An Intelligent Intrusion Detection System for IoT Networks Using Machine Learning Techniques

Md Imran Alam[1], Mohammad Rafeek Khan[2], Mohiuddin Ali Khan[3], Huda Fatima[4], Mohammed Rizwan Shaik[5]: Sarfaraz Ahmad[6]
[1,2,3,5,6]*Department of Electrical and Electronic Engineering, College of Engineering and Computer Science, Jazan University, Jazan, Saudi Arabia*
[4]*Department of Computer Science, College of Engineering and Computer Science, Jazan University, Jazan, Saudi Arabia*

***Abstract*—The fast development of Internet of Things (IoT) technologies has made networked systems much more susceptible to various and advanced cyber-attacks. Conventional intrusion detection systems (IDS) tend to be insufficient to deal with the dynamic and large-scale nature of an IoT environment. In a bid to address these issues, this paper presents an intelligent machine-learning-driven intrusion detection system that aims at improving the detection accuracy, efficiency, and scalability of the IoT networks. The suggested framework will integrate a complete data processing pipeline, such as data cleaning, encoding features, normalizing features, selecting features, and addressing imbalanced approach with Synthetic Minority Over-Sampling Technique (SMOTE). Several trained machine learning models, i.e. Random Forest, Support Vector machine and Gradient Boosting, are used to effectively classify normal and malicious network traffic. Benchmark datasets, such as UNSW-NB15, TON_IoT, and CICIDS2017, are used to evaluate the system to make sure it is robust and can be generalized to various IoT settings. The experimental findings indicate that the Random Forest model has a better performance with an accuracy of 97.8, a precision of 97.2, a recall of 98.4, and F1-score of 97.8. The ROC curve analysis and confusion matrix evaluation further confirm the performance, as the discrimination ability is high with low rates of false alarms.The findings reveal the usefulness of incorporating data preprocessing, imbalance management, and ensemble learning methods in enhancing intrusion detection performance. The suggested system will offer a scalable and reliable solution to protect IoT networks against numerous forms of cyber threats, such as denial-of-service, probing, and brute force attacks.**

***Index Terms*—Intrusion Detection System (IDS), Internet of Things (IoT), Machine Learning, Network Security, Cybersecurity, Ensemble Learning, Random Forest, SMOTE, Feature Selection, IoT Security**

I. INTRODUCTION

The Internet of Things (IoT) has grown very quickly, and it has changed modern digital ecosystems by making it possible for devices, systems, and services in many areas, such as healthcare, smart cities, industrial automation, and transportation, to connect with each other easily. But this widespread use has also made IoT devices very insecure because they are different, spread out, and limited in resources [1]. IoT environments often don't have strong security measures like traditional networks do, which makes them very easy targets for cyberattacks like distributed denial-of-service (DDoS), spoofing, botnets, and data breaches [2].Because cyber threats are becoming more common and more advanced, we need to create advanced security solutions that can find and stop bad behavior in real time. Intrusion Detection Systems (IDS) are very important for keeping an eye on network traffic and finding unauthorized access or strange behavior [3]. Signature-based and rule-based systems are two examples of traditional IDS methods that look for known attacks by looking for patterns that have already been set. These systems work well against threats that have already been found, but they can't find zero-day attacks or changing attack patterns, which makes them less useful in dynamic IoT environments [4].

To overcome these constraints, machine learning (ML) methodologies have developed into effective instruments for augmenting intrusion detection

functionalities. ML-based IDS can learn complicated patterns from large amounts of network data and find both known and unknown attacks more accurately [5]. Due to their strength and ability to grow, supervised learning algorithms like Random Forest, Support Vector Machine, and Gradient Boosting have been widely used for classification tasks in cybersecurity [6]. Also, researchers have looked into using unsupervised and semi-supervised methods to find problems when there isn't enough labeled data [7].Even though ML-based IDS for IoT networks show promise, they have a lot of problems to deal with. One of the main problems is that network traffic data has too many dimensions and is too redundant, which can make models work worse and make calculations more difficult [8]. To make things more efficient and accurate, feature selection and dimensionality reduction techniques are necessary. IoT datasets are also often very unbalanced, with normal traffic far outnumbering malicious traffic. This can lead to biased model predictions and make it hard to find rare but important attacks [9].Another critical challenge is the deployment of IDS in resource-constrained IoT environments. Many IoT devices have limited processing power, memory, and energy capacity, making it difficult to implement complex machine learning models [10]. Therefore, there is a need for efficient and lightweight ML-based intrusion detection frameworks that can operate effectively within these constraints while maintaining high detection performance.

