International Journal of Synergy in Engineering and Technology Vol.6 No.2 (2025) 1-15

International Journal of Synergy in Engineering and Technology IJSET

Journal homepage: https://tatiuc.edu.my/ijset/index.php/ijset/ e-ISSN: 2735-0622

Quality-Aware Data Aggregation in Environmental WSNs Using Autoencoder Filtering

A.O. Adedokun[1*], F. M. Dahunsi[1], J. J. Popoola[1]

[1]Federal University of Technology Akure, Nigeria

[*]Corresponding author email: aayoadedokun@gmail.com

KEYWORDS

Wireless Sensor Networks
Environmental Monitoring
Data Aggregation
Autoencoder
Data Integrity

ABSTRACT

Environmental monitoring using Wireless Sensor Networks (WSNs) requires accurate and reliable data to support real-time decision-making. However, traditional data aggregation techniques often suffer from issues such as redundancy, noise, outliers, and unreliable readings, which compromise data integrity. This paper proposes a hybrid quality-aware data aggregation method that integrates autoencoder-based anomaly detection with rule-based filtering to enhance data fidelity. The autoencoder identifies anomalies through reconstruction error analysis, while the rule-based system applies sensor-specific thresholds to eliminate invalid values. The approach was implemented and simulated in MATLAB using real-world sensor data patterns. Performance was evaluated against conventional LEACH, PEGASIS, K-LEACH, and MABRL protocols based on metrics such as aggregation accuracy, packet delivery ratio, data loss rate, routing overhead, and end-to-end delay. Results show that the proposed method consistently achieves higher accuracy across various sensor types and significantly improves network reliability and communication efficiency. It also reduces unnecessary transmissions, leading to lower energy consumption and extended network lifespan. This work presents a robust solution for ensuring high-quality data aggregation in environmental WSN applications, contributing to more accurate and sustainable environmental monitoring systems.

Received 27 May 2025; Revised 14 August 2025; Accepted 31 August 2025; Published 31 October 2025.

1.0 INTRODUCTION

Environmental monitoring using Wireless Sensor Networks (WSNs) plays a pivotal role in collecting and analysing data critical for assessing air quality, temperature, humidity, and other ecological parameters [1]. These systems are especially useful in remote or hazardous areas where manual data collection is infeasible. However, the accuracy and reliability of environmental monitoring data depend heavily on the quality of the aggregated information from spatially distributed sensor nodes [2]. Ensuring data integrity is essential for effective environmental policy-making, pollution control, and early warning systems [3].

Despite advancements in sensor technology, WSNs are still prone to several data quality issues. These include data redundancy caused by overlapping sensing regions, sensor noise due to environmental interference, outliers introduced by hardware faults, and missing or delayed packets due to intermittent communication links [4], [5]. Such anomalies can distort decision-making processes, reduce model performance, and compromise the dependability of the monitoring system.

Traditional data aggregation techniques, such as averaging, median filtering, and majority voting, have been widely used to reduce transmission cost and compress data [6]. While these techniques help in reducing bandwidth and energy usage, they often compromise data quality by overlooking context, discarding useful variations, or retaining noise [7]. Furthermore, they are not adaptive to varying environmental conditions or sensor dynamics, which limits their effectiveness in real-world deployments.

To address these shortcomings, recent research efforts have shifted towards intelligent data preprocessing methods that combine statistical, machine learning, and signal processing techniques [8]. Among these, autoencoder-based filtering presents a promising solution. Autoencoders are unsupervised neural networks capable of learning compact representations of input data and reconstructing them with minimal loss [9]. Their reconstruction capability allows them to distinguish between normal and anomalous patterns, making them suitable for filtering noisy or corrupted sensor readings.

By learning the latent structure of environmental data, autoencoders can serve as dynamic filters that suppress noise, correct outliers, and preserve valid data variations. This approach not only improves the accuracy of aggregated data but also enables downstream tasks such as anomaly detection, environmental modelling, and predictive analytics to perform more effectively. Additionally, when combined with rule-based heuristics that incorporate domain knowledge (e.g., threshold-based outlier rules), the system can achieve enhanced robustness and interpretability.

This study aims to develop a quality-aware data aggregation framework using autoencoder-based filtering to enhance the accuracy and reliability of environmental monitoring in wireless sensor networks (WSNs). The framework is designed to intelligently identify and eliminate noise, outliers, and redundant data that commonly degrade the quality of sensor readings. It will be implemented and simulated in MATLAB, incorporating realistic environmental sensor data to assess its effectiveness. The performance of the proposed model will be compared with traditional aggregation methods, focusing on metrics such as aggregation accuracy, outlier detection, and data delivery reliability. Additionally, the framework is tested under various challenging conditions, including noisy and redundant data scenarios, to demonstrate its robustness.

By improving the fidelity of sensor data before transmission, the proposed solution helps reduce energy consumption and enhances the operational reliability of WSN-based environmental monitoring systems.

This paper is structured as follows: the introduction presents the motivation and relevance of quality-aware aggregation in WSNs. The related work section reviews existing aggregation methods and their limitations. The methodology section details the design and implementation of the autoencoder-based framework. The results and discussion section analyses the performance outcomes, and the final section provides conclusions and suggestions for future work.

.

2.0 RELATED WORKS

Wireless Sensor Networks (WSNs) have become indispensable in environmental monitoring due to their capability to collect and transmit data from diverse and often inaccessible terrains [10]. However, ensuring the quality and reliability of the aggregated data remains a significant challenge. This section reviews existing data aggregation techniques, highlighting their strengths and limitations, and explores the integration of machine learning methods, particularly autoencoders, to enhance data integrity in WSNs. One of the pioneering protocols in WSN data aggregation is the Low-

Energy Adaptive Clustering Hierarchy (LEACH) [11]. LEACH employs a hierarchical clustering approach where nodes self-organize into local clusters, with one node acting as the cluster head. These cluster heads aggregate data from their respective clusters and transmit the summarized information to the base station. While LEACH effectively reduces energy consumption and extends network lifetime, it has notable drawbacks [12]. The random selection of cluster heads can lead to uneven energy distribution, and the protocol does not account for the residual energy of nodes, potentially causing premature node failures and reduced data reliability [13].

