International Journal of Multidisciplinary Research and Growth Evaluation.

**Ferdowsi University Mashhad/ The second International Conference
Artificial Intelligence: Between Scientific Innovation and Human Responsibility**

Intelligent Detection and Prediction of DDoS Attacks Using Machine Learning and Network Traffic Analysis

Ghaith Mousa Hamzah Amlak
General Directorate of Education Babylon, Ministry of Education, Terbia Street, Hila, Iraq

* Corresponding Author: **Ghaith Mousa Hamzah Amlak**

Article Info

ISSN (Online): 2582-7138
Impact Factor (RSIF): 7.98
Volume: 06
Issue: 06
November - December 2025
Received: 21-09-2025
Accepted: 23-10-2025
Published: 20-11-2025
Page No: 144-150

Abstract
One of the most dangerous attacks that is increasing day by day is the distributed denial of service (DDoS) attack. Therefore, it is necessary to develop a model to detect and predict these attacks. To this end, a model was created to detect and prevent DDoS. This model begins by using the distributed DDoS database available on the internet. The next step is data preprocessing to obtain high-quality data. Then, a neural network algorithm is applied to detect DDoS. In the final step, infected data is predicted by taking the infected data that was classified in the previous step and then applying a K-means clustering algorithm. The results obtained with this model yielded an accuracy of 0.99% with a neural network algorithm. The prediction values obtained with the hierarchical clustering algorithm were as follows: {C0: 11748, C1: 9349, C3:4571}.

Keywords: DDoS Attack, Neural Network, Clustering, Feature Engineering, Network Traffic

1. Introduction

In the era of rapid digital development, cyber-attacks have become one of the most prominent challenges facing organizations and individuals alike. Among these threats, the Distributed Denial of Service (DDoS) attack stands out as one of the most common and dangerous forms of cyberattacks [1]. This type of attack floods servers, networks, or websites with massive connection requests from multiple sources, with the goal of disrupting service and preventing legitimate users from accessing it [2]. The danger of DDoS attacks stems from their ability to completely paralyze targeted systems, causing significant financial losses and negatively impacting organizations' digital reputations. These attacks often don't require advanced technical skills, as they can be executed using tools available online or via remotely controlled botnets. With the increasing reliance on electronic services, it has become essential to understand the nature of these attacks, their implementation mechanisms, and ways to prevent them, in order to enhance the security of digital infrastructure and ensure the continuity of services [3]. A DDoS attack on the internet is a serious threat, aiming to inundate the network, server, and targeted computers with fake traffic, requests, and corrupted data packets. This leads to the targeted service being shut down. Today, the internet is the target of millions of cyberattacks. This research contributes to the design and implementation of a model to accurately detect new and different types of DDoS attacks on the network, and to predict the type of each attack.

