levels are directly proportional to the measured temperature levels; therefore, due to the chemical properties of ammonia, which absorbs heat, the ambient temperature surrounding the ammonia source increases. Therefore, as the temperature of the area where the panel is located increases, the efficiency of the panel decreases as a result of the temperature increase.

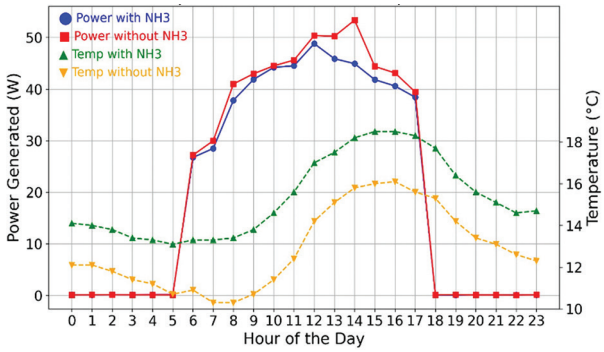


Fig. 11. Comparison of Generated Power and Temperature with and without NH_3

Fig. 11 shows the decrease in power due to the increase in temperature caused by the ammonia levels, with a minimum decrease of 2-3% of the power generated concerning the panel free of exposure with low exposure levels, with an average decrease of 5.5% and under certain conditions, a maximum decrease of 10% is reached.

The analysis of variables such as temperature, solar irradiation, voltage, current, and power enables a comparative evaluation of the measured data to determine whether there is a variation in the generated power as a function of certain variables affected by the presence of NH_3.

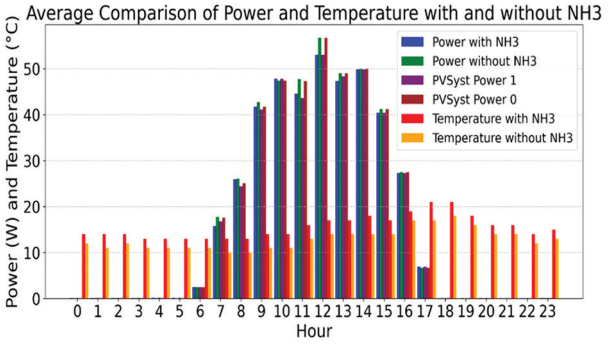


Fig. 12. Comparison of Generated Power and Temperature with and without NH_3. Measured and simulated

Fig. 12 shows a negative impact of NH_3 on the energy conversion efficiency of the photovoltaic panel. Comparing the "Power with NH_3" and "Power without NH_3" bars, it can be observed that, during peak solar irradiation hours (approximately 10:00 a.m. to 2:00 p.m.), the power generated without ammonia is higher than that generated with ammonia. This same trend is reflected in the data simulated using the PVSyst software. Examining the "PVSyst Power 0" (with NH_3) and "PVSyst Power 1" (without NH_3) bars, it is confirmed that the simulation model also predicts a power reduction when the presence of ammonia, represented by the increase in temperature, is considered.

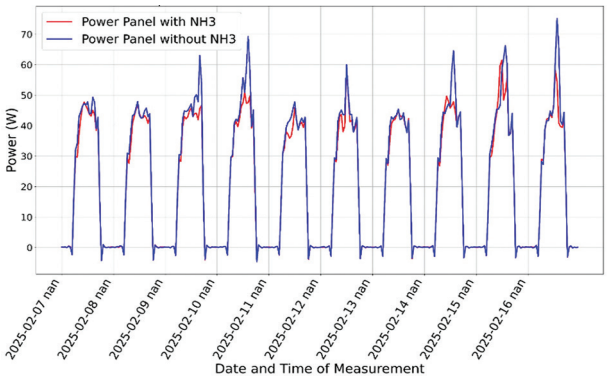


Fig. 13. Generated Power with and without NH_3

Fig. 13 shows the power generated by the panel exposed to ammonia and compared with the power produced by the prototype free of exposure, for a consecutive period of ten days, in which it can be verified that the power generated by the panel free of exposure to ammonia is slightly higher than the power generated in the photovoltaic panel exposed to levels between 3-5 parts per million (ppm) of NH_3.

Table 1. Analysis of results obtained

Day	Average Power, With NH_3(W)	Average Power, Without NH_3(W)	% Loss Power (W)
1	40.44	42.38	4.81
2	39.45	40.84	3.51
3	40.52	44.30	9.32
4	41.43	45.65	10.20
5	37.79	39.78	5.25
6	40.51	41.44	2.29
7	38.69	39.28	1.51
8	40.99	43.66	6.49
9	43.30	44.90	3.68
10	41.14	44.60	8.41
% Total Power Loss (W)			**55.52**
% Average Power Loss per day (W)			**5.55**

Based on the results obtained and their subsequent analysis, Table 1 shows the percentage loss of power generated under the influence of ammonia compared to the power produced by the panel free of exposure, determining that there is an average power loss of 5.5%, a maximum loss of 10.20% and a minimum loss of 1.51%. Although other factors may affect the performance or decrease the energy conversion efficiency, it is considered that the two prototypes are exposed to the same amount of solar irradiation. Through consecutive measurements taken over a 10-day period, a com-

parative analysis was carried out on the main variables involved in the energy conversion process, including solar irradiance, temperature, and power output. Based on this analysis, it was determined that power decreases as temperature increases (NH_3 increases).

Although other parameters, such as dust and dirt, can affect system performance, the photovoltaic panels were properly cleaned to eliminate these factors from influencing the energy conversion process. Additionally, both prototypes were installed under identical conditions in terms of location and tilt angle to ensure that no external variables would introduce inconsistencies or errors in the research results.

4. CONCLUSIONS.

The study confirms that NH_3 exposure negatively impacts the performance of photovoltaic panels. The observed efficiency losses suggest that NH_3 contamination should be considered in PV system design and maintenance, particularly in regions with significant ammonia emissions. Further research is recommended to explore mitigation strategies and protective coatings to minimize NH_3-induced efficiency degradation. Additionally, long-term studies are needed to assess the cumulative effects of NH_3 exposure over extended periods.

The implementation of data acquisition systems enables the analysis of comparative data to determine factors that influence or affect certain variables that may impact a process or phenomenon. In this case, it determines the influence of ammonia on the energy conversion efficiency of photovoltaic solar panels.

The results of this research determine the impact of ammonia (NH_3) on the energy conversion efficiency of solar panels by comparing the power generated in each of the proposed scenarios. It was found that the presence of this gas alters the ambient temperature near its source due to its characteristics, causing the temperature to rise. This increase in temperature leads to a decrease in the power generated by the photovoltaic solar panel exposed to significant levels of ammonia.

Based on the results obtained, the experimentation should be replicated on a larger scale to determine the level or degree of impact on the efficiency of a photovoltaic system. Additionally, a cloud-based data storage system should be implemented to create a database of the measured parameters for projects related to machine learning and data science.

Following the experimental tests carried out, it was determined that NH_3, present in bird feces, indirectly affects power generation in solar panels by increasing the temperature levels in the areas surrounding the source of this gas. Therefore, as the temperature rises, there is a slight decrease in the amount of power generated by photovoltaic solar panels.

5. ACKNOWLEDGMENT.

The present project is part of the research about the Impact of Ammonia (NH_3) on the Energy Conversion Efficiency of Photovoltaic Panels, which was carried out under the direction and supervision by the Postgraduate School of Instituto Superior Tecnológico Rumiñahui.

6. REFERENCES:

[1] M. R. Braga, W. N. D. Silva, A. F. L. Almeida, F. N. A. Freire, P. A. C. Rocha, "Análise Bibliométrica das Inovações em Tecnologias de Geração de Energia Solar na Base Scopus", presented at the Anais CBENS 2024, 2024.

[2] J. N. Zatsarinnaya, D. I. Amirov, L. Zemskova, "Analysis of the environmental factors influence on the efficiency of photovoltaic systems", IOP Conference Series: Materials Science and Engineering, Vol. 552, No. 1, 2019, p. 012033.

[3] K. M. Alawasa, R. S. AlAbri, A. S. Al-Hinai, M. H. Albadi, A. H. Al-Badi, "Experimental Study on the Effect of Dust Deposition on a Car Park Photovoltaic System with Different Cleaning Cycles", Sustainability, Vol. 13, No. 14, 2021, p. 7636.

[4] A. Juaidi, H. H. Muhammad, R. Abdallah, R. Abdalhaq, A. Albatayneh, and F. Kawa, "Experimental validation of dust impact on-grid connected PV system performance in Palestine: An energy nexus perspective", Energy Nexus, Vol. 6, 2022, p. 100082.

[5] D. Matusz-Kalász, I. Bodnár, "Operation Problems of Solar Panel Caused by the Surface Contamination", Energies, Vol. 14, No. 17, 2021, p. 5461.

[6] T. Rahman et al. "Investigation of Degradation of Solar Photovoltaics: A Review of Aging Factors, Impacts, and Future Directions toward Sustainable Energy Management", Energies, Vol. 16, No. 9, 2023, p. 3706.

[7] R. J. Mustafa, M. R. Gomaa, M. Al-Dhaifallah, H. Rezk, "Environmental Impacts on the Performance of Solar Photovoltaic Systems", Sustainability, Vol. 12, No. 2, 2020, p. 608.

[8] A. Aslam, N. Ahmed, S. A. Qureshi, M. Assadi, N. Ahmed, "Advances in Solar PV Systems; A Comprehensive Review of PV Performance, Influencing Factors, and Mitigation Techniques", Energies, Vol. 15, No. 20, 2022, p. 7595.

[9] P. W. Khan, Y. C. Byun, O.-R. Jeong, "A stacking ensemble classifier-based machine learning model for classifying pollution sources on photovoltaic panels", Scientific Reports, Vol. 13, No. 1, 2023, p. 10256.

[10] F. Shaik, S. S. Lingala, P. Veeraboina, "Effect of various parameters on the performance of solar PV power plant: a review and the experimental study", Sustainable Energy Research, Vol. 10, No. 1, 2023, p. 6.

[11] S. Z. Said, S. Z. Islam, N. H. Radzi, C. W. Wekesa, M. Altimania, J. Uddin, "Dust impact on solar PV performance: A critical review of optimal cleaning techniques for yield enhancement across varied environmental conditions", Energy Reports, Vol. 12, 2024, pp. 1121-1141.

[12] S. Nwokolo, A. Obiwulu, S. Amadi, J. Ogbulezie, "Assessing the Impact of Soiling, Tilt Angle, and Solar Radiation on the Performance of Solar PV Systems", Trends in Renewable Energy, Vol. 9, No. 1, 2023.

[13] B. O. Olorunfemi, N. I. Nwulu, O. A. Ogbolumani, "Solar panel surface dirt detection and removal based on arduino color recognition", MethodsX, Vol. 10, 2023, p. 101967.

[14] I. Al Siyabi, A. Al Mayasi, A. Al Shukaili, S. Khanna, "Effect of Soiling on Solar Photovoltaic Performance under Desert Climatic Conditions", Energies, Vol. 14, No. 3, 2021, p. 659.

[15] A. D. Dhass, N. Beemkumar, S. Harikrishnan, H. M. Ali, "A Review on Factors Influencing the Mismatch Losses in Solar Photovoltaic System", International Journal of Photoenergy, Vol. 2022, 2022, pp. 1-27.

[16] K. Olcay, S. G. Tunca, M. A. Özgür, "Forecasting and Performance Analysis of Energy Production in Solar Power Plants Using Long Short-Term Memory (LSTM) and Random Forest Models", IEEE Access, Vol. 12, 2024, pp. 103299-103312.

[17] M. Z. Farahmand, M. E. Nazari, S. Shamlou, M. Shafie-khah, "The Simultaneous Impacts of Seasonal Weather and Solar Conditions on PV Panels Electrical Characteristics", Energies, Vol. 14, No. 4, 2021, p. 845.

[18] A. Aldawoud, A. Aldawoud, Y. Aryanfar, M. E. H. Assad, S. Sharma, R. Alayi, "Reducing PV soiling and condensation using hydrophobic coating with brush and controllable curtains", International Journal of Low-Carbon Technologies, Vol. 17, 2022, pp. 919-930.

[19] M. K. Hassan, I. M. Alqurashi, A. E. Salama, A. F. Mohamed, "Investigation the performance of PV solar cells in extremely hot environments", J. Umm Al-Qura Univ. Eng.Archit., Vol. 13, No. 1-2, 2022, pp. 18-26.

[20] M. K. Al-Ghezi, R. T. Ahmed, M. T. Chaichan, "The Influence of Temperature and Irradiance on Performance of the photovoltaic panel in the Middle of Iraq", International Journal of Renewable Energy Development, Vol. 11, No. 2, 2022, pp. 501-513.

[21] D. Yadav et al. "Analysis of the Factors Influencing the Performance of Single- and Multi-Diode PV Solar Modules", IEEE Access, Vol. 11, 2023, pp. 95507-95525.

Fast and Accurate Design of BLDC Motors Using Bayesian Neural Networks

Original Scientific Paper

Son T. Nguyen *
Hanoi University of Science and Technology
School of Electrical and Engineering,
Faculty of Electrical Engineering
Dai Co Viet Street, Hanoi, Vietnam
son.nguyenthanh@hust.edu.vn

Trung T. Cao
Hanoi University of Science and Technology
School of Electrical and Engineering,
Faculty of Electrical Engineering
Dai Co Viet Street, Hanoi, Vietnam
trung.caothanh@hust.edu.vn

Tu M. Pham
Hanoi University of Science and Technology
School of Electrical and Engineering,
Faculty of Electrical Engineering
Dai Co Viet Street, Hanoi, Vietnam
tu.phamminh@hust.edu.vn

Tinh V. Lai
Hanoi University of Science and Technology
School of Electrical and Engineering,
Faculty of Electrical Engineering
Dai Co Viet Street, Hanoi, Vietnam
tinh.lv240456e@sis.hust.edu.vn

Anh Hoang
Hanoi University of Science and Technology
School of Electrical and Engineering,
Faculty of Electrical Engineering
Dai Co Viet Street, Hanoi, Vietnam
anh.hoang@hust.edu.vn

Hoang Q. Ha
Hanoi University of Science and Technology
School of Electrical and Engineering,
Faculty of Electrical Engineering
Dai Co Viet Street, Hanoi, Vietnam
hoang.hq240414e@sis.hust.edu.vn

*Corresponding author

Abstract *– Brushless direct current (BLDC) motors are gaining popularity over traditional direct current (DC) motors due to their higher efficiency, compact size, and precise control capabilities. This study proposes a fast and accurate approach to BLDC motor design using a Bayesian neural network (BNN). The BNN, a specialized form of the multi-layer perceptron (MLP), offers strong resistance to overfitting and performs effectively with noisy or limited datasets, making it well-suited for complex motor design problems. In the proposed method, the BNN is applied within an inverse modeling framework to map desired motor performance parameters to the corresponding design variables. A dataset for an outer-rotor BLDC motor—containing both design parameters and the resulting output torque—is generated through finite element analysis (FEA). Finally, a demonstration of BLDC motor design using the BNN validates the effectiveness of the proposed approach.*

Keywords: *BLDC motors, Bayesian neural networks, finite element analysis*

Received: March 18, 2025; Received in revised form: August 19, 2025; Accepted: October 1, 2025

1. INTRODUCTION

Brushless DC (BLDC) motors have been extensively studied in recent decades due to their high efficiency, reliability, and precise motion control capabilities. Their compact design and lightweight construction facilitate accurate speed and torque regulation, making them well-suited for modern engineering applications [1]. Unlike traditional brushed DC motors, BLDC motors employ electronic commutation instead of mechanical commutators. Consequently, they have been widely adopted in diverse fields such as industrial automation, electric vehicles, drones, medical devices, and home appliances, where precise speed control, low maintenance, and high efficiency are critical requirements.

Finite element analysis (FEA) is essential for designing and optimizing electromagnetic devices such as BLDC motors. It enables engineers to evaluate motor performance under various operating conditions, and

by simulating electromagnetic behaviors and mechanical stresses, it allows for precise design adjustments before production. This approach reduces development costs while improving efficiency, reliability, and overall performance [2].

A major challenge in BLDC motor design is cogging torque, which affects smooth operation and overall efficiency. Extensive research has been conducted to analyze and mitigate this issue. Studies indicate that factors such as stator tooth width and slot–pole alignment significantly influence cogging torque and can be optimized to enhance motor performance [3]. In outer-rotor BLDC motors, optimizing the stator core design is an effective strategy for reducing cogging torque [4], while in inner-rotor BLDC motors, segmenting the rotor's permanent magnets is commonly employed to minimize cogging effects [5]. Furthermore, field-oriented control (FOC) is an advanced technique that reduces cogging torque by incorporating dominant harmonics from the cogging torque waveform into the q-axis current reference, thereby counteracting torque ripples and minimizing speed variations [6].

BLDC motors with trapezoidal back electromotive force (BEMF) traditionally require six rotor position signals for inverter control, typically detected by Hall-effect sensors embedded in the motor. While effective, these sensors increase manufacturing costs and are sensitive to temperature variations, which can reduce system reliability. To overcome these limitations, sensorless control techniques have been extensively developed over the past two decades. A common approach estimates rotor position and regulates speed by detecting BEMF zero crossings from terminal voltages [7]; however, this method performs poorly at low speeds due to weak induced voltages. To address this issue, a novel sensorless position detection technique based on a speed-independent position function has been proposed [8], significantly improving estimation accuracy and enhancing system performance across a wide speed range.

The design of electromagnetic devices—such as electric motors, transformers, and inductors—is a complex process that requires balancing multiple performance criteria, including efficiency, thermal management, weight, and material costs [9, 10]. This challenge is often formulated as an optimization problem, where the objective is to minimize or maximize a specific cost function, such as power loss, torque ripple, or electromagnetic interference. Stochastic optimization methods, including genetic algorithms (GA) [11], particle swarm optimization (PSO) [12], and simulated annealing (SA) [13], are widely applied because of their effectiveness in exploring complex, multi-dimensional design spaces. These methods use iterative procedures to evaluate design parameters at each step, progressively refining solutions to approach an optimal configuration.

Artificial neural networks (ANNs) are transforming the design of electromagnetic devices by enhancing simulation efficiency, optimizing design parameters, and addressing complex inverse problems. These computational models can learn and represent intricate nonlinear relationships, making them especially valuable in scenarios where conventional physics-based methods are limited or computationally expensive [14, 15]. A key advantage of ANNs is their ability to process large datasets and detect patterns that traditional methods may overlook, enabling more accurate and faster performance predictions for electromagnetic devices. When integrated with FEA, ANNs facilitate more efficient and precise optimization of permanent magnet (PM) motors [16]. For example, ANNs trained on FEA-generated data can model complex electromagnetic behaviors and predict motor performance under varying conditions. This hybrid approach allows for rapid evaluation of design alternatives, significantly reducing the time and cost associated with physical prototyping and extensive simulation runs.

Bayesian neural networks (BNNs) extend traditional multi-layer perceptron (MLP) neural networks by incorporating principles of Bayesian inference [17]. Unlike conventional MLPs, which learn fixed point estimates for their weights and biases, BNNs treat these parameters as probability distributions, enabling the quantification of uncertainty in predictions. In this study, BNNs are applied to the accurate design of an outer-rotor BLDC motor. FEA was performed to generate a dataset for training the network. Once trained, the BNN computes optimal motor design parameters to achieve a specified target torque. The remainder of this paper is organized as follows: Section 2 discusses the theory of BNNs and their application to regression problems; Section 3 introduces FEA for electromagnetic devices, with emphasis on BLDC motors; Section 4 details the proposed design methodology for the outer-rotor BLDC motor using BNNs; and Section 5 presents the conclusions and outlines directions for future research.