To address some of LEACH's limitations, the Power-Efficient Gathering in Sensor Information Systems (PEGASIS) protocol was introduced [14]. PEGASIS forms a chain among sensor nodes so that each node communicates only with its closest neighbor, and a designated leader node transmits the aggregated data to the base station. This approach reduces the number of transmissions per node, thereby conserving energy [15]. However, PEGASIS can introduce delays due to the sequential data transmission along the chain, which may not be suitable for time-sensitive environmental monitoring applications [16].

Further enhancements led to the development of protocols like K-LEACH and MABRL. K-LEACH incorporates knowledge-based clustering, where nodes with higher residual energy and better communication capabilities are preferentially selected as cluster heads [17]. This strategy aims to balance energy consumption across the network. MABRL integrates machine learning techniques, specifically reinforcement learning, to optimize routing decisions dynamically. By learning from the network's operational patterns, MABRL can adapt to changing conditions, improving data delivery rates and reducing energy consumption [18].

Traditional aggregation protocols primarily focus on energy efficiency, often overlooking data quality aspects such as redundancy, noise, and outliers. To enhance data integrity, quality-aware aggregation methods have been proposed. Rule-based approaches utilize predefined thresholds to filter out anomalous data. While straightforward, they lack adaptability to dynamic environmental conditions [19].

Statistical methods, including moving averages and Kalman filters, estimate true sensor readings by smoothing out noise. These techniques can effectively handle random fluctuations but may struggle with sudden changes or non-linear patterns in the data [20].

Fuzzy logic-based aggregation introduces a degree of uncertainty handling, allowing for more flexible data interpretation [21]. By assigning degrees of membership to data points, fuzzy systems can better manage imprecise information. However, designing appropriate membership functions and rules can be complex and application-specific [22].

The advent of machine learning has opened new avenues for data validation and cleaning in WSNs. Supervised learning algorithms, such as Support Vector Machines (SVM) and Decision Trees, have been employed to classify and filter sensor data [23]. These methods require labeled datasets for training, which may not always be available in environmental monitoring scenarios [24].

Unsupervised learning techniques, particularly clustering algorithms like K-Means and DBSCAN, can identify patterns and anomalies without labeled data [25]. These methods are useful for detecting outliers and grouping similar data points, aiding in noise reduction and data consistency [26].

Deep learning models, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), have demonstrated superior performance in handling complex and high-dimensional data [27]. Their ability to learn hierarchical representations makes them suitable for capturing intricate patterns in sensor data. However, their computational demands pose challenges for deployment in resource-constrained WSN environments [28].

Autoencoders, a type of unsupervised neural network, have gained prominence in anomaly detection and data reconstruction tasks within WSNs [29]. An autoencoder learns to compress input data into a lower-dimensional representation and then reconstruct it, minimizing the reconstruction error. Data points with high reconstruction errors are considered anomalies, indicating potential faults or unusual events [30].

In environmental monitoring, autoencoders can effectively identify and filter out anomalous readings caused by sensor malfunctions or external interferences [31]. Their ability to learn from unlabeled data makes them particularly valuable in scenarios where labeled datasets are scarce. Moreover, autoencoders can adapt to changing environmental conditions by retraining on new data, ensuring sustained performance over time [32].

Recent advancements have led to the development of hybrid models combining autoencoders with other techniques to enhance anomaly detection capabilities. For instance, integrating autoencoders with Generative Adversarial Networks (GANs) has shown promise in improving detection accuracy and robustness [33]. These hybrid models leverage the strengths of both architectures, enabling more precise identification of anomalies in complex datasets [34].

Furthermore, the incorporation of autoencoders into routing protocols has been explored to improve data aggregation processes [35]. By filtering out anomalous data at the node level before aggregation, these approaches can enhance the overall quality of the transmitted information, reduce unnecessary data transmission, and conserve energy.

The evolution of data aggregation techniques in WSNs reflects a shift from solely energy-focused approaches to methods that also prioritize data quality. While traditional protocols like LEACH and PEGASIS laid the groundwork for energy-efficient communication, their limitations in handling data anomalies necessitated the development of more sophisticated methods. The integration of machine learning, particularly autoencoder-based models, offers a promising avenue for enhancing data integrity in environmental monitoring applications. These models' ability to detect anomalies and reconstruct accurate data positions them as valuable tools in the pursuit of reliable and efficient WSN operations.

3.0 METHODOLOGY

3.1 Mathematical Framework

Quality data delivery in Wireless Sensor Networks (WSN) involves ensuring that the transmitted data is accurate, reliable, and meets certain quality-of-service (QoS) criteria. In wireless sensor networks (WSNs) for environmental monitoring, redundant and erroneous data reduce energy efficiency and reliability. To counter this, a Quality-Aware Data Aggregation and Delivery approach is used at the Cluster Head (CH) level. It combines autoencoder-based anomaly detection with rule-based filtering to ensure high data fidelity. The autoencoder identifies unknown anomalies like air quality drift, while rule-based filters remove known errors using predefined thresholds. This hybrid method effectively minimizes data redundancy and transmission overhead, especially under challenging environmental conditions such as high humidity or fluctuating temperatures, thereby improving overall network performance and data reliability.

Given:
The enhanced environmental formulations of the first-order radio energy model for transmitter and receiver are presented in Equations (1) and (2) respectively.

$$E_{Tx}^{k}(L,d) = E_{lec} * L * (1 + \delta_T * T + \delta_C * C) + \varepsilon_{amp} * L * \left(d * \varphi_H\right)^2 * \left(1 + \delta_{AQ} * AQ\right) \quad (1)$$

$$E_{Rx}^{k}(L) = E_{lec} * L * (1 + \delta_T * T + \delta_C * C) \quad (2)$$

since:

$$\varphi_H = 1 + \vartheta_H * H \tag{3}$$

where L represents the packet length in bits, and d is the distance to the receiver. T, C, and AQ denote the normalized temperature, light intensity, and air quality index respectively. δT is the temperature degradation coefficient, accounting for energy inefficiency at higher temperatures. δC represents the light intensity impact factor, which models the effect of low light on solar-powered node efficiency, potentially increasing the duty-cycle energy. δAQ is the air quality priority factor, reflecting increased transmission frequency when air quality is poor. The humidity correction factor, $\varphi_H = 1 + \vartheta_H * H$, adjusts for signal attenuation due to humidity, where αH is the experimentally derived humidity sensitivity coefficient. ε_{amp} is the amplifier energy coefficient used in the radio model.