2. Related Work

Studied on the several DDoS attack categories and families to propose a new DDoS taxonomy for the application layer. The main contribution of this work is: i) reviewing the existing datasets comprehensively and propose a new taxonomy for DDoS attacks. ii) generating a new dataset, namely CICDDoS2019, which remedies all current shortcomings. iii) using the generated dataset, we propose a new detection and family classification approach based on a set of network flow features. iv) providing the most important feature sets to detect different types of DDoS attacks with their corresponding weights. The proposed system used four common machine learning algorithms namely ID3, Random Forest (RF), Naive Bayes, and logistic regression along. According to the weighted average of the three-evaluation metrics (Precision (Pr), Recall (Rc), F-Measure (F1)), the highest accuracy belongs to random forest with Pr =0.77, Rc=0.56, and F1=0.62 [4]. Suggest the Convolutional Neural Network CNN model to detect DDoS attacks. Authors have compared their proposed model with the classification algorithms like decision tree (D-Tree), support vector machine (SVM), K-nearest neighbors (K-NN), and neural network (NN) over two datasets, dataset 1 (simulated network traffic) captured from simulated MCC network by Wireshark and dataset 2 (NSL-KDD). The class of attack dealt in this study are Dataset 1: TCP and HTTP Flood DDoS Attack; NSL-KDD: DoS, Probe, R2L, U2RIt.It has been observed that the proposed model performed well compared to the other four classification algorithms such as like DT, SVM, KNN, and NN and gives an accuracy of 99% on both datasets. In this approach one-column padding has been used to convert the data into matrix form. Thus, it can affect the learning of the model [5]. Proposed a DAD-MCNN (i.e. multichannel CNN) framework to detect DDoS attacks. The number of feature groups decides the number of channels. The authors have split the features into different levels, like packet level, host level, and traffic level. The authors have used the incremental training approach to train MC-CNN. The authors have conducted a sequence of tests over KDDCUP99, CICIDS2017 datasets for binary classification in both datasets and multiclass category in KDDCUP99 only. The classes of DDoS attack that used are: KDDCUP99 class: Normal, DoS, R2L, U2R, Probe and CICIDS2017 class: DoS/DDoS: Hulk, Heartbleed, slowloris, Slowhttptest, GoldenEye. The authors compared MC-CNN with CNN, LSTM (3 layers), and other shallow ML methods (RF, SVM, C4.5, and KNN). The results showed that MC-CNN outperformed the state-of-art methods for all binary and multiclass classification. Further, the authors have also changed the training dataset size and evaluated the CNN and MC-CNN. The results showed that MC-CNN with ccuracy: KDDCUP99 (2 class) = 99.18%, KDDCUP99 (5 class) = 98.54%, CICIDS2017 = 98.87% and these results proved the proposed method is better in the restricted dataset and helpful in building DDoS detection systems when the training data are relatively insufficient. There is no much difference in the results of multichannel and single channel models. Also, the multichannel models will increase the complexity and thus might not be suitable when validated over real-time scenarios [6]. Proposed a new model in order to detect malicious activities on a network by using machine learning and deep learning technologies. The proposed model combined Autoencoder based deep neural network algorithm to separate normal network traffic from the attacks, including Analysis, Fuzzers, Generic, DoS, Backdoors, Exploits, Shellcode, Worms, and Reconnaissance classes. The main objective of the proposed model is to use a hybrid model approach for exact classification of malicious network flow from packets. The autoencoder layer of the model learns the representation of the network flows. The second layer (deep neural network model) tries to find out the exact malicious activity class. The proposed model does not rely on signature-based detection or deep packet inspection methods. The authors have evaluated their model on the UNSWNB15 dataset and KDDCUP99 with different activation functions. The results obtained the best F1 results with ReLu activation function, i.e. 0.8985. The overall accuracy and precision for KDDCUP'99 are approximately 99% for activation functions softplus, softsign, ReLu, tanh. In this article, the focus is only on the activation functions [7]. recommended a study about the problem of DDoS attack detection in a Cloud environment by considering the most popular CICIDS 2017 benchmark dataset and applying multiple regression analysis for building a machine learning model to predict DDoS and Bot attacks by considering a Friday afternoon traffic logfile. The dataset chosen for experimentation consisted of five-day log records from Monday to Friday in csv format and it consists two class labels are Benign (Normal) and DDoS (attack). The total number of traffic packets in the log file included 225,746 traffic pack. The results show the ensemble model for Friday morning class, a prediction accuracy of 97.86% is achieved. Similarly, for the Friday afternoon log file, the prediction accuracy is obtained as 73.79% for 16 attributes obtained through information gain-based feature selection and regression analysis-based machine learning model. This work thus paved a way to show the importance of regression analysis in building a machine learning [8]. developed a hybrid intrusion detection mechanism, called LAE-BLSTM, for the detection of botnets in IoT networks. The LAE-BLSTM mechanism uses deep Bidirectional Long Short-Term Memory (BLSTM) and Long Short-Term Memory Auto encoder (LAE). The LAE is used for the dimensionality reduction of the feature, while the BLSTM is used to identify the traffic of botnet attacks from benign traffic in IoT networks. The Bot-IoT dataset used in the evaluation of performance, which demonstrates that the LAE-BLSTM mechanism reached a data size reduction ratio of 91.89%. As the data size of big network traffic features becomes smaller, the implementation of DL method in memory-constraint IoT devices seems to be more practicable for efficient botnet detection. In addition, deep BLSTM model, which were trained to analysis the long-term inter-related changes in low-dimensional feature set produced by LAE, demonstrated robustness against model under-fitting and over-fitting as well as good generalization ability. Therefore, LAE-BLSTM has proven to be efficient for botnet attack detection in IoT networks [9]. Suggested a DDoS attack detection method based on information entropy and deep learning. Firstly, suspicious traffic can be inspected through information entropy detection by the controller. Then, fine-grained packet-based detection is executed by the convolutional neural network (CNN) model to distinguish between normal traffic and attack traffic. Finally, the controller performs the defense strategy to intercept the attack. The proposed method implentaion on CICIDS2017 dataset for detection 15 DDoS classes which are : Benign, BForce, SFTP and SSH, slowloris, Slowhttptest, Heartbleed, Web BForce, Hulk, GoldenEye, XSS and SQL Inject, Infiltration Dropbox Download, Botnet ARES, Cool disk, DDoS LOIT, PortScans. The authors have compared their method with the DNN, SVM, and DT. The CNN achieved higher precision, accuracy, F1-score, and recall among them. The accuracy of it is 98.98%. The ROC curve of CNN is steeper than DNNs, SVM, and DT. The AUC of CNN is 0.949. There is a need to set the threshold value for the detection method based on information entropy [10].

