**International Journal of Applied Mathematics**

**Volume 38 No. 1s, 2025**

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

# BEHAVIORAL ANALYSIS OF MALICIOUS ACTIVITIES IN NETWORK TRAFFIC USING MACHINE LEARNING ALGORITHMS

**Naved Raza Q. Ali[1], Dr. Sandhya Arora[2], Dr. Parikshit N. Mahalle[3]**

[1]Research Scholar, Smt. Kashibai Navale College of Engineering, Savitribai Phule Pune University, Maharashtra, India
[2]Cummins College of Engineering for Women, Savitribai Phule Pune University, Maharashtra, India
[3]Vishwakarma Institute of Technology, Savitribai Phule Pune University, Maharashtra, India
nvdrazaqali@gmail.com, sandhya.arora@cumminscollege.in, aalborg.pnm@gmail.com

**ABSTRACT**

Ensuring network security is a critical concern in today's increasingly digital world, requiring sophisticated methods to identify malicious activities. This research employs machine learning methodologies, specifically 'Random Forest, Decision Tree, Naïve Bayes, and AdaBoost' algorithms, to conduct a comprehensive behavioral evaluation of malicious activities in network traffic. The study underscores the critical importance of data sampling in datasets for augmenting the effectiveness of intrusion detection systems (IDS). The dataset utilized for experimentation includes the CICIDS2017 and CICIoT2023 datasets, offering a varied and authentic portrayal of network traffic situations. Using Random Forest and Decision Tree algorithms helps model complex relationships in the data, while Naïve Bayes and AdaBoost algorithms add diversity to the analysis. The study examines how data sampling techniques affect the performance of algorithms, highlighting the significance of a carefully curated dataset in developing strong intrusion detection models. The experiment results reveal the pros and cons of each algorithm, providing insight into their efficacy in detecting malicious activities. This study's results contribute significantly to the advancement of behavioral analysis in intrusion detection by providing recommendations on algorithm choice and data preprocessing methods to achieve optimal performance in practical network security situations. The results highlight the important function of machine learning in strengthening network security and offer useful suggestions for creating strong IDS systems.

**Keyword:** Intrusion Detection Systems (IDS), Machine Learning Algorithms, Behavioral Analysis, Network Security, Internet of Things (IoT), Data Sampling, Cyber Threats.

Received: July 19, 2025

International Journal of Applied Mathematics

Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

## 1. INTRODUCTION

In today's rapidly changing information technology environment, where digital connectivity is widespread, it is crucial to strengthen network security against malicious activities. The rise of cyber threats, which are becoming more complex and adaptable, demands the development and implementation of advanced IDS Systems that can accurately detect security breaches [1]. This study delves into a thorough examination of behavioral analysis in network security, utilizing machine learning algorithms to identify patterns that suggest malicious activity in network traffic. This study is motivated by the necessity for a proactive and adaptive intrusion detection approach due to cyber adversaries constantly changing their tactics to exploit vulnerabilities and compromise digital systems [2].

Cyber threats have evolved from basic, easily identifiable attacks to more complex and elusive strategies. It is crucial for intrusion detection mechanisms to advance in response to adversaries using sophisticated methods like polymorphic malware and zeroday exploits. Conventional signature-based IDS Systems demonstrate strong defenses against familiar attacks, however, they face challenges in keeping pace with the rapidly evolving terrain of contemporary cyber attacks [3]. Behavioral analysis represents a promising approach that concentrates on detecting abnormal patterns and deviations from typical network behavior. Incorporating ML algorithms into IDS enables a more flexible and intelligent method, enabling the immediate examination of network traffic and identification of novel threats [4].

This research focuses on developing more accurate and efficient methods for detecting network intrusions through examining four key machine learning methodologies as 'Random Forest, Decision Tree, Naïve Bayes, and AdaBoost'. Every algorithm has its own distinct strengths and methods for modeling complex relationships in network traffic data. Random Forest is particularly adept at managing datasets with many dimensions and is resistant to overfitting. The Decision Tree offers a clear depiction of decision-making processes due to its tree-like structure. Naïve Bayes, utilizing probabilistic reasoning, provides a straightforward and efficient approach. AdaBoost utilizes an iterative boosting method to merge the capabilities of base classifiers and create a robust ensemble classifier [5]. This paper represents a systematic assessment of the performance of these algorithms in identifying network security threats, through a careful examination of their results.

The efficacy of machine learning-based IDS systems is profoundly affected by the reliability and integrity of the training dataset. The findings underscore the significance of data sampling methods in influencing the strength of IDS system. Selecting the right dataset is vital because it profoundly influences the algorithm's ability to recognize and detect diverse types of malicious behavior. Two datasets, the CICIDS2017 and CICIoT2023 datasets, are used for a thorough evaluation in this research. The datasets offer a varied and comprehensive collection of network traffic scenarios, providing an authentic depiction of the difficulties faced in real-world cybersecurity. Integrating data sampling considerations into the analysis improves the

Received: July 19, 2025

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

practical significance of the research findings, emphasizing the necessity of a well-maintained dataset for developing efficient intrusion detection models [6].

The 'Canadian Institute for Cybersecurity' developed the CICIDS2017 dataset., contains labeled. network traffic data covering different attack scenarios. Using this dataset enables the researchers to assess the algorithms in various attack scenarios, ensuring a comprehensive evaluation of their resilience. The CICIoT2023 dataset, created to simulate Internet of Things (IoT) settings, adds an extra level of intricacy. Adding this dataset expands the study's scope, recognizing the changing environment of network-connected devices and the distinct challenges they present for intrusion detection.

As the research progresses, a crucial part of the study involves evaluating how data sampling methods affect the effectiveness of the selected machine learning algorithms [7]. Data sampling techniques like oversampling, under-sampling, and synthetic data generation are important for dealing with class imbalances, which are typically prevalent in intrusion detection datasets due to the prevalence of legitimate network traffic compared to malicious instances. The research seeks to analyze the impact of various data sampling methods on algorithmic performance to find the optimal balance between accurate detection and real-world dataset limitations [8], [9].