2. BAYESIAN NEURAL NETWORKS FOR REGRESSION PROBLEMS

2.1. MULTI-LAYER PERCEPTRON NETWORKS

A MLP neural network takes a vector of real-valued inputs and computes one or more activation values for the output layer. In a network with a single hidden layer, as illustrated in Fig. 1, the activation of the output layer is calculated as follows:

$$a_k(x) = b_k + \sum_{j=1}^{m} w_{kj} \tanh\left(\bar{b}_j + \sum_{i=1}^{d} \bar{w}_{ji} x_i\right) = b_k + \sum_{j=1}^{m} w_{kj} y_j \quad (1)$$

Here, x_i are real inputs, $\bar{w}_{ji}$ is the weight on the connection from the input unit i to the hidden unit j; similarly, w_{kj} is the weight on the connection from the hidden unit j to the output unit k. The $\bar{b}_j$ and b_k are the biases of the hidden and output units. These weights and biases are the parameters of the MLP neural network. Then the activation $a_k(x)$ are used to compute the outputs of the output layer by using a "linear" activation function as follows:

$$z_k = a_k(x) \tag{2}$$

The training of an MLP neural network aims to minimize a data error function structured in the following way:

$$E_D = \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{c}\left(z_k^n - t_k^n\right)^2 \tag{3}$$

In which z_k^n is the k-th output corresponding to the n-th training pattern and t_k^n is the k-th target corresponding to the n-th training pattern.

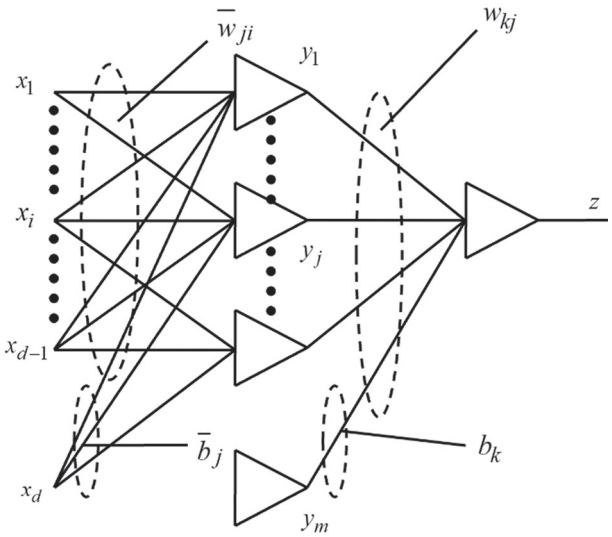


Fig. 1. Structure of MLP neural networks

2.2. NETWORK REGULARIZATION

In MLP neural networks, regularization is employed to prevent any weights and biases from becoming excessively large, as large weights and biases can lead to poor generalization on new test cases. To address this issue, a weight function, E_W, is added to the error function to penalize large weights and biases. Specifically, for regression problems, this approach is utilized to enhance model performance and a total error function, $S(w)$, is defined as follows:

$$S(w) = \beta E_D + \alpha E_W \tag{4}$$

Where β and α are non-negative parameters, also known as " hyperparameters ", need to be determinded. The weight function, E_W, usually originates from the theory of weight priors having the following form:

$$E_W = \frac{1}{2}\|w\|^2 \tag{5}$$

Where w is the vector of the weights and biases in the network.

2.3. BAYESIAN INFERENCE

In Bayesian inference for MLP neural networks, β and α can be automatically determined. This process considers the Gaussian probability distributions of the weights and biases which can give the best generalization. In particular, the vector of weights and biases, w, in the network is adjusted to their most probable values given the training data D. Specifically, the posterior distribution of the vector of weights and biases can be computed using Bayes' rule as follows:

$$p(w|D) = \frac{p(D|w)\,p(w)}{p(D)} \tag{6}$$

In this formula, $p(D|w)$ represents the likelihood function, which captures the information derived from observations. In contrast, the prior distribution $p(w)$ incorporates information based on background knowledge. The denominator, $p(D)$, known as the evidence for the network.

The requirement of small values of weights and biases in the MLP neural network suggests the use of a Gaussian prior distribution for the vector of weights and biases as follows:

$$p(w) = \frac{1}{Z_W(\alpha)}\exp\left(-\frac{\alpha}{2}\|w\|^2\right) \tag{7}$$

Where $ZW(\alpha)$ is a normalization constant given by:

$$Z_W(\alpha) = \left(\frac{2\pi}{\alpha}\right)^{W/2} \tag{8}$$

In equation (8), W is the number of weights and biases in the network. The likelihood function is given by:

$$p(D|w) = \frac{1}{Z_D(\beta)}\exp\left(-\frac{\beta}{2}E_D\right) \tag{9}$$

Where $Z_D(\beta)$ is a normalization factor given by:

$$Z_D(\beta) = \left(\frac{2\pi}{\beta}\right)^{N/2} \tag{10}$$

In equation (10), N is the number of training patterns. At the most probable vector of the weights and biases, , the Hessian matrix of the total error function, A, can be evaluated as follows:

$$A = \nabla\nabla S(w_{MP}) = \beta H + \alpha I \tag{11}$$

Where I is the identity matrix. $H=\nabla\nabla E_D(w_{MP})$ is the Hessian matrix of the data error function at the most probable vector, w_{MP} of weights and biases.

If the posterior distribution of weights and biases is assumed as a Gaussian, then it is given by:

$$p(w|D) = \frac{1}{Z_S}\exp\left(-S(w_{MP}) - \frac{1}{2}\Delta w^T A \Delta w\right) \tag{12}$$

Where $\Delta w = w - w_{MP}$ and Z_S is a normalization constant given by:

$$Z_S = \exp\left(-S(w_{MP})\right)(2\pi)^{W/2}\left(\det A\right)^{-1/2} \tag{13}$$

Re-arranging (6) gives:

$$p(D) = \frac{p(D|w)\,p(w)}{p(w|D)} \tag{14}$$

Substituting (7), (9) and (12) into (14) results in:

$$p(D)=\exp\left(-S(w_{MP})\right)\left(\left(\frac{\beta}{2\pi}\right)^{N/2}\right)\left(\alpha^{W/2}\right)\left(\det A\right)^{-1/2} \quad (15)$$

Taking the logarithm of (15) gives:

$$\ln p(D)=-S(w_{MP})+\frac{N}{2}\ln(\beta)-\frac{N}{2}\ln(2\pi)+ +\frac{W}{2}\ln(\alpha)-\frac{1}{2}\ln(\det A) \quad (16)$$

In this context, the variable *ln p* (*D*) is referred to as the "log evidence". To optimize the log evidence *ln p* (*D*) with respect to α, it is necessary to compute a partial derivative of the log evidence as follows:

$$\frac{\partial \ln p(D)}{\partial \alpha}=-E_W(w_{MP})+\frac{W}{2\alpha}-\frac{1}{2}\frac{\partial\left(\ln(\det A)\right)}{\partial \alpha} \quad (17)$$

In (17), $\frac{\partial(ln(det\,A))}{\partial \alpha}$ is computed as follows:

$$\frac{\partial\left(\ln(\det A)\right)}{\partial \alpha}=\sum_{i=1}^{W}\frac{\alpha}{\lambda_i+\alpha} \quad (18)$$

Where λ_i are the eigenvalues of the Hessian matrix of the data error function, $H=\nabla\nabla E_D(w_{MP})$. Substituting (18) into (17) gives:

$$\frac{\partial \ln p(D)}{\partial \alpha}=-E_W(w_{MP})+\frac{W}{2\alpha}-\frac{1}{2}\sum_{i=1}^{W}\frac{\alpha}{\lambda_i+\alpha} \quad (19)$$

The optimal value of α is determined when the right-hand side of equation (19) is equal to zero to obtain the following equation:

$$2\alpha E_W(w_{MP})=W-\sum_{i=1}^{W}\frac{\alpha}{\lambda_i+\alpha} \quad (20)$$

The right-hand side of equation (20) is equal to a value γ defined as follows:

$$\gamma=W-\sum_{i=1}^{W}\frac{\alpha}{\lambda_i+\alpha}=\sum_{i=1}^{W}\frac{\lambda_i}{\lambda_i+\alpha} \quad (21)$$

If $\lambda_i \gg \alpha$, γ is approximately equal to 1. Whereas, if $\lambda_i \leq \alpha$, γ is near to 0. γ is used to measure the number of "well-determined" parameters in the network.

From equations (20) and (21), α can be determined as follows:

$$\alpha=\frac{\gamma}{2E_W(w_{MP})} \quad (22)$$

Similarly, to optimize the log evidence *ln p* (*D*) with respect to β, it is also necessary to compute a partial derivative of the log evidence as follows:

$$\frac{\partial \ln p(D)}{\partial \beta}=-E_D(w_{MP})+\frac{N}{2\beta}-\frac{1}{2}\frac{\partial\left(\ln(\det A)\right)}{\partial \beta} \quad (23)$$

In (23), $\frac{\partial(ln(det\,A))}{\partial \beta}$ is computed as follows:

$$\frac{\partial\left(\ln(\det A)\right)}{\partial \beta}=\frac{1}{\beta}\sum_{i=1}^{W}\frac{\alpha}{\lambda_i+\alpha} \quad (24)$$

Substituting (24) into (23) gives:

$$\frac{\partial \ln p(D)}{\partial \beta}=-E_D(w_{MP})+\frac{N}{2\beta}-\frac{1}{2\beta}\sum_{i=1}^{W}\frac{\lambda_i}{\lambda_i+\alpha} \quad (25)$$

The optimal value of β is determined when the right-hand side of equation (25) is equal to zero to obtain the following equation:

$$2\beta E_D(w_{MP})=N-\sum_{i=1}^{W}\frac{\lambda_i}{\lambda_i+\alpha}=N-\gamma \quad (26)$$

From equation (26), β can be determined as follows:

$$\beta=\frac{N-\gamma}{2E_D(w_{MP})} \quad (27)$$

Choosing the number of input, hidden, and output nodes in a BNN for regression is similar in principle to standard neural networks, but the Bayesian framework adds a probabilistic perspective that helps control overfitting and provides uncertainty estimates.

- The number of input nodes is equal to number of features (independent variables) in dataset.
- A single output node (the predicted continuous value).
- Number of hidden nodes can be determined from heuristic starting points. For example, the number of hidden nodes can be computed by the following rule:

$$N_{hidden}=\frac{N_{in}+N_{out}}{2} \quad (28)$$

Where N_{in}, N_{out} and N_{hidden} are numbers of input nodes, number of output nodes and number of hidden nodes, respectively.

2.4. THE QUASI-NEWTON METHOD

The training of an MLP neural network involves minimizing the total error function, $S(w)$, through an iterative process. The quasi-Newton method, which extends the gradient descent method, can be effectively used to minimize this error function. In Newton method, the vector of weights and biases of the network can be updated as follows:

$$w_{m+1}=w_m-A_m^{-1}g_m \quad (29)$$

The vector $-Am^{-1}\,g_m$ is called the "Newton direction" or the "Newton step". However, the evaluation of the Hessian matrix, Am^{-1}, can be very computational. From equation (29), we can form the relationship between the weight vectors at steps *m* and *m*+1 as follows:

$$w_{m+1}=w_m-\alpha_m F_m g_m \quad (30)$$

From equation (30), if $\alpha_m=1$ and $F_m=A_m^{-1}$, we have the Newton method, while if $Fm=I$, we have the gradient descent method with the learning rate α_m.

F_m can be chosen to approximate the Hessian matrix. In addition, F_m must be positive definite so that for small α_m we can obtain a descent method. In practice, the value of α_m can be found by a "line search". Equation (30) is known as the quasi-Newton condition. The most successful method to compute F_m is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) formula as follows:

$$F_{m+1} = F_m + \frac{pp^T}{p^T v} - \frac{(F_m v)v^T F_m}{v^T F_m v} + \left(v^T F_m v\right)uu^T \quad (31)$$

Where p, v and u are defined as:

$$p = w_{m+1} - w_m \quad (32)$$

$$v = g_{m+1} - g_m \quad (33)$$

$$u = \frac{p}{p^T v} - \frac{F_m v}{v^T F_m v} \quad (34)$$

Finally, training MLP neural networks using Bayesian inference involves several key steps as follows:

Step 1: Initialize the weights and biases for the network, and initialize the values for the hyperparameters β and α.

Step 2: Minimize the cost function $S(w)$ (4) using the quasi-Newton method and calculate γ as follows:

$$\gamma_{old} = \sum_{i=1}^{W} \frac{\lambda_i}{\lambda_i + \alpha_{old}} \quad (35)$$

Where λ_i are the eigenvalues of the Hessian matrix of the data error function, $H=\nabla\nabla E_D$.

Step 3: When the cost function has reached a local minimum, re-estimate the values of the hyperparameters as follows:

$$\alpha_{new} = \frac{\gamma_{old}}{2E_W} \quad (36)$$

$$\beta_{new} = \frac{N - \gamma_{old}}{2E_D} \quad (37)$$

Step 4: Repeat steps 2 and 3 until the convergence.

3. FEA FOR BLDC MOTORS

The operating principle of many electromechanical devices is based on electromagnetic theory, and these devices can often be mathematically described using partial differential equations (PDEs). Analyzing such devices therefore requires methods for solving PDEs. Since analytical techniques often fail to provide accurate solutions, the finite element method (FEM) has emerged as a powerful tool for addressing PDEs in the analysis of electromagnetic systems. The FEA process for a specific electromagnetic device typically involves four key steps:

- Discretize the solution region into finite elements.
- Derive the governing equations for each individual element.
- Assemble all the finite elements within the solution region.
- Solve the resulting set of equations.

In this study, Finite Element Method Magnetics (FEMM), a free and open-source computational tool, is utilized to analyze the performance and characteristics of a BLDC motor. FEMM has gained popularity in both academic research and industrial applications due to its accessibility, efficiency, and capability to handle complex electromagnetic problems using FEM [18].

To set up FEMM, the problem type must first be defined, typically as a planar or axisymmetric magnetic problem with appropriate units. The geometry of the device, such as the stator, rotor, air gap, and windings, is then drawn or imported, and each region is assigned suitable material properties from the FEMM library or custom definitions. Boundary conditions, excitations (such as currents or permanent magnet properties), and circuit parameters are specified to represent the operating conditions. The model is then discretized using a finite element mesh, refined in critical regions like the air gap for higher accuracy. Finally, the analysis is run, and the postprocessor is used to visualize field distributions and extract key performance quantities such as flux linkage, torque, or inductance.

A key advantage of FEMM is its ability to integrate with MATLAB. This interaction enables users to define complex electromagnetic problems, execute simulations, and extract results programmatically within the MATLAB environment. Through MATLAB commands, researchers can automate the analysis process, perform parametric studies, and optimize motor designs more efficiently. This integration is particularly valuable for iterative design workflows, where multiple simulations are required to evaluate the impact of design parameter variations on motor performance. Fig. 2 shows a commercial outer-rotor BLDC motor, while Fig. 3 illustrates its 2D finite element analysis (FEA) conducted using FEMM.

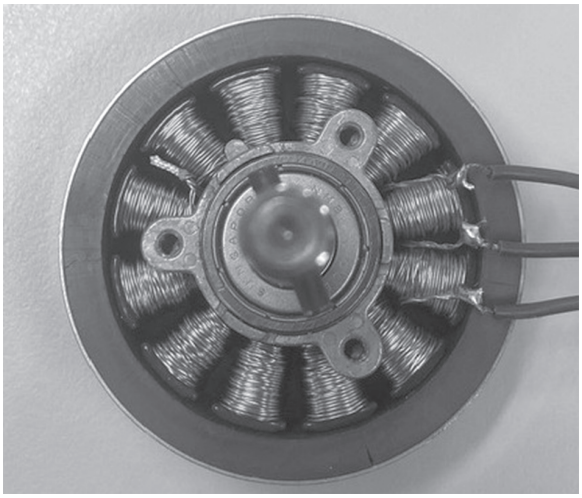


Fig. 2. Commercial outer-rotor BLDC motor

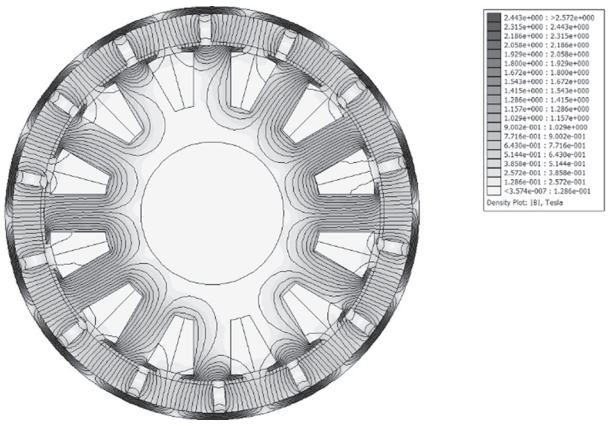


Fig. 3. 2D-FEA of a commercial outer-rotor BLDC motor

4. DESIGN OF BLDC MOTORS USING BAYESIAN NEURAL NETWORKS

This section presents a detailed procedure for designing a small-scale outer-rotor BLDC motor. The motor's dimensional and electrical parameters are as follows:

- Stator outer radius (*rso*).
- Stator inner radius (*rsi*).
- Magnet thickness (*dm*).
- Can thickness (*dc*).
- Depth of slot opening (*ds*).
- Pole fraction spanned by the magnet (*fm*).
- Pole fraction spanned by the iron (*fp*).
- Width of the tooth as a fraction of the pole pitch at the stator (*ft*).
- Back iron thickness as a fraction of tooth thickness (*fb*).
- Stator to magnet mechanical clearance (*go*).
- Axial length of the machine (*hh*).
- Peak current density in the winding (*Jpk*).

Fig. 4 illustrates the process of generating the dataset used for training the Bayesian neural network (BNN). The dataset was constructed to capture the complex nonlinear relationships between the design parameters of the BLDC motor and the corresponding output torque values. To ensure diversity and representativeness, multiple variations of key design parameters—such as stator and rotor dimensions, air gap length, winding configurations, and material properties—were systematically simulated. Each configuration was analyzed using FEA, producing torque outputs under specified operating conditions. This process resulted in a dataset consisting of 2000 distinct patterns ($N = 2000$), derived from a commercial outer-rotor BLDC motor. The large number of samples provides sufficient coverage of the design space, allowing the BNN to learn both linear and nonlinear dependencies effectively. By incorporating such a dataset, the network is better equipped to generalize across unseen configurations, thereby improving its predictive accuracy and robustness in motor design optimization. The ranges for the design parameters and output torque of the motor are provided in Table 1. Lastly, the dataset was utilized to train the BNN according to the procedure outlined in Fig. 5. The BNN has the following structure:

- An input corresponding to the desired output torque
- Six units in the hidden layer
- Twelve outputs in the output layer, representing the twelve design parameters that need to be determined

In this research, the training algorithm of the BNN is based on the quasi-Newton optimization method. This approach is chosen because it provides a balance between computational efficiency and convergence accuracy. Unlike traditional gradient descent methods, which may suffer from slow convergence or becoming trapped in local minima, quasi-Newton methods approximate the Hessian matrix of second-order derivatives to achieve faster and more stable convergence. By leveraging curvature information of the error surface, the algorithm can adjust the learning step more intelligently, thereby reducing the number of iterations required to reach an optimal solution. This makes the quasi-Newton method particularly suitable for training complex models such as BNNs, where robustness and efficiency are essential in handling high-dimensional parameter spaces and ensuring reliable generalization.

The number of hidden nodes is initially determined using a heuristic approach, guided by the sizes of the input and output layers. This provides a reasonable starting point, ensuring that the network has sufficient capacity to capture the underlying nonlinear relationships without becoming overly complex. To enable effective learning from the training data, the number of training epochs is set to 1000. This choice balances providing enough iterations for convergence with avoiding excessive training that could lead to overfitting. By combining an informed initialization of hidden nodes with an adequate number of training epochs, the BNN is structured to achieve reliable performance, accurately capturing the mapping between design parameters and output responses with both precision and stability.

Table 2 presents the variations in hyperparameters across different re-estimation periods. These adjustments result from the Bayesian optimization process, which systematically refines hyperparameters to enhance model accuracy and stability. By periodically re-estimating and updating these parameters, the BNN sustains optimal performance throughout training, leading to more accurate predictions and improved generalization to unseen data.

Once trained, the BNN acts as an effective mapping tool, translating desired output torque values into corresponding motor design parameters. Fig. 5 illustrates the relationship between target torque and the associated design variables. BNN's ability to accurately map

these relationships is crucial for optimizing BLDC motor designs, as it streamlines the design process and reduces the need for extensive trial-and-error experimentation.

