International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 11 | Nov - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Lightweight Machine Learning-Based DOS Attack Detection in Wireless Sensor Networks Using Decision Tree and Information Gain

Iswarya J
Department of CSE
Arunachala College of Engineering for Women
jiswarya2503@gmail.com@gmail.com

Dr. T. V. Chithra,M.E, Ph.D
Associate Professor,Department of CSE
Arunachala College of Engineering for Women
chithratvakul@gmail.com

Abstract—Wireless Sensor Networks (WSNs) are rapidly expanding across various domains due to their unique characteristics and performance capabilities. However, these networks are highly vulnerable to a range of security threats, particularly Denial-of-Service (DoS) attacks, which are among the most common in WSNs. This paper explores the vulnerabilities of WSNs, focusing on DoS threats, and reviews current techniques for their detection. It introduces a lightweight machine learning-based approach using a decision tree (DT) algorithm with the Information Gain (IG) feature selection method for efficient DoS detection. Tested on an enhanced WSN-DS dataset, the proposed method demonstrated high accuracy and minimal processing time compared to other classifiers, such as XGBoost, and RF. This efficiency makes the proposed method well-suited for real-time DoS attack detection in resource-constrained WSNs.

Keywords—Machine Learning,Decision Tree (DT),Information Gain (IG),DoS attack detection

I. INTRODUCTION

Wireless Sensor Networks (WSNs) are a significant area of research in computer science, with widespread applications across various fields, including healthcare, telecommunications, and disaster management. WSNs are particularly useful for monitoring natural disasters such as flooding, earthquakes, and volcanic activity [1]. However, the extensive use of WSNs introduces a range of security risks. These networks are vulnerable to various types of attacks due to limitations like restricted processing power, battery life, and storage capacity. Among these, Denial of Service (DoS) attacks are the most prevalent. DoS attacks disrupt network services by flooding the network with excessive fake requests, which can overwhelm system resources and prevent legitimate traffic from being processed. To mitigate security threats in Wireless Sensor Networks (WSNs), security policies are typically structured around the principles of the Confidentiality, Integrity, and Availability (CIA) triad. While significant efforts have been focused on safeguarding the confidentiality and integrity of data, less emphasis has been placed on addressing threats that undermine the availability of network resources. Attacks targeting resource depletion, such as Denial of Service (DoS), Distributed Denial of Service (DDoS), and jamming attacks, are designed to deplete the limited resources of sensor nodes, disrupting network operations and potentially causing node failures or shutdowns. To protect the network, an Intrusion Detection System (IDS) is essential to detect and mitigate such threats, ensuring the security and reliability of the WSN.

Although various research approaches have been proposed to address security challenges in Wireless Sensor Networks (WSNs), developing an effective solution remains a complex task. One promising area of focus is the integration of blockchain technology to improve security, particularly in the context of authentication protocols. Dener et al. [3] introduced a blockchain-based authentication protocol for WSNs. However, the implementation of blockchain in these networks faces several challenges, such as high processing demands, storage requirements, and increased power consumption. Blockchain technology typically requires substantial computational resources and energy, which are limited in the low-capacity nodes of WSNs. Additionally, significant attention has been directed towards device-free localization (DFL) technology, which has become more feasible for WSNs due to advancements in wireless sensing technologies.

Machine learning (ML) techniques have emerged as a powerful tool to enhance the performance of Intrusion Detection Systems (IDS) in identifying and detecting attackers. In particular, ML classification methods have been applied to detect Denial of Service (DoS) attacks in Wireless Sensor Networks (WSNs). This study explores several classification models, including Naïve Bayes, Neural Networks, Support Vector Machines, Decision Trees, and

Random Forests, to determine which model provides the best classification performance. These techniques are evaluated based on various metrics, such as accuracy, precision, and recall, using a specialized WSN dataset [2], WSN-DS, which includes both normal and attack scenarios. The aim is to assess the effectiveness of these models in accurately detecting DoS attacks within WSNs

Recent studies, such as those by Atitallah et al. [5], explored deep learning techniques, combining convolutional neural networks (CNNs) to detect DoS attacks. While this approach showed good accuracy, it is computationally expensive and may not be suitable for resource-constrained WSNs. Similarly, Vinayakumar et al. [4] employed a deep neural network (DNN) classifier, achieving high accuracy on several datasets, but without addressing the significant computational cost involved. A comparison by Otoum et al. [6] between deep learning and machine learning-based intrusion detection systems revealed that while both approaches had similar detection accuracy, deep learning models require significantly more computational power and time, making them less practical for real-time WSN applications.