This research. aims to significantly enhance the field. of intrusion. detection. by thoroughly examining behavioral analysis in network security. The study has three primary objectives: First, to provide practical recommendations for developing effective IDS system by exploring the capabilities of. ML. methodologies, like 'Random Forest, Decision Tree, Naïve Bayes, and AdaBoost'. Second, to examine data sampling intricacies using the CICIDS2017 and CICIoT2023 datasets. Third, to provide cybersecurity practitioners with valuable insights to effectively combat the challenges posed by modern adversaries as the cyber threat landscape continues to change.

## 2. Literature Review

The The literature on cybersecurity has witnessed a surge in research endeavors addressing the multifaceted challenges associated with data security and intrusion detection. Recent work [10] emphasizes the critical need for securing data transfers within the IoT, acknowledging the vulnerability of IoT networks to cyber threats. In a related vein, the study [11] delves into reversible data hiding techniques, employing histogram and prediction error mechanisms to safeguard secure data. Complementing this, research [12] introduces a CNN for detecting potential concealed data. in images of spatial. domain., highlighting the role of advanced neural network architectures in uncovering concealed information. in pictures of the spatial domain.

In terms of IDSs, a novel approach is presented in [13] exploring the fusion of kernel methods and extreme learning machines, demonstrating the potential for improved detection accuracy. Federated Learning (FL) takes center stage in the study by Agrawal et al. [14], where FL is applied to intrusion detection systems. This innovative approach leverages collaborative learning across distributed nodes to effectively mitigate the complexities arising from diverse

Received: July 19, 2025

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

and geographically dispersed datasets pertaining to threat detection. Building upon this theme, Otoum et al. [15] introduce DLIDS, a DL based threat detection framework developed specifically to safeguard IoT environments. The integration of deep learning techniques signifies a shift towards more adaptive and context-aware intrusion detection systems.

Hybrid approaches are explored in [5], where a hybrid approach integrates deep learning with binary techniques for anomaly categorization for optimization within the intrusion detection system. This innovative fusion demonstrates the potential for leveraging diverse algorithmic strengths to enhance overall detection capabilities. Machine learning algorithms take center stage in the work by Abdulmajeed et al. [16], providing a comprehensive exploration of various techniques and datasets essential to contemporary IDS designing. This survey offers key perspectives on the landscape of ML applications in intrusion recognition, informing the selection of appropriate algorithms for specific security requirements.

Network level IDS based on Software Defined Networking are surveyed in [17], highlighting the growing significance of SDN in enhancing the flexibility and responsiveness of IDS. The study explores ML approaches within the SDN framework, emphasizing the potential for dynamic and context aware security solutions. A broader perspective on ML and DL methods for IDS are presented in [13], providing an in-depth examination of current developments and challenges in intrusion detection research. The work identifies the importance of ongoing research to address emerging threats and adapt intrusion detection mechanisms to evolving cyber landscapes.

Recent research in cybersecurity has focused on detecting cyberattacks through the analysis of abnormal network traffic behavior. Statistical techniques have been employed to identify deviations in network traffic, aiding in the early detection of potential cyber threats [18]. Further studies have explored the behavior of malicious software, categorizing different detection techniques while addressing emerging trends and challenges in this domain [19]. To facilitate the discernment of malicious and benign users, user behavior analysis has increasingly incorporated ML methodologies and data analytical techniques, enhancing the precision of detection systems [20]. Another approach involves analyzing DNS behavior to uncover malicious activities, with efforts to generate benchmark datasets through application layer traffic analysis [21]. Also, advanced methods such as combining multi-head self-attention and graph CNN mechanisms have been proposed to boost the efficacy of malicious traffic detection, showcasing the potential for more robust security systems [22], [23].

The literature review delves into a specific application of machine learning, wherein a tailored ML method is employed to. analyze. the ‘NSL-KDD dataset’ to examine malware and threat recognition [24], [25], [26]. This. research. showcases the implementation of ML. methodologies on authentic datasets, providing a deeper understanding of the performance and applicability of intrusion detection algorithms. As we navigate the diverse landscape of literature in cybersecurity and intrusion detection, these works collectively contribute to the

Received: July 19, 2025

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

ongoing discourse, offering a rich tapestry of methodologies, applications, and insights that collectively shape the trajectory of research in this critical domain.

## 3. Proposed Methodology

In the proposed work, the main objective is to analyze malicious activities in network traffic. Machine learning techniques are crucial for enhancing the performance of IDS. The study utilizes four potential classifiers – 'Random Forest, Decision Tree, Naïve Bayes, and AdaBoost' to identify patterns that suggest malicious behavior in network traffic.

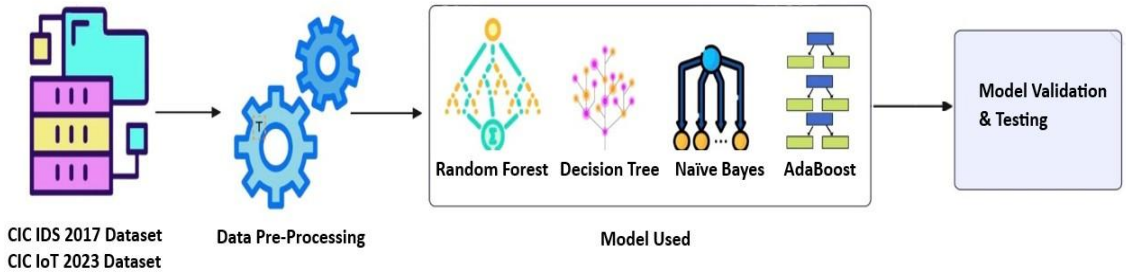



Figure 1. Proposed methodology for detection of malicious activities

The data sets available in online repositories are typically raw and contain dirty data, which must be cleaned before being provided to machine learning algorithms. It is important to perform Exploratory data analysis [14] and data preprocessing.

