IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 6, December 2025, pp. 4913∼4922
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i6.pp4913-4922

Comparative evaluation of machine learning models for intrusion detection in WSNs using the IDSAI dataset

Mansour Lmkaiti^1, Houda Moudni^2, Hicham Mouncif^1
^{1}LIMATI Laboratory, Faculty of Polydisciplinary, University Sultan Moulay Slimane, Beni-Mellal, Morocco
^{2}TIAD Laboratory, Faculty of Sciences and Technology, University Sultan Moulay Slimane, Beni-Mellal, Morocco

Article Info

Article history:

Received Aug 27, 2024
Revised Oct 24, 2025
Accepted Nov 8, 2025

Keywords:

Gradient boosting
IDSAI
Intrusion detection system
Logistic regression
Machine learning
Random forest
Wireless sensor networks

ABSTRACT

This paper provides comparative assessment of three lightweight machine learning (ML) models (logistic regression (LR), random forest (RF), and gradient boosting (GB)), which are employed to detect intrusions in wireless sensor networks (WSNs) using the IDSAI dataset. The goal is to determine the most effective and deployable classifier within the constraints of WSN resources. In order to prevent data leakage and report accuracy, precision, recall, F1-score, and receiver operating characteristic-area under the curve (ROC-AUC) with mean±SD, we implement stratified 5-fold cross validation with in fold preprocessing. The results indicate that RF provides the most optimal generalization and overall performance (accuracy 0.9994 ± 0.0001, precision 0.9995 ± 0.0001, recall 0.9994 ± 0.0001, F1-score 0.9994 ± 0.0001, ROC–AUC 0.9998 ± 0.0000). RF is closely followed by GB (accuracy 0.9990 ± 0.0001, precision 0.9995 ± 0.0001, recall 0.9985 ± 0.0001, F1-score 0.9990 ± 0.0001, ROC-AUC ≈ 1.0000). LR demonstrates limitations in linearly overlapping classes, as evidenced by its high precision but reduced recall (accuracy 0.9167 ± 0.0010, precision 0.9829 ± 0.0002, recall 0.8481 ± 0.0018, F1-score 0.9105 ± 0.0011, ROC–AUC 0.9707 ± 0.0001). In order to evaluate deployability, we characterize the inference throughput on a modest PC: LR $\sim 6.5 \times 10^5$ samples/s, GB $\sim 2.2 \times 10^5$ samples/s, and RF $\sim 1.3 \times 10^5$ samples/s, indicating a tiered intrusion detection system (IDS) (LR at sensors, RF at cluster-heads, and GB at the gateway). We also address the potential dangers of overfitting that may arise from the cleanliness of the dataset and provide a roadmap for future validation on a more diverse set of traffic. The research establishes a baseline for lightweight IDS in actual WSNs that is deployable and reproducible.

Corresponding Author:

Mansour Lmkaiti
LIMATI Laboratory, Faculty of Polydisciplinary, University Sultan Moulay Slimane
Beni-Mellal, Morocco
Email: lamkaitimansour@gmail.com

1. INTRODUCTION

This paper introduces a methodical approach that uses cutting-edge machine learning (ML) [1] algorithms to thoroughly assess the effectiveness of intrusion detection systems (IDS) [2]. Ensuring strong network security is crucial in the quickly changing cybersecurity landscape of today, which is marked by an increase in cyberthreats and the widespread integration of internet of things (IoT) devices [3]. IDS [4] are essential for protecting networks because they keep an eye on traffic patterns and spot possible harmful activity.

However, depending on the detection methods utilized and the caliber of the training and evaluation datasets, IDS efficacy might vary greatly [4].

Our study suggests an organized strategy that includes several crucial steps to fully address these issues: careful dataset preparation [4], stringent feature selection and engineering procedures, extensive model training and evaluation techniques, reliable cross-validation procedures, and in-depth performance analysis. Our research intends to improve the accuracy and dependability of IDS implementations by utilizing the variety of real-world intrusion scenarios captured in the IDSAI dataset [4]. IDS models are developed and evaluated using ML [1] algorithms, including gradient boosting (GB), random forest (RF) [5], and logistic regression (LR) [6], in order to improve their ability to effectively detect and mitigate security breaches [7].

By using this systematic and empirical research, our study aims to offer detailed insight into the advantages and disadvantages of ML based IDS [8], [9] techniques. We help develop more flexible and effective security measures suited to the complex dynamics of modern networks and the changing terrain of cyberthreats by critically assessing the performance of different algorithms against benchmark datasets and a range of attack scenarios [10]. Previously, classical classifiers were frequently evaluated on synthetic or restricted IoT datasets without taking into account genuine wireless sensor network (WSN) constraints. This work addresses that lacuna by introducing a robust statistical validation (mean ± SD, 95% confidence interval), assessing computational feasibility at the sensor, cluster-head, and gateway levels, and benchmarking three interpretable, lightweight models on the IDSAI dataset.