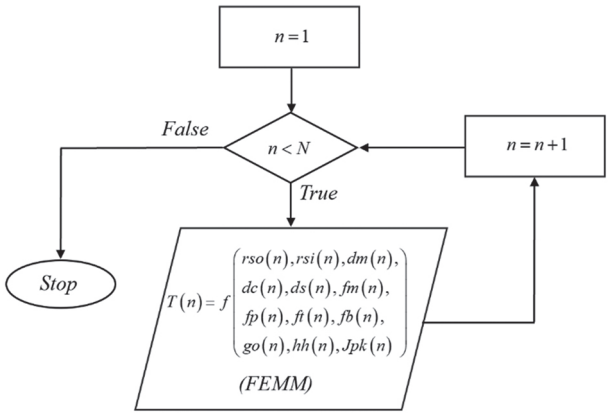


Fig. 4. Principle of generating the dataset for the BNN training

Table 1. Ranges of the design parameters and output torque of the motor

Design Parameters	Ranges
rso(mm)	[22.5004 27.4921]
rsi(mm)	[9.0005 10.9982]
dm(mm)	[3.6000 4.3998]
dc(mm)	[0.9001 1.0996]
ds(mm)	[0.4500 0.5500]
fm	[0.7715 0.9428]
fp	[0.6302 0.7700]
ft	[0.9000 1.0998]
fb	[0.9000 1.0999]
go(mm)	[0.4500 0.5499]
hh(mm)	[22.5019 27.4992]
Jpk(*MA*/m^2)	[0.7201 0.8798]
Output Torque (N.m)	[0.0470 0.2275]

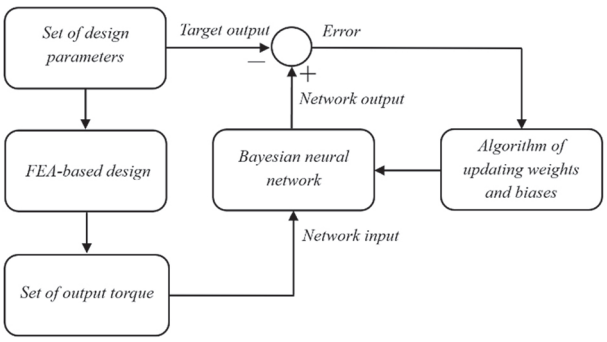


Fig. 5. Principle of updating the weights and biases of the BNN

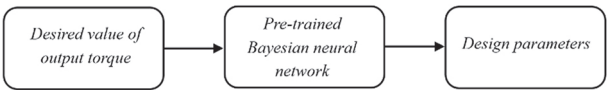


Fig. 6. Principle of calculating the design parameters of the motor using the pre-trained BNN

Table 2. Change of the hyperparameters according to the periods of re-estimation

Re-estimation Periods	α	β
1	0.3304	479.271
2	1.2190	479.4954
3	2.5027	479.3550

The final motor design parameters, obtained after completing the optimization process, are outlined in Table 3. These parameters represent the optimized configuration derived from the BNN's predictions, reflecting a balance between performance objectives and design constraints. The information in Table 3 serves as a comprehensive summary of the key design variables, providing a clear reference for evaluating the effectiveness of the BNN-driven optimization approach.

Table 3. Design parameters of the motor obtained after the design process

Parameters	Values
rso(mm)	25.1423
rsi(mm)	10.0374
dm(mm)	3.9904
dc(mm)	1.0000
ds(mm)	0.4988
fm	0.8576
fp	0.6998
ft	0.9950
fb	0.9980
go(mm)	0.4985
hh(mm)	25.0006
Jpk(*MA*/m^2)	0.8011

Table 4 compares the target output torque with the actual output torque, showing only a very small percentage difference. This minimal error confirms that the BNN can accurately predict motor performance by capturing the complex relationships between design parameters and torque. The low error rate demonstrates both high precision and strong reliability, ensuring that the predicted torque is almost identical to the desired target. Such accuracy is crucial for optimizing BLDC motor designs, as it enables improved performance, reduces the number of design iterations, and increases confidence in the model's predictions.

This research does not incorporate optimization techniques for BLDC motor design—such as cost minimization, compact dimensions, or material efficiency—into its framework. Instead, the focus is placed on developing a fast and accurate design methodology. While this approach provides valuable insights into the design process, the absence of optimization considerations limits its applicability in scenarios where economic feasibility, space constraints, or manufacturing efficiency are critical. Therefore, integrating optimization strategies in future studies would enhance the practical relevance of the proposed design method.

Table 4. Comparison between the target and true torques

Target Torque (N.m)	True Torque (N.m)	Error (%)
0.1	0.0967	3.3
0.125	0.1256	-0.48
0.175	0.1751	-0.0571
0.2	0.2018	-0.9

5. CONCLUSIONS

This study highlights the effectiveness of BNNs in optimizing the design of BLDC motors, demonstrating their potential as a powerful alternative to conventional optimization methods. To support the training process, FEA was employed to generate a comprehensive dataset that accurately captures the complex nonlinear relationships between design parameters—such as dimensions, material properties, and winding configurations—and motor performance characteristics, including torque, efficiency, and thermal behavior. This data set enables BNN to learn these intricate mappings and provide reliable performance predictions across a wide range of operating conditions.

A key contribution of this research is the integration of Bayesian optimization for hyperparameter tuning of the multi-layer perceptron (MLP) structure underlying the BNN in BLDC motor design. Unlike manual trial-and-error methods, Bayesian optimization systematically explores the hyperparameter space in a data-driven manner, leading to improved training accuracy, enhanced stability, and reduced computational cost. This approach also minimizes the risk of overfitting while significantly improving the model's ability to generalize to unseen design scenarios.

Looking forward, future research will extend the application of BNNs in BLDC motor design optimization, with a particular focus on addressing practical constraints such as reducing material costs, improving energy efficiency, and enhancing manufacturability. Moreover, incorporating multi-objective optimization—balancing trade-offs between cost, weight, torque ripple, and thermal performance—could further advance the applicability of BNN-based methods in industrial motor design.

ACKNOWLEDGEMENT

This research is funded by Hanoi University of Science and Technology (HUST) under project number T2024-PC-059.

6. REFERENCES:

[1] D. Mohanraj, R. Aruldavid, R. Verma, K. Sathiyasekar, A. B. Barnawi, B. Chokkalingam, "A Review of BLDC Motor: State of Art, Advanced Control Techniques, and Applications", IEEE Access, Vol. 10, 2022, pp. 54833-54869.

[2] T.-Y. Lee, M.-K. Seo, Y.-J. Kim, S.-Y. Jung, "Motor Design and Characteristics Comparison of Outer-Rotor-Type BLDC Motor and BLAC Motor Based on Numerical Analysis", IEEE Transactions on Applied Superconductivity, Vol. 26, No. 4, 2016, pp. 1-6.

[3] Y. L. Karnavas, I. D. Chasiotis, A. D. Gkiokas, "An Investigation Study Considering the Effect of Magnet Type, Slot Type and Pole-Arc to Pole-Pitch Ratio Variation on PM Brushless DC Motor Design", Proceedings of the 5th International Conference on Mathematics and Computers in Sciences and Industry, Corfu, Greece, 25-27 August 2018, pp. 7-13.

[4] K.-J. Han, H.-S. Cho, D.-H. Cho, H.-K. Jung, "Optimal core shape design for cogging torque reduction of brushless DC motor using genetic algorithm", IEEE Transactions on Magnetics, Vol. 36, No. 4, 2000, pp. 1927-1931.

[5] R. Setiabudy, H. Wahab, Y. S. Putra, "Reduction of cogging torque on brushless direct current motor with segmentation of magnet permanent", Proceedings of the 4th International Conference on Information Technology, Computer, and Electrical Engineering, Semarang, Indonesia, 18-19 October 2017, pp. 81-86.

[6] M. Sumega, P. Rafajdus, G. Scelba, M. Stulrajter, "Control Strategies for the Identification and Reduction of Cogging Torque in PM Motors", Proceedings of the International Conference on Electrical Drives & Power Electronics, The High Tatras, Slovakia, 24-26 September 2019, pp. 74-80.

[7] X. Song, B. Han, K. Wang, "Sensorless Drive of High-Speed BLDC Motors Based on Virtual Third-Harmonic Back EMF and High-Precision Compensation", IEEE Transactions on Power Electronics, Vol. 34, No. 9, 2019, pp. 8787-8796.

[8] T. Li, J. Zhou, "High-Stability Position-Sensorless Control Method for Brushless DC Motors at Low Speed", IEEE Transactions on Power Electronics, Vol. 34, No. 5, 2019, pp. 4895-4903.

[9] Y. Zhao, S. L. Ho, W. N. Fu, "A Novel Fast Remesh-Free Mesh Deformation Method and Its Application to Optimal Design of Electromagnetic Devices", IEEE Transactions on Magnetics, Vol. 50, No. 11, 2014.

[10] H. R. E. H. Bouchekara, "Optimal design of electromagnetic devices using a black-hole-based opti-

mization technique", IEEE Transactions on Magnetics, Vol. 49, No. 12, 2013, pp. 5709-5714.

[11] J. Gao, L. Dai, W. Zhang, "Improved genetic optimization algorithm with subdomain model for multi-objective optimal design of SPMSM", CES Transactions on Electrical Machines and Systems, Vol. 2, No. 1, 2018, pp. 160-165.

[12] J. Hwan Lee, J.-W. Kim, J.-Y. Song, D.-W. Kim, Y.-J. Kim, S.-Y. Jung, "Distance-Based Intelligent Particle Swarm Optimization for Optimal Design of Permanent Magnet Synchronous Machine", IEEE Transactions on Magnetics, Vol. 53, No. 6, 2017.

[13] T. Renyuan, S. Jianzhong, L. Yan, C. Xiang, "Optimization of electromagnetic devices by using intelligent simulated annealing algorithm", IEEE Transactions on Magnetics, Vol. 34, No. 5, 1998, pp. 2992-2995.

[14] D. Cherubini, A. Fanni, A. Montisci, P. Testoni, "Inversion of MLP neural networks for direct solution of inverse problems", IEEE Transactions on Magnetics, Vol. 41, No. 5, 2005, pp. 1784-1787.

[15] I. Marinova, C. Panchev, D. Katsakos, "A neural network inversion approach to electromagnetic device design", IEEE Transactions on Magnetics, Vol. 36, No. 4, 2000, pp. 1080-1084.

[16] L. Hadjout, N. Takorabet, R. Ibtiouen, S. Mezani, "Optimization of instantaneous torque shape of PM motors using artificial neural networks based on FE results", IEEE Transactions on Magnetics, Vol. 42, No. 4, 2006, pp. 1283-1286.

[17] S. T. Nguyen, T. M. Pham, A. Hoang, L. V. Trieu, T. T. Cao, "Bayesian Inference for Regularization and Model Complexity Control of Artificial Neural Networks in Classification Problems", Bayesian Inference-Recent Trends, IntechOpen, 2023.

[18] Finite Element Method Magnetics, https://www.femm.info/wiki/HomePage (accessed: 2025)

A Secure Data Aggregation for Clustering Routing Protocols in Heterogenous Wireless Sensor Networks

Original Scientific Paper

Basim Abood*
Department of Communication Engineering,
College of Engineering, University of Sumer,
Thi-Qar 64001, Iraq
basim.alkhafaji@uos.edu.iq

Hayder Kareem Amer
Department of Computer Technology Engineering,
College of Technical,
Imam Ja'afar Al-Sadiq University, Thi-Qar 64001, Iraq
hayder.kareem@ijsu.edu.iq

Wael Abd Alaziz
Department of Computer Information Systems,
College of Computer Science & Information Technology,
University of Sumer, Thi-Qar 64001, Iraq
w.abdalaziz@uos.edu.iq

Hussain K. Chaiel
Department of Communication Engineering,
College of Engineering, University of Sumer,
Thi-Qar 64001, Iraq
hussain.chaiel@uos.edu.iq

*Corresponding author

Abstract *– The paper presents a broadly elaborated, secure, and energy-efficient data aggregation scheme of the heterogeneous wireless sensor networks (HWSNs). This is motivated by two consistent shortcomings of existing work: (i) clustering-based routing algorithms like LEACH, SEP, and FSEP are inadequate on balancing the energy usage when there is a disparity in the node capabilities, and (ii) most ECC-based security systems create too much computation overhead to extend network lifetime. To satisfy such gaps, the given framework integrates the Spider Monkey Optimization Routing Protocol (SMORP) with a compact cryptographic implementer including the Improved Elliptic Curve Cryptography (IECC) and El Gamal Digital Signature (ELGDS) scheme. SMORP gives maximum consideration to cluster forming and multi hop forwarding and the IECC-ELGDS module that provides all the above data confidentiality, authentication and data integrity at a lower cost of computation. As compared to the previous strategies, the combination of routing optimization and elliptic-curve-based secure aggregation facilitates energy efficiency and high-security assurance in the resource-constrained nodes. MATLAB models show that the offered framework can boost network life up to 27 percent, residual energy up to 32 percent, and get a 96 percent packet-delivery ratio relative to LEACH, SEP, and FSEP. Moreover, the IECC-ELGDS module will need less time in encryption/decryption by 22-35 percent in comparison with ECC-HE, IEKC and ECDH-RSA. These findings support the idea that the SMORP-IECC-ELGDS is a viable and fast architecture to secure aggregation in the real-life HWSN deployment.*

Keywords: *Wireless sensor networks, Lifetime prolonging, Data aggregation Security, Spider Monkey Optimization, Elliptic Curve, EL Gamal Digital Signature algorithm, cryptography, Routing clustering.*

Received: October 11, 2025; Received in revised form: January 8, 2026; Accepted: January 12, 2026

1. INTRODUCTION

Typically, low cost and easy scale-up characteristics have made Wireless Sensor Networks (WSNs) a base technology in large scale environmental monitoring, automation in industrial settings and Autonomous operation in harsh environments or remote settings that do not require continuous human supervision. New applications require round-the-clock sensing, time-sensitive data streaming and unattended long-term operation, which puts intense limitations on both network lifetime and energy expenditure. The same requirements are further complicated in the heterogeneous WSNs (HWSNs) where there is differences in hardware capacity and battery resources provided by the sensor nodes, communication range, and processing power. These heterogeneous architectures facilitate more differing deployments, as well as result in drawsive mismatched energy depletion, uneven routing loads, and enhance susceptibility to communication issues [1-3]. Alongside energy constraints, security is one of the most often challenged issues in deployments of clustered WSN, as sensor nodes are frequently deployed in hostile physical conditions and they use broadcast wireless networks, similar to those used by eavesdropping, packet manipulation,

identity spoofing, replay attacks and malicious node injection. Providing multi-hop aggregation with confidentiality, authentication and integrity of the data is thus important to mission critical applications, especially where the aggregated data has a direct impact on control or situational awareness [4-6]. Nonetheless, even classical forms of public-key cryptography are computationally infeasible on the lean sensor nodes, and lightweight cryptography (elliptic-curve cryptography) and optimized digital signature designs are made use of to mitigate the impact of computation overheads and offer high levels of security assurance[7-9]. These two issues, energy efficiency and secure data aggregation, have led to more recent studies that focus on integrated solutions, which combine routing and security together, instead of focusing on them as different layers. Existing clustering-based routing schemes such as LEACH, SEP, and FSEP (introduced in [10-12]) provide strong baselines for energy-aware operation but do not incorporate end-to-end security. Similarly, modern lightweight security frameworks such as ECC-HE, IEKC, and ECDH-RSA (examined in [13-15]) respectively, improve confidentiality and authentication but do not address energy balancing or cluster-head (CH) overloading during repeated aggregation cycles. Therefore, there is a clear need for a unified framework that simultaneously ensures secure data aggregation and minimizes routing-related energy consumption across heterogeneous sensing tiers. *To address this need, this paper proposes an integrated SMORP–IECC–ELGDS framework that jointly optimizes energy-aware routing and secure ciphertext aggregation in heterogeneous wireless sensor networks.* The remainder of this paper is organized as follows. Section 2 presents the related works, covering recent advances in energy-efficient routing, lightweight cryptographic mechanisms, and integrated energy–security frameworks in heterogeneous WSNs. Section 3 describes the proposed methodology, including the enhanced SMORP-based clustering and routing process together with the integrated IECC–ELGDS security architecture for secure data aggregation. Section 4 outlines the simulation environment, the network and radio-energy models, and the performance metrics used in the evaluation. Section 5 provides a detailed discussion and analysis of the obtained results and compares the proposed framework with existing routing and security schemes. Finally, Section 6 concludes the paper and highlights prospective directions for future research.

Objectives, Contributions, and Novelty

To bridge this gap, the present work introduces a unified secure-and-energy-efficient architecture that integrates a biologically inspired optimization-based routing protocol with a lightweight hybrid cryptographic mechanism. Specifically, the study proposes a combined Spider Monkey Optimization Routing Protocol (SMORP) and Improved Elliptic Curve Cryptography with ElGamal Digital Signature (IECC–ELGDS) framework that jointly optimizes cluster formation, forwarding decisions, secure ciphertext aggregation, and authenticated delivery. The objectives of this work are threefold:

1. **Design an energy-efficient routing mechanism** capable of maintaining balanced energy consumption across heterogeneous sensor tiers through adaptive CH selection and optimized multi-hop forwarding.
2. **Develop a lightweight, secure aggregation framework** that ensures confidentiality, integrity, and authentication without imposing prohibitive computational overhead on sensor nodes.
3. **Integrate routing and security into a single operational pipeline**, eliminating the traditional separation between network-layer optimization and cryptographic protection.

The novelty of the proposed SMORP–IECC–ELGDS architecture lies in:

- The initial closely coordinated model with energy-conscious routing and hybrid lightweight security strengthening other instead of acting as separate layers.
- An aggregated workflow of ciphertexts, such that CHs are able to aggregate encrypted readings without decryption and this decreases the computational cost and removes any plaintext exposure.
- Concurrent engineering of energy metrics and security-aware communication structure is a dual-fitness routing scheme modulated by both- an element unattainable in previous SMORP-based research and ECC-based aggregation plan.
- Improved security strength based on a hybrid encryption and signature check by elliptic curves and maintains scalability with dense HWSNs.

Full MATLAB simulations indicate that the suggested framework has a substantial impact on network lifetime, distribution of residual-energy, secure aggregation overhead, and delivery reliability over the state-of-the-art routing and security baselines

2. RELATED WORKS

Recent developments in the area of heterogeneous wireless sensor networks (HWSNs) have increased the pressure on the design of routing protocols and security solutions that could meet both energy constraints and data privacy. The current research activities can be approximately divided into two directions that are complementary (i) energy-conscious clustering and routing algorithms aimed at extending network lifetime and (ii) lightweight cryptographic and authentication systems aimed at ensuring in-network data aggregation security. This part presents a selected collection of the recent literature, focusing on their methodology, performance, and limitations when used in scalable and secure HWSN implementation.