In the view of this, the first step in the proposed methodology is data acquisition as shown in Figure 1. Data preprocessing is the second step, where operations like data cleaning are performed, including removal of null values, missing values, and inconsistent values. Then, the data is analyzed again to understand it syntactically and semantically. Hereafter, visualization is carried out to generate narratives and understand prominent features of the dataset (timestamps, flow duration, Header Length, Protocol Type, Duration, HTTP, HTTPS, DNS, Telnet, SMTP, Number and Weight) that contribute to the objective. The visualization also helps this study understand the behavioral analysis of the dataset prior to model building.

Subsequently, sampling of the dataset is carried out to analyse a subset of the data and uncover meaningful information as well as achieve dimensionality reduction. Hereafter, train-test splitting is carried out as part of the proposed methodology, with standard splitting used in the study 80% training and 20% testing. The last stage involves building the machine learning model on this proposed dataset, followed by model validation and testing.

### 3.1. Datasets

Selecting the right dataset is essential because it has a profound effect on the algorithm's ability to learn patterns and detect various types of malicious behavior. Two datasets, the CICIDS2017 and CICIoT2023 datasets, are used for a thorough evaluation in this research. The datasets offer

Received: July 19, 2025

International Journal of Applied Mathematics

Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

a varied and comprehensive collection of network traffic scenarios, providing an authentic depiction of the difficulties faced in real-world cybersecurity.

The CICIDS2017 dataset constitutes an exhaustive repository of network traffic data, specifically tailored for advancing research in intrusion detection systems (IDS). This dataset includes network traffic captures of both benign and malicious activities, consisting of 28,30,743 number of rows (approximately 2.8 million) and 80 features (columns). The proposed work used the 2GB size of data set as csv file. 77 features extracted from network traffic captures (e.g., source IP, port numbers, destination IP, packet sizes, timestamps, etc.), 2 labels (benign or malicious) and 1 additional feature (the attack type, if malicious). Missing values are handled using imputation techniques, such as mean or median replacement, to ensure that the dataset remained comprehensive and usable. The dataset is labeled, with each record marked as either benign or malicious (including DoS, DDoS, scanning, exploitation, malware-related activities, and botnet attacks). This labeling enables supervised learning approaches and facilitates the assessment of IDS models.

The CICIoT2023 dataset provides a thorough repository of diverse network traffic patterns, specifically designed to advance the development of robust intrusion detection systems for IoT ecosystems. It has a total size of approximately 17.4 GB in CSV format. The dataset comprises 10 million network traffic flows from various loT devices with 47 features (columns) extracted from network traffic captures, including ts (timestamps), flow duration, Header Length, Protocol Type, Duration, Rate, HTTP, HTTPS, DNS, Telnet, SMTP AVG, Std, Tot size, Number, Weight, etc.

This dataset includes 2 labels (benign or malicious) and an additional feature indicating the attack type (if malicious). These attacks are defined in seven main classes, namely Reconnaissance, .Brute Force, DDoS, Web-based attacks, DoS, Mirai-based threats, and Spoofing, which collectively comprise 33 distinct attack categories. The dataset's missing values are addressed through imputation techniques, ensuring its comprehensiveness and usability. Each record is labeled, enabling supervised learning approaches, and facilitating the evaluation of IDS models for loT networks.

### 3.2. Algorithm

Machine learning methodologies are crucial for improving the performance of IDS System. This study utilizes the four prominent algorithms- ‘Random Forest, Decision Tree, Naïve Bayes, and AdaBoost’ to identify patterns that suggest malicious behavior in network traffic.

Algorithm 1. Malicious activity detection in network traffic

**Input:** CICIDS2017 and CICIoT2023 datasets

**Output:** Detection of malicious network traffic using ML models

1. **Data Acquisition**
2. Load the dataset ← CICIDS2017, CICIoT2023.
3. **Data Preprocessing**

Received: July 19, 2025

International Journal of Applied Mathematics

Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4. a. Clean the dataset:
5. - Remove ← null values, missing values, and inconsistent data.
6. - Drop irrelevant columns
7. b. Handle missing values
8. c. Encode categorical variables
9. **Data Sampling**
10. *Apply* ←data sampling techniques to balance the dataset.
11. **Data Split**
12. *Split* ← training (80%) and testing (20%) sets.
13. **Feature Selection**
14. *Select* ←key features for malicious activity detection
15. **Machine Learning Model Training**
16. a. Initialize classifiers:
17. - 'Random Forest, Decision Tree, Naïve Bayes, AdaBoost'
18. b. Train each model on the training data.
19. **Model Evaluation**
20. For each trained model:
21. - *Calculate. performance metrics* ←.Accuracy, ..Precision, .Recall, .F1-Score.
22. **Prediction**
23. **Classify** ← network traffic as malicious or normal.

---

### 3.2.1. Random Forest

This technique involves creating 'multiple decision trees' and merging their predictions to achieve higher accuracy and lower overfitting. In the context of behavioral analysis of malicious network activities, Random Forest excels at handling complex, high-dimensional datasets like CICIDS2017 and CICIoT2023. Its ability to model intricate patterns in network traffic helps in accurately detecting both known and novel malicious behaviors by analyzing diverse features simultaneously as shown in Equation 1.

$$RF(x) = \frac{1}{N}\sum_{i=1}^{N} T_i(x) \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where, $RF(x)$= "Random forest prediction output for input x", N= "no. of decision trees", $T_i(x)$= "prediction of the i$^{th}$ decision tree"

### 3.2.2. Decision Tree

This algorithm splits the data into subsets by evaluating feature values, resulting in a hierarchical, tree-like structure. It provides a simple yet effective solution for classifying network traffic into benign or malicious categories. However, Decision Tree models often plagued by overfitting, particularly with inaccurate or imbalanced datasets, which makes it less

Received: July 19, 2025

International Journal of Applied Mathematics

Volume 38 No. 1s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

robust in certain network environments, such as IoT scenarios, where the variety of attacks is broader and the data is more complex.

