Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Detecting DDoS Attacks using Machine Learning: Survey

Sarah Zghair Arrak [1], *Rana Jumma Surayh Al- Janabi* [2]

1 College of Computer science & Information Technology , University of Al –Qadisiyah , Al –Diwaniyah , Iraq, it.mast.23.9@qu.edu.iq

2 College of Computer science & Information Technology , University of Al –Qadisiyah , Al –Diwaniyah , Iraq, Rana.aljanaby@qu.edu.iq

ARTICLE INFO

Article history:
Received: 13 /03/2024
Rrevised form: 30 /05/2024
Accepted : 26 /06/2024
Available online: 30 /06/2024

Keywords:

DDoS attack,

distributed denial of service,

detection,

machine learning algorithms(ML),

deep learning.

ABSTRACT

Phishing attacks have increased dramatically in recent years affecting many areas of society. Phishing attempts often use DDoS attacks to flood a server with too many requests, overwhelming it. DDoS attacks represent a major threat to cybersecurity and pose a significant risk to computer networks. Creating a solid defense system against these attacks is essential but complex due to the wide range of attack methods and complex networks and communication protocols. Ransom demands, revenge, rivalry, or other motives may trigger attacks. This survey discusses DDoS attacks, the advantages and disadvantages of detecting DDoS using machine and deep learning, and a framework for detection using machine learning and deep learning. And use their classifiers to detect DDoS attacks. Furthermore, we explore datasets used in related works. This research is necessary because DDoS attacks are diverse and pose a significant threat to computer networks.

MSC..

https://doi.org/10.29304/jqcsm.2024.16.21548

A. Introduction:

DDoS (Distributed Denial of Service) is an advanced DoS attack where malicious requests originate from multiple sources instead of just one. The DDoS attack involves multiple scattered sources worldwide that simultaneously target the server. The server now has to handle various sources. As a result, the server will get overloaded and deplete all its resources, including bandwidth, disc space, and memory capacity, leading to a denial of service for legitimate users. How does the attacker manage to enlist numerous machines in a DDoS attack? The assailant creates malicious software. He now needs to install this program on other internet users' PCs. He disseminates harmful software by infecting online pages or sending it as an email attachment to other people. When a user accesses these compromised

∗Corresponding author : sarah zghair arrak

Email addresses: it.mast.23.9@qu.edu.iq

Communicated by 'sub etitor': Dr. Rana Jumma surayh Al-Janabi websites or opens the email attachment, the software is installed on their computer without their awareness. Infected machines with malware installed are recruited to form an army to carry out attacks. This army is referred to as Botnet. This Botnet functions as a collective of computers poised to receive commands from its controller to launch attacks. The attacker schedules a specific date and time for the malware attack. When the scheduled time arrives, the Botnet launches simultaneous attacks on the server, causing it to crash. A DDoS attack can persist for hours or days, depending on the attacker's goal [1]. The rest of the paper is organized as follows: In Section II, Types of DDoS are explained in Section III. IDS is described in Section IV. Machine learning and deep learning are described in Section V. Related works are presented. In Section VI, the Dataset is explained. In Section VII, the discussion is presented. The VIII section contains the conclusion.

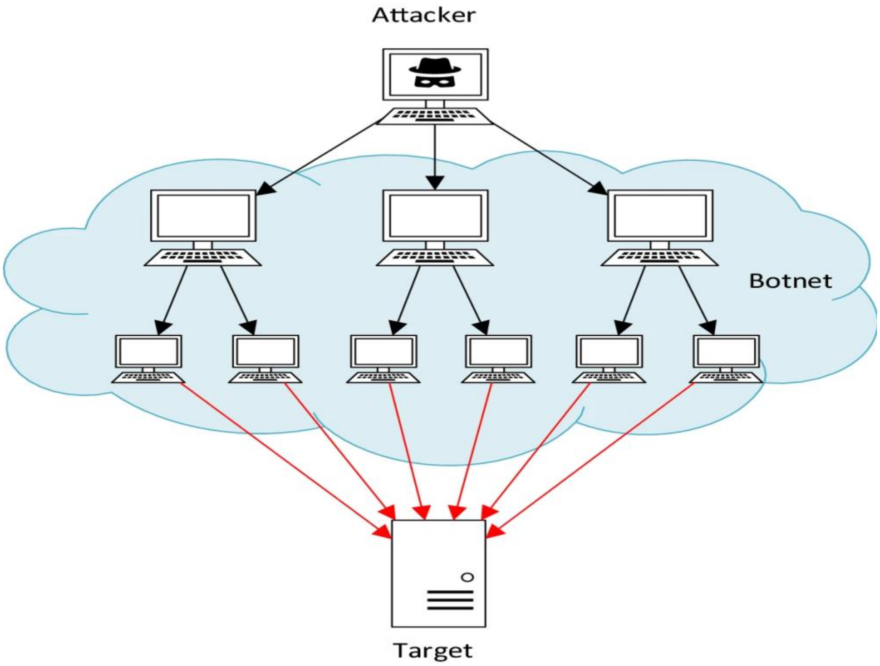

Fig.1 DDoS attack implementation scheme using a botnet [2].

B. Common types of DDoS attacks:

- HTTP flooding attack: This type of attack overwhelms a web server by sending large amounts of requests to it.
- User Datagram Protocol (UDP) flood attack: In this attack, many UDP packets from many zombies are sent to the victim's port. The victim's system searches for a response but cannot locate any applications associated with these UDP packets. This will return packets with unreachable destinations, confusing the victim's system. Thus, the opportunity to respond to legitimate requests is lost.
- ICMP (Ping) Flood Attack: In this type of attack, numerous ICMP echo requests or ping commands are sent, disrupting the victim's network and using its total capacity.
- The SYN flood attack capitalizes on weaknesses in the three-way handshake protocol, causing the zombie to send several SYN requests to initiate a TCP connection. The victim replies to the requests by sending SYN/ACK signals and anticipates a response from the zombie, which never materializes.