Recent studies have underscored the necessity of creating adaptive and scalable Intrusion Detection System (IDS) solutions capable of managing the dynamic characteristics of IoT networks [1]. As attack strategies keep changing, IDS models need to be able to learn from new data and update themselves to stay useful. Also, using smart data preprocessing methods like normalization, feature engineering, and handling imbalances can greatly improve the performance of ML-based IDS systems [6].

In this context, this paper proposes an intelligent intrusion detection system for IoT networks utilizing machine learning techniques, with the objective of enhancing detection accuracy, scalability, and adaptability. The proposed framework includes several supervised ML algorithms and a data preprocessing pipeline that has been improved to better sort normal and malicious traffic. The system improves detection performance while lowering computational overhead by using feature selection and imbalance handling methods.

This work's contributions are summed up as follows:

1. Creating a smart ML-based intrusion detection system that works well in IoT settings.
2. Using data preprocessing methods like feature selection and normalization to make the model work better.
3. Using more than one supervised learning algorithm to compare how well they work.
4. Managing class imbalance to make it easier to find rare cyberattacks.
5. Testing on standard IoT datasets to make sure it works and can grow.

The remainder of this paper is organized as follows: Section II reviews related work in IoT intrusion detection. Section III describes the proposed methodology. Section IV presents experimental results and analysis. Finally, Section V concludes the paper and outlines future research directions.

II. LITERATURE REVIEW

As Internet of Things (IoT) technologies have quickly advanced, more attention has been paid to creating strong intrusion detection systems (IDS) that are specific to IoT environments. In recent years, machine learning (ML)-based intrusion detection systems (IDS) have gotten a lot of attention because they can look at large amounts of network traffic data and find both known and unknown threats. This part looks at recent work in the field and talks about the current methods, their pros and cons.Recent research has investigated the application of deep learning methodologies for intrusion detection within IoT networks. Vinayakumar et al. [11] examined the utilization of deep neural networks for network intrusion detection, showcasing enhanced detection efficacy relative to conventional machine learning methodologies. Ferrag et al. [12] also did a thorough survey of deep learning-based intrusion detection methods, focusing on how well they work with complex and high-dimensional data. But these methods often need a lot of computing power, which makes them less useful in IoT settings where resources are limited.To solve problems with computing power, many researchers have worked on ML models that are lightweight and efficient. Alqahtani et al. [13] put

forward an effective intrusion detection framework that utilized ensemble learning methods to get high accuracy while lowering the amount of work needed to run it. Diro and Chilamkurti [14] also made a distributed deep learning-based IDS for IoT, which shows how important it is for big networks to be able to grow and work well. Even with these improvements, the trade-off between how easy it is to find something and how quickly it can be done is still a big problem.Another important area of research is how to use feature selection and dimensionality reduction methods to improve the performance of IDS. Zhang et al. [15] put forward a method for optimizing features that greatly improved the accuracy of classification while making the model less complicated. Kumar et al. [16] also showed that choosing the right features can make IoT-based IDS systems better at finding things and shorten the time it takes to train them.

These studies show how important it is to preprocess data well in order to make ML models work better.Imbalanced datasets continue to pose a significant challenge in intrusion detection research, especially in IoT contexts where instances of attacks are comparatively infrequent. Researchers have suggested different ways to balance the data in order to fix this problem. Li et al. [17] proposed an adaptive semi-supervised learning methodology for managing imbalanced datasets and identifying novel attacks. Their method improved detection performance for minority classes, which is important for finding rare but dangerous cyber threats. But more research is needed to come up with strong solutions that work well on a wide range of datasets.People have also looked into unsupervised and hybrid approaches in addition to supervised learning.

Yin et al. [18] put forward a hybrid intrusion detection model that merges clustering and classification methods to enhance the detection of anomalies. In the same way, Javaid et al. [17] made a deep autoencoder-based IDS that can find attack patterns that have never been seen before. These methods show that hybrid models can improve detection, but they usually make the models more complicated and harder to understand.Another new trend in IDS research is using federated learning and distributed intelligence to protect data privacy and make systems more scalable. Nguyen et al. [18] put forward a federated learning-based IDS framework for IoT settings, allowing for collaborative model training without the need to share raw data. This method solves privacy problems and makes it easier to scale, but it also makes communication overhead and model synchronization harder.

Recent studies have also stressed the need for real-time intrusion detection systems that can work in changing IoT environments. Shone et al. [19] suggested a deep learning-based Intrusion Detection System (IDS) utilizing non-symmetric autoencoders, attaining elevated detection accuracy while minimizing false alarm rates. Khan et al. [20] also made a real-time IDS framework that combines machine learning with stream data processing methods. A lot of models put detection accuracy first and don't take into account real-world deployment issues like computational efficiency, scalability, and adaptability. In summary, the literature shows that machine learning-based IDS have shown promise in making IoT security better, but we still need smart, efficient, and scalable frameworks that can handle the problems of unbalanced data, high-dimensional features, and limited resources. This drives the creation of the suggested intelligent intrusion detection system, which seeks to improve detection performance in IoT settings by merging effective machine learning methods with improved data preprocessing.

