MJSAT

Malaysian Journal of Science and Advanced Technology

journal homepage: https://mjsat.com.my/

A Comparative Review on DDoS Attack Detection Using Machine Learning Techniques

Zerin Hasan Sahosh[1], Azraf Faheem[1], Marzana Bintay Tuba[1], Md. Istiaq Ahmed[1], and Syeda Anika Tasnim*[1]

[1] *Department of Computer Science, American International University-Bangladesh. Dhaka, Bangladesh.*

KEYWORDS

Artificial Intelligence
Machine Learning
DDoS Attack
SVM
Random Forest

ARTICLE HISTORY

Received 27 October 2023
Received in revised form 16 February 2023
Accepted 16 February 2024
Available online 9 March 2024

ABSTRACT

The rapid growth of the internet and the increasing reliance on digital infrastructures have posed significant challenges to cybersecurity. Among the other variants of attacks, Distributed Denial of Service (DDoS) attacks have emerged as one of the most destructive and common threats. These attacks disrupt or slow down network services by overwhelming the network infrastructure with a massive volume of malicious traffic. To effectively identify and mitigate DDoS attacks, machine learning techniques have been extensively employed in intrusion detection systems. Machine learning approaches offer the advantage of automating the detection process by learning patterns and characteristics of DDoS attacks from historical data. Researchers have explored various machine learning algorithms such as K-Nearest Neighbours (KNN), Support Vector Machine (SVM), Random Forest (RF), and Naïve Bayes to classify and detect DDoS attacks. These algorithms leverage features extracted from network traffic data, including packet size, packet delay patterns, and traffic behaviour, to differentiate between normal and malicious traffic.

1. INTRODUCTION

The exponential rise of the internet and people's reliance on digital infrastructures have presented cybersecurity with new challenges in recent years. Internet services have become a vital need for both enterprises and individuals. With the growing demand for network-based services, network trespassers have also increased the volume of attacks on the network infrastructure to discontinue the response of services to genuine users. The attacks that prevent or slow down any network's services are commonly known as Denial of Service (DoS) attacks [1]. The first Denial of Services (DoS) attack was monitored back in 1974. Since DoS attack has upgraded to Distributed Denial of Services (DDoS) it has become more destructive [2].

Traffic-based attackers target the victim by sending large volumes of TCP and UDP packets via botnet to damage theperformance of the network [3]. To plot DDoS assaults, hackers frequently create software applications placed on computers, which they refer to as botnets. Another term for malware or an infected network (computer) from which DDoS attacks are conducted is botnets, which are controlled by hackers [4]. However, of a botnet's distribution strategy is spent setting the bots to assist potential future exploitation [5]. Therefore, numerous security tools such as antivirus and firewalls should be installed in computer networks to protect important data and services from trespassers.

DDoS attacks can be categorized into a number of different types, although they are typically divided into three classes. They are the following: application assault, bandwidth/volume attack, and traffic/fragmentation attack [6]. The two most common application layer attacks today are HTTP flood and UDP flood. Using HTTP GET or POST requests, an attacker can target a web server or application in an HTTP flood [7]. To defend against DDoS assaults, the intrusion detection system mainly depends on machine learning algorithms. Smurf assaults, HTTP POST/GET attacks, SQL Injection DoS, and other types of DDoS attacks are only a few of the many DDoS attack types that are currently being used. Numerous datasets that are available to the public are outdated and lack the most recent attack flow [8].

The following diagram shows the simplified architecture of DDoS attack:

*Corresponding author:
E-mail address: Syeda Anika Tasnim <anika.tasnim@aiub.edu>.
https://doi.org/10.56532/mjsat.v4i2.208

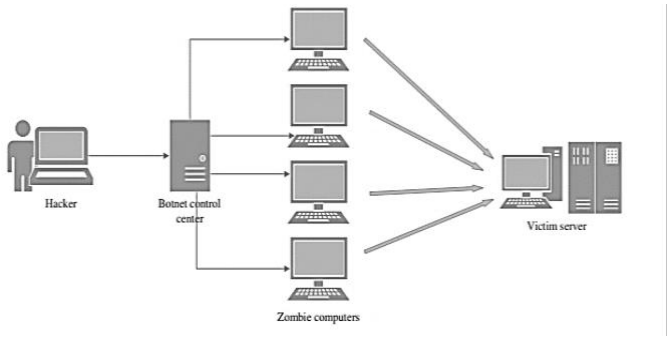

Fig. 1. DDoS Architecture[9].

DDoS attack occurs in a very short period. Due to the rise in DDoS assaults and the outdated network-based methods' inability to detect them, it is crucial to create new attack detection systems. The detection and prevention of various DDoS assaults have been proposed using data mining techniques and machine learning-based algorithms [10]. So, the detection system must be very fast and should include all the new patterns of DDoS attacks. The usage of the machine learning approach helps to classify the attack patterns and the characteristics. Machine learning (ML) is used to overcome the issue because it can produce more accurate findings [11]. Machine learning or deep learning techniques are necessary to build a system that can understand the traffic behaviour, that can detect the attacker's packet size, that can recognize the pattern of the packet delay, and that can detect the attack efficiently.

2. BACKGROUND STUDY

Both the scientific community and industry have been researching DDoS detection and mitigation for an extended period [12]. Nowadays, DDoS assaults are the most frequent and vulnerable attacks on network systems, costing commercial and individual infrastructures millions of dollars. These attacks are controlled by Botnets. A botnet is a group of several computers that have been infected with malware and cooperate to carry out repetitive tasks. A single attacking group in charge of the botnet is known as the "bot-master" [13] The biggest threat to the IT sector is DDoS attacks, and their frequency is rising drastically each year [14]. The greatest DDoS attack to date, according to Amazon Web Services (AWS), was a 2.3 Tbit/s attack in 2020 [15]. By 2023, there will be 15.4 million DDoS attacks worldwide, predicts Cisco, and 50% of DDoS attacks, were directed against financial institutions [16]. According to studies, almost 50 billion IoT devices have been used by 2020 and most of the IoT devices carry the risk of getting infected with malware [17].

