RESEARCH ARTICLE

An Optimized Ensemble Approach for Securing Wireless Sensor Networks Against Attacks

Anshika Sharma[1] **and Shalli Rani**[1,*]

[1] Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

Abstract

Wireless Sensor Networks (WSNs) are prone to different security threats because of their open communication environment, distributed architecture, and resource constraints. For the security and integrity of a network to be ensured, robust intrusion detection systems (IDS) are required. The WSN-DS dataset has been used to provide an effective machine learning (ML) and Deep Learning (DL) based IDS and attack detection technique for WSNs. Several learning models, including K-Nearest Neighbors (KNN), Random Forest (RF), Decision Tree (DT), Convolutional Neural Networks (CNN), Support Vector Machine (SVM), Logistic Regression (LR), and Neural Networks (NN), are compared in terms of performance. Preprocessing methods of data encoding, normalization, and data splitting are applied to the dataset to improve classification performance. The effectiveness of these models is compared using large trials with significant performance metrics such as ROC-AUC, F1-score, accuracy, precision, and recall. The results indicate that the Optimized RF model has been optimized to achieve the optimal accuracy of 99.71%, which outperforms other state-of-the-art approaches. Apart from pointing out the importance of ML in detecting WSN attacks, this research provides a promising way forward for enhancing network security through effective detection methods.

Keywords: wireless sensor networks, intrusion detection system, security, machine learning, deep learning, WSN-DS dataset.

1 Introduction

Since WSNs have been widely applied in many areas, including industrial automation, healthcare, military reconnaissance and environmental sensing, they have been subject to a great deal of attention [1, 2]. WSNs consist of small, low-power, wirelessly communicating sensor nodes (SN) that collect and transmit data [3, 4]. But they are highly vulnerable to security threats, such as denial-of-service (DoS) attacks, blackhole attacks, sinkhole attacks, and other malicious activities, because of their resource-constrained nature, distributed structure, and open communication medium [5, 6]. Robust IDS are required for the effective detection and prevention of these attacks to ensure the reliability and security of WSNs [7, 8]. Using data-driven approaches, ML-based IDS have emerged as a promising choice for detecting anomalies and attacks in WSNs [19, 20].

Submitted: 10 March 2025
Accepted: 26 May 2025
Published: 19 June 2025

Vol. 1, **No.** 1, 2025.
10.62762/TWN.2025.109626

***Corresponding author:**
✉ Shalli Rani
shallir79@gmail.com

Citation
Sharma, A., & Rani, S. (2025). An Optimized Ensemble Approach for Securing Wireless Sensor Networks Against Attacks. *ICCK Transactions on Wireless Networks*, 1(1), 5–15.

© 2025 by the Authors. Published by Institute of Central Computation and Knowledge. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

Although learning-based models can monitor traffic patterns in real-time and detect anomalous behaviour, conventional rule-based IDS may struggle to adapt to new and evolving threats [17, 18]. A few ML and DL models are implemented here to design an effective IDS for WSNs including LR, DT, RF, SVM, KNN, CNN, and NN. These models are trained and evaluated using the WSN-DS dataset, a popular benchmark dataset for WSN-IDS.

Comparing how well different learning models detect WSN threats is the main goal of this study, which focuses on important evaluation metrics including accuracy, recall, precision, F1-score, and ROC-AUC. This work's contributions include:

- Using the WSN-DS dataset, a thorough assessment of various ML and DL models for WSN intrusion detection.
- Random Forest model optimization to improve detection precision and lower false positives.
- A comparison with other state-of-the-art methods demonstrates how successful the proposed strategy is.
- A discussion of how various performance criteria affect the selection of the best model for practical WSN security applications.

This paper's remaining sections are organized as follows: Related work on ML and DL-based IDS in WSNs is presented in Section 2. Section 3 presents an architecture of WSN in attack detection. The WSN-DS dataset, data preprocessing methods and evaluation metrics are explained in Section 3 with the proposed methodology. The experimental results and comparative analysis are presented in Section 5, and Section 6 concludes and outlines the next directions.

2 Related Work

Tan et al. [1] suggest balancing the dataset using the synthetic minority oversampling technique (SMOTE) and training the intrusion detection classifier using the RF algorithm to secure the WSN against the attacks. A benchmark intrusion dataset is used for the simulations, and the RF's accuracy of 92.39% is higher than that of other comparative techniques. The accuracy of the RF and SMOTE combined has grown to 92.57% after the minority samples were oversampled. Salmi et al. [2] used a computer-generated WSN-DS dataset in this investigation. Using network simulator NS-2, which is based on the LEACH routing protocol, the WSN was simulated. Data was collected from the network and preprocessed to create 23 features that classified the condition of each sensor and mimicked 5 different types of DoS attacks. On a scale of 0 to 1, the created CNN-LSTM model is further assessed over 25 epochs, yielding accuracy, precision, and recall scores of 94.4%, 95.9%, and 92.2%, respectively.

Ifzarne et al. [3] introduced an incremental ML-based intelligent IDS approach. Based on a cluster WSN topology, the model detects the existence of an intrusion and instantly classifies the type of attack. A dataset from a WSN-DS was used for the investigation. With a 96% detection rate, the suggested model ID-GOPA can identify if the network is under attack or operating normally. Saleh et al. [4] used ML methods, specifically the Stochastic Gradient Descent (SGD) and Gaussian Nave Bayes (GNB) algorithms. By incorporating context awareness, recommendation systems become more effective. They initially analyze the raw traffic data using PCA and SVD to reduce the computational load. The proposed SG-IDS model surpassed state-of-the-art algorithms, which attained a 96% accuracy rate on the WSN-DS dataset. The SG-IDS demonstrated excellent performance in an examination of an IoMT dataset, achieving an accuracy of 87% and a precision of 100% in ID tasks.