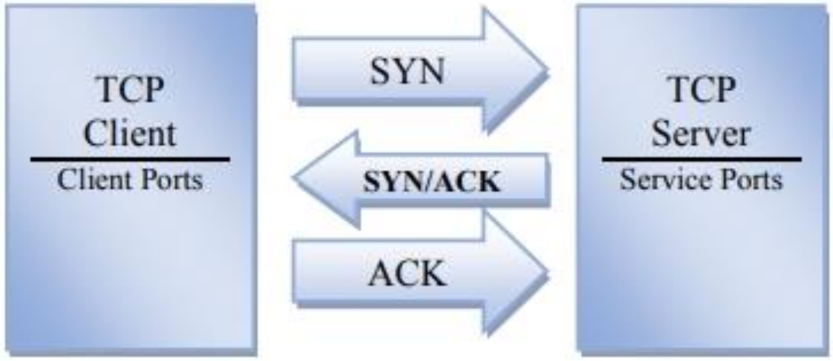

Fig. 2 Hand Shake Protocol[3].

C. Intrusion Detection System:

An intrusion detection system is a system whose primary purpose is to monitor the network for any anomalies and raise the alarm when found. In addition, an IDS is application software that constantly checks for security flaws during a network scan. The network administration or central collection system handles these vulnerabilities and is notified of intrusions. They can be divided into the following two main categories: IDSs.

1) Host-based intrusion detection system: A system installed on the host computer. It tracks network activity on its application files. As well as the operating system. In addition, the application and system files maintain a record called an audit trail. It can detect unusual traffic before data is sent and received, but for it to work correctly, the host system must also implement it. Because host-based IDS can monitor low-level system activity, impractical for NIDS, they can improve attack detection.

2) Network-based Intrusion Detection System (IDS): This system looks at all network activity, both incoming and outgoing, on network links in specific designated areas, which reduces the cost and time of installing software on different computers. The main drawback of this system is that it is vulnerable to attacks originating from within the network [4].

D. Machine Learning and Deep Learning for Detecting DDoS Attacks:

Machine learning (ML) is used to classify anomalies by allowing computers to learn without requiring explicit programming. Using a large data set, the system is trained to detect intruders and keep the network vigilant against potential threats. This algorithm is continuously trained to enhance its classification skill with the new data it processes. While machine learning shows the potential to address many cybersecurity challenges, it also suffers from multiple limitations. One problem is that a malicious data packet can be mistakenly identified as harmless. Another problem is that the large volume of packets on the network requires large-scale processing, hindering packet analysis and affecting the computer system's performance. AI experts are investigating applying deep learning techniques, a branch of machine learning, to improve intrusion detection methods. The efficacy of deep learning algorithms correlates directly with the amount of data being analyzed, while machine learning techniques typically plateau with time.

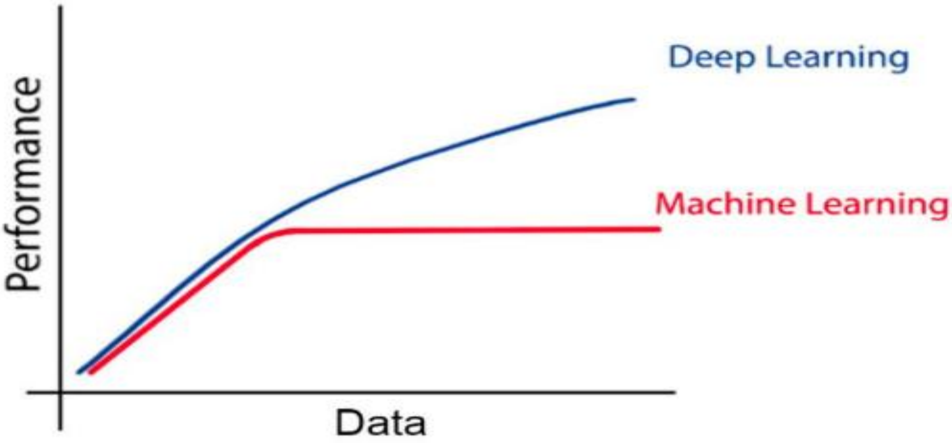

Fig .3 Machine learning vs deep learning

Regarding intrusion detection, DL is superior to ML in many ways.

Size of data: While ML algorithms function best with small datasets, DL algorithms perform significantly better with millions of data points.

Time: While training DL algorithms takes longer, the extra time is made up for in the real-time production and operation phase.

Dedication: Security professionals interpret the outcomes based on their methodology, while DL algorithms independently choose the features (inputs). On the other hand, ML algorithms need to define their features and labels (outputs) [5].

E .Related Work:

There are a plethora of current methods that rely on various concepts and algorithms. Even though some of them do pretty well in accuracy, they lengthen the processing time to find the DDoS assault. Let us have a look at the works that already exist.
The study conducted by V. Deepa et al. I focused on ensemble learning methods to identify Distributed Denial of Service (DDoS) attacks in Software Defined Networking (SDN) settings. This method uses machine learning algorithms such as K-Nearest Neighbor, Naive Bayes, Support Vector Machine, Self-Organizing Maps, and the CAIDA 2016 dataset. The clustering method outperforms individual learning algorithms, with the SVM-SOM model achieving the highest accuracy of 98.12% and a detection rate of 97.14%. The paper highlights the importance of ensemble methods in identifying DDoS attacks in SDN controllers [6].

