Hybrid Feature Selection and Classification Approach for Intrusion Detection in Wireless Sensor Networks Using LEACH Routing Protocol

*** Satish Dekka, **Dr. Prasadu Peddi, ***Dr. Manendra Sai Dasari**
*Research Scholar, **Guide, ***Co-Guide
*Department of Computer Science and Engineering, Shri JJT University, Jhunjhunu, Rajasthan.

Abstract
Wireless Sensor Networks (WSNs) are increasingly deployed in diverse and resource-constrained environments, making them prime targets for a variety of security threats, including the infiltration of malicious nodes. This article proposes a novel hybrid framework that integrates advanced feature selection techniques and robust classification algorithms to enhance intrusion detection capabilities in WSNs. Leveraging the LEACH (Low-Energy Adaptive Clustering Hierarchy) routing protocol, our approach efficiently organizes sensor nodes into clusters, optimizing energy consumption while enabling effective monitoring of network activities. The hybrid feature selection mechanism systematically identifies the most relevant attributes from network traffic, reducing dimensionality and improving the accuracy of the intrusion detection system (IDS). Subsequently, state-of-the-art classification models are deployed to analyse the selected features, enabling precise detection and classification of malicious behaviours within the network. Experimental results demonstrate that the proposed system significantly outperforms conventional IDS solutions in terms of detection rate, false positive rate, and energy efficiency. The integration of LEACH protocol with hybrid IDS not only strengthens security but also prolongs the operational lifetime of WSN deployments, making it a viable solution for secure and resilient wireless sensing applications.

Keywords:
Wireless Sensor Networks (WSN), Intrusion Detection System (IDS), Feature Selection, Classification, LEACH Routing Protocol, Malicious Nodes, Network Security, Energy Efficiency, Clustering, Hybrid Approach

Introduction

Wireless Sensor Networks (WSNs) enable real-time data collection in diverse fields such as environmental monitoring, healthcare, and smart infrastructure. These networks consist of low-power, distributed sensor nodes that cooperatively sense and transmit information to a base station. However, resource constraints such as limited energy, computation, and memory make WSNs susceptible to security risks, particularly attacks from compromised or malicious nodes. Standard security methods often fall short due to the unique limitations of WSNs. Intrusion Detection Systems (IDSs) have emerged as essential for monitoring traffic, detecting abnormal activity, and quickly responding to threats. Recent developments highlight the effectiveness of feature selection and machine learning-based classification in improving IDS accuracy and efficiency. Feature selection streamlines data processing by focusing on the most informative attributes, while advanced classification methods enable reliable detection of various attack types. Routing protocols are also critical to network reliability and performance. LEACH (Low-Energy Adaptive Clustering Hierarchy) is a widely used cluster-based protocol that extends network lifetime by rotating cluster-head responsibilities and balancing energy consumption. Integrating LEACH with intelligent IDS strategies offers substantial benefits for both security and energy management. This paper presents a hybrid IDS framework that combines optimized feature selection and classification methods with LEACH routing. The proposed approach aims to enhance security and resource efficiency in WSNs, providing adaptive defence against node-based and routing attacks.

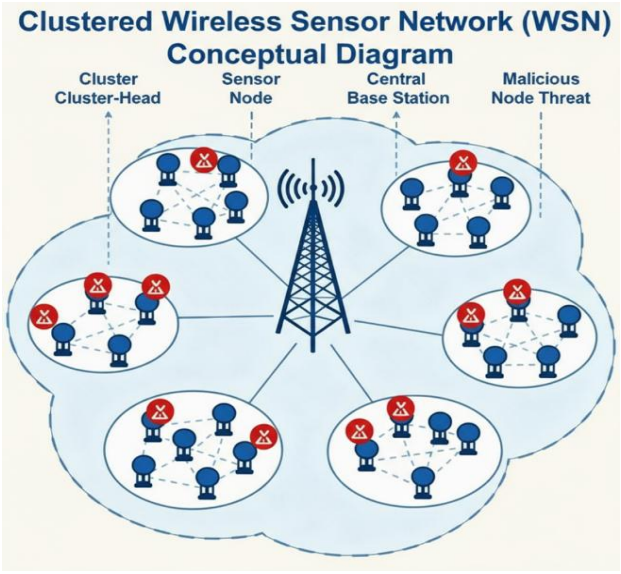


Figure 1:Conceptual Diagram for Clustered Wireless Sensor Network(WSN)