### 3.2.3. Naïve Bayes

This probabilistic classifier utilizes Bayes' Theorem, relying on the assumption of feature independence as formulated in equation 2. In the behavioral analysis of network traffic, this model is quick and efficient, making it suitable for real-time detection of malicious activities.

$$P(H/E) = \frac{P(E/H)*P(H)}{P(E)} \quad \text{……...……..……} (2)$$

P(H/E): Posterior probability - The probability of event H occurring given that event E has occurred.
P(E/H): The probability of event E occurring subsequent to the realization of event H.
P(H): Prior probability - represents the probability of event H occurring before the observation of event E.
P(E): Evidence probability - denotes the unconditional probability of event E, quantifying its likelihood irrespective of event H.

### 3.2.4. AdaBoost

This algorithm aggregates weak learners via ensemble methodology to enhance predictive performance. In network traffic analysis, AdaBoost iteratively focuses on misclassified instances, enhancing the detection of hard-to-identify malicious activities. By boosting weak classifiers, it becomes highly effective in detecting subtle deviations in traffic behavior, making it particularly valuable in environments like IoT where attack patterns can be varied and harder to detect. It is represented as Equation (3).

$$F(x) = \sum_{t=1}^{T} \alpha_t f_t(x) \quad \text{……} (3)$$

Where F(x)= "final prediction", $\alpha_t$= "boosting weight for the $t^{th}$ weak classifier", T= "number of boosting iterations

Algorithm 1 applies machine learning techniques to analyze and detect suspicious activities in network traffic. This approach involves data preprocessing, feature selection, and training of multiple ML techniques, such as 'Random Forest, Decision Tree, Naïve Bayes, and AdaBoost'. Each algorithm is applied to the CICIDS2017 and CICIoT2023 datasets, with the goal of improving IDS by identifying abnormal network behavior patterns and providing a thorough analysis of their performance in cyber threat detection.

## 4. Results And Discussion

The outcomes of different machine. learning. models. evaluated on the 'CICIDS2017 dataset' are presented in Table.1 and Figure.2, showcasing their performance metrics, comprising

Received: July 19, 2025

International Journal of Applied Mathematics
Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

'accuracy, precision, recall, and F1 Score'. A critical step in assessing these models requires separating the dataset into training and testing sets. This division enables the utilization of the training subset for model development and the testing subset for evaluating performance on new data. Typically, the dataset is split, with 80% allocated to training and 20% held back for testing purposes.

Table 1. Performance metrics of ML models on CICIDS2017.

| ML Models | Accuracy | Precision | Recall | F1Score |
|---|---|---|---|---|
| Random Forest | 98 | 97 | 97 | 97 |
| Decision Tree | 95 | 97 | 92 | 94 |
| Naïve Bayes | 93 | 93 | 89 | 91 |
| AdaBoost | 97 | 97 | 95 | 96 |

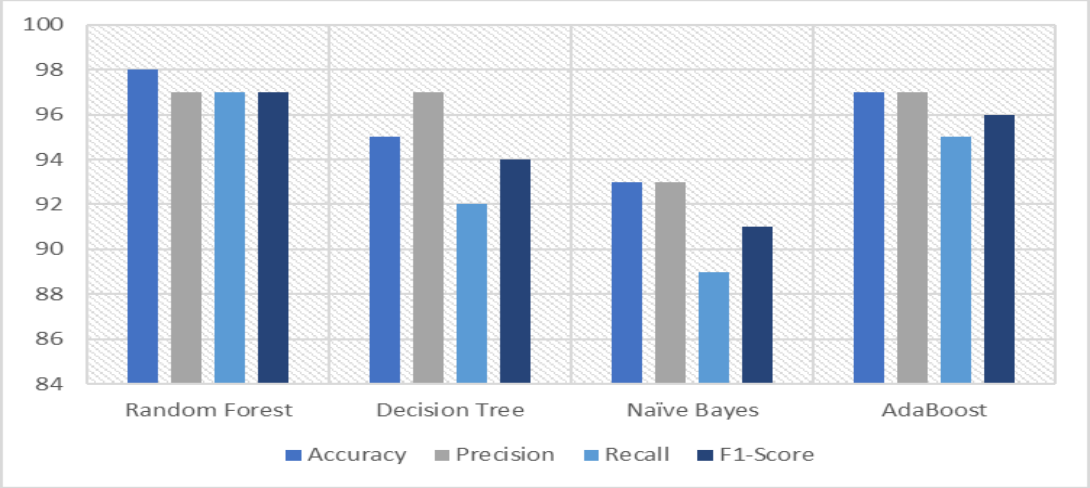



Figure 2. Evaluation parameter comparison of various ML models on CICIDS2017

The results of CICIDS2017 dataset reveal that Random Forest achieves the highest accuracy rate of 98%, complemented by precision, recall, and F1-Score values of 97%. Decision Tree follows closely, demonstrating an accuracy of 95%, and notable precision, recall, and F1-Score values of 97%, 92%, and 94%, respectively. Naïve Bayes also demonstrates a commendable performance, achieving an accuracy of 93% and corresponding precision, recall, and F1-Score values of 93%, 89%, and 91%. Furthermore, AdaBoost showcases strong performance, with an accuracy of 97% and precision, recall, and F1-Score values of 97%, 95%, and 96%, respectively.