Sadia et al. [5] introduced a sophisticated NIDS to protect Wi-Fi-based WSNs from cyber threats such as impersonation, flooding, and injection attacks. From 154 features, 76 are selected and cut down to 13 key features for efficient analysis. This paper proposes a CNN-based technique for effective ID and prevention in WSN, using common scaler functions for feature scaling and preprocessing. The study compares CNN, DNN and LSTM networks to improve detection accuracy, minimize loss values, and lower false alarm rates. Model performance is assessed using metrics such as precision, recall, support, F1 score, and macroaverage. The research results in a CNN model with an astounding 97% accuracy rate, 0.14 loss measure, and negligible False Alarm Rate. Using a specialized WSN-DS dataset, Tabba et al. [6] investigate the use of various kinds of online ensembles in sensory data analysis to categorize four types of attacks: flooding, scheduling, grayhole, and blackhole attacks among regular network traffic. The heterogeneous ensemble including an Adaptive RF (ARF) in conjunction with the Hoeffding Adaptive Tree (HAT) algorithm and the homogeneous ensemble HAT, which consists of 10 models, both achieved superior detection rates of 96.84% and 97.2%, respectively, among

the suggested novel online ensembles. When considering the resource limitations of WSNs, the aforementioned approaches are productive and successful in addressing concept drift.

Ahmed et al. [7] suggest intrusion detection using ML for WSN. SVM paired with stochastic gradient descent (SGD) improves detection precision. The research also suggests integrating context information, or context awareness, to improve recommendation systems by taking into consideration user preferences and system characteristics or conditions. Principal component analysis (PCA) and singular value decomposition (SVD) decrease first traffic data to decentralize the system's computational load. A VG-IDS model categorizes network threats. The recommended WSN-DS method had a 96% accuracy rate and outperformed other sophisticated algorithms tested using the WSN-DS dataset. Accuracy, recall, and F1-measure rates improved to 98%, 96%, and 97%. Biswas et al. [8] demonstrated using ensemble RF (ERF) in WSN for anomaly detection. As the ensemble's basic learners, they select KNN, DT and NB. The RF was also built using bootstrap sampling. The activity recognition based on multi-sensor data fusion (AReM) dataset is a real-world sensor dataset that they used to test the ERF algorithm. Different performance metrics demonstrate that the ERF outperforms the base learners alone.

3 Architecture of Wireless Sensor Network for Attack Detection

A WSN is a dispersed network made up of many tiny, lightweight sensor nodes that connect wirelessly to observe and gather information about their environment [9, 10]. In many applications, such as industrial automation, healthcare, military surveillance, and environmental monitoring, these networks are essential [11]. A key component of WSNs' dependability and effectiveness is attack detection because of their wireless nature and resource constraints, which make them extremely vulnerable to security attacks. Multiple layers make up a WSN architecture, which facilitates processing, data aggregation, and transmission [12, 13]. The following elements are present in a WSN architecture:

- Sensor Node (SN): These are tiny, battery-operated gadgets with sensors to gather information about the environment, including motion, pressure, temperature, and humidity. SN exchanges information with nearby nodes to send data to the BS.
- Cluster Head (CH): Sensor nodes are frequently arranged in clusters, with a CH in charge of data aggregation and forwarding for each cluster. By serving as a bridge between SN and the base station (BS), CH lowers communication overhead.
- Base Station/Gateway: Data from several SNs is gathered and processed by the gateway node, also known as the BS. After processing, it sends the data to a computer network or a centralized server for additional examination.
- Computer Network and Server: The gathered information is transmitted to distant computers for analysis and storage to make decisions. Real-time access and interpretation of WSN data is made possible via the computer network.
- Attacker Nodes: WSNs are susceptible to some attacks, including node compromise, jamming, Sybil attacks, blackhole attacks, and sinkhole attacks, because they function in open and frequently hostile environments.

Figure 1 shows the architecture of WSN, including its architectural components and a possible attack scenario. WSNs are made up of a computer network, a gateway, a server, SNs, and CHs. To guarantee effective communication and data aggregation, this network's operation is organized hierarchically. WSN communication is hierarchical, with SN sending data to the appropriate CHs. After combining the received data, these CHs forward it to the gateway, which forwards the processed data to the server for storage and additional analysis. The computer network makes the gathered data available, enabling remote decision-making and monitoring. In a cluster, each SN is in charge of collecting particular environmental characteristics and sending the information to the CH. By cutting down on unnecessary transmissions and making sure that only pertinent, aggregated data is sent to the server, the CH are essential in maximizing energy usage. The gateway ensures smooth data transfer between SN and centralized processing units by serving as a bridge between the WSN and the outside network. Despite its effectiveness, a WSN's wireless communication medium and weak security measures make it extremely susceptible to hackers. To interfere with operations, attackers frequently target weak places in the network. The attacker as depicted in Figure 1 is an outside party that locates a vulnerable or exposed sensor node and initiates an attack to undermine its operation.