2.1. POWER-SAVING CLUSTERING AND ROUTING IN HWSNS.

In the heterogeneous wireless sensor networks (HWSNs), energy-efficient clustering and routing are still fundamental issues due to the underlying heterogeneity of the nodes, that is, they are not equal in terms of their initial energy and differing levels of computational power. Energy-conscious communication The classical clustering algorithms, including LEACH [10], SEP [11], and FSEP [12] achieved the benchmark of energy-optimal algorithms through localized data-aggregating and periodic rotation of CH. LEACH proposed a probabilistic mechanism of CH election that reduces the transmission overhead whereas SEP generalized this designation to unequal deployments by weighting probabilities of CH election by the initial battery level of each node. FSEP

also improved heterogeneity support by adding two sensor classes (L- nodes and H- nodes) which gave a better stability of the networks whose energy distribution was in multilevel. Such classical clustering protocols have been the usual benchmark models on performance comparison in current WSN studies because of its straightforwardness, reproducibility, and behavioral understanding in a heterogeneous environment. In addition, FSEP can also be of relevance in the case of HWSNs since its two level energy model is quite consistent with the heterogeneity assumptions typically utilized in large scale simulation research. Based on these classical models, optimization-based routing schemes have been proposed to overcome the constraints of these classical ones. A typical example is the Spider Monkey Optimization Routing Protocol (SMORP) proposed by Jabbar and Alshawi [16], which provides swarm-intelligence behavior to achieve the stability of CH selections, more evenly distributes the load, and delays the energy depletion SMORP consistently outperforms LEACH, SEP and FSEP on various measures; but is strictly an energy centric approach. It lacks cryptographic protection, in-network aggregation security or authentication, making it susceptible to manipulation in routing and tampering data in hostile conditions. More recently, trust-based routing as well as optimization-assisted routing strategies have been considered in order to increase reliability and resilience Muneeswari *et al*. in [17], introduced a Trust- and Energy-Aware Routing Protocol which compares the credibility of nodes to prevent malicious relays and enhance the reliability of packet delivery. Although these benefits are evident, the computation of trust is associated with much overhead when the network density is large. At the same time, Balan *et al*. in [18], came up with a Taylor-based Gravitational Search Algorithm (TBGSA) of multi-hop routing, which realized better load balancing and network lifetime. Nevertheless, it cannot be used in a hostile environment due to the lack of cryptographic or secure aggregation measures. Similarly, direction-aware multicast routing scheme was suggested by Lekshmi and Suji Pramila [19], to serve a vehicular sensor network with focus on stability in fast mobility. Though this model works in dynamic situations, it is not applicable to static HWSNs as well as confidentiality or authentication are not considered. More developments in optimization of clustering have also been reported based on metaheuristic methods. To get a more homogenous distribution of the residual-energy and minimize irrelevant re-clustering, Reddy *et al*. proposed a better way to get a better Grey Wolf Optimization (IGWO) that results in better distributions [20]. The approach that Jibreel *et al*. came up with is HMGEAR, which is a heterogeneous gateway-assisted routing protocol; it addresses the energy holes surrounding the base station, involving the combination of multi-hop and adaptive head in its selection [21]. Tabatabaei also illustrated the approach whereby optimization of bacterial foraging along with the mobile sink can minimize routing bottlenecks and increase the the network lifetime [22]. These strategies like SMORP did not provide support to security, which they were very effective in maximizing energy consumption. Notably, the new research carried out in [17-19], is a significant step forward regarding the trust-based routing, optimization-based clustering, and reliability-based communication. Nonetheless, all these works do not offer primitives of lightweight cryptography or authenticated aggregation of data, which are crucial in providing a reliable operation in adversarial HWSN setting. The continued divide highlights the necessity of having an integrated energy-security routing architecture, which inspired the proposed framework of integrated SMORP-IECC-ELGDS, reported in this paper.

2.2. LIGHTWEIGHT CRYPTOGRAPHIC AND SECURE AGGREGATION TECHNIQUES IN HWSNS

Heterogeneous wireless sensor networks (HWSNs) are constantly faced with the issue of security because the sensor nodes pose harsh requirements on the system since they have a minimal calculation ability, limited memory storage and lack of a power source that can be recharged. Although widely known to provide high levels of security with the use of less key, elliptic curve cryptography (ECC) based on public-key encryption is relatively costly in terms of its computational attributes, thus rendering its conventional implementations costly in energy-limited systems. In turn, there is a significant amount of literature devoted to the creation of lightweight cryptography, the optimization of ECC implementation, hybrid encryption schemes, authenticated communication schemes specific to WSNs. However, these methods have significant weaknesses that do not allow them to fit in clustering-based routing schemes or privacy ductile data aggregation chains in HWSNs. Among the first models, which have incorporated the use of ECC in terms of secure data forwarding, there is the ECC-Homomorphic Encryption (ECC-HE) model by Elhoseny *et al*. [13]. They can be cryptically aggregated to perform elliptic curve encryption and additive homomorphic operations, and their design supports it. Even though the approach can guarantee high confidentiality and allow the aggregation of results at intermediate nodes, without decryption, the homomorphic component greatly expands the size of ciphertext and the computational burden. Homomorphic addition and multiplicative operations are expensive which results in high processing latency, increases energy consumption, and reduces bandwidth. These inefficiencies make ECC-HE inappropriate in units whose battery is of low power like L-nodes in heterogeneous environments and its implementation is not practical in a network that needs long lifetime stability. Simultaneously, a number of works have tried to trim down the cryptographic weight load by suggesting lightweight or better ECC versions. Ramadevi *et al*. [14] brought the improvements aimed at major management efficiency and arithmetic reduction on a modular basis. Likewise, Hammi *et al*. in [23] and Mahlak *et al*. in [24] suggested the lightweight ECC techniques in which the complexity of scalar multiplication-the most prevalent cost in ECC operations-is minimized. Although these enhancements provide significant improvements in terms of encryption time and energy expenditure, they pay more attention to key exchange or node authentication. Notably, these works consider no authenticated secure aggregation, and they have no provision of checking integrity of aggregated data in the CHs. As a result, such plans do not fit well into hierarchical routing schemes whereby multi-level aggregation and authentication must be performed simultaneously. The literature has also covered hybrid cryptographic architectures. In particular, one should reference the ECDH-RSA model proposed by Abood *et al*. [15], that is, the diffusion of hardware via the Elliptic-Curve Diffie-Hellman of a secure key exchange strategy with the encryption of the payload using RSA. Despite the enhanced confidentiality and immunity to key compromise in hybrid designs, the RSA element

creates excessive modular exponentiation a highly power-intensive function in asymmetric cryptology. That is why ECDH-RSA cannot be used with HWSNs where the CHs have to work with data aggregation of multiple nodes subject to strict energy constraints. Moreover, such hybrid models do not have a lightweight signature mechanism, and therefore, they will not be able to authenticate aggregated data or provide multi-hop integrity. Other methods have sought to increase sensor network authentication. The commonly used digital signature schema has been suggested by Bashirpour *et al*. (2018) in [25], which provided a better authentication scheme on broadcasting using ECC-based signatures. Although the scheme provides good integrity and avoids the broadcast of unauthorised messages, the repetition of generation of signatures as well as their validation has heavy computational requirements. More important, this scheme is not applicable to the clustered routing architectures as well as to secure in-network aggregation. Consequently, the model does not match the operational specifications of heterogeneous and cluster-based WSNs even though it has a robust cryptographic basis. In this literature, some recurrent gaps can be seen to exist with regard to Major Shortcomings in Existing Security Models.

1. **High computational overhead**: Homomorphic ECC and RSA-based hybrids require excessive time and energy for cryptographic operations.
2. **Lack of integrated authentication and aggregation**: Most techniques address either confidentiality or authentication, but not both in one unified architecture.
3. **Incompatibility with clustered HWSNs**: Existing schemes are not designed for hierarchical routing structures where CHs perform multi-level aggregation.
4. **Absence of lightweight digital signatures**: ECC-based signatures remain costly and impractical for repeated verification at CH and BS levels.
5. **No optimization for heterogeneity**: Most models treat nodes as homogeneous, ignoring the energy imbalance inherent in HWSNs.
6. **Scalability concerns**: homomorphic systems do not scale efficiently in dense deployments.

Table 2 provides a comparative analysis of major lightweight cryptographic and secure aggregation schemes relevant to heterogeneous WSNs.

Novelty and Distinct Contribution

The novelty of the proposed SMORP–IECC–ELGDS framework lies in combining optimized energy-efficient routing with lightweight cryptographic protection in a single integrated architecture tailored for heterogeneous WSNs. In contrast to the previous SMORP-based works which solely optimize energy, the suggested design also presents the concept of security-conscious routing, where the selection of the CH factors in the residual energy and cryptography preparedness. The second contribution is the use of a lightweight IECC ciphertext-aggregation procedure to enable CHs without the need to decrypt encrypted input and ciphertext in a two-way communication to multi-hop aggregate ciphertext ECC-HE has been found to be computationally expensive, and plaintext exposure during multi-hop address this issue. In addition, the suggested ELGDS signature mechanism allows aggregation of authenticated results at relatively reduced cost compared to ECC-based signatures like those suggested by Bashirpour *et al*. [25], that is inappropriate in a clustered context as it involves repeated verification at high cost. Combined with the foregoing, these contributions can present the first framework where SMORP energy balancing and lightweight security are mutually influencing alongside one another therefore generating quantifiable advancements in lifetime, secure aggregation cost, and reliability of delivery.

2.3. INTEGRATED ENERGY–SECURITY FRAMEWORKS IN WSNS

Although both energy efficient routing and light-weight cryptographical schemes have been made with huge progress, not much literature has aimed at combining the two dimensions into a single architecture of heterogeneous wireless sensor networks (HWSNs). The current hybrid designs typically seek to integrate the secure communications models with routing protocols, but they are limited in scope, scalability, or application to clustered, multi-hop aggregation spaces. In [26], implemented one of the initial lightweight secure routing protocols in the IoT-oriented WSNs, a protocol combining the crypto-operations with multi-hop routing to reduce the black-hole and sinkhole attacks. The model supports only route's reliability though it fails to support hierarchical clustering or secure in-network aggregation, so it can only be applicable to HWSNs. Equally, [27] introduced an authenticated routing scheme that uses hashing primitives, which are used to maintain the integrity of the message and validation of the route. Although the model has a very high safeguard against packet tampering, repeated hashing and verification bring non-negligible overhead on the CHs and absent confidentiality-preserving aggregation, which makes the scheme inapplicable to hierarchies that are energy-sensitive. Liu [28] tried to make integration more security-conscious by integrating elliptic curve cryptography into a reliable routing protocol, to enhance link-level privacy and authentication. Although this design uses ECC to decrease key size and computational cost, it does not support ciphertext aggregation or lightweight digital signatures, two requirements in supporting multi-hop secure data fusion. As a result, even with security provided by ECC, the absence of the aggregation-aware optimization limits the framework to be used effectively within the densely populated or heterogeneous deployment. The overall result of these hybrid solutions shows increased popularity of using a combination of security and routing but they are not capable of providing a tightly integrated solution that may deliver encrypted aggregation, multi-level authentication, and optimization of energy consumption at the same time. No of the analyzed literature have a combined design of routing choices, cryptographic force, and signature examination in a heterogeneous cluster-based design. This deficiency highlights the necessity of a common architecture like the suggested SMORP-IECC-ELGDS concept-in which energy efficient routing and low weight security work together towards attainment of the rare needs of safe and scalable HWSNs. The table 1 recapsulates the main related works that are discussed in Section 2.1 and 2.2. Energy-efficient routing strategies are represented by rows 1-5, lightweight security and cryptographic mechanisms findings are summarized by rows 6-10, and the suggested SMORP-IECC-ELGDS model is mentioned in row 11.

Table 1. Unified Comparison of Energy-Efficient Routing and Lightweight Security Mechanism

Method	Technique Category	Key Idea	Strength	Limitation	Relevance to Proposed Work
TEARP (Muneeswari *et al*., 2023) [17]	Energy-efficient routing	Trust and energy-aware CH selection	Improves reliability and stability	High overhead in dense networks	Baseline for energy improvements
Taylor-GSA (Balan *et al*., 2023) [18]	Optimization-based routing	Taylor-based GSA for multihop load balancing	Good scalability	Parameter sensitivity	Energy comparison baseline
Direction-aware V2V (Vanitha & Prakash, 2024) [19]	Mobility-aware routing	Directional multicast routing	Robust to topology changes	Not suitable for static HWSNs	Shows limits of mobility-based models
SMORP (Jabbar & Alshawi, 2021) [16]	Metaheuristic clustering	Spider Monkey Optimization for CH rotation	Strong energy balancing	No security integration	Energy base protocol for integration
Reddy *et al*. (IGWO-CH, 2023) [20]	Metaheuristic-based clustering	Applies an improved Grey Wolf Optimization for energy-aware cluster-head selection	Improves energy balance and network lifetime	Does not consider security or secure data aggregation	Serves as an energy-efficient clustering reference motivating secure optimized routing
ECC Digital Signature (Bashirpour *et al*., 2018) [25]	Security authentication	ECC-based broadcast authentication	Strong integrity	High signature overhead	Security baseline for comparison
ECC-HE (Elhoseny *et al*., 2016) [13]	Homomorphic encryption	Encrypted aggregation using ECC-HE	Confidentiality + aggregation	Large ciphertext and high cost	Aggregation security comparison
Ramadevi *et al* ., 2023 IKEC [14]	ECC key exchange, Lightweight cryptography	Improved ECC key management and Reduced-complexity crypto for WSN	Lightweight key handling for Low computational cost	No aggregation support for Limited authentication features	Cryptographic complement baseline by Supports lightweight design rationale
ECDH–RSA (Abood *et al*., 2022) [15]	Hybrid cryptography	ECDH + RSA for secure transmission	Strong confidentiality	RSA overhead heavy for CHs	Motivation for lightweight hybrid
Proposed SMORP–IECC–ELGDS	Integrated routing + security	Energy-aware routing + hybrid ECC security	Unified secure aggregation + efficiency	—	Main contribution

3. PROPOSED METHODOLOGY

The research design adopted in this study is structured around an integrated workflow that links routing optimization with secure data aggregation. The routing layer is first responsible for cluster formation and multi-hop data forwarding, while the security layer operates concurrently to protect the transmitted data without interfering with routing decisions. This design ensures that energy efficiency and data security are addressed within a single operational process rather than as separate or sequential stages.

3.1. INTEGRATED ENERGY-EFFICIENT ROUTING AND SECURE DATA AGGREGATION METHODOLOGY

This part provides a holistic approach that combines an optimization-based clustering and routing framework with a lightweight cryptography framework in order to provide secure and energy-efficient data aggregation in heterogeneous wireless sensor networks (HWSNs). The new framework will utilize the (SMORP) to build dynamic cluster topology and balanced multi-hop routing paths, and a new hybrid security model, which consists of (IECC) and a hybrid security model (ELGDS) will be used to provide end-to-end confidentiality, integrity, and authentication. The proposed model integrates cluster formation, route stabilization, ciphertext aggregation and signature verification in a single operational pipeline, in contrast with traditional methods where routing and security processes have been engineered like applications without connection to each other. We have summarized the interactions between these components and the sequential execution of them conceptually in Fig. 1 and elaborated on each in the following section. Fig. 1 illustrates the interaction between SMORP clustering, optimized routing, IECC encryption, ciphertext aggregation, and ELGDS authentication within the integrated framework.

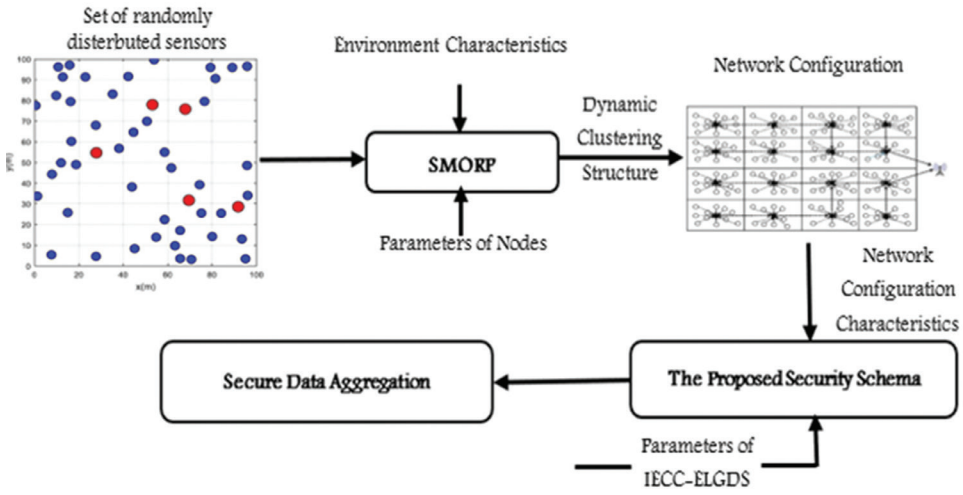


Fig. 1. Proposed Secure Data Aggregation Workflow Schema

3.2. SMORP-BASED CLUSTERING AND ROUTING PROTOCOL

The core procedure that is utilized in the development of the energy-balanced clusters and calculating the optimal multi-hop paths through the heterogeneous wireless sensor network is the Spider Monkey Optimization Routing Protocol (SMORP). The protocol is inspired by these social behaviors of spider monkeys, namely the fission-fusion foraging, subgroup form and rotation of leaders, are the elements that help maximize the energy efficiency. Under the proposed framework, SMORP has the responsibility of CH selection, election of a leader, formation of subgroups and refurbishment of routes as the residual energy goes down, and/or intra/inter-cluster distance. SMORP works in a series of iterative phases which entail network start, Local Leader Phase (LLP), Global Leader Phase (GLP), Local Leader Updating, Global Leader Updating and termination. All the stages help in the selection of balance CHs and construction of strong routing paths towards the sink.

3.2.1. Network Initialization and Node Evaluation

In the starting stage, the positional coordinates, residual energy and neighbor-list information of each sensor node are broadcast to give the initial network state involved in SMORP activities. It is on the basis of this information that candidate forwarding nodes are obtained and their suitability evaluated to proceed with being part of the routing structure. Evaluation is then done spatially to calculate the closeness of each node to the sink, as a node that is closer to sink usually takes lesser cost of transmission. Given the coordinates (x_s, y_s) of the sink and (x_l, y_l) of the candidate node l, the Euclidean distance is computed as:

$$\boldsymbol{d(l)} = \sqrt{(\boldsymbol{x_s} - \boldsymbol{x_l})^2 + (\boldsymbol{y_s} - \boldsymbol{y_l})^2} \quad (1)$$

This distance measure along with the nodes residual energy along with intra/inter cluster distance forms directly part of the calculate of the fitness value which rules routing potential of every node. The fitness function has the definition of:

$$\boldsymbol{fitness}\ (\boldsymbol{l}) = \boldsymbol{\alpha} \times \boldsymbol{RE(l)} + (\boldsymbol{\beta} \times \frac{\mathbf{1}}{\boldsymbol{D_1}} + \boldsymbol{\gamma} \times \frac{\mathbf{1}}{\boldsymbol{D_2}}) \quad (2)$$

- $RE(l)$ is the residual energy of node l.
- $D_1$is the distance between an L-sensor node and its associated CH.
- D_2 is the distance between the CH and the sink,
- (α, β, γ) are weighting coefficients regulating the impact of each parameter.

The fittest nodes are said to be the most suitable in terms of serving as nodes of CHs or forwarding nodes. This evaluation step gives SMORP an energy aware, spatially efficient, analysis of the network structure that can be utilized in effective decision-making during later local and global decision-making steps in choosing local and global leaders and routing states.

3.2.2. Fitness Evaluation and Forwarding Candidate Assessment

SMORP routing is based on a systematic analysis of forwarding candidate evaluation criteria depending on the availability of energy, spatial proximity, and cluster-specific metrics. Once initialized each node keeps current data on its remaining energy, its distance to the CH to which it belongs, and the distance between the CH and the sink. The metrics allow the protocol to build a spatially efficient and an energy-balanced forwarding infrastructure. At every expansion phase, candidate nodes are analyzed in order to be considered suitable to add to the routing path. A Euclidean distance d (l) of a candidate node l and sink calculated above in Eq. (1) is one of the basic spatial descriptors. The fitness in Eq. 2 is a combination of this distance and the nodes energy and distances associated with clusters produced and a total routing utility score. After the computing of the values of fitness of all the nearby candidates, the Global Leader Spider Monkey (GLSM) will examine them during which the forwarder with the most promise is identified. The forwarding possibility of a candidate node l_i is determined as

$$P(l_i) = \frac{fitness(l_i)}{\sum_{j=1}^{N} fitness(l_j)} \quad (3)$$

Where:

- $P(l_i)$ is the forwarding probability of node l_i.
- $fitness(l_i)$ is the fitness value of node l_i.

- N is the number of neighboring nodes considered in the expansion stage.

The candidate nodes that are found in the same iteration are successors to the expanded node and custodians of it by way of pack-pointers. This design allows SMORP to build a hierarchical expansion tree effectively searching the possibilities of routing. The growth will be repeated until the sink is reached and all data sensed will be sent via the optimal path. These forwarding measures are the basis of the process of leader coordination where multi-level leaders optimize the routing search and direct the expansion in the direction of the sink. The sequential interactions between the Local Leaders (LLs), their subgroup members (LLSMs) and the Global Leader (GLSM) are expounded in the subsequent section.

3.2.3. Leader Hierarchy and Sub-Group Formation in SMORP

SMORP arranges sensor nodes in a hierarchical leader-member framework which creates the opportunity to explore forwarding paths in a coordinated manner and equally balanced energy use. This is built by repeated estimation of the fitness of nodes when by the nodes with high fitness level become leaders of their local neighborhoods. Every neighborhood of nodes comprises a Local Leader Sub-Group (LLSG). In every LLSG, the node that has the largest fitness score is made the Local Leader (LL), and the rest of the nodes the Local Leader Sub-Group Members (LLSMs). The LL is able to examine several forwarding opportunities in its immediate environment. This design can guarantee that routing choices is not constrained on a particular node and is robust to local failures or fast failure of energy sources. At an international level, the node with the highest global fitness in the network is made the GLSM. The GLSM manages the further upper hierarchical advancement of the routing search and directs the choice of the most promising next level of expansion towards the sink. This strictly hierarchical team structure, where LLSGs develop into LLs and then into LLSMs overseen by the GLSM, lets SMORP build up a multi-level strategy of exploration. The LLSM oversees global refinement, the LLs control interaction between subgroups, and LLSMs are involved in the assessment of candidate successors. This multi-level coordination is the structural basis to the determination of the most suitable forwarding path. The complete operation of this mechanism is summarized in Algorithm 1 below.