The key contributions of this study are as follows. First, an evaluation approach that can be replicated for lightweight IDS benchmarking using the IDSAI dataset is proposed. Second, LR, RF, and GB are integrated in a hierarchical IDS architecture for scalable WSN security. Third, statistical tests, including Friedman and Wilcoxon, are used to validate the robustness of the model. Fourth, computational footprint and inference throughput are included in the deployment analysis. Finally, a plan for upcoming validation with a variety of scenarios is outlined.

2. RELATED WORK

Much research has been done in the field of computer security [8], especially in WSNs, to address the changing challenges posed by security threats [9]. Numerous strategies for improving WSN security, such as intrusion detection, encryption methods and secure routing protocols, have been examined in earlier research [10], [11]. The creation and assessment of IDS [8] designed especially for WSNs constitutes a substantial field of study [12]. By keeping an eye on network traffic and spotting unusual activity suggestive of malicious activity, these systems are essential in detecting and preventing security breaches [7] within WSNs.

Numerous studies have used a variety of datasets and evaluation metrics to assess the effectiveness of IDS in WSNs [13], [14]. These tests seek to determine how well IDS identify different kinds of assaults, such as routing attacks, data manipulation and denial-of-service attacks [14], [15]. Researchers have shed important light on the advantages and disadvantages of current IDS in WSNs by comparing various to common datasets and attack scenarios.

Additionally, research has focused on creating lightweight security measures that are suited for WSN devices with limited resources [16]. These safeguards are designed to reduce energy usage and computational overhead while offering strong defense against security risks [16]. To address the particular security issues presented by WSNs, methods like energy-efficient key management techniques, secure routing protocols, and lightweight cryptography have been developed [12]. Additionally, research has looked into how to incorporate cutting-edge technologies like machine intelligence and blockchain into WSN security designs [16]–[20]. While ML [18], [19], [21], [22] methods provide for adaptive and autonomous intrusion detection capabilities, blockchain [17]-based techniques provide decentralized and tamper-resistant mechanisms for safeguarding WSN data and transactions. Recent research has investigated the use of anomaly-based methods (e.g., one-class support vector machine (SVM), isolation forest) and long short-term memory (LSTM), as well as autoencoders, for the purpose of intrusion detection in IoT/WSN. Although these methods are frequently precise, they typically necessitate substantial refining and higher compute and energy budgets. We concentrate on the development of interpretable and lightweight models that are appropriate for decentralized WSN nodes. In the future, we will investigate the integration of on-node lightweight classifiers with gateway-level deep feature extraction. Table 1 summarizes recent IDS studies in WSN and IoT environments.

Table 1. Comparative summary of recent IDS studies in WSNs and IoT environments

Study	Dataset	Model	Accuracy (%)	Main limitation
Dharini *et al.* [9]	WSN-LEACH	XGBoost	98.7	High computational cost
Meenakshi and Karunkuzhali [10]	IoT-Custom	GAN-VAE	99.2	Complex training
Ajmi *et al.* [16]	Hardware IDS	CNN	96.8	Not energy-efficient
This study	IDSAI	RF / GB / LR	99.9	Dataset simplicity

3. TYPES OF CYBERSECURITY ATTACKS

WSNs are at risk from a number of frequent assaults, such as ARP spoofing, SYN/ACK flooding, and ICMP Echo floods, which can overwhelm nodes and reroute communication channels. Brute-force SSH attempts and UDP port scans take use of service flaws, while distributed denial-of-service (DDOS) attacks use a lot of network capacity by making a lot of requests at once. These assaults demonstrate the need for effective and portable IDS that can identify anomalous activity at several network tiers. As shown in Figure 1, the main dataset parameters and attack classes are illustrated.

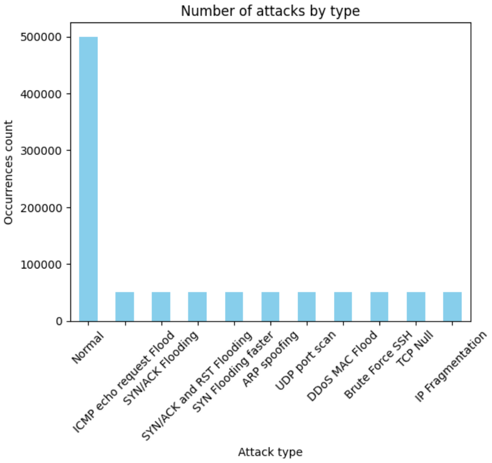


Figure 1. Dataset parameters

4. METHODOLOGY

Using the IDSAI dataset [4], we employed a systematic evaluation procedure in this study to evaluate the effectiveness of IDS models based on ML algorithms [20], [23]. Dataset preparation, feature engineering and selection, model training, cross-validation, and performance evaluation are the five primary processes of the methodology. The hierarchical IDS architecture employed in this investigation is depicted in Figure 2, with RF functioning at the cluster-head level, LR at the sensor layer, and GB at the gateway for retraining and validation.