Autoencoders are trained using normal sensor data to learn a compressed representation. During inference, reconstruction error is used to detect anomalies.

Reconstruction Error (RE):

$$RE_{rr} = \left\| B_i - \widehat{B_i} \right\| \tag{4}$$

where B_i is the original input vector from sensor node, $\widehat{B_i}$ is the reconstructed output produced by the autoencoder, and RE_{rr} represents the reconstruction error. If the reconstruction error RE_{rr} exceeds a predefined threshold θ, the corresponding data is flagged as anomalous.
To ensure high data fidelity and minimize the influence of sensor inaccuracies and environmental noise, this component applies a set of deterministic rules during data aggregation at the Cluster Head (CH). These rules are derived from established sensor error models and environmental operating constraints, providing a simple but effective mechanism to discard invalid, noisy, or redundant sensor readings before aggregation.

The following are the filtering rules employed in this research to ensure the integrity and reliability of sensor data during aggregation:

i. If the air quality index AQ_i falls outside the valid operational range defined by the sensor specifications, the reading B_i is considered invalid due to possible sensor drift or out-of-range values and is discarded.
ii. If the temperature reading T_i falls outside the acceptable environmental range defined by the application or sensor specifications, the data B_i is considered invalid and is discarded to avoid thermal noise or faulty measurements.
iii. If the humidity value H_i exceeds the operational humidity tolerance limits of the sensor or falls outside the expected environmental bounds, the data B_i is flagged as unreliable and discarded due to potential condensation effects or sensor saturation.
iv. If the light intensity C_i is below the minimum detection threshold or above the sensor's calibrated maximum range, the reading B_i is treated as out-of-bounds and discarded to prevent misinterpretation caused by low-light noise or sensor overexposure.

After filtering, data from multiple nodes is aggregated using:

$$\widehat{B_i} = \frac{1}{n}\sum_{i=1}^{n} B_i \tag{5}$$

where n is the number of valid readings in the cluster and $\widehat{B}_i$is the aggregated value sent to the base station. The evaluation metrics of Packet Delivery Ratio (PDR), Data Loss Rate (DLR), and Aggregation Accuracy (AA) are formally defined in Equations (6), (7), and (8), respectively.

$$PDR = \frac{Packet\ Received}{Packet\ Sent} \tag{6}$$

$$DLR = 1 - PDR \tag{7}$$

$$AA = \frac{\|\widehat{B}_i - B_0\|}{B_0} \tag{8}$$

where $B_0$is the ground-truth environmental reading.

Algorithm: Quality-Aware Data Aggregation (Hybrid AAD + RBF)
Require:

i. Sensor data matrix $X = [X1, X2, \ldots, Xn]$
ii. Autoencoder model $AE()$
iii. Anomaly threshold θ
iv. Rule-based filter $R()$
v. Number of sensor nodes per cluster n

Ensure

- Aggregated quality data $\widehat{B}_i$ with high fidelity

1: Initialize valid data set $V \leftarrow \varnothing$
2: for $i = 1$ to n do
3: Compute reconstruction error $RE\_rr \leftarrow \|B\_i - AE(B\_i)\|^2$
4: if $RE\_rr < \theta$ then
5: if $R(B\_i) == TRUE$ then
6: $V \leftarrow V \cup \{B\_i\}$
7: end if
8: end if
9: end for
10: if $|V| > 0$ then
11: Compute aggregated data: \[\bar{B} = \frac{1}{|V|} \sum_{B_i ∈ V} B_i \]
12: else
13: Flag missing data and request re-transmission
14: end if
15: Transmit \(\bar{B}\) to base station
16: Return \(\bar{B}\)

This algorithm performs Quality-Aware Data Aggregation by first filtering sensor readings using an autoencoder-based anomaly detection mechanism and a rule-based validation function. Only data with low reconstruction error and that passes predefined rules are included in the valid set V, which is then averaged to produce an aggregated value $\widehat{B}_i$ for transmission to the base station.

3.2 Autoencoder Architecture and Training Configuration

The autoencoder used in the proposed Quality-Aware Data Aggregation framework was designed as a lightweight, resource-efficient neural network tailored for deployment on wireless sensor network aggregator nodes. It adopts a symmetric encoder–decoder structure to compress input data and reconstruct it with minimal loss, enabling anomaly detection through reconstruction error analysis.

In the architecture, the autoencoder processes four sensor features: Temperature, Humidity, Air Quality Index (AQI), and Light Intensity. Table 1 presents the layer configuration, neuron counts, and activation functions. The bottleneck layer captures the most essential information from the input while discarding noise, thereby facilitating efficient anomaly detection.

Table 1: Autoencoder Layer Configuration for 4 Sensor Features

Layer No.	Layer Type	No. of Neurons	Activation Function
Input	Dense	4 (sensor features: temperature, humidity, AQI, LI)	—
1	Encoder	3	ReLU
2	Encoder	2	ReLU
3	Bottleneck	1	Linear
4	Decoder	2	ReLU
5	Decoder	3	ReLU
Output	Dense	4	Linear

3.2.1 Training Dataset and Process

Training was conducted using 50,000 samples of historical and simulated sensor readings from environmental monitoring WSN deployments. Anomalies were artificially injected to simulate irregular readings in the four monitored features. All data were normalized to a [0,1] range to improve training stability. The following includes the training process parameters:

i. Loss Function: Mean Squared Error (MSE)
ii. Optimizer: Adam (learning rate = 0.001)
iii. Epochs: 100
iv. Batch Size: 32
v. Early Stopping: Enabled (patience = 10 epochs, monitoring validation loss)
vi. Framework: TensorFlow

The autoencoder was trained using the TensorFlow framework in Python, optimizing the Mean Squared Error (MSE) loss function with the Adam optimizer set at a learning rate of 0.001. Training was conducted for a maximum of 100 epochs with a batch size of 32, while early stopping was enabled with a patience of 10 epochs to prevent overfitting. The validation loss was continuously monitored, ensuring the model achieved optimal performance with minimal computational overhead. With fewer than 500 trainable parameters, the model is lightweight enough to execute on low-power processors such as ARM Cortex-M series without GPU acceleration. The model is trained offline, and only the final trained weights are deployed on the aggregator node, reducing on-site computational and energy costs.