This paper offers a thorough examination of the constraints, vulnerabilities, and attack classifications in Wireless Sensor Networks (WSNs). It explores recent approaches to counter Denial of Service (DoS) attacks, with the goal of developing a lightweight detection method that aligns with the unique constraints of WSNs. The key contributions of this study are as follows:

- A comprehensive analysis of WSN constraints, vulnerabilities, and attack classifications.
- Investigation of recent methods to detect and mitigate DoS attacks in WSNs.
- Development of a lightweight detection solution for DoS attacks, optimized to meet the resource limitations of WSNs.
- Comparing the proposed detection approach with several recently introduced methods for countering DoS attacks. The results of this comparison demonstrate that the proposed approach outperforms other classifiers in terms of effectiveness and efficiency.

II. LITERATURE SURVEY

Research in Wireless Sensor Networks (WSNs) focuses on lightweight machine learning (ML) models to detect Denial-of-Service (DoS) attacks, given the resource constraints of sensor nodes. Traditional models like decision trees, support vector machines (SVM), and k-nearest neighbors (k-NN) are commonly used, though more complex models, such as deep learning, can be too resource-intensive. Lightweight approaches, including simplified decision trees, random forests, and feature selection methods, are preferred for balancing accuracy and computational efficiency in WSNs.

Mohammed Faris' [7] paper reviews the security challenges in WSNs, which are critical for applications like surveillance and healthcare. It highlights WSN limitations, such as low memory and processing power, and categorizes security threats while presenting algorithmic solutions and a framework for intrusion detection systems (IDS). Shahzad Ashraf's [8] paper introduces the Bodacious-instance Coverage Mechanism (BiCM), which improves WSN coverage by redeploying sensor nodes, outperforming other algorithms in coverage range and performance.

Salim Salmi's [9] work tackles security challenges in WSNs, particularly DoS attacks. Due to resource constraints, traditional IDS are often inadequate for evolving attacks. Salmi proposes a deep learning-based IDS trained on a specialized dataset (WSN-DS) to detect multiple DoS attack types. Shereen Ismail's [10] paper examines cybersecurity in WSNs and the Internet of Things (IoT), focusing on machine learning (ML) and blockchain (BC) techniques for enhanced security. She proposes a lightweight framework integrating ML and BC for attack detection and prevention.

Virendra Dani's [11] paper addresses vulnerabilities in wireless ad hoc networks, particularly regarding routing and DoS attacks. He proposes a Trust-Aware Routing Approach to mitigate bogus route requests in the Ad hoc On Demand Vector (AODV) protocol. The method improves network security and performance in simulations. Murat Dener's [6] paper discusses security in WSNs, particularly the limitations of existing authentication protocols. Dener introduces a blockchain-based authentication model, using private blockchain to enhance data and node security.

Muawia A. Elsadig's [12] paper reviews the security challenges in WSNs, which are vital for military, environmental, and healthcare applications. The paper discusses the constraints, vulnerabilities, and common security

threats in WSNs, and emphasizes the need for improved solutions. Mamoona Majid's [13] work highlights the role of WSNs and IoT in Industry 4.0, addressing research gaps in security, deployment, and design of networks. A systematic review of over 130 articles is presented, discussing the challenges and future directions for Industry 4.0 automation.

Van-Linh Nguyen's [14] paper reviews energy depletion attacks (EDAs) in Low Power Wireless (LPW) networks, critical for IoT connectivity. These networks are vulnerable to EDAs that exploit communication flaws, leading to battery drain and network failure. The paper reviews existing defense mechanisms and suggests future research directions to strengthen LPW protocol security.

Kamran Shaukat's [15] paper provides a comprehensive review of machine learning (ML) techniques applied to cybersecurity, focusing on detecting various cyber threats such as fraud, intrusion, spam, and malware. With the increasing reliance on the internet and rising cybersecurity risks, the paper explores how ML models can help address these challenges. It offers a detailed comparison of commonly used ML models, analyzing their performance based on datasets and types of cyber threats. The review also discusses the time complexity of these models, highlighting their strengths and limitations in cybersecurity applications. Additionally, the paper addresses the current challenges and constraints in applying ML to cybersecurity, offering insights into areas for further improvement.