Tanaphon Roempluk et al. conducted a study using two datasets, KDD CUP 1999 and NSL-KDD, to detect DDoS attacks. This experiment tested only three algorithms: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Multilayer Perceptron (MLP). The K-Nearest Neighbors (KNN) algorithm outperformed the other algorithms mentioned above in detecting attacks and achieved an accuracy of 99.99%. This study had the advantages of improving training and increasing accuracy using the KNN algorithm. The study's narrow focus on limited data sets and examination of only three algorithms [7] .

Research by Swathi Sambangi et al. used a dataset, the CICIDS 2017, and multiple regression analysis to identify DDoS attacks in cloud computing. Their experiment achieved an accuracy rate of 73.79 using the 16 selected features. This demonstrates the potential of multiple linear regressions in identifying attacks. One of the advantages of this research is the use of machine learning algorithms in a safe way to identify attacks in the cloud environment. At the same time, its disadvantages include the need to improve the accuracy rate of 73.79[8].

Amer A. Sallam et al. They conducted a study to evaluate machine learning algorithms Random Forest, Decision Tree, K-Nearest Neighbors, AdaBoost, and Native Bayes for detecting DDoS attacks using the CICIDS 2017 dataset provided by the Canadian Cybersecurity Institute, and the accuracy rate was respectively 95.31%, 95.68%, 92.61%, 95.69%, and 83.11%. This study emphasized using machine learning techniques to identify attacks and increase system security through accurate classification [9].

Suman Nandi et al. She investigated five feature selection techniques (Information et al.) to identify and classify the most essential features from the NSL KDD dataset. Intrusion detection datasets include anomalous and normal packets, although not all anomalies are associated with DDoS attacks. We separated DDoS class packets from the Dataset based on selected features. The new KDD DDoS dataset was prepared and analyzed using machine learning methods such as Naive Bayes, Bayes Net, Decision Table, J48, and Random. The J48 algorithm obtained the best accuracy of 99.72% of the mentioned algorithms[10].

Chukwuemeka Christian UGwu et al. developed a deep learning-based DDoS attack detection system using a long short-term memory (LSTM) model with singular value decomposition. The model outperformed traditional methods, achieving 94.28% accuracy on UNSW-NB15 and 90.59% on NSL-KDD[11].

Sagar Panda et al. They used a random forest approach and the NSL-KDD dataset to focus their machine-learning techniques on detecting DDoS attacks. By separating attack cases from standard cases, their technique achieved an accuracy rate of 99.76%. The main benefit of their research is that DDoS attacks can be identified with high accuracy, demonstrating the effectiveness of machine learning in network security applications. Disadvantages: Reliance on specific datasets is not applicable in different network contexts, and processing and analysis costs are high[12].

Anupama Mishra's paper mainly focused on supervised learning to identify and classify DDoS attacks in cloud computing using Naive Bayes, Random Forest, and K-Nearest Neighbor algorithms. Considering the KDDCUP99 dataset, among the above algorithms, Random Forest achieved the highest accuracy rate of 99.76%. This study highlighted the usefulness of machine learning methods in improving cloud security, achieving high detection accuracy with low false positives. Disadvantages: Focusing on cloud computing may limit its application in other fields. Excessive precision may lead to concerns about overfitting to the given data set [13].

A hybrid DDoS detection system is described in the work by A. Rama Krishna Kowsik et al., which uses self-organizing maps (SOM) and artificial neural networks (ANN) in cloud computing. When applied to the CICIDS 2017 dataset, their method achieves an impressive accuracy of 99.66% and a minimum false positive rate of 0.004. The advantage of this study is that it reduces the false positive rate to improve the reliability of DDoS attack detection. Disadvantage: The complex nature of the hybrid model may make it challenging to implement[14].

Abdullah Emir Cil et al . Introduced a deep neural network (DNN) model in their paper to classify network traffic and detect DDoS attacks. They achieved a detection success rate of 99.99% and an accuracy of 94.57% in classifying attack types using the CICDDoS2019 dataset. Advantages of this study: reaching the highest level of accuracy in identification and classification. Illustrating the effectiveness of deep learning in cybersecurity. Disadvantages: Model complexity may require significant processing resources[15].

In the study conducted by Ismail et al. Machine learning methods, specifically Random Forest and XGBoost classifiers, were used to classify and predict various DDoS attacks using the UNWS-np-15 dataset. XGBoost achieved an accuracy of 90%. The advantages of the study stem from the effective use of supervised learning models for classification. The disadvantage of this technique is its limitations as it relies on pre-defined data sets and may ignore some attack methods. Complex models may raise concerns about computer efficiency and scalability in practical applications [16].

Vimal Gaur et al. presented a hybrid approach to identify distributed denial of service (DDoS) attacks on Internet of Things (IoT) devices. Using ANOVA for XGBoost, the methodology provides superior performance, with 82.5% feature reduction and 98.34% accuracy. The advantages of this study on the CICDDos2019 dataset are the high detection accuracy, which indicates the effectiveness of the hybrid approach, and the simplification of the model through feature reduction. The disadvantage was that the model was updated constantly to adapt to new attack patterns[17].

The study presented by Chandan et al. highlights the use of machine learning methods, including random forest, nearest neighbors, and support vector machine, along with fuzzy inference rules to classify malicious or benign nodes with around 94% accuracy using a recently generated AVV DDos dataset which includes HTTP floods attacks and UDP Flood. The benefits of this study are the excellent classification accuracy due to the use of a sophisticated attack detection approach that combines fuzzy logic and machine learning[18].

Deepak Kumar et al. They investigated DDoS attack detection using a deep learning long short-term memory (LSTM) model with the CICDDoS2019 dataset. The study shows that LSTM can detect DDoS threats with up to 98% accuracy. One of the advantages of this study is its high accuracy in detecting DDoS attacks, as well as the efficient LSTM processing of sequential data. Disadvantage: LSTM and similar deep learning models may require significant computational resources, and the training data set's quality and comprehensiveness substantially impact model performance [19].