Received: July 19, 2025

**International Journal of Applied Mathematics**

**Volume 38 No. 1s, 2025**

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

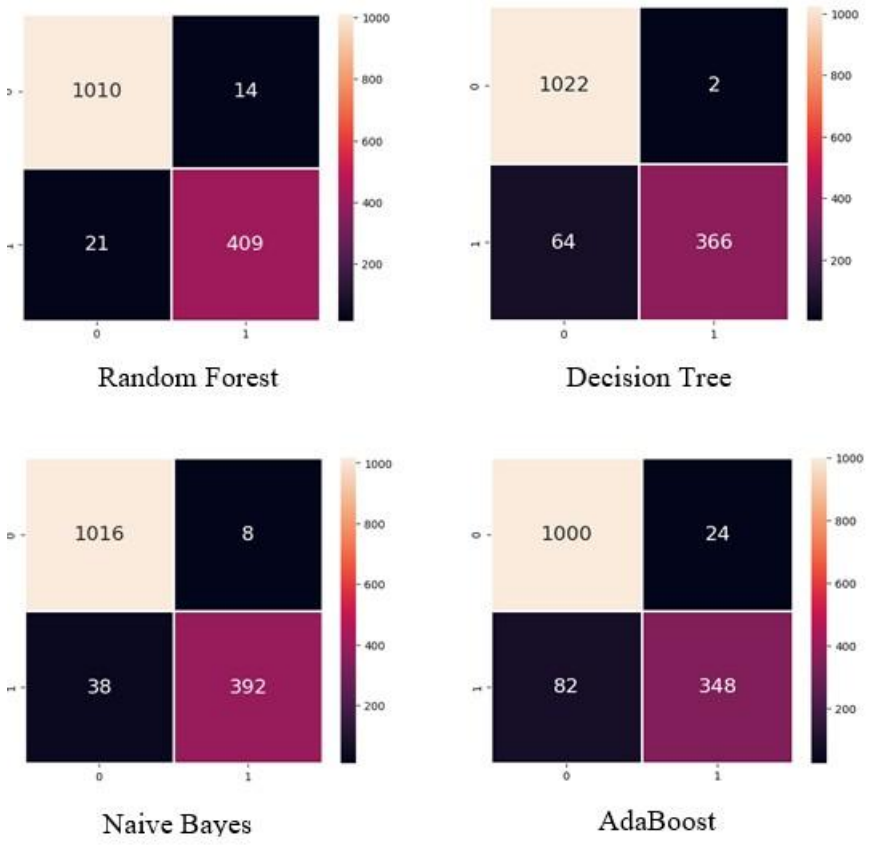



Figure 3. Confusion matrix of CICIDS2017 dataset

Figure 3 presents the confusion matrix for the machine learning algorithms evaluated on the CICIDS2017 dataset. Furthermore, the ROC curves of these algorithms are shown in Figures 4A, 4B, 4C, and 4D.

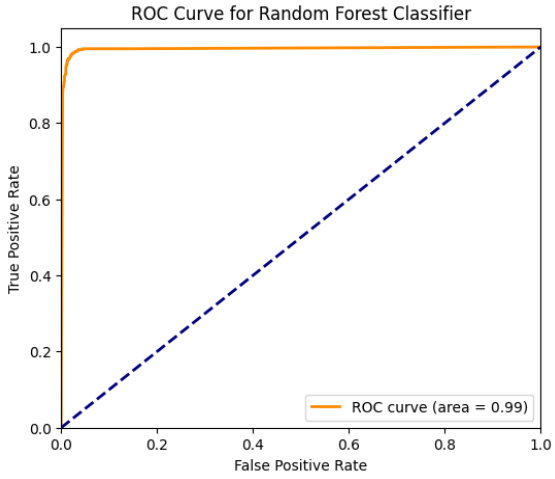



Figure 4A. ROC Curve of Random Forest

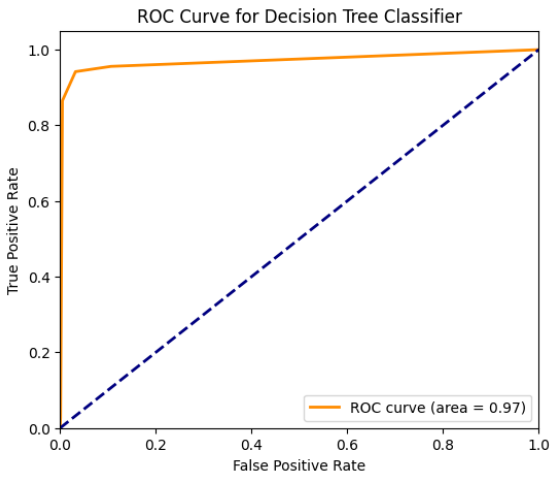



Figure 4B. .ROC. Curve of .Decision Tree

Received: July 19, 2025

**International Journal of Applied Mathematics**

**Volume 38 No. 1s, 2025**

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

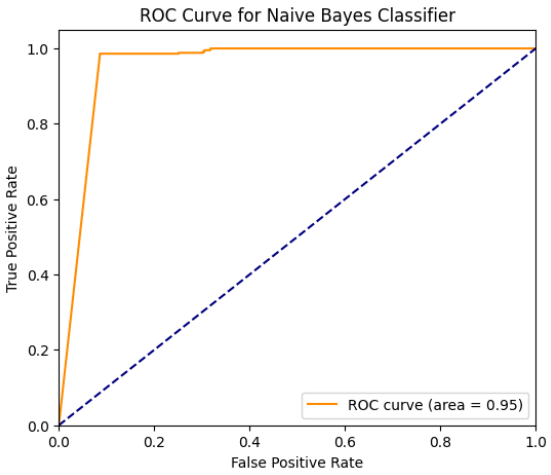



Figure 4C. .ROC Curve of Naïve Bayes

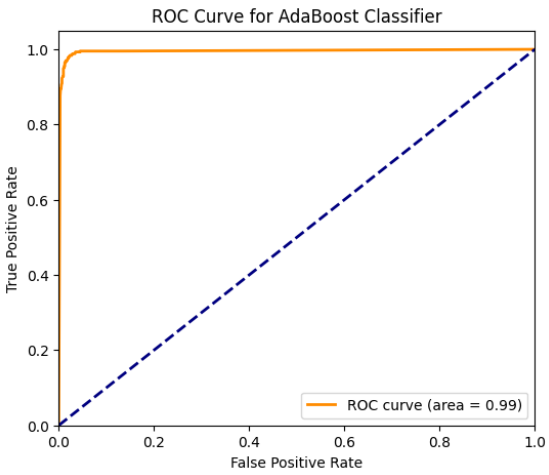



Figure 4D. ROC Curve of AdaBoost

.The performance metrics of ML models on the CICIoT2023 dataset.are presented in Table 2 and illustrated in Figure 5. These results provide comprehensive comparison of the models' .accuracy, precision, recall, and .F1-Score.