Algorithm 1. SMORP-Based Packet Forwarding Procedures in HWSNs

Input:

- Set of sensor nodes N with positions and residual energy $RE(l)$
- Distances D_1 (L-sensor → CH) and D_2 (CH → $Sink$)
- Fitness parameters α, β, γ
- Sink node S

Output:

- Optimal forwarding path from source node to sink

1. Initialize network state and compute distances $d(l)$ to the sink for all nodes using Eq. (1).
2. Compute $fitness(l)$ for each node using Eq. (2).
3. Form Local Leader Sub-Groups (LLSGs) based on neighbourhood proximity.
4. For each LLSG do
5. Identify Local Leader (LL) as the node with maximum fitness.
6. Assign remaining nodes in the sub-group as LLSMs.
7. End for
8. Determine the Global Leader Spider Monkey (GLSM) as the node with highest global fitness.
9. Set current node ← source node.
10. Initialize Forwarding Path.
11. While current node ≠ Sink do
12. Extract neighbour set L of current node.
13. For each node $l_i \in L$ do
14. Compute forwarding probability $P(l_i)$ using Eq. (3).
15. End for
16. Select next node ← argmax $P(l_i)$, $l_i \in L$.
17. Set pack-pointer (next node) ← current node.
18. Append next node to Forwarding Path.
19. Update current node ← next node.
20. End while
21. Return Forwarding Path.

The steps outlined in Algorithm 1 describe how routing decisions are progressively refined based on node fitness and forwarding probability. By prioritizing nodes with higher residual energy and favorable spatial positions, the routing process avoids overloading specific nodes and maintains balanced energy consumption across the network. This procedural design supports stable multi-hop communication while preserving the energy efficiency required for long-term HWSN operation. In the enhanced formulation of SMORP adopted in this work, several structural and operational refinements are incorporated to overcome the limitations of the classical SMORP routing mechanism and to better accommodate the requirements of heterogeneous wireless sensor networks. Among the significant enhancements, it is possible to list the following:

- **Multi-Metric Fitness Evaluation**: The distance-centric SMORP model is expanded by including a composite fitness functional which takes into account jointly the residual energy and the L-sensor-to-CH distance as well as the CH-to-Sink distance. This multi-parameter assessment makes the forwarding decisions more balanced and avoids the early exhaustion of the critical nodes.
- **Probability-Driven Forwarder Selection**: To better improve on heuristic exploration, instead of using a simple heuristic exploration, candidate forwarding nodes are being selected based on a normalized probability that is based on the fitness values of the candidate forwarding node. This deterministic choice allows contributing to the ability of routing stability and the reduction of the risk of repetitive selection of the same nodes in the subsequent round.
- **Refined Multi-Level Leadership Hierarchy**: It is an extension of organizational hierarchy explicitly expanding it to encompass (LLSMs), (LLs), (LLSGs) and (GLSM). Such high-level refinement enhances coordination and decentralized decision-making in heterogeneous nodes in subgroups.
- **Heterogeneity-Aware Role Assignment**: The improved SMORP is able to incorporate the distinct roles

of L-sensor nodes and CHs as part of the optimization cycle. This is to make sure that nodes that have less energy or ones with lesser communication ability are not overwhelmed, which means that routing performance of HWSNs can be made more sustainable.

Security-Compatible Routing Design: The routing paths resulting under the enhanced SMORP are built in a way to allow ciphertext forwarding and authenticated aggregation, which make them easy to integrate with the IECC-ELGDS security architecture presented in the following section. The traditional SMORP formulation lacks such a compatibility. All of these improvements make SMORP an energy-conscious, heterogeneity-aware, and security-enabled routing mechanism than the conventional SMORP model that was applied in previous works. As an explanation to an outline of the hierarchical coordination process applied in the enhanced SMORP formulation, Fig. 2 represents the multi-level leadership structure that is adopted in subgroup construction and route construction.

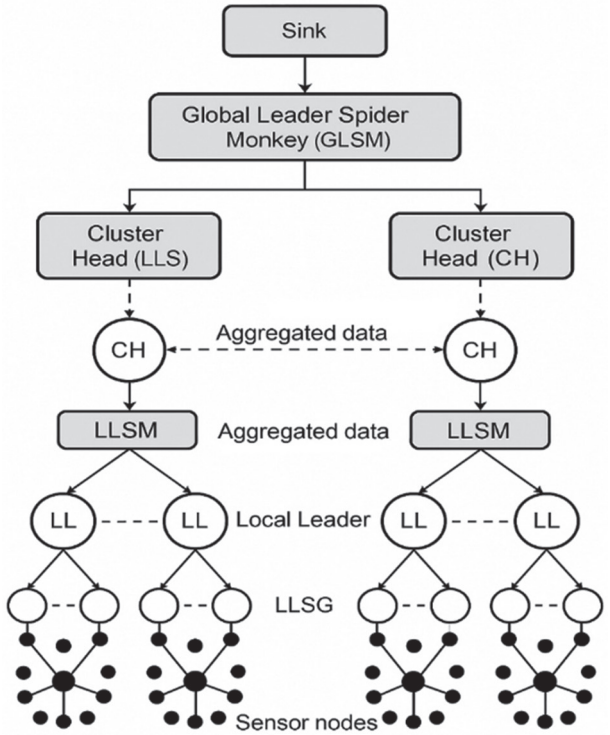


Fig. 2. Multi-Level SMORP Hierarchy

In this architecture the sensor nodes will first be clustered into (LLSGs), which are controlled by (LL) that would facilitate localized decision making. Higher on, several LLs will be assigned to (LLSM), whereby the consolidation of reports on subgroups is made, and the routing activities are coherent among distributed routing activities. The topmost position of decision-making is controlled by (GLSM), and it is the one that coordinates inter-cluster communication, and guides the construction of the ultimate multi-hop routing path to the sink. This hierarchy allows forwarding candidate evaluation in distributed fashion that is scalable, routing overhead reduction and also improves stability of constructed paths. Also, the illustration points out the smoothness of the interaction between these layers of leadership and the underlying cluster-based architecture of the heterogeneous network and which basis the structural foundation of the optimized routing process illustrated in the previous subsections.

3.1. INTEGRATED SECURITY ARCHITECTURE USING IECC AND ELGDS

To ensure confidentiality, integrity and authenticity at proposed routing framework, the proposed work uses dual layer light weight security architecture through (IECC) scheme of data encryption and (ELGDS) scheme of authentication. These two should be used in conjunction to make the routing energy efficient as provided by SMORP and computationally manageable when facing a mixed population of the wireless sensor nodes with limited processing and energy capabilities.

The routing mechanism will operate in parallel to the security architecture where L-sensor nodes will encrypt the sensed data with IECC and the CHs will have access to the encrypted data but not to the decrypted one. Such a design makes the intermediary nodes unable to access plaintext values, and decreases eavesdropping or tampering. Besides, a digital signature is generated to every encrypted packet by ELGDS to ensure that the sink can perform end-to-end authentication of the data authenticity and prevent any form of data manipulation in the event that the packet is forwarded through a multihop. This is ensured by the combination of IECC and ELGDS to ensure that safe data aggregation is carried out effectively and efficiently without imposing excessive computing load on low-power nodes. The elements of the proposed security architecture are discussed in the subsections below starting with description of encrypted communications to be used in the system i.e. the IECC encryption model, and then the description of the ELGDS signature mechanism and finally the inbuilt workflow of secure aggregation.

3.3.1. IECC-Based Lightweight Encryption Model

The Improved Elliptic Curve Cryptography (IECC) model suggested to be used as the first element of the proposed security architecture is used to deliver lightweight and energy-efficient data confidentiality to heterogeneous wireless sensor networks. IECC has been chosen because it can provide high cryptography security certain key sizes, which are small enough to be applicable to sensor nodes with small computational capacity and restricted battery power. Under the proposed framework, an elliptic-curve public-private key pair is produced by every L-sensor node and the sinks public key is used to encrypt transmission sensory data by the sending node before transmission. This guarantees that the original plaintext can only be garnered by the sink which holds the corresponding private key. The encryption process of the IECC is as follows: the sensed data of the sensor nodes is first mapped to a point in the elliptic curve and a scalar multiplication with the sinks public key is performed to obtain a pair of ciphertext elements. Such ciphertext values are then sent across the intermediate nodes and CHs without being decrypted and therefore they cannot be accessed by unauthorized users in the multi-hop routing. The energy overhead of encryption is further diminished because the head of the cluster can provide aggregation of ciphertext directly; this means that the energy cost of encryption is only realized once at the sensing node thereby minimizing the total level of computational overhead experienced by secure data aggregation. The IECC model can provide confidentiality with the, in comparison, very small key sizes of elliptic-curve operations, which does not compromise the long-term viability of the heterogeneous sensor nodes.

This lightweight encryption mechanism forms the foundation for the authenticated secure aggregation workflow described in the subsequent subsections. "The operational steps of the IECC key generation and encryption process at each L-sensor node are summarized in Algorithm 2.

Algorithm 2. IECC Key Generation and Encryption at L-Sensor Node

Input:

- Elliptic curve parameters (p, a, b), base point G of order n,
- sink public key Q_{sink}, plaintext message M.

Output:

- Ciphertext pair (C_1, C_2).

% Offline key generation phase (performed once per L-sensor node)

1. Select an elliptic curve E over a finite field F_p defined by
- $E: y^2 = x^3 + ax + b \pmod p$, where a and b are integers such that E is non-singular.
2. Choose a base point $G \in E(F_p)$ with large prime order n.
3. Select a private key d_{node} randomly such that $1 \le d_{node} \le n - 1$.
4. Compute the corresponding public key of the node as $Q_{node} = d_{node} \cdot G$.

% Online encryption phase at the L-sensor node

5. Represent the sensed data as a point M on the elliptic curve E.
6. Select a fresh random integer k such that $1 \le k \le n - 1$.
7. Compute the first ciphertext component as $C_1 = k \cdot G$.
8. Compute the second ciphertext component as $C_2 = M + k \cdot Q_{sink}$.
9. Form the IECC ciphertext as the pair $C = (C_1, C_2)$.
10. Transmit the ciphertext C to the cluster head or next-hop node.
11. Return (C_1, C_2).

3.3.2. ELGDS Digital Signature and Authentication

The second component of the proposed security architecture is (ELGDS) scheme, which is employed to ensure end-to-end data authenticity and integrity throughout the multi-hop transmission process. CO The flow of control in ELGDS is similar to that of IECC in that the sink can confirm that every received ciphertext was produced by a trustful source and no alterations were made to the message when it was forwarded. Such a two-layer design will keep the enemies off-balance-sheet as they cannot send spoofed packets, modify encrypted values, or repeat already transmitted messages in the network. Since the digital signature is generated with the help of the private signing key of each L-sensor node in the suggested framework, each node uses its own key to create a signature to every packet encrypted with the help of IECC. The signature is calculated against a hashed version of the ciphertext so that subtle changes in a cipher-text payload will spoil the signature. The signature pair that is obtained is added to the ciphertext prior to sending, allowing the intermediate nodes to transmit the information without doing any authentication. The computational load of signature generation is therefore restricted to the source L-sensor nodes since CHs are only used as a point to aggregate, and are never involved in the authentication. When the sink receives a ciphertext packet made up of aggregates, it decrypts the packet with the public verification keys supplied to the sink to confirm the ciphertext signatures attached to the packet. Effective verification guarantees that the ciphertext elements were created by honest nodes and in addition to that, they were not distorted during the routing. This end-to-end authentication mechanism eliminates impersonation, tampering and replay attack, thereby strengthening the security guarantees of the proposed secure data aggregation model without imposing excessive computational overhead on intermediate nodes.

Algorithm 3. ELGDS Key Generation and Signature Generation at L-Sensor Node

Input:

- Large prime modulus p, generator g of Zp^*,
- private signing key x ($1 < x < p - 1$),
- hash function $H(\cdot)$, message m.

Output:

- Public verification key y, digital signature (r, s) for m.

% Offline key generation phase (executed once per L-sensor node)

1. Select a large prime number p and a generator g of the multiplicative group Zp^*.
2. Choose a private signing key x such that $1 < x < p - 1$.
3. Compute the corresponding public verification key as $y = g^x \bmod p$.

% Online signature generation phase (executed whenever a message m is sent)

4. Compute the message hash $h = H(m)$, where h is mapped into $Zp-1$.
5. Repeat
6. Select a random ephemeral key k such that $1 < k < p - 1$.
7. Until $gcd(k, p - 1) = 1$.
8. Compute $r = g^k \bmod p$.
9. Compute the modular inverse k^{-1} of k *modul o* $(p - 1)$.
10. Compute the second signature component as $s = k^{-1} \times (h - x \cdot r) \bmod (p - 1)$.
11. Output the public verification key y and the digital signature pair (r, s).

It is worth noting that the key generation phase in Algorithm 3 is executed infrequently and can be performed offline, ensuring that only lightweight signing operations are carried out during regular sensing rounds.

3.3.3. Integration of IECC and ELGDS for Secure Data Aggregation

The concluding phase of the presented security architecture offers the assurance of confidentiality (IECC) with the assistance of the authentication and integrity services given by the (ELGDS) scheme to create a single effective wise data aggregation pipeline. Each L- sensor node in this model optimizes the sinks IECC public key with its encrypted data and

then entities a digital signature over the encrypted data. This joint encryption-signing process also provides confidentiality, authenticity and integrity of data are applied before a packet is sent by the sending node. In the routing step, packet ciphertexts and their signatures are sent out by intermediate nodes such as CHs and are not decrypted or validated. This design inhibits plaintext exposure and does not distribute computationally expensive cryptographic operations across resource restricted forwarding nodes. CHs perform ciphertext-preserving aggregation, which is a process that allows data forwarding over multi-hops and keeps the encrypted version of the information all the way across the routing path. Since aggregation is performed directly on ciphertext, no intermediate node would have access to the underlying sensing data, which practically performs leakage elimination even in case compromised forwarding nodes. When the aggregated ciphertext and the corresponding set of signatures are received the sink starts to run a two-stage recovery process. To ascertain an authenticity and integrity of every encrypted contribution, first, ELGDS public verification keys are applied. Block ciphertexts that do a pass are only stored to undergo further processing after signature check passes. Second, timely verification is performed, and secondary to it is IECC decryption, as a result of which the sink constructs the organized plaintext.

A verification-first architecture ensures that manipulated or replayed ciphertext is completely dropped before decryption and thus prevents impersonation attacks, tampering and fake-contributions to the network. The integrated IECC-ELGDS model achieves a strong end-to-end end-authentication and confidentiality with the use of the L-sensor nodes and the sink alone and conserves the energy resources of the intermediate nodes. Algorithms 4, essentially express the whole workflow of the new scheme, including generation of ciphertext, building signature, middle-level forwarding, signature verification and recovery of plaintext.

Algorithm 4. Integrated IECC–ELGDS Secure Aggregation

Input:

- For each L-sensor node i: plain text message Mi, IECC parameters (p, a, b, G, n),
- sink public key Q_{sink}, ELGDS parameters (p, g, x_i, y_i), hash function $H(\cdot)$.

Output:

- At the sink: verified aggregated plaintext M_{agg}.

% Phase 1: IECC encryption and ELGDS signing at each L-sensor node

1. For each L-sensor node i do
2. Map the sensed data to a point M_{ion} the elliptic curve E over F_p.
3. Select a random k_i such that $1 \le k_i \le n - 1$.
4. Compute $(C_1)_i = k_i \times G$.
5. Compute $(C_2)_i = M_i + k_i \times Q_{sink}$.
6. Form the IECC ciphertext $C_i = ((C_1)_i, (C_2)_i)$.
7. Compute the message hash $h_i = H((C_1)_i, (C_2)_i)$.
8. Select a random ephemeral key k_s such that $1 < k_s < p - 1$ and $gcd(k_s, p - 1) = 1$.
9. Compute $r_i = g^{(ks)} \bmod p$.
10. Compute $k_{s(-1)}$ as the modular inverse of k_s modulo $(p - 1)$.
11. Compute $s_i = k_{s(-1)} \times (h_i - x_i \times r_i) \bmod (p-1)$.
12. Attach the signature $\sigma_i = (r_i, s_i)$ to the ciphertext Ci.
13. Transmit the packet $P_i = ((C_1)_i, (C_2)_i, r_i, s_i)$ to the corresponding cluster head.
14. End for

% Phase 2: *Ciphertext* forwarding and aggregation at intermediate nodes / *CHs*

15. For each cluster head CH do
16. Collect incoming packets P_i from associated L-sensor nodes.
17. Perform ciphertext aggregation:
    $(C_1)_{agg} = f_1\{(C_1)_i\}$, $(C_2)_{agg} = f_2\{(C_2)_i\}$,
    where f_1 and f_2 preserve *ciphertext* structure.
18. Forward aggregated *ciphertext* $C_{agg} = ((C_1)_{agg}, (C_2)_{agg})$
    Along with signatures(σ_i) toward the sink.
19. End for

% Phase 3: Signature verification and IECC decryption at the sink

20. Upon receiving C_{agg} and the set (σ_i), the sink performs:
21. For each node i do
22. Recompute $h_i = H((C_1)_i \,||\, ((C_2)_i)$.
23. Compute $(v_1)_i = (yi^{(ri)} \times r_i^{(si)}) mod\ p$.
24. Compute $(v_2)_i = g^{(hi)} \bmod p$.
25. If $(v_1)_i \neq (v_2)_i$ then
26. Discard the corresponding ciphertext contribution.
27. End if
28. End for
29. Apply IECC decryption to recover aggregated plaintext (M_{agg})
    $M_{agg} = (C_2)_{agg} - d_{sink} \times (C_1)_{agg}$.
30. Output the verified aggregated plaintext M_{agg}.

3.4. NOVELTY AND DISTINCT DESIGN CONTRIBUTIONS

The suggested framework presents a collection of unique design provisions that will separate it with the current routing and security plans in heterogeneous wireless sensor networks (HWSNs). In contrast to the original SMORP formulation in Jabbar and Alshawi (2021) [16], where the protocol is concerned only with the formation of clusters through energy-efficient routing and multi-hop routing, but no cybersecurity integrity version is introduced-the methodology established in this paper inserts a full cryptographic pipeline directly into the SMORP framework of operation. The improved model adds confidentiality-saving IECC encryption to SMORP, end-to-end signature enforcement by ELGDS, ciphertext-presaving CH aggregation and secure signatures propagated on-top of the hierarchical LLSM-GLSM routing process. This forms a hybrid between SMORP as a simple optimization-driven routing protocol, and as a resilient, secure-by-design communication architecture, which can support resilient multi-hop data forwarding in adversarial environments. Compared to the single ELGDS digital signature protocol modeled by Bashirpour *et al*. in [25], where user authentication is enabled by the protocol, but routing is not, multi-hop data aggregation, and the ability to adapt to the resource constraints inherent to HWSNs, the proposed framework integrates lightweight encryption and authentication into an energy-aware communication substrate. IECC-ELGDS hybrid mechanism is specially designed to be implemented in heterogeneous environment of sensors so that all the cryptographic actions are performed at L- sensor

nodes and at the sink. This design reduces the calculational burden of computing nodes and allows relaying encrypted and signed packets without the decryption or verification steps in the middle of the path. The innovative character of the offered approach is thus in three aspects:

- Co-design in routing-security networks, SMORPs leader-based optimization structure has been generalized to support ciphertext routing, signature propagation, and secure aggregation with no modification of protocols energy-efficiency goals.
- Multi-hop aggregation uses ciphers, such that the CHs are allowed to receive the aggregation of encrypted numbers and ensure the utmost confidentiality of sensor data.
- A verification-first decryption model, whereby the sink validates all received ciphertext elements with ELGDS prior to the IECC decryption, which gives a high level of resistance to tampering, replay and impersonation.