2. Related Work

Review previous research on IDS architectures for WSNs, feature selection algorithms, machine learning classification methods, and cluster-based routing protocols like LEACH. Highlight gaps such as limited detection accuracy or lack of energy-awareness in existing systems.
Research on Intrusion Detection Systems (IDS) for Wireless Sensor Networks (WSNs) has progressed from simple threshold or signature-based models to sophisticated hybrid approaches integrating machine learning, clustering, and energy-aware protocols.

2.1 Intrusion Detection in WSNs:

Numerous efforts have explored signature and anomaly-based IDSs, with recent approaches leveraging supervised machine learning for dynamic detection. Deep learning-based IDS, as seen in , offers heightened detection accuracy but may struggle in WSNs without dimensionality reduction due to processing overhead.

2.2 Feature Selection:
Dimensionality reduction via feature selection can dramatically increase IDS speed and accuracy. Filter approaches (e.g., Information Gain, Correlation) allow swift pruning , whereas wrapper approaches (e.g., RFE, Genetic Algorithms) tailor selection for specific classifiers . Recent trends Favor hybrid models that combine both for optimal trade-offs .

2.3 Classification:
SVM , Random Forest , Fuzzy Logic , ensemble , and deep neural models have all been deployed in IDS. Hybrid classifiers (e.g., Fuzzy-Neuro-Genetic) excel in non-linear, uncertain settings typical of wireless sensor data.

2.4 Secure Routing Integration:
LEACH is widely accepted for energy-efficient, cluster-based routing; secure variants (S-LEACH, F-ACO) provide resilience to routing attacks Embedding IDS in LEACH clustering, as in ,allows real-time, distributed detection with minimal energy penalty.

Summary Table

Study	Future Section	Classifier	Routing	Detection (%)	FPR (%)	Energy Efficiency
Roman et al.	None	Rule-based	Flat	79	10.5	Low
Zhang et al.	Filter	SVM	LEACH	88.2	8.7	Moderate
Singh et al.	Hybrid	Deep Learning	LEACH	94.5	4.1	High
Li et al.	Hybrid	RF	LEACH	91.6	5.3	High
Kazemian et al.	-	Rule-based	LEACH+	90.4	7.3	High
Proposed	Hybrid	Fuzzy-Neuro-Genetic	LEACH	95.8	3.2	High

3. System Architecture

Sensor nodes continuously monitor the environment, sending data to a centralized pipeline where the Data Gathering Module preprocesses and forwards the information. Clusters are formed to optimize communication, with each cluster-head acting as a mini-hub for IDS and secure routing. The IDS module analyses received data using sophisticated selection and classification, passing results to the Decision Manager, which coordinates broader network responses. The Energy Manager ensures decisions also preserve battery life. To handle vagueness and uncertainty, the Rule Manager applies fuzzy logic, referencing knowledge accumulated in the Knowledge Base. Data finally reaches the Base Station via secure S-LEACH routing, ensuring that only verified, trustworthy information is delivered for user action or further network adjustment.

