***DOI: 10.15849/IJASCA.251130.15***

***Int. J. Advance Soft Compu. Appl., Vol. 17, No. 3, November 2025***
***Print ISSN: 2710-1274, Online ISSN: 2074-8523***
***Copyright © Al-Zaytoonah University of Jordan (ZUJ)***

# Comparative Analysis of Machine Learning Algorithms for Intrusion Detection in IoT Networks

**Mohammad AbdulJawad[1], and Amal Ahmad[2]**

[1]Software Engineering Department, Faculty of Science and Information Technology, Al-Zaytoonah University of Jordan, Amman, Jordan; M.Abduljawad@Zuj.edu.jo

[2]Electrical Engineering Department, Faculty of Engineering and Technology, Al-Zaytoonah University of Jordan, Amman, Jordan; Amal.Q@zuj.edu.jo

## Abstract

***The growing threat of cyberattacks poses significant risks to the security and privacy of the Internet of Things (IoT), affecting everything from devices to networks. In response to these threats, research has focused on developing effective countermeasures. Intrusion Detection Systems (IDSs), particularly those utilizing Machine Learning (ML) techniques for faster attack detection, are now recognized as some of the most powerful solutions for safeguarding the IoT environment. This study evaluates the effectiveness of various supervised Machine Learning techniques, specifically K-Nearest Neighbors (KNN), Random Forest (RF), Decision Trees (DT), and Support Vector Machines (SVM), in detecting anomalies within IoT networks. The performance of these algorithms was assessed using the BoTNeTIoT-L01-v2 dataset. The results show notable performance differences before and after normalization. SVM emerged as the top performer, achieving 100% validation accuracy both before and after normalization. In contrast, RF and DT classifiers saw improved accuracy after normalization, reaching 92%, while KNN's accuracy decreased post-normalization. Additionally, clustering techniques, including K-Means and soft clustering, were used to categorize network behavior. By comparing multiple ML approaches, this study makes a valuable contribution to advancing IoT security through enhanced intrusion detection methods.***

**Keywords:** ***Intrusion Detection System (IDS), Machine Learning Classification, Data Preprocessing and Normalization, IoT Security and Botnet Attacks, Supervised and Unsupervised Learning***

***Received 4 May 2025; Accepted 10 September 2025***

# 1. Introduction

Internet of Things (IoT) Intrusion Detection Systems (IDSs) are designed to detect and prevent unauthorized use or malicious activities on networks, systems, or applications. As cyberattacks have expanded exponentially, the need for IDSs has also increased further, with current security measures not being adequate enough. IDSs are crucial in their ability to detect attacks and potential vulnerabilities, which allows companies to act to prevent total damage [1].

Machine Learning (ML) enhances the performance of IDS by enabling models to process huge volumes of data to identify patterns indicative of attacks. Using supervised learning, unsupervised learning, and reinforcement learning methods, ML-based IDSs can detect anomalies and malicious behavior. Various common ML algorithms are used in IDS implementations, including Support Vector Machines (SVM), K-Nearest Neighbors (KNN), Decision Trees (DT), and Random Forests (RF). Each algorithm operates differently: SVM discovers an optimal hyperplane to classify network traffic, DT and RF employ decision rules to detect anomalies, and KNN classifies data based on proximity to known samples [2] [34].

Effective data preprocessing is essential for improving IDS performance, particularly through normalization, which standardizes data and reduces biases caused by varying feature scales. This study evaluates how normalization influences classification accuracy across different models. Results indicate that SVM achieved 100% accuracy both before and after normalization, while RF and DT exhibited improved performance following normalization [3].

In addition, an ML-driven IDS model has been proposed for IoT networks that can be trained on multiple sources from large and classified datasets. The model effectively classifies small-sized data in the target domain. Another IDS technique uses enhanced transient search optimization to optimize hyperparameters of ML models. Experimental results indicate that this method of optimization improves accuracy while reducing the false alarm rate compared to other IDS implementations [4].

This research investigates a number of ML classifiers for detecting anomalies in IoT networks with and without normalization. This research examines the impact of clustering algorithms, such as K-Means and soft clustering, on classifying network behavior and compares the classification accuracy of each approach. The findings highlight the importance of preprocessing steps in augmenting IDS capability and mitigating new cyber threats.

# 2. Related Works

Numerous studies have reflected on how Machine Learning (ML) has been applied for the implementation of Intrusion Detection Systems (IDS) in the Internet of Things (IoT). They reflect on sophisticated algorithmic and technological capability in the context of recognizing attacks on networks, systems, and applications. These researchers have enumerated numerous approaches including supervised learning, unsupervised learning, deep models of learning, as well as statistical algorithms for anomaly detection.