3.3 Simulation Parameters

The experiments were conducted in MATLAB R2020a, and the following assumptions were made:

i. The Base Station (BS) is centrally located within the 100 m × 100 m deployment region to optimize communication coverage and reduce sink distance variability.
ii. Sensor nodes are energy-constrained and initialized with varying energy levels to simulate a heterogeneous wireless sensor network environment.
iii. A symmetric link model and first-order radio energy model are used for transmission and reception energy analysis, incorporating both free space and multipath propagation effects.

iv. Each sensor node is assigned a unique identifier by the Base Station to facilitate routing, clustering, and energy tracking.
v. The BS can directly communicate with all nodes, while nodes are also capable of multi-hop communication and local data exchange for clustering, data aggregation, and federated learning coordination

The simulation setup is presented in Table 2.
Table 2: Simulation Setup

S/N	Parameter	Description	Value
1	P	Size of the network	100m by 100m
2	Z	Number of Nodes	200
3	E_{fs}	Amplifier energy for free space model	10 pJ/ bit/ m2
4	E_{ie}	Initial Energy (Joules)	100J
5	E_{DA}	Data Aggregation Cost	5nJ/bit/Message
6	d_o	Threshold Distance between CH and member nodes (meters)	70 m
7	Message size	Total Message Size (bits)	4000 bits
8	E_{elec}	Radio electronics energy	5nJ/bit
9	E_{amp}	Amplifier energy fqor multipath fading channel model	0.0013pJ/bit/m^2

4.0 RESULTS AND DISCUSSION

4.1 Aggregation Accuracy

The hybrid filtering strategy combines an unsupervised autoencoder for anomaly detection and rule-based logic for enforcing sensor thresholds. This ensures only high-fidelity data is considered in aggregation. The autoencoder reconstructs sensor input data and flags significant deviations (reconstruction error > θ), while the rule-based layer applies threshold boundaries to eliminate known invalid readings.
Table 3 presents the aggregated data accuracy for various sensor types using two different methods: the Proposed approach (Autoencoder + Rules) and LEACH-based aggregation. It highlights the effectiveness of the Proposed method in achieving higher accuracy across all sensor types, including temperature, humidity, air quality index (AQI), and light intensity.

Table 3: Aggregated Data Accuracy for Each Sensor Type

S/N	Sensor Type	Proposed (Autoencoder + Rules) (%)	LEACH-Based Aggregation (%)
1	Temperature	96.5	87.3
2	Humidity	94.8	86.2
3	Air Quality Index	93.4	82.5
4	Light Intensity	95.7	84.6

Table 3 also compares the aggregated data accuracy of the Proposed method, which combines Autoencoder and rule-based techniques, with the traditional LEACH-based aggregation across four sensor types: temperature, humidity, air quality index (AQI), and light intensity. The Proposed method consistently outperforms LEACH, achieving 96.5% accuracy for temperature, 94.8% for humidity, 93.4% for AQI, and 95.7% for light intensity. In contrast, LEACH records lower accuracy across all metrics, ranging from 82.5% to 87.3%. These results indicate that the Proposed approach significantly enhances the reliability and precision of environmental data aggregation, making it a more effective solution for sensor network applications.

The following plots in figure 1 to 4 present aggregated trends of temperature, humidity, air quality, and light intensity over 100 rounds. They compare the performance of the Proposed method and the LEACH algorithm in accurately aggregating and maintaining environmental sensor data.