4.1. Dataset preparation

The IDSAI dataset, which replicates actual cyberattacks in WSNs, we employed. The IDSAI dataset was originally introduced by Fernando *et al.* [4], and it has been widely used for evaluating IDS in IoT environments. Labeled traffic cases from a range of attack methods, including DDoS, ARP spoofing, and port scanning, are including this collection. The IDSAI dataset comprises more than eighty thousand labeled flows that encompass both standard traffic and a variety of attack categories, including DDoS, ARP spoofing, SYN flooding, and port scanning. In order to maintain the distribution of labels across folds, we employed stratified sampling and verified class proportions. Variance analysis revealed low noise and partial feature redundancy,which may result in prominent metrics being inflated. Consequently, we provide fold-wise statistics and explicitly address overfitting risks in sections 5.1 and 5.2.

4.2. Feature selection and engineering

We used filter-based feature selection techniques, such as mutual information and variance thresholding, to increase model efficiency and decrease overfitting. Based on their contribution to classification performance, we kept the most discriminative features that were pertinent to intrusion detection. This choice enhances interpretability, reduces redundancy, and expedites training without compromising the quality of detection.

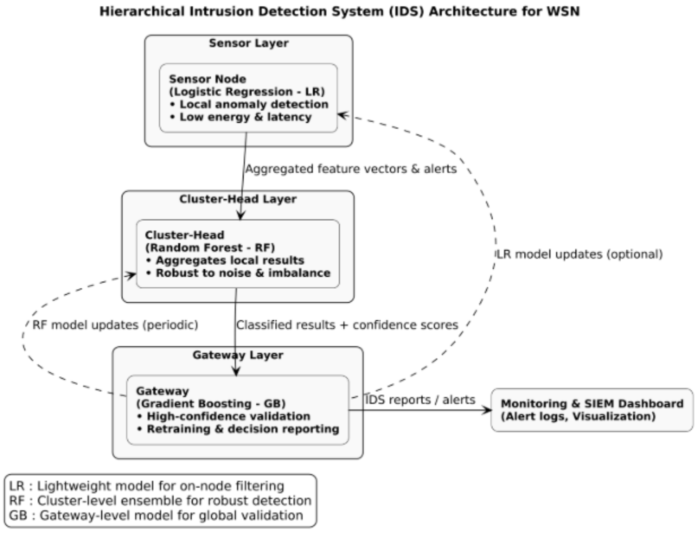

Hierarchical Intrusion Detection System (IDS) Architecture for WSN
Sensor Layer
Sensor Node
(Logistic Regression - LR)
• Local anomaly detection
• Low energy & latency
Aggregated feature vectors & alerts
Cluster-Head Layer
Cluster-Head
(Random Forest - RF)
• Aggregates local results
• Robust to noise & imbalance
LR model updates (optional)
RF model updates (periodic)
Classified results + confidence scores
Gateway Layer
Gateway
(Gradient Boosting - GB)
• High-confidence validation
• Retraining & decision reporting
IDS reports / alerts
Monitoring & SIEM Dashboard
(Alert logs, Visualization)
LR : Lightweight model for on-node filtering
RF : Cluster-level ensemble for robust detection
GB : Gateway-level model for global validation

Figure 2. Hierarchical IDS architecture for WSN

4.3. Model training

We chose three popular ML classifiers: RF [23], LR [24], and GB [25]. L2 regularization was used to train the LR model. Grid search was used to determine the inverse of regularization strength (C).

4.4. Mathematical overview

The LR model minimizes the ℓ_2-regularized negative log-likelihood:

$$J(\boldsymbol{\theta}) = -\frac{1}{m}\sum_{i=1}^{m}\Big[y_i \log h_{\boldsymbol{\theta}}(\mathbf{x}_i) + (1-y_i)\log\big(1-h_{\boldsymbol{\theta}}(\mathbf{x}_i)\big)\Big] + \lambda\|\boldsymbol{\theta}\|_2^2, \tag{1}$$

with $h_{\boldsymbol{\theta}}(\mathbf{x}) = \frac{1}{1+\exp(-\boldsymbol{\theta}^\top\mathbf{x})}$. RF aggregates T decision trees $\{f_t\}_{t=1}^{T}$ by majority vote, $\hat{y} = \mathrm{mode}\big(f_1(\mathbf{x}), \ldots, f_T(\mathbf{x})\big)$. GB builds an additive model $F_M(\mathbf{x}) = \sum_{m=1}^{M} \eta\, h_m(\mathbf{x})$, where h_m are shallow trees fitted stage-wise to the negative gradients of the loss, and η is the learning rate. Grid search selected hyperparameters to balance bias and variance.

4.5. Cross-validation and generalization assessment

During model evaluation, we used stratified 5-fold cross-validation to guarantee the robustness and generalizability of our findings. In order to monitor training behavior and identify any possible overfitting or underfitting tendencies, learning curves were created. All preprocessing stages (feature selection, scaling when applicable, and model fitting) were executed within each training fold of the stratified 5-fold cross-validation pipeline. Test folds were concealed until the final scoring in order to prevent optimistic bias.

4.6. Performance metrics

Precision, recall, F1-score, accuracy, and receiver operating characteristic-area under the curve (ROC-AUC) are common classification metrics that we used to evaluate the performance of the models. Prediction probability histograms, ROC curves, and precision-recall curves were used to display these metrics, which were calculated for every model. Every outcome was examined in light of the models interpretability and usefulness for IDS deployment in WSNs [26].

