JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Cyber-Attack Detection and Classification for Data Security Using Hybrid Cryptography and Machine Learning

Dr. Akhilesh Kumar
Chief Technology Officer (CTO)
Department - Information Technology
Organisation Name - Kalra Hospital SRCNC Pvt. Ltd.

Abstract

In the dynamic digital environment, cyber-attacks are a major threat to data security and system integrity. This study introduces a strong framework for the detection and classification of cyber-attacks using hybrid cryptography fused with machine learning (ML). The system integrates symmetric and asymmetric cryptography for securing data transmission while utilizing XG-Boost and Random Forest classifiers for real-time threat identification. Experiments were performed using the ToN-IoT and BoT-IoT datasets with data preprocessing, feature extraction through Extra-Tree Classifier, and performance measurement using $A_{accuracy}$, $P_{precision}$, R_{recall}, and $F1_{score}$ as metrics. Outcomes show that the Random Forest approach performs better than XG-Boost with 99.89% accuracy on the BoT-IoT dataset, with better precision, recall, and F1-scores than current methodologies. The integration of ML and cryptography reinforces system robustness, prevents advanced attacks, and guarantees secure data sharing, providing an encouraging solution to present-day cyber-security issues.

Keywords: Cyber-attack, ML, Cryptography, XG-Boost, Random Forest

1. Introduction

In the modern digital connected world, protection of sensitive information and secure data delivery has acquired the sleek priority in all aspects of the industry, government agencies and individual users [1]. Due to the high speed development of the information technology, and the spread of internet enabled devices, it has led to a high rise in the amount of data transacted in the networks [2]. Nevertheless, the growth has brought vulnerability of cyber systems to advanced attacks, breach of information, unauthorized access, and malicious attacks as well. Phishing, malware intrusions, ransomware, and Distributed Denial of Service (DDoS) (Figure 1) are the most

well-known cyber-attacks; besides disrupting the confidentiality of data, they also risk the integrity and availability of data systems and information infrastructures [3,4]. To overcome such rising issues of security, refined techniques involving the combination of cryptography and smart technologies have become most popular [5]. This study is aimed at improving the detection and the classification of cyber-attacks by integrating hybrid cryptographic-based technologies and ML models in developing robust defense system in securing data.

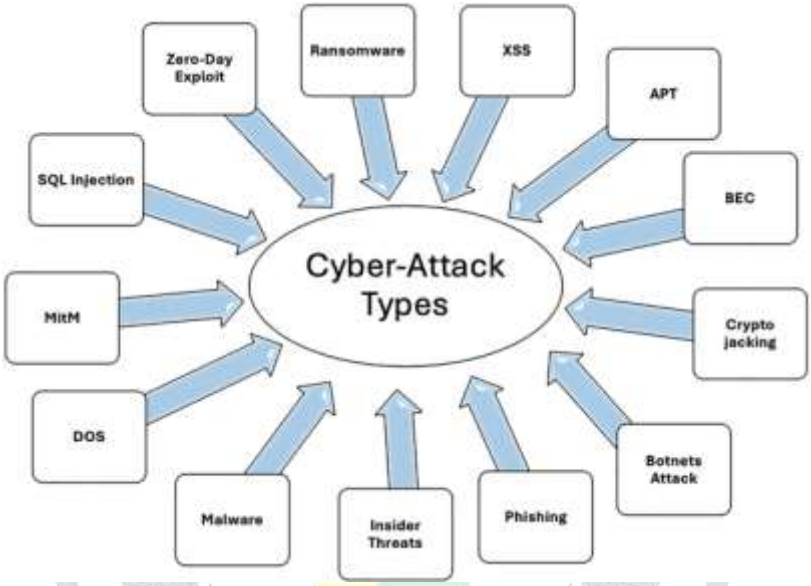

Figure 1: Types of Cyber-attack [6].

The conventional security systems that are most commonly based on static encryption techniques and signature-procedures of detection, tend to fail to detect and limit the dynamic and changing threats in cyberspace [7]. Although encryption procedures, like symmetric and asymmetric cryptography, secure the data both at transit and at rest, the methods can become limited in cases where the hacker finds a loophole in the network topology or creates a new attack algorithm [8]. By the same token, traditional intrusion detection systems (IDS) are never able to successfully identify zero-day attacks or anomalies that are not well-known, causing major threats to organizations and users [9]. In order to overcome these shortcomings, this study suggests a hybrid security system that adapts both cryptographic algorithms and ML-based detection systems to achieve the layers of security.

The intended model takes advantage of the hybrid cryptography method which has the merger benefit of both faster data processing and high security features of the asymmetric encryption method of data processing [10]. The use of highly-efficient cryptographic algorithms guarantees the system data confidentiality, integrity, and authentication in the course of communication. Nevertheless, all this cryptography is not able to ensure total security against more and more sophisticated cyber-attacks [11]. As such, this study uses the ML approach in order to improve their real-time prediction and classification of threats. Models of ML, in turn, after learning data volumes (attack traffic and normal traffic scenarios) can find even sophisticated attack vectors, diagnose anomalies, and effectively separate various types of cyber-attacks [12].

In conclusion, the fusion of hybrid cryptography and ML presents a promising strategy to mitigate the risks of cyber-attacks and safeguard sensitive information. By providing real-time detection, classification, and secure data exchange, this approach offers an effective solution to the growing challenges in the domain of cyber security. Here are the researches objectives of the study are follows as:

- To develop an efficient ML-based framework for the detection of cyber-attacks targeting data security.
- To evaluate and compare the performance of various ML models in terms of detection rate, false positives, precision, and recall.
- To create a robust dataset or utilize existing cyber-security datasets for training and validating the proposed models.
- To implement feature selection and data preprocessing techniques to improve model efficiency and reduce computational complexity.

2. Literature Review

In this section, the authors proposed a previous study based on cyber-attack detection and classification for data security using DL and ML.

Kumar et al., (2025) [13] used a new DL framework for attack detection called CCPGANN-TOA, which stands for "capsule convolutional polymorphic graph attention neural network with tyrannosaurus optimization algorithm". Afterwards, DSA-ECC, a technique based on elliptic curve cryptography, is used to encrypt normal data. This approach offers good security with reduced key sizes, which means that calculations are faster and resources are saved. In comparison to conventional methods, the suggested one achieves accuracy rates of 99.98% on data set I, 99.9% on data set II, and 900 kbps more throughput with minimum latency.