For instance, configuring a router with a Cisco vendor and configuring a Juniper router are two different processes. In a complicated computer network, there is a significant likelihood of human mistakes or configuration errors. To solve this problem, the Software-Defined Network (SDN) was developed. SDN divides the vertical abstractions in conventional network devices into two fundamental layers: the controller and the forwarding layer. When a centralized control system is implemented in this manner, a security vulnerability is created that an attacker could exploit to change the topology directly connected to a controller. DDoS attacks can therefore be carried out both locally and globally [18].

To increase the number of successful attack chances malicious attackers redesign and update the sizes, volumes, and frequencies of their attacks. So, organizations that rely on IT to conduct their operations must come forward to protect themselves from this kind of cyber-attacks [19]. It is impossible to identify DDoS attacks manually [20]. However, as the whole world is becoming automated, introducing machine learning methods to detect this type of attack would be revolutionary. K-Nearest Neighbor (KNN), Logistic Regression, Random Forest (RF), Support Vector Machine (SVM), Naive Based Classifiers, etc. are only a few examples of the many machine learning approaches [21]. To build a system that can track DDoS attacks rapidly and efficiently by observing feature extraction and classification, size of the server-traffic, request types and protocols, it is necessary to use different machine learning methods.

3. COMPARATIVE REVIEW BASED ON METHODS

Numerous researchers brought up multiple ML-based DDoS attack detection models [22]. The CIC IDS 2017 dataset is used in [23] to identify DDoS attacks. To evaluate several models, it applies data pre-processing and K-fold cross-validation. The study concludes that the Random Forest algorithm outperforms other models in swiftly recognizing DDoS attacks based on evaluation metrics including recall, accuracy, precision, and FAR (False Alarm Rate). The analysis carried out in [24] sheds light on various DDoS attack intentions and launch techniques. Offering a thorough overview of the changing DDoS assault scene and defence techniques, it also analyses various intrusion detection methodologies. The authors of [25] offer two approaches for identifying DDoS attacks. Using a mathematical model, the relationship between the inter-arrival time of requests and network performance is first established. Additionally, based on throughput analysis, a Machine Learning Model is built using Logistic Regression and Naive Bayes methods to detect DDoS assaults. The combination of these models offers a thorough method for detecting DDoS attacks. The CCIDS2017 dataset is used by the authors of [20] to train an algorithm for the categorization of DDoS assaults. The suggested technique uses the SVM classification algorithm to obtain an accuracy of 99.68% while considering parameters like packet size, packet length, flow time, forward and backward packets, and other packet properties. This shows how reliable the method is in correctly identifying DDOS assaults. Like that, the research described in [2] focuses on the CCIDS2017 dataset for the identification of DDoS assaults.

The feature vector dimensions were decreased using PCA, and the neural network model was made shorter to reduce the time complexity. When choosing the output dimension, PCA is more versatile than Linear Discriminant Analysis (LDA) and other linear dimensionality reduction techniques [26]. The dataset includes a variety of packet attributes, including TCP flags, flow time, header length, and packet length. The suggested technique achieves a high accuracy of 99.68% in identifying DDOS traffic by training a Support Vector Machine (SVM) classification algorithm on this dataset, proving its efficacy in separating attack traffic from regular traffic. The suggested study in [27] uses a variety of machine learning

models to assess the input dataset to detect DDoS attacks. By utilizing Principal Component Analysis (PCA) and Random Forest classifier to validate the feature significance, the authors increase accuracy. In terms of accuracy, the Decision Tree model outperforms the other classifiers that were tested, showcasing its potential as an effective tool for DDoS attack detection.

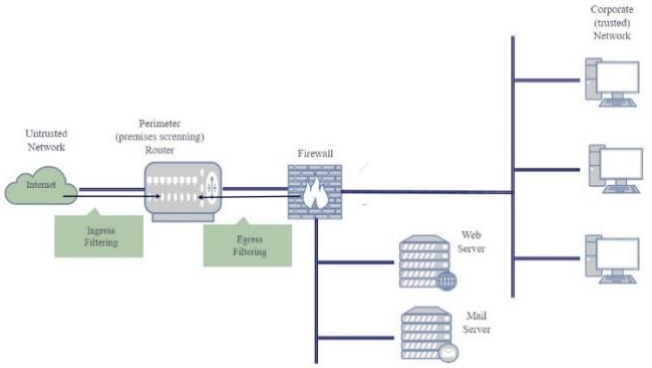

Fig. 2. The Ingress/Egress Method of Packet Filtering to Prevent a DDoS Attack [19].

The necessity of prevention in reducing DDoS attacks is emphasized by [19]. The authors suggest using packet filtering to detect and remove malicious packets to avoid DDoS assaults. The need for proactive defensive tactics is shown by the discussion of a few techniques, including ingress/egress packet filtering, router-based packet filtering, and statistical approaches like Packets Core.

4. COMPARATIVE REVIEW BASED ON FRAMEWORKS

To generate DDoS Detection Models, Alghoson et al. introduced the Light Gradient Boosting framework [43] which is a framework for gradient boosting that is small, quick, dispersed, and effective. It has the potential to be utilized for a range of machine learning applications, which comprises classification and ranking. It operates by dividing the tree with the best match into its individual branches. On the other hand, other boosting algorithms, divide the tree in two ways: depth-wise and leaf-wise [16].

Rahman, M. A. developed a framework to track anomalies that includes several steps, such as feature selection, data pre-processing, data analysis to apply various machine learning algorithms, training the dataset to algorithms, testing the dataset, and contrasting the results with various algorithmic approaches [18]. the creation of a classifier that can tell malicious packets apart from normal ones. The assault is initially detected by this model's detectors, which then stop it or lessen its strength. When this detector gets a request from a web client, it may determine if the request corresponds to the DDoS class or not by attempting to identify malicious packet. This has been detected because the request did not behave as anticipated.

