WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Detection of Attacks in Wireless Sensor Networks Using Machine Learning Algorithms

Dr. B. Karthikeyan[1], MS. M. Kamali [2]

[1]Assistant Professor (SG), Department of Computer Science, Bishop Heber College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli - 621 017, Tamil Nadu, India, bkarthikeyanphd@gmail.com

[2]Department of Computer Science, Bishop Heber College (Autonomous), Affiliated to Bharathidasan University, Tiruchirappalli - 621 017, Tamil Nadu, India, kamalimuruganantham171@gmail.com

ABSTRACT—

In today's digital age, the widespread use of internet-based services, including social media, online transactions, and room bookings, has heightened the risk of cyber-attacks like Denial of Service (DoS), Distributed Denial of Service (DDoS), Man-in-the-Middle, and SQL Injection. Wireless Sensor Networks (WSNs), which rely on interconnected sensor nodes to transmit environmental data wirelessly, are particularly vulnerable to these threats. This project aims to enhance intrusion detection in WSNs by analyzing network traffic and classifying cyber-attacks using advanced machine learning techniques, specifically Random Forest and XGBoost algorithms. The dataset is categorized into three main attack types: DoS, Probe, and R2L (Remote to Local). Unlike existing methods that use Multi-Layer Perceptron (MLP), Decision Tree (DT), and Support Vector Machines (SVM), this study leverages the superior performance of Random Forest and XGBoost to improve classification accuracy and reduce detection time. The models' effectiveness will be assessed through graphical representations and confusion matrices, offering a detailed performance comparison. The proposed approach aims to strengthen the security and resilience of WSNs against emerging cyber threats.

Keywords—Random Forest, XGBoost, Attack Detection, Classification Accuracy, Confusion Matrix.

I. INTRODUCTION

The Wireless Sensor Networks (WSNs) consist of multiple Sensor Nodes (SNs) that wirelessly collect and transmit environmental data. These nodes are compact, low-powered devices with sensing, processing, and communication units, widely used in applications such as industrial monitoring, healthcare, agriculture, and smart cities. Despite their versatility, WSNs face security challenges due to limited power, memory, and computational resources, making them vulnerable to failures and cyber-attacks. Security concerns are heightened by factors like limited transmission range, wireless communication mediums, and decentralized deployment. They are particularly susceptible to Denial of Service (DoS), Probe, and Remote to Local (R2L) attacks. Traditional security methods using Multi-Layer Perceptron (MLP), Decision Tree (DT), and Support Vector Machines (SVM) have limitations in accuracy and efficiency. This paper proposes enhancing WSN security by leveraging advanced Machine Learning (ML) algorithms, specifically Random Forest (RF) and XGBoost, known for their superior classification accuracy and efficiency. The study aims to detect and classify DoS, Probe, and R2L attacks more effectively than existing models while reducing detection time. Performance evaluation will be conducted using graphical analysis and confusion matrices, ensuring a detailed comparison with traditional methods. The objective is to develop a reliable and efficient intrusion detection system that enhances the security and resilience of WSNs against emerging cyber threats.

Types of Attacks in WSNs

WSNs have limited computational resources compared to other sensory devices, making them cost-effective but vulnerable to security threats. Additionally, their deployment in remote and unattended locations increases their susceptibility to physical tampering and unauthorized access. This study focuses on three primary types of attacks:

1. DoS (Denial of Service) Attack: These attacks disrupt network services by overwhelming the system, preventing legitimate users from accessing the network.

2. Probe Attack: These attacks involve gathering information about the network or computer system, often as a precursor to more severe security breaches.

3. R2L (Remote to Local) Attack: In this type of attack, an external intruder gains unauthorized access to a local system within the network.