Recommended an efficient deep learning-based DDoS attack detection framework in 5G and B5G environments. The proposed framework is developed by concatenating two differently designed Deep Neural Network DNN models, coupled with a feature extraction algorithm using Pearson Correlation Coefficient PCC. It is built to detect the DDoS attacks and the type of DDoS attacks encountered. The authors evaluated the proposed framework using four different scenarios over an industry-recognize CICDDoS2019 dataset with UDP LAG, SYN, DNS, MSSQL, NTP, SSDP, TFTP, NetBIOS, LDAP, UDP and Benign classes. Results illustrated that the framework could detect DDoS attacks with an accuracy of 99.66% and a loss of 0.011. Furthermore, the proposed detection framework results were compared with the existing approaches, i.e. KNN, SVM, DeepDefense, and CNN ensemble. The proposed framework outperformed all except the CNN ensemble. The CNN ensemble has better precision and recall than the proposed framework. The proposed model has a complex structure so it can take more detection time and thus can affect the model's performance in a real-time scenario [11]. This study was designed to detect malware using data mining techniques to provide a secure cloud environment. The Meraz 18 dataset was used in this study. To obtain the highest quality, the data was preprocessed and the top 10 relevant features were selected using the Pearson method. Two classification algorithms were used: random forest and neural network, which achieved an accuracy of 98.88% and 97.37%, respectively. Finally, the attack type was predicted using hierarchical clustering [12].

3. Methodology

This section explains the proposed model for pre-detection of DDos attacks. This is done by tracking the traffic that provides information about attacks in the proposed framework. Classification algorithms are used to detect these attacks, then predict this data and finally evaluate its effectiveness. This is done through several stages starting with data collection and then pre-processing this data. The next step is to classify this data by applying the neural network algorithm and then evaluate its performance and predict it using a K-means clustering algorithm.

3.1. Dataset Collection

The proposed model uses the DDos dataset [13]. This dataset is available online, making it easy for researchers to access and use for research and evaluation purposes. The uploaded dataset contains (225708) samples, of which 97683 are benign and 128025 are DDoS-infected, as shown in Figure 1. This dataset is considered relatively balanced, making it ideal for training and testing machine learning models. This dataset is classified as benign or DDoS-infected.

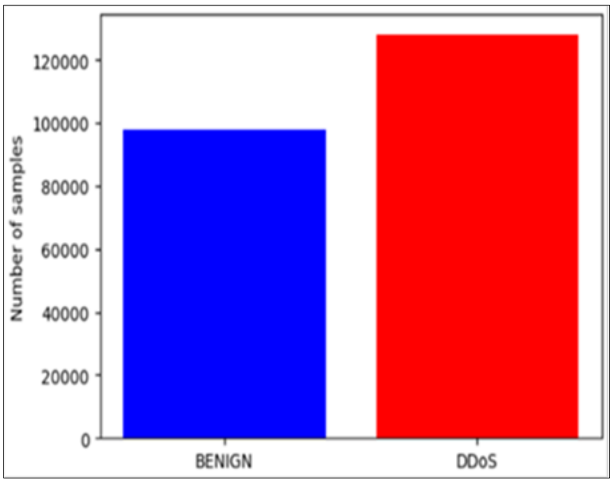

Fig 1: DDoS vs benign sample counts

3.2. Pre-processing Data:

Preprocessing a dataset is an important step in data mining. Figure 2 illustrates the steps involved in preprocessing, which involves analyzing the data to find missing values, inconsistent values, and even out-of-range values.