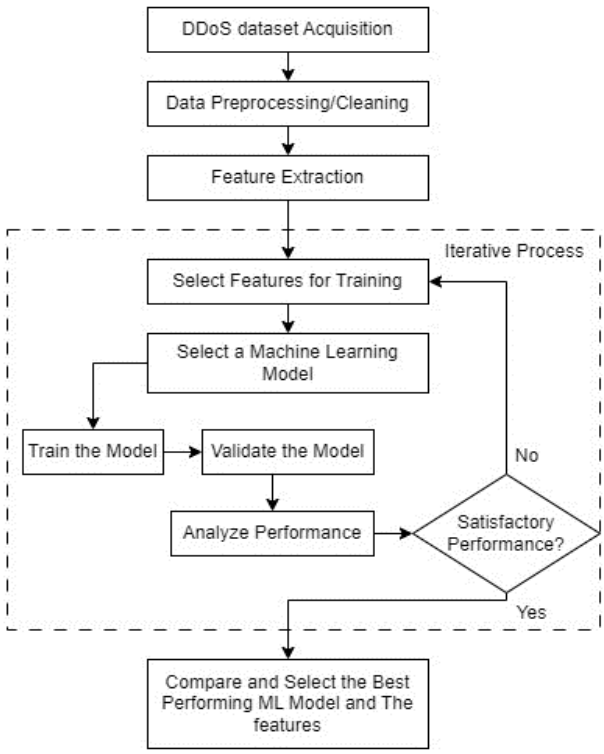

Fig. 3. Generalized Framework for ML-Based DDoS Detection System.

The generalized framework presented in Fig. 3. contains the essence of various research conducted on machine learning-based DDoS detection systems.

5. COMPARATIVE REVIEW BASED ON ALGORITHMS

Examples of machine learning algorithms include Support Vector Machines, Artificial Neural Networks, Genetic Algorithms, k-means, AdaBoost, Apriori, Cluster Analysis, C4.5, k-nearest Neighbors, and Naive Bayes [28].

5.1 SVM (Support Vector Machine)

A learning technique called Support Vector Machine (SVM) is based on statistical learning theory which applies the supervised method to perform classification and regression. An SVM algorithm generates a design that predicts the new example falling into one of the two categories based on a set of trained examples, each of which is designated as a method and split into two classifications [29]. It doesn't need a lot of training data to generate effective classification results. The SVM classifier divides the input data into several groups and produces a hyperplane. The best possible data separation is attempted by this hyperplane. This strategy was initially proposed by Vapnik, and it has subsequently produced good results to increase interest in ML research. Using supervised learning, SVM can do regression and classification [30]. C M. [23] used SVM algorithm in their proposed method. Mahmood [31] proposed a method for creating DDoS attack detection that used SVM. Ashutosh [2] and Jiangtao [20] also used SVM in their proposed methods. Eventually, according to paper [32] SVM is more stable than other machine learning techniques.

5.2 KNN (K-Nearest Neighbours)

The k-NN (k-Nearest Neighbors) approach is a well-known similarity-based learning algorithm that excels in a wide range of problem areas, including classification challenges [33]. It categorizes test data observations into groups according to how closely they are related to other individuals in the same class. KNN employs a non-parametric strategy to categorize samples. The unlabelled points are then allocated to the neighbor class K. Mona [34] and Jeswin [35] also utilized KNN in their model for identifying DDoS attack detection. The distance between unique places is estimated using the input vectors.

5.3 Naïve Bayes

The Naive Bayes probabilistic classifier is a user-friendly tool. It is based on the idea that the values of other variables have no bearing on the effects of one variable on a particular class. This presumption is known as class conditional independence [24]. This approach, which is based on Bayesian networks, is used to carry out the classification operation. Naive Bayes (NB) is frequently acknowledged as the most fundamental and straightforward method for creating classifiers. The class labels for issue scenarios are determined by classifiers which shows feature value vectors after that. A few sets will have been used to produce the class labels. Parvinder [36] described a technique for detecting DDoS attacks that statistically analysed network traffic using NB. The NB classifier was also used to detect DDoS attacks as a fully designed, operational model.

5.4 RF (Random Forest)

The Random Forest classifier is an array of classifiers that uses various decision tree processing techniques and classifies the results according to the mood of each individual tree [37]. Leo Breiman created the popular machine learning technique known as random forest (RF), which is used for classification. Different decision trees are produced by the random forest where each tree is created using an independent bootstrap test and a tree classification method using the first set of data [38]. The Random Forest algorithm is made up of decision trees that could be applied for classification and regression. For classification, decision trees are used to make a majority of the predictions whereas the average of the tress's output is considered as the result of regression. RF models include a few benefits, including the quickest model training time, the capacity to handle inconsistent datasets, a classification method for embedded features, and internal metrics for assessing the influence of features [34]. RF is also called random decision forest and represents a supervised machine-learning algorithm used to classify regression problems. It's method of operation entails building several decision trees during the training phase and delivering a class output, which is either the mode of the examined classes or an average forecast of the particular trees. RF algorithm was employed by C M [23] in their suggested technique. To identify DDoS attacks, Mona [34] and Jeswin [35] used RF in their model.

5.5 LR (Logistic Regression)

Logistic regression is used for binary classification tasks. It is a modified version of linear regression which applies a logistic function known as sigmoid function to transform the output into a probability score between 0 and 1. This probability decides a data point belonging to a certain class. Olga [9] and Neeraj [13] proposed a method for creating DDoS attack detection that used LR.

Table 1. Summary of algorithms used in reviewed papers

Paper References	Algorithms							
	J48	SVM	NB	RF	CNN	KNN	DT	LR
[2]		✓						
[9]		✓				✓		✓
[13]		✓				✓	✓	✓
[14]		✓	✓			✓		
[15]		✓			✓			
[16]			✓			✓	✓	✓
[18]		✓		✓		✓		
[19]				✓	✓			
[20]		✓		✓				
[23], [31]		✓	✓	✓		✓	✓	✓
[24]		✓	✓			✓		
[25]			✓					✓
[27]					✓	✓		
[31]		✓	✓				✓	
[34]				✓		✓		✓
[35]		✓		✓		✓		
[39]	✓		✓					
[40]					✓			
[41]		✓					✓	
[42]						✓		
[43]				✓	✓		✓	
[44]		✓						
[45]				✓				
[46]		✓						
[47]				✓				
[48]						✓		

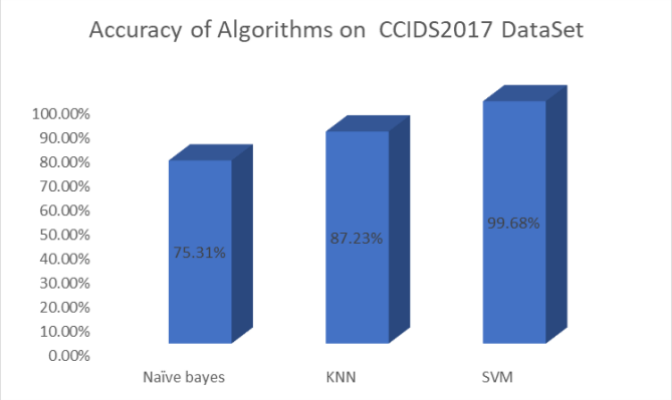

Fig. 4. Accuracy According to Different Algorithms on CCIDS2017 Dataset [2].