Zhang et al. [5] surveyed the use of machine learning for intrusion detection in the last decade, classifying methods into traditional machine learning, ensemble learning, and deep learning. Comparative experiments were performed on Decision Trees, Naïve Bayes, Support Vector Machines, Random Forests, XGBoost, Convolutional Neural Networks, and Recurrent Neural Networks using the KDD CUP99 and NSL-KDD datasets. Results indicated that ensemble learning is generally more accurate, while Naïve Bayes is best at detecting new attacks since it has a fast training speed.

Saran and Kesswani in [6] performed analysis to evaluate the performance ML classifier techniques. Overall the experiments result obtained higher rates of accuracy. The accuracy of experimented Intrusion Detection System has 97.76%, 97.80%, 97.58%, 99.98%, 99.98%, and 97.58% using k-Nearest Neighbour (k-NN), Support Vector Machine (SVM), Naive Bayes (NB), Random Forest (RF), Decision Tree (DT) and Stochastic Gradient Descent (SGD) classifiers respectively.

Saranya et al. [7] compared certain ML classifiers including Linear Discriminant Analysis (LDA), Classification and Regression Trees (CART), RF, Support Vector Machines (SVM), NB, DT, and LR using the KDDcup99 dataset. RF achieved a 99.81% accuracy due to a filter-based feature selection technique reducing the feature set from 42 to 20. However, multiclass classification methods were not studied in the research.

Baich et al. [8] researched over two dozen articles to draw a comparison across various ML models and concluded that Decision Trees (DT), Support Vector Machines (SVM), and Convolutional Neural Networks (CNN) were the top-most implemented models used in intrusion detection. Interestingly, Decision Trees showed the best performance on the Sensor480 dataset. They also tested two feature selection techniques (Pearson Correlation and Fisher Score) and used Principal Component Analysis (PCA) for feature extraction. The experiment aimed to assess both binary classification (identifying malicious vs. benign traffic) and multi-class classification (detecting various attack types like DoS, Probe, U2R, and R2L). The results revealed that the Decision Tree algorithm, combined with the Fisher Score feature selection, achieved an impressive accuracy of 99.26% and the fastest prediction time of just 0.4 seconds.

Other scientists attempted different ML approaches for IDS. In [9], the authors designed a feature clustering technique based on Flow, MQTT, and TCP using the UNSW-NB15 dataset. It was attempted to overcome issues like overfitting, the curse of dimensionality, and class imbalance within the dataset. Supervised machine learning algorithms, including Random Forest, SVM, and Artificial Neural Networks, were applied to these clusters. The results showed that Random Forest achieved an accuracy of 98.67% for binary classification and 97.37% for multi-class classification. Additionally, the proposed feature clusters provided higher accuracy and required less training time compared to other state-of-the-art supervised ML approaches. In contrast in [10], cyberattacks have been categorized using the UNSW-NB15 dataset through the implementation of NB, RF, J48, and Zero algorithms. Two clusters were generated with K-means and expectation-maximization clustering algorithms on attack or normal traffic. RF and J48 achieved 97.59% and 93.78% accuracy, respectively.

Zwane et al. [11], evaluated seven machine learning classifiers: Multi-Layer Perceptron, Bayesian Network, Support Vector Machine (SMO), Adaboost, Random Forest, Bootstrap Aggregation, and Decision Tree (J48). The WEKA tool was used to create and assess the classifiers. When comparing detection accuracy metrics such as AUC, TPR, and FPR, the results showed that ensemble-based learning methods beat single learning methods. Ensemble classifiers, on the other hand, are more time-consuming to design and test.

In [12], the authors focused on enhancing security for Internet of Things (IoT) devices, which face critical security issues. They presented a deep learning-based technique, i.e., Convolutional Neural Networks (CNN), to detect anomalies and possible intrusions in IoT systems. The technique is able to effectively analyze traffic flows in IoT networks. The model was trained and tested with two datasets, NID and BoT-IoT, and obtained 99.51% and 92.85% accuracy, respectively.

In [13], a modified IDS utilizing RF on the IoTID20 dataset reached 96.5% accuracy after applying feature selection and transforming categorical values into numeric values. While most studies focused on binary classification, some suggested expanding classification to multiple categories (9–10 classes).

These researches highlight the significance of ML classifiers and feature selection techniques towards improving IDS performance. Although previous work primarily concentrated on classifier evaluation and feature selection optimization, this work further investigates the impact of normalization on classification accuracy among supervised learning methods, i.e., KNN, RF, DT, SVM and clustering, to develop more effective intrusion detection strategies for IoT networks.