III. Proposed Methodology

Security in Wireless Sensor Networks (WSNs) requires meeting several key parameters, including integrity, availability, confidentiality, and authentication, while minimizing computational overhead due to resource constraints. WSNs are highly vulnerable to various security attacks, categorized in different ways, including internal vs. external and layer-based classifications. Internal attacks originate within the network, while external attacks come from outside sources. Layer-based classifications, based on the OSI model, show that network layer attacks are the most common.

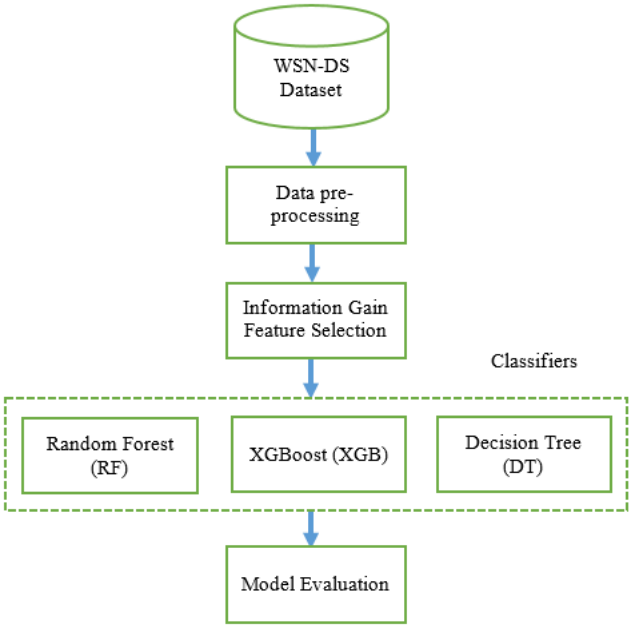

Figure 1 Block Diagram of the proposed work

Denial of Service (DoS) attacks are among the most prevalent threats in WSNs, targeting network availability and preventing legitimate users from accessing network services. These attacks are particularly damaging due to the resource-constrained and distributed nature of WSNs. DoS attacks can occur at various layers and are characterized by network performance degradation, packet loss, unresponsiveness of network components, and increased spam. The objective of DoS attacks is to disrupt normal network operations by exhausting or blocking access to sensor node resources.

Figure 1 shows the Block Diagram of the Proposed Work. The WSN-DS dataset serves as the foundational data for the proposed methodology. The process begins with data preprocessing, which involves cleaning and preparing the dataset for further analysis. Next, the Information Gain feature selection technique is applied to identify the most relevant features, reducing dimensionality and enhancing the model's efficiency. Following this, the data is split into training and testing sets, ensuring a robust evaluation of the model's performance. The training set is used to train the Decision Tree (DT) model, which is then tested using the testing set to assess its accuracy in detecting DoS attacks. Finally, the trained model is used to make predictions, classifying the data into two categories: DoS or Normal. This structured approach enables the effective detection of DoS attacks while

optimizing computational resources in the context of Wireless Sensor Networks (WSNs).

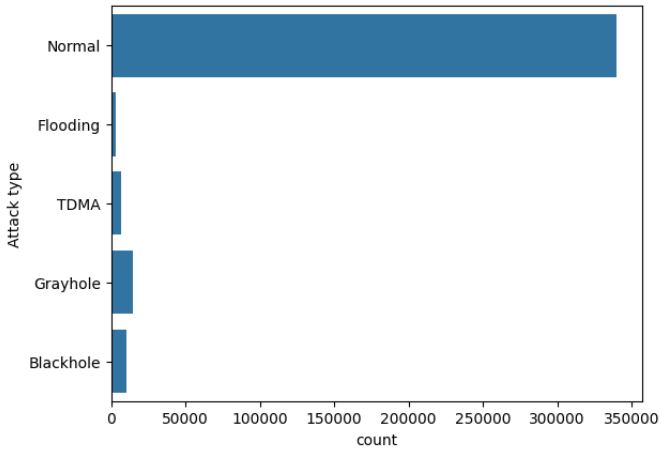

Figure 2. Attack types

A. DATASET

This paper uses the WSN-DS dataset, created by Almomani et al., [2] which includes four types of DoS attacks: blackhole, flooding, grayhole, and scheduling. The dataset, available on Kaggle, contains 18 features, and an enhanced version was developed using feature selection to improve performance. Both versions were used to train and test machine learning models. Preprocessing steps included handling missing values and normalizing feature scales. Model performance was evaluated based on accuracy, precision, recall, and F1-score to assess the effectiveness of different algorithms in detecting DoS attacks in WSNs.