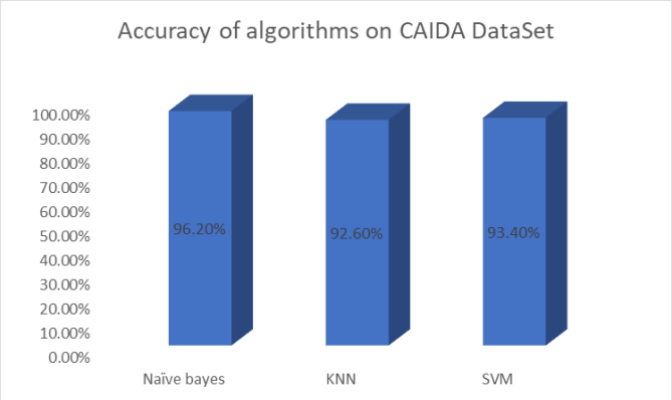

Fig. 5. Accuracy According to Different Algorithms on CAIDA Dataset [24].

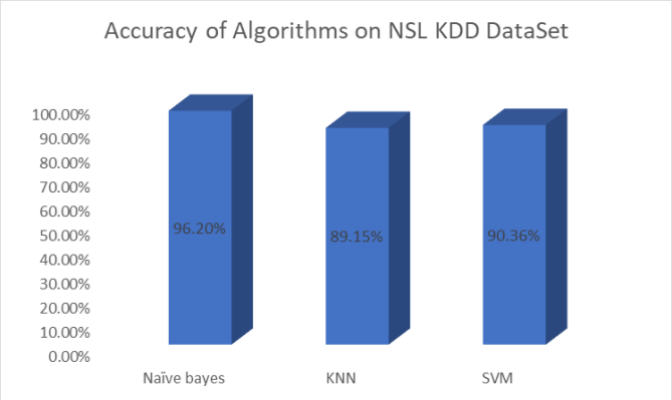

Fig. 6. Accuracy According to Different Algorithms on NSL-KDD Dataset [13].

6. Comparative Review Based on Datasets

CIC IDS 2017 dataset was generated by interpreting the behaviour of twenty-five users over the course of five days utilizing several application layer protocols, including HTTP, FTP, HTTPS, and SSH [49]. Before initiating the training of a machine learning model, the data must be pre-processed. The data set was pre-processed using a variety of methods. Other pre-processing methods include category value encoding, addressing missing values, and null value elimination. Categorical variables lacking numerical values, including source-destination IP and protocol, were encoded using one-hot encoding [50].

The CIC IDS 2017 dataset, according to [25], offers network traffic analysis findings gathered based on variables including source IP, destination IP, source and destination ports, timestamp, protocols, and more. It is frequently used to categorize whether DDoS or non-DDoS assaults are present [19]. The study used the CCIDS2017 dataset, which includes safe and current frequent assaults that resemble data from the actual world (PCAPs).

Time stamps, source and destination IP addresses, source and destination ports, protocols, and attack-based flows are all included in this data collection. It also includes the outcomes of the network traffic analysis performed by CICFlowMeter. More than 80 network flow features are available [41]. Results on DDoS assaults and network traffic categorization using an ANN-based model were given by Perakovic et al. in 2017. On multiple datasets, their model had the greatest accuracy of 95.6% [25].

Using JNNS and snort AI, Sabah M. Alzahrani et al. (2017) achieved 98% accuracy for DDoS attack detection on UDP and TCP protocols in another research [23]. A heuristic method based on Single Value Decomposition (SVD) was used by Bin Jia et al. (2017), surpassing random forest, KNN, and bagging with a precision of 99.84% [31].

Relevant datasets were obtained to train and evaluate the machine learning models, including the CIC-DDoS2019 dataset, which presents current and benign DDoS assaults that are representative of actual attack data [24]. Initial dataset screening involved removing incomplete and missing data, followed by feature extraction using RF selection methodology. The machine learning models were trained on the selected DDoS dataset, and the outcomes were contrasted to identify the top models and associated feature sets. The present study's major focus is on processing data for DDoS attacks against SDN. The relevant dataset includes malicious and friendly TCP, UDP, and ICMP traffic, such as TCP SYN, UDP floods, and ICMP attacks. It possesses 23 traits that w[1]. The first Denial of Servi which was accessible via the Kaggle dataset repository, was used to evaluate the method's performance. The dataset includes instances from the following network data types: Normal, DoS, Probing, User to Root (U2R), and Remote to User (R2L) [40].

The researchers used the NSL-KDD dataset, a well-known benchmark dataset for DDoS detection [44], to assess the effectiveness of the proposed model. The NSL-KDD dataset is made up of a small sample of network traffic that has been segmented into a variety of groups, such as regular traffic, DoS assaults, probing attacks, user-to-root (U2R) attacks, and remote-to-user (R2L) attacks.

The data set was already divided into two subsets: a training dataset with 125,973 rows and a test dataset with 22,544 rows. This data collection included a thorough selection of network traffic situations, enabling efficient training and assessment of the suggested model. The authors' study concentrated on the processing of information primarily relevant to DDoS assaults on Software-Defined Networks (SDN) [20]. The researchers used a dataset generated specifically for this that includes both regular TCP, UDP, and ICMP traffic as well as malicious traffic, such as TCP SYN,

UDP flood, and ICMP assaults. To give extensive details on network traffic, 23 attributes from switches were added to the dataset. The dataset has 104,345 rows, 63,335 of which have been classified as benign, and 40,504 of which have been classified as malicious. The researchers were able to use this information to investigate and create a model that is specially designed for identifying DDoS assaults in SDN setups.