Intrusion Detection Systems (IDSs) are a critical component of networked environment security, particularly within the rapidly growing Internet of Things (IoT) ecosystem [15]. In smart applications, such as the Fog-based IoT security framework proposed by [16], IDSs play a vital role in safeguarding sensor networks and communication channels from cyber threats. This framework integrates a specialized crawler at the Fog layer for real-time monitoring and a Machine Learning (ML)-based behavior analyzer to detect malicious activities. Given the increasing sophistication of cyberattacks, ensuring secure data transmission is as important as maintaining system efficiency. Therefore, employing IDS models supported by advanced ML classifiers—such as Extra Trees, Random Forest, and Decision Trees—enhances both the reliability and resilience of IoT systems [17]. This dual focus on security and performance not only strengthens detection accuracy while minimizing false alarms but also ensures that innovative IoT applications are protected against malicious intrusions in distributed Fog computing environments [18].

Comparative research within the field comes under two general types:

1. Existing Work Comparison: Comparing methodologies, datasets, strengths, and weaknesses of existing IDS research papers by reading and analyzing them.
2. Proposed Models Comparison: Comparing of newly proposed IDS models, with focus on different ML classifiers, feature selection techniques, datasets, and experimental conditions.

The following table provides an overview of well-known comparative studies on IDSs according to ML approaches, categorized by focus areas:

Table 1: Summary of Comparative Studies on Intrusion Detection Systems (IDS) Based on Machine Learning Techniques

| **Study** | **Focus** | **Key Findings** |
|---|---|---|
| **Category: Existing Work Comparison** | | |
| Bhatia et al. [18] | Review of intrusion detection techniques (2016-2020) | Identified strengths and weaknesses of various approaches, provided accuracy comparison. |
| Kathiresan et al. [19] | Comparison of ML classification methods for network intrusion detection | Systematic evaluation of classifiers, highlighting benefits and challenges. |
| Boyanapalli & Shanthini [20] | Categorization of IDS research in IoT environments | Analyzed datasets used, highlighted IDS model features and effectiveness. |
| Anushiya & Lavanya [21] | Comparison of traditional, ML, and Deep Learning detection techniques | Evaluated advantages and limitations of different approaches. |
| Sangwan & Chhillar [22] | Review of ML algorithms for IoT security | Compared accuracy of classification techniques and discussed IDS limitations. |
| **Category: Proposed Models Comparison** | | |
| Agrawal & Singh [23] | Evaluation of SVM kernel functions for anomaly detection | Linear SVM kernel achieved 99.99% accuracy using NSL-KDD dataset. |
| Aswal et al. [24] | Classification of botnet attacks in IoV environments using ML classifiers | KNN and CART achieved highest accuracy (99.79% and 99.97%) after reducing features from 71 to 35. |
| Ibrahim & Thanon [25] | Effectiveness of ANOVA F-test & Recursive Feature Elimination (RFE) | RF classifier performed best across both full and reduced feature sets using NSL-KDD dataset. |
| Mondal & Singh [26] | Evaluation of 8 ML classifiers on network traffic data | Decision Tree and Gradient Boosting achieved highest accuracy. |
| Manvith et al. [27] | Performance analysis of SVM, LR, and RF classifiers | RF outperformed others in detecting network attacks on KDDCup99 dataset. |
| Dwibedi et al. [28] | Analysis of UNSW-NB15, Bot-IoT, and CSE-CIC-IDS2018 datasets | RF, SVM, Deep Learning, and XGBoost classifiers were tested on key dataset features. |
| Kilincer et al. [29] | ML model performance on 5 intrusion detection datasets | Decision Tree consistently achieved highest accuracy. |

| Study | Focus | Key Findings |
|---|---|---|
| Al Fayoumi et al. [30] | Impact of split ratios (0.5, 0.4, 0.3) on ML classifiers for email phishing detection | SVM achieved highest accuracy with a 0.3 train-test split. |

# 4. Proposed Model

The Intrusion Detection System (IDS) workflow typically consists of four main stages: data preprocessing, Machine Learning (ML) classification, clustering, and model evaluation. Each of these stages plays a critical role in determining the overall effectiveness of the system. Many existing studies focus on evaluating different ML classifiers but often neglect the impact of data preprocessing. Some research compares classifiers without considering how preprocessing influences their performance, while others use a fixed preprocessing method without analyzing its effect on accuracy. Additionally, most studies prioritize binary classification, overlooking how these techniques perform in multiclass classification scenarios.