Soumyajit Das et al. They proposed a method based on the random forest classifier model. After training and learning, the model predicts whether the network traffic flow is typical. Based on the results, it can be inferred that the Random Forest Classifier model outperforms an accuracy rate of 99.9997% and a minimal false alert rate. Using the CICDDoS2019 dataset containing DDoS attacks, the advantages of this work are the minimal false alarms in distinguishing between normal and malicious traffic. Disadvantages of this study: The study was conducted on a clean and limited data set, which may need to accurately reflect the complexity of real-world network traffic[20].

Neeta Chavan et al., using the NSL-KDD dataset, have developed a system based on machine learning methodology to prevent bots and detect DDoS attacks. With accuracy rates of 82.28%, 89.15%, 90.4%, and 90.36%, respectively. Four algorithms were implemented: Decision tree classifier, K-nearest neighbour, logistic regression, and support vector machine. One of the advantages of this study is the high accuracy in detecting DDoS attacks Attacks. The disadvantages are that the datasets used (NSL-KDD) for DDoS and phishing detection Site URL dataset to prevent bots) does not cover all attack vectors [21].

Using the CICDDoS2019 dataset, Fathima Nazarudeen et al. They created a model that uses XGBoost, Random Forest, and Decision Tree methods to detect DDoS attacks. After selecting the features, XGBoost obtained a verification accuracy of 98.72%. Benefits from this study: Improved model performance due to a comprehensive feature selection process. Cons: The complex nature of the model and strict feature selection procedures may make implementation more difficult [22].

In the study conducted by M. Kavitha et al. He used the KDDCUP 99 dataset and focused on applying ML methods to identify DDoS attacks in Software Defined Networks (SDN). including Decision Tree, K-Nearest Neighbor Classifier, and Logistic Regression. The decision tree was superior to other algorithms, achieving an accuracy of 99.90%. The benefits of this study were improving detection capabilities based on machine learning and using good data for testing and training. A drawback of this study is that the possibility of overfitting the model may hinder its applicability to new or different types of attacks[23].

In the study conducted by Ghazia Qaiser et al., vulnerabilities that harm industrial Internet services were investigated. This study evaluated the ability of six machine learning algorithms to detect attacks using the CIC-IDS2017 dataset. These algorithms are Naive Bayes, SMO, J48, DecisionTable, AdaBoost, and SimpleLogistic. The J48 and DecisionTable algorithms achieved the best accuracy of 98.9%. This study aims to enhance the security of industrial services by demonstrating the effectiveness of machine learning in detecting DDoS attacks[24].

Using machine learning classification algorithms, the study by Srinivas Mikala et al. focuses on identifying NetBIOS DDoS attacks. The study uses correlation methods to analyze uncorrelated features using the NetBIOS_DrDoS dataset from CICCDDoS2019. It tests a range of algorithms, including Logistic Regression, Decision Tree, Random Forest, Ada Boost, Gradient Boost, K-Nearest Neighbour, Naïve Bayes, and Multilayer Perceptron. With an accuracy of 99.26%, multilayer perceptron with uncorrelated Pearson features was the most accurate. The benefit is precise detection, and a good selection of features can reduce the computational burden. Disadvantages: It cannot be applied to other types of attacks [25].

R. Sahila Devi et al. Use the CICDoS2019 dataset to analyze machine learning techniques such as K-Nearest Neighbors, Random Forest (RF), Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), and Naïve Bayes to detect DDoS attacks. They applied these algorithms and modified their parameters to develop the hybrid algorithm. Advantages of this study: The hybrid algorithms demonstrated 100% accuracy and an F1 score of 1 in DDoS detection, displaying exceptional accuracy. One of the drawbacks of this study is that it addresses the complexity of the hybrid algorithm in terms of implementation and modification[26].

The CICDDoS2019 and CICIDS2017 datasets were used for training and testing in Mahrukh Ramzan et al.'s study to identify DDoS attacks using deep learning models, such as recurrent neural network (RNN), long short-term memory (LSTM) and gradient recursion module. (GRU) with an accuracy rate of 99.99%. GRU showed a shorter execution time than LSTM and RNN, indicating its ability to detect real-time intrusion. Advantages of this study: Both datasets have excellent accuracy. Plus, faster boot times, especially for GRU. Disadvantages: Since the scope of the study is limited to the datasets used, generalizability may be affected [27].

Vanlalruata Hnamte et al. I recommend using a deep neural network (DNN) to identify DDoS attacks through training and testing with the CICIDS2017 and CICDDos2019 datasets. This approach achieved 99.9% accuracy using 81 features in the CICDDs2017 dataset and 67 features in the CICDDos2019 dataset [28].

S. Santhosh et al. looked at how well Random Forest, XGBoost, and a modified version of XGBoost worked with the CICDoS2019 dataset. The modified version of XGBoost achieved an amazing 97% accuracy, which was better than the original algorithm's 88% accuracy. Benefits of this work: The adjusted XGBoost model yields exceptional precision. Drawbacks: The algorithm presents challenges in terms of tuning overfitting due to high accuracy could limit the model's applicability to other datasets [29].

Hafiz Amaad et al. used the CIC-DDoS2019 dataset in their study to apply ensemble machine learning approaches, namely adaptive boosting classifiers, random forest, and histogram-based gradient boosting, to detect DDoS attacks. With a 99.9887% detection accuracy, this method outperformed earlier studies in the field. Using a large dataset guarantees the model's robustness. Still, the complexity of ensemble models may require a significant computational effort, and their high precision raises the possibility of overfitting in dynamic assault scenarios[30].