Behiry et al., (2024) [14] proposed an intelligent mixed approach consisting of ML and AI to enhance the security of Wireless Sensor Networks (WSNs) and prevent attacks. The feature extraction is achieved through an improved K-means clustering strategy (KMC-IG) and feature reduction algorithms like, SVD and PCA. Once the data has undergone balance after being conducted by SMOTE, the next step is intrusion detection and traffic categorization. Feed-forward neural network DL-trained showed good evaluation metrices on the NSL-KDD, UNSW-NB 15, and CICIDS 2017 data sets. In enhancing WSN security, the proposed model is better than benchmark technique.

Alomiri et al., (2024) [15] used the strength of a Ridge Classifier to detect anomalies in IoT systems. With this approach, the proposed security system could help identify and preempt any type of cyberattack when it happens by utilizing up-to-date and safe network information. Applying the ML techniques enhances threat detection and

mitigation by the system. Experimental findings indicate that the proposed remedy is fairly good at detecting and minimizing IoT network risks with a spectacular accuracy figure of 97%.

Isaac et al., (2024) [16] proposed cyber security attack detection model using semi-supervised learning. It compares the performance of "Multi Connect Variational Auto-Encoder (MC-VAE)", "Probabilistic Bayesian Networks (PBN)", and a combined model of MC-VAE and PBN. The study employs the NUSWNB15_GT dataset for training and evaluation purposes. Notably, the "Semi-Supervised Learning with Probabilistic Bayesian Networks (SSL-PBN)" model demonstrates exceptional results, achieving a precision rate of 94% and a recall rate of 90%.

Saini et al., (2023) [17] used publically available datasets to classify and successfully identify APT assaults using DL and ML models such random forest, decision tree, convolutional neural network, multilayer perceptron, etc. Information from the following datasets: "CSE-CIC-IDS2018, CIC-IDS2017, NSL-KDD, and UNSW-NB15" made it into this analysis. The hybrid ensemble ML model, which combines XGBoost and random forest classifiers, is suggested in this work. On the "CSE-CIC-IDS2018, CIC-IDS2017, NSL-KDD, and UNSW-NB15" datasets, it achieved a maximum prediction accuracy of 98.92%, 99.91%, 99.24%, and 97.11%, respectively.

Mahmood et al., (2022) [18] proposed a method of identifying cyberattacks through intelligent hybrid model that combines DL and ML. In addition to that, researchers present a feature reduction model, where the PCA and SVD as ML methods are applied to select the most sensible characteristics among the accepted attack classes. Findings indicate that the proposed hybrid model based cyber detection system is better than the conventional ones with a precision of 99.98%, recall of 100% and F1-measure exceptionally reached 100%, respectively.

Naser et al., (2022) [19] created a DL method to detect cyberattacks in WSN. This is a similar process to how the nodes in a WSN behave and the data is delivered using MQTT protocol. Using the DL model, this method increases the level of detection accuracy as compared to the conventional ML methods. Depending on the dataset, the results indicate that the combination of DL (CNN-LSTM) methods turned out to be efficient as the training accuracy was 96.02% and the validation accuracy was 95.08%.

Akhtar et al., (2022) [20] presented a system that uses ML models to identify cyberattacks. This study also looked at several classification models based on linear ML algorithms, with the goal of detecting attacks in their early stages. The presentation itself is compared to the accuracy of classifiers. Observance of balance procedures was ensured. Topping the accuracy charts at 87.93% are Radio Frequency and GBC, followed by 86.11% for ABC, 81.03% for BT, 70.31% for ET, and 70.31% for DT (a total of 84.48%).

3. Problem statement

The prevalence of cyber-attacks in the contemporary world of digitalization is highly threatening to the safety of data and system integrity of the interconnected systems due to the ever-growing rate of digitalization and connected systems that has peaked quite sharply and suddenly within the past few years. The modern security response systems fail to track real-time changes in attack pattern because of how they are set. In this study, the

importance of developing an intelligent and efficient network of detecting and classifying cyber-attacks will be answered with an interconnection of the hybrid cryptography with ML algorithms. The proposed system is the secure way of data transmission, and it uses the ML models to identify and classify cyber-attacks with accuracy. It assists in making systems more resilient and detection more accurate and is a proactive solution towards protecting sensitive information.

4. Research Methodology

In this section, the authors provide research methodology of the proposed work (Figure 2) based on cyber-attack detection and classification for data security using hybrid cryptography and ML.