B. PREPROCESSING

The preprocessing steps for the dataset involved two main tasks: handling missing values and normalizing feature scales. First, missing values were addressed using imputation techniques. For continuous features, missing values were replaced with the mean or median of the respective columns, while for categorical features, the most frequent value (mode) was used. In cases where imputation was not feasible due to the extent of missing data, rows with missing values were removed to maintain dataset integrity. This ensured that the dataset was complete and ready for analysis without introducing bias.

Next, feature normalization was performed to ensure that all features contributed equally to the machine learning models. Since the dataset included features with different units and ranges, normalization was necessary to prevent features with larger numerical values from dominating the learning process. Min-Max normalization was applied to scale all features to a common range, typically between 0 and 1, using the formula $\frac{X-X_{min}}{X_{max}-X_{min}}$, where X is the feature value and X_{min} and X_{max} are the minimum and maximum values for that feature. In some cases, Z-score standardization was also applied to ensure features had a mean of 0 and a standard deviation of 1. These preprocessing steps helped standardize the dataset, making it suitable for machine learning models by reducing biases from varying scales and improving classification accuracy

C. FEATURE SELECTION

Feature selection is a critical step in improving the performance of classification models, particularly in environments with limited computational resources, such as Wireless Sensor Networks (WSNs). By selecting only the most informative features, feature selection not only reduces the dimensionality of the data but also improves the accuracy and efficiency of the model. This is especially important in resource-constrained applications where minimizing computational load is crucial for real-time performance. In this study, after evaluating several feature selection methods, the authors opted for Information Gain due to its superior performance in selecting the most relevant features while maintaining model efficiency.

Information Gain is a widely used metric for feature selection in decision trees and other classification algorithms. It measures the reduction in entropy (uncertainty) achieved by partitioning the data based on a given feature. The higher the Information Gain of a feature, the more it contributes to reducing the uncertainty in the classification task. By calculating the Information Gain for each feature, the algorithm can identify the most valuable features for building an efficient classification model. This process helps eliminate irrelevant or redundant features, which are less informative, thus improving model performance.

The Information Gain for a feature is calculated as the difference between the entropy of the dataset before and after the split based on that feature. In mathematical terms, Information Gain (IG) for a feature AAA is given by:

$$IG(D,A) = H(D) - \sum_{v \in Values(A)} \frac{|D_v|}{|D|} H(D_v) \qquad (1)$$

Where:

- $H(D)$ is the entropy of the dataset D,

- $Values(A)$are the possible values that feature A can take,
- D_v is the subset of data for which feature A has value v,
- $|D|$ and $|D_v|$ are the sizes of the dataset and subset, respectively,
- $H(D_v)$is the entropy of the subset D_v

The Entropy$H(D)$of a dataset D is defined as:

$$H(D) = -\sum_{i=1}^{c} p_i log_2(p_i) \quad (2)$$

Where:

- c is the number of classes,
- $p_i$is the proportion of samples in class i within the dataset.

In this work, after applying Information Gain as the feature selection method, the authors successfully reduced the number of features, maintaining only the most relevant attributes. This process not only enhanced the model's prediction accuracy but also decreased the training time and computational complexity. By retaining only the features that contribute the most to class separation, the system becomes more efficient, reducing the risk of overfitting and improving its ability to generalize to new, unseen data. This is especially important for Denial-of-Service (DoS) attack detection and other real-time applications in dynamic WSN environments, where quick decision-making is critical

D. Decision Tree (DT)

A Decision Tree (DT) is a widely used supervised machine learning algorithm for classification and regression tasks, known for its simplicity and interpretability. It works by recursively splitting the dataset into subsets based on feature values, with the goal of creating a tree-like structure where each node represents a decision rule and each leaf node represents the final output or classification label. At the root, the tree selects the feature that provides the best separation of the data according to Information Gain. The process of splitting continues recursively, with each internal node further dividing the data based on the next most informative feature, until certain stopping conditions are met, such as reaching a predefined tree depth or achieving a node that contains only data points from a single class. The **leaf nodes** represent the final prediction: for classification, it is the majority class, and for regression, it is the average or median of the values in that leaf.