The study by Mamoon Saeed et al. She evaluated the effectiveness of machine learning algorithms in classifying DDoS attacks. Algorithms include Random Forest, Support Vector Machine, Native Bayes, XGBoost, and Decision Tree. The Random Forest algorithm obtained a maximum accuracy of 99.954% when used on the CICDDOS 2019 dataset [31].

The study by Tamanna Fardusy et al. The semi-supervised method detects DDoS attacks using Support Vector Machine (SVM) and AutoEncoder (AE) architectures. On the CICDDoS2019 dataset, the proposed AE+SVM model achieves the highest accuracy. The advantages of this study are the following: The proposed semi-supervised model outperforms many supervised and semi-supervised models with high accuracy (99.57%), precision, recall, and F1 score of more than 99%. The disadvantage is that the focus is on binary classification without distinguishing between several DDoS assaults [32].

In this paper, Zhenpeng Liu et al. presented a machine learning and feature engineering-based approach to detecting DDoS attacks in SDN using the CSE-CIC-IDS2018 dataset, The data set was trained and tested using decision trees, random forests (RF), support vector machines (SVMs), k-nearest (k-NN) classifiers, and XGBoost classifiers. The results showed that the highest accuracy for detecting DDoS was achieved by XGBoost, which amounted to 0.969. A drawback of this study is the need for extensive computational resources for processing and analysis due to the computational complexity of the optimization algorithm [33].

Table 1: Summary of relevant work carried out in the field of DDoS attack detection.

Ref	Dataset	Algorithm	Accuracy
[6]	CAIDA2016	K-Nearest Neighbor (KNN),Naive Bayes(NB), Support Vector Machine(SVM), and Self Organizing Maps(SOM)	SVM-SOM gave better results Accuracy = 98.12%.
[7]	KDD CUP1999 And NSL-KDD	Multilayer Perceptron (MLP), Support Vector Machine (SVM), and K-Nearest Neighbor (KNN)	K-Nearest Neighbor (KNN) gave better results: Accuracy = 99.99%.
[8]	CICIDS 2017	multiple regression analysis	Accuracy = 73.79%.
[9]	CICIDS 2017	Random Forest(RF), AdaBoost, Decision Tree(DT), K-Nearest Neighbors(KNN), and Naive Bayes(NB)	AdaBoost gave better results: Accuracy = 95.69%.
[10]	NSL KDD And UNSW-NB15	Naive Bayes(NB), Bayes Net(BN), Decision Table, J48, and Random Forest	J48 gave better results: Accuracy = 99.72%.
[11]	NSLKDD and KD DDOS	singular value decomposition (SVD) with long short-term memory (LSTM) mode	Accuracy = 94.28%.
[12]	NSL-KDD	random forest(RF)	Random forest gave better results Accuracy =99.76%.
[13]	KDDCUP99	Random Forest(RF), K-Nearest	Random Forest gave better

		Neighbor(KNN) and Naive Bayes(NB)	results: Accuracy =99.76%.
[14]	CICIDS 2017	self-organizing maps (SOM) and artificial neural networks (ANN)	Accuracy = 99.66%.
[15]	CICDDoS2019	Deep neural network	Deep neural network Gave results: Accuracy = 94.57%.
[16]	UNWS-np-15	Random Forest and XGBoost classifiers	XGBoost gave better results: Accuracy = 90%.
[17]	CICDDos2019	XGBoost	Accuracy = 98.34%.
[18]	AVV DDos	random forest, support vector machine, and K-nearest neighbors, with fuzzy inference rules	Accuracy = 94%.
[19]	CICDDoS2019	Long short-term memory (LSTM)	Long short-term memory (LSTM)gave results: Accuracy = 98%.
[20]	CICDDoS2019	Random Forest Classifier	Random Forest Classifier gave better results: Accuracy = 99.9997%.

[21]	NSL-KDD And URL	decision tree classifier, K-nearest neighbor, logistic regression, and support vector machine	Support vector machine gave better results: Accuracy = 90.36%False
[22]	CICDDoS2019	XGBoost, Random Forest, and Decision Tree	XGBoost gave better results: Accuracy =98.72%.
[23]	KDDCUP 99	Decision Tree, K-Nearest Neighbor Classifier, and Logistic Regression	Decision Tree gave better results: Accuracy =99.90%.
[24]	CIC-IDS2017	Naive Bayes, SMO, J48, DecisionTable, AdaBoost, and Simple Logistic.	J48 and Decision Table gave better results: Accuracy =98.9%.
[25]	CICDDoS2019	Logistic Regression, Decision Tree, Random Forest, Ada Boost, Gradient Boost, K-Nearest Neighbors, Naïve Bayes, and Multilayer Perceptron	Multilayer Perceptron gave better results: Accuracy =98.9%.
[26]	CICDDoS2019	Support Vector Machine (SVM), Naïve Bayes, Stochastic Gradient Descent (SGD), K-Nearest Neighbors, and Random Forest (RF), Hybrid algorithm	Hybrid algorithm gave better results: Accuracy = 100% F1-Score=1.0.
[27]	CICDDoS2019 And	Recurrent neural network (RNN), long short-term	

	CICIDS2017	memory (LSTM), and gradient recursion module. (GRU)	Accuracy = 99.99%.
[28]	CICDDos2019 And CICDDs2017	Deep neural network (DNN)	DNN gave better results: Accuracy = 99.9%.
[29]	CICDDoS2019	Random Forest, XGBoost, and a modified version of XGBoost	A modified version of XGBoost gave better results: Accuracy = 97%.
[30]	CICDDoS2019	Boosting classifiers, Random forest, Histogram-based gradient boosting, and Ensemble Model	Ensemble Model gave better results: Accuracy = 99.9887%.
[31]	CICDDOS201	Random Forest, Support Vector Machine, Native Bayes, XGBoost, and Decision Tree	Random Forest (RF) gave better results: Accuracy = 99.954%.
[32]	CICDDoS2019	AutoEncoder (AE) and Support Vector Machine (SVM)	AE+SVM gave results: Accuracy = 99.57%.
[33]	CSE-CIC-DS2018	XGBoost,,decision trees,K-nearest classifier,random forest ,support vector machines (SVMS)	XGBoost gave better results: Accuracy = 96.9

F. Dataset

1-KDDCUP

Since 1999, KDD'99 has been the most widely utilized Dataset for evaluating anomaly detection systems. The Dataset was compiled by Stolfo et al. using data collected during the DARPA'98 IDS evaluation program.