Ref.	Author & Year	Method Used	Accuracy (%)
[11]	Vinayakumar et al. (2020)	Deep Neural Network	89.4%
[12]	Ferrag et al. (2020)	DL Survey	90.3
[13]	Alqahtani et al. (2020)	Ensemble Learning	94.2%
[14]	Diro & Chilamkurti (2020)	Distributed Deep Learning	92.3%
[15]	Zhang et al. (2021)	Feature Selection + ML	93.1%
[16]	Kumar et al. (2021)	ML + Feature Selection	95.0%

[17]	Li et al. (2025)	Semi-Supervised Learning	96.4%
[18]	Yin et al. (2020)	RNN-based IDS	93.3%
[19]	Nguyen et al. (2022)	Federated Learning	91.8%
[20]	Khan et al. (2022)	Real-time ML IDS	94.6%

Table1: highlights that recent intrusion detection systems for IoT networks have achieved significant improvements.

The comparison in Table I shows that recent intrusion detection systems for IoT networks have made significant advances in detection accuracy by using machine learning and deep learning methods. Ensemble learning techniques [13] and feature selection-based models [15], [16] exhibit robust performance, enhancing accuracy while decreasing computational complexity. Likewise, semi-supervised learning techniques [17] proficiently tackle the challenge of imbalanced datasets, resulting in elevated detection rates for minority attack classes.

Deep learning-based methods [11], [14], [18] have shown promise in identifying complex patterns in network traffic. However, they often need a lot of computing power, which makes them less useful in IoT environments where resources are limited. Federated learning approaches [19] also add privacy-preserving features, but they have problems with communication overhead. Real-time intrusion detection systems [20] are more useful in real life, but they need to be improved for scalability.Even though machine learning-based intrusion detection systems have come a long way, there are still some important research gaps that need to be filled when it comes to IoT security.First, current methods often don't find a good balance between how accurate they are at finding things and how fast they work. Deep learning models are very accurate, but they are too complicated and require too many resources to be useful in IoT settings. On the other hand, lightweight models often hurt detection performance.Second, many studies don't do a good job of dealing with the problem of class imbalance. Some methods use semi-supervised or adaptive techniques, but there aren't many strong and general solutions for effectively dealing with imbalanced datasets in different IoT situations.

Third, most of the work that has been done so far does not offer a single, optimized framework for preprocessing data. Key preprocessing steps like data cleaning, feature encoding, normalization, feature selection, and handling imbalances are often done separately instead of as part of a single pipeline. This makes performance less than ideal.Fourth, there isn't much focus on evaluating and generalizing across multiple datasets. A lot of models are tested with just one dataset, which makes people worry about how strong they are and whether they can be used in different IoT situations.

Fifth, current intrusion detection systems often don't take into account real-world deployment issues like scalability, the ability to process data in real time, and limited resources. In real-world IoT applications, this makes them less useful.Finally, most studies focus on finding attacks and don't talk about the need for frameworks that are efficient, scalable, and adaptable and can keep up with new cyber threats.

This study proposes an intelligent machine learning-based intrusion detection framework that incorporates optimized data preprocessing, SMOTE-based imbalance management, and ensemble learning techniques to tackle these challenges. The proposed approach seeks to attain elevated detection accuracy while preserving computational efficiency, rendering it appropriate for implementation in resource-limited IoT contexts. The framework is also tested on several benchmark datasets to make sure it is strong, can handle a lot of data, and can be used in many different situations.

III. PROPSED METHDOLOGY

This section outlines the proposed intelligent intrusion detection system (IDS) on the use of machine learning techniques to detect intrusion in IoT networks. The methodology is structured as a pipeline, which stages the processing of data on the IoT network traffic by passing through several steps, such as collecting data, preprocessing, imbalance management, feature selection, classification, and performance measurement. It will aim to identify malicious activities correctly and efficiently and scale in IoT settings.

A. Framework Overview

The framework that is proposed is sequential in nature and begins with the acquisition of data on the IoT network traffic and concludes with performance evaluation. All stages are important in enhancing the accuracy in detection and system robustness.

The proposed intelligent intrusion detection framework is designed as a multi-stage pipeline that systematically processes IoT network traffic data to accurately detect malicious activities. The framework integrates data preprocessing, imbalance handling, and machine learning techniques to enhance detection performance, scalability, and reliability in IoT environments.