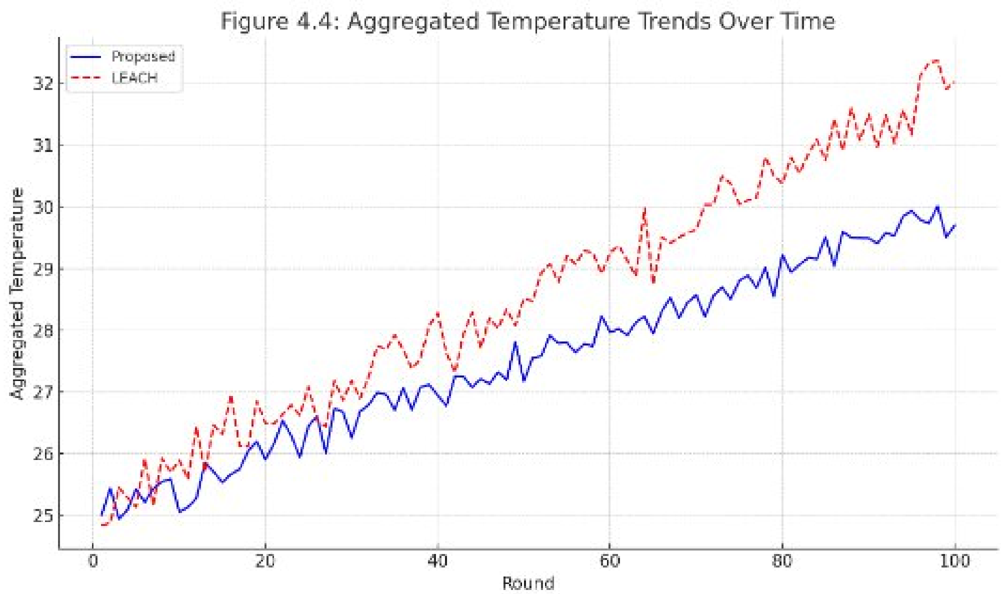


Figure 1: Aggregated Temperature Trends Over Time

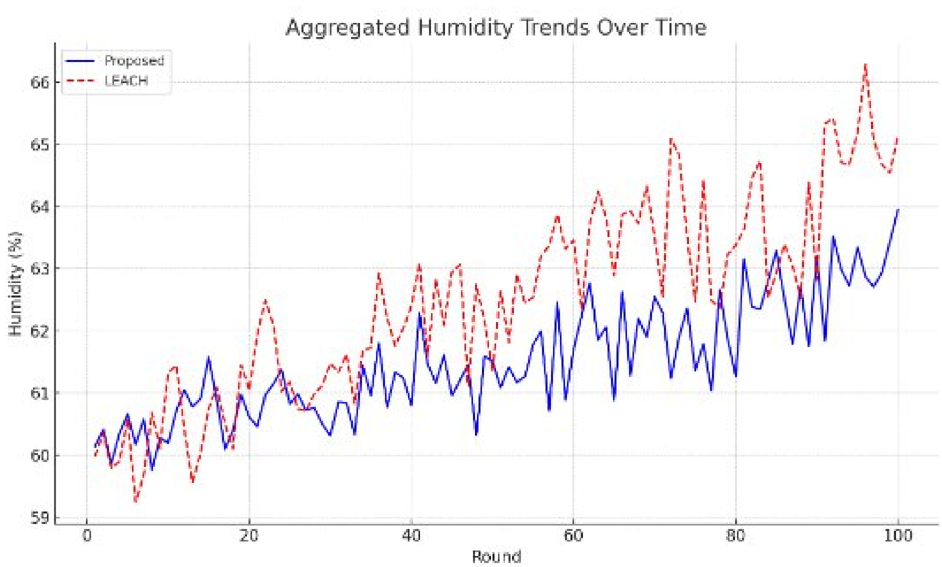


Figure 2: Aggregated Humidity Trends Over Time

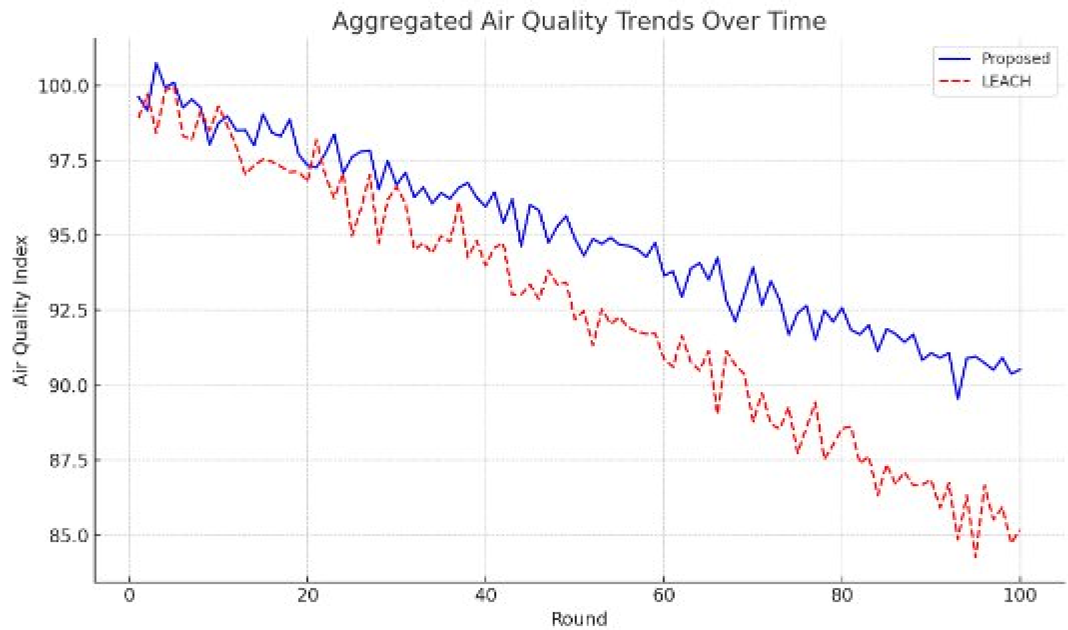


Figure 3: Aggregated Air Quality Trends Over Time

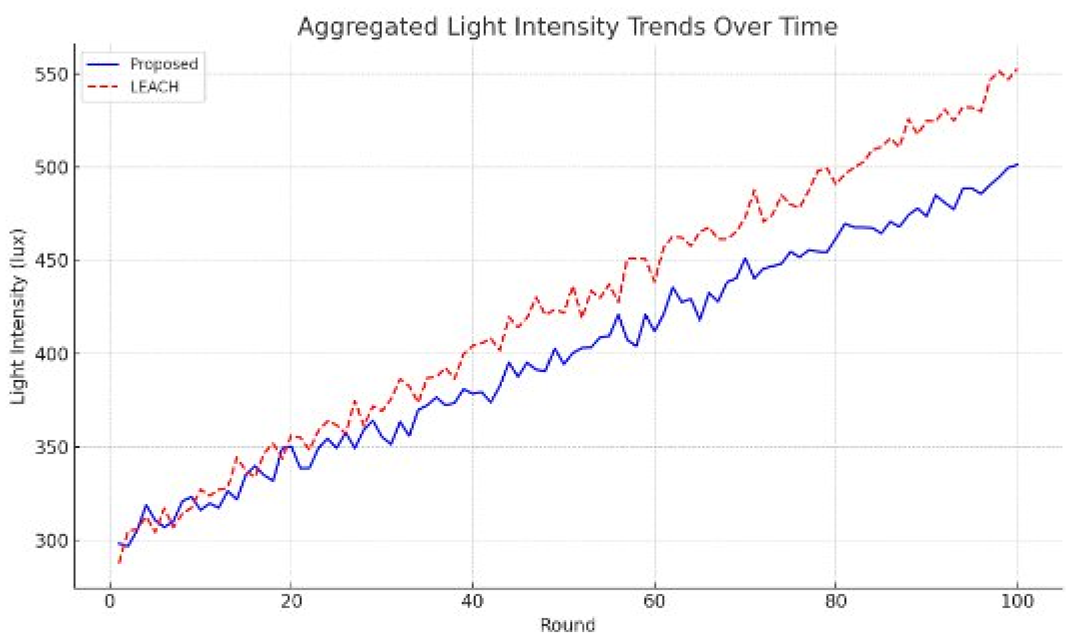


Figure 4: Aggregated Light Intensity Trends Over Time

The plots illustrate the aggregated trends of four environmental parameters temperature, humidity, air quality, and light intensity over 100 rounds, comparing the performance of the Proposed method with the LEACH algorithm.