II. LITERATURE REVIEW

Neha Jagwani et al.[1] Emphasize how Wireless Sensor Networks (WSNs) are used to monitor environmental factors including motion, temperature, and humidity. In order to improve WSN security, they suggest using Machine Learning (ML) methods and talk about the difficulties presented by security threats such as Denial of Service (DoS), Probe, R2L (Remote to Local), and U2R (User to Root) assaults. Along with comparing different machine learning algorithms using performance metrics including MCC, ROC, precision, recall, and F-1 scores, they also emphasize the use of the SMOTE approach.[2024]

***Priyanka Shah et al.*[2]** Emphasize that Wireless Sensor Networks (WSNs) are popular due to their low cost, power efficiency, and ease of implementation, but they face significant security challenges. These networks are particularly vulnerable to attacks like Sybil, Black-hole, and Denial of Service (DoS). This paper explores detecting these attacks and enhancing security using the AODV protocol. The RSA algorithm is proposed to identify malicious sensor nodes and secure message transmission, thereby strengthening the AODV protocol. The simulations are conducted using the MATLAB tool to evaluate the proposed approach.

***B.J Santhosh Kumar et al.*[3]** Highlight that Internet-integrated Wireless Sensor Networks (WSNs) significantly impact daily life, but they face security challenges, particularly from Denial of Service (DoS) attacks. These attacks are difficult to defend against due to their various forms. Traditional encryption-based methods have proven ineffective in protecting WSNs from DoS attacks. This paper reviews current Intrusion Detection and Prevention systems against flooding and jamming attacks in WSNs and proposes a new approach for more effective detection and prevention of these severe threats.

***Shereen Ismail et al*.[4]** Propose that Wireless Sensor Networks (WSNs) are crucial for low-cost IoT systems but are vulnerable to cyber-attacks. This paper proposes a lightweight multi-layer machine learning detection system to counter internal attacks in WSNs using a mobile robot. The system uses Naive Bayes for first-layer binary classification and LightGBM for second-layer multi-class classification. It detects four network-layer DoS attacks from the WSN-DS dataset. The mobile robot aids by routing updates to the Base Station for deeper investigation when an attack is detected.

***Shalini Swami et al.*[5]** The authors examine the security challenges faced by Wireless Sensor Networks (WSNs), particularly their susceptibility to Denial of Service (DoS) attacks. To address these issues, this paper introduces a fast and efficient anomaly detection system that combines a rule-based approach with Machine Learning (ML) techniques. The rule-based method initially filters normal traffic, activating ML models only when suspicious flows are detected. This layered approach optimizes system efficiency and reduces unnecessary processing. To accurately identify various types of DoS attacks, including TDMA, Blackhole, Grayhole, and Flooding, the system employs Decision Tree (DT) and Support Vector Machine (SVM) classifiers. Among these, DT demonstrates superior accuracy, effectively enhancing network security while preserving system speed and performance.

***Muawia A. Elsadig et al.*[6]** Highlight that Wireless Sensor Networks (WSNs) are widely used but vulnerable to security threats, especially Denial-of-Service (DoS) attacks. This study explores WSN limitations and recent DoS detection methods, highlighting their strengths and weaknesses. It proposes a lightweight machine learning detection approach using a Decision Tree (DT) algorithm with Gini feature selection to detect DoS attacks. The model, trained on an enhanced WSN-DS dataset, achieved 99.5% accuracy with minimal processing time compared to Random Forest, XGBoost, and KNN classifiers. The approach effectively balances high accuracy and low overhead, addressing WSN constraints efficiently.

***Somnath Sinha et al.*[7]** Assert that Security is a major concern in the Internet of Things (IoT), with the RPL protocol being vulnerable to various attacks. One common threat is the Denial-of-Service (DoS) attack, affecting all layers of the IoT architecture. This research examines the impact of selective forwarding attacks by varying the number and position of malicious nodes. It proposes a unique detection mechanism that calculates a detecting factor based on energy consumption and packets received. The approach enhances security by identifying attacks area-wise in the IoT network.

***Sayamuddin Ahmed Jilani et al.*[8]** The authors explain that Wireless Sensor Networks (WSNs) are cost-effective and provide solutions to real-world problems, yet they remain vulnerable to attacks due to the resource limitations of their nodes. Common threats, such as DoS, black hole, and wormhole attacks, can compromise data by either replicating or destroying it. This paper evaluates these security threats and emphasizes the necessity for real-time intrusion detection systems. The authors propose a novel detection algorithm designed to function as a firewall, identifying intrusions in a progressive manner. Additionally, the paper outlines directions for future research to further enhance network security.