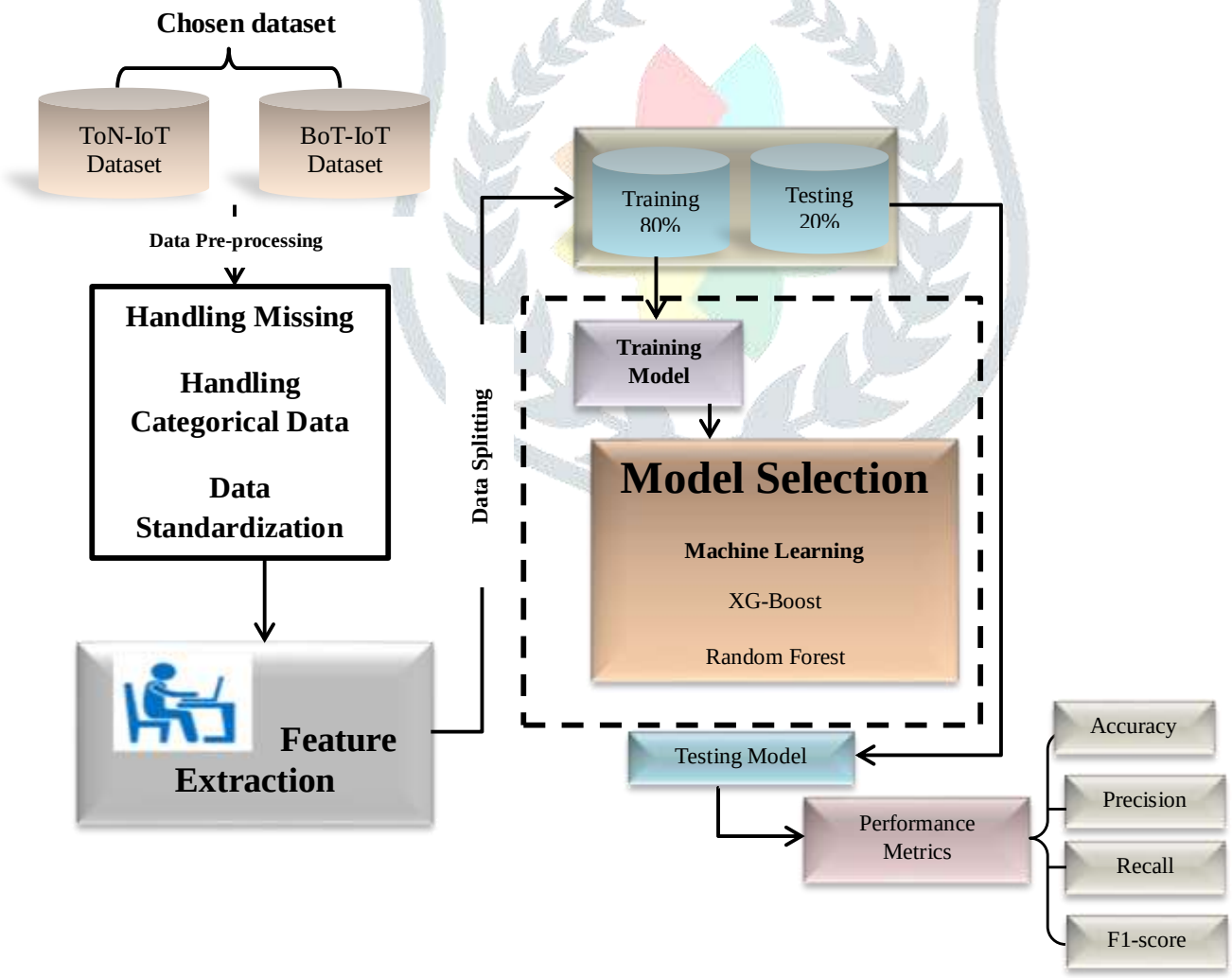

Figure 2: Flowchart of proposed work

4.1 Dataset Used

a) ToN-IoT Dataset

The ToN-IoT dataset compiles data of both the Internet of Things (IoT) and the Industrial Internet of Things (IIoT) to evaluate the practice of cybersecurity alleviation to implement artificial intelligence (AI) [21]. As part of

its ability to simulate actual network set-ups using VMs, cloud layers, and physical systems, it uses information on connected devices, system logs (Windows/Linux) and network traffic. An attribute count of 45, covering timestamps, IPs, ports, protocols, service data, connection statuses, anomalies, labelled attack type covers 461,043 entries in CSV format. Various, real-world data on network systems and IoT services could be used by researchers to examine threats and create proper cybersecurity protection measures.

b) BoT-IoT Dataset

The BoT-IoT dataset, developed by the Cyber Range Lab at UNSW Canberra, represents a realistic network setting with both legitimate and malicious activity [22]. Data exfiltration, keylogging, DDoS, and OS/service scanning are just a few of the attack methods covered. It's available in pcap, argus, and CSV formats. There are around 3 million records spread across four files (~1.07 GB) with 46 characteristics in a 5% sampled subset that was utilized for this investigation. Among these are labelled attack categories, timestamps, IP addresses, protocols, packet counts, connection statuses, and statistical metrics. Cybersecurity models based on artificial intelligence (AI) could be built and tested using this dataset to identify various network assaults.

4.2 Data Preprocessing

a) Handling Missing

Handling missing values is the initial stage in data pre-processing. Statistical methods, such as the mean or standard deviation, are typically used to replace them. Since there were few missing values in the ToN-IoT and BoT-IoT datasets, they were considered clean for this investigation. To prevent overfitting—when models discover patterns associated with certain IP addresses or times that don't transfer to fresh data—attributes like timestamps and IP addresses were deleted. The predictive power of the model can be diminished by the introduction of noise caused by irrelevant information, such as IPs.

b) Handling Categorical Data

ML models often struggle with datasets containing string attributes. Python's scikit-learn library, specifically "sklearn.preprocessing," provides tools to handle this using label encoding. In order to prepare data for model training, this method transforms categorical text input into numerical values according to class labels. To guarantee efficient transformation and boost speed, the suggested approach uses label encoding.

c) Data Standardization

Features that are not model-friendly such as the IP source and destination addresses are found in ToN-IoT and BoT-IoT data sets. Labels are encoded after which values need to be normalized to Standard Scaler. A correlation matrix is used to establish suitable attributes which are to be standardized. The most appropriate traits to normalize IPs with ToN-IoT data are to be conn_state and the most appropriate traits to normalize IPs with the data of BoT-IoT are to be proto_number because they will aid in the boosting of the model performance.