Decision Trees offer several advantages, including their ease of understanding and visualization, as the decision-making process is explicitly laid out in the tree structure. They also do not require feature scaling, making them computationally efficient, and can handle both categorical and numerical data.

E. VALIDATION

The classification accuracy for each model was calculated using the formula:

$$Accuracy = \frac{TP + TN}{FN + TP + FP + TN} \quad (2)$$

Where TP (True Positive), TN (True Negative), FP (False Positive), and **FN** (False Negative) represent the respective counts of correctly and incorrectly predicted cases. A confusion matrix was used to evaluate model performance, with additional metrics such as precision, recall, and F1-score calculated from the following formulas:

$$Recall = \frac{TP}{FN + TP} \quad (3)$$

$$Precision = \frac{TP}{FP + TP} \quad (4)$$

$$F1score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Furthermore, ROC curves were plotted to provide a more comprehensive assessment of each classifier's performance.

IV. Results and Discussion

To assess the effectiveness of the enhanced dataset, all classifiers (XGBoost, DT and RF) were trained and tested using both the original and the enhanced versions of the WSN-DS dataset, which was improved using the information gain feature selection method. The results, shown in Figure 3, demonstrate that the accuracy of all classifiers on enhanced datasets. However, the enhanced version reduced computational time, making it more suitable for resource-constrained WSNs. This reduction in overhead is a significant improvement, and based on the results, the authors recommend using the enhanced dataset with information gain feature selection, as it did not negatively impact accuracy unlike other feature selection methods tested.

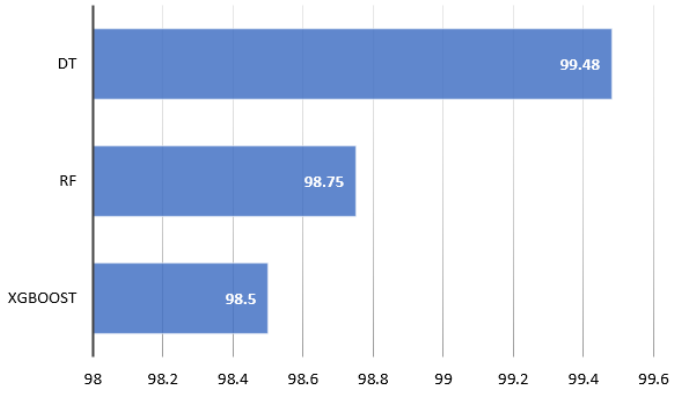

FIGURE 3. The accuracy of all classifiers

Table 1. Performance Metrics

Methods	Accuracy	Precision	Recall	F1 Score
XGBoost	98.5	98.5	98.5	98.5
RF	98.75	98.81	98.75	98.77
DT	99.48	99.48	99.48	99.48

The results from the evaluation of the three machine learning models—XGBoost, Random Forest (RF), and Decision Tree (DT)—demonstrate strong performance in detecting Denial of Service (DoS) attacks in Wireless Sensor Networks (WSNs). All models achieved high accuracy, with the Decision Tree (DT) model leading at 99.48%, followed closely by Random Forest (RF) at 98.75% and XGBoost at 98.5%. Despite the small difference in accuracy, it is evident that all three models perform well in correctly classifying both DoS and normal instances. In terms of precision, RF slightly outperformed the other models with 98.81%, while DT and XGBoost both achieved 98.5%, indicating that all models are effective at minimizing false positives. When assessing recall, which measures the ability to correctly identify DoS attacks, the models again performed excellently, with DT leading at 99.48%, followed by RF and XGBoost with 98.75% and 98.5%, respectively. This suggests that the models are highly efficient at detecting DoS attacks and have minimal false negatives. The F1 scores, which balance precision and recall, were also highest for the DT model at 99.48%, while RF and XGBoost achieved 98.77% and 98.5%, respectively. These results suggest that the DT model provides the best trade-off between precision and recall, offering a more robust solution for DoS attack detection in WSNs.