***Kosaraju Chaitanya et al.*[9]** Emphasize that Wireless Sensor Networks (WSNs) play a crucial role in data collection but are vulnerable to intrusions due to their disorganized layout. The rise in digital interactions, especially during the Covid-19 pandemic, has increased cyberattacks like Distributed Denial of Service (DDoS) and Distributed Reflective Denial of Service (DRDoS). Traditional detection systems struggle to identify new DDoS attacks due to outdated techniques and the exponential growth of data. This research reviews deep learning models for detecting DDoS and DRDoS attacks, analyzing their working mechanisms and limitations. It highlights the need for advanced detection methods for early attack identification.

III. PROPOSED SYSTEM

The proposed system enhances network security in Wireless Sensor Networks (WSNs) by effectively detecting network intrusions, focusing on DoS (Denial of Service), Probe, and R2L (Remote to Local) attacks. It classifies 37 subtypes of these attacks using advanced machine learning algorithms—Random Forest and XGBoost. In performance evaluation, Random Forest achieved 90.6% accuracy, surpassing XGBoost's 85%. Detection accuracy, precision, recall, and F1 score were analyzed using a confusion matrix. The system addresses the growing threat of network intrusions, which often lead to personal data theft. Unlike existing solutions that struggle with high processing times and delayed detection, this system reduces processing time for

real-time attack detection. It leverages the strengths of both Random Forest and XGBoost to optimize detection accuracy while minimizing false positives and negatives. The Intrusion Detection System (IDS) is designed for robust and fast classification of network attacks with minimal computational overhead, enhancing overall network security. The dataset for intrusion detection was split into 80% for training and 20% for testing, ensuring effective model evaluation and validation.

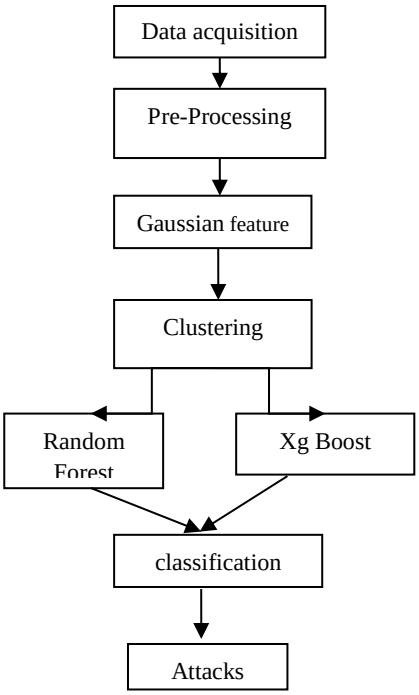

Fig 1: Proposed Architecture

IV. IMPLEMENTATION METHODOLOGY

A. Data Acquisition:

This module involves collecting network traffic data from Wireless Sensor Networks (WSNs) to detect potential security threats. The data is gathered from various sources, including packet captures, network logs, and sensor nodes, ensuring a comprehensive dataset. It focuses on obtaining real-time and historical data to train and test the Intrusion Detection System (IDS). Proper data acquisition is essential for accurately identifying attack patterns and normal network behavior. In this system, the WSN-DS dataset is utilized to provide diverse and reliable data for effective model training.

B. Data Preprocessing:

Data preprocessing is crucial for cleaning and organizing the acquired data, ensuring accuracy and efficiency in attack detection. It involves removing noise, handling missing values, and normalizing data to a consistent format. This step also includes feature selection and dimensionality reduction, enhancing computational efficiency. Proper preprocessing reduces biases and improves the performance of machine learning models. By transforming raw data into a structured format, the system achieves better detection accuracy and reduced processing time.