This study proposes a new approach that integrates supervised learning and clustering techniques to identify the most effective IDS configuration for IoT networks. The proposed framework systematically examines the impact of data normalization on classification accuracy across different ML models.

To begin, we used the **BoTNeTIoT-L01-v2 dataset** from [31] [33] and applied four widely used supervised ML classifiers:

- K-Nearest Neighbors (KNN)
- Random Forest (RF)
- Decision Tree (DT)
- Support Vector Machine (SVM)

Additionally, we incorporated unsupervised clustering methods:

- K-Means Clustering
- Soft Clustering

We initially evaluated these models using a confusion matrix to measure how well each classifier distinguishes between normal and malicious network activity.

In the second phase, we processed the dataset by transforming it into a normal (Gaussian) distribution to assess how standardizing the data affects classification performance. The same ML algorithms were then re-evaluated using confusion matrices to compare results before and after normalization.

# 5. Experiments

The experiments were conducted on a system equipped with an Intel® Core™ i7 processor, 16GB RAM, and a 64-bit Windows operating system. The analysis was performed using Matlab.

## 5.1. Data Collection

This study utilized the BoTNeTIoT-L01-v2 dataset, a refined version of the BoTNeTIoT dataset, which integrates network traffic data from various IoT devices. The dataset includes two primary types of botnet attacks, Mirai and Gafgyt, alongside normal traffic. It comprises 23 engineered statistical features extracted from packet capture (.pcap) files, ensuring a structured representation of network behavior over a 10-second time window. In this dataset, class labels are assigned as follows: 0 for attack instances and 1 for normal network activity.

### 5.1.1 Mirai and Gafgyt Attacks

Mirai and Gafgyt are two well-known IoT-based botnet attacks that exploit vulnerabilities in unsecured Internet of Things (IoT) devices. Both attacks have caused significant disruptions in various networks, including large-scale Distributed Denial of Service (DDoS) attacks. Mirai Botnet is a malware strain that infects IoT devices such as cameras, routers, and DVRs, and turns them into a botnet that can be controlled remotely. The attack primarily focuses on launching DDoS attacks by flooding targeted servers with massive amounts of traffic, leading to service outages. The malware exploits weak default passwords to gain control over the devices. On the other hand, Gafgyt Botnet, also known as “Bashlite,” is a family of botnets that targets IoT devices. Similar to Mirai, Gafgyt leverages weak credentials to compromise devices and use them for DDoS attacks. However, Gafgyt is also known for its smaller attack size and less sophisticated structure compared to Mirai [32]. The following table contain a comparison between these two attacks in IoT Environment.

Table 2: Comparison of Mirai and Gafgyt Botnet Attacks in IoT Environments

| **Attack Type** | **Mirai Botnet** | **Gafgyt Botnet** |
|---|---|---|
| **Primary Target** | IoT devices (IP cameras, routers, DVRs) | IoT devices (IP cameras, routers, DVRs) |
| **Exploited Vulnerability** | Weak default passwords and unsecured devices | Weak default passwords and unsecured devices |
| **Attack Method** | Large-scale DDoS using amplification techniques | DDoS using smaller-scale flood attacks |
| **Traffic Generated** | High-volume DDoS attacks | Smaller-scale DDoS attacks, primarily targeting specific devices |
| **Complexity** | High, with advanced techniques (e.g., DNS amplification) | Low to moderate, simpler attack methods |
| **Historical Impact** | Used in some of the largest DDoS attacks, including the Dyn attack in 2016 | Frequently used for smaller-scale attacks, but still effective in causing disruption |

### 5.1.2 Preprocessing

Preprocessing data is a necessary step to improve Machine Learning Model accuracy. In accordance with this research, there were two experiment settings conducted to confirm if preprocessing contributed to the variation in classification accuracy. First, the dataset was used as it was for classification. Then data were transformed to the normal pattern of distribution (nearly comparable to Gaussian) to check the contribution of normalization to the model.

Normalization was used to normalize feature values so that every attribute had a similar contribution in the classification task. This procedure was especially vital due to the statistical nature of the data as it assisted in eliminating biases presented by features whose scales varied greatly. The dataset was then separated into training and test sets in order to compare classifier performance.