DARPA'98 comprises around four terabytes of compressed raw binary data. Seven weeks of network traffic captured in tcpdump format can be converted into around 5 million connection records, each containing about 100 data. The test data has nearly 2 million connection records during the two weeks. The KDD training dataset comprises over 4,900,000 individual connection vectors, each containing 41 features and classified as either standard or an attack, with precisely one specified attack type. The simulated attacks can be categorized into four groups:

1) A Denial of Service Attack (DoS) is an attack where the attacker overwhelms a computer or memory resource to the point where it cannot handle valid requests or denies access to legitimate users.

2) Remote to Local Attack (R2L): This type of attack happens when an attacker, An individual who can transmit data packets to a computer on a network, even without having an account on that system, exploits a vulnerability to get local access as a user of that computer.

3) User-to-Root attack (U2R) is an exploit in which the attacker gains access to a standard user account, typically acquired by password sniffing, dictionary attacks, or social engineering methods. Subsequently, the assailant capitalizes on weakness to obtain privileged root access to the system.

4) A probing attack intentionally collects information about a computer network to bypass its security measures [34].

2-NSL-KDD

The NSL-KDD dataset includes four types of attacks: Denial of Service (DoS), Probe, Remote to Local (R2L), and User to Root (U2R). The Dataset consists of 41 features categorized into three groups: fundamental features, traffic features, and content features. The Dataset has 148,517 records across the training and testing sets. This Dataset addresses issues found in previous datasets, including KDD Cup'99 and DARPA'98, such as duplicate records.

3-UNSW-NB15

The UNSW-NB15 dataset includes nine categories of contemporary attacks and novel standard traffic patterns. The 49 features are categorized into five groups: Flow features, Basic features, Content features, Time features, and Additional produced features. The collection has 257,705 entries labeled either by an assault type or a standard label. Sixteen thousand three hundred fifty-three items, accounting for 6.34% of the Dataset, are related to the DDoS attack[35].

4-CICIDS2017

The CICIDS2017 dataset, which simulates actual real-world data (PCAPs), includes common early and recent attacks. Besides flows classified based on timestamps, source and destination IP addresses, source and destination ports, protocols, and attacks, it also contains results of network traffic analysis using CICFlowMeter (CSV files). A definition of the extracted features is also given. Data was collected for 5 days, beginning on Monday, July 3, 2017, at 9 a.m. and ending on Friday, July 7, 2017, at 5 p.m. Monday is a typical day with low traffic. A few types of attacks are Botnet, web attacks, SSH, DDoS, DoS, Heartbleed, hacking, and FTP brute force. They were executed on Tuesdays, Wednesdays, and in the afternoon. Thursday and Friday [36].

5-CSE-CIC-IDS2018

The Canadian Establishment for Cybersecurity (CIC) released the realistic cyber Dataset CSE-CIC-IDS2018 in 2018. Globally, datasets from CIC and ISCX have been used to predict malware and detect intrusion. This Dataset includes the seven unique attack scenarios: web attacks, inside network infiltration, brute force, Botnet, heartbleed, distributed denial of service, and denial of service. This specific Dataset is generated as a CSV document, with 80 attributes categorized as Protocol and six important features labeled SourceIP, FlowID, DestinationIP, SourcePort, and Destination[37].

6-CICDDoS2019

In 2019, Sharafaldin et al. from the University of New Brunswick's Canadian Institute for Cybersecurity offered the CICDDoS2019 dataset, which is accessible to the public. Its numeric feature set is identical to that of CICIDS2017; however, preprocessing considerations are necessary because some of its records include the infinite value. More attack methods, such as CharGen, NetBIOS, LDAP, and SSDP reflection assaults, as well as more conventional attacks like UDP and SYN flood attacks, are included in CICDDoS2019 compared to CICIDS2017. The collection contains roughly 46 million entries and is accessible in CSV files with records of several assault types and some with benign labels [38].

G. Discussion

These studies reveal a broad and evolving landscape in the field of DDoS attack detection, taking advantage of machine learning and deep learning techniques. Here are several key insights from the reviewed studies: Ensemble and hybrid approaches. Several studies have highlighted the effectiveness of ensemble and hybrid approaches in achieving high accuracy rates in attack detection, e.g., SVM-SOM, SVM, and ANN. In addition to feature selection and optimization, studies have shown that using techniques such as ANOVA Gain and Chi-Square for feature reduction and binary optimization for gray wolves has an essential role in improving model performance. Selecting efficient features enhances accuracy and reduces computational complexity, making models more efficient and faster to train. As well as the advantages and challenges of deep learning, LSTM and DNN models have shown promising results in detecting DDoS attacks with high accuracy rates. The discussion suggests several future directions for research in DDoS attack detection. Diversifying the Dataset The development and use of more diverse datasets that reflect a broader range of attack scenarios and network conditions can help improve model generalizability. Simplifying models to reduce computational requirements without significantly compromising accuracy can make ML/DL techniques accessible to more users.