C. Gaussian Feature Extraction:

This module extracts relevant features from the preprocessed data using Gaussian distribution techniques. It identifies patterns and correlations in network traffic data to differentiate between normal and malicious behavior. Gaussian feature extraction enhances the model's ability to recognize subtle variations in attack patterns. This step ensures that only the most significant features are used, reducing complexity and improving detection accuracy. It also optimizes computational resources by minimizing redundant data inputs.

D. Random Forest and XGBoost:

This module uses Random Forest and XGBoost algorithms for attack classification. Random Forest is chosen for its high accuracy and robustness in handling imbalanced datasets, achieving a detection accuracy of 90.6%. XGBoost (Extreme Gradient Boosting) is utilized for its fast processing speed and effectiveness in handling complex data patterns, achieving 84% accuracy. The combination of these algorithms enhances prediction accuracy while minimizing false positives and negatives. Performance is evaluated using metrics such as precision, recall, and F1 score. This approach provides a balanced trade-off between speed and accuracy in real-time intrusion detection.

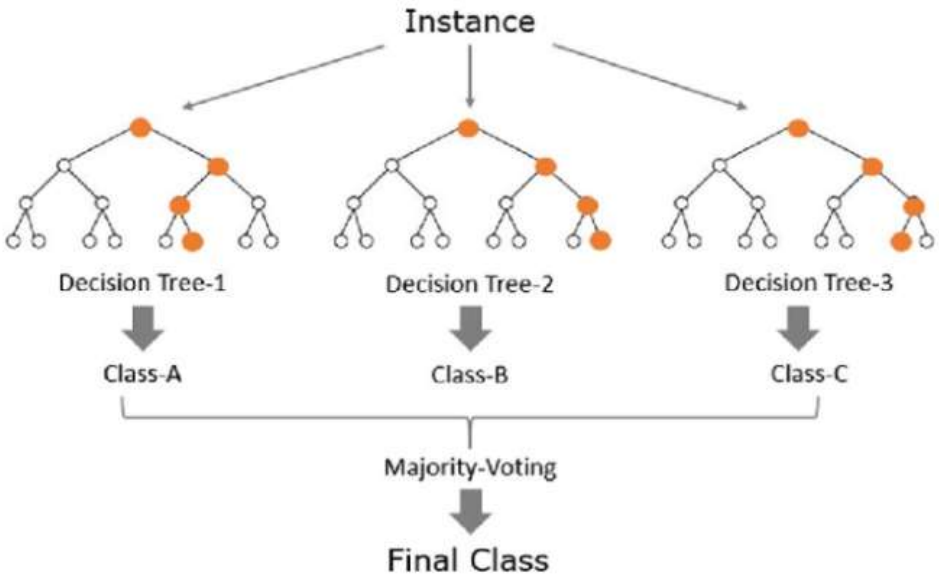

Fig 2 Random Forest

E. ***Attacks Prediction:***

In this module, the trained models predict different types of attacks, including DoS (Denial of Service), Probe, and R2L (Remote to Local) attacks. By analyzing the extracted features, the system identifies and classifies 37 subdivisions of these attacks. Real-time prediction enables immediate response to potential threats, enhancing network security. The system compares the performance of Random Forest and XGBoost in terms of prediction accuracy and efficiency. This module ensures robust detection with minimal computational overhead, optimizing network safety against evolving cyber threats.

F. ***Confusion Matrix****:*

A confusion matrix is used to evaluate the performance of a classification model by comparing the actual and predicted outputs. It shows how many predictions were correct and where errors occurred, providing a clear picture of the model's accuracy. The matrix includes four components: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). TP and TN indicate correct predictions, while FP and FN show where the model made mistakes. By analyzing these components, the confusion matrix helps measure accuracy, precision, recall, and F1 score, offering valuable insights into the model's performance.

TP-True Positive: The predicted output is positive, and the actual output is also positive.

TN-True Negative: The predicted output is negative, and the actual output is also negative.

FP-False Positive: The predicted output is positive, but the actual output is negative.

FN-False Negative: The predicted output is negative, but the actual output is positive.