5. RESULTS OF DATASET

To achieve our research objectives, we propose a systematic methodology for evaluating IDS performance with ML algorithms. Figure 3 illustrates the precision-recall curves for LR, RF, and GB models, highlighting that GB and RF maintain near-perfect precision across almost the entire recall range, whereas LR shows a noticeable drop in performance as recall increases. Figure 4 demonstrates that while LR suffers from overfitting, RF generalizes effectively with growing data. GB improves validation scores while providing a balanced performance.

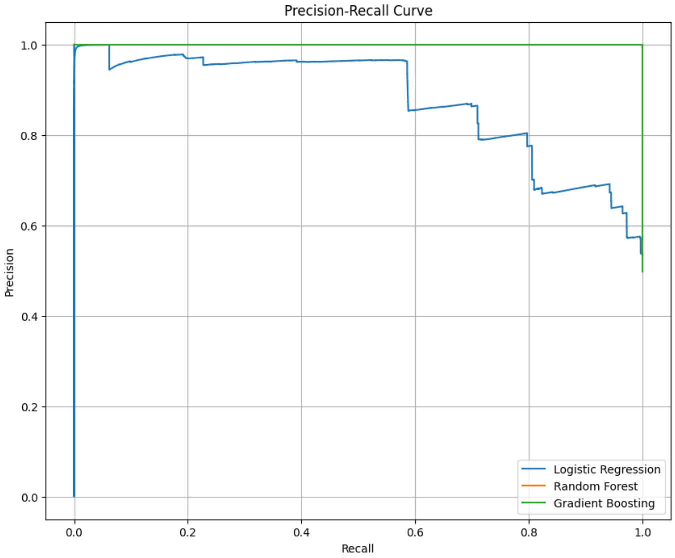


Figure 3. Precision-Recall curve for LR, RF, and GB models

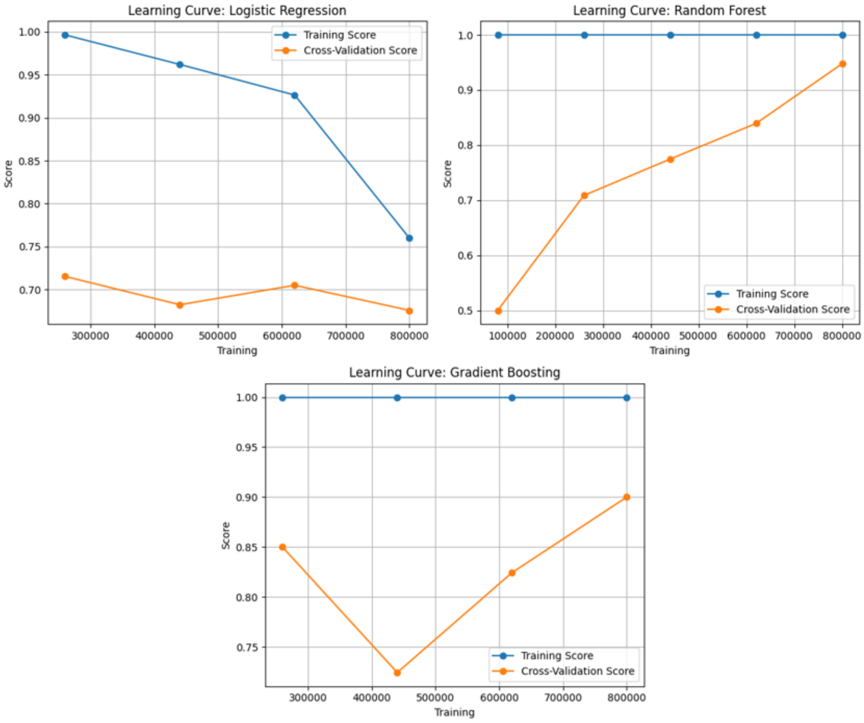


Figure 4. Learning curves (train vs. CV) showing: LR overfits as data grows (gap persists), RF maintains high and stable generalization, and GB improves validation steadily—aligning with the metric ranking in Table 1

Figure 5 illustrates the learning trajectories of the three models. RF demonstrates a high degree of generalization, with validation scores that continue to improve as the training set expands. GB closely matches RF and enhances validation performance, whereas LR overfits in late regimes (training validation), which explains its lower recall and F1-score.