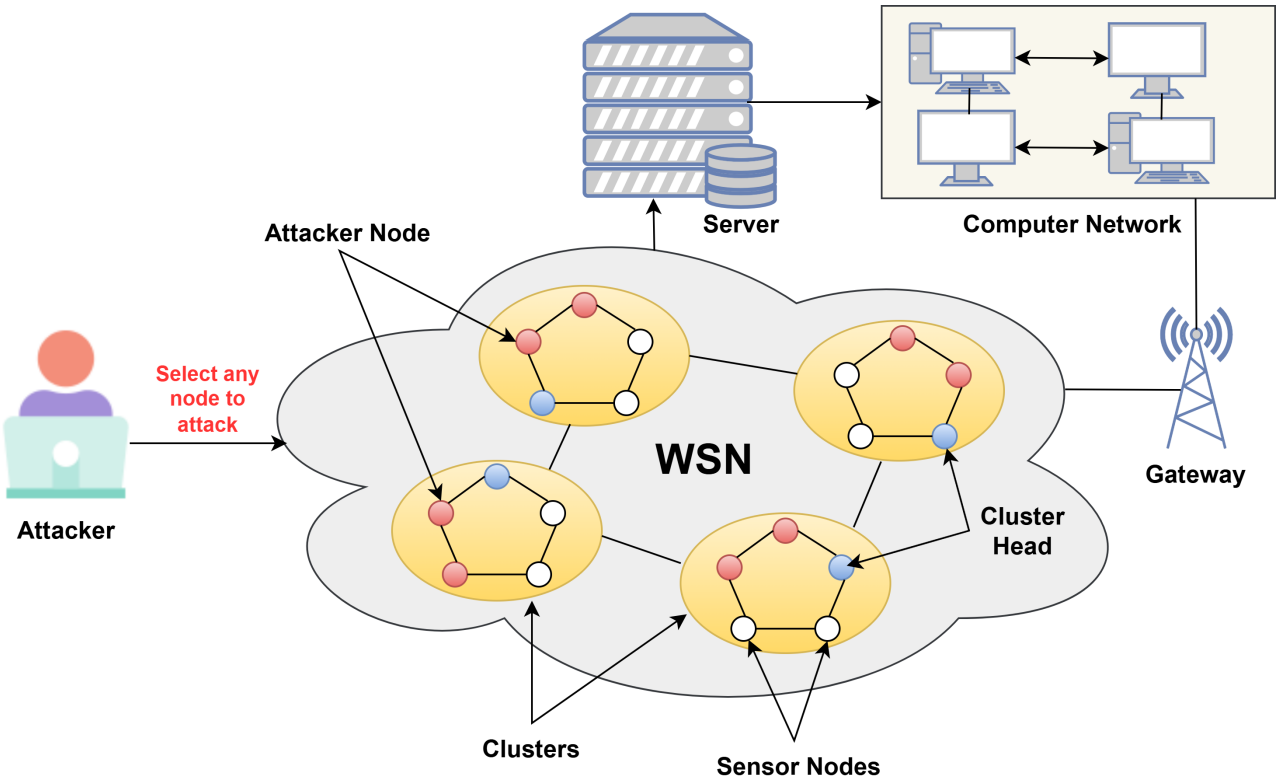


Figure 1. Architecture of wireless sensor networks.

Using security flaws like inadequate encryption, a lack of authentication, or energy fatigue, the attacker first chooses a target node inside a cluster. The compromised nodes begin acting maliciously, interfering with the regular operation of the uncompromised nodes, allowing the attacker to carry out a variety of malicious actions. Several nodes across different clusters can ultimately be impacted as the attack continues, leading to severe network failure, high latency, and security violations. In the presence of an attacker node within the WSN, there can be severe issues with data integrity, increased connection time, and reduced network performance. A compromised SN would be a point of entry for further exploitation, allowing the attacker greater control and power over more segments of the network. Effective detection mechanisms such as anomaly detection models and ML-based IDS are essential in mitigating these threats. These mechanisms can detect abnormal network traffic, identify compromised nodes, and prevent attacks before they cause extensive damage.

4 Proposed Methodology

Figure 2 illustrates the overall workflow of our proposed approach for securing WSNs against attacks. The methodology consists of four main stages: data preprocessing, model training, evaluation, and optimization. Each stage is carefully designed to ensure high detection accuracy while maintaining computational efficiency.

As shown in Figure 2, the process begins with data collection and preprocessing, followed by training multiple machine learning models. The best-performing model is then optimized and evaluated using various performance metrics. This systematic approach ensures reliable detection of different types of attacks in WSNs.

4.1 Dataset

A labeled dataset named the Wireless Sensor Network Dataset (WSN-DS) [1] was developed to evaluate IDS in WSN. To support the development of ML and DL-based security models, it simulates an actual WSN, recording both normal and attack situations. Various network-related and energy-based features in the dataset support the analysis of various types of WSN attacks. This dataset, which was created using NS-2 simulations, contains 374,662 samples in total with 19 features as shown in Table 1, divided into four different traffic types:

- Normal Traffic: Indicates benign network activity.