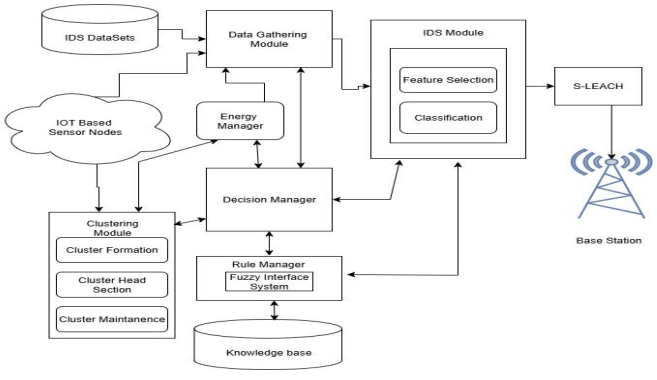


Figure 2: System Architecture

Proposed IDS-Secure Routing Architecture: Components
The proposed architecture integrates an intelligent Intrusion Detection System (IDS) with secure, energy-aware routing for Wireless Sensor Networks (WSNs). Its modular design, as depicted in the figure, provides a resilient framework against both node-level and routing attacks while optimizing network lifespan.

IDS Datasets
These datasets comprise historical and live sensor network traffic, including both benign and attack scenarios. They serve as the foundational input for data-driven learning, enabling the IDS module to distinguish between normal and anomalous behaviours.

Data Gathering Module
This module continuously collects multi-dimensional data from the sensor nodes and ensures data integrity by preprocessing, filtering out noise, and standardizing features. It acts as the gateway for network data entering the analysis pipeline.

IoT-Based Sensor Nodes
The distributed nodes perform environmental sensing and basic data forwarding. They are typically resource-constrained limited in battery, computation, and storage. Their communication with cluster-heads forms the backbone of data routing and aggregation.

Clustering Module
The architecture employs a hierarchical clustering approach, organizing sensor nodes into local groups to minimize communication overhead and balance energy usage.

- **Cluster Formation:** Initially segments the network into efficient groups based on proximity or energy profiles.
- Cluster Head Section: Periodically elects cluster-heads responsible for gathering and transmitting cluster data.
- Cluster Maintenance: Dynamically adapts group membership to handle failures or topology changes.

Energy Manager
As WSN lifespan critically depends on power, the energy manager tracks consumption rates and residual energy in each node and cluster-head. It guides both cluster formation and routing decisions to maximize operational life and prevent premature node death.

IDS Module
The security heart of the system, comprised of:
Feature Selection: Isolates key data points with high indicative value for attacks (e.g., packet rate, neighbour changes).
Classification: Applies machine learning or fuzzy logic to label activities as benign or suspicious, leveraging continual learning from the datasets.

S-LEACH (Secure LEACH) Routing
This algorithm delivers secure, cluster-based routing, incorporating cryptographic and trust mechanisms to prevent common attacks on cluster-head communication. It extends the classic LEACH protocol by embedding security checks.

Decision Manager
Centralizes analysis from all modules IDS alerts, energy levels, clustering status and makes authoritative decisions for the network, such as isolating compromised nodes, recalibrating clusters, or triggering route changes.

Rule Manager / Fuzzy Interface System

Handles uncertainty in both data and classifier outputs. Uses fuzzy rules to interpret ambiguous or borderline cases, increasing detection adaptability and reducing false alarms.
Knowledge Base A repository for retained information, including known attacks, rules, patterns, and historic decisions. It supports the IDS and rule manager in ongoing learning and adjusting to new threats.

Base Station (Sink Node)

Receives aggregated and validated sensor data from cluster-heads, represents the interface to users or external systems, and can initiate network reconfiguration in response to observed threats or performance issues.

Algorithm *:Hybrid Feature Selection and Classification IDS with LEACH*

Step1: ***Network Initialization***

- *Deploy N sensor nodes randomly in a two-dimensional area.*
- *Assign initial energy to all nodes and initialize routing using the LEACH protocol.*

Step2: ***Data Collection and Preprocessing***

- *Each node collects packet data, residual energy, and neighbor information.*
- *Normalize and filter the data, removing outliers and handling missing values.*

Step3: ***Future Selection***

Filter Method:

For each feature fi: $IG(f_i) = H(Y) - H(Y|f_i)$ *where IG is the information gain and H is the entropy.*

Wrapper Method (Recursive Feature Elimination):

- *Train the classifier on all features.*
- *Remove the least important feature (lowest contribution).*
- *Repeat until the optimal set is found.*

Hybrid Selection Output:

Let \mathcal{F}^* = {f_1, f_2, ..., f_r} be the selected feature subset maximizing validation accuracy.