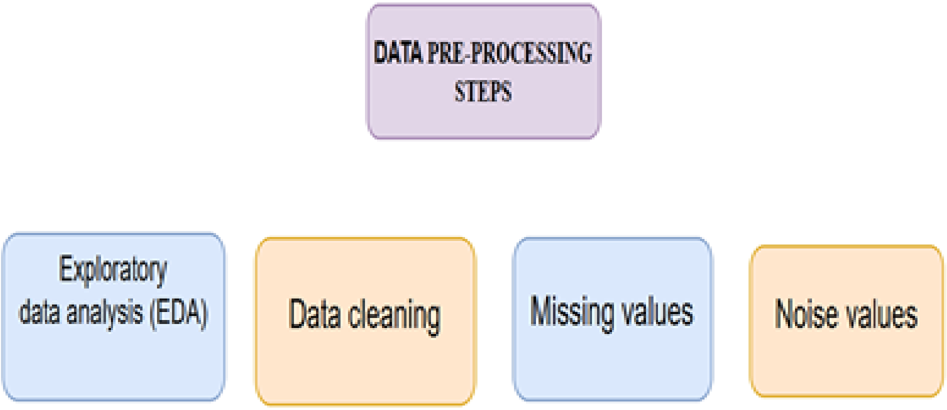

Fig 3: Pre- Processing Stage

These steps are explained as follows:

1. **Exploratory data analysis (EDA):** It represents a basic step after collecting data, so it is considered an important step in examining data for anomalies and extreme values. This is done by observing and understanding the data, so it helps the researcher analyze the data and identify natural patterns [14].
2. **Data Cleaning:** The process of identifying, correcting, or removing inaccurate, irrelevant, or incomplete data to ensure data accuracy and usability, making the data more suitable for analysis and decision-making, is a fundamental step in machine learning, providing reliable insights into the data [15].
3. **Removing Missing Values:** The presence of missing values negatively impacts the results and accuracy of classification due to the amount of missing information. Therefore, it is considered an essential step in the preprocessing process. Missing values are removed by identifying the row containing these values and excluding them entirely [16].
4. **Processing Noise Values:** Noise in data represents the presence of incorrect or irrelevant values in the data. There are many examples of irrelevant data, such as data with fixed (non-changing) values, as well as random values such as timestamp numbers and row numbers. All of this data is unnecessary in the classification process [17].

3.3. Classification

The dataset was divided into a training set and a test set, with the training set being 80% and the test set being 20%. The result of the split was (180566) training samples and (45142) test samples, as shown in Figure 3. This set was shuffled each time, and the randomness was set to (42). After that, a neural network algorithm was applied, and the accuracy, confusion matrix, precision, recall, and RMSE were calculated.