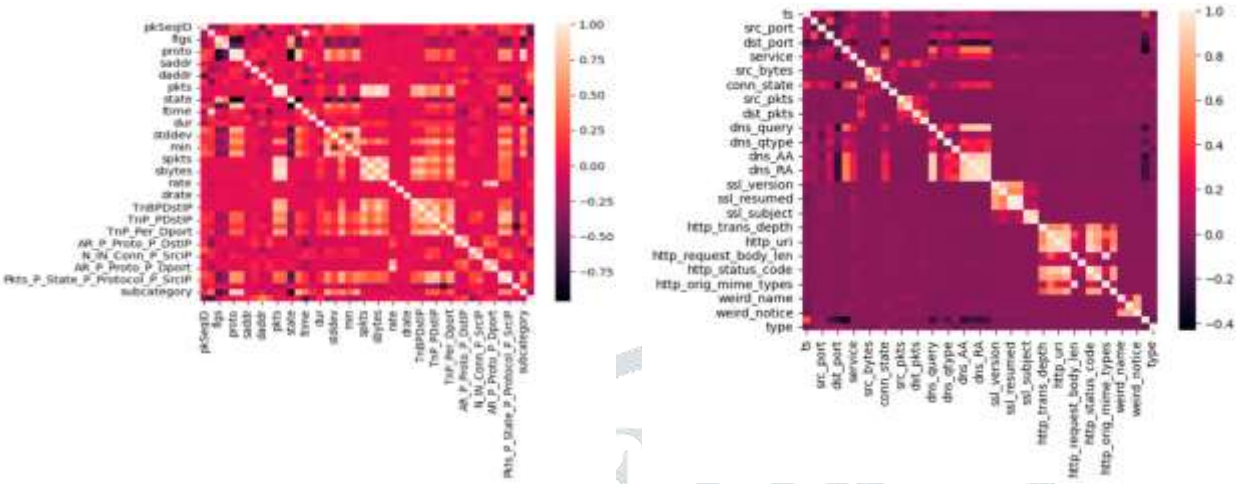

Figure 3: Correlation matrix of proposed dataset

4.3 Feature Extraction

By applying the ExtraTreeClassifier method, the authors selected a set of critical properties of the ToN-IoT and BoT-IoT databases to ensure that overfitting did not lead to the incorrect evaluation of the proposed model. They considered the subset of 20 significant features by utilizing the ExtraTreeClassifier and discarding all the others since the traditional classifiers using a large number of features have a tendency to overfit. Due to the ability of this algorithm to handle and process high-dimensional, cross-type information without the need to feature-scale data, it is ideal in solving DDoS identification. It is computationally economical in real-time peers and it is particularly effective with skewed datasets which are common in case of DDoS scenarios. ExtraTreeClassifier captures non-linear and wicked feature associations to help identify the pattern of attacks within normal flow of traffic traffic. It is also noise and overfitting resistant, due to its random choice of features and ensemble structure and this adds accuracy and reduces false positives or negatives. The introduction of security systems is possible without any hassle. To increase the performance of the model, ExtraTreeClassifier using sklearn.ensemble was selected to identify the top 15 features of both data as Gini significance was used.

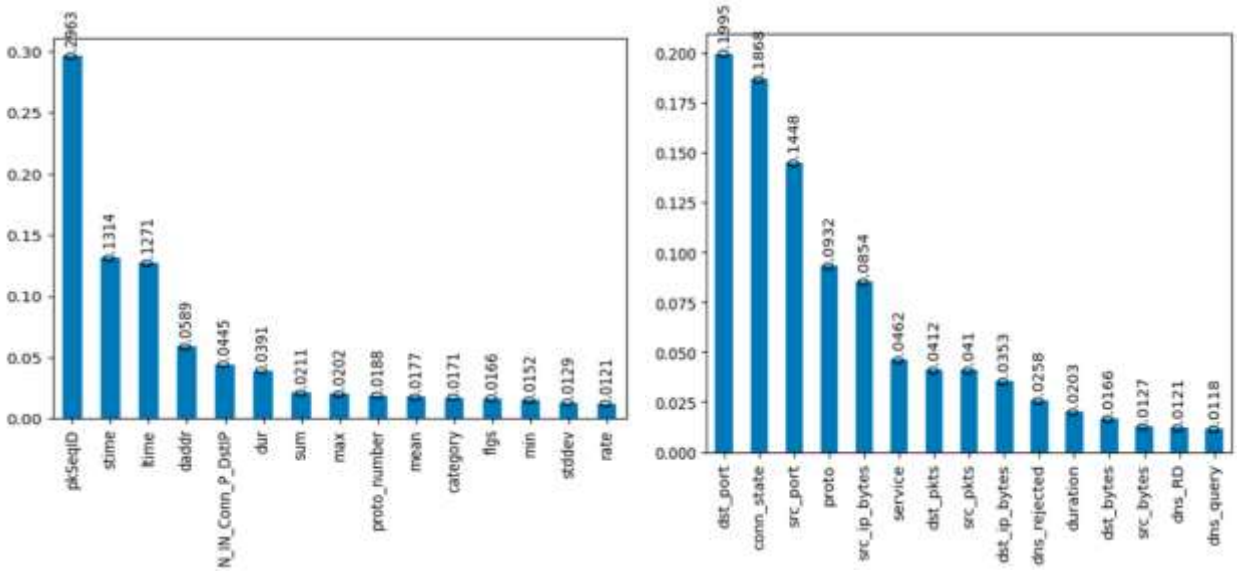

Figure 4: Fifteen important features of proposed dataset

4.4 ML Classifications

In this section, the authors proposed ML classification models for cyber-attack detection and classification.

a) XG-Boost Model

XG-Boost is a high-performance and computationally efficient ML algorithm following the gradient boosting paradigm [23]. The fluctuation of a gradient boost is considerably more regulated. While it employs more complex smoothing techniques (L1 and L2), its performance is noticeably superior to gradient boost methods. It is one of the most famous algorithms in the machine-learning field because it is consistently better than other strategies [24]. The rate at which it runs is rapid. The XG-Boost algorithm is different from the traditional gradient boosting in that it uses a novel method to build trees, where the gain and similarity score decide the best node splits [25].

$$The\ similarity\ Score = \frac{(\sum_{i=1}^{n} Residual_i)^2}{\sum_{i=0}^{n}[Previous\ Probability_i^*(1-Previous\ Probability_i)]+\lambda} \quad (3)$$

b) Random Forest (RF) Model

RF model are utilized in both classification and regression analyses. Predictions are generated with a tree-structured representation of the input [26]. RF approach could produce consistent results on big datasets, even in the absence of a substantial number of record entries. The decision tree can archive its findings and apply them to other datasets [27]. There are two parts to a RF: first, making the random forest itself; and second, making a prediction based on the classifications generated in the first part.

$$Gini = 1 - \sum_{i=1}^{n}(p_i)^2 \quad (4)$$

The value of p_i in equation (4) represents the object's classification probability according to a given characteristic or class.

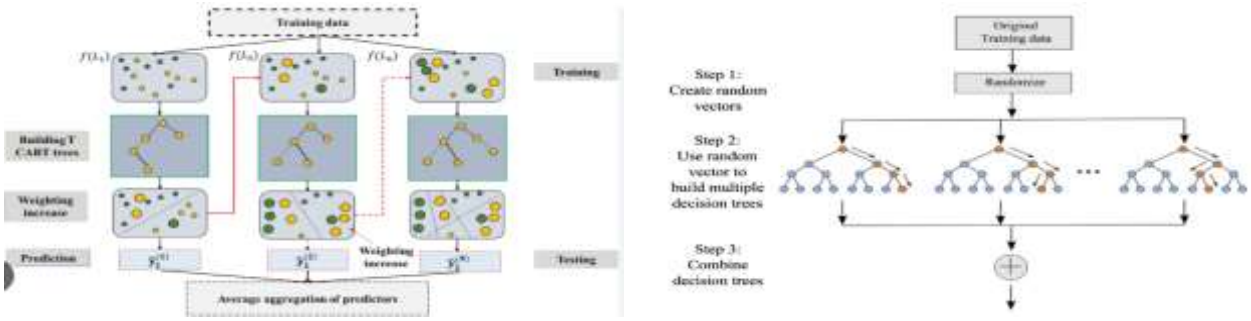

Figure 5: Architecture of ML model [28].