The CIC-DDoS2019 dataset, which provides a useful collection of recent and benign DDoS attacks with traits like genuine attack data, was used by the researchers in their study [20]. The dataset was accessed online and underwent initial screening to remove incomplete and missing data. Feature extraction was performed using the RF selection methodology, and a heat map was generated to visualize the extracted features. The features were divided into three main sets, and different machine-learning models were trained and evaluated on each feature set using the selected DDoS dataset.

The dataset includes a variety of DDoS assaults, including those on DNS, UDP, TFTP, LDAP, MSSQL, Net-BIOS, SNMP, Syn, NTP, UDPLag, and WebDDoS. The researchers were able to properly train the machine learning models and evaluate their performance by dividing the dataset into training and testing groups. Several datasets were utilized to train and analyse the machine learning models used to assess DDoS attacks [3]. These databases, which are accessible online, include details on different DDoS assaults. The current study, however, concentrated on handling DDoS attack data in the context of Software-Defined Networks (SDN). A link to a particular dataset that fits with the study's goals was supplied by the authors[20]. This dataset primarily focuses on TCP, UDP, and ICMP protocols and contains both benign and harmful network traffic. It comprises 23 characteristics that were taken from switches and provides thorough details on network traffic. To effectively analyze and detect DDoS assaults in SDN systems, the dataset was properly tagged, with 63,335 occurrences classed as benign and 40,504 instances classified as malicious.

The researchers used the NSL-KDD dataset, which has been widely used in the DDoS detection industry [18]. A training dataset with 125,973 rows and a test dataset with 22,544 rows had previously been produced from the entire data set. This dataset contains a variety of examples, including common, DoS, probing, user-to-root, and remote-to-user (R2U) attacks. It is a modified version of the 1999 KDD Cup dataset. The NSL-KDD dataset, which has served as a benchmark dataset in several research projects, is a useful resource for assessing the effectiveness of DDoS detection systems.

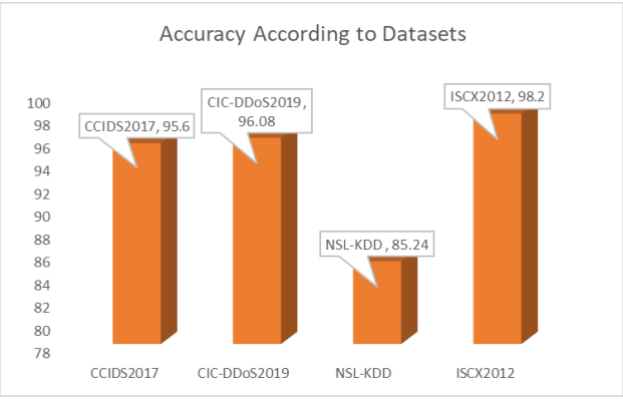

Fig. 7. Accuracy According to Datasets (In Percentage).

7. COMPARATIVE REVIEW BASED ON RESULTS

Several algorithms have been evaluated for their accuracy in various research papers. Naive Bayes, as described in Paper [2], achieved an accuracy of 75.31%, while Paper [24] reported an impressive accuracy of 96.2%. However, in general, Naive Bayes tended to achieve lower accuracies compared to other algorithms mentioned in the table.

Support Vector Machine (SVM) showed promising results across multiple papers. Paper [2] reported an accuracy of 99.68%, showcasing its effectiveness. Similarly, Paper [16] achieved a respectable accuracy of 93.4%. In another study, Paper [18] achieved accuracies of 91% with SVM_OVO and 96% with SVM_POLY. Furthermore, Paper [6] obtained an accuracy of 90.36%, further solidifying the overall success of SVM.

K-Nearest Neighbors (KNN) also exhibited good accuracy, although slightly lower when compared to SVM and Random Forest. Paper [16] achieved an accuracy of 92.6%, while Paper [6] reported an accuracy of 89.15% for KNN.

Random Forest, as evaluated in various papers, demonstrated exceptional performance. Paper [12] achieved an astounding accuracy of 99.99%, while Paper [18] reported an accuracy of 96%. These results indicate that Random Forest consistently delivered high accuracies and proved to be a robust algorithm.

Convolutional Neural Network (CNN), as described in Paper [12], achieved an accuracy of 99.53%, which is comparable to the performance of Random Forest. This suggests that CNN is a powerful algorithm for the tasks investigated in the respective paper.

Logistic Regression yielded mixed results in different studies. Paper [5] achieved an accuracy of 65%, indicating relatively lower performance. However, Paper [6] reported an accuracy of 90.4%, showcasing higher accuracy for Logistic Regression in that study.

ADA Boost, while not reaching the high accuracies of some other algorithms, achieved an accuracy of 84.57% in Paper [6]. This suggests that ADA Boost can still provide reasonably good results, although it may not be the top-performing algorithm in every scenario.

Lastly, Decision Tree achieved moderate accuracy in Paper [6], with an accuracy of 82.28%. Although not as high as some other algorithms, Decision Trees can still be a useful tool in certain contexts.

Table 2. Summary of different algorithms and their accuracy

Paper Ref.	Paper Title	Used Algorithm	Accuracy
[2]	DDOS Attack Detection Using Machine Learning	Naive Bayes	75.31 %
[2]	DDOS Attack Detection Using Machine Learning	SVM	99.68 %
[9]	Comprehensive DDoS Attack Classification Using Machine Learning Algorithms	Logistic Regression	65 %
[9]	Comprehensive DDoS Attack Classification Using Machine Learning Algorithms	Naïve Bayes	59 %

[13]	DDoS Attack Detection and Botnet Prevention using Machine Learning	Logistic Regression	90.4 %
[13]	DDoS Attack Detection and Botnet Prevention using Machine Learning	SVM	90.36 %
[13]	DDoS Attack Detection and Botnet Prevention using Machine Learning	KNN	89.15 %
[13]	DDoS Attack Detection and Botnet Prevention using Machine Learning	ADA Boost	84.57 %
[13]	DDoS Attack Detection and Botnet Prevention using Machine Learning	Decision Tree	82.28 %
[19]	Detecting Distributed Denial of Services Using Machine Language Learning Techniques	Random Forest	99.99 %
[19]	Detecting Distributed Denial of Services Using Machine Language Learning Techniques	CNN	99.53 %
[24]	DDOS Attack Detection Using Machine Learning for Network Performance Improvement	Naive Bayes	96.2 %
[24]	DDOS Attack Detection Using Machine Learning for Network Performance Improvement	SVM	93.4 %
[24]	DDOS Attack Detection Using Machine Learning for Network Performance Improvement	KNN	92.6 %
[27]	The Study of DDOS Attacks and Classification Performance Using Machine Learning Techniques	SVM_OVO	91 %
[27]	The Study of DDOS Attacks and Classification Performance Using Machine Learning Techniques	SVM_Poly	96 %

Overall, SVM, Random Forest, and CNN consistently performed well, achieving high accuracies in multiple papers. Naive Bayes generally had lower accuracies, while KNN showed decent performance. Logistic Regression, ADA Boost, and Decision Tree had mixed results, with varying accuracies across different papers. It's important to note that the accuracy numbers provided in the table represent specific studies and may not be directly comparable due to variations in datasets, features, and experimental setups.