### 5.1.3 ML Classifiers and Clustering Techniques

Data preprocessing is among the significant procedures to improve the performance of machine learning models. In the present study, two experimental settings were designed to study the impact of preprocessing on accuracy in classification. Initially, the dataset was utilized in its natural form for classification. Subsequently, the data were normalized into a normal distribution (i.e., more similar to Gaussian) to verify the impact of normalization on model performance. Six machine learning techniques were employed to analyze the dataset:

**Supervised Classification Models:**

- K-Nearest Neighbors (KNN): A distance-based algorithm that assigns labels based on the majority class of the nearest neighbors.
- Random Forest (RF): An ensemble learning model that constructs multiple decision trees to improve classification robustness.
- Decision Tree (DT): A rule-based classifier that partitions data based on feature values.
- Support Vector Machine (SVM): A margin-based classifier that identifies an optimal hyperplane to separate different classes.

**Unsupervised Clustering Methods:**

- K-Means Clustering: A centroid-based algorithm that groups data into K clusters by minimizing intra-cluster variance.
- Soft Clustering: A probabilistic clustering method that assigns data points to multiple clusters with varying degrees of membership.

Each classifier was evaluated before and after normalization to determine how preprocessing influenced classification performance. The effectiveness of the machine learning models was assessed using the confusion matrix, and the accuracy derived from the following formula, before and after data normalization.

**Accuracy:** The proportion of correctly classified instances among all samples

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

SVM achieved 100% accuracy both in pre-normalization and post-normalization steps. RF and DT achieved nearly 92% post-normalization, while KNN's performance was low, indicating that distance-based systems are not necessarily likely to benefit from normalization. The results show the significant role played by data preprocessing in the IDS workflow. The next section provides a detailed discussion.

## 6. Results and Discussion

This section presents and analyzes the experimental results obtained from applying various supervised machine learning (ML) algorithms to the BoTNeTIoT-L01-v2 dataset. The analysis compares the performance of the selected algorithms before and after normalizing the dataset to approximate a Gaussian distribution. The effect of normalization on classification accuracy is discussed, along with a detailed examination of the confusion matrix for each algorithm. Tables 3 and 4 summarize the obtained results.

Table 3: Pre-Normalization Results

| **Algorithm** | **Accuracy** | **Confusion Matrix (Normal, Gafgyt, Mirai)** |
|---|---|---|
| **KNN** | 100% | 55259 0 0<br>0 2837711 561<br>0 2293 3666109 |
| **Random Forest (RF)** | 71.2% | 222373 0 333559<br>0 1135309 1702963<br>0 2 3668400 |
| **Decision Tree (DT)** | 90.6% | 555932 0 0<br>0 2838272 0<br>0 3 3668399 |
| **Support Vector Machine (SVM)** | 99.9% | 555932 0 0<br>0 2835936 2336<br>0 2074 3666328 |
| **K-mean Cluster** | 71.8% | 0 01991 553941<br>0 162522 2675750<br>0 1151 3667251 |
| **Soft-cluster** | 68.4% | 0 45076 510856<br>759998 2631 2075643<br>510084 1575713 1582605 |

Table 4: Post-Normalization Results

| **Algorithm** | **Accuracy** | **Confusion Matrix (Normal, Gafgyt, Mirai)** |
|---|---|---|
| **KNN** | 86.8% | 555932 0 0<br>0 1907163 931109<br>0 0 3668402 |
| **Random Forest (RF)** | 92.1% | 555932 0 0<br>0 2623508 214764<br>0 328668 3339734 |
| **Decision Tree (DT)** | 92.4% | 555932 0 0<br>0 2623522 214750<br>0 328316 3340086 |
| **Support Vector Machine (SVM)** | 100% | 555932 0 0<br>0 2832579 5693<br>0 2941 3665461 |
| **K-mean Cluster** | 63% | 34966 493701 27265<br>1205 2835679 1388<br>0 2093373 1575029 |
| **Soft-cluster** | 57.7% | 34966 493701 27265<br>1205 2835679 1388<br>0 2093373 1575029 |

Impact of Normalization on Accuracy:

- SVM: Achieved the highest accuracy both before and after normalization, with a minor improvement in accuracy post-normalization, reaching 100%. This suggests that normalization had little impact on the SVM model, which shows its robustness in handling the dataset.
- DT and RF: Both algorithms showed significant improvements after normalization, with accuracy increasing from 90.6% to 92.4% for DT and from 71.2% to 92.1% for RF. This indicates that the feature scaling benefits tree-based classifiers and improves their performance.
- KNN: The accuracy of the KNN model dropped from 100% to 86.8% after normalization. This decline suggests that distance-based classifiers like KNN are sensitive to feature scaling, and the Gaussian-like distribution after normalization may have disrupted the distance measurements.
- K-mean Cluster: This model showed a reduction in accuracy after normalization, decreasing from 71.8% to 63%. It highlights that clustering models may not always benefit from normalization, as they rely on the distribution of data in ways that might be altered by scaling.