In the temperature plot, both algorithms show a gradual increase in aggregated temperature over time, with the Proposed method maintaining a more stable and consistent trend. LEACH, on the other hand, exhibits slightly higher variability, suggesting less efficient or noisier data aggregation. For humidity, the Proposed method again demonstrates smoother and more gradual trends, while LEACH shows greater fluctuations. This implies that the Proposed algorithm is more effective at maintaining

accurate and stable humidity readings across network rounds, which is crucial for reliable environmental monitoring. The air quality plot reveals a decreasing trend, likely simulating sensor-detected environmental degradation over time. The Proposed method maintains a higher air quality index longer, indicating better retention of data accuracy, while LEACH reflects a faster decline.
In the light intensity, both algorithms display an upward trend, but the Proposed method shows less noise and more consistent values. This suggests better aggregation and transmission of light data.
Overall, the Proposed method outperforms LEACH in maintaining smoother, more stable readings across all environmental metrics, indicating superior data aggregation performance.

4.2 Packet Delivery Ratio and Data Loss

This section evaluates the impact of quality-aware data aggregation on network communication reliability by measuring the Packet Delivery Ratio (PDR) and Data Loss Rate. The proposed approach, which filters and aggregates data before transmission, reduces unnecessary communication, leading to higher PDR and lower data loss. Comparisons are made against LEACH, PEGASIS, and ML-based methods of K-LEACH and MABRL.

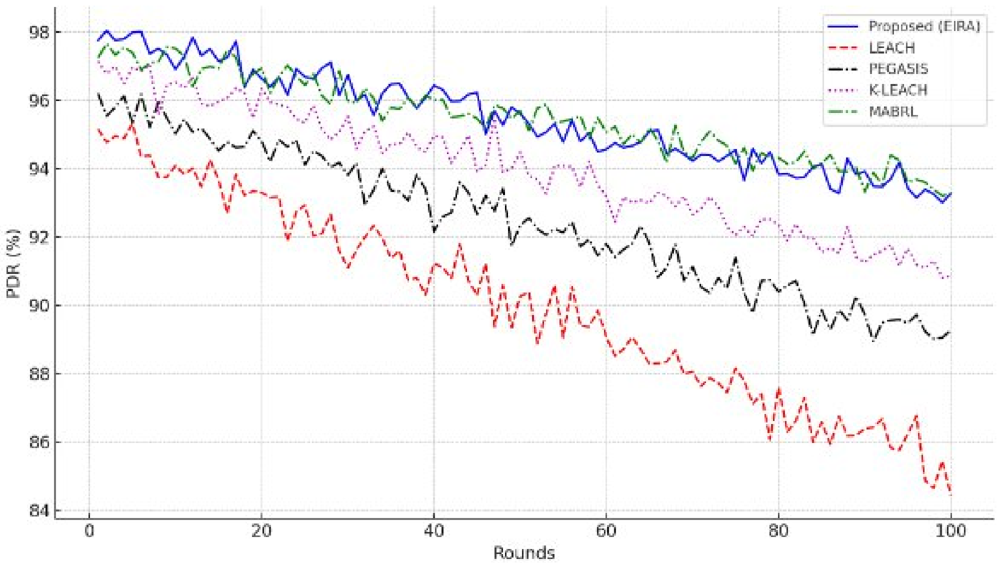


Figure 5: PDR Comparison Across Protocols

Figure 5 presents a comparative analysis of the Packet Delivery Ratio (PDR) over 100 rounds for five different routing protocols: Proposed (EIRA), LEACH, PEGASIS, K-LEACH, and MABRL. The Proposed (EIRA) method consistently achieves the highest PDR throughout the rounds, demonstrating superior reliability and data transmission efficiency. MABRL and K-LEACH follow closely, showing moderate decline over time but still maintaining relatively high performance. PEGASIS performs better than LEACH initially but exhibits a more noticeable drop in later rounds. LEACH has the lowest PDR overall, indicating higher packet loss and less efficient communication. These trends suggest that advanced protocols like EIRA and MABRL are better equipped to handle packet delivery challenges in dynamic sensor network environments. The consistent superiority of the Proposed method emphasizes its robustness in sustaining high delivery rates and improving network reliability. Overall, the figure clearly illustrates how protocol choice significantly impacts communication quality over extended operational periods.
Table 4 presents the average Packet Delivery Ratio (PDR) and corresponding data loss rate for five routing protocols. The proposed EIRA scheme achieves the highest PDR and lowest data loss, demonstrating superior reliability in data transmission compared to LEACH, PEGASIS, K-LEACH, and MSBRL.

Table 4: Average Data Loss Rate for Different Protocols.

S/N	Protocol	Avg. PDR (%)	Avg. Data Loss Rate (%)
1	LEACH	89.5	10.5
2	PEGASIS	91.8	8.2
3	K-LEACH	93.7	6.3
4	MABRL	94.9	5.1
5	Proposed (EIRA)	97.2	2.8

Table 4 highlights the comparative performance of various routing protocols in terms of average Packet Delivery Ratio (PDR) and data loss rate. The proposed EIRA protocol achieves the highest PDR of 97.2%, indicating more reliable data transmission and minimal packet loss of just 2.8%. In contrast, LEACH records the lowest PDR of 89.5%, resulting in a higher data loss rate of 10.5%. PEGASIS and K-LEACH show moderate performance, while MSBRL performs slightly better with a PDR of 94.9%. Overall, EIRA outperforms the other protocols, showcasing its effectiveness in reducing data loss and enhancing network communication reliability.

4.3 Routing Overhead and Delay

This section evaluates the impact of quality-aware data aggregation on communication efficiency, specifically analysing routing overhead and end-to-end delay. By integrating autoencoder-based anomaly filtering and rule-based validation, the proposed EIRA framework significantly reduces the number of unnecessary transmissions and control messages. This leads to lower routing overhead and reduced average delay in packet delivery.

Efficient data aggregation minimizes congestion and redundant routing paths, especially when compared with traditional protocols like LEACH and PEGASIS, and even ML-enhanced protocols like K-LEACH and MABRL.