1) ***Accuracy****:* Accuracy measures the proportion of correct predictions out of the total predictions made by the model. It is useful for getting a general idea of model performance but can be misleading with imbalanced datasets.

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

2) ***Precision***: Precision calculates how many of the predicted positive instances are actually positive. It is crucial when minimizing false positives is important, like in spam detection.

$$\boldsymbol{Precision} = \frac{Tp}{TP + FP}$$

3) ***Recall****:* Recall measures how well the model identifies all actual positive cases. It is important when it is critical not to miss positive cases, such as in medical diagnostics.

$$\boldsymbol{Recall} = \frac{TP}{TP + FN}$$

4) ***F1-Score***: F1-Score balances precision and recall by calculating their harmonic mean. It is useful when both false positives and false negatives are crucial, especially with imbalanced data.

$$\boldsymbol{F1-Score} = \frac{2.Precision.Recall}{Precision + Recall}$$

5) ***Support***: Support shows how many instances of a particular class there are in the dataset. It aids in the comprehension of class distribution and is especially crucial when working with datasets that are unbalanced.

Support = total number of real instances of the class in the dataset

V. RESULT AND DISCUSSION

1. Confusion Matrix Analysis:

The confusion matrix evaluates the performance of XGBoost and Random Forest by showing True Positives, False Positives, False Negatives, and True Negatives. Both algorithms demonstrate high detection rates with few false negatives, ensuring accurate attack identification. Random Forest maintains a better balance between precision and recall, reducing false positives while being highly sensitive to attacks.

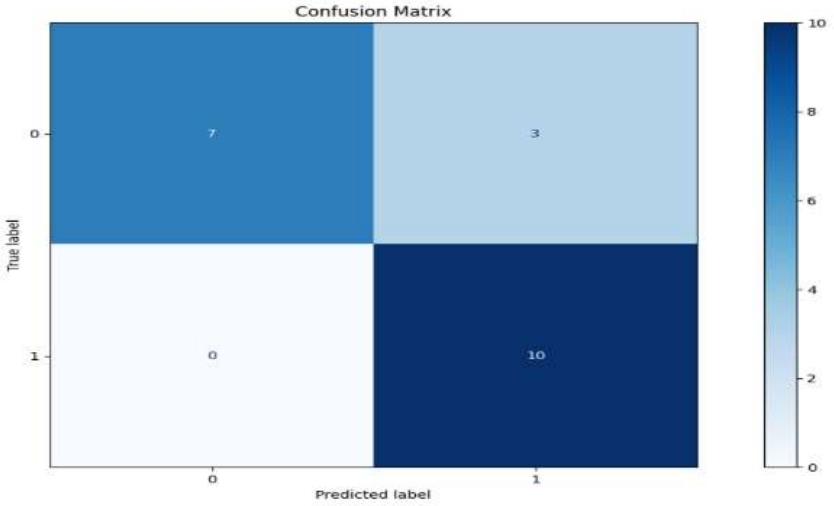

Fig 3: Confusion Matrix Analysis

2. Comparison Bar Graph:

The bar graph compares the performance metrics of Random Forest and XGBoost, including accuracy, precision, recall, and F1 score. Random Forest outperforms XGBoost in accuracy and recall, indicating better generalization across attack types. XGBoost shows slightly higher precision but at the expense of missing some attack instances.