8. DISCUSSION

In the area of cybersecurity, the use of machine learning (ML) algorithms for DDoS attack detection has shown substantial potential. Our review article investigates the challenging field of ML-based DDoS attack detection and offers an overview of several strategies and their effectiveness.

Researchers have adapted a range of techniques for the detection of DDoS attacks, the review highlights the importance of machine learning algorithms and proactive defence strategies such as packet filtering and statistical approaches, including ingress/egress packet filtering and router-based packet filtering. These findings collectively emphasize the significance of preventative measures and ongoing advancements in DDoS detection methodologies for enhancing network security.

The review of machine learning algorithms applied in DDoS attack detection reveals a diverse spectrum of techniques employed in different studies. To detect anomalies in network traffic Support Vector Machines (SVM) have gained significant attention in this domain due to their capacity for effective classification and regression.

The review of DDoS detection frameworks introduces two noteworthy approaches that contribute to the development of efficient detection models. Alghoson et al. proposed the utilization of the Light Gradient Boosting framework which focuses on branch division and facilitates efficient model creation [18]. In a complementary approach, Rahman, M. A. designed a comprehensive framework encompassing feature selection, data pre-processing, machine learning algorithm application, dataset training, and testing, as well as outcome comparison aiming to develop a robust classifier capable of distinguishing between malicious and benign packets, serving as an early detection system for DDoS attacks. It reflects a proactive approach to identify and mitigate threats by recognizing anomalous behaviours in web client requests [18]. These frameworks provide valuable tools for enhancing DDoS detection and exemplify the ongoing pursuit of effective solutions in the field.

The discussion on datasets in the systematic literature review highlights the diverse range of data sources and attributes used for DDoS attack detection research. The CIC IDS 2017 dataset, derived from the activities of twenty-five users over a five-day period and spanning various application layer protocols, is a common choice. Researchers employ data preprocessing techniques, including category value encoding and one-hot encoding, to prepare this dataset for machine learning model training. The dataset's attributes, such as source and destination IP addresses, ports, timestamps, and protocols, serve as a foundation for classifying DDoS and non-DDoS attacks.

The review of results across multiple research papers highlights the varying performance of machine learning algorithms in DDoS attack detection. It is evident that Naive Bayes generally tends to achieve lower accuracies compared to other algorithms discussed. Support Vector Machine (SVM) consistently demonstrated promising results across different studies. Nearest Neighbors (KNN) showed good accuracy, although slightly lower than SVM and Random Forest, with accuracies of 92.6% and 89.15%, respectively. Random Forest consistently delivered exceptional performance, with Paper [19] achieving an outstanding accuracy of 99.99%.

This review paper demonstrates the potential of machine learning (ML) for DDoS detection while also emphasizing the challenges that need to be addressed through ongoing research, along with the necessity for ongoing research to address its problems. The use of sophisticated ML algorithms in DDoS protection systems is anticipated to play a crucial role in protecting vital digital infrastructures as the threat landscape continues to change.

9. CONCLUSION

An article claims that it is a common problem in a dispersed network architecture to identify DDoS assaults. Since this type of attack disables access to cloud services, it is important to recognize it [34]. DDoS attacks based on zombies now affect normal traffic. Therefore, it is extremely difficult to

identify such an attack, even in the presence of stored attack traffic signatures.

Differentiating between authorized traffic and DDoS attack traffic is difficult. Due to their high memory requirements for data storage, existing algorithms for data classification are ineffective when used for real-time WBAN streaming data [51]. However, it could be challenging to differentiate between DDoS attacks with different rates and patterns and regular traffic. Throughout the years, a wide range of researchers have suggested numerous effective ML/DL methods for detecting DDoS attacks. Machine learning algorithms are used to categorize the requests, and a smart detection system makes use of this model to alert network administrators to dangerous requests. The most effective methods for using already-gathered data are those that use machine learning. With more data available, categorization accuracy improves[52] However, problems including a lack of datasets, hostile assaults, and model robustness continue to exist. Further investigation is required due to the requirement for interpretable ML models for security experts and regulatory compliance.

In this study, we have found some algorithms and frameworks. It might be useful to use those algorithms and frameworks to identify DDoS attacks. In the future, we will be focusing on the study to develop a unified autonomous model to detect any type of intrusion and network anomaly.

ACKNOWLEDGMENT

We would like to express our heartfelt appreciation to our mentors, supervisors, and the research community for their helpful suggestions and contributions to our review work on "DDoS Attack Detection Using ML." We also acknowledge the institutions and organizations that provided crucial resources. Furthermore, we appreciate the patience and support of our families and friends throughout this endeavour. This paper owes its existence to the collaborative efforts of these individuals and entities. Thank you all for your essential support in making this work possible.

REFERENCES

[1] Institute of Electrical and Electronics Engineers, Institute of Electrical and Electronics Engineers. Delhi Section, and I. INDIAcom (Conference) (14th : 2020 : New Delhi, 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom).

[2] A. Nath Rimal, R. Praveen, M. Tech Cyber Security Student, and A. Professor, "Issue 6 www.jetir.org (ISSN-2349-5162)," JETIR, 2020. [Online]. Available: www.jetir.org

[3] Mahāwitthayālai Songkhlānakharin. College of Computing, C. Electrical Engineering/Electronics, IEEE Thailand Section, and Institute of Electrical and Electronics Engineers, The 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology : ECTI-CON 2020 : 24-27 June 2020, virtual conference hosted by College of Computing, Prince of Songkla University.