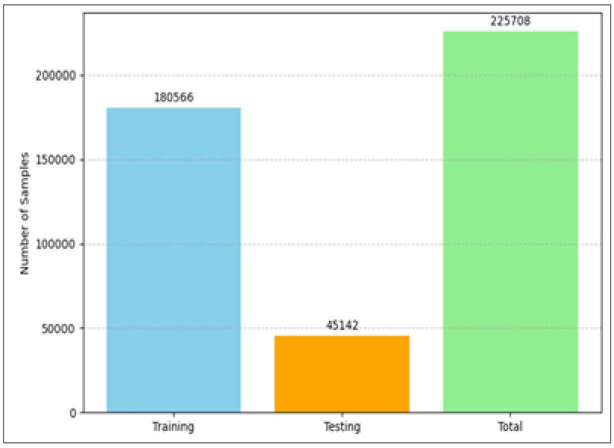

Fig 3: Training and testing samples number

3.4. Clustering

In the final stage of the proposed model, the K-means clustering algorithm is applied to generate clusters of DDoS attacks, each cluster being distinct based on its behavior in network traffic. The K-means clustering algorithm requires a pre-determination of the number of clusters to be formed, denoted by K. A selected element represents the cluster itself. When testing k from 1 to 4, with k = 3 being the optimal choice, each element in the dataset is assigned to its nearest center based on distance. These clusters are useful for more accurate analysis of DDoS attack behavior patterns. Clustering is used to predict similar patterns between attacks and group them according to their characteristics. This helps understand threat behavior and promotes a higher level of network. The proposed system is designed to be applied to any data flowing to the device, in order to prevent any DDoS attacks. The results obtained are as follows: The NN algorithm achieved an accuracy of (0.99) and an RMSE of (0.02) as shown in Figure 4, with a true positive sample of (19468), a true negative sample of (25655), a false positive sample of (16), and a false negative sample of (3). Figure 5 shows the NN confusion matrix with an accuracy of (1.00), a recall of (1.0), and an F1 score of (1.0).

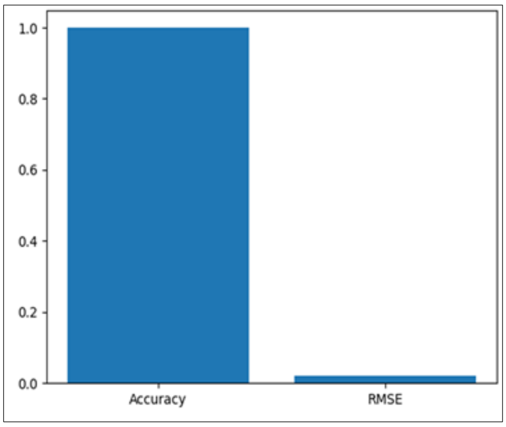

Fig 4: Accuracy and RMSE of NN

This section of the paper presents experiments using the proposed approach and discusses the results. The proposed model was evaluated using Jupiter, and a DDoS dataset was used to test the proposed model.

The neural network algorithm achieved an accuracy of 0.99 and RMSE of 0.01, as shown in Figure 6, with true positive samples (19480), true negative samples (25656), false negative samples (2), and false positive samples (4). Figure 7 shows the neural network algorithm confusion matrix with a recall of 1.00, a precision of 1.00, and an F1 score of 1.00.