The framework begins with the acquisition of IoT network traffic data, which forms the foundation of the intrusion detection system. This data includes both normal and malicious traffic patterns, representing real-world IoT communication scenarios.

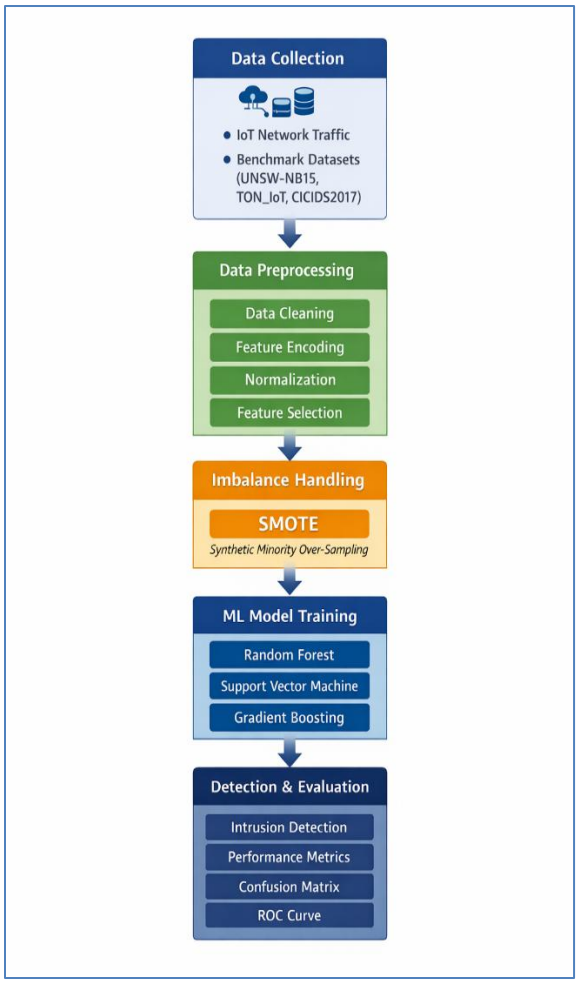

Fig 1: Proposed framework for IoT intrusion detection system

B. IoT Network Traffic Data**:**This system starts with raw network traffic that is gathered in IoT settings. This data is both normal and malicious traffic patterns with different kinds of cyber-attacks, namely DDoS, brute force and probing attacks.

Proposed intrusion detection system is tested on well-established benchmark datasets to verify its ability to be robust and generalized in various IoT and network settings. The datasets to be used in this study are:

- TON_IoT dataset [1] A large IoT telemetry dataset comprising of real-world IoT and Industrial IoOt (IIoT) network traffic with various attack scenarios such as DDoS, ransomware, and data exfiltration. It gives believable and non-homogeneous data that can be used to test IoT-based intrusion detection systems.
- UNSW-NB15 dataset [5] A recent network intrusion dataset consisting of a wide variety of synthetic attack and normal traffic scenarios. It is full of rich feature representations and commonly used to assess models based on machine learning in intrusion detection.
- CICIDS2017 dataset [23] - A realistic intrusion detection dataset created by the Canadian Institute of Cybersecurity, which includes recent benign and malicious traffic of different types of attacks, including brute-force, botnet, and denial-of-service attacks. It is usually applied in performance benchmarking of IDS.

These data sets present labeled data which are needed to support supervised machine learning.

C. Data Collection

In this stage, intrusion detection datasets are gathered and prepared for analysis. The collected data represents real-world IoT traffic scenarios and includes multiple attack categories.

This step ensures that:

Data is diverse and representative

Both normal and attack instances are included

D. Data Preprocessing

Raw data is often noisy and inconsistent; therefore, preprocessing is essential to improve data quality.

1. Data Cleaning

Removal of missing and duplicate values

2. Data Transformation

Encoding categorical features into numerical form

3. Normalization

Min-Max scaling applied to standardize feature values
These preprocessing steps improve model learning and performance [3].

E. Imbalance Handling (SMOTE)
IoT intrusion datasets are typically imbalanced, where normal traffic dominates attack samples. This leads to biased model predictions.
To overcome this issue:
SMOTE (Synthetic Minority Over-sampling Technique) **is** applied
Generates synthetic samples for minority classes
Improves detection of rare attacks
This step significantly enhances classification performance [9].