Table 2. Performance metrics of ML models on CICIoT2023

| ML Models | Accuracy | Precision | Recall | F1Score |
|---|---|---|---|---|
| Random Forest | 99 | 99 | 99 | 99 |
| Decision Tree | 54 | 42 | 54 | 45 |
| Naïve Bayes | 88 | 91 | 88 | 88 |
| AdaBoost | 97 | 97 | 95 | 96 |

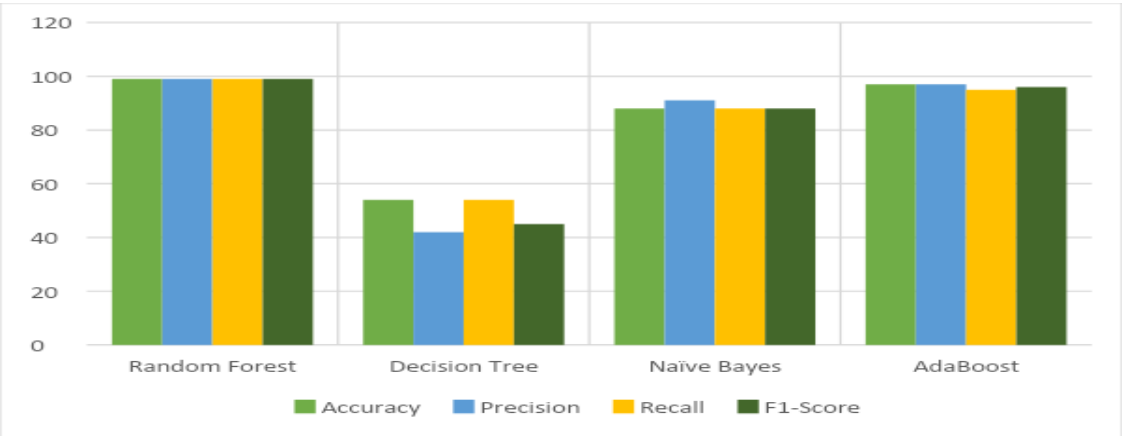



Figure 5. Evaluation parameter comparison of various ML models on CICIoT2023

Received: July 19, 2025

**International Journal of Applied Mathematics**

**Volume 38 No. 1s, 2025**

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Transitioning to the CICIoT2023 dataset, Random Forest establishes itself as the standout performer, achieving a remarkable accuracy of 99%, along with precision, recall, and F1Score all at 99%. In contrast, Decision Tree faces challenges on this dataset, recording lower metrics with an accuracy of 54%, precision of 42%, recall of 54%, and F1Score of 45%. Naïve Bayes demonstrates robust performance with an accuracy of 88%, precision of 91%, recall of 88%, and F1Score of 88%. AdaBoost maintains its strong performance, consistent with the results on the CICIDS2017 dataset, with an accuracy of 97%, precision of 97%, recall of 95%, and F1Score of 96%. Figure 6 shows the ROC curves of machine learning algorithms on the CICIoT2023 dataset.

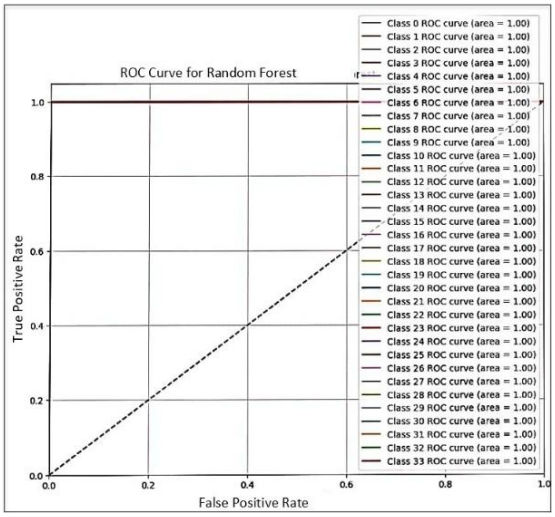



Figure 6A. Random Forest ROC Curve

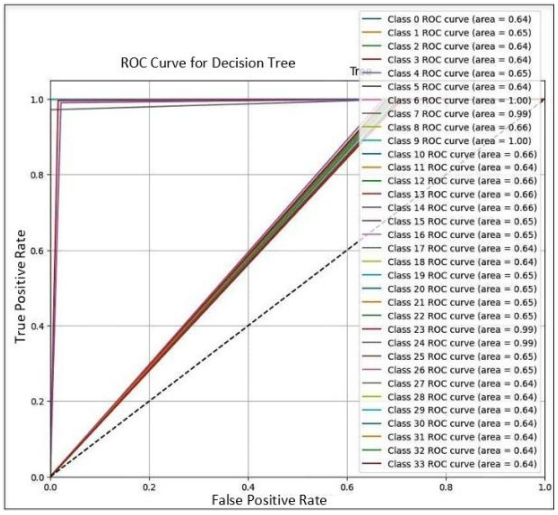
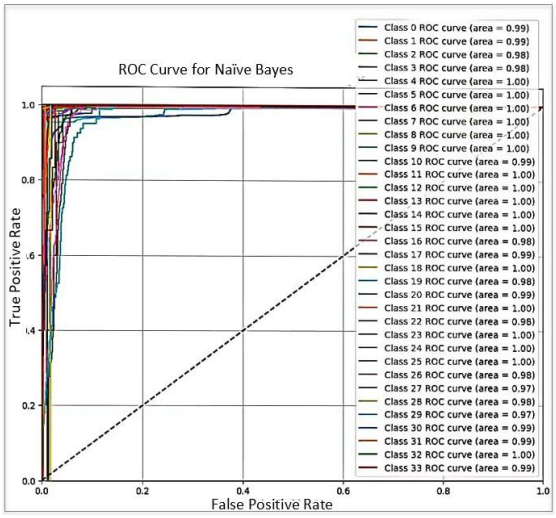



Figure 6B. Decision Tree ROC Curve.