[1] https://www.kaggle.com/datasets/bassamkasasbeh1/wsnds

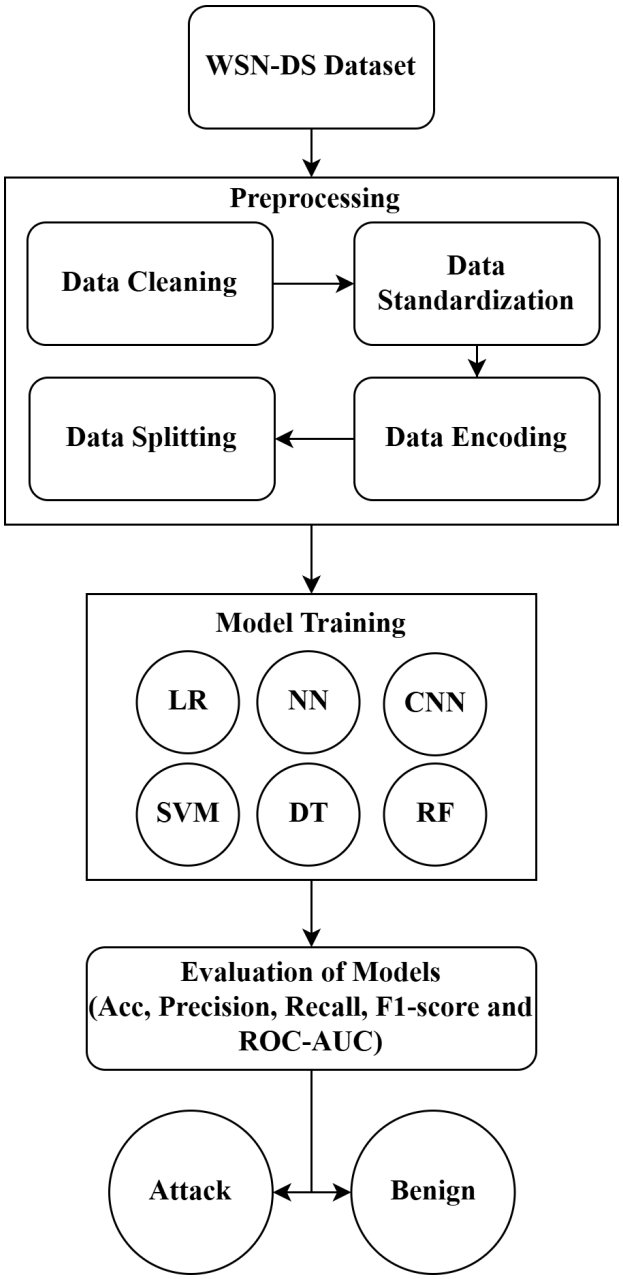


Figure 2. Proposed methodology for WSN attack detection.

- Blackhole: A situation in which malicious nodes consume all data packets and prevent them from accessing their intended destinations is simulated.
- Flooding: Shows instances in which attackers transmit too many packets to the network, overloading it and depleting its resources.
- Scheduling Attack: Nodes disrupt the communication schedule in a cluster-based WSN.

4.2 Data Preprocessing

The study's dataset has been preprocessed to guarantee high-quality training and assessment data. Initially, incomplete records are eliminated or missing data is imputed using the proper statistical techniques to handle missing values [14, 15]. To transform the dataset's categorical attributes into numerical representations, label encoding is used. A StandardScaler is used to standardize the feature values and guarantee uniformity. It converts the numerical values to have an average and standard deviation of 0 and 1 respectively. To provide a fair representation of attack and normal occurrences in both subsets, the dataset is divided into training and testing subsets after data normalization, in an 80:20 ratio. Data loosing is avoided and efficient model evaluation is made possible by this division. To facilitate multi-class classification, the target variable is one-hot encoded for DL models in the presence of several attack classes. The integer format of the categorical labels is maintained for ML models [16]. The accuracy and dependability of WSN attack detection systems are improved by this systematic data preparation method, which guarantees that the dataset is well-optimized for training different ML and DL models.

4.3 Model Training

The models are trained to categorize various threats in the WSN following data preprocessing. To discover patterns and connections between features and attack types, the processed dataset is fed into a variety of ML and DL models during the training phase. Hyperparameter tuning is applied to each model to maximize performance as shown in Table 2. While DL models including CNN and NN use multi-layer architectures to extract complex patterns from data, ML models including LR, DT, RF, SVM, and KNN are trained using conventional classification techniques. Choosing the optimal parameters, avoiding overfitting with regularization and dropout strategies, and guaranteeing strong generalization to unknown attack patterns are all steps in the training process. The best-performing method for WSN attack detection is identified by comparing the models' performance after they have been trained on an unseen test dataset. The different ML and DL models used for this study are explained here:

4.3.1 Logistic Regression

LR is a basic but effective classification algorithm that models the likelihood of a given input belonging to a specific class. It is based on the logistic (sigmoid) function, which converts input information into a likelihood score from 0 to 1. For multi-class

Table 1. Features and the description of the WSN-DS dataset.

Feature	Description
id	Unique identifier for each SN.
Time	Timestamp of the node.
Is-CH	Whether a node is a CH (1: Yes, 0: No).
who-CH	CH ID to which the node belongs.
Dist-ToCH	Distance of the SN from its respective CH.
Adv-S	Advertisement messages sent by the node.
Adv-R	Advertisement messages received by the node.
Join-S	Join requests sent by the node to a CH.
Join-R	Join responses received by the node from a CH.
Sch-S	Schedule messages sent by the node.
Sch-R	Scheduled messages received by the node.
Rank	Position or priority of the node in the network hierarchy.
Data-S	Data packets sent by the node.
Data-R	Data packets received by the node.
Data-SentToBS	Amount of data transmitted from CH to the BS.
Dist-CHToBS	Distance between the CH and the BS.
Send-Code	Code indicating the type of message sent by the node.
Expanded-Energy	Energy consumption of the node during communication.
Attack-Type	Indicates whether the node is under attack and the type of attack.

classification, softmax regression is used, which involves training several LR models to predict different attack types. It is computed by Eq. 1

$$P(v=1|u) = \frac{1}{1+e^{-(\beta_0+\beta_1 u_1+\beta_2 u_2+\ldots+\beta_n u_n)}} \quad (1)$$

where u_i, β_0 and β_i are the feature values, intercept and coefficients respectively.