H. Conclusion and Future Work:

A comprehensive analysis of existing research emphasizes the use of deep learning and machine learning techniques to detect DDoS attacks. This study consistently shows high accuracy rates, with K-nearest neighbors achieving an accuracy rate of 99.99%, random forest achieving 99.9997%, ensemble model achieving 99.9887%, decision tree achieving 99.90%, and the hybrid method achieving 100% accuracy. Moreover, the combination of recurrent neural networks, long short-term memory, and gradient recursion modules resulted in a remarkable accuracy rate of 99.99%. These achievements demonstrate the ability of both deep learning models and machine learning techniques to distinguish between benign and malicious network traffic. However, there remain numerous challenges that require resolution. There are questions regarding the generalizability and use of these strategies in other network contexts due to their reliance on specific datasets for model testing and training. Due to their complexity, some models particularly those involving deep learning requires significant computing power and efficiency in model tuning and execution. In future work, we propose to hybridise two datasets, CICDDoS2019 and NSL-KDD, and use deep learning algorithms to detect DDoS attacks.

References

[1] Raj, R. & Singh Kang, S. Mitigating DDoS Attack using Machine Learning Approach in SDN. Proc. - 2022 4th Int. Conf. Adv. Comput. Commun. Control Networking, ICAC3N 2022 462–467 (2022) doi:10.1109/ICAC3N56670.2022.10074307.

.

[2] Najafimehr, M., Zarifzadeh, S. & Mostafavi, S. A hybrid machine learning approach for detecting unprecedented DDoS attacks. J. Supercomput. 78, 8106–8136 (2022)..

[3] Ashi, Z. Fast and Reliable DDoS Detection using Dimensionality Reduction and Machine Learning. (1959) doi:10.23919/ICITST51030.2020.9351347.

[4] Jyoti, N. & Behal, S. A meta-evaluation of machine learning techniques for detecting DDoS attacks. Proc. 2021 8th Int. Conf. Comput. Sustain. Glob. Dev. INDIACom 2021 522–526 (2021) doi:10.1109/INDIACom51348.2021.00093.

[5] Garcia, J. F. C. & Blandon, G. E. T. A Deep Learning-Based Intrusion Detection and Prevention System for Detecting and Preventing Denial-of-Service Attacks. IEEE Access 10, 83043–83060 (2022).

[6] Deepa, V., Sudar, K. M. & Deepalakshmi, P. Design of Ensemble Learning Methods for DDoS Detection in SDN Environment. Proc. - Int. Conf. Vis. Towar. Emerg. Trends Commun. Networking, ViTECoN 2019 1–6 (2019) doi:10.1109/ViTECoN.2019.8899682.

[7] Roempluk, T. & Surinta, O. A machine learning approach for detecting distributed denial of service attacks. ECTI DAMT-NCON 2019 - 4th Int. Conf. Digit. Arts, Media Technol. 2nd ECTI North. Sect. Conf. Electr. Electron. Comput. Telecommun. Eng. 146–149 (2019) doi:10.1109/ECTI-NCON.2019.8692243.

[8] Sambangi, S. & Gondi, L. A Machine Learning Approach for DDoS (Distributed Denial of Service) Attack Detection Using Multiple Linear Regression. 51 (2020) doi:10.3390/proceedings2020063051.

[9] Sallam, A. A., Kabir, M. N., Alginahi, Y. M., Jamal, A. & Esmeel, T. K. IDS for Improving DDoS Attack Recognition Based on Attack Profiles and Network Traffic Features. Proc. - 2020 16th IEEE Int. Colloq. Signal Process. Its Appl. CSPA 2020 255–260 (2020) doi:10.1109/CSPA48992.2020.9068679.

[10] Nandi, S., Phadikar, S. & Majumder, K. Detection of DDoS Attack and Classification Using a Hybrid Approach. ISEA-ISAP 2020 - Proc. 3rd ISEA Int. Conf. Secure. Priv. 2020 41–47 (2020) doi:10.1109/ISEA-ISAP49340.2020.234999.

[11] Ugwu, C. C., Obe, O. O., Popoola, O. S. & Adetunmbi, A. O. A distributed denial of service attack detection system using long-term memory with Singular Value Decomposition. Proc. 2020 IEEE 2nd Int. Conf. Cyberspace, CYBER Niger. 2020 112–118 (2021) doi:10.1109/CYBERNIGERIA51635.2021.9428870.

[12] Pande, S., Khamparia, A., Gupta, D. & Thanh, D. N. H. DDOS Detection Using Machine Learning Technique. tudies in Computational Intelligence vol. 921 (Springer Singapore, 2021).

[13] Mishra, A., Gupta, B. B., Perakovic, D., Penalvo, F. J. G. & Hsu, C. H. Classification Based Machine Learning for Detection of DDoS attack in Cloud Computing. Dig. Tech. Pap. - IEEE Int. Conf. Consum. Electron. 2021-Janua, 2–5 (2021).

[14] Kowsik, A. R. K., Pateriya, R. K. & Verma, P. A Deep Learning-based Hybrid Approach for DDoS Detection in Cloud Computing Environment. 2021 IEEE 4th Int. Conf. Comput. Power Commun. Technol. GUCON 2021 1– 6 (2021) doi:10.1109/GUCON50781.2021.9573817.

[15] Cil, A. E., Yildiz, K. & Buldu, A. Detection of DDoS attacks with feed-forward based deep neural network odel. Expert Syst. Appl. 169, 114520 (2021).

[16] Ismail et al. A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks. IEEE Access 10, 21443–21454 (2022).