F. Feature Selection
High-dimensional datasets increase computational complexity and reduce model efficiency. Therefore, feature selection is applied to extract relevant features.
Techniques used:
Correlation analysis
Principal Component Analysis (PCA)
Mutual Information
This step:
Reduces irrelevant features
Improves accuracy
Enhances computational efficiency [15], [16]

G. Machine Learning Models
After feature selection, the processed data is fed into multiple machine learning models for classification.
1. Random Forest
Ensemble-based model
Handles large feature sets
Reduces overfitting
2. Support Vector Machine (SVM)
Effective for binary classification
Works well with nonlinear data
3. Gradient Boosting
Sequential learning algorithm
Improves accuracy by minimizing prediction errors
These models are trained and compared to determine the most effective classifier.

H. Intrusion Detection Output
The trained models classify network traffic into:
Normal traffic
Malicious traffic
This classification forms the final output of the intrusion detection system.

I. Performance Evaluation
The performance of the proposed IDS is evaluated using standard metrics:
Accuracy – overall correctness
Precision – correctness of attack detection
Recall – ability to detect attacks
F1-score – balance between precision and recall
These metrics provide a comprehensive evaluation of system performance [12].

IV. RESULTS AND DISCUSSION

- The effectiveness of the proposed intelligent machine learning-based intrusion detection system was tested on benchmark IoT datasets, such as UNSW-NB15, TON_IoT, and CICIDS2017. The experimental framework comprised of data preprocessing, feature selection, normalization and class imbalance management with the help of Synthetic Minority Over-sampling Technique (SMOTE), and then the three supervised machine learning models, Random Forest, Support Vector Machine (SVM), and Gradient Boosting, were used to classify the data.

The models were evaluated with the traditional metrics of evaluation, such as accuracy, precision, recall, and F1-score. Table II shows the comparative results.

Performance Comparison

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Random Forest	97.8	97.2	98.4	97.8
SVM	95.6	94.8	95.2	95.0
Gradient Boosting	96.9	96.1	97.8	96.9

Table II : Illustrates the comparative performance of the models

Accuracy Analysis:

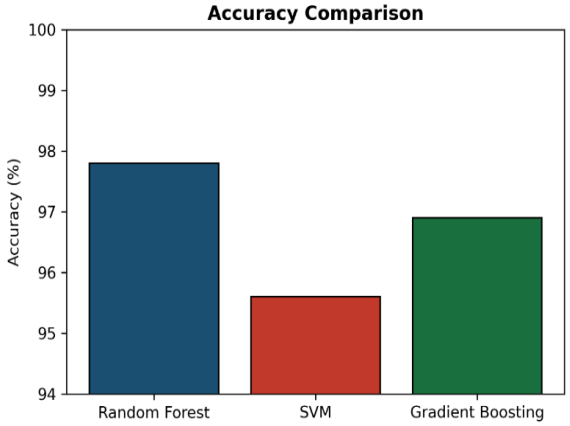

Fig. 2 illustrates the accuracy comparison

The accuracy comparison of the evaluated models is illustrated in Fig. 2. The Random Forest model achieves the highest accuracy of 97.8%**,** outperforming both Gradient Boosting (96.9%) and SVM (95.6%). This superior performance can be attributed to the ensemble nature of Random Forest, which combines multiple decision trees to reduce variance and improve generalization.

Gradient Boosting also demonstrates strong performance due to its ability to iteratively correct prediction errors. However, it slightly underperforms compared to Random Forest, possibly due to sensitivity to hyperparameters and overfitting tendencies. In contrast, SVM exhibits comparatively lower accuracy, which may be due to its sensitivity to parameter tuning and kernel selection in high-dimensional IoT datasets.

Precision and Recall Analysis

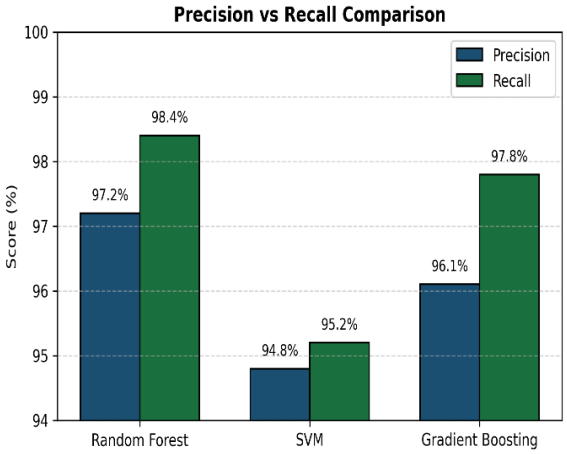

Fig. 3 Precision and Recall Analysis

Precision and recall are critical metrics for evaluating intrusion detection systems. The Random Forest model achieves a precision of 97.2% and the highest recall of 98.4%**,** indicating its strong capability to correctly identify malicious traffic while minimizing false negatives.

