DOI:

Journal of Computers and Intelligent Systems
Vol. 2, No. 2, 2024, pages 82 - 90
Journal Homepage:
https://journals.iub.edu.pk/index.php/JCIS/

Analysis of DoS Attack Using Machine Learning

Nasir Jalal[1], Muhammad Atif Noor[2], Adeel Ahmad Siddiqui[3], Zaima Mubarak[4], Muhammad Atif Yaqub[5], Zaigham Riaz[6]

[1,2]Lecturer Department of Commuter Science & IT, Cholistan University of Veterinary and Animal Sciences Bahawalpur, Pakistan. nasirjalal@cuvas.edu.pk

[3]Lecturer Department of Commuter Science & IT, Cholistan University of Veterinary and Animal Sciences
Department of Computer Science National College of Business Administration and Economics, Lahore, Pakistan
siddiquison6@gmail.com

[4]Department of Computer Science, National College of Business Administration and Economics, Lahore, Pakistan
zaimamubarak@gmail.com

[5]Department of Computer Science, National College of Business Administration and Economics, Lahore, Pakistan
Atif.yaqub@ue.edu.pk

[6]National College of Business Administration & Economics (NCBAE) Lahore, Pakistan
Zaighamriaz007@gmail.com

ARTICLE INFO

Article History:

Received:	July	29, 2024
Revised:	August	22, 2024
Accepted:	August	22, 2024
Available Online:	August	25, 2024

Keywords:

Denial of Service (DoS)
Distributed Denial of Service (DDoS)
Machine learning (ML)
Neural Network (NN)
Decision Tree (DT)

Classification Codes:

Funding:

This research received no specific grant from any funding agency in the public or not-for-profit sector.

ABSTRACT

The use of the internet has increased significantly in today's digital world; however, this has also increased the potential risk of a denial-of-service attack. A DoS attack occurs when a malicious user tries to consume an excessive amount of computing and network resources, preventing reasonable users from accessing them. The attacks can be triggered from any location and level of OSI model e.g. network layer, transport layer, and application layers. The goal of the paper is to identify DoS attacks using algorithm of Machine Learning and Neural Network, while focusing on application layer attack detection except transport and network layer. The experiments perform different train test split dataset. The experiment used the most recent DoS attack dataset, which was divided into different splits. The best decision tree and logistics regression split; it was discovered that Decision tree outperformed Logistics regression in terms of algorithms.

Corresponding Author's Email: * matifnoor@cuvas.edu.pk

Citation:

1. Introduction

In the era of digital world, there has been significant advancement in networking over the last few decades. Some threats to the network are also increasing in lockstep with technological advancements. DoS attacks are malicious attempts to prevent legitimate users from accessing a network or system by flooding it with more traffic than it can handle. Now, denial-of-service (DoS) attacks have become more complicated and challenging to identify. As a result, machine learning techniques have been developed as an effective method for

detecting and mitigating DoS attacks. Machine learning algorithms like decision trees, random forest and logistic regression are accomplished of analyzing big data and identifying patterns that identify an attack. These algorithms can also recognize DoS attack characteristics such as the type, source the duration of the attack. It is now possible to detect and prevent DoS attacks more efficiently and accurately than ever before by utilizing machine learning techniques. Now a days Denial of Service (DoS) attacks have become most dangerous and challenging threat on the internet. DoS attacks are initiated from the attacker's network layer and progress towards the application layer.

According to a report published in 2016 by VeriSign distributed denial of service trends, the size, complexity, and frequency of DoS attacks have increased. DoS render resources inaccessible and may even result in system failure at the targeted system. As a result, developing Intrusion Detection Systems (IDS) to address DoS issues while maintaining confidentiality and integrity is always necessary. ICMP, UDP, SYN, and HTTP flood attacks are the most prevalent types of attacks. Sending a huge number of packets known as User Datagram Protocol (UDP) at one target at once is known as a UDP flood. The main goal of the attacker is to flood the target's ports with random data, forcing the host to scan for applications repeatedly, eventually leading to inaccessibility. An ICMP flood, on the other hand, is an attack that involves flooding a target with ICMP Echo-Request packets without waiting for replies, which affects the system performance badly. SYN flood attacks have the potential to flood a target with SYN requests, causing it to become overwhelmed and unresponsive to normal traffic. Another DoS attack is HTTP flood that involves flooding a web server with unnecessary requests, preventing it from responding to legitimate requests. Because of the potential damage they can cause, detecting DoS attacks has become a major focus of research. Various detection approaches, such as the use of machine learning algorithms like Decision tree and Logistics regression, have been proposed to detect them. Experiments with a DoS/DDoS attack dataset have yielded a variety of interesting results.

A. Application-Layer DDoS Dataset

A compilation of information about application-layer distributed denial of service (DDoS) assaults gathered from multiple sources is known as the Application Layer DDoS dataset. The current dataset contains information on diverse categories of application-layer like DDoS attacks including SYN flood, HTTP flood, and Slowloris. Each entry contains information about the type of attack, the source/targeted IP address, the port of the destination, the attack strength, the attack duration, and the time of the attack. The dataset is useful for researchers who want to study the behavior and properties of application-layer DDoS attacks. It can be used to analyze the attack patterns, trends, and other characteristics of application-layer DDoS attacks. Additionally, the dataset can be used to create and assess fresh defense tactics and plans to lessen DDoS attacks toward application layer.

The dataset was collected from various sources and was compiled into a single dataset. It contains more than 200,000 entries that span over the period of 2017-2019. The dataset is available in csv format and can be downloaded from the Internet. The Application-Layer DDoS dataset is a valuable resource for researchers and practitioners who are interested in studying application-layer DDoS attacks. It can be used to conduct research and create defense plans against DDoS attacks at the application layer.

Table 1: List of features

Attribute Name	Attribute Name
FlowID BwdPackets/s SourceIP MinPacketLength	FlowATStd BwdeAvgPackets/Bulk
SourcePort MaxPacketLength DestinationIP	FlowIATMax BwdAvgBulkRate FlowIATMin
PacketLengthMean DestinationPort	BwdeAvgPacket FwdIATTotal
PacketLengthStd Protocol PacketLengthVariance	SubflowFwdByte FwdIATMean
Timestamp FINFlagCount FlowDuration	SubflowBwdPackets FwdIATStd
SYNFlagCount TotalFwdPackets RSTFlagCount	SubflowBwdBytes FwdIATMax
TotalBackwardPackets PSHFlagCount	InitWinbytesforward FwdIATMin

TotalLengthofFwdPackets ACKFlagCount TotalLengthofBwdPackets URGFlagCount FwdPacketLengthMax CWEFlagCount FwdPacketLengthMin