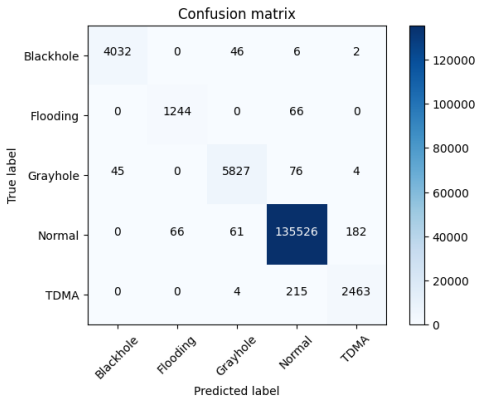

Figure 4. Confusion Matrix of DT

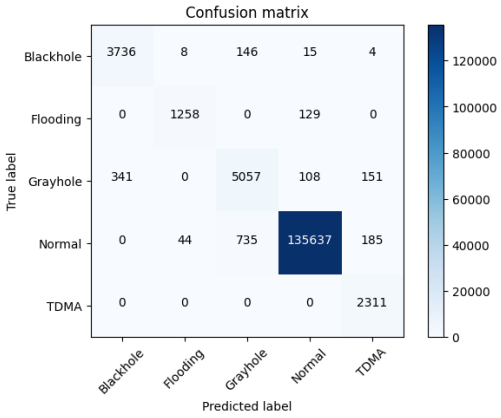

Figure 5. Confusion Matrix of RF

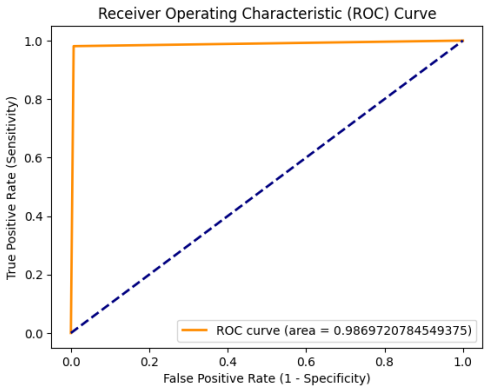

Figure 6. ROC Curve of DT

The confusion matrices for both the Decision Tree (DT) and Random Forest (RF) models, shown in Figures 4 and 5, provide further insight into the models' performance in detecting DoS attacks in Wireless Sensor Networks (WSNs). Specifically, the DT model shows an almost perfect classification with no false positives or negatives, further

confirming its exceptional accuracy. The RF model also demonstrates strong performance, with only a slight increase in false positives and false negatives compared to DT. Overall, the confusion matrices reinforce the high classification accuracy of both models, highlighting their effectiveness in detecting DoS attacks while maintaining low error rates.

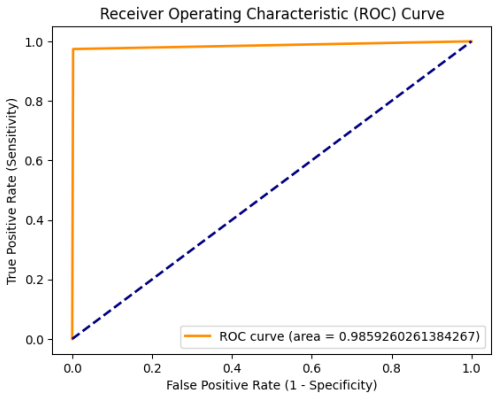

Figure 7. ROC Curve of RF

The Receiver Operating Characteristic (ROC) curves for both the Decision Tree (DT) and Random Forest (RF) models, shown in Figures 6 and 7, For the DT model (Figure 6), the ROC curve shows a high Area Under the Curve (AUC) of 0.986, indicating excellent discrimination between DoS attacks and normal instances. A higher AUC implies that the DT model is very effective at distinguishing between the two classes, with fewer misclassifications. Similarly, the RF model (Figure 7) has a very close AUC of 0.9859, demonstrating nearly identical performance in terms of separating attack and normal traffic.The slight difference in AUC between the two models is negligible, indicating that both the DT and RF models are highly reliable for detecting DoS attacks, with the RF model offering a marginally better performance

V. Conclusion

This paper proposes a lightweight solution for detecting DoS attacks using a Decision Tree (DT) classifier enhanced with the information gain feature-selection method. The method was tested on an improved version of the WSN-DS dataset, achieving a high accuracy rate of 99.48%. It outperformed other classifiers, including Random Forest (RF) and Extreme Gradient Boosting (XGBoost), while significantly reducing processing time. This approach used only 9.7%, 13%, and 2% of the processing time required by RF, XGBoost, and DT, respectively, making it highly efficient for resource-constrained WSNs. However, the study was conducted on a single dataset, and future work will focus on evaluating the method across different datasets and addressing the dataset's imbalance to improve accuracy further.