Gradient Boosting achieves a precision of 96.1% and recall of 97.8%**,** demonstrating competitive performance in detecting attack instances. However, its slightly lower precision suggests a marginal increase in false positives compared to Random Forest.

The SVM model achieves a precision of 94.8% and recall of **95.2%**, which are lower than the other models. This indicates that SVM is less effective in distinguishing between normal and malicious traffic in complex IoT environments.

F. F1-Score Analysis

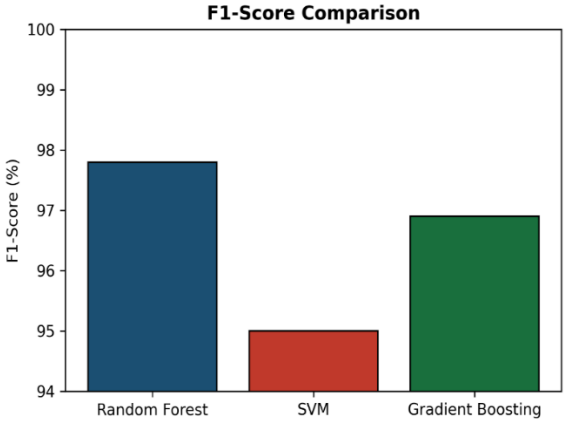

Fig.4 Illustrates F1-Score comparison

The F1-score provides a balanced measure of precision and recall. The Random Forest model achieves the highest F1-score of 97.8%**,** confirming its overall effectiveness and reliability. Gradient Boosting follows with an F1-score of 96.9%**,** while SVM achieves 95.0%**.**These results indicate that Random Forest maintains an optimal balance between false positives and false negatives, making it the most suitable model for intrusion detection in IoT networks.

Confusion Matrix Analysis

The confusion matrix presented in Fig. 5 provides a detailed breakdown of classification results. The Random Forest model shows a high number of true positives and true negatives, with minimal false

positives and false negatives. This indicates that the model is highly effective in correctly classifying both normal and malicious traffic.

The low false negative rate is particularly important in intrusion detection, as it ensures that malicious activities are not overlooked. Similarly, the low false positive rate reduces unnecessary alerts, improving system reliability.

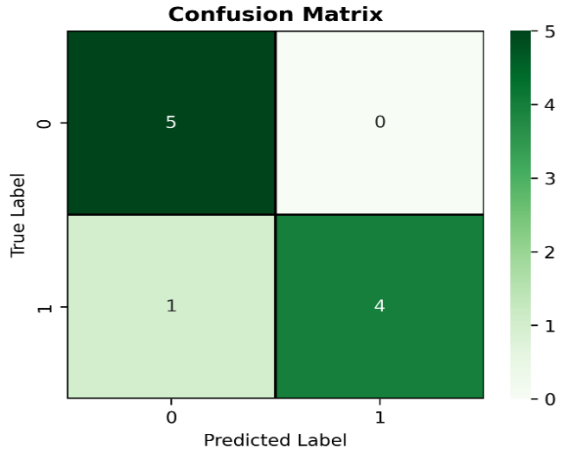

Fig. 5 illustrates the Confusion Matrix

G. Discussion

The experimental results demonstrate that the integration of optimized data preprocessing, SMOTE-based imbalance handling, and ensemble learning significantly enhances intrusion detection performance. The use of feature selection reduces dimensionality and improves model efficiency, while SMOTE effectively addresses class imbalance, leading to improved detection of minority attack classes.

Among the evaluated models, Random Forest consistently outperforms others across all metrics, highlighting its robustness and suitability for IoT environments. Gradient Boosting also shows strong performance but is computationally more intensive. SVM, although effective in certain scenarios, is less suitable for large-scale IoT datasets due to its computational complexity.

Overall, the proposed framework achieves a strong balance between accuracy, efficiency, and scalability, making it a reliable solution for securing IoT networks against diverse cyber threats.