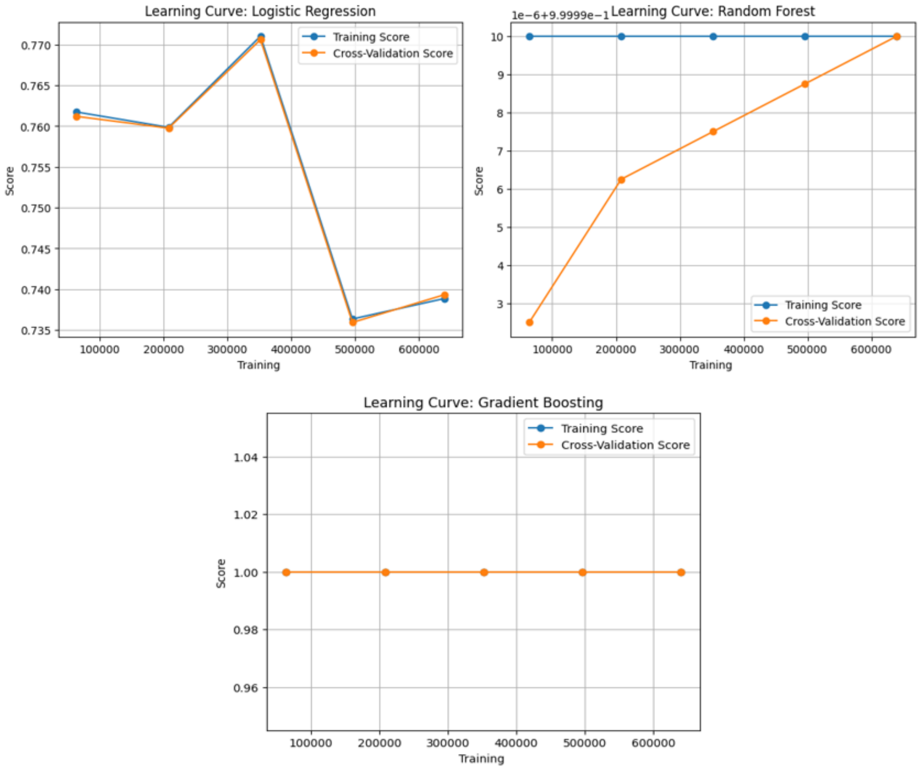


Figure 5. Learning curves confirm RF's superior generalization; GB is a close second; LR exhibits persistent train–validation gap (overfitting), consistent with its lower recall and F1-score

5.1. Prediction probability distribution

Figure 6 shows that ensemble models (RF and GB) give more confident predictions, while LR shows more uncertainty in its probability estimates. Table 2 summarizes the performance of the three classifiers (mean ± SD over 5 stratified folds). RF achieves the best overall performance across all metrics; GB is a close, well-balanced runner-up. LR attains high precision but noticeably lower recall, leading to a lower F1-score than the ensemble methods. To prevent leakage, all preprocessing is conducted within each training fold, and these intervals are calculated over five stratified folds.

We report 95% confidence intervals together with fold-wise means and standard deviations for all metrics over five stratified folds. For accuracy, RF attains the highest score (0.9994 ± 0.0001; 95% CI [0.9994, 0.9995]), GB is close (0.9990 ± 0.0001; [0.9989, 0.9991]), while LR is lower (0.9167 ± 0.0010; [0.9158, 0.9175]). LR's lower recall leads to a lower F1-score than the ensemble methods. In order to evaluate the performance disparities among models, a non-parametric Friedman test was implemented on fold-wise F1-scores. The results indicated that at least one model performed differently, as evidenced by the significant aggregate difference ($\chi^2(2) = 10.00$, $p = 0.0067$). The post-hoc Wilcoxon signed-rank tests with Holm correction revealed no significant difference between RF and GB ($p_{\text{adj}} > 0.05$), thereby confirming that both ensemble models obtain consistently high performance. This statistical consistency emphasizes the reliability and robustness of the data, thereby bolstering the credibility of the comparative framework.

5.2. Overfitting consideration

The unusually high values of precision and recall (near 1.0) require critical consideration. These may be due to: i) well-separated class boundaries in the IDSAI dataset, ii) feature redundancy or low noise, and iii) lack of real-world diversity in attack vectors. We mitigated overfitting risks through 5-fold stratified cross-validation and comparison across multiple metrics and plots. Nonetheless, future work will integrate more challenging datasets to further assess generalizability.

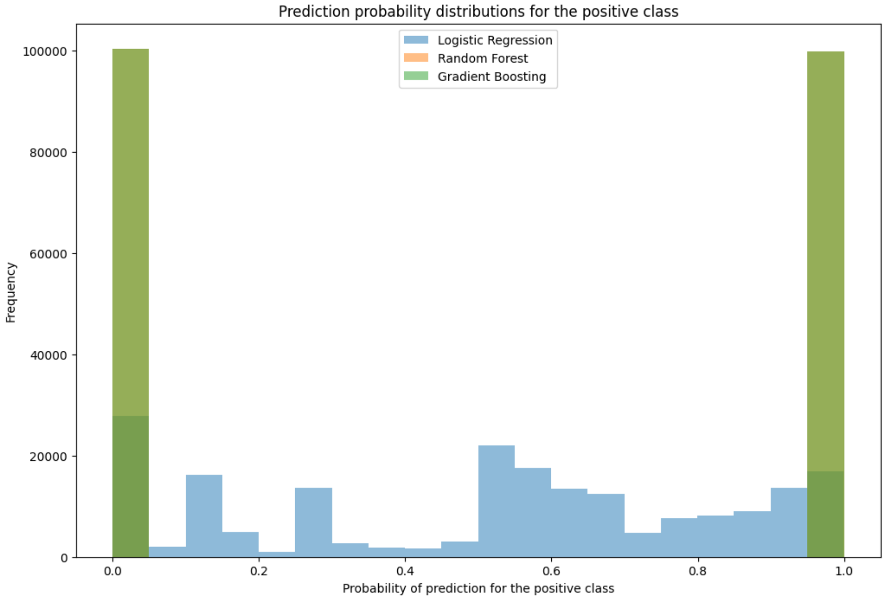


Figure 6. Prediction probability distributions: RF/GB yield confident, well-separated posteriors; LR shows broader uncertainty, consistent with its reduced recall