References

[1] N. A. A. Aziz and K. A. Aziz, “Managing disaster with wireless sensor networks,” in International Conference on Advanced Communication Technology, ICACT, 2011, pp. 202–207.
[2] I. Almomani, B. Al-Kasasbeh, and M. Al-Akhras, “WSN-DS: A Dataset for Intrusion Detection Systems in Wireless Sensor Networks,” J. Sensors, vol. 2016, 2016, doi: 10.1155/2016/4731953.
[3] M. Dener and A. Orman, ‘‘BBAP-WSN: A new blockchain-based authentication protocol for wireless sensor networks,’’ Appl. Sci., vol. 13, no. 3, p. 1526, Jan. 2023, doi: 10.3390/app13031526
[4] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, ‘‘Deep learning approach for intelligent intrusion detection system,’’ IEEE Access, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/ACCESS.2019.2895334.
[5] S. B. Atitallah, M. Driss, W. Boulila, and I. Almomani, ‘‘An effective detection and classification approach for DoS attacks in wireless sensor networks using deep transfer learning models and majority voting,’’ in Advances in Computational Collective Intelligence, vol. 1653
[6] S. Otoum, B. Kantarci, and H. T. Mouftah, ‘‘On the feasibility of deep learning in sensor network intrusion detection,’’ IEEE Netw. Lett., vol. 1, no. 2, pp. 68–71, Jun. 2019, doi: 10.1109/LNET.2019.2901792.
[7] Faris, M., Mahmud, M. N., Salleh, M. F. M., & Alnoor, A. (2023). Wireless sensor network security: A recent review based on state-of-the-art works. In *International Journal of Engineering Business Management* (Vol. 15). SAGE Publications Inc. https://doi.org/10.1177/18479790231157220
[8] Ashraf, S., Alfandi, O., Ahmad, A., Khattak, A. M., Hayat, B., Kim, K. H., & Ullah, A. (2020). Bodacious-instance coverage mechanism for wireless sensor network. *Wireless Communications and Mobile Computing*, *2020*. https://doi.org/10.1155/2020/8833767
[8] Salmi, S., & Oughdir, L. (2023). Performance evaluation of deep learning techniques for DoS attacks detection in wireless sensor network. *Journal of Big Data*, *10*(1). https://doi.org/10.1186/s40537-023-00692-w
[10] Ismail, S., Dawoud, D. W., & Reza, H. (2023). Securing Wireless Sensor Networks Using Machine Learning and Blockchain: A Review. In Future Internet (Vol. 15, Issue 6). MDPI. https://doi.org/10.3390/fi15060200
[11] Dani, V., Vaishnav, S., & Vishwavidyalaya, V. (2023). Detection of Denial-of-Service Attack Using Weight based Trust Aware Routing Approach. In Journal of Information Assurance and Security (Vol. 18).
[12] Dener, M., & Orman, A. (2023). BBAP-WSN: A New Blockchain-Based Authentication Protocol for Wireless

Sensor Networks. *Applied Sciences (Switzerland)*, *13*(3). https://doi.org/10.3390/app13031526

[13] Elsadig, M. A., Altigani, A., & Baraka, M. A. A. (2019). Security issues and challenges on wireless sensor networks. *International Journal of Advanced Trends in Computer ScienceandEngineering*,*8*(4),1551–1559. https://doi.org/10.30534/ijatcse/2019/78842019

[14]Majid, M., Habib, S., Javed, A. R., Rizwan, M., Srivastava, G., Gadekallu, T. R., & Lin, J. C. W. (2022). Applications of Wireless Sensor Networks and Internet of Things Frameworks in the Industry Revolution 4.0: A Systematic Literature Review. *Sensors*, *22*(6). https://doi.org/10.3390/s22062087

[15] Nguyen, V. L., Lin, P. C., & Hwang, R. H. (2019). Energy depletion attacks in low power wireless networks. *IEEE Access*,51915–51932. https://doi.org/10.1109/ACCESS.2019.2911424

[16] Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., Chen, S., Liu, D., & Li, J. (2020). Performance comparison and current challenges of using machine learning techniques in cybersecurity. In *Energies* (Vol. 13, Issue 10). MDPI AG. https://doi.org/10.3390/en13102509