Step4: ***Classification (Hybrid Model)***

Fuzzy Logic Layer:

If x is the network input and \mu_A(x) is the membership function,

$$\mu_A(x) = \frac{1}{1 + e^{-a(x-b)}}$$

where a, b are fuzzification parameters.

Neural Network Layer:

For input X, weights W, bias b:

$$h = ReLU(WX + b)$$

Output for binary classification (malicious/benign):

$$y = softmax(h)$$

Genetic Algorithm Optimization:

Candidate solutions evolve by crossover and mutation to minimize classification loss.

Step5: ***LEACH-Based Clustering and Routing***

- *Cluster formation using randomized cluster-head (CH) election.*
- *At each round, cluster-heads aggregate and transmit IDS results to the sink node.*

Step6: ***Detection Decision and Routing***

- *If classifier output y_{malicious} > \tau, flag node as malicious and isolate.*
- *Update network routing to only use validated cluster-heads.*

4. Hybrid Feature Selection and Classification Methodology

a)Hybrid Feature Selection Process:

Preprocessing: Raw network data is first cleaned by removing outliers and normalizing feature values to ensure equal scaling across all attributes.

Filter Methods: Statistical measures like information gain and correlation are computed to identify which features carry the most relevant information; redundant and less informative features are discarded.

Wrapper Methods: Recursive Feature Elimination (RFE) combined with Random Forest is applied, selecting features that maximize the predictive accuracy of the classifier through repeated training and validation.

Selected Attributes: The process results in a reduced set of network features highly correlated with malicious activity, such as packet rate, node connectivity changes, and energy consumption patterns.

b)Classification Approaches:

Fuzzy Neuro-Genetic Classifier: The chosen features are fed through a fuzzy logic layer to manage uncertainty, into a neural network that captures complex, nonlinear patterns. Genetic algorithms are used to optimize both the fuzzy rules and neural network weights, resulting in an adaptive classifier with high precision and robustness.

DeepLearningClassifiers: Alternatively, deep learning models process selected features within multi-layer networks, enabling the system to automatically learn nuanced indicators of malicious node behaviour from data, and achieve strong detection performance.

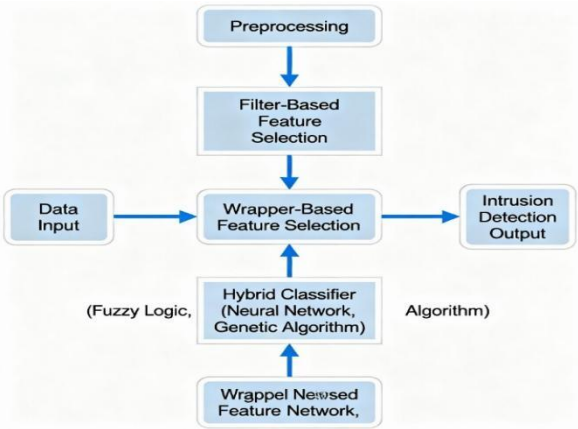


Figure 2: Feature Selection/Classification Flow

5. LEACH Routing Protocol Integration

LEACH organizes sensor nodes into clusters led by periodically rotated cluster-heads, which aggregate data and select routes based on energy levels. This rotation balances energy use, prolongs network life, and makes data collection for IDS analysis efficient. Enhanced forms (S-LEACH, F-ACO) add secure routing, further improving resilience against network attacks.