[17] Gaur, V. & Kumar, R. Analysis of Machine Learning Classifiers for Early Detection of DDoS Attacks on IoT Devices. Arab. J. Sci. Eng. 47, 1353–1374 (2022).

[18] Chandan, Kumar, S. & Sinha, S. The study by Chandan et al. highlights machine learning methods, including random forest, support vector machine, and nearest neighbors, along with fuzzy inference rules to classify malicious or benign nodes with around 94% accuracy. Proc. 2022 Int. Conf. Intell. Innov. Eng. Technol. ICIIET 2022 294–300 (2022) doi:10.1109/ICIIET55458.2022.9967543.

[19] Kumar, D., Pateriya, R. K., Gupta, R. K., Dehalwar, V. & Sharma, A. DDoS Detection using Deep Learning. Procedia Comput. Sci. 218, 2420–2429 (2022).

[20] Das, S., Dayam, Z. & Chatterjee, P. S. Application of Random Forest Classifier for Prevention and Detection of Distributed Denial of Service Attacks. Proc. - 2022 OITS Int. Conf. Inf. Technol. OCIT 2022 380–384 (2022) doi:10.1109/OCIT56763.2022.00078.

[21] Chavan, N., Kukreja, M., Jagwani, G., Nishad, N. & Deb, N. DDoS Attack Detection and Botnet Prevention using Machine Learning. 8th Int. Conf. Adv. Comput. Commun. Syst. ICACCS 2022 1, 1159–1163 (2022).

[22] Nazarudeen, F. & Sundar, S. Efficient DDoS Attack Detection using Machine Learning Techniques. 2022 IEEE. Power Renew. Energy Conf. IPRECON 2022 1–6 (2022) doi:10.1109/IPRECON55716.2022.10059561

.

[23] Kavitha, M. et al. Machine Learning Techniques for Detecting DDoS Attacks in SDN. Int. Conf. Autom. Comput Renew. Syst. ICACRS 2022 - Proc. 634–638 (2022) doi:10.1109/ICACRS55517.2022.10029110.

[24] Qaiser, G., Chandrasekaran, S., Chai, R. & Zheng, J. In the study presented by Ghazia Qaiser et al., I worked to investigate vulnerabilities that negatively affect industrial Internet services. This study evaluated the ability of six machine learning algorithms to detect attacks using the CIC-IDS2017 dataset. T. 2023 15th Int. Conf. Comput. Autom. Eng. ICCAE 2023 546–550 (2023) doi:10.1109/ICCAE56788.2023.10111178.

[25] Mekala, S. & Dasari, K. B. NetBIOS DDoS Attacks Detection with Machine Learning Classification Algorithms. 2023 Int. Conf. Adv. Comput. Comput. Technol. InCACCT 2023 176–179 (2023) doi:10.1109/InCACCT57535.2023.10141815.

[26] Devi, R. S., Bharathi, R. & Kumar, P. K. Investigation on Efficient Machine Learning Algorithm for DDoS Attack Detection. ICCECE 2023 - Int. Conf. Comput. Electr. Commun. Eng.

[27] Ramzan, M. et al. Distributed Denial of Service Attack Detection in Network Traffic Using Deep Learning Algorithm. Sensors (Basel). 23, 1–24 (2023).

[28] Hnamte, V. & Hussain, J. DDoS Detection Using Hybrid Deep Neural Network Approaches. 2023 IEEE 8th Int.Conf. Converg. Technol. I2CT 2023 1–8 (2023) doi:10.1109/I2CT57861.2023.10126434.

[29] Santhosh, S., Sambath, M. & Thangakumar, J. Detection of DDOS Attack using Machine Learning Models. Proc. 1st IEEE Int. Conf. Netw. Commun. 2023, ICNWC 2023 1–6 (2023) doi:10.1109/ICNWC57852.2023.10127537.

[30] Amaad, H. & Mughal, H. Experimenting Ensemble Machine Learning for DDoS Classification: Detection of DDoS Using Large Scale Dataset. 2023 4th Int. Conf. Adv. Comput. Sci. ICACS 2023 - Proc. 1–7 (2023) doi:10.1109/ICACS55311.2023.10089656.

[31] Saeed, M. M. et al. Machine Learning Techniques for Detecting DDOS Attacks. 2023 3rd Int. Conf. Emerg. Smart Technol. Appl. eSmarTA 2023 1–6 (2023) doi:10.1109/eSmarTA59349.2023.10293366.

[32] Fardusy, T., Afrin, S., Sraboni, I. J. & Dey, U. K. An Autoencoder-Based Approach for DDoS Attack Detection Using Semi-Supervised Learning. 2023 Int. Conf. Next-Generation Comput. IoT Mach. Learn. NCIM 2023 1–7 (2023) doi:10.1109/NCIM59001.2023.10212626.

[33] Liu, Z. et al. A DDoS Detection Method Based on Feature Engineering and Machine Learning in Software- Defined Networks. Sensors 23, (2023).

[34] Tavallaee, M., Bagheri, E., Lu, W. & Ghorbani, A. A. A detailed analysis of the KDD CUP 99 data set in Computational Intelligence for Security and Defense Applications. Comput. Intell. Secure. Def. Appl. 1–6 (2009).

[35] dhammad, M., Afdel, K. & Belouch, M. Semi-supervised machine learning approach for DDoS detection. Appl Intell. 48, 3193–3208 (2018).

[36] https://www.unb.ca/cic/datasets/ids-2017.html.

[37] Kanimozhi, V. & Jacob, T. P. Artificial Intelligence outflanks all other machine learning classifiers in Network Intrusion Detection System on the realistic cyber Dataset CSE-CIC-IDS2018 using cloud computing. ICT Express 7, 366–370 (2021).

[38] https://www.unb.ca/cic/datasets/ddos-2019.html.