- Soft-cluster: Similarly, Soft-cluster showed a significant decrease in accuracy post-normalization, from 68.4% to 57.7%. The results suggest that certain types of clustering algorithms may be less effective when the dataset is normalized.

The results highlight the superiority of SVM in detecting botnet attacks in IoT networks, with its classification accuracy remaining almost perfect throughout the experiment. Both Decision Tree and Random Forest classifiers significantly benefited from normalization, while KNN and clustering techniques (K-mean and Soft-cluster) showed a decline in performance post-normalization. When compared to other existing methods, our approach demonstrated superior accuracy, especially in detecting specific attack types like Mirai and Gafgyt.

This study emphasizes the critical role of data preprocessing in intrusion detection systems (IDS) for IoT environments and provides valuable insights into choosing the best classifier based on dataset characteristics and preprocessing techniques.

## 7. CONCLUSION AND FUTURE WORK

This study explored the impact of dataset normalization on the performance of several supervised machine learning classifiers applied to the BoTNeTIoT-L01-v2 dataset for IoT botnet attack detection. The results highlighted the significance of preprocessing, especially normalization, on the accuracy of classifiers. Key findings include:

- Support Vector Machine (SVM) demonstrated the highest performance, achieving near-perfect accuracy both before and after normalization, making it the most robust model for anomaly detection in this scenario.
- Decision Tree (DT) and Random Forest (RF) classifiers showed significant improvements in accuracy after normalization, indicating that feature scaling positively influences tree-based models.
- K-Nearest Neighbors (KNN), however, exhibited a decrease in accuracy post-normalization, suggesting that distance-based classifiers are highly sensitive to feature scaling.
- K-Means and Soft-cluster methods showed moderate results, with accuracy generally lower than supervised models, especially after normalization.

These findings underscore the importance of choosing the right classifier for different types of attack detection in IoT networks and the pivotal role of preprocessing techniques like normalization. The results can be used to guide the development of more effective intrusion detection systems (IDS) for IoT environments, providing better detection capabilities and more reliable performance.

While this study focused on the impact of normalization on the classification performance, future work could explore several directions:

- Exploring Other Normalization Techniques: Investigating alternative normalization and preprocessing techniques could provide further insight into their effect on classifier performance, especially in handling complex IoT data.

- Hybrid Models: Future research could examine hybrid approaches, combining multiple machine learning models or integrating unsupervised learning techniques with supervised classifiers to improve detection accuracy.
- Real-Time Application: Implementing the models in real-world IoT environments for real-time attack detection would be valuable. Testing the classifiers in diverse, dynamic networks could help understand their scalability and adaptability to different attack scenarios.
- Deep Learning: Leveraging deep learning techniques, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), could be explored to detect complex patterns in IoT botnet attacks and enhance the accuracy and speed of detection.
- Evaluation Metrics Expansion: Expanding the evaluation metrics beyond accuracy, such as precision, recall, and F1-score, can provide a more comprehensive understanding of classifier performance in detecting both normal and attack data.

Through these approaches, future studies can further improve the efficacy of intrusion detection systems, ensuring that IoT networks remain secure against evolving botnet threats.

## ACKNOWLEDGEMENTS

This research is supported by the Deanship of Scientific Research and Innovation, at Al-Zaytoonah University of Jordan (Project No. 36/13/2024– 2025).

# References

[1] Khraisat, A., & Alazab, A. (2021). A critical review of intrusion detection systems in the internet of things: techniques, deployment strategy, validation strategy, attacks, public datasets and challenges. *Cybersecurity*, 4, 1-27.

[2] Mohammed, M. S., & Talib, H. A. (2024). Using machine learning algorithms in intrusion detection systems: a review. *Tikrit Journal of Pure Science*, 29(3), 63-74.

[3] Paulauskas, N., & Auskalnis, J. (2017, April). Analysis of data pre-processing influence on intrusion detection using NSL-KDD dataset. *In 2017 open conference of electrical, electronic and information sciences (eStream)* (pp. 1-5). IEEE.

[4] Atul, D. J., Kamalraj, R., Ramesh, G., Sankaran, K. S., Sharma, S., & Khasim, S. (2021). A machine learning based IoT for providing an intrusion detection system for security. *Microprocess. Microsystems*, 82, 103741.

[5] Zhang, C., Jia, D., Wang, L., Wang, W., Liu, F., & Yang, A. (2022). Comparative research on network intrusion detection methods based on machine learning. *Computers & Security*, 121, 102861.