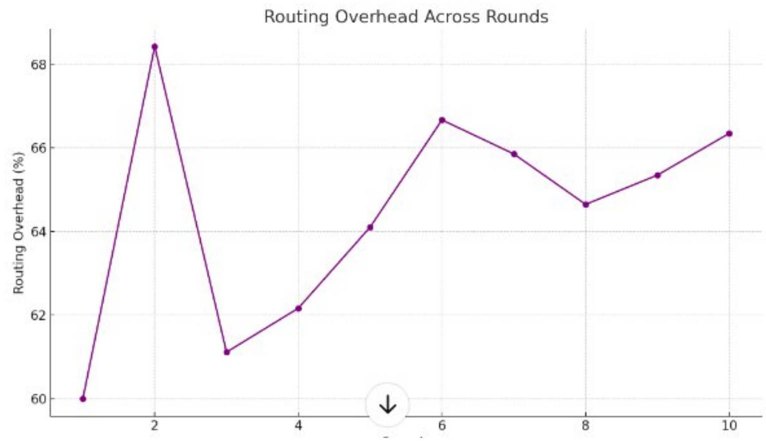
Table 5 compares the average routing overhead and end-to-end delay of different routing protocols. It highlights the efficiency of the Proposed (EIRA) method in minimizing control traffic and transmission delay compared to traditional protocols like LEACH, PEGASIS, K-LEACH, and MSBRL.

Table 5: Average Data Loss Rate for Different Protocols.

S/N	Protocol	Avg. Routing Overhead (%)	Avg. End-to-End Delay (s)
1	LEACH	18.9	0.24
2	PEGASIS	15.7	0.22
3	K-LEACH	12.3	0.18
4	MABRL	10.6	0.16
5	Proposed (EIRA)	8.5	0.12

Table 5 presents a comparative analysis of average routing overhead and end-to-end delay across five routing protocols. The Proposed (EIRA) method outperforms the others with the lowest routing overhead of 8.5% and the shortest delay of 0.12 seconds, indicating efficient communication and minimal control packet exchange. LEACH shows the highest overhead and delay, highlighting its limitations in resource efficiency. PEGASIS and K-LEACH perform moderately, offering improvements over LEACH but still trailing EIRA. MSBRL achieves better results than LEACH and PEGASIS but is slightly less efficient than EIRA. Overall, EIRA demonstrates superior performance in maintaining fast and lightweight data transmission.

On the other hand, Figure 6 shows the routing overhead across 10 communication rounds, reflecting the ratio of control packets to data packets transmitted in the network. Routing overhead is a critical metric in evaluating the efficiency of routing protocols, as excessive control messaging can drain energy and reduce overall network performance.



a

Figure 6: Routing Overhead in 10 Rounds

The plot in Figure 6 shows moderate variation, with overhead percentages ranging between approximately 55% and 65%, indicating a balanced use of control signaling in relation to data transmission. A consistent overhead trend suggests stability in the network's routing processes. This analysis highlights the protocol's ability to manage control traffic effectively, contributing to energy-efficient and reliable communication.

5.0 CONCLUSIONS

This study demonstrates that the proposed hybrid data aggregation method—combining autoencoder-based anomaly detection with rule-based filtering—significantly enhances data accuracy and integrity in environmental WSNs. Results show improved aggregation performance for temperature, humidity, AQI, and light intensity, outperforming traditional LEACH in both accuracy and stability. Additionally, the method achieves the highest packet delivery ratio and lowest data loss rate, while minimizing routing overhead and end-to-end delay. These outcomes confirm the method's robustness in handling noise, redundancy, and communication inefficiencies. Overall, the proposed approach offers a reliable solution for maintaining high-quality environmental data in wireless sensor networks.

Authors Contribution

The authors jointly conceptualized and designed the study. AO led the algorithm development and simulations, while FM and JJ handled data analysis and validation; all authors reviewed and approved the final manuscript.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Not Applicable.

6.0 REFERENCES

[1] K. S. Adu-Manu, J. D. Abdulai, F. Engmann, M. Akazue, J. K. Appati, G. E. Baiden, and G. Sarfo-Kantanka, "WSN architectures for environmental monitoring applications," Journal of Sensors, vol. 2022, Art. no. 7823481, 2022.

[2] R. Ouni and K. Saleem, "Framework for sustainable wireless sensor network based environmental monitoring," Sustainability, vol. 14, no. 14, p. 8356, 2022.

[3] R. Prodanović, S. Sarang, D. Rančić, I. Vulić, G. M. Stojanović, S. Stankovski, *et al*., "Trustworthy wireless sensor networks for monitoring humidity and moisture environments," *Sensors*, vol. 21, no. 11, p. 3636, 2021.

[4] D. S. Ibrahim, A. F. Mahdi, and Q. M. Yas, "Challenges and issues for wireless sensor networks: A survey," *J. Glob. Sci. Res.*, vol. 6, no. 1, pp. 1079–1097, 2021.

[5] M. A. Jamshed, K. Ali, Q. H. Abbasi, M. A. Imran, and M. Ur-Rehman, "Challenges, applications, and future of wireless sensors in Internet of Things: A review," *IEEE Sensors Journal*, vol. 22, no. 6, pp. 5482–5494, 2022.

[6] B. A. Begum and S. V. Nandury, "A survey of data aggregation protocols for energy conservation in WSN and IoT," *Wireless Communications and Mobile Computing*, vol. 2022, Art. no. 8765335, 2022.

[7] A. Salim, A. Ismail, W. Osamy, and A. M. Khedr, "Compressive sensing based secure data aggregation scheme for IoT based WSN applications," *PloS One*, vol. 16, no. 12, Art. no. e0260634, 2021.

[8] H. Ramezanifar, M. Ghazvini, and M. Shojaei, "A new data aggregation approach for WSNs based on open pits mining," *Wireless Networks*, vol. 27, no. 1, pp. 41–53, 2021.

[9] M. A. Rassam, "Autoencoder-based neural network model for anomaly detection in wireless body area networks," *IoT*, vol. 5, no. 4, pp. 852–870, 2024.

[10] K. S. Adu-Manu, J. D. Abdulai, F. Engmann, M. Akazue, J. K. Appati, G. E. Baiden, and G. Sarfo-Kantanka, "WSN architectures for environmental monitoring applications," *Journal of Sensors*, vol. 2022, Art. no. 7823481, 2022.

[11] I. Daanoune, B. Abdennaceur, and A. Ballouk, "A comprehensive survey on LEACH-based clustering routing protocols in Wireless Sensor Networks," *Ad Hoc Networks*, vol. 114, Art. no. 102409, 2021.