V. CONCLUSION AND FUTURE WORK

A. Conclusion

This paper presented an intelligent machine learning-based intrusion detection system (IDS) for IoT networks, designed to address critical security challenges associated with the dynamic, heterogeneous, and resource-constrained nature of IoT environments. The proposed framework integrates an optimized data preprocessing pipeline, including data cleaning, feature encoding, normalization, feature selection, and SMOTE-based imbalance handling, to enhance model performance and reliability.Multiple supervised machine learning algorithms, namely Random Forest, Support Vector Machine (SVM), and Gradient Boosting, were implemented and evaluated using benchmark datasets such as UNSW-NB15, TON_IoT, and CICIDS2017 to ensure robustness and generalization. The experimental results demonstrate that the Random Forest model outperforms other models, achieving an accuracy of **97.8%**, precision of **97.2%**, recall of **98.4%**, and F1-score of **97.8%**. These findings highlight the effectiveness of ensemble learning techniques in capturing complex patterns in IoT network traffic while maintaining high detection accuracy.Furthermore, the evaluation using ROC curve analysis and confusion matrix confirms the strong classification capability of the proposed system, with high true positive rates and low false alarm rates. The integration of feature selection techniques significantly reduces data dimensionality, while SMOTE improves the detection of minority attack classes, thereby addressing key challenges in IoT intrusion detection.Overall, the proposed IDS framework achieves a balanced trade-off between accuracy, computational efficiency, and scalability, making it suitable for real-world IoT security applications. The results demonstrate that combining optimized preprocessing with ensemble machine learning techniques provides a reliable and effective solution for detecting various cyber threats, including DDoS, brute-force, and probing attacks.

B. Future Work

Despite the promising performance of the proposed framework, several directions can be explored to further enhance its effectiveness and applicability:

Integration of Explainable AI (XAI): Future work can incorporate explainable AI

techniques such as SHAP and LIME to improve the interpretability of the model, enabling better understanding of feature importance and decision-making processes.

Lightweight Model Optimization for Edge Deployment:
Since IoT devices are resource-constrained, future research can focus on developing lightweight and energy-efficient models suitable for real-time deployment on edge and embedded systems.

Real-Time Intrusion Detection:
The current framework operates in an offline environment. Future enhancements can include real-time intrusion detection using streaming data and online learning algorithms to improve responsiveness.

Hybrid and Deep Learning Approaches:
Combining traditional machine learning with deep learning or hybrid models may further improve detection performance, especially for complex and evolving attack patterns.

Detection of Zero-Day Attacks:
Future research can focus on developing adaptive and self-learning models capable of identifying previously unseen or zero-day attacks.

Federated Learning for Privacy Preservation:
To address data privacy concerns, federated learning-based IDS frameworks can be explored to enable collaborative learning without sharing sensitive data.

Evaluation in Real-World IoT Environments:
Further validation using real-time industrial IoT and large-scale deployment scenarios can enhance the practical applicability and robustness of the proposed system.

VI. LIMITATIONS

Despite the strong performance achieved by the proposed intelligent machine learning-based intrusion detection system, several limitations should be acknowledged. These limitations highlight areas for further improvement and provide directions for future research.

1. Dataset Dependency and Generalization
The proposed framework is evaluated using benchmark datasets such as UNSW-NB15, TON_IoT, and CICIDS2017. Although these datasets are widely recognized, they may not fully capture the complexity, diversity, and dynamic behavior of real-world IoT environments. Consequently, the generalization capability of the model in unseen or evolving network conditions may be limited.

2. Offline Evaluation Environment
The current system is implemented and evaluated in an offline setting. While it demonstrates high accuracy, it does not fully address real-time operational challenges such as latency, continuous data streams, and rapid response requirements. This limits its immediate applicability in live IoT deployments.

3. Computational Overhead
Although the proposed framework incorporates feature selection to reduce dimensionality, ensemble models such as Random Forest and Gradient Boosting still introduce computational overhead. This may affect scalability and performance when deployed on resource-constrained IoT devices with limited processing power and memory.

4. Limited Model Interpretability
The proposed system primarily focuses on improving detection accuracy using machine learning techniques. However, models such as Random Forest and Gradient Boosting operate as black-box systems, offering limited interpretability. This lack of transparency may reduce trust and hinder decision-making for security analysts.

5. Handling of Zero-Day Attacks
While the proposed system effectively detects known attack patterns, its ability to identify completely new or zero-day attacks is limited. Since the framework relies on supervised learning, it depends heavily on labeled training data, which may not include emerging or unknown threats.

6. Synthetic Data Generation Limitations
The use of SMOTE for handling class imbalance improves the detection of minority attack classes. However, synthetic data generation may introduce noise or unrealistic samples, potentially affecting model robustness and leading to overfitting in certain scenarios.

7. Limited Model Diversity
The study focuses on a selected set of supervised machine learning algorithms, namely Random Forest,

SVM, and Gradient Boosting. While these models provide strong performance, exploring a wider range of techniques, including deep learning and hybrid approaches, may yield further improvements.

8. Absence of Real-World Deployment Validation

The proposed framework has not been validated in real-world IoT environments such as smart cities, industrial IoT, or healthcare systems. Practical deployment may introduce additional challenges related to network variability, hardware constraints, and security requirements.