Table 2. Classification performance (mean ± SD over 5 stratified folds)

Model	Accuracy	Precision	Recall	F1-score	ROC–AUC
LR	0.9167 ± 0.0010	0.9829 ± 0.0002	0.8481 ± 0.0018	0.9105 ± 0.0011	0.9707 ± 0.0001
RF	0.9994 ± 0.0001	0.9995 ± 0.0001	0.9994 ± 0.0001	0.9994 ± 0.0001	0.9998 ± 0.0000
GB	0.9990 ± 0.0001	0.9995 ± 0.0001	0.9985 ± 0.0001	0.9990 ± 0.0001	1.0000 ± 0.0000

Low label noise and well-separated classes are corroborated by the nearly flawless metrics. In spite of this, learning curves indicate disparities in capacity (LR overfitting in late regimes versus RF stability). There are plans for future validation on a wider range of traffic, such as unseen devices, and blended protocols, to stress-test generalization. The computational footprint of each model is summarized in Table 3. Training is performed offline; inference reflects on-device cost in deployment.

Interpretation: LR is the optimal choice for embedded filtering at sensor nodes due to its ability to generate the quickest inference ($\sim 6.5 \times 10^5$ samples/s). RF achieves a robust accuracy-cost trade-off ($\sim 1.3 \times 10^5$ samples/s) that is suitable for cluster-heads, while GB remains viable at the gateway with competitive inference speed ($\sim 2.2 \times 10^5$ samples/s) despite being more expensive to train. These on-device inference costs are the primary constraint for real-world deployment, as training is conducted offline.

Table 3. Computational footprint on a modest PC

Model	Training time (s)	Inference on 9.98×10^5 samples (s)	Throughput (samples/s)
LR	823.40	1.54	6.48×10^5
RF	125.02	7.48	1.33×10^5
GB	460.33	4.47	2.23×10^5

6. DISCUSSION OF THE RESULTS

The results demonstrate the strong capabilities of ML algorithms in detecting intrusions in WSNs. RF achieves the strongest overall performance; GB is a close, well-balanced alternative. LR attains high precision but lower recall, leading to a lower F1-score than the ensemble methods. However, learning curves reveal overfitting as dataset size increases, limiting its generalization. GB is a well-balanced, close alternative to RF, which obtains the strongest overall performance. The F1-score is lower than that of the ensemble methods due to the fact that LR achieves high precision but a reduced recall. GB offered balanced performance, with high confidence in predictions and competitive scores across all metrics. In order to minimize false negatives in IDS, precision-recall curves verify that RF and GB have good precision even as recall rises. Additionally, prediction probability distributions demonstrate that, in contrast to LR, RF, and GB offer more certain classifications . The near-perfect metrics may be indicative of minor overfitting or dataset simplicity, despite the highly encouraging results. Therefore, these discoveries should be verified in future research using datasets that are more intricate and diverse. In general, GB provides a balanced performance, LR is appropriate for straightforward scenarios, and RF remains the most scalable and dependable option for intrusion detection in real WSNs

7. CONCLUSION AND DEPLOYMENT INSIGHTS

We recommend a hierarchical IDS, which consists of LR at sensor nodes (negligible latency), RF at cluster-heads (robust aggregation), and GB at the gateway (validation and periodic retraining). This approach minimizes communication overhead and concentrates heavier computation in areas where resources are less constrained. Using the IDSAI dataset, this study offered a systematic assessment of ML-based IDS in WSNs. We illustrated the benefits of each model by contrasting RF, GB, and LR. On IDSAI, RF exhibited superior generalization and stability, while GB was a close, well-balanced second. The F1-score was lower than that of the ensemble methods due to the fact that LR achieved high precision but a reduced recall. The findings highlight how crucial it is to choose ML models based on the particular deployment context, whether that context is one of scalability, accuracy, or interpretability. In order to increase IDS reliability, the study also examined overfitting concerns and the requirement for realistic, diverse attack data. To sum up, our results confirm the importance of ML in protecting WSNs and recommend that more sophisticated and adaptable methods that can manage dynamic and diverse IoT settings be explored in future studies. In order to further enhance IDS flexibility in dynamic IoT contexts, future research will concentrate on merging federated and hybrid learning methodologies.

ACKNOWLEDGMENTS

The authors would like to thank the LIMATI Laboratory and the Faculty of Polydisciplinary at University Sultan Moulay Slimane for their scientific guidance, technical resources, and continuous support during the preparation of this work.