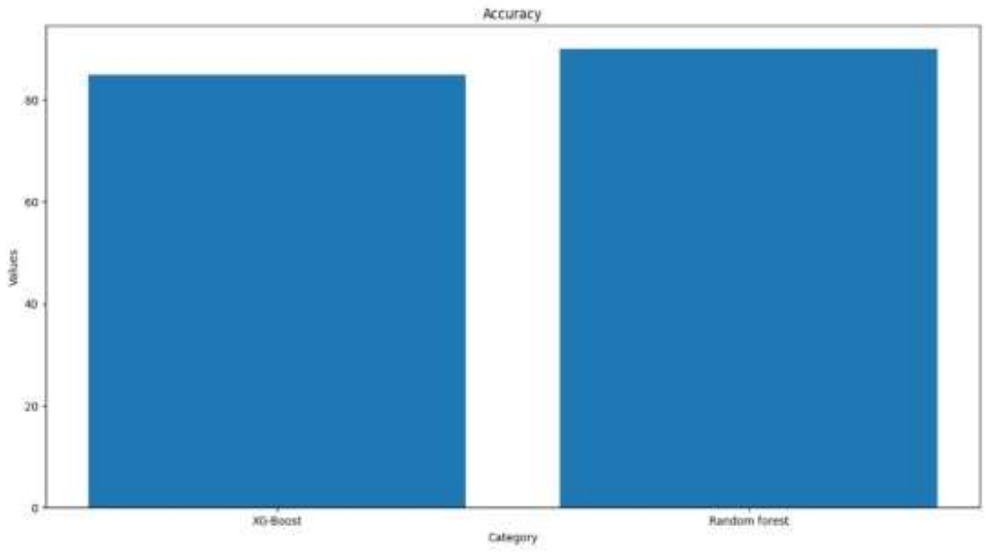

Fig 4: Comparison Bar Graph

3. ***Line Graph Comparison with Existing Algorithms:***

This graph compares five algorithms MLP, Decision Tree, SVM, Random Forest, and XGBoosthighlighting the superiority of the proposed models. Random Forest consistently performs well across all metrics, while XGBoost achieves competitive accuracy with slightly lower recall. This demonstrates the effectiveness of advanced ensemble techniques in enhancing attack detection.

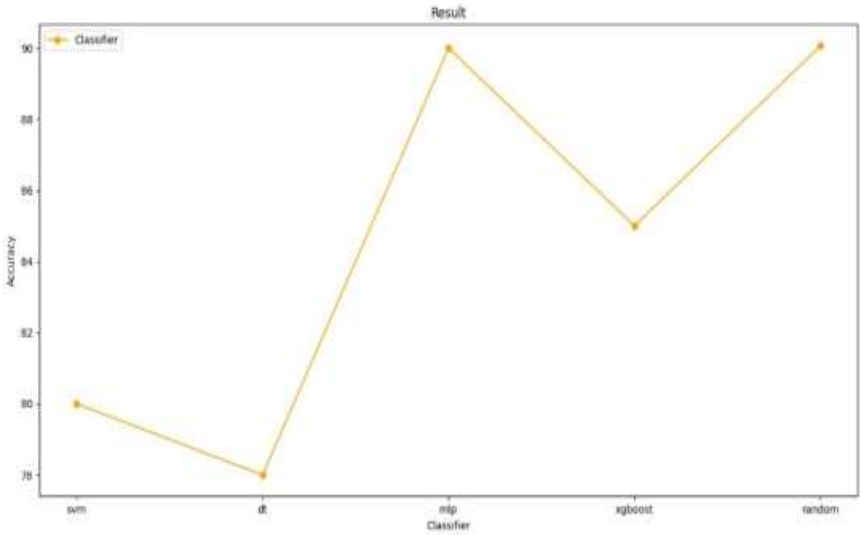

Fig 4: Line Graph Comparison with Existing Algorithms

4. ***Time Graph Analysis:*** The time graph shows the processing times for training and prediction across five algorithms. While traditional models like MLP and Decision Tree are faster on small datasets, Random Forest and XGBoost are optimized for larger datasets, balancing speed and accuracy. Random Forest provides the best trade-off, whereas XGBoost takes slightly longer due to its complex gradient-boosting process.