[4] S. Sarraf, "Analysis and Detection of DDoS Attacks Using Machine Learning Techniques," American Scientific Research Journal for Engineering, [Online]. Available: http://asrjetsjournal.org/

[5] M. H. Aysa, A. A. Ibrahim, and A. H. Mohammed, "IoT Ddos Attack Detection Using Machine Learning," in 4th International Symposium on Multidisciplinary Studies and Innovative Technologies, ISMSIT 2020 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Oct. 2020. doi: 10.1109/ISMSIT50672.2020.9254703.

[6] S. Sambangi and L. Gondi, "A Machine Learning Approach for DDoS (Distributed Denial of Service) Attack Detection Using Multiple Linear Regression," MDPI AG, Dec. 2020, p. 51. doi: 10.3390/ proceedings2020063051.

[7] I. Sofi, A. Mahajan, and V. Mansotra, "Machine Learning Techniques used for the Detection and Analysis of Modern Types of DDoS Attacks," International Research Journal of Engineering and Technology, 2017, [Online]. Available: www.irjet.net

[8] Institute of Electrical and Electronics Engineers, 2020 European Conference on Networks and Communications (EuCNC).

[9] O. Ussatova, A. Zhumabekova, Y. Begimbayeva, E. T. Matson, and N. Ussatov, "Comprehensive DDoS Attack Classification Using Machine Learning Algorithms," Computers, Materials and Continua, vol. 73, no. 1, pp. 577–594, 2022, doi: 10.32604/cmc.2022.026552.

[10] G. Lucky, F. Jjunju, and A. Marshall, "A Lightweight Decision-Tree Algorithm for detecting DDoS flooding attacks," in Proceedings - Companion of the 2020 IEEE 20th International Conference on Software Quality, Reliability, and Security, QRS-C 2020, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 382–389. doi: 10.1109/QRS-C51114.2020.00072.

[11] K. S. Hoon, K. C. Yeo, S. Azam, B. Shunmugam, and F. De Boer, "Critical review of machine learning approaches to apply big data analytics in DDoS forensics," in 2018 International Conference on Computer Communication and Informatics, ICCCI 2018, Institute of Electrical and Electronics Engineers Inc., Aug. 2018. doi: 10.1109/ICCCI.2018.8441286.

[12] F. S. De Lima Filho, F. A. F. Silveira, A. De Medeiros Brito Junior, G. Vargas-Solar, and L. F. Silveira, "Smart Detection: An Online Approach for DoS/DDoS Attack Detection Using Machine Learning," Security and Communication Networks, vol. 2019, 2019, doi: 10.1155/2019/1574749.

[13] N. Patil, "DDoS Attack Detection and Botnet Prevention using Machine Learning," International Research Journal of Engineering and Technology, 2022, [Online]. Available: www.irjet.net

[14] D. S. Rajput, A. K. Upadhyay, M. Statistician, and E. Applications, "Hybrid Technique for DDOS Attack Detection Using Machine Learning," vol. 71, no. 4, 2022, [Online]. Available: http://philstat.org.phhttp://philstat.org.ph

[15] B. Nugraha and R. N. Murthy, "Deep Learning-based Slow DDoS Attack Detection in SDN-based Networks," in 2020 IEEE Conference on Network Function Virtualization and Software Defined Networks, NFV-SDN 2020 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Nov. 2020, pp. 51–56.

[16] K. B. Dasari and N. Devarakonda, "Detection of DDoS Attacks Using Machine Learning Classification Algorithms," International Journal of Computer Network and Information Security, vol. 14, no. 6, pp. 89–97, Dec. 2022, doi: 10.5815/ijcnis.2022.06.07.

[17] Institute of Electrical and Electronics Engineers, 2020 European Conference on Networks and Communications (EuCNC).

[18] F. D. Setiawan Sumadi and C. S. Kusuma Aditya, "Comparative Analysis of DDoS Detection Techniques Based on Machine Learning in OpenFlow Network," in 2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 152–157. doi: 10.1109/ISRITI51436.2020.9315510.

[19] B. Fakiha, "DETECTING DISTRIBUTED DENIAL OF SERVICES USING MACHINE LANGUAGE LEARNING TECHNIQUES," Xinan Jiaotong Daxue Xuebao/Journal of Southwest Jiaotong University, vol. 57, no. 5, pp. 675–688, Oct. 2022, doi: 10.35741/issn.0258-2724.57.5.55.

[20] J. Pei, Y. Chen, and W. Ji, "A DDoS Attack Detection Method Based on Machine Learning," in Journal of Physics: Conference Series, Institute of Physics Publishing, Jul. 2019. doi: 10.1088/1742-6596/1237/3/032040.

[21] D. Kumar, R. K. Pateriya, R. K. Gupta, V. Dehalwar, and A. Sharma, "DDoS Detection using Deep Learning," Procedia Comput Sci, vol. 218, pp. 2420–2429, 2023, doi: 10.1016/j.procs.2023.01.217.

[22] A. A. Saeed and N. G. M. Jameel, "Intelligent feature selection using particle swarm optimization algorithm with a decision tree for ddos attack detection," International Journal of Advances in Intelligent Informatics, vol. 7, no. 1, pp. 37–48, 2021, doi: 10.26555/ijain.v7i1.553.

[23] C. M. NalayiniI, J. Katiravan, and A. Professor, "Detection of DDoS Attack using Machine Learning Algorithms," 2022. [Online]. Available: https://ssrn.com/abstract=4173187

[24] D. Lunkad, G. Singh, and M. T. Student, "DDOS Attack Detection Using Machine Learning For Network Performance Improvement," 2020. [Online]. Available: www.ijcrt.org

[25] K. Kumari and M. Mrunalini, "Detecting Denial of Service attacks using machine learning algorithms," J Big Data, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40537-022-00616-0.

[26] Q. Li, L. Meng, J. Yan, and Y. Zhang, "DDoS Attacks Detection using Machine Learning Algorithms." [Online]. Available: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

[27] D. Sudheer et al., "The Study of DDOS Attacks and Classification Performance Using Machine Learning Techniques."

[28] A. Sanmorino, "A study for DDOS attack classification method," in Journal of Physics: Conference Series, Institute of Physics Publishing, Jun. 2019. doi: 10.1088/1742-6596/1175/1/012025.