FUNDING INFORMATION

The author(s) received no financial support for the research, authorship, and/or publication of this article.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Mansour Lmkaiti	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	
Houda Moudni		✓		✓	✓	✓	✓			✓		✓		
Hicham Mouncif	✓	✓		✓	✓	✓	✓			✓		✓		

C	: **C**onceptualization	I	: **I**nvestigation	Vi	: **Vi**sualization
M	: **M**ethodology	R	: **R**esources	Su	: **Su**pervision
So	: **So**ftware	D	: **D**ata Curation	P	: **P**roject Administration
Va	: **Va**lidation	O	: Writing - **O**riginal Draft	Fu	: **Fu**nding Acquisition
Fo	: **Fo**rmal Analysis	E	: Writing - Review & **E**diting		

CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

DATA AVAILABILITY

The data underpinning the results of this study are accessible from the corresponding author upon reasonable request.

REFERENCES

[1] C. S. W. Ng, M. N. Amar, A. J. Ghahfarokhi, and L. S. Imsland, “A survey on the application of machine learning and metaheuristic algorithms for intelligent proxy modeling in reservoir simulation,” *Computers & Chemical Engineering*, vol. 170, Feb. 2023, doi: 10.1016/j.compchemeng.2022.108107.

[2] S. Tabbassum and R. K. Pathak, “Effective data transmission through energy-efficient clustering and fuzzy-Based IDS routing approach in WSNs,” *Virtual Reality & Intelligent Hardware*, vol. 6, no. 1, pp. 1–16, Feb. 2024, doi: 10.1016/j.vrih.2022.10.002.

[3] B. Suresh and G. S. C. Prasad, “An energy efficient secure routing scheme using LEACH protocol in WSN for IoT networks,” *Measurement: Sensors*, vol. 30, Dec. 2023, doi: 10.1016/j.measen.2023.100883.

[4] G.-P. Fernando, A.-A. H. Brayan, A. M. Florina, C.-B. Liliana, A.-M. H. -Gabriel, and T.-S. Reinel, “Enhancing intrusion detection in IoT communications through ML model generalization with a new dataset (IDSAI),” *IEEE Access*, vol. 11, pp. 70542–70559, 2023, doi: 10.1109/ACCESS.2023.3292267.

[5] S. Nieland, R. Oostendorp, M. Heinrichs, and R. Cyganski, “Transferability analysis of user groups in travel behaviour surveys using a random forest classification model,” *Transportation Research Procedia*, vol. 76, pp. 81–95, 2024, doi: 10.1016/j.trpro.2023.12.040.

[6] B. Kolukisa, B. K. Dedeturk, H. Hacilar, and V. C. Gungor, “An efficient network intrusion detection approach based on logistic regression model and parallel artificial bee colony algorithm,” *Computer Standards & Interfaces*, vol. 89, Apr. 2024, doi: 10.1016/j.csi.2023.103808.

[7] N. Balakrishnan, A. Rajendran, D. Pelusi, and V. Ponnusamy, “Deep belief network enhanced intrusion detection system to prevent security breach in the internet of things,” *Internet of Things*, vol. 14, Jun. 2021, doi: 10.1016/j.iot.2019.100112.

[8] T. Nandy, R. Md Noor, R. Kolandaisamy, M. Y. I. Idris, and S. Bhattacharyya, “A review of security attacks and intrusion detection in the vehicular networks,” *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 2, Feb. 2024, doi: 10.1016/j.jksuci.2024.101945.

[9] N. Dharini, J. Katiravan, S. D. M. Priya, and S. V. A. Sneghaa, “Intrusion detection in novel WSN-Leach Dos attack dataset using machine learning based boosting algorithms,” *Procedia Computer Science*, vol. 230, pp. 90–99, 2023, doi: 10.1016/j.procs.2023.12.064.

[10] 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, Mar. 2024, doi: 10.1016/j.csi.2023.103802.

[11] D. A. J. Rajan and E. R. Naganathan, “Trust based anonymous intrusion detection for cloud assisted WSN-IOT,” *Global Transitions Proceedings*, vol. 3, no. 1, pp. 104–108, Jun. 2022, doi: 10.1016/j.gltp.2022.04.022.

[12] S. Md Zin, N. B. Anuar, M. L. M. Kiah, and I. Ahmedy, “Survey of secure multipath routing protocols for WSNs,” *Journal of Network and Computer Applications*, vol. 55, pp. 123–153, Sep. 2015, doi: 10.1016/j.jnca.2015.04.018.

[13] R. Yadav, I. Sreedevi, and D. Gupta, “Augmentation in performance and security of WSNs for IoT applications using feature selection and classification techniques,” *Alexandria Engineering Journal*, vol. 65, pp. 461–473, Feb. 2023, doi: 10.1016/j.aej.2022.10.033.

[14] M. R. Kadri, A. Abdelli, J. B. Othman, and L. Mokdad, “Survey and classification of Dos and DDos attack detection and validation approaches for IoT environments,” *Internet of Things*, vol. 25, Apr. 2024, doi: 10.1016/j.iot.2023.101021.