These collectively enhanced advances make the proposed system have a single secure routing and aggregation pipeline that has never existed in any other SMORP-based research, or ECC-based authentication system. This combined design is the basic contribution of the work and it is used in the development of the better performance and security properties in the further parts.

3.5. INTEGRATION OF SMORP WITH THE IECC–ELGDS SECURITY MECHANISM

The presented framework ensures the integration of SMORPs optimization-based routing framework and lightweight IECC-ELGDS security framework to offer an integrated and thorough approach to data aggregation of heterogeneous wireless sensor networks both in energy efficiency and full security. In contrast to a more traditional design where routing and security are separate, integration here means the implementation of confidentiality, authentication and ciphertext aggregation directly into the SMORPs multi-level Leadership and forwarding work. This allows the routing operation to be energy sensitive and at the same time minimize delays on how data is sent without the wrongdoer violating the data integrity against eavesdropping, manipulation, and impersonation. Each L-sensor at the sensing layer ciphers its result with the sinks IECC public key and creates an ELGDS signature over the resultant ciphertext before participating in the SMORP routing workflow. This will make sure that the packets played in the forwarding procedure are already encrypted. Just like in the original SMORP, the encrypted packets and the respective SMORP signature are propagated in the same route as the optimization-built routes in the hierarchy of (LLs), (LLSM), and (GLSM). Importantly, cutting points such as head of a cluster only do forwarding actions without decryption or validation of signature. This architecture avoids exposing plaintext at energy-constrained nodes, as well as maintains the lightness of the routing substrate. The forwarding mechanism of every SMORP level is unchanged in that the suitability of forwarding candidates is still using the formulation of energy-aware fitness presented earlier in the form of Eq. (2), except that the fitness of forwarding is now determined using residual energy and cluster-specific distance measures(D_1, D_2). By keeping the original routing utility measure introduced in Section 3.2.1, the integration will preserve the efficiency of SMORPs without compromising the efficiency of the security layer in any way. In multi-hop propagation, elements of the ciphertext, namely (C_1, C_2) in the structure of ciphertext in Section 3.3.1, are propagated, and CHs do ciphertext-preserving aggregation using the homomorphic addition property of the IECC construction. This enables a direct aggregation of encrypted values to be done without any loss of its complete confidentiality.

The sink starts a two-step recovery process after aggregated ciphertext contributions have been received. Signatures authentication is initially carried out with the ELGDS verification condition in algorithm 4 of Section 3.3.3. Only ciphertext blocks whose signatures satisfy the relation (v_1, v_2)are accepted for further processing. Second, the validated ciphertext is decrypted using the IECC private key to reconstruct the aggregated plaintext. This verification-first model prevents forged or manipulated ciphertext from entering the decryption pipeline and enhances the system's resilience against replay, impersonation, and tampering attacks. By integrating the routing utility of Eq. (2), the IECC ciphertext formulation of Section 3.3.1., and the ELGDS signature verification rule in Algoritm 4, the proposed system produces a cohesive secure-SMORP framework capable of delivering energy-efficient, confidential, and authenticated multi-hop communication. The complete operational flowchart of this integrated model is illustrated in Fig. 3, which summarizes the interaction between SMORP routing stages and the IECC–ELGDS cryptographic operations.

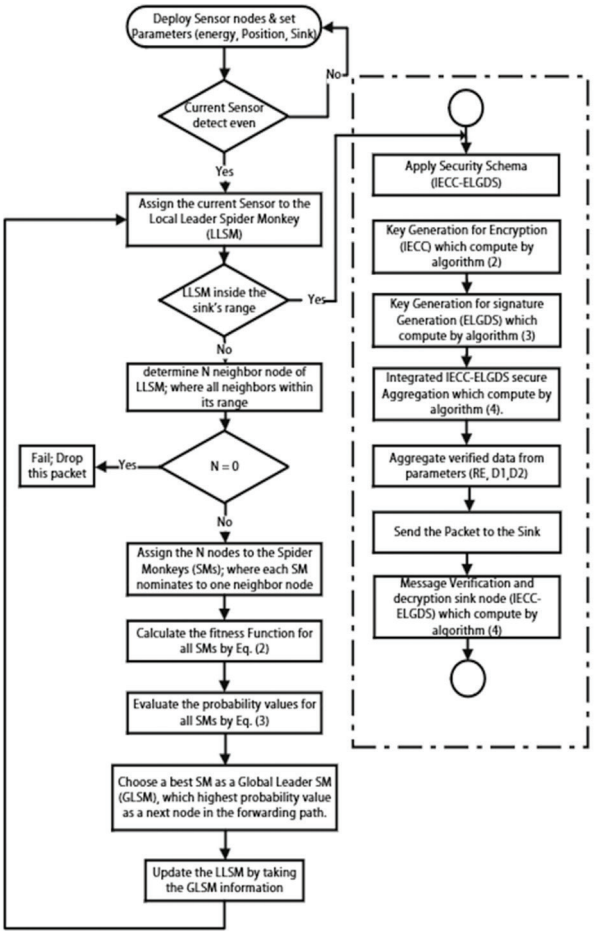


Fig. 3. Flowchart of Proposed method (Security schema IECC-ELGDS) in SMORP for HWSNs

4. SIMULATION ENVIRONMENT AND PERFORMANCE EVALUATION

A structured simulation environment was established to rigorously examine the performance of the proposed SMORP routing and IECC–ELGDS security mechanisms. This section outlines the evaluation framework, including deployment assumptions, communication model, parameter settings, and metrics employed to assess efficiency and robustness. The simulation assumes static sensor nodes and ideal channel conditions; therefore, the obtained results reflect performance under controlled network scenarios.

4.1. NETWORK DEPLOYMENT

The deployment of heterogeneous wireless sensor network is under a square sensing area of (100 m by 100 m). One hundred L-sensor nodes and five CHs are randomly distributed throughout the field to have a realistic and non-uniform spatial distributions. Fig. 4 gives a structure, in a schematic way, of the heterogeneous network layout that represents the spatial distribution of L- sensor nodes, hierarchical arrangement of CHs, and the location of sink such that it gives a vivid visualization of the deployment structure assumed in the work.

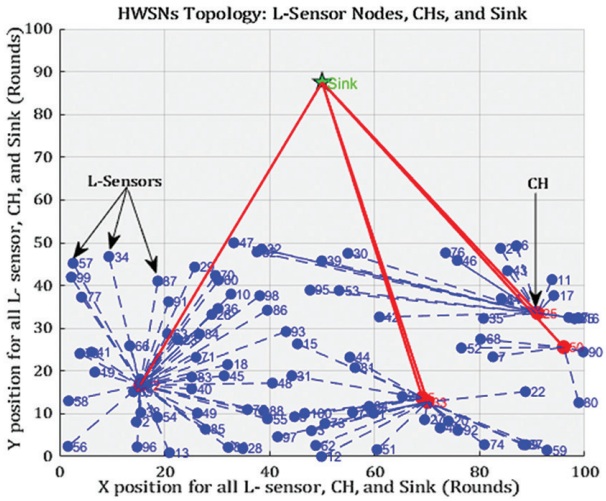


Fig. 4. HWSN topology showing L-sensor nodes, cluster heads (CHs), and sink placement

The nodes are stationary during the simulation and the geography is assumed to be known. One sink node is deployed at (50 m, 85 m), which is at the boundary that is close to the upper limit of the field to create routing asymmetry and resembles real-life multi-hop communication pattern. The L-sensors carry sensed information into their corresponding CH, where ciphertext aggregation is carried out into the sink. The transmission distances of L-sensors and CHs are set to 20 m and 80 m respectively, and this allows the creation of a three-level routing topology, which is based on the dispersity of the energies of nodes. Energy levels will be configured initially to 0.5 J/L-sensors and 2.5 J/CHs, which are consistent with the standard specifications of the heterogeneous WSN hardware platforms. The game continues up to 2000 rounds, with each round consisting of one full sensing-aggregation-transmission cycle on the network. This implementation scheme aligns with real-world use of HWSN deployments in environmental surveillance, and smart-city systems, where nodes will be heteronymously deployed, and will not be moved once in place. With such a set-up, realistic energy-depleting behavior, variability of routes, and the joint effect of SMORP routing and IECCELGDS secure data aggregation on that of the entire network is measured.

4.2. RADIO ENERGY MODEL

The energy consumption of wireless communication in the heterogeneous sensor network is modeled using the first-order radio model, which is widely employed in WSN performance evaluation and remains consistent with foundational studies such as LEACH [10]. This model provides an analytically tractable and experimentally validated representation of radio dissipation, making it suitable for both short-range L-sensor transmissions and *long-range CH*-to-sink links within heterogeneous architectures. In this model, the energy required to transmit a k-bit packet over a distance d depends on whether the communication operates in the free-space regime or the multipath-fading regime.

$$E_n T(k) = \begin{cases} k \times \left(E_{ele} + E_{fs} \times d^2\right) & if\ d < d_0 \\ k \times \left(E_{ele} + E_{fs} \times d^4\right) & if\ d \geq d_0 \end{cases} \quad (4)$$

Where:

- E_{elec} is the per-bit electronic circuitry cost.
- E_{fs} and E_{mp} represent the free-space and multipath amplifier coefficients, respectively.

The threshold distance that separates the free-space and the multi-path fading channel models is:

$$d_0 = \sqrt{\frac{E_{fs}}{E_{mp}}} \quad (5)$$

The energy consumed to receive a k-bit packet is defined by:

$$E_n R(k) = k \times E_{elec} \quad (6)$$

This model follows the formulation introduced by Heinzelman *et al.* [10], and it provides a widely accepted abstraction for radio communication energy in wireless sensor networks. Its linear-plus-distance-dependent structure accurately reflects the physical behavior of low-power transceivers and ensures fair comparison with prior routing- and clustering-based WSN protocols. In heterogeneous sensing environments, L-sensor nodes perform primarily short-range transmissions to their nearest CHs, while CHs conduct longer-range forwarding toward the sink. This asymmetry is best represented by the adopted dual-regime model in which short range transmissions would be within the free-space region whereas CH-to-sink links would often induce the multipath model because of larger transmission distances. This difference can accurately estimate node level energy consumption, visible energy dynamics and network lifetime characteristics using SMORP routing with the built in IECC-ELGDS secure aggregation. Fig 5. and Table 1 summarizes the radio-model parameters used in the simulations, including (E_{elec}, E_{fs}, E_{mp}) and E_{DA}. These parameters are identical for both L-sensor nodes and CHs, since they represent hardware-level characteristics of the transceiver module used across all nodes. The sole differences between L-sensors and CHs are their initial energy capabilities and range of transmission which are indicated separately in Table 2.

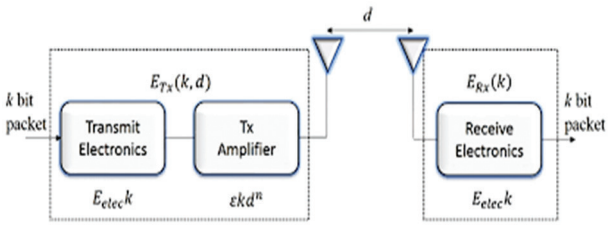


Fig. 5. first-order radio model

Table 2. Specifications of the Initial Radio Model for Both L-Sensors and CH

Parameter	Description	Value
E_{elec}	Energy for T_x/R_x electronics	50 nJ/bit
E_{fs}	Free-space amplifier coefficient	10 $pJ/bit/m^2$
E_{mp}	Multipath amplifier coefficient	0.0013 $pJ/bit/m^4$

4.3. SIMULATION PARAMETERS

Every simulation test was performed in the MATLAB R2023a under the singular assessment atmosphere so that each and every scheme was evaluated equitably and reproducibly. The complete operational cycle; the sensing stage, the aggregating stage, the process of secure processing, the routing stage, the radio-energy updating stage occur in each simulation round, and the overall assessment is 2000 rounds. Every routing and security protocol functions within the same communication constraints of the first-order radio energy model of Section 4.2. All protocols use a fixed value of 2 kB packet-size to ensure uniformity when being evaluated in terms of transmission-cost. All the comparative benchmark protocols were implemented with identical node deployment, sink position, radio parameters and initial energy settings. The most widely used baseline routing algorithms are LEACH [10], SEP [11], and FSEP [12], whereas the security-oriented schemes are ECC-HE [13], IEKC [14], and ECDH-RSA [15]. These protocols are also popular reference models in the optimization of WSNs and secure data aggregation, and their presence in the evaluation guarantees that the success of the suggested SMORP routing and IECC-ELGDS security framework are performance contributions of the evaluation. Cluster heads use a fixed cost of data-aggregation (E_{DA})and L-sensor nodes send un aggregated values to the corresponding CHs before they are processed securely. In order to reduce bias in statistics due to randomly selected nodes or the sequence of events, every experiment was repeated a couple of times and the mean of the outcomes was published. The same three-tier hierarchy of communication L-sensors, CHs and sink was determined in all simulations, and the ranges of L-sensors and CHs transmission were 20 m and 80 m, respectively, to indicate the heterogeneous network energy capacity. Table 3 gives a full overview of all the parameters of the simulation considered in the evaluation.

Table 3. Parameters of the simulation

	Parameters	Value
	Topographical area (meters)	(100 m×100 m)
	Sink location (meters)	(50 m×85 m)
	Control packet length	2 k
L-Sensors	No. of transmission packets (rounds)	2×10^3
	No. of SMORP , FSEP, and LEACH	100
	Distance limit for transmission	20 m
	Initial energy	0.5 J
CH	No. of SMORP and SEP	5
	Distance limit for transmission	80 m
	Initial energy	2.5 J
	Energy data aggregate	5 nJ/bit

4.4. PERFORMANCE METRICS

In order to provide a rigorous and reproducible analysis of the suggested SMORP-IECC-ELGDS framework, the following subsection offers the performance metrics applied in the course of the simulation study. Both metrics will be given a definition, a mathematical formula, and a clear explanation of all the variables. All of these measures evaluate the efficiency of routing, energy sensitivity, latency response, network lifetime, and computational cost associated with the built-in security solutions.

4.4.1. Network Lifetime

Network lifetime is a measure reflecting the efficiency of the routing structure concerning the consumption of energy and the balancing of consumption amongst heterogeneous nodes. Two indicators are adopted, which are:

- **First Node Dead (FND)**: the round at which the first sensor exhausts its energy, reflecting the stability period of the network. The earliest time at which any L-sensor exhausts its energy as shown in Eq. (7)

$$FND = r\{\min|E_i(r) = 0, i = 1, \ldots, N\}\cdot(7) \quad (7)$$

Where $E_i(r)$ denotes the residual energy of node i at round r, and N is the total number of deployed sensors. FND is especially important in HWSNs where the loss of even a single L-sensor creates a sensing void.

- **Last Node Dead (LND):** Denotes the round index at which the final remaining node exhausts its residual energy. With the help of this metric, the maximum sustainable lifetime of the network can be measured and how well the energy consumption is distributed between L-sensors and CHs. A larger LND means better load balancing and greater duration of full-network operation as includes in the Eq. 8.

$$LND = r\{\max|E_i(r) = 0\} \quad (8)$$

4.4.2. Average Residual Energy (ARE)

In the same way that equation (9) is used to compute the arithmetic mean of the remaining energy of all the nodes in each round, we get a worldwide view of what network sustainability is doing. Higher ARE values indicate that the proposed routing and secure-aggregation processes avoid concentrating energy consumption on specific nodes, especially CHs or high-traffic forwarders, which is critical for prolonging system lifetime in heterogeneous WSN environments.

$$ARE(r) = \frac{1}{N}\sum_{i=1}^{N} E_i(r) \quad (9)$$

4.4.3. Packet Delivery Ratio (PDR)

Packet Delivery Ratio quantifies reliability by measuring the ratio between the number of received packets and the number of packets generated by sensing nodes. It is defined as:

$$PDR = \frac{P_{recv}}{P_{sent}} \quad (10)$$

Where: P_{recv} is number of packets correctly received at the sink, and P_{sent} is total packets transmitted by L-sensors. A higher PDR reflects robustness against packet loss, interference, and route instability, as shown in Eq. (10).

4.4.4. End-to-End Delay (E2E)

End-to-End Delay measures the total time required for a data packet to travel from an L-sensor node to the sink through multi-hop aggregation. Let $t_{recv}(p)$ be the packet reception time at the sink and $t_{send}(p)$ be the packet transmission time at the source. E2E is defined as Eq. (11):

$$E2E = t_{recv}(p) - t_{send}(p) \tag{11}$$

Where: $t_{send}(p)$ is a time of the packet is generated, and $t_{recv}(p)$ is a time of the sink receives the packet. Lower delay indicates better routing efficiency and reduced congestion.

4.4.5. Data Aggregation Security Overhead

This measure represents the incremental cost of the security layer of the IECC-ELGDS compared to the base communication and calculation cost of the operation done by the scorecard through the aggregation and routing of this service. The overhead encompasses the power consumption of elliptic-curve-based encryption, generation of digital signature, verification of digital signatures, and aggregation of ciphertext undertaken during the routing path. The overall security overhead per message E_s in the form of equations is expressed as E_q (12):

$$E_{sec} = E_{enc} + E_{sig} + E_{ver} + E_{agg} \tag{12}$$

Where:

- E_{enc}: energy consumed by IECC encryption.
- E_{sig}: energy required for ELGDS signature generation.
- E_{ver}: verification cost at the sink.
- E_{agg}: cost of ciphertext aggregation at CHs.

A reduced security overhead gives a more efficient cryptography structure of heterogeneous WSNs. Comparative costs of the proposed IECC-ELGDS scheme and competent techniques (ECC-HE, IEKC, ECDH-RSA) are shown in Fig. 6.

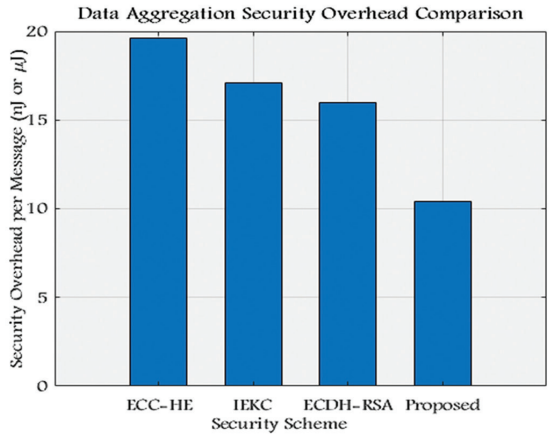


Fig. 6. Data Aggregation Security Overhead Components

The Fig. 6, integrates the security-processing terms and highlights their combined contribution to the overall secure-aggregation overhead, offering a concise visual summary of the protocol's security-related energy footprint.

4.5. COMPARATIVE EVALUATION FRAMEWORK

A single evaluation framework is used to conduct all routing and security schemes evaluated in this study to have a scientifically rigorous and unbiased comparison. The analysis represents the dualistic nature of the proposed solution-energy-efficient routing with the help of SMORP and secure data aggregation with the help of the IECC-ELGDS mechanism-and makes sure that the two are considered under the same identical and reproducible conditions. In the case of routing layer, SMORP has been compared to three of the well-known clustering based protocols: LEACH [10], SEP [11] and FSEP [12]. These baselines are heterogeneous-WSN fundamental routing methods, and employ similar radio-energy assumptions. The routing protocols are all carried out in the same deployment setting, initial energy distribution, transmission radii, and first-order radio parameters as in Sections 4.1 to 4.3. This ensures that performance difference is only due to behavior in an algorithm and not due to environmental variation. In case of the security layer, the IECC-ELGDS framework will be assessed on three exemplary examples of the cryptographic schemes that use the elliptic-curve: ECC-HE [13], IEKC [14] and ECDH-RSA [15]. These techniques are commonly used in the lightweight secure aggregation step and hence form the right baselines. Each security scheme uses the same message size, computation assumptions and traffic loads, which makes one directly compare the cost of encryption, signature-generation overhead, verification effort, and a general impact on network lifetime. Comparative analysis will be based on the standardized performance measures reported in section 4.4. such as network lifetime, residual-energy distribution, packet-delivery behavior, end-to-end delay and total security-processing overhead. The evaluation framework offers a clear and reproducible framework on isolating the actual performance contribution of the both SMORP routing and the IECC-ELGDS secure aggregation because all the simulated methods were fully parametrically consistent. Simulation environment, radio-energy model, parameter institution and evaluation metrics collectively provide a single and methodologically equal platform of assessment. Every routing and security plan has been implemented under exactly the same conditions in order to post fairness, transparency and reproducibility. Having this background, the following section forms the results of detailed performance and analysis of results and comparison of the performance between the proposed SMORP routing strategy and the IECC-ELGDS secure aggregation mechanism.