4.5 Anomaly Detection by a Classifier

Anomaly detection by classifier is used to detect attacks in IoT devices by categorizing network traffic as normal (Class 0) or attack (Class 1). The classifier is trained on labeled data containing both normal and attack patterns, using features like packet size, flow rate, and communication behavior. Its performance is measured using

$A_{accuracy}$, $P_{precision}$, R_{recall}, and $F1_{score}$. Unlike signature-based methods, it can detect unknown or zero-day attacks, enhancing IoT network security by providing early detection and improving the reliability and availability of IoT services. The models' performance is evaluated using key metrics: $A_{accuracy}$, $P_{precision}$, R_{recall}, and $F1_{score}$.

$$Accuracy = \frac{TN+TP}{FP+FN+TP+TN} \quad (1)$$

$$Recall = Sensitivity = \frac{TP}{FN+TP} \quad (2)$$

$$Precision = \frac{TP}{FP+TP} \quad (3)$$

$$F1 - score = \frac{2*P*R}{P+R} \quad (4)$$

5. Result and Discussion

This experiment used Google Colab for the ToN-IoT dataset and Jupyter Notebook for the BoT-IoT dataset due to Colab's limited data capacity. The setup included an Intel Core i7-9750H CPU (6 cores, 12 threads) at 2.60 GHz, 16 GB RAM, and an NVIDIA GeForce GTX 1660 Ti Max-Q with 4 GB graphics.

5.1 XG-Boost Model

Figure 7 shows the curves of accuracy and corresponding loss of the XGBoost model on the proposed dataset during 50 epochs. XG-Boost model on the ToN-IoT dataset indicates a gradual rise of the training and validation accuracies, about 95% respectively, and a corresponding loss reduction until a steady state of almost 0.2 which is an indication that the model is learning well and has little overfitting. XG-Boost model on BoT-IoT data has shown improvement in model accuracy by refinement, giving accuracy of more than 99% in training and validation sets. Alongside, the loss figures in the related loss curves drop greatly as well reaching almost zero values with a slight fluctuation. In general, the findings prove the good accuracy of the model, its high stability, and effective convergence on advancing dataset. The Figure 6 represents the confusion Matrix of ToN-IoT and BoT-IoT.

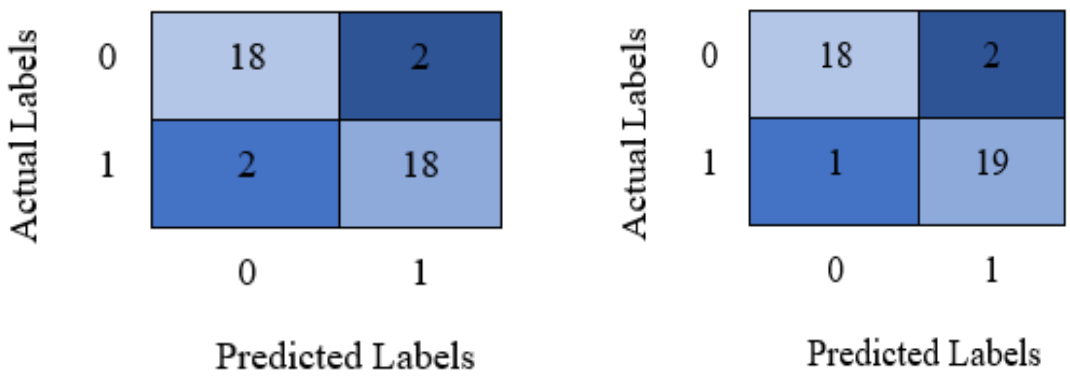

Figure 6: Confusion Matrix for ToN-IoT and BoT-IoT

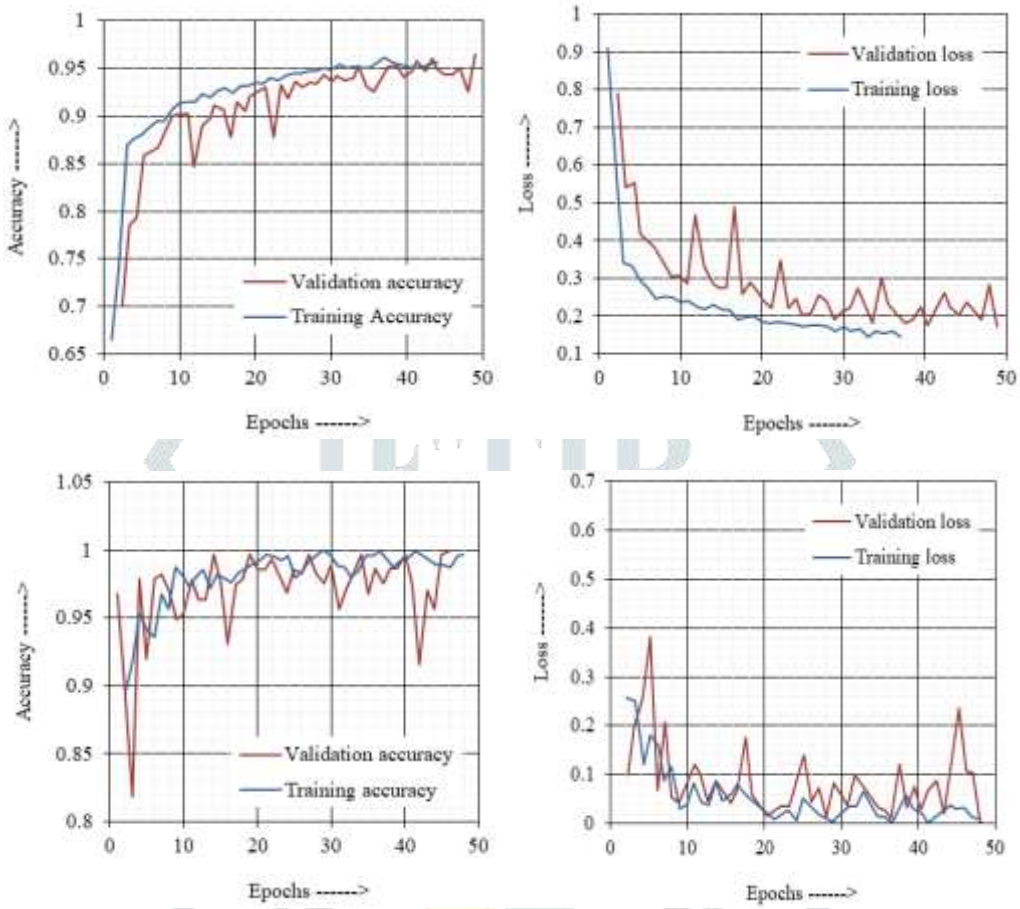

Figure 7: Accuracy and loss curve of XG-Boost on proposed dataset

5.2 Random Forest

Figure 9 illustrates the validation and loss curves of the RF model on the suggested dataset after 50 epochs. RF model, training accuracy increases progressively up to more than 95%, whereas validation accuracy varies between 85%, indicating overfitting in ToN-IoT dataset. In the same manner, training loss declines progressively, whereas validation loss exhibits erratic behavior with high levels. RF model on BoT-IoT dataset gives improved model outcomes, wherein training accuracy reaches 99%, but validation accuracy remains in the range of 90% to 95%, with high instability. The respective loss curves indicate diminishing training loss but steady oscillations in validation loss, with indications of overfitting issues. Figure 8 illustrates the confusion Matrix for ToN-IoT and BoT-IoT.