[15] T. T. Lai, T. P. Tran, J. Cho, and M. Yoo, “DoS attack detection using online learning techniques in wireless sensor networks,” *Alexandria Engineering Journal*, vol. 85, pp. 307–319, Dec. 2023, doi: 10.1016/j.aej.2023.11.022.

[16] H. Ajmi *et al.*, “Efficient and lightweight in-memory computing architecture for hardware security,” *Journal of Parallel and Distributed Computing*, vol. 190, Aug. 2024, doi: 10.1016/j.jpdc.2024.104898.

[17] M. Kharjana, F. H. Pohrmen, S. C. Sahana, and G. Saha, “Blockchain-based key management system in named data networking: A survey,” *Journal of Network and Computer Applications*, vol. 220, Nov. 2023, doi: 10.1016/j.jnca.2023.103732.

[18] L. Ecke, M. Magdolen, S. Jaquart, R. Andre, and P. Vortisch, “A case study of checking national household travel survey data with machine learning,” *Transportation Research Interdisciplinary Perspectives*, vol. 24, Mar. 2024, doi: 10.1016/j.trip.2024.101078.

[19] D. Ma, X. Li, J. Liang, Z. Wang, and W. Yang, “Distilling seed-assisted zeolite synthesis conditions by machine learning,” *Microporous and Mesoporous Materials*, vol. 339, Jul. 2022, doi: 10.1016/j.micromeso.2022.112029.

[20] C. Alex, G. Creado, W. Almobaideen, O. A. Alghanam, and M. Saadeh, “A comprehensive survey for IoT security datasets taxonomy, classification and machine learning mechanisms,” *Computers & Security*, vol. 132, Sep. 2023, doi: 10.1016/j.cose.2023.103283.

[21] S. M. S. Bukhari, M. H. Zafar, M. A. Houran, Z. Qadir, S. K. R. Moosavi, and F. Sanfilippo, “Enhancing cybersecurity in Edge IIoT networks: An asynchronous federated learning approach with a deep hybrid detection model,” *Internet of Things*, vol. 27, Oct. 2024, doi: 10.1016/j.iot.2024.101252.

[22] D. Karunkuzhali, K. P. Arunachalam, R. Ramamoorthi, and R. K. Kadu, “Cyber-physical system for enhanced WSN-IoT security using spherical graph triple convolutional neural network with planet optimization algorithm,” *Progress in Engineering Science*, vol. 2, no. 3, Sep. 2025, doi: 10.1016/j.pes.2025.100108.

[23] I. Saadi, A. Mustafa, J. Teller, and M. Cools, "A bi-level random forest based approach for estimating O-D matrices: Preliminary results from the Belgium national household travel survey," *Transportation Research Procedia*, vol. 25, pp. 2566–2573, 2017, doi: 10.1016/j.trpro.2017.05.301.

[24] M. O. M. Mohammed, "Prevalence and risk factors associated with under-five years children diarrhea in Malawi: Application of survey logistic regression," *Heliyon*, vol. 10, no. 7, Apr. 2024, doi: 10.1016/j.heliyon.2024.e29335.

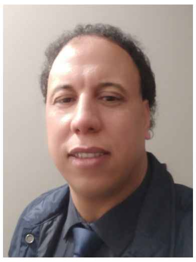
[25] A. Manoharan, K. M. Begam, V. R. Aparow, and D. Sooriamoorthy, "Artificial neural networks, gradient boosting and support vector machines for electric vehicle battery state estimation: A review," *Journal of Energy Storage*, vol. 55, Aug. 2022, doi: 10.1016/j.est.2022.105384.

[26] M. Lmkaiti, I. Larhlimi, M. Lachgar, H. Moudni, and H. Mouncif, "Advanced optimization of RPL-IoT protocol using ML algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 16, no. 2, 2025, doi: 10.14569/IJACSA.2025.01602135.

BIOGRAPHIES OF AUTHORS

Mansour Lmkaiti is from Department of Computer Mathematics, Faculty of Polydisciplinary, University Sultan Moulay Slimane, Morocco. His domains of interests is high-performance computer systems and networks: theory, machine learning algorithms; high performance in WSNs; and cybersecurity in wireless sensor networks. He can be contacted at email: lamkaitimansour@gmail.com.

Houda Moudni is currently working as theis an Assistant Professor at the National School of Business and Management, Sultan Moulay Slimane University, Béni Mellal, Morocco. She received the Ph.D. degree in Computer Sciences from the Faculty of Sciences and Technology of Beni Mellal in 2019. She developed a strong interest in computer networking. Her research work primarily focuses on securing routing protocols in mobile Ad Hoc networks (MANET), wireless sensor networks (WSN), and the internet of things (IoT). She can be contacted at email: h.moudni@usms.ma.



Hicham Mouncif is from Department of Computer Mathematics, University Sultan Moulay Slimane, Morocco. He is currently working as the Professor at the Department of Mathematics and Informatics. His research interests include computer networking, communication engineering, and securing routing protocols in wireless sensor networks. His domains of interests is high-performance computer systems and networks: theory, machine learning algorithms; high performance in WSNs and cybersecurity. He can be contacted at email: h.mouncif@usms.ma.