5. RESULT AND PERFORMANCE ANALYSIS

In this section, the performance outcomes of the recommended integrated framework are presented in the single simulation scenario that is described by Section 4. Each and every routing scheme and security scheme were tested under the same deployment conditions and radio-energy parameters in order to compare the schemes fairly. The evaluation is based on routing efficiency, energy behavior, delay behavior, packet-delivery reliability, and secure aggregation impact, which allow giving a clear understanding of the gains made by using SMORP routing along with the IECC-ELGDS security mechanism.

5.1. SMORP ROUTING PERFORMANCE

The proposed SMORP-based forwarding architecture is evaluated in terms of the routing throughput with three established clustering protocols: LEACH [10], FSEP [12], and SEP [11]. Each of the approaches has been implemented using the same simulation conditions and radio-energy parameters in order to make sure that the difference in the performance is due to the routing logic and not related to the environment itself. The assessment concerns three main metrics of routing effectiveness, namely network lifetime, residual-energy behavior and end-to-end delay.

5.1.1. Network Lifetime

Figs. 7 and 8 show the number of active L-sensor nodes and CHs over successive transmission rounds. SMORP demonstrates a substantially longer operational duration compared to LEACH, SEP, and FSEP.

Table 4. Number of rounds with the first dead node based on the four approaches

Approaches	LEACH	FSEP	SEP	SMORP
First dead L-sensor lifetime (Rounds)	176	293	—	975
First dead CHs Lifetime (Rounds)	—	—	377	1046

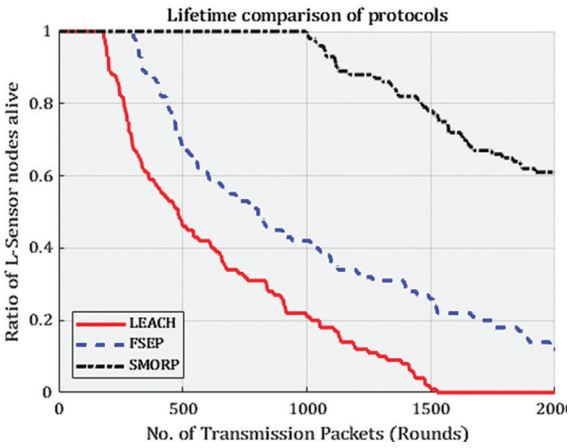


Fig. 7. Ratio of L-sensors still alive on different approaches (LEACH, FSEP, and proposed)

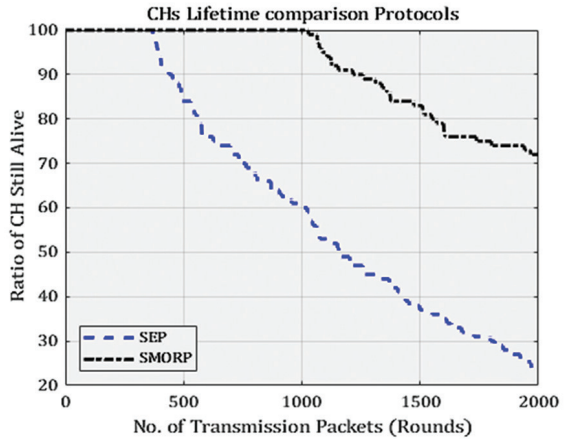


Fig. 8. Ratio of CHs still alive on different approaches (SEP, and proposed)

For L-sensors, the first node dies after 975 rounds, representing an improvement of approximately (+35% and +47%) over the approaches (FSEP and LEACH) respectively. For CHs, SMORP prolongs the first-dead-node lifetime by nearly 34% relative to SEP. These numerical results are summarized in Table 4 that obviously demonstrates that the proposed SMORP protocol provides the longest first-dead lifetime of all the considered strategies which proves its better capability to postpone the early failures of nodes and provide stable sensing coverage. These improvements directly align with the analytical formulations presented in Section 4. In particular, the prolonged survival time of SMORP nodes is explained by the radio-energy dissipation model (Eqs. (4)–(6)), where transmission cost grows quadratically with distance. Because SMORP continuously selects forwarding nodes with favorable spatial positions and sufficient residual energy according to the fitness and probability functions defined in Eqs. (1)–(3) the protocol naturally avoids high-cost transmissions and balances energy depletion across the network. This analytical grounding clarifies why SMORP maintains a larger population of active nodes across all rounds and achieves significantly longer network lifetime than LEACH, SEP, and FSEP.

Fig. 7 illustrates the ratio of active L-sensor nodes over the simulation rounds for the evaluated routing schemes. It can be observed that the proposed SMORP-based approach maintains a higher number of alive L-sensors compared to LEACH and FSEP throughout the network operation. This behavior indicates a more balanced energy consumption pattern, where forwarding and clustering decisions avoid overburdening individual nodes, thereby delaying early node failures and extending the stability period of the network.

5.1.2. Residual Energy Behavior

Figs. 9 and 10 illustrate the remaining energy ratio for L-sensors and CHs under the evaluated protocols. SMORP consistently maintains a higher residual-energy profile throughout the simulation.

This is attributed to:

- multi-hop forwarding guided by fitness-based decisions.
- adaptive subgroup organization.
- balanced load distribution between L-sensors and CHs.

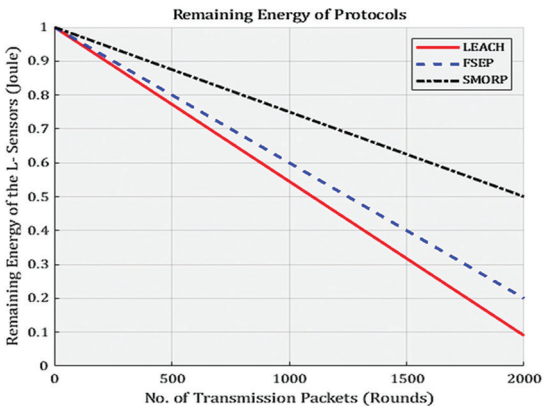


Fig. 9. Ratio remaining energy of L-sensors on different approaches (LEACH, FSEP, and proposed)

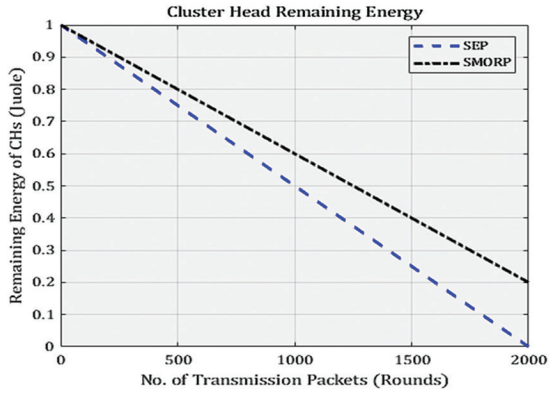


Fig. 10. Ratio remaining energy of H-sensors on the approaches (SEP, and proposed)

The results verify that SMORP achieves a more uniform energy-depletion pattern, preventing premature exhaustion of heavily loaded nodes and ensuring stable cluster performance.

Fig. 9 presents the remaining energy ratio of L-sensor nodes under different routing approaches. The proposed SMORP-based routing maintains a higher residual energy level throughout the simulation compared to LEACH and FSEP. This trend indicates that energy consumption is more evenly distributed among L-sensors, reducing excessive energy drain on individual nodes and supporting prolonged network operation.

5.1.3. End-to-End Delay and Transmission Efficiency

Figs. 11 and 12 demonstrate that SMORP significantly reduces simulation time and end-to-end delay compared to the baseline protocols.

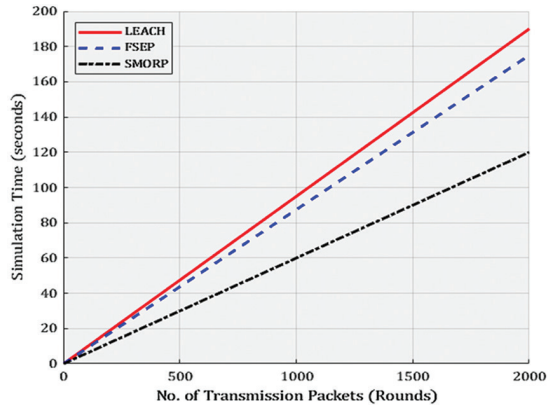


Fig. 11. Data transmission delay (simulation time for all packets) on different approaches (LEACH, FSEP, and proposed)

Packets experience fewer retransmissions and shorter forwarding paths due to:

- optimal next-hop selection via fitness evaluation.
- avoidance of overloaded or low-energy nodes.
- hierarchical pack-pointer–based path construction.

Lower delay directly translates to reduce per-packet energy expenditure, reinforcing the protocol's overall efficiency.

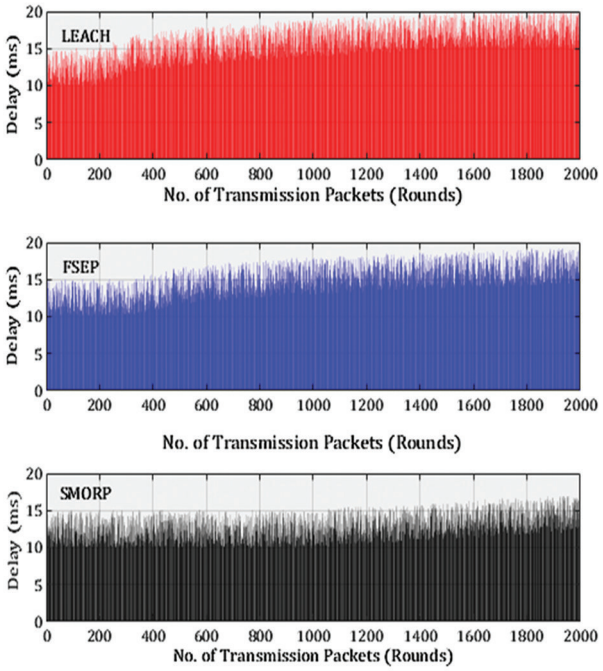


Fig. 12. Number of hops (end to end delay) on different approaches (LEACH, FSEP, and proposed)

5.1.4. Cluster Stability

Fig. 13 shows that SMORP preserves a near-optimal number of CHs over time, unlike LEACH and FSEP, which exhibit unstable CH formation patterns. Stable CH counts result in:

- predictable cluster structures.
- efficient aggregation.
- reduced control-message overhead.

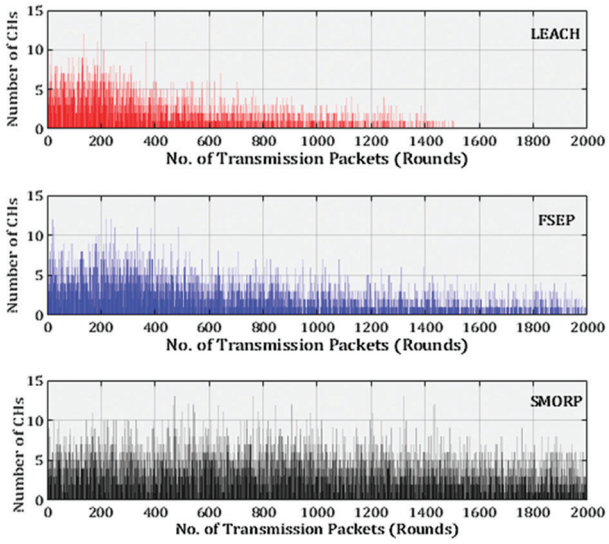


Fig. 13. Number of CHs on different approaches (LEACH, FSEP, and proposed)

5.1.5. Packet Delivery Dynamics

Figs. 14 and 15 indicate that SMORP achieves higher packet-delivery rates to both CHs and the sink. This reflects effective routing-path stability and reduced node failures, enabling reliable data flow across the network.

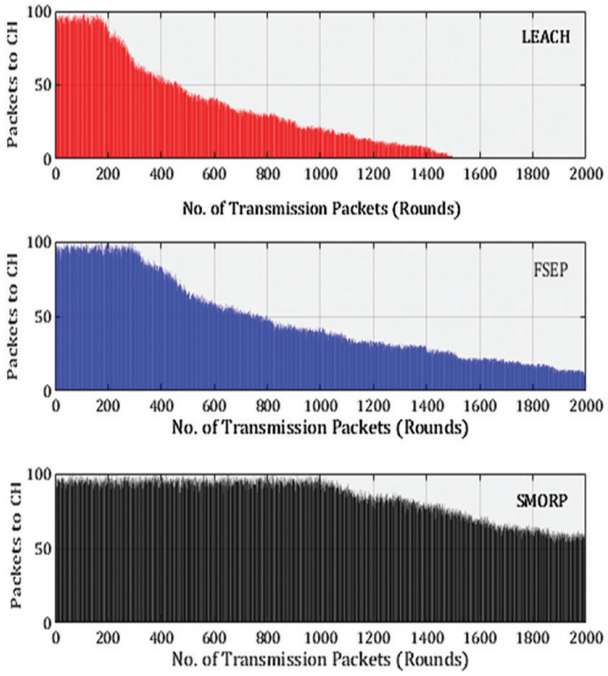


Fig. 14. Number of packets transmitted to CHs on different approaches (LEACH, FSEP, and proposed)

The collective results confirm that SMORP outperforms LEACH, SEP, and FSEP across all major routing-performance metrics. Its hierarchical leader-coordination, fitness-based next-hop evaluation, and balanced energy exploitation significantly enhance network lifetime, stability, and data-delivery reliability.

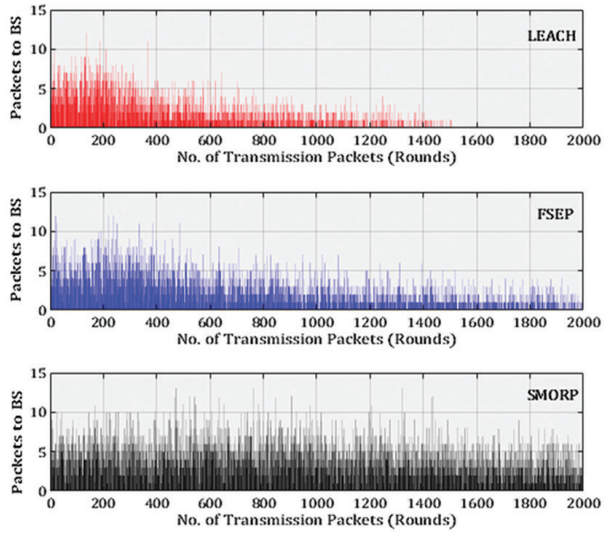


Fig. 15. Number of packets transmitted to BS on different approaches (LEACH, FSEP, and proposed)

5.2. IECC–ELGDS SECURITY PERFORMANCE

The effectiveness of the proposed IECC–ELGDS security mechanism is evaluated by comparing it with three well-known elliptic-curve-based security schemes: ECC-HE[13], IEKC [14], and ECDH-RSA [15]. All approaches were executed under identical simulation conditions, traffic load, aggregation structure, and cryptographic assumptions to ensure that performance differences arise solely from each method's security-processing efficiency. The evaluation focuses on four primary indicators: secure network lifetime, residual energy sustainability, encryption/decryption computational cost, and packet-delivery behavior under secure transmission.

5.2.1. Secure Network Lifetime

Fig. 16 shows the number of active sensor nodes under each security scheme. The proposed IECC–ELGDS mechanism maintains a significantly larger population of active nodes across all simulation rounds. The lifetime of the first-dead-node of the proposed method is 1223 rounds; this improvement is about (+44%), (+39%), and (+28%) over ECC-HE, IEKC, and ECDH-RSA. This finding, summarized in Table 5, affirms that both the lightweight Ness of the scalar- multiplication of the IECC and the lower computational intensity of ELGDS reduce security overheads and prevent node death alike.

Table 5. Number of rounds with the first dead node based on the four approaches

Approaches	ECC-HE	IEKC	ECDH-RSA	Proposed
First dead Lifetime (rounds)	342	451	659	1223

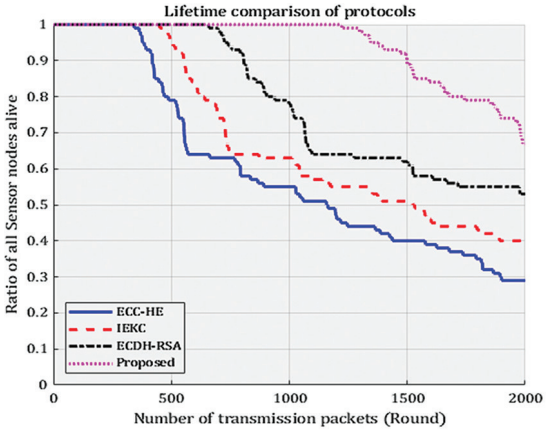


Fig. 16. Ratio of all sensor nodes still alive on different approaches (ECC-HE, IEKC, ECDH-RSA, and proposed)

5.2.2. Residual Energy Behavior Under Secure Processing

The residual energy trajectory, presented in Fig. 17, demonstrates that IECC–ELGDS preserves energy more effectively than the benchmark schemes.

ECC-HE and ECDH-RSA incur substantially higher cryptographic costs due to their use of homomorphic.

Fig. 16 shows the proportion of sensor nodes remaining alive under different security mechanisms. The proposed IECC–ELGDS scheme sustains a larger number of active nodes over the simulation rounds compared to ECC-HE, IEKC, and ECDH-RSA. This outcome reflects the reduced computational and energy overhead of the proposed security design, which limits premature energy depletion caused by cryptographic operations. Conversely, the optimized elliptic-curve

operations in IECC, combined with the single-round signature generation of ELGDS, reduce the per-packet cryptographic burden. This efficient processing yields a smoother energy-decline pattern and delays the onset of critical-energy states across sensor nodes.

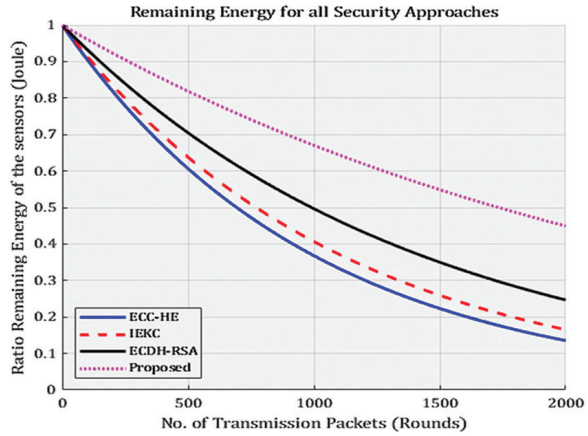


Fig. 17. Ratio remaining energy of all sensor nodes on different approaches (ECC-HE, IEKC, ECDH-RSA, and proposed)

5.2.3. Encryption/Decryption Cost Analysis

Fig. 18 compares the computational cost associated with ciphertext generation and recovery. The proposed IECC–ELGDS method consistently achieves the smallest processing cost for all evaluated data sizes. The improved ECC scalar multiplication in IECC and the two-step linear-modular computation of ELGDS require fewer arithmetic operations per message than the multi-layer encrypt–aggregate–decrypt structure used in ECC-HE and the RSA-based verification in ECDH-RSA. This lightweight operation significantly lowers both encryption and decryption delays, enabling faster secure forwarding and reduced energy expenditure.

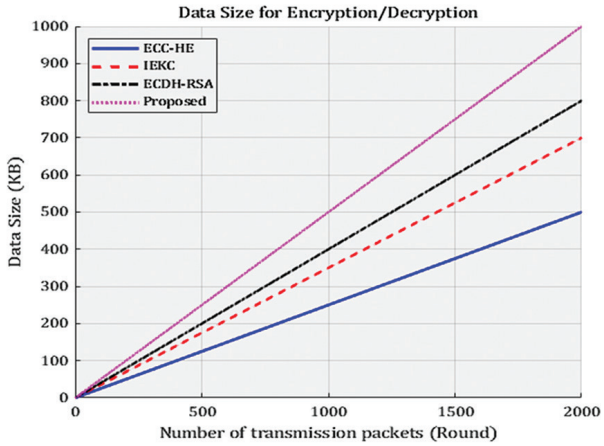


Fig. 18. Data Size for Encryption / Decryption on different approaches (ECC-HE, IEKC, ECDH-RSA, and proposed)

Fig. 18 compares the encryption and decryption cost of different security schemes. The proposed IECC–ELGDS approach exhibits lower computational overhead than ECC-HE and ECDH-RSA due to lightweight elliptic-curve operations and reduced cryptographic processing.