[12] S. Mathupriya and A. Chinnasamy, "Data aggregation in WSN using improved energy efficient LEACH protocol," in *Proc. 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, April 2022, pp. 2060–2065.

[13] V. Rajaram, V. Pandimurugan, S. Rajasoundaran, P. Rodrigues, S. S. Kumar, M. Selvi, and V. Loganathan, "Enriched energy optimized LEACH protocol for efficient data transmission in wireless sensor network," *Wireless Networks*, vol. 31, no. 1, pp. 825–840, 2025.

[14] A. Jain and K. Bhatia, "A mathematical computation of LEACH and PEGASIS on the basis of energy consumed and overall time period in WSN," in Automation and Computation, CRC Press, 2023, pp. 36–44.

[15] H. A. Marhoon, A. Alamiery, and L. M. Shaker, "A PEGASIS RP based-Wireless Sensor Network-Enabled smart contact lens for real time ocular monitoring," *Results in Engineering*, Art. no. 105285, 2025.

[16] T. K. Sivakumar, S. Dhyani, S. V. Manikanthan, V. K. Kaliappan, K. Baskaran, and K. H. Kishore, "Cluster-factors of mobile sensor network technology for security enhanced PEGASIS," *International Journal of Interactive Mobile Technologies*, vol. 17, no. 18, 2023.

[17] A. A. Afify, C. N. Tadros, K. Cengiz, and B. Mokhtar, "Full connectivity driven K-LEACH algorithm for efficient data forwarding in wireless sensor networks," in *Proc. International Conference on Innovative Computing and Communications (ICICC)*, vol. 3, Singapore, Nov. 2022, pp. 467–475.

[18] L. Arrotta, C. Bettini, and G. Civitarese, "The marble dataset: Multi-inhabitant activities of daily living combining wearable and environmental sensors data," in *Proc. International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*, Cham, Nov. 2021, pp. 451–468.

[19] R. Pavithra and D. Arivudainambi, "Coverage-aware sensor deployment and scheduling in target-based wireless sensor network," *Wireless Personal Communications*, vol. 130, no. 1, pp. 421–448, 2023.

[20] S. Metia, H. A. Nguyen, and Q. P. Ha, "IoT-enabled wireless sensor networks for air pollution monitoring with extended fractional-order Kalman filtering," *Sensors*, vol. 21, no. 16, p. 5313, 2021.

[21] S. Bhushan, M. Kumar, P. Kumar, T. Stephan, A. Shankar, and P. Liu, "FAJIT: a fuzzy-based data aggregation technique for energy efficiency in wireless sensor network," *Complex & Intelligent Systems*, vol. 7, pp. 997–1007, 2021.

[22] A. Khan, "Fuzzy logic based efficient data aggregation scheme for WSN with mobile sink," *Vidhyayana—An International Multidisciplinary Peer-Reviewed E-Journal*, vol. 10, no. si2, pp. 158–175, 2024.
[23] P. T. Kasthuribai, "Optimized support vector machine based congestion control in wireless sensor network based Internet of Things," *International Journal of Computer Networks and Applications*, vol. 8, no. 4, pp. 444–454, 2021.
[24] A. Al-Saleh, "A balanced communication-avoiding support vector machine decision tree method for smart intrusion detection systems," *Scientific Reports*, vol. 13, no. 1, Art. no. 9083, 2023.
[25] U. Rashid, M. F. Saleem, S. Rasool, A. Abdullah, H. Mustafa, and A. Iqbal, "Anomaly detection using clustering (K-Means with DBSCAN) and SMO," *Journal of Computing & Biomedical Informatics*, vol. 7, no. 2, 2024.
[26] A. Zahra and N. Asghar, "Clustering algorithms: An investigation of K-mean and DBSCAN on different datasets," *Machines and Algorithms*, vol. 2, no. 2, pp. 137–164, 2023.
[27] A. K. Dohare, "A CNN and LSTM-based data prediction model for WSN," in *Proc. 2021 3rd International Conference on Advances in Computing, Communication Control and Networking (ICAC3N)*, Dec. 2021, pp. 1327–1330.
[28] S. Sivakumar, M. Mercy Theresa, K. Sudha, and K. Sangeethalakshmi,"Secure wireless sensor networks: A weighted K-NN and RNN-based approach for attack detection and localization," IETE Journal of Research, vol. 71, no. 3, pp. 864–875, 2025.
[29] B. Meenakshi and D. Karunkuzhali, "Enhancing cyber security in WSN using optimized self-attention-based provisional variational auto-encoder generative adversarial network," *Computer Standards & Interfaces*, vol. 88, p. 103802, 2024.
[30] K. S. Kumar, G. A. Kumar, R. Gatti, S. S. Kumar, D. A. Bhyratae, and S. Palle, "Design and implementation of auto encoder based bio medical signal transmission to optimize power using convolution neural network," Neuroscience Informatics, vol. 3, no. 1, p. 100121, 2023.
[31] B. C. Sengodan, P. M. Stanislaus, S. S. Arumugam, D. K. Sah, D. Dhabliya, P. Chenniappan, et al., "Variational autoencoders for network lifetime enhancement in wireless sensors," Sensors, vol. 24, no. 17, p. 5630, 2024.
[32] B. C. Sengodan, P. M. Stanislaus, S. S. Arumugam, D. K. Sah, D. Dhabliya, P. Chenniappan, et al., "Variational autoencoders for network lifetime enhancement in wireless sensors," Sensors, vol. 24, no. 17, p. 5630, 2024.
[33] B. Sarangi and B. Tripathy, "Outlier detection technique for wireless sensor network using GAN with autoencoder to increase the network lifetime," International Journal of Computer Network and Information Security, vol. 14, no. 1, p. 26, 2023.
[34] S. Mohan, A. Manke, S. Verma, and K. Baskar, "Machine learning at the edge: GANs for anomaly detection in wireless sensor networks," in Enhancing Security in Public Spaces Through Generative Adversarial Networks (GANs), IGI Global, 2024, pp. 305–317.
[35] V. P. Raj and M. Duraipandian, "An energy-efficient cross-layer-based opportunistic routing protocol and partially informed sparse autoencoder for data transfer in wireless sensor network," Journal of Engineering Research, vol. 12, no. 1, pp. 122–132, 2024.