4.3.2 *Support Vector Machine*

SVM is an effective classifier that identifies an ideal hyperplane to distinguish various classes in high-dimensional space. It operates by optimizing the margin between distinct classes through support vectors given by Eq. 2. Kernel functions Radial Basis Function (RBF), are employed to transform non-linearly separable data into a higher-dimensional space, facilitating linear separability.

$$\min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2 \quad (2)$$

subject to Eq. 9:

$$y_i(\mathbf{w}^T X_i + b) \geq 1, \quad \forall i \quad (3)$$

4.3.3 *Neural Network*

A NN is a DL architecture modelled after the human brain. It comprises several layers, including input, hidden, and output layers, with neurons that execute weighted computations succeeded by activation functions. Every layer acquires the ability to extract hierarchical features from the incoming data. Nevertheless, they necessitate meticulous adjustment of hyperparameters, including the number of layers, neurons, and learning rates, to get maximum performance. A fully connected NN computes the output as given by Eq. :

$$a^{(l)} = f(W^{(l)}a^{(l-1)} + b^{(l)}) \quad (4)$$

where $W^{(l)}$ and $b^{(l)}$ are the weight matrix and bias for layer l, and f is the activation function.

4.3.4 *Decision Tree*

A DT is a rule-based classifier that partitions data into hierarchical decision nodes according to the most salient attributes. The model recursively partitions the dataset into smaller subgroups utilizing Gini impurity or entropy as the criteria for division, hence creating a tree-like structure as computed in Eq. 5. Pruning methods can be utilized to restrict tree depth and enhance generalization.

$$H(S) = -\sum_{i=1}^{c} p_i \log_2 p_i \quad (5)$$

Table 2. Hyperparameter tuning of different learning models.

Model	Hyperparameters
LR	Solver=lbfgs, Penalty=l2, C=1.0
DT	Criterion=entropy, MaxDepth=20, MinSamplesSplit=2, MinSamplesLeaf=1
SVM	Kernel=RBF, C=1.0, Gamma=scale
KNN	NumberofNeighbors=5, DistanceMetric=Euclidean, WeightFunction=Uniform
CNN	NumberofLayers=4, ActivationFunction=ReLU, Optimizer=Adam , BatchSize=32, LearningRate=0.001
NN	NumberofHiddenLayers=3, ActivationFunction=ReLU, Optimizer=Adam, BatchSize=64, LearningRate=0.001
RF	NumberofEstimators=200, Criterion=Entropy, MaxDepth=50, MinSamplesSplit=4, MinSamplesLeaf=2, Bootstrap=True, MaxFeatures=sqrt

where p_i is the probability of i^{th} class .
And the Information Gain (IG) is calculated by Eq. 6:

$$IG = H(S) - \sum_{j=1}^{m} \frac{|S_j|}{|S|} H(S_j) \tag{6}$$

where $H(S)$ is the entropy of the set and $H(S_j)$ is the entropy of each subset.

4.3.5 K-Nearest Neighbour

KNN is a distance-based classification technique that allocates labels to a sample according to the predominant class of its K closest neighbours in feature space. The Euclidean distance metric is frequently employed to assess similarity between occurrences. It is a non-parametric model, indicating that it does not presume any particular data distribution. The Euclidean distance d is calculated by Eq. 7 between points v and w.

$$d(v, w) = \sqrt{\sum_{i=1}^{n} (v_i - w_i)^2} \tag{7}$$

4.3.6 Convolutional Neural Network

A CNN is a DL architecture commonly employed for feature extraction and classification tasks. The CNN architecture comprises convolutional layers that use filters to detect patterns, pooling layers that reduce dimensionality, and fully connected layers that execute classification. The output feature map is calculated by Eq. 8.

$$O(i, j) = \sum_{m} \sum_{n} U(i + m, j + n) K(m, n) \tag{8}$$

where $O(i, j)$, U and K are the output feature map, input and kernel respectively.

4.3.7 Random Forest

RF is an ensemble method that develops several DTs and combines their predictions to increase accuracy and reduce overfitting. Each tree in the forest is trained on a selected subset of data and attributes, and the ultimate prediction is obtained through majority voting (in classification). The final prediction is obtained using majority voting for classification by Eq. 9.

$$\hat{v} = \text{mode}(T_1(u), T_2(u), ..., T_k(u)) \tag{9}$$

where $T_k(u)$ is k-th tree prediction.

4.4 Evaluation Metric

This part discusses key evaluation measures that provide a comprehensive understanding of the strengths and weaknesses of the model in WSN attack detection, such as accuracy, precision, recall, F1-score, and ROC-AUC.

4.4.1 Accuracy

Accuracy measures the proportion of correctly classified instances out of all instances. It is formulated by Eq. 10.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

4.4.2 Precision

The precision determines the correctness of the positive predictions made by the model. A high precision value indicates that the model makes fewer FP errors, meaning higher reliability in its positive predictions. It is formulated by Eq. 11.

$$Precision = \frac{TP}{TP + FP} \tag{11}$$

Table 3. Comparative analysis of the different learning models with regards to performance metrics.

Model	Training Time	Testing Time	Accuracy	Balanced Accuracy	Precision	Recall	F1-score	ROC-AUC
LR	36.71	0.01	0.9517	0.8946	0.9634	0.9517	0.9559	0.9883
SVM	6.27	0.01	0.9664	0.8118	0.9642	0.9664	0.9644	0.9888
NN	162.25	2.88	0.9822	0.9055	0.9851	0.9822	0.9822	0.9943
DT	1.66	0.01	0.9943	0.9698	0.9944	0.9943	0.9943	0.9848
KNN	1.19	38.32	0.9939	0.9545	0.9939	0.9939	0.9939	0.9927
CNN	277.63	4.48	0.9873	0.9206	0.9878	0.9873	0.9822	0.9947
RF	38.02	1.06	0.9972	0.9815	0.9972	0.9972	0.9971	0.9966

4.4.3 Recall

Recall measures the model's effectiveness in detecting correct positive instances. It is formulated by Eq. 12.

$$Recall = \frac{TP}{TP + FN} \tag{12}$$

4.4.4 F1-score

F1-score is a harmonic mean of precision and recall, which balances their trade-off effectively. It is formulated by Eq. 13.

$$F1\text{-}score = \frac{2PR}{P + R} \tag{13}$$

where P and R are the precision and recall respectively.