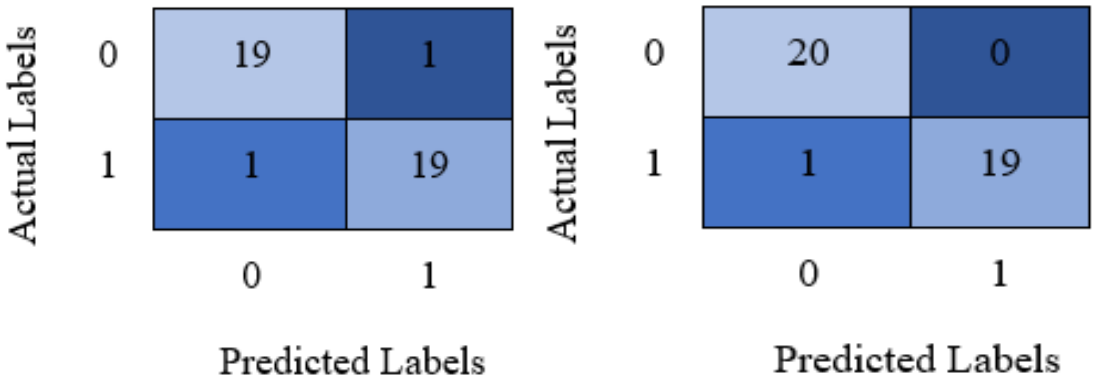

Figure 8: Confusion Matrix for TON-IOT and BOT-IOT

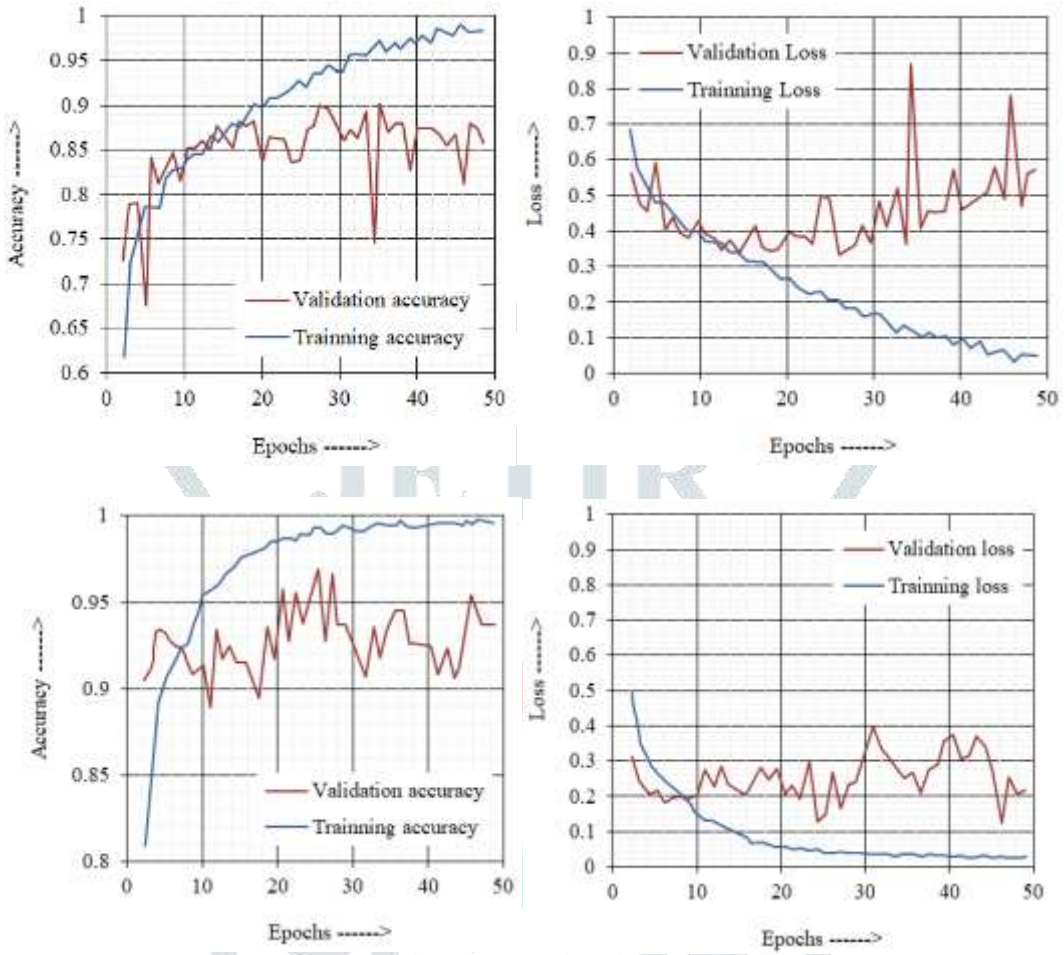

Figure 9: Accuracy and loss curve of RF on proposed dataset

5.3 Comparison Analysis

The table 1 compares the performance of XG-Boost and Random Forest (RF) models on the TON-IOT and BOT-IOT datasets using key evaluation metrics. For the TON-IOT dataset, RF outperforms XG-Boost, achieving higher accuracy (97.50% vs. 96.79%), precision (95.25% vs. 94.27%), recall (100% vs. 95.74%), and F1-score (97.56% vs. 95%). Similarly, for the BOT-IOT dataset, RF achieves superior results with 99.89% accuracy, 97.70% precision, 98.64% recall, and 98.17% F1-score, compared to XG-Boost's 97.29% accuracy and slightly lower precision, recall, and F1-score. Overall, RF demonstrates better performance and robustness across both datasets.