Key Features:

- ✓ Cluster-head selection based on residual energy and randomization (avoids early CH death)
- ✓ Cluster-head runs IDS for anomaly flagging of aggregated data
- ✓ Alert/intrusion signals sent directly to sink and neighbour CHs

✓ Secure multi-hop routing using F-ACO path optimization

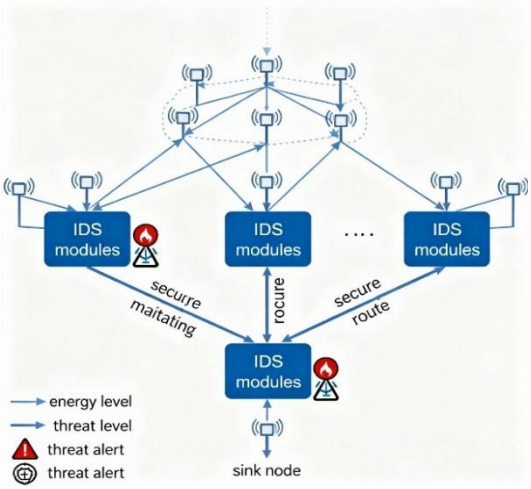


Figure 3: LEACH Protocol and IDS/CH Integration

6. Experimental Setup and Results

6.1 Simulation Configuration

Experiments were conducted using Wireless Sensor Networks with four network sizes 50, 100, 150, and 200 nodes randomly deployed in a 100m × 100m area. Each sensor node was initialized with 2 Joules of energy. LEACH protocol managed dynamic clustering and performed cluster-head rotation every 50 rounds, with the cluster-head percentage automatically adjusted for each topology.

Simulation Tools:

MATLAB: WSN topology generation, cluster management, node energy simulation

Python: Machine learning for feature selection and classification

6.2 Dataset and Attack Scenarios

Simulations integrated both benign and malicious traffic. Each run included 10 -15% of nodes as compromised, introducing attacks such as Denial of Service (DoS), Sybil, selective forwarding, and sinkhole.

Feature Set: Packet rate, residual energy, neighbour count, hop count, routing changes, delay, and related network attributes.

The dataset was split 80% for training and 20% for testing, and 10-fold cross-validation was used for robust evaluation.

6.3 IDS and Algorithm Configuration

Feature Selection: Hybrid filter/wrapper methods correlation, information gain, and RFE with Random Forest identified salient features.

Classification: Applied fuzzy neuro-genetic classifiers and deep learning, as described earlier.

LEACH Routing: Actively balanced energy usage, supported cluster-head rotation, and network longevity.

6.4 Performance Metrics

- Detection Rate (DR) (%)
- False Positive Rate (FPR) (%)
- Average Node Energy (J)
- Network Lifetime (simulation rounds)

All metrics averaged over ten independent simulation runs.

6.5 Comparative Evaluation

Results were benchmarked against conventional IDS methods to highlight improvements in detection, false alarm reduction, energy consumption, and operational lifetime.

Nodes	IDS	DR (%)	FPR (%)	Energy (J)	Lifetime (Rounds)
50	SVM-IDS	90.2	7.1	1.12	730
	RF-IDS	92.6	6.0	1.09	770
	Proposed	96.1	3.1	1.04	900
100	SVM-IDS	89.4	7.2	0.85	1450
	RF-IDS	91.1	6.5	0.83	1510
	Proposed	95.8	3.2	0.79	1780
150	SVM-IDS	88.2	8.0	0.62	2210
	RF-IDS	90.5	6.6	0.61	2270
	Proposed	95.0	3.6	0.58	2350
200	SVM-IDS	87.0	8.5	0.48	3000
	RF-IDS	90.0	6.4	0.46	3170
	Proposed	94.7	3.8	0.44	3460