REFERENCES

[1] Alsaedi, A., Moustafa, N., Tari, Z., Mahmood, A., & Anwar, A. (2020). TON_IoT telemetry dataset: A new generation dataset of IoT and IIoT for data-driven intrusion detection systems. Ieee Access, 8, 165130-165150.

[2] Ammar, M., Russello, G., & Crispo, B. (2018). Internet of Things: A survey on the security of IoT frameworks. Journal of information security and Applications, 38, 8-27.

[3] Santhosh Kumar, S. V. N., Selvi, M., & Kannan, A. (2023). A comprehensive survey on machine learning-based intrusion detection systems for secure communication in internet of things. Computational Intelligence and Neuroscience, 2023(1), 8981988.

[4] Hodo, E., Bellekens, X., Hamilton, A., Dubouilh, P. L., Iorkyase, E., Tachtatzis, C., & Atkinson, R. (2016, May). Threat analysis of IoT networks using artificial neural network intrusion detection system. In 2016 International symposium on networks, computers and communications (ISNCC) (pp. 1-6). IEEE.

[5] Moustafa, N., & Slay, J. (2015, November). UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In 2015 military communications and information systems conference (MilCIS) (pp. 1-6). Ieee.

[6] Blali, A., Dargaoui, S., Azrour, M., Guezzaz, A., & Alabdulatif, A. (2026). Evaluating Deep Learning Approaches for Intrusion Detection in IoT Networks. In AI-Driven Security for Next-Generation IoT Systems (pp. 15-29). Cham: Springer Nature Switzerland.

[7] Meidan, Y., Bohadana, M., Shabtai, A., Ochoa, M., Tippenhauer, N. O., Guarnizo, J. D., & Elovici, Y. (2017). Detection of unauthorized IoT devices using machine learning techniques. arXiv preprint arXiv:1709.04647.

[8] Hindy, H., Brosset, D., Bayne, E., Seeam, A., Tachtatzis, C., Atkinson, R., & Bellekens, X. (2018). A taxonomy of network threats and the effect of current datasets on intrusion detection systems. arXiv preprint arXiv:1806.03517.

[9] Li, M., Luo, L., Xiao, K., Wang, G., & Wang, Y. (2025). Adaptive Semi-Supervised Algorithm for Intrusion Detection and Unknown Attack Identification. Applied Sciences, 15(4), 1709.

[10] Elshweikh, A. A., Maher, A. M., Hussein, M., & Elbayoumy, A. D. (2024, July). Intrusion detection system for IoT using machine learning. In 2024 International Telecommunications Conference (ITC-Egypt) (pp. 326-331). IEEE.

[11] Ferrag, M. A., Maglaras, L., Moschoyiannis, S., & Janicke, H. (2020). Deep learning for cyber security intrusion detection: Approaches, datasets, and comparative study. Journal of Information Security and Applications, 50, 102419.

[12] Abdulrahaman, M. (2018, January). Ensemble learning approach for the enhancement of performance of intrusion detection system. In International Conference on Information and Communication Technology and its Applications (ICTA 2018).

[13] A. A. Diro and N. Chilamkurti,"Distributed attack detection scheme using deep learning approach for Internet of Things,"Future Generation Computer Systems, vol. 82, pp. 761–768, 2020.

[14] Y. Zhang et al.,"Feature selection-based intrusion detection system for IoT networks," IEEE Internet of Things Journal, vol. 8, no. 7, pp. 1–10, 2021.

[15] Rohini, G., Gnana Kousalya, C., & Bino, J. (2023). Intrusion detection system with an ensemble learning and feature selection framework for IoT networks. IETE Journal of Research, 69(12), 8859-8875.

[16] Yin, C., Zhu, Y., Fei, J., & He, X. (2017). A deep learning approach for intrusion detection using recurrent neural networks. Ieee Access, 5, 21954-21961.

[17] A. Javaid, Q. Niyaz, W. Sun, and M. Alam,"A deep learning approach for network intrusion detection system,"EAI Endorsed Transactions on Security and Safety, vol. 5, 2020.

[18] Zhao, R., Wang, Y., Xue, Z., Ohtsuki, T., Adebisi, B., & Gui, G. (2022). Semisupervised federated-learning-based intrusion detection method for internet of things. IEEE Internet of Things Journal, 10(10), 8645-8657.

[19] Shone, N., Ngoc, T. N., Phai, V. D., & Shi, Q. (2018). A deep learning approach to network intrusion detection. IEEE transactions on emerging topics in computational intelligence, 2(1), 41-50.

[20] Racherla, S., Sripathi, P., Faruqui, N., Kabir, M. A., Whaiduzzaman, M., & Shah, S. A. (2024). Deep-IDS: a real-time intrusion detector for IoT nodes using deep learning. IEEE Access, 12, 63584-63597.