4.4.5 ROC-AUC

The ROC curve plots the true positive rate (TPR) against the false positive rate (FPR) at multiple threshold levels, providing information about the model's performance at various categorization thresholds. The AUC (Area Under the Curve) measure ranges from 0 to 1, with a higher value indicating better discriminatory power.

5 Results and Analysis

This section analyzes and compares various ML and DL methods to determine how effectively the proposed models detect attacks in WSN. Accuracy, balanced accuracy, precision, recall, F1-score, and ROC-AUC are some of the evaluation metrics utilized to compare the performance of these models. A comparative study of various learning models is illustrated in Table 3, highlighting their performance in terms of classification accuracy and computing time.

It is evident from Figure 3 that RF has performed better than the rest of the models, providing the best accuracy of 99.72% along with better F1-score, precision, and recall. With an accuracy of 99.43%, the DT classifier

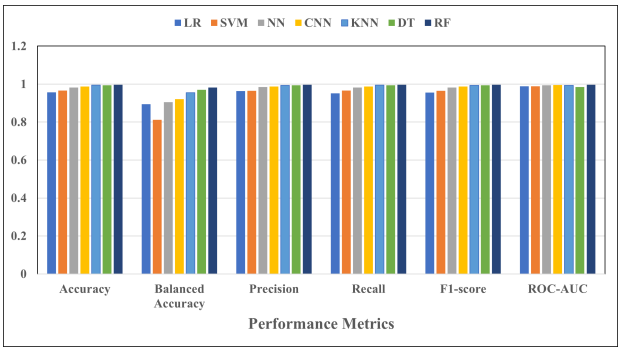


Figure 3. Comparative analysis of different learning models with regards to different performance metrics.

also exhibits remarkable performance, proving its capacity to successfully categorize attacks in WSN. With an accuracy of 99.39%, the KNN algorithm performs well; nevertheless, it is less effective for real-time applications due to its much longer testing time than other models. With an accuracy of 98.73%, the CNN model outperforms the NN, which has an accuracy of 98.22% among the DL models. With an accuracy of 96.64%, the SVM model performs admirably; nevertheless, its balanced accuracy is lower, which would suggest that it has some trouble managing class imbalances. With the lowest accuracy of 95.17%, LR may not be as good at identifying WSN attacks as more sophisticated models.

To reduce FP in attack detection, precision is essential. The highest precision is shown by RF, DT, and KNN, with respective values of 99.72%, 99.44%, and 99.39%. The precision values of the DL models CNN and NN, are likewise quite high, approaching 98.78% and 98.51%, respectively. Compared to other models, LR and SVM have the lowest precision values, 96.34% and 96.42%, respectively, suggesting a higher likelihood of misclassification. RF, DT, and KNN attained near-perfect recall values, showing that they successfully classified practically all attack cases. CNN and NN likewise maintain high recall

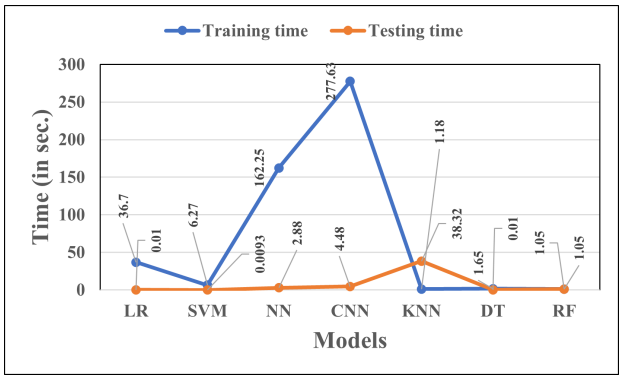


Figure 4. Comparative analysis of different learning models with regards to training and testing time.

values of 98.73% and 98.22%, respectively. Relatively low recall indicates that SVM and LR might not be able to identify every attack scenario as effectively as the best-performing models. With an F1-score of 99.71%, RF has the highest score, closely followed by DT (99.43%) and KNN (99.39%). CNN (98.22%) and NN (98.22%), two DL models, also score well, but marginally worse than conventional ML models. Due to their relatively poorer ability to balance FP and FN, SVM (96.44%) and LR (95.59%) have the lowest F1-scores. The model's capacity to discriminate across classes is gauged by the ROC-AUC. The highest ROC-AUC score is obtained by RF (99.66%), demonstrating its higher discriminatory capability. Closely behind are CNN (99.47%), NN (99.43%), and KNN (99.27%). With somewhat lower ROC-AUC, LR (98.83%) and SVM (98.88%) are less successful at differentiating between normal and attack cases.

Figure 4 represents the training and testing time of these learning models. While RF has a moderate training time of 38.02 seconds, it performs the best in classification. DT and KNN are computationally efficient since they have the shortest training times of 1.66s and 1.19s, respectively. KNN is unsuitable for real-time applications because of its longest testing duration of 38.32s. High computational costs are indicated by the much longer training times of 277.63s and 162.25s, respectively required for models such as CNN and NN. The model that performs the best overall is RF, which strikes a balance between high accuracy, precision, recall, and computational efficiency. DL models are also effective, although they are less useful for real-time WSN attack detection due to their lengthy training periods.