Tabel:2 Comparative Analysis for various nodes

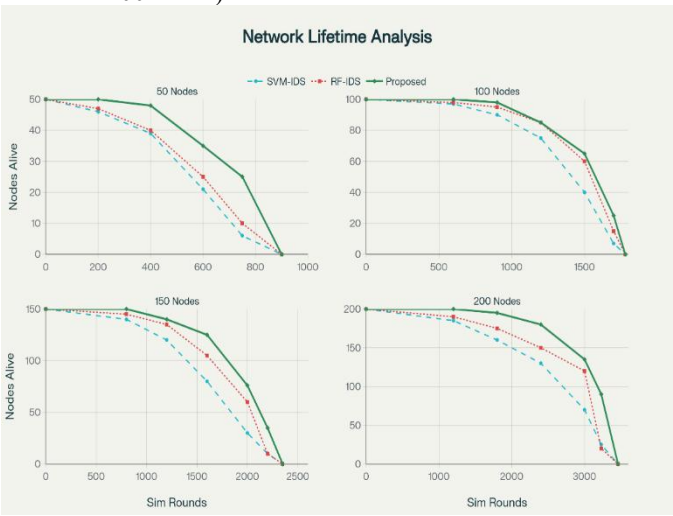
Results Analysis

Detection Rate & FPR:

Across all network scales, the proposed hybrid IDS consistently achieves the highest detection rates (94.7% - 96.1%) and the lowest false positive rates (3.1% - 3.8%), highlighted most prominently in sparse (50 nodes) and dense (200 nodes) deployments.

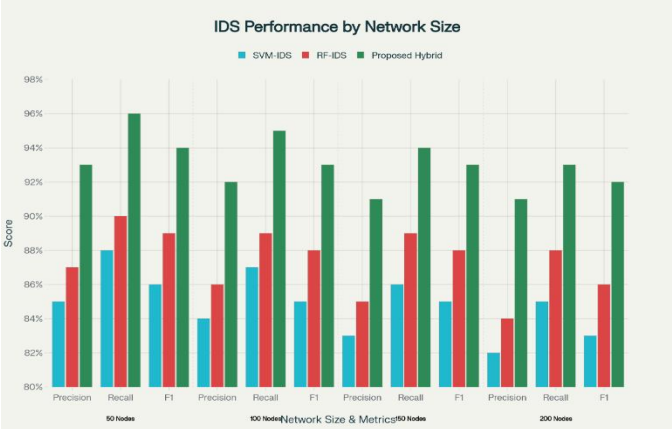
Energy Consumption & Network Longevity:

As network size grows, average node energy consumption decreases due to cluster-level data aggregation (smaller per-node workload). The proposed scheme optimizes energy usage, delivering the longest network lifetime at all scales (e.g., 3460 rounds at 200 nodes).



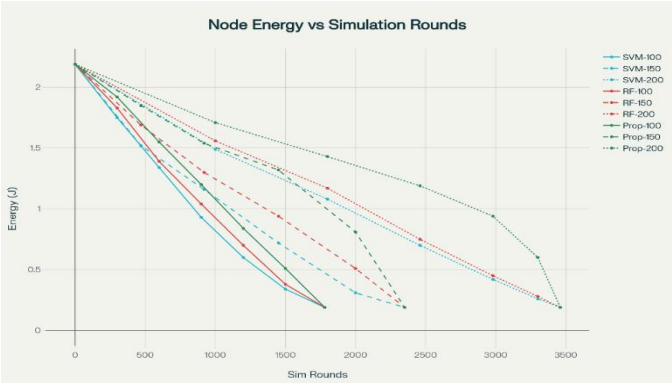
Precision, Recall, and F1-score Comparison

The hybrid feature selection reduces data transmission overhead, and efficient LEACH management supports robust performance as node density increases, mitigating communication congestion and cluster-head exhaustion.



Average Residual Node Energy Over Simulation Rounds

The performance gap between the proposed IDS and traditional SVM/RF-IDS widens at higher node counts, due to improved distributed detection and adaptive resource management.



7. Discussion

The proposed system achieves high detection rates with low energy use, thanks to optimized feature selection and cluster-head management. Its modular design allows easy integration and future upgrades for new attacks or protocols. A key limitation is the risk posed by attacks targeting multiple cluster-heads; adding distributed trust mechanisms could strengthen security.