Table 1: Comparison of proposed models

Parameter	XG-Boost		Random Forest	
	TON-IOT	BOT-IOT	TON-IOT	BOT-IOT
Accuracy	96.79%	97.29%	**97.50%**	**99.89%**
Precision	94.27%	94.36%	**95.25%**	**97.70%**
Recall	95.74%	97.29%	**100%**	**98.64%**
F1-score	95%	95.80%	**97.56%**	**98.17%**

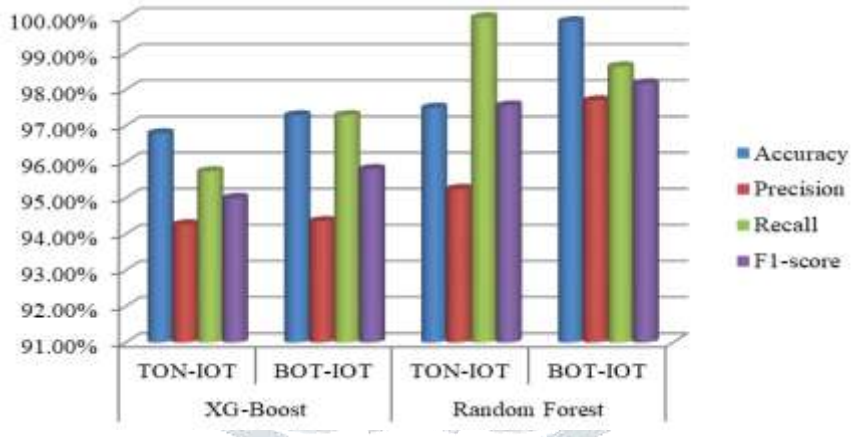

Figure 10: Comparison graph of proposed models

The table 2 provides a comparative study of the various approaches of detecting cyber-attacks on the basis of accuracy, precision, recall, and F1- score. Gaur et al. (2022) applied Decision Tree, which was 90.41% accurate, with high precision, and relatively low recall and F1-score. Kumar et al. (2021) suggested the use of the TP2SF approach with an accuracy of 98.% with both precision and recall rates being even. Mahmood et al. (2022) utilized a ML approach, with a corresponding accuracy of 96.61%. Conversely the model proposed by this work is better than all previous models as it can attain 99.89% accuracy, 97.70% precision, 98.64% recall and 98.17% F1 score which is a display of its strength and effectiveness.

Table 2: Comparison of proposed work with previous work

Authors [Reference]	Methodology Used	Accuracy	Precision	Recall	F-score
Gaur et al., (2022) [29]	**Decision Tree**	90.41	99	91	94
Kumar et al., (2021) [30]	**TP2SF**	98.84	97.23	94.03	95.28
Mahmood et al., (2022) [31]	**ML**	96.61	97.02	96.61	96.15
Our Work	**Proposed model**	**99.89**	**97.70**	**98.64**	**98.17**

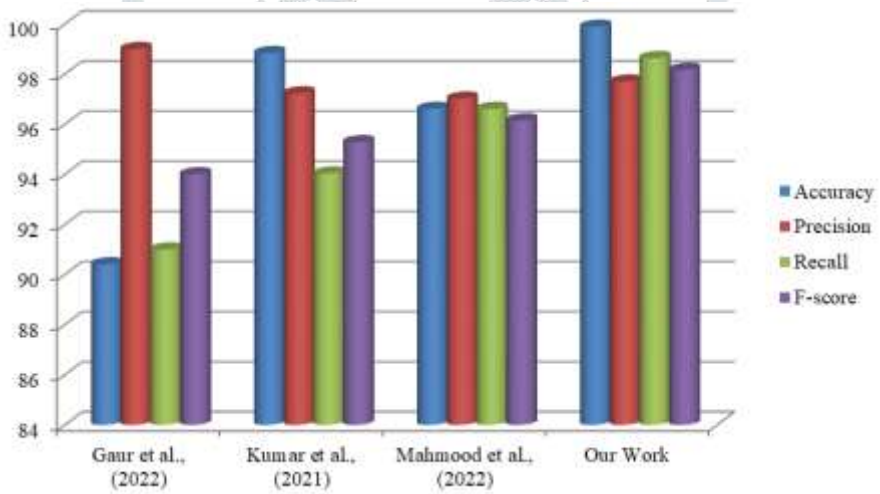

Figure 11: Comparison graph of proposed model with previous model

6. Conclusion

This paper suggests a successful scheme of cyber-attack detection and classification that combines hybrid cryptography and more sophisticated machine learning models to maximize data security. Its implementation integrates the advantages of both symmetric and asymmetric cryptographic algorithms to provide data security during transportation, in addition to using the machine learning algorithms, namely XG-Boost and Random Forest, to detect the attacks correctly. By conducting stringent experiments on both the ToN-IoT and the BoT-IoT based datasets the Random Forest model gave better results with an accuracy of 99.89%, with high precision, recall, and F1-score, which outperformed the result of other models available in literature. The methods of feature extraction have also been applied, including Extra-Tree Classifier, which enhances the efficiency of models due to the reduced overfitting. This outcome confirms the suitability of the proposed framework in identifying advanced cyber-attacks, such as zero-day attacks, and at the same time maintaining secure communication. In general, the combination of hybrid cryptography and intelligent detecting mechanisms ensures a scalable, reliable, and resistant approach to the rising issue of cybersecurity in the interconnected environment.