5.2.4. Secure Packet-Delivery Characteristics

Fig. 19 presents the secure packet-delivery time for all approaches. The IECC–ELGDS mechanism demonstrates the shortest end-to-end secure transmission delay, attributable to two primary factors:

1. Intermediate nodes forward ciphertext without decryption, eliminating the overhead associated with hop-by-hop key operations.
2. Signature verification is restricted to the sink, reducing per-hop processing costs and mitigating congestion on forwarding nodes.

Consequently, the suggested approach will have a better delivery ratio even under conditions of multipath forwarding. The level of packet-delivery is always higher as it is impossible to conduct opportunistic manipulation: ciphertext aggregation into CH prevents the exposure of plaintexts, whereas ELGDS authentication precludes replay, impersonation, and other forgery. Across all performance indicators—energy sustainability, secure lifetime, processing cost, and secure delivery behavior—the **IECC–ELGDS** framework consistently outperforms existing ECC-based security schemes. These results validate that the combined lightweight elliptic-curve encryption and efficient digital-signature generation deliver strong confidentiality and authentication guarantees while preserving network longevity in heterogeneous WSN environments.

5.3. COMPARATIVE DISCUSSION

Section 5.1 results, together with those in 5.2, reveal that the suggested system is able to simultaneously enhance the routing efficiency and secure data aggregation two notions that are usually at odds in resource-limited wireless sensor networks.

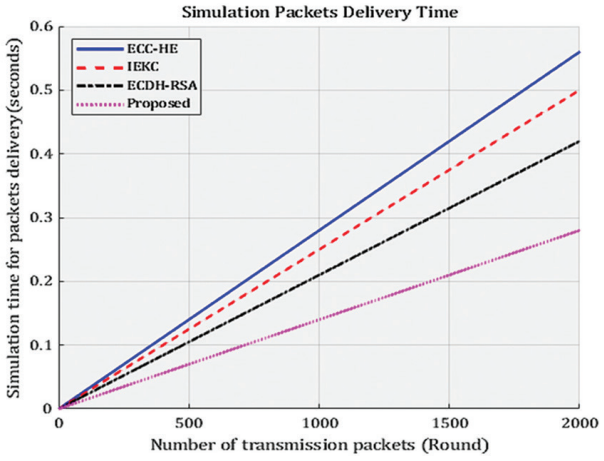


Fig. 19. Simulation Packets Delivery Time on different approaches (ECC-HE, IEKC, ECDH-RSA, and proposed)

Compared with the previous strategy of using idle routes to balance the end-to-end delay, SMORP-based routing strategy offers a longer lifetime of operation as indicated by the long first-dead-node intervals and clearer residual-energy curves in Figs. 7-14. At the same time, the IECC-ELGDS security layer can provide a high grade of confidentiality, as well as authentication assurances without impacting the energy-awareness of the underlying routing structure, which is

superior to ECC-HE, IEKC, and ECDH-RSA in all the security-related metrics than in Figs. 16-19. Of critical observation is the fact that the implementation of SMORP and IECC-ELGDS does not present a harmful trade-off between the performance and security- a problem that is common in the context of WSNs. Rather, the energy and computational cost of the secure protocols is significantly lowered by the lightweight elliptic-curve encryption and single round ElGamal signature generation. This compression makes SMORP retain its routing performance even as they operate in the secure mode, and the network can deliver a high ratio of packets and less latency than traditional schemes. Moreover, the suggested system makes sure that the aggregation of ciphertext at CHs is non-decrypted so that no data loss at intermediary nodes can be realized, not to create bottlenecks in cryptography. This design means that the multi-hop forwarding paths can have low levels of congestion, which combined with the delay performance shown in Fig. 12 and Fig. 19 increase concurrently. The overall gains, backed with the performances in Tables 3 and 4, prove the fact that the suggested integration is no longer than the sum of two mechanisms but an integrated structure in which routing intelligence and lightweight security improve one another. In general, the comparative results demonstrate that the presented framework of the SMORP-IECC-ELGDS approach provides a balanced and scalable solution that could be used to maintain the secure and energy-efficient functioning throughout the period of network existence. This twin improvement denotes the appropriateness of this proposed model to the scenario of heterogeneous sensing surroundings that are high reliability and high security warranties.

5.4. OVERALL INTERPRETATION

The combined experimental results which are obtained indicate that a balanced improvement in both energy conservation as well as safe data aggregation-two goals which generally clash in heterogeneous WSNs is realized when SMORP routing is integrated with the IECC-ELGDS security framework. SMORP tremendously enhances stabilization of routing, balances energy loss and prolongs the life of both L-sensor nodes and CHs and the IECC-ELGDS mechanism presents high level of confidentiality and authentication with low level of computer calculations. The joint effect of the optimization-based routing behavior and the lightweight elliptic-curve cryptography is in facilitating security in communication without negatively impacting network responsiveness or delay. The profile of the overall performance shows that despite the severe energy and security conditions, the proposed architecture is stable and proves its applicability to the long-term sensing applications in the heterogeneous environment with resource restrictions.

6. CONCLUSION

The obtained simulation results confirm that the proposed SMORP–IECC–ELGDS framework improves network lifetime, energy distribution, and security efficiency when compared with existing routing and cryptographic schemes. This paper presented a combined architecture that takes the SMORP complimentary routing protocol founded on optimization and the IECC-ELGDS delicate security tool to handle the two fold difficulty in terms of energy conservation as well as safe mantle of data accumulation within a heterogeneous wireless sensor network. The proposed architecture can improve the stability of clustering, load balancing through forwarding, and end-to-end confidentiality and authentication without causing too much load on the resource-limited nodes. The experimental findings illustrate a evident improvement in performance: SMORP allows increasing the first-dead-node lifetime of L-sensors and CHs by up to 47 and 34 percent respectively in comparison to LEACH, FSEP, and SEP, whereas the scheme of IECC-ELGDS can enhance the length of the safe network by a factor of 28-44 percent relative to ECC-HE, IEKC, and ECDH-RSA. Such enhancements verify the supportability of the integration of optimization-based routing and lightweight elliptic-curve security. In spite of the fact that the estimation is based on the simulation analysis and presupposes that the nodes remain still with an idealized behavior of the channels, the real-hardware validation, the adversarial attack models, and the scenarios of dynamic networks are to be included in the range of the future work to assess scalability and resilience further. Future work may consider extending the proposed framework to dynamic network scenarios, incorporating mobile sinks, and evaluating performance under realistic channel and attack models.

7. REFERENCES

[1] L.-L. Hung, F.-Y. Leu, K.-L. Tsai, C.-Y. Ko, "Energy-efficient cooperative routing scheme for heterogeneous wireless sensor networks", IEEE Access, Vol. 8, 2020, pp. 56321-56332.

[2] H. Qabouche, A. Sahel, A. Badri, "Hybrid energy efficient static routing protocol for homogeneous and heterogeneous large scale WSN", Wireless Networks, Vol. 27, No. 1, 2021, pp. 575-587.

[3] S. K. Chaurasiya, S. Mondal, A. Biswas, A. Nayyar, M. A. Shah, R. Banerjee, "An energy-efficient hybrid clustering technique (EEHCT) for IoT-based multilevel heterogeneous wireless sensor networks", IEEE Access, Vol. 11, 2023, pp. 25941-25958.

[4] U. Chatterjee, S. Ray, M. K. Khan, M. Dasgupta, C.-M. Chen, "An ECC-based lightweight remote user authentication and key management scheme for IoT communication in context of fog computing", Computing, Vol. 104, No. 6, 2022, pp. 1359-1395.

[5] S. X. Pushpa, S. K. S. Raja, "Enhanced ECC based authentication protocol in wireless sensor network with DoS mitigation", Cybernetics and Systems, Vol. 53, No. 8, 2022, pp. 734-755.

[6] S. Hu, L. Liu, L. Fang, F. Zhou, R. Ye, "A novel energy-efficient and privacy-preserving data aggregation for WSNs", IEEE Access, Vol. 8, 2019, pp. 802-813.

[7] M. A. Khan, M. T. Quasim, N. S. Alghamdi, M. Y. Khan, "A secure framework for authentication and encryption using improved ECC for IoT-based medical sensor data", IEEE Access, Vol. 8, 2020, pp. 52018-52027.

[8] H. Y. Adarbah, M. F. Moghadam, R. L. R. Maata, A. Mohajerzadeh, A. H. Al-Badi, "Security challenges of selective forwarding attack and design a secure ECDH-based authentication protocol to improve RPL security", IEEE Access, Vol. 11, 2022, pp. 11268-11280.

[9] X. Yang *et al*. "Blockchain-based secure and lightweight authentication for Internet of Things", IEEE Internet of Things Journal, Vol. 9, No. 5, 2021, pp. 3321-3332.

[10] W. B. Heinzelman, A. P. Chandrakasan, H. Balakrishnan, "An application-specific protocol architecture for wireless microsensor networks", IEEE Transactions on Wireless Communications, Vol. 1, No. 4, 2002, pp. 660-670.

[11] G. Smaragdakis, I. Matta, A. Bestavros, "SEP: A stable election protocol for clustered heterogeneous wireless sensor networks", Boston University, Computer Science Department, Boston, MA, USA, 2004.

[12] A. Ali *et al*. "Enhanced fuzzy logic zone stable election protocol for cluster head election (E-FLZSEPFCH) and multipath routing in wireless sensor networks", Ain Shams Engineering Journal, Vol. 15, No. 2, 2024, p. 102356.

[13] M. Elhoseny, H. Elminir, A. Riad, X. Yuan, "A secure data routing schema for WSN using elliptic curve cryptography and homomorphic encryption", Journal of King Saud University-Computer and Information Sciences, Vol. 28, No. 3, 2016, pp. 262-275.

[14] P. Ramadevi, S. Ayyasamy, Y. Suryaprakash, C. Anilkumar, S. Vijayakumar, R. Sudha, "Security for wireless sensor networks using cryptography", Measurement: Sensors, Vol. 29, 2023, p. 100874.

[15] B. Abood, A. N. Faisal, Q. A. Hamed, "Data transmitted encryption for clustering protocol in heterogeneous wireless sensor networks", Indonesian Journal of Electrical Engineering and Computer Science, Vol. 25, No. 1, 2022, pp. 347-357.

[16] A. H. Jabbar, I. S. Alshawi, "Spider monkey optimization routing protocol for wireless sensor networks", International Journal of Electrical & Computer Engineering, Vol. 11, No. 3, 2021, pp. 2432-2442.

[17] G. Muneeswari, A. Ahilan, R. Rajeshwari, K. Kannan, C. J. C. Singh, "Trust and energy-aware routing protocol for wireless sensor networks based on secure routing", International Journal of Electrical and Computer Engineering Systems, Vol. 14, No. 9, 2023, pp. 1015-1022.

[18] S. Balan, D. Champla, M. Pushpavalli, A. Ahilan, "Energy Efficient Multi-hop routing scheme using Taylor based Gravitational Search Algorithm in Wireless Sensor Networks", International Journal of Electrical and Computer Engineering Systems, Vol. 14, No. 3, 2023, pp. 333-343.

[19] V. Lekshmi, "Increasing efficiency and reliability in multicast routing based V2V communication for direction-aware cooperative collision avoidance", International Journal of Electrical and Computer Engineering Systems, Vol. 15, No. 2, 2024, pp. 145-153.

[20] M. Rami Reddy, M. Ravi Chandra, P. Venkatramana, R. Dilli, "Energy-efficient cluster head selection in wireless sensor networks using an improved grey wolf optimization algorithm", Computers, Vol. 12, No. 2, 2023, p. 35.

[21] F. Jibreel, E. Tuyishimire, M. I. Daabo, "An enhanced heterogeneous gateway-based energy-aware multi-hop routing protocol for wireless sensor networks", Information, Vol. 13, No. 4, 2022, p. 166.

[22] S. Tabatabaei, "Provide energy-aware routing protocol in wireless sensor networks using bacterial foraging optimization algorithm and mobile sink", Plos one, Vol. 17, No. 3, 2022, p. e0265113.

[23] B. Hammi, A. Fayad, R. Khatoun, S. Zeadally, Y. Begriche, "A lightweight ECC-based authentication scheme for Internet of Things (IoT)", IEEE Systems Journal, Vol. 14, No. 3, 2020, pp. 3440-3450.

[24] N. Mahlake, T. E. Mathonsi, D. Du Plessis, T. Muchenje, "A Lightweight Encryption Algorithm to Enhance Wireless Sensor Network Security on the Internet of Things", Journal of Communications, Vol. 18, No. 1, 2023, pp. 47-57.

[25] H. Bashirpour, S. Bashirpour, S. Shamshirband, A. T. Chronopoulos, "An improved digital signature protocol to multi-user broadcast authentication based on elliptic curve cryptography in wireless sensor networks (WSNS)", Mathematical and Computational Applications, Vol. 23, No. 2, 2018, p. 17.

[26] B. R. Rao, B. Sujatha, "A hybrid elliptic curve cryptography (HECC) technique for fast encryption of data for public cloud security", Measurement: Sensors, Vol. 29, 2023, p. 100870.

[27] I. Ahmad *et al*. "Adaptive and Priority-Based Data Aggregation and Scheduling Model for Wireless Sensor Network", Knowledge-Based Systems, Vol. 303, 2024, p. 112393.

[28] X. Liu, J. Yu, K. Yu, G. Wang, X. Feng, "Trust secure data aggregation in WSN-based IIoT with single mobile sink", Ad Hoc Networks, Vol. 136, 2022, p. 102956.

INTERNATIONAL JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING SYSTEMS

Published by Faculty of Electrical Engineering, Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia.

About this Journal

The International Journal of Electrical and Computer Engineering Systems publishes original research in the form of full papers, case studies, reviews and surveys. It covers theory and application of electrical and computer engineering, synergy of computer systems and computational methods with electrical and electronic systems, as well as interdisciplinary research.

Topics of interest include, but are not limited to:

- Power systems
- Renewable electricity production
- Power electronics
- Electrical drives
- Industrial electronics
- Communication systems
- Advanced modulation techniques
- RFID devices and systems
- Signal and data processing
- Image processing
- Multimedia systems
- Microelectronics
- Instrumentation and measurement
- Control systems
- Robotics
- Modeling and simulation
- Modern computer architectures
- Computer networks
- Embedded systems
- High-performance computing
- Parallel and distributed computer systems
- Human-computer systems
- Intelligent systems
- Multi-agent and holonic systems
- Real-time systems
- Software engineering
- Internet and web applications and systems
- Applications of computer systems in engineering and related disciplines
- Mathematical models of engineering systems
- Engineering management
- Engineering education

Paper Submission

Authors are invited to submit original, unpublished research papers that are not being considered by another journal or any other publisher. Manuscripts must be submitted in doc, docx, rtf or pdf format, and limited to 30 one-column double-spaced pages. All figures and tables must be cited and placed in the body of the paper. Provide contact information of all authors and designate the corresponding author who should submit the manuscript to https://ijeces.ferit.hr. The corresponding author is responsible for ensuring that the article's publication has been approved by all coauthors and by the institutions of the authors if required. All enquiries concerning the publication of accepted papers should be sent to ijeces@ferit.hr.

The following information should be included in the submission:

- paper title;
- full name of each author;
- full institutional mailing addresses;
- e-mail addresses of each author;
- abstract (should be self-contained and not exceed 150 words). Introduction should have no subheadings;
- manuscript should contain one to five alphabetically ordered keywords;
- all abbreviations used in the manuscript should be explained by first appearance;
- all acknowledgments should be included at the end of the paper:
- authors are responsible for ensuring that the information in each reference is complete and accurate. All references must be numbered consecutively and citations of references in text should be identified using numbers in square brackets. All references should be cited within the text;
- each figure should be integrated in the text and cited in a consecutive order. Upon acceptance of the paper, each figure should be of high quality in one of the following formats: EPS, WMF, BMP and TIFF;
- corrected proofs must be returned to the publisher within 7 days of receipt.

Peer Review

All manuscripts are subject to peer review and must meet academic standards. Submissions will be first considered by an editor-in-chief and if not rejected right away, then they will be reviewed by anonymous reviewers. The submitting author will be asked to provide the names of 5 proposed reviewers including their e-mail addresses. The proposed reviewers should be in the research field of the manuscript. They should not be affiliated to the same institution of the manuscript author(s) and should not have had any collaboration with any of the authors during the last 3 years.

Author Benefits

The corresponding author will be provided with a .pdf file of the article or alternatively one hardcopy of the journal free of charge.

Units of Measurement

Units of measurement should be presented simply and concisely using System International (SI) units.

Bibliographic Information

Commenced in 2010.
ISSN: 1847-6996
e-ISSN: 1847-7003

Published: semiannually

Copyright

Subscription Information

The annual subscription rate is 50€ for individuals, 25€ for students and 150€ for libraries.

Postal Address

Faculty of Electrical Engineering,
Computer Science and Information Technology Osijek,
Josip Juraj Strossmayer University of Osijek, Croatia
Kneza Trpimira 2b
31000 Osijek, Croatia

IJECES Copyright Transfer Form

(Please, read this carefully)

This form is intended for all accepted material submitted to the IJECES journal and must accompany any such material before publication.

TITLE OF ARTICLE (hereinafter referred to as "the Work"):

COMPLETE LIST OF AUTHORS:

The undersigned hereby assigns to the IJECES all rights under copyright that may exist in and to the above Work, and any revised or expanded works submitted to the IJECES by the undersigned based on the Work. The undersigned hereby warrants that the Work is original and that he/she is the author of the complete Work and all incorporated parts of the Work. Otherwise he/she warrants that necessary permissions have been obtained for those parts of works originating from other authors or publishers.

Authors retain all proprietary rights in any process or procedure described in the Work. Authors may reproduce or authorize others to reproduce the Work or derivative works for the author's personal use or for company use, provided that the source and the IJECES copyright notice are indicated, the copies are not used in any way that implies IJECES endorsement of a product or service of any author, and the copies themselves are not offered for sale. In the case of a Work performed under a special government contract or grant, the IJECES recognizes that the government has royalty-free permission to reproduce all or portions of the Work, and to authorize others to do so, for official government purposes only, if the contract/grant so requires. For all uses not covered previously, authors must ask for permission from the IJECES to reproduce or authorize the reproduction of the Work or material extracted from the Work. Although authors are permitted to re-use all or portions of the Work in other works, this excludes granting third-party requests for reprinting, republishing, or other types of re-use. The IJECES must handle all such third-party requests. The IJECES distributes its publication by various means and media. It also abstracts and may translate its publications, and articles contained therein, for inclusion in various collections, databases and other publications. The IJECES publisher requires that the consent of the first-named author be sought as a condition to granting reprint or republication rights to others or for permitting use of a Work for promotion or marketing purposes. If you are employed and prepared the Work on a subject within the scope of your employment, the copyright in the Work belongs to your employer as a work-for-hire. In that case, the IJECES publisher assumes that when you sign this Form, you are authorized to do so by your employer and that your employer has consented to the transfer of copyright, to the representation and warranty of publication rights, and to all other terms and conditions of this Form. If such authorization and consent has not been given to you, an authorized representative of your employer should sign this Form as the Author.

Authors of IJECES journal articles and other material must ensure that their Work meets originality, authorship, author responsibilities and author misconduct requirements. It is the responsibility of the authors, not the IJECES publisher, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it.

- The undersigned represents that he/she has the authority to make and execute this assignment.
- For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.
- The undersigned agrees to indemnify and hold harmless the IJECES publisher from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.

_____ _____

Author/Authorized Agent **Date**

CONTACT
International Journal of Electrical and Computer Engineering Systems (IJECES)
Faculty of Electrical Engineering, Computer Science and Information Technology Osijek
Josip Juraj Strossmayer University of Osijek
Kneza Trpimira 2b
31000 Osijek, Croatia
Phone: +38531224600,
Fax: +38531224605,
e-mail: ijeces@ferit.hr