[29] M. Arshi, M. D. Nasreen, and K. Madhavi, "A Survey of DDOS Attacks Using Machine Learning Techniques," in E3S Web of Conferences, EDP Sciences, Aug. 2020. doi: 10.1051/e3sconf/202018401052.

[30] V. N. Vapnik, "An Overview of Statistical Learning Theory," 1999.

[31] M. A. Al-Shareeda, S. Manickam, and M. A. Saare, "DDoS attacks detection using machine learning and deep learning techniques: analysis and comparison," Bulletin of Electrical Engineering and Informatics, vol. 12, no. 2, pp. 930–939, Apr. 2023, doi: 10.11591/eei.v12i2.4466.

[32] K. M. Sudar, M. Beulah, P. Deepalakshmi, P. Nagaraj, and P. Chinnasamy, "Detection of Distributed Denial of Service Attacks in SDN using Machine learning techniques," in 2021 International Conference on Computer Communication and Informatics, ICCCI 2021, Institute of Electrical and Electronics Engineers Inc., Jan. 2021. doi: 10.1109/ICCCI50826.2021.9402517.

[33] M. Zekri, S. El Kafhali, N. Aboutabit, and Y. Saadi, "DDoS Attack Detection using Machine Learning Techniques in Cloud Computing Environments."

[34] M. Alduailij, Q. W. Khan, M. Tahir, M. Sardaraz, M. Alduailij, and F. Malik, "Machine-Learning-Based DDoS Attack Detection Using Mutual Information and Random Forest Feature Importance Method," Symmetry (Basel), vol. 14, no. 6, Jun. 2022, doi: 10.3390/sym14061095.

[35] J. Johnson, S. George Associate Professor, and C. Dept, "Review on DDOS Detection using Machine Learning." [Online]. Available: www.ijert.org

[36] Institute of Electrical and Electronics Engineers, Institute of Electrical and Electronics Engineers. Delhi Section, and I. INDIAcom (Conference) (14th : 2020 : New Delhi, 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom).

[37] Anna University. Madras Institute of Technology, Anna University. Madras Institute of Technology. Department of Electronics Engineering, Institute of Electrical and Electronics Engineers. Madras Section., and Institute of Electrical and Electronics Engineers, 2017 Fourth International Conference on Signal Processing, Communication and Networking (ICSCN) : 16-18 March 2017.

[38] S. Pande, A. Khamparia, D. Gupta, and D. N. H. Thanh, "DDOS Detection Using Machine Learning Technique," in Studies in Computational Intelligence, Springer Science and Business Media Deutschland GmbH, 2021, pp. 59–68. doi: 10.1007/978-981-15-8469-5_5.

[39] A. Prasad, S. Prasad, K. Arockiasamy, and X. Yuan, "International Journal of Intelligent Systems and Applications in Engineering Detection of DDoS Attack in Software-Defined Networking Environment and Its Protocol-wise Analysis using Machine Learning." [Online]. Available: www.ijisae.org

[40] T. E. Ali, Y. W. Chong, and S. Manickam, "Machine Learning Techniques to Detect a DDoS Attack in SDN: A Systematic Review," Applied Sciences (Switzerland), vol. 13, no. 5, Mar. 2023, doi: 10.3390/app13053183.

[41] Z. He, T. Zhang, and R. B. Lee, "Machine Learning Based DDoS Attack Detection From Source Side in Cloud."

[42] D. Kumar, R. K. Pateriya, R. K. Gupta, V. Dehalwar, and A. Sharma, "DDoS Detection using Deep Learning," Procedia Comput Sci, vol. 218, pp. 2420–2429, 2023, doi: 10.1016/j.procs.2023.01.217.

[43] E. S. Alghoson and O. Abbass, "Detecting Distributed Denial of Service Attacks using Machine Learning Models." [Online]. Available: www.ijacsa.thesai.org

[44] M. A. Rahman, "Detection of Distributed Denial of Service Attacks based on Machine Learning Algorithms," International Journal of Smart Home, vol. 14, no. 2, pp. 15–24, Oct. 2020, doi: 10.21742/IJSH.2020.14.2.02.

[45] M. Shurman, R. Khrais, and A. Yateem, "DoS and DDoS attack detection using deep learning and IDS," International Arab Journal of Information Technology, vol. 17, no. 4A Special Issue, pp. 655–661, 2020, doi: 10.34028/iajit/17/4A/10.

[46] J. Ye, X. Cheng, J. Zhu, L. Feng, and L. Song, "A DDoS Attack Detection Method Based on SVM in Software Defined Network," Security and Communication Networks, vol. 2018, Apr. 2018, doi: 10.1155/2018/9804061.

[47] L. Chen, Y. Zhang, Q. Zhao, G. Geng, and Z. Yan, "Detection of DNS DDoS Attacks with Random Forest Algorithm on Spark," in Procedia Computer Science, Elsevier B.V., 2018, pp. 310–315. doi: 10.1016/j.procs.2018.07.177.

[48] S. Dong and M. Sarem, "DDoS Attack Detection Method Based on Improved KNN with the Degree of DDoS Attack in Software-Defined Networks," IEEE Access, vol. 8, pp. 5039–5048, 2020, doi: 10.1109/ACCESS.2019.2963077.

[49] 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA).

[50] H. Kamel and M. Z. Abdullah, "Distributed denial of service attacks detection for software defined networks based on evolutionary decision tree model," Bulletin of Electrical Engineering and Informatics, vol. 11, no. 4, pp. 2322–2330, Aug. 2022, doi: 10.11591/eei.v11i4.3835.

[51] H. Abbas, R. Latif, S. Latif, and A. Masood, "Performance evaluation of Enhanced Very Fast Decision Tree (EVFDT) mechanism for distributed denial-of-service attack detection in health care systems," Annales des Telecommunications/Annals of Telecommunications, vol. 71, no. 9–10, pp. 477–487, Oct. 2016, doi: 10.1007/s12243-016-0495-x.

[52] S. Peneti and Hemalatha, "DDOS Attack Identification using Machine Learning Techniques," in 2021 International Conference on Computer Communication and Informatics, ICCCI 2021, Institute of Electrical and Electronics Engineers Inc., Jan. 2021.