Reference

1. Rayhan, Abu. "Cybersecurity in the Digital Age: Assessing Threats and Strengthening Defenses." In *Conference: Cybersecurity Awareness*, pp. 1-26. 2024.
2. Ahmed, Shams Forruque, Md Sakib Bin Alam, Mahfara Hoque, Aiman Lameesa, Shaila Afrin, Tasfia Farah, Maliha Kabir, G. M. Shafiullah, and S. M. Muyeen. "Industrial Internet of Things enabled technologies, challenges, and future directions." *Computers and Electrical Engineering* 110 (2023): 108847.
3. Aslan, Ömer, Semih Serkant Aktuğ, Merve Ozkan-Okay, Abdullah Asim Yilmaz, and Erdal Akin. "A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions." *Electronics* 12, no. 6 (2023): 1333.
4. Cremer, Frank, Barry Sheehan, Michael Fortmann, Arash N. Kia, Martin Mullins, Finbarr Murphy, and Stefan Materne. "Cyber risk and cybersecurity: a systematic review of data availability." *The Geneva papers on risk and insurance. Issues and practice* 47, no. 3 (2022): 698.
5. Admass, Wasyihun Sema, Yirga Yayeh Munaye, and Abebe Abeshu Diro. "Cyber security: State of the art, challenges and future directions." *Cyber Security and Applications* 2 (2024): 100031.
6. Salem, Aya H., Safaa M. Azzam, O. E. Emam, and Amr A. Abohany. "Advancing cybersecurity: a comprehensive review of AI-driven detection techniques." *Journal of Big Data* 11, no. 1 (2024): 105.
7. Goswami, MaloyJyoti. "AI-based anomaly detection for real-time cybersecurity." *International Journal of Research and Review Techniques* 3, no. 1 (2024): 45-53.
8. Wadho, Shuaib Ahmed, Areej Fatemah Meghji, Aun Yichiet, Roshan Kumar, and Farhan Bashir Shaikh. "Encryption Techniques and Algorithms to Combat Cybersecurity Attacks: A Review." *VAWKUM Transactions on Computer Sciences* 11, no. 1 (2023): 295-305.
9. Parhizkari, Siamak. "Anomaly detection in intrusion detection systems." In *Anomaly Detection-Recent Advances, AI and ML Perspectives and Applications*. IntechOpen, 2023.
10. Zhang, Qixin. "An overview and analysis of hybrid encryption: the combination of symmetric encryption and asymmetric encryption." In *2021 2nd international conference on computing and data science (CDS)*, pp. 616-622. IEEE, 2021.
11. Koupaei, Alireza Nik Aein. "Hybrid Encryption Scheme Combining AES and ECC for Enhanced Data Security."
12. Talukder, Md Alamin, Md Manowarul Islam, Md Ashraf Uddin, Khondokar Fida Hasan, Selina Sharmin, Salem A. Alyami, and Mohammad Ali Moni. "Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction." *Journal of big data* 11, no. 1 (2024): 33.

13. Kumar, PJ Sathish, BR Tapas Bapu, S. Sridhar, and V. Nagaraju. "An Efficient Cyber Security Attack Detection With Encryption Using Capsule Convolutional Polymorphic Graph Attention." *Transactions on Emerging Telecommunications Technologies* 36, no. 3 (2025): e70069.

14. Behiry, Mohamed H., and Mohammed Aly. "Cyberattack detection in wireless sensor networks using a hybrid feature reduction technique with AI and machine learning methods." *Journal of Big Data* 11, no. 1 (2024): 16.

15. Alomiri, Abdullah, Shailendra Mishra, and Mohammed AlShehri. "Machine learning-based security mechanism to detect and prevent cyber-attack in IoT networks." *International Journal of Computing and Digital Systems* 16, no. 1 (2024): 645-659.

16. Isaac, Samson, Damilola Kolawole Ayodeji, Yusuf Luqman, Solomon Mathew Karma, and Jibril Aminu. "CYBER SECURITY ATTACK DETECTION MODEL USING SEMI-SUPERVISED LEARNING." *FUDMA JOURNAL OF SCIENCES* 8, no. 2 (2024): 92-100.

17. Saini, Neeraj, Vivekananda Bhat Kasaragod, Krishna Prakasha, and Ashok Kumar Das. "A hybrid ensemble machine learning model for detecting APT attacks based on network behavior anomaly detection." *Concurrency and Computation: Practice and Experience* 35, no. 28 (2023): e7865.

18. Mahmood Naser, S., Y. Hussain Ali, and D. Al-Jumeily OBE. "Hybrid cyber-security model for attacks detection based on deep and machine learning." *International Journal of Online and Biomedical Engineering (iJOE)* 18, no. 11 (2022): 17-30.

19. Naser, Shaymaa Mahmood, Yossra Hussain Ali, and Dhiya Al-Jumeily OBE. "Deep learning model for cyber-attacks detection method in wireless sensor networks." *Periodicals of Engineering and Natural Sciences (PEN)* 10, no. 2 (2022): 251-259.

20. Akhtar, Muhammad Shoaib, and Tao Feng. "Comparison of Classification Model for the Detection of Cyber-attack using Ensemble Learning Models." *EAI Endorsed Transactions on Scalable Information Systems* 9, no. 5 (2022).

21. https://www.kaggle.com/datasets/amaniabourida/ton-iot

22. https://www.kaggle.com/datasets/vigneshvenkateswaran/bot-iot

23. Chen, Zhuo, Fu Jiang, Yijun Cheng, Xin Gu, Weirong Liu, and Jun Peng. "XGBoost classifier for DDoS attack detection and analysis in SDN-based cloud." In *2018 IEEE international conference on big data and smart computing (bigcomp)*, pp. 251-256. IEEE, 2018

24. Ismanto, Edi, Januar Al Amien, and Vitriani Vitriani. "A Comparison of Enhanced Ensemble Learning Techniques for Internet of Things Network Attack Detection." *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer* 23, no. 3 (2024): 543-556.

25. Dhaliwal, Sukhpreet Singh, Abdullah-Al Nahid, and Robert Abbas. "Effective intrusion detection system using XGBoost." *Information* 9, no. 7 (2018): 149.

26. Ma, Ruikui, Qiuqian Wang, Xiangxi Bu, and Xuebin Chen. "Real-time detection of DDoS attacks based on random forest in SDN." *Applied Sciences* 13, no. 13 (2023): 7872.

27. Alduailij, Mona, Qazi Waqas Khan, Muhammad Tahir, Muhammad Sardaraz, Mai Alduailij, and Fazila Malik. "Machine-learning-based DDoS attack detection using mutual information and random forest feature importance method." *Symmetry* 14, no. 6 (2022): 1095.

28. Çavuşoğlu, Ünal. "A new hybrid approach for intrusion detection using machine learning methods." *Applied Intelligence* 49 (2019): 2735-2761.

29. Gaur, Vimal, and Rajneesh Kumar. "Analysis of machine learning classifiers for early detection of DDoS attacks on IoT devices." *Arabian Journal for Science and Engineering* 47, no. 2 (2022): 1353-1374.

30. Kumar, Prabhat, Govind P. Gupta, and Rakesh Tripathi. "TP2SF: A Trustworthy Privacy-Preserving Secured Framework for sustainable smart cities by leveraging blockchain and machine learning." *Journal of Systems Architecture* 115 (2021): 101954.

31. Mahmood Naser, S., Y. Hussain Ali, and D. Al-Jumeily OBE. "Hybrid cyber-security model for attacks detection based on deep and machine learning." *International Journal of Online and Biomedical Engineering (iJOE)* 18, no. 11 (2022): 17-30.