A comparison between the suggested optimized RF model and several cutting-edge methods for WSN attack detection is shown in Table 4. The comparison

Table 4. Comparative analysis of the proposed model with the state-of-the-art approaches.

Ref.	Techniques	Dataset	Accuracy
[1]	RF	Benchmark Intrusion	92.57%
[2]	CNN-LSTM	WSN-DS, IoMT	94.4%
[3]	ID-GOPA	WSN-DS	96%
[4]	SGD, GNB	WSN-DS	96%
[5]	CNN	AWID	97%
[6]	ARF, HAT	WSN-DS	97.2%
[7]	SVM, SGD	WSN-DS	98%
Proposed Work	Optimized RF	WSN-DS	99.71%

is predicated on several DL and ML techniques used on benchmark datasets, specifically WSN-DS. On a benchmark incursion dataset, Tan et al. [1] applied a standard RF model and obtained an accuracy of 92.57%, which is much lower than the proposed optimized RF model. Using a CNN-LSTM model on WSN-DS and IoMT datasets, Salmi et al. [2] achieved 94.4% accuracy, suggesting reasonable performance in complicated settings. Ifzarne et al. [3] achieved 96% accuracy using the ID-GOPA anomaly detection methodology on WSN-DS, which is comparable to the GNB and SGD methods introduced by Saleh et al. [4]. On the AWID dataset, the CNN-based model put out by Sadia et al. [5] achieved 97% accuracy, demonstrating advancements in DL-based models. By using ARF and HAT classifiers, Tabbaa et al. [6] were able to further improve, achieving 97.2% accuracy on WSN-DS. One of the best methods was used by Ahmed et al. [7], who used SVM and SGD on WSN-DS and achieved 98% accuracy. With an accuracy of 99.71% on WSN-DS, the proposed optimized RF model performs noticeably better than any of these methods, proving its higher effectiveness in identifying attacks with few errors as shown in Figure 5. This demonstrates how well the proposed strategy works to increase security in the WSN.

6 Conclusion

Using the WSN-DS dataset, this study examines how well different ML and DL models identify intrusions in WSNs. To guarantee network dependability and data integrity, WSNs must develop strong and effective IDS in response to growing security threats. Key performance metrics like accuracy, precision, recall, F1-score, and ROC-AUC are used to evaluate several

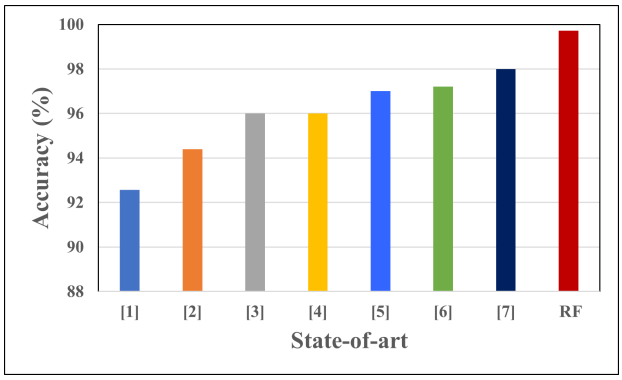


Figure 5. Comparative analysis of proposed optimized RF model with state-of-art approaches.

models, including LR, DT, RF, SVM, KNN, CNN and NN. Based on experimental data, the optimized RF model outperforms other state-of-the-art methods with a maximum detection accuracy of 99.71%. The efficacy of the method is demonstrated by its superior performance in differentiating between benign and malevolent actions as well as its computational efficiency. A comparison with existing techniques shows that WSN security has significantly improved. By offering a thorough performance comparison of ML and DL-based IDS models for WSNs and demonstrating the advantages of improving learning techniques for attack detection, the methodology the proposed methodology contributes to the field. There are still some issues, such as the need for real-time detection in settings with limited resources and the defence against hostile attacks that try to circumvent IDS systems. To further improve detection accuracy, future developments might investigate hybrid ML and DL models. A more secure and robust WSN could result from the integration of federated learning and edge computing approaches, which could enhance the scalability and real-time capabilities of IDS in WSNs.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Tan, X., Su, S., Huang, Z., Guo, X., Zuo, Z., Sun, X., & Li, L. (2019). Wireless sensor networks intrusion detection based on SMOTE and the random forest algorithm. *Sensors, 19*(1), 203. [CrossRef]

[2] Salmi, S., & Oughdir, L. (2022). CNN-LSTM based approach for DOS attacks detection in wireless sensor networks. *International Journal of Advanced Computer Science and Applications, 13*(4). [CrossRef]

[3] Ifzarne, S., Tabbaa, H., Hafidi, I., & Lamghari, N. (2021). Anomaly detection using machine learning techniques in wireless sensor networks. *Journal of Physics: Conference Series, 1743*(1), 012021. [CrossRef]

[4] Saleh, H. M., Marouane, H., & Fakhfakh, A. (2024). Stochastic gradient descent intrusions detection for wireless sensor network attack detection system using machine learning. *IEEE Access, 12*, 3825-3836. [CrossRef]

[5] Sadia, H., Farhan, S., Haq, Y. U., Sana, R., Mahmood, T., Bahaj, S. A. O., & Rehman, A. (2024). Intrusion Detection System for Wireless Sensor Networks: A Machine Learning based Approach. *IEEE Access*. [CrossRef]

[6] Tabbaa, H., Ifzarne, S., & Hafidi, I. (2023). An Online Ensemble Learning Model for Detecting Attacks in Wireless Sensor Networks. *Computing and Informatics, 42*(4), 1013-1036. [CrossRef]