.Figure 3C. .Naïve Bayes .ROC Curve

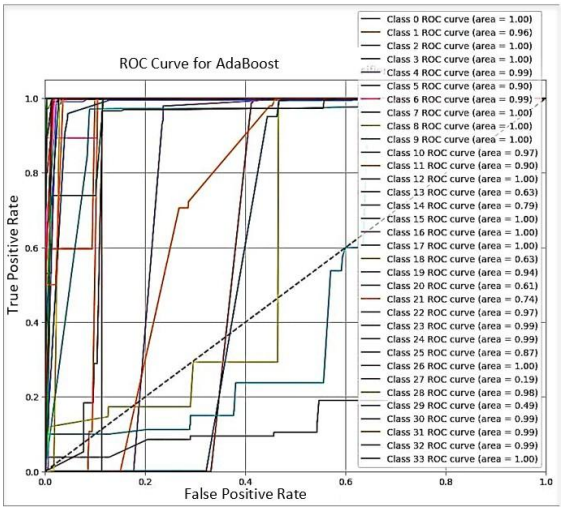



Figure 3D. AdaBoost ROC Curve.

Received: July 19, 2025

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

These results underscore the importance of dataset-specific considerations when deploying machine intelligence models for intrusion detection. While certain algorithms consistently perform well across both datasets, the variable performance of others highlights the need for customized algorithm choice based on the distinct attributes of the dataset in use. The robust accuracy and precision metrics achieved by Random Forest and AdaBoost across both datasets position them as promising candidates for intrusion detection applications, warranting further exploration and consideration in real-world cybersecurity implementations.

## 5. Conclusion and Future Work

This research focused on behavioral analysis for detecting malicious activities in network traffic using various predictive models, such as *Random* 'Forest, Decision Tree, Naïve Bayes, and AdaBoost'. Furthermore, testing these algorithms on the CICIDS2017 and CICIoT2023 datasets showed detailed performances, highlighting the significance of dataset-specific factors. Notably, Random Forest and AdaBoost consistently demonstrated strong performance in both datasets, emphasizing their potential for effective intrusion detection. Additionally, this research highlights the importance of data sampling methods in influencing the efficiency of intrusion detection models and stresses the necessity of a carefully curated dataset for training and evaluating machine learning algorithms effectively. Ultimately, the comparative analysis of algorithms and datasets provides valuable understanding into their respective merits and limitations, thereby informing the selection of optimal algorithms for addressing specific cybersecurity challenges.

Future research directions could explore ensemble approaches that combine the strengths of diverse algorithms to enhance the performance of intrusion detection systems. Moreover, incorporating explainability techniques can enhance comprehension of the decision-making mechanisms of machine learning models, thus tackling issues regarding model interpretability in practical scenarios. Additionally, adapting the suggested methodology to changing cybersecurity environments, considering new threats and various network structures, is an area that warrants further investigation. Furthermore, this study establishes a basis for enhancing behavioral analysis in intrusion detection, providing a groundwork for future research focused on strengthening network security against evolving cyber threats.

### ACKNOWLEDGEMENTS

We would like to express our deepest gratitude to our institutions and loved ones for their support throughout this research. We are grateful to Smt. Kashibai Navale College of Engineering and its research center for providing the necessary resources and infrastructure to facilitate this research. I would like to thank my co-authors, Dr. Sandhya Arora and Dr. Parikshit N. Mahalle, for their valuable contributions, guidance, and support throughout this research.

Received: July 19, 2025

**Volume 38 No. 1s, 2025**

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

## REFERENCES

[1] P. Parkar and A. Bilimoria, "A survey on cyber security IDS using ML methods," in *Proc. 5th Int. Conf. I1ntell. Comput. Control Syst. (ICICCS),* pp. 352–360, doi:10.1109/ICICCS51141. 20 21.9 4322 10, 2021.

[2] S. M. Kasongo and Y. Sun, "A Deep Learning Method With Wrapper Based Feature Extraction For Wireless Intrusion Detection System," *Comput. Secur*., vol. 92, doi: 10.1016/j.cose.2020.101752, 2020.

[3] M. Lyu, H. H. Gharakheili, C. Russell, and V. Sivaraman, "Enterprise DNS Asset Mapping and Cyber-Health Tracking via Passive Traffic Analysis," *IEEE Trans. Netw. Serv. Manag.,* vol. 20, no. 3, pp. 3699–3716, doi: 10.1109/TNSM.2022.3221981, 2023.

[4] N. R. Q. Ali, S. Arora and P.N. Mahalle, "Intrusion Detection and Prevention in Wireless Network A Survey of the State-Of-The-Art", in *International Conference on Information and Communication Technology for Intelligent Systems (ICTIS),* vol 1111, pp. 575-585, Springer Nature, Singapore, 2024.

[5] I. Abrar, "A Machine Learning Approach for Intrusion Detection System on NSL-KDD Dataset," in *Proc. Int. Conf. Smart Electron. Commun. (ICOSEC)*, pp. 825–830, doi: 10.1109/ ICOSEC49089.9215358, 2020.

[6] J. Feng, J. Zhang, W. Zhang, and G. Han, "Detecting malicious encrypted traffic with privacy set intersection in cloud-assisted industrial internet," *J. Inf. Secur. Appl.*, vol. 85, p. 103831, 2024, doi: 10.1016/j.jisa.103831, 2024.

[7] K. Atefi, H. Hashim, and T. Khodadadi, "A Hybrid Anomaly Classification with Deep Learning (DL) and Binary Algorithms (BA) as Optimizer in the Intrusion Detection System (IDS)," *Proc. -16th IEEE Int. Colloq. Signal Process. its Appl. CSPA*, no. Cspa, pp. 29–34, doi: 10.1109/CSPA48992.2020.9068725, 2020.