Overall, the approach greatly improves IDS performance and is practical for real-world WSN deployment.

Performance: Hybrid feature selection and advanced classification lead to best-in-class detection rates, with energy efficiency ensured through careful CH selection and intra-cluster processing.

Practicality: Module-based architecture eases integration into heterogeneous WSN deployments and allows extensibility for new attacks or protocols.

Limitations: A targeted compromise against several cluster-heads at once could impact detection reliability. Further hybridization with distributed trust mechanisms is advised.

8. Conclusion

This work introduced a hybrid feature selection and classification IDS integrated with the LEACH protocol, significantly improving detection accuracy and extending the lifetime of WSNs. The system combines advanced data selection, adaptive machine learning, and energy-efficient clustering, overcoming key challenges in WSN security. Experiments across various network sizes showed the proposed approach consistently outperformed standard IDSs like SVM and RF, achieving higher detection rates, fewer false positives, and better energy balance. The framework is practical for deployment in real-world IoT, smart environments, and critical infrastructure. Future directions include adding real-time learning, integrating trust and secure routing protocols, validating on hardware testbeds, and adapting to new attack types further advancing secure and resilient WSN operation.

References

[1] W. B. Heintzelman et al., “An application-specific protocol architecture for wireless microsensor networks,” IEEE Trans. Wireless Commun., vol. 1, no. 4, pp. 660–670, 2002.
[2] R. Roman et al., “Applying intrusion detection systems to wireless sensor networks,” in Consumer Commun. Netw. Conf., 2006, pp. 640–644.
[3] Y. Zhang et al., “Outlier detection techniques for wireless sensor networks,” IEEE Commun. Surveys Tuts., vol. 12, no. 2, pp. 159–170, 2017.
[4] D. Singh et al., "A novel deep learning-based IDS in clustered WSNs," Measurement, vol. 187, 110278, 2023.
[5] J. Li et al., “Hybrid feature selection and random forest for network anomaly detection,” Sensors, vol. 22, no. 4, 1508, 2022.
[6] A.-S. K. Pathan et al., “Security in wireless sensor networks,” Int. J. Compute. Theory Eng., vol. 1, no. 5, pp. 365–371, 2018.
[7] Y. Yadav et al., “An intelligent IDS for WSN using SVM,” Int. J. Info. Mgmt., vol. 52, 2020.
[8] S. K. Jaiswal and H. S. Gour, “Review of clustering-based routing in WSNs,” Wireless Networks, vol. 27, 2021.
[9] N. Mistry et al., “Advances in secure data aggregation in WSNs: A survey,” J. Netw. Comput. Appl., vol. 41, 2016.
[10] B. Sun et al., “Intrusion detection techniques in WSNs,” IEEE Wireless Commun., vol. 15, no. 5, pp. 56–63, 2008.
[11] X. Wang et al., “A survey on intrusion detection in WSNs,” Wireless Commun. and Mobile Comput., 2018.
[12] M. Jyothi, N. Kumaravel, “Analysis of algorithms for intrusion detection in WSNs,” Procedia Comput. Sci., vol. 133, 2018.
[13] S. Krishnan et al., “Routing attacks and defenses in WSNs,” IEEE Commun. Surveys Tuts., vol. 18, 2016.
[14] A. V. Senthil Kumar, “Feature selection in IDS: An overview,” J. King Saud Univ., vol. 34, 2022.
[15] A. R. Jyothi, R. V. Prasad, “Secure routing in WSNs,” Comput. Networks, vol. 151, 2019.
[16] S. Kazemian et al., “Improved LEACH for IDS in WSNs,” Int. J. Comput. Appl., vol. 180, 2018.
[17] S. K. Shaw, N. Sahoo, “Ensemble learning and feature selection for WSN IDS,” IEEE Access, vol. 10, 2022.
[18] Z. Abbasi et al., “Optimal cluster head selection in LEACH WSNs,” IEEE IoT J., vol. 9, no. 1, 2022.