[7] Ahmed, O. (2024). Enhancing Intrusion Detection in Wireless Sensor Networks through Machine Learning Techniques and Context Awareness Integration. *International Journal of Mathematics, Statistics, and Computer Science, 2*, 244–258. [CrossRef]

[8] Biswas, P., & Samanta, T. (2021). Anomaly detection using ensemble random forest in wireless sensor network. *International Journal of Information Technology, 13*(5), 2043–2052. [CrossRef]

[9] Ghadi, Y. Y., Mazhar, T., Al Shloul, T., Shahzad, T., Salaria, U. A., Ahmed, A., & Hamam, H. (2024). Machine learning solutions for the security of wireless sensor networks: A review. *IEEE Access, 12*, 12699–12719. [CrossRef]

[10] Gowdhaman, V., & Dhanapal, R. (2022). An intrusion detection system for wireless sensor networks using deep neural network. *Soft Computing, 26*(23), 13059–13067. [CrossRef]

[11] Haque, A., Chowdhury, N. U. R., Soliman, H., Hossen, M. S., Fatima, T., & Ahmed, I. (2023). Wireless sensor networks anomaly detection using machine learning: a survey. *Intelligent Systems Conference*, 491–506. [CrossRef]

[12] Jeevaraj, D. (2023). Feature selection model using naive bayes ML algorithm for WSN intrusion detection system. *International Journal of Electrical and Computer Engineering Systems, 14*(2), 179–185. [CrossRef]

[13] Karthikeyan, M., Manimegalai, D., & RajaGopal, K.

(2024). Firefly algorithm based WSN-IoT security enhancement with machine learning for intrusion detection. *Scientific Reports*, *14*(1), 231. [CrossRef]

[14] Quincozes, S. E., Kazienko, J. F., & Quincozes, V. E. (2023). An extended evaluation on machine learning techniques for denial-of-service detection in wireless sensor networks. *Internet of Things*, *22*, 100684. [CrossRef]

[15] Raveendranadh, B., & Tamilselvan, S. (2023). An accurate attack detection framework based on exponential polynomial kernel-centered deep neural networks in the wireless sensor network. *Transactions on Emerging Telecommunications Technologies*, *34*(3), e4726. [CrossRef]

[16] Sakthimohan, M., Deny, J., & Rani, G. E. (2024). Secure deep learning-based energy efficient routing with intrusion detection system for wireless sensor networks. *Journal of Intelligent & Fuzzy Systems*, *46*(4), 8587-8603. [CrossRef]

[17] Sivagaminathan, V., Sharma, M., & Henge, S. K. (2023). Intrusion detection systems for wireless sensor networks using computational intelligence techniques. *Cybersecurity*, *6*(1), 27. [CrossRef]

[18] Subbiah, S., Anbananthen, K. S. M., Thangaraj, S., Kannan, S., & Chelliah, D. (2022). Intrusion detection technique in wireless sensor network using grid search random forest with Boruta feature selection algorithm. *Journal of Communications and Networks*, *24*(2), 264–273. [CrossRef]

[19] Bukhari, S. M. S., Zafar, M. H., Abou Houran, M., Moosavi, S. K. R., Mansoor, M., Muaaz, M., & Sanfilippo, F. (2024). Secure and privacy-preserving intrusion detection in wireless sensor networks: Federated learning with SCNN-Bi-LSTM for enhanced reliability. *Ad Hoc Networks*, *155*, 103407. [CrossRef]

[20] Aljebreen, M., Alohali, M. A., Saeed, M. K., Mohsen, H., Al Duhayyim, M., Abdelmageed, A. A., Drar, S., & Abdelbagi, S. (2023). Binary chimp optimization algorithm with ML based intrusion detection for secure IoT-assisted wireless sensor networks. *Sensors*, *23*(8), 4073. [CrossRef]



Anshika Sharma is pursuing her PhD in Computer Science at Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India. She received an M.E. degree in Data Science from Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India in 2024 and an M.Sc. degree in Mathematics from Banasthali Vidyapith in 2022. Her main areas of interest and research are Cybersecurity, Machine Learning and the Internet of Things. (Email: anshika.sharma@chitkara.edu.in)



Dr. Shalli Rani (Director, Research) completed her Post-doc from Manchester Metropolitan University, UK in June 2023. She is a Professor at Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab, India. She has 18+ years of teaching experience. She received an MCA degree from Maharishi Dayanand University, Rohtak in 2004 and an M.Tech. degree in Computer Science from Janardan Rai Nagar Vidyapeeth University, Udaipur in 2007 and a PhD degree in Computer Applications from Punjab Technical University, Jalandhar in 2017. Her main areas of interest and research are Wireless Sensor Networks, Underwater Sensor Networks, Machine Learning and Internet of Things. She has published/accepted/presented more than 100+ papers in international journals /conferences (SCI+Scopus) and edited/authored five books with international publishers. She is serving as the associate editor of IEEE Future Directions Letters. She served as a guest editor in IEEE Transaction on Industrial Informatics, Hindawi WCMC and Elsevier IoT Journals. She has also served as a reviewer in many repudiated journals of IEEE, Springer, Elsevier, IET, Hindawi and Wiley. She has worked on Big Data, Underwater Acoustic Sensors and IoT to show the importance of WSN in IoT applications. She received a Young Scientist award in Feb. 2014 from the Punjab Science Congress, a Lifetime Achievement Award and a Supervisor of the Year award from Global Innovation and Excellence, in 2021. Her work has gained global and reputed recognition, and she has been nominated as one of the top 2% of scientists in her field by Stanford University. (Email: shallir79@gmail.com)