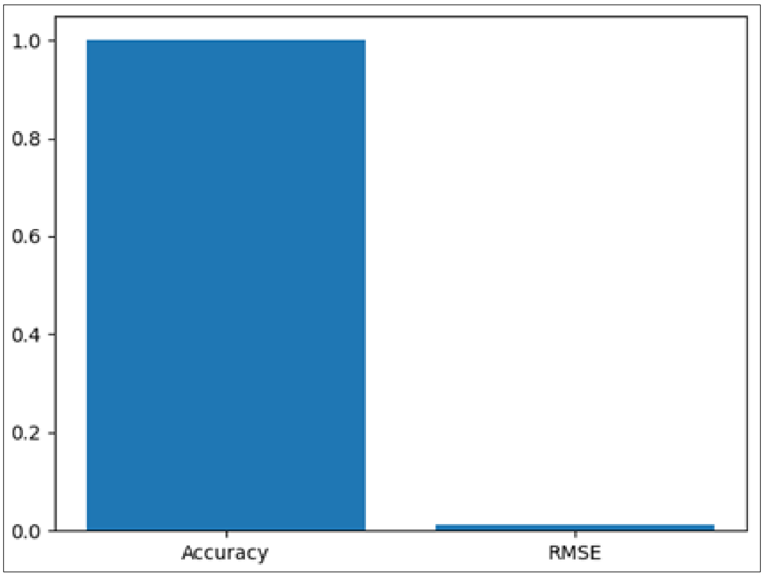

Fig 6: Accuracy and RMSE of neural network

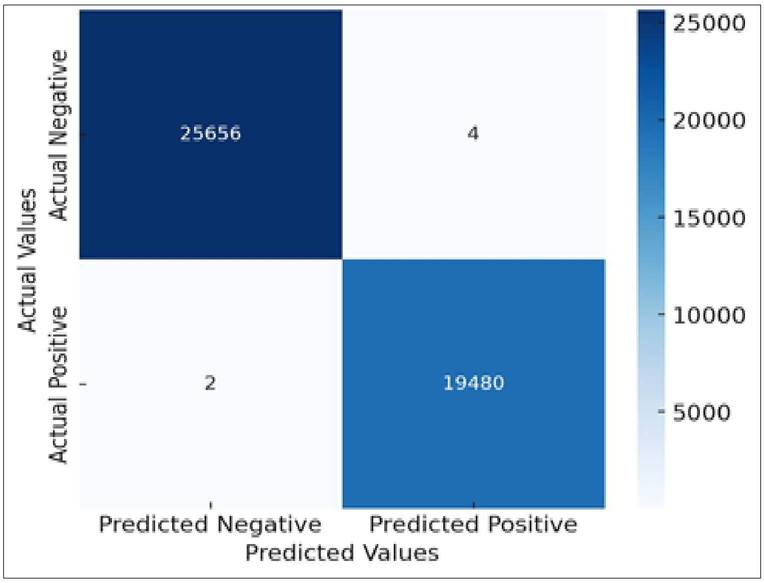

Fig 7: Neural network Confusion matrix

Figure 8 shows the results of applying the K-means clustering algorithm to the pre-classified DDoS attack data. The clustering algorithm was applied to the samples within the test set output from the neural network that were classified as DDoS attacks, representing 20% of the original data. The clustering results show that the samples were distributed into three clusters: A0, A1, and A2. A0 contains 11,748 samples, A1 contains 9,349, and A2 contains 4,571 samples, indicating the presence of three main expected attack patterns.

Figure 8 shows the results of applying the K-means clustering algorithm to the pre-classified DDoS attack data. The clustering algorithm was applied to the samples within the test set output from the neural network that were classified as DDoS attacks, representing 20% of the original data. The clustering results show that the samples were distributed into three clusters: A0, A1, and A2. A0 contains 11,748 samples, A1 contains 9,349, and A2 contains 4,571 samples, indicating the presence of three main expected attack patterns.

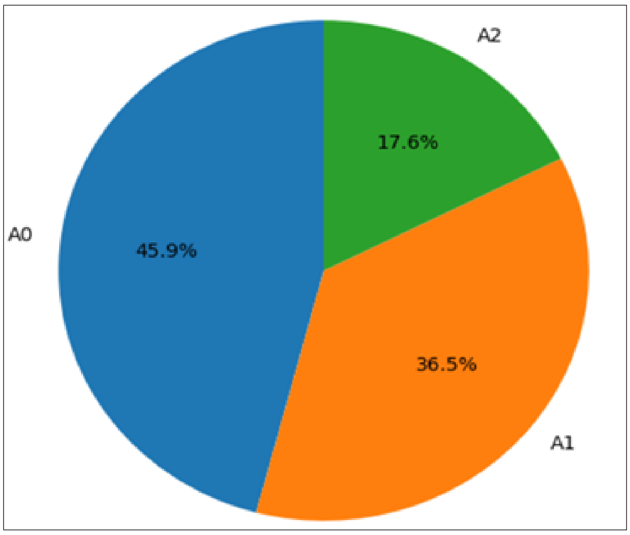

Fig 8: Analysis of Three K-means clustering

Cluster A0 contains the largest percentage of malicious samples compared to the other clusters. The overall statistics show that most traffic characteristics, particularly flow duration, packet length, and access times, are relatively low in this cluster. A1 includes a sample pool that is fairly moderate in most flow and packet-level metrics, representing moderate DDoS activity. Cluster A2 has the fewest samples, but has high values for flow duration, idle time, and byte count. In this case, distributed denial-of-service (DDoS) attacks consume more resources and can cause greater disruption. Therefore, this cluster is more dangerous.

4. Conclusion

In this research, distributed denial-of-service (DDoS) attacks were detected using machine learning, which is capable of accurately distinguishing between normal traffic and DoS traffic. Different machine learning classifiers were tested on a dataset to identify normal traffic and DDoS traffic. The results obtained from the, a neural network algorithm showed the highest accuracy is 0.99. Clustering was then used to place the attack into similar groups, with a K-means clustering algorithm clustering was used to predict the attack type. These preliminary results encourage further research in the field of anomaly detection using machine learning to protect networks from DDoS attacks.

5. Acknowledgement

The author would like to thank General Directorate of Education Babylon, Ministry of Education Iraq, for its support in this work.

Conflicts of Interest

The author declares no conflict of interest.