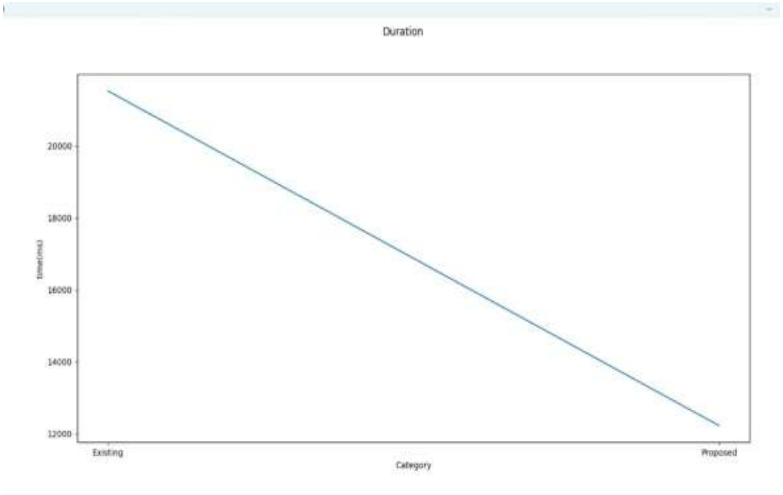

Fig 5: Line Graph Comparison with Existing Algorithms

	precision	accuracy	error rate
MLP	0.92	88	12
DT	0.85	78	22
SVM	0.90	81	19
XGBoost	0.81	85	15
Random Forest	**0.95**	91	9

Fig 6 Real time data analysis of comparison system

VI. CONCLUSION

This study demonstrates the effectiveness of Random Forest and XGBoost in detecting and classifying network attacks in Wireless Sensor Networks (WSNs). Comprehensive performance evaluations using confusion matrices, bar graphs, and line graphs reveal that the proposed models outperform traditional algorithms in accuracy, precision, recall, and F1 score. Random Forest emerges as the most reliable model due to its strong generalization ability and balanced performance across all metrics. XGBoost, although slightly lower in recall, achieves high precision, effectively minimizing false positives. Additionally, the proposed models enhance computational efficiency, as shown by time graph analysis, reducing detection time compared to existing methods. Overall, Random Forest and XGBoost provide a robust and efficient approach to intrusion detection, significantly improving the security and resilience of WSNs against evolving cyber threats. These findings offer valuable insights for optimizing network security systems and lay the groundwork for future advancements in attack detection methodologies.

IX. REFERENCES

1. Neha Jagwani 1 , Dr. Poornima G .Machine Learning Algorithms to Detect Attacks in Wireless Sensor Networks .
2. Priyanka Shah; Tanmay Kasbe. Detecting Sybil Attack, Black Hole Attack and DoS Attack in VANET Using RSA Algorithm. 2021 Emerging Trends in Industry 4.0 (ETI 4.0).
3. B.J Santhosh Kumar; Somnath Sinha An Intrusion Detection and Prevention System against DOS Attacks for Internet-Integrated WSN 2022 7th International Conference on Communication and Electronics Systems (ICCES).
4. Shalini Swami; Pushpa Singh; Sansar Singh Chauhan An Integrated Rule-Based and Machine Learning Technique for Efficient DoS Attack Detection in WSN 2024 2nd International Conference on Disruptive Technologies (ICDT)
5. Muawia A. Elsadig Detection of Denial-of-Service Attack in Wireless Sensor Networks: A Lightweight Machine Learning Approach IEEE Access (Volume: 11)
6. Somnath Sinha; Sindhu. B Impact of DoS attack in IoT system and identifying the attacker location for interference attacks 2021 6th International Conference on Communication and Electronics Systems (ICCES).
7. Sayamuddin Ahmed Jilani; Chandan Koner; Shovon Nandi Security in Wireless Sensor Networks: Attacks and Evasion 2020 National Conference on Emerging Trends on Sustainable Technology and Engineering Applications (NCETSTEA)
8. Hriday Banerjee; Surendra Yadav Energy-efficient Security Technique Implementation for Selective Forwarding Attack in WSN 2023 11th International Conference on Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON)
9. Kosaraju Chaitanya; Sankara Narayanan Security and Privacy in Wireless Sensor Networks Using Intrusion Detection Models to Detect DDOS and Drdos Attacks: A Survey 2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)
10. Shereen Ismail; Diana Dawoud; Hassan Reza A Lightweight Multilayer Machine Learning Detection System for Cyber-attacks in WSN 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)