[6] Saran, N., & Kesswani, N. (2023). A comparative study of supervised Machine Learning classifiers for Intrusion Detection in Internet of Things. *Procedia Computer Science*, 218, 2049-2057.

[7] Saranya, T., Sridevi, S., Deisy, C., Chung, T. D., & Khan, M. A. (2020). Performance analysis of machine learning algorithms in intrusion detection system: A review. *Procedia Computer Science*, 171, 1251-1260.

[8] Baich, M., Hamim, T., Sael, N., & Chemlal, Y. (2022). Machine Learning for IoT based networks intrusion detection: a comparative study. *Procedia Computer Science*, 215, 742-751.

[9] Ahmad, M., Riaz, Q., Zeeshan, M., Tahir, H., Haider, S. A., & Khan, M. S. (2021). Intrusion detection in internet of things using supervised machine learning based on application and transport layer features using UNSW-NB15 data-set. *EURASIP Journal on Wireless Communications and Networking*, 2021, 1-23.

[10] Hammad, M., El-Medany, W., & Ismail, Y. (2020, December). Intrusion detection system using feature selection with clustering and classification machine learning algorithms on the unsw-nb15 dataset. *In 2020 international conference on innovation and intelligence for informatics, computing and technologies (3ICT)* (pp. 1-6). IEEE.

[11] Zwane, S., Tarwireyi, P., & Adigun, M. (2018, December). Performance analysis of machine learning classifiers for intrusion detection. *In 2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC)* (pp. 1-5). IEEE.

[12] Saba, T., Rehman, A., Sadad, T., Kolivand, H., & Bahaj, S. A. (2022). Anomaly-based intrusion detection system for IoT networks through deep learning model. *Computers and Electrical Engineering*, 99, 107810.

[13] Hussein, A. Y., & Sadiq, A. T. (2022). Meerkat clan-based feature selection in random forest algorithm for IoT intrusion detection. *Iraqi Journal of Computers, Communications, Control and Systems Engineering*, 22(3), 15-24.

[14] Hukkeri, G. S., Ankalaki, S., Goudar, R. H., & Hadimani, L. (2024). The Impact of Protocol Conversions in the Wireless Communication of IOT Network. *International Journal of Advances in Soft Computing & Its Applications*, 16(1).

[15] Benhmad, T., Rhaimi, C. B., Alomari, S., & Aljuhani, L. (2024). Design and Implementation of an Integrated IoT and Artificial Intelligence System for Smart Irrigation Management. *International Journal of Advances in Soft Computing & Its Applications*, 16(1).

[16] Alia, M., Jaradat, Y., Masoud, M., Swais, K., Manasrah, A., & others. (2024). Low-cost IoT-based charging management system for electric vehicles: Design guidelines. *International Journal of Advances in Soft Computing and Its Applications*, 16(3), Article 4. https://doi.org/10.15849/IJASCA.240330.04.

[17] Masoud, M., Jaradat, Y., Manasrah, A., & Jannoud, I. (2019). Sensors of smart devices in the Internet of Everything (IoE) era: Big opportunities and massive doubts. *Journal of Sensors*, 2019, 6514520. https://doi.org/10.1155/2019/6514520.

[18] Bhatia, V., Choudhary, S., Ramkumar, K.R. (2020). A comparative study on various intrusion detection techniques using machine learning and neural network. *In 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), IEEE*, pp. 232-236. https://doi.org/10.1109/ICRITO48877.2020.9198008.

[19] Kathiresan, V., Karthik, S., Divya, P., Rajan, D.P. (2022). A comparative study of diverse intrusion detection methods using machine learning techniques. *In 2022 International Conference on Computer Communication and Informatics (ICCCI), IEEE*, pp. 1-6. https://doi.org/10.1109/ICCCI54379.2022.9740744

[20] Boyanapalli, A., Shanthini, A. (2021). A comparative study of techniques, datasets and performances for intrusion detection systems in IoT. *In Artificial Intelligence Techniques for Advanced Computing Applications: Proceedings of ICACT 2020, Springer Singapore*, pp. 225-236. https://doi.org/10.1007/978-981-15-5329-5_22

[21] Anushiya, R., Lavanya, V.S. (2021). A comparative study on intrusion detection systems for secured communication in internet of things. *ICTACT Journal on Communication Technology*, 6948: 2527-2537. https://doi.org/10.21917/ijct.2021.0373

[22] Sangwan, U., Chhillar, R.S. (2022). Comparison of various classification techniques in cyber security using Iot. *International Journal of Intelligent Systems and Applications in Engineering*, 10(3): 334-339.