6. References

1. Pandey P, Kapoor A. Cybercrime in the digital era: impacts, awareness, and strategic solutions for a secure future. Sachetas. 2025;4(1):32-37.
2. Qamar R, Hussain Z, Arain AA, Zardari BA, Siraj S, Khan F. Distributed denial of service (DDoS) attacks technique to interruption the system's service and identification. Int J Comput Sci Netw Secur. 2022;22(10):153-161.
3. The danger of DDoS attacks stems from their ability to completely paralyze targeted systems, causing significant financial losses and negatively impacting organizations' digital reputations.
4. Sharafaldin I, Lashkari AH, Hakak S, Ghorbani AA. Developing a realistic distributed denial of service (DDoS) attack dataset and taxonomy. In: 2019 International Carnahan Conference on Security Technology (ICCST); 2019 Oct 1-3; Chennai, India. IEEE; 2019. p. 1-8. doi: 10.1109/ICCST.2019.8888418.
5. Shaaban AR, Abd-Elwanis E, Hussein M. DDoS attack detection and classification via convolutional neural network (CNN). In: 2019 Ninth International Conference on Intelligent Computing and Information Systems (ICICIS); 2019 Dec 8-10; Cairo, Egypt. IEEE; 2019. p. 233-238. doi: 10.1109/ICICIS46948.2019.9014819.
6. Chen J, Yang YT, Hu KK, Zheng HB, Wang Z. DAD-MCNN: DDoS attack detection via multi-channel CNN. In: Proceedings of the 11th International Conference on Machine Learning and Computing (ICMLC 2019); 2019 Feb 22-24; Zhuhai, China. New York: ACM; 2019. p. 484-488. doi: 10.1145/3318299.3318379.
7. Catak FO, Mustacoglu AF. Distributed denial of service attack detection using autoencoder and deep neural networks. J Intell Fuzzy Syst. 2019;37(3):3969-3979. doi: 10.3233/JIFS-190667.
8. Sambangi S, Gondi L. A machine learning approach for DDoS (distributed denial of service) attack detection using multiple linear regression. Proceedings. 2020;63(1):51. doi: 10.3390/proceedings2020063051.
9. Popoola SI, Adebisi B, Hammoudeh M, Gui G, Gacanin H. Hybrid deep learning for botnet attack detection in the internet-of-things networks. IEEE Internet Things J. 2021;8(6):4944-4956. doi: 10.1109/JIOT.2020.3034760.
10. Wang L, Liu Y. A DDoS attack detection method based on information entropy and deep learning in SDN. In: 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC); 2020 Jun 12-14; Chongqing, China. IEEE; 2020. p. 1084-1088. doi: 10.1109/ITNEC48623.2020.9084798.
11. Amaizu GC, Nwakanma CI, Bhardwaj S, Lee JM, Kim DS. Composite and efficient DDoS attack detection

framework for B5G networks. Comput Netw. 2021;188:107871. doi: 10.1016/j.comnet.2021.107871.

12. Amlak GMH, Al-Saedi KHK, Aljanabi KB. Cyber attacks detection and type prediction for cloud system using machine learning techniques. In: AIP Conference Proceedings 3207; 2024 Sep. Melville (NY): AIP Publishing; 2024. 040003. doi: 10.1063/5.0221634.
13. Kfoury EF, Crichigno J, Bou-Harb E. An exhaustive survey on P4 programmable data plane switches: taxonomy, applications, challenges, and future trends. IEEE Access. 2021;9:87094-87155. doi: 10.1109/ACCESS.2021.3089390.
14. Komorowski M, Marshall DC, Salciccioli JD, Crutain Y. Exploratory data analysis. In: Secondary Analysis of Electronic Health Records. Cham: Springer; 2016. p. 185-203. doi: 10.1007/978-3-319-43742-2_15.
15. Mumuni A, Mumuni F. Automated data processing and feature engineering for deep learning and big data applications: a survey. J Inf Intell. 2025;3(2):113-153. doi: 10.1016/j.jii.2024.100093.
16. Palanivinayagam A, Damaševičius R. Effective handling of missing values in datasets for classification using machine learning methods. Information. 2023;14(2):92. doi: 10.3390/info14020092.
17. Oleghe O. A predictive noise correction methodology for manufacturing process datasets. J Big Data. 2020;7:89. doi: 10.1186/s40537-020-00365-3.