[8] N. Sultana, N. Chilamkurti, W. Peng, and R. Alhadad, "Survey on SDN based network intrusion detection system using machine learning approaches," *Peer-to-Peer Netw. Appl.*, vol. 12, no. 2, pp. 493–501, doi: 10.1007/s12083-017-0630-0, 2019.

[9] H. R. Maier, S. Razavi, Z. Kapelan, L. S. Matott, J. Kasprzyk, and B. A. Tolson, "Introductory overview: Optimization using evolutionary algorithms and other metaheuristics," *Environ. Model. Softw.*, vol. 114, no. pp. 195–213, 2019, doi: 10.1016/j.envsoft.2018.11.018, December 2018.

[10] O. Belej and L. Tamara, "Security of data transfer to the internet of things," *Electron. Prof. Sci. Ed. Cybersecurity Educ. Sci. Tech.*, vol. 2, no. 6 SE-Статті, pp. 6–18, doi: 10.28925/2663-4023.2019.6.618, December 2019.

Received: July 19, 2025

**Volume 38 No. 1s, 2025**

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[11] C. C. Islamy, T. Ahmad, and R. M. Ijtihadie, "Reversible data hiding based on histogram and prediction error for sharing secret data," *Cybersecurity*, vol. 6, no. 1, doi: 10.1186/s42400-023-00147-y, 2023.

[12] J. D. L. C. Ntivuguruzwa and T. Ahmad, "A convolutional neural network to detect possible hidden data in spatial domain images," *Cybersecurity*, vol. 6, no. 1, doi: 10.1186/s42400-023-00156-x, 2023.

[13] L. Lv, W. Wang, Z. Zhang, and X. Liu, "A novel intrusion detection system based on an optimal hybrid kernel extreme learning machine," *Knowledge-Based Syst.*, vol. 195, doi: 10.1016/j.knosys.2020.105648, 2020.

[14] S. Agrawal, "Federated Learning for intrusion detection system: Concepts, challenges and future directions," *Comput. Commun.*, vol. 195, no. September, pp. 346–361, doi: 10.1016/j.comcom.2022.09.012, 2022.

[15] Y. Otoum, D. Liu, and A. Nayak, "DL-IDS: a deep learning–based intrusion detection framework for securing IoT," *Trans. Emerg. Telecommun. Technol.*, vol. 33, no. 3, pp. 1–14, doi: 10.1002/ett.3803, 2022.

[16] I. A. Abdulmajeed and I. M. Husien, "Machine Learning Algorithms and Datasets for Modern IDS Design," *Proc. - 2022 IEEE Int. Conf. Cybern. Comput. Intell. Cybern.*, pp. 335–340, 2022, doi: 10.1109/CyberneticsCom55287.2022.9865255, *2022.*

[17] G. Kocher and G. Kumar, "Machine learning and deep learning methods for intrusion detection systems: recent developments and challenges," *Soft Comput.*, vol. 25, no. 15, pp. 9731–9763, doi: 10.1007/s00500-021-05893-0, 2021.

[18] Z. Hu et al., "Statistical techniques for detecting cyberattacks on computer networks based on an analysis of abnormal traffic behavior," *Int. J. Comput. Netw. Inf. Secur.,* vol. 12, no. 6, pp. 1–13, doi: 10.5815/ijcnis.2020.06.01, 2020.

[19] P. Maniriho, A. N. Mahmood, and M. J. M. Chowdhury, "A study on malicious software behaviour analysis and detection techniques: Taxonomy, current trends and challenges," *Futur. Gener. Comput. Syst.,* vol. 130, pp. 1–18, doi: https://doi.org/10.1016/j. future. 2021.11.030, 2022.

[20] R. Ranjan and S. S. Kumar, "User behaviour analysis using data analytics and machine learning to predict malicious user versus legitimate user," *High-Confidence Comput.,* vol. 2, no. 1, p. 100034, doi: https://doi.org/10.1016/j.hcc.2021.100034, 2022.

[21] M. Shafi, A. H. Lashkari, and H. Mohanty, "Unveiling malicious DNS behavior profiling and generating benchmark dataset through application layer traffic analysis," *Comput. Electr. Eng.,* vol. 118, p. 109436, doi: https://doi.org/10.1016/j. compeleceng. 2024.109436, 2024

Received: July 19, 2025

Volume 38 No. 1s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

[22] J. Chen, H. Xie, S. Cai, L. Song, B. Geng, and W. Guo, "GCN-MHSA: A novel malicious traffic detection method based on graph convolutional neural network and multi-head self-attention mechanism," *Comput. Secur.*, vol. 147, p. 104083, doi: https://doi.org/10.1016/j.cose.2024.104083, 2024.

[23] Shinde, G.R., Majumder, S., Bhapkar, H.R., Mahalle, P.N, "Exploratory Data Analysis. In: Quality of Work-Life During Pandemic Studies in Big Data", vol 100. Springer, Singapore, 2022.

[24] K. Xu, Z.-L. Zhang, and S. Bhattacharyya, "Internet Traffic Behavior Profiling for Network Security Monitoring," *IEEE/ACM Trans. Netw.,* vol. 16, no. 6, pp. 1241–1252, doi: 10.1109/TNET.2007.911438, 2008.

[25] S. Ali, O. Abusabha, F. Ali, M. Imran, and T. Abuhmed, "Effective Multitask Deep Learning for IoT Malware Detection and Identification Using Behavioral Traffic Analysis," *IEEE Trans. Netw. Serv. Manag.,* vol. 20, no. 2, pp. 1199–1209, doi: 10.1109/TNSM.2022.3200741, 2023.

[26] H. N. Nguyen, F. Abri, V. Pham, M. Chatterjee, A. S. Namin, and T. Dang, "MalView: Interactive Visual Analytics for Comprehending Malware Behavior," *IEEE Access,* vol. 10, no. September, pp. 99909–99930, doi: 10.1109/ACCESS.2022.3207782, 2022.

Received: July 19, 2025