[23] Agrawal, A.P., Singh, N. (2021). Comparative analysis of SVM kernels and parameters for efficient anomaly detection in IoT. *In 2021 5th International Conference on Information Systems and Computer Networks (ISCON), IEEE*, pp. 1-6.

https://doi.org/10.1109/ISCON52037.2021.9702398

[24] Aswal, K., Dobhal, D.C., Pathak, H. (2020). Comparative analysis of machine learning algorithms for identification of BOT attack on the internet of vehicles (IoV*). In 2020 International Conference on Inventive Computation Technologies (ICICT), IEEE*, pp. 312-317. https://doi.org/10.1109/ICICT48043.2020.9112422

[25] Ibrahim, Z.K., Thanon, M.Y. (2021). Performance comparison of intrusion detection system using three different machine learning algorithms. *In 2021 6th International Conference on Inventive Computation Technologies (ICICT), IEEE*, pp. 1116-1124. https://doi.org/10.1109/ICICT50816.2021.9358775

[26] Mondal, B., Singh, S.K. (2022). A comparative analysis of network intrusion detection system for iot using machine learning. *In Internet of Things and Its Applications: Select Proceedings of ICIA 2020, Singapore: Springer Nature Singapore*, 825: 211-221. https://doi.org/10.1007/978-981-16-7637-6_19

[27] Manvith, V.S., Saraswathi, R.V., Vasavi, R. (2021). A performance comparison of machine learning approaches on intrusion detection dataset. *In 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), IEEE*, pp. 782-788.

https://doi.org/10.1109/ICICV50876.2021.9388502

[28] Dwibedi, S., Pujari, M., Sun, W.Q. (2020). A comparative study on contemporary intrusion detection datasets for machine learning research. *In 2020 IEEE International Conference on Intelligence and Security Informatics (ISI), IEEE,* pp. 1-6.

https://doi.org/10.1109/ISI49825.2020.9280519

[29] Kilincer, I.F., Ertam, F., Sengur, A. (2021). Machine learning methods for cyber security intrusion detection: Datasets and comparative study. *Computer Networks*, 188: 107840.

https://doi.org/10.1016/j.comnet.2021.107840

[30] Al Fayoumi, M., Odeh, A., Keshta, I., Aboshgifa, A., AlHajahjeh, T., Abdulraheem, R. (2022). Email phishing detection based on naïve Bayes, random forests, and SVM classifications: A comparative study. *In 2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC), IEEE*, pp. 0007-0011. https://doi.org/10.1109/CCWC54503.2022.9720757

[31] IOT Dataset, https://www.kaggle.com/datasets/azalhowaide/iot-dataset-for-intrusion-detection-systems-ids

[32] Aldawod, R., Alsaleh, N., Aldalbahi, N., Alqahtani, R., & Sakri, S. (2022, December). Smart prediction system for classifying MIRAI and GAFGYT attacks on IOT devices. In 2022 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 1216-1222). IEEE.

[33] Jawad, M. A., García, J., & Masoud, M. (2026). A survey of Network Intrusion Detection Systems (NIDS) based on Machine Learning Algorithms for Industrial Internet of Things (IIoT). Lecture Notes in Networks and Systems, Springer, 1–12. https://doi.org/10.1007/978-3-031-96631-6_16

[34] Jawad, M. A., Masoud, M. Z., Alesanco, Á., & García, J. (2025). A distributed brute-force attack and misleading countermeasure for securing ModbusTCP implementations in PLCs. *Proceedings of the 12th International Conference on Information Technology: Innovation Technologies (ICIT 2025),* 1–8. IEEE. https://doi.org/10.1109/ICIT64950.2025.11049239

## Notes on contributors

***Mohammad AbdulMohdi AbdulJawad*** is a faculty member at Al-Zaytoonah University of Jordan, serving as the Head of the Councils Secretariat and the Director of the Follow-Up and Accreditation of Scientific Journals Office. He holds a M.Sc. in Computer Science and a B.Sc. in Computer Information Systems from Al-Zaytoonah University of Jordan. His work focuses on Artificial Intelligence, Genetic Algorithms, Networks, E-Learning, Networks and Cybersecurity, with several funded projects and publications in International Journals and Conferences, in addition to professional experience in academia and software development.

***Amal Qassed Mahmoud Ahmad*** is an Instructor and Deputy Director of the Accreditation and Quality Assurance Office at Al-Zaytoonah University of Jordan. She holds a M.Sc. in Computer Engineering and Networks from the University of Jordan. Her work focuses on quality assurance, academic development, and research in data security, E-Learning, network performance analysis, and machine learning, with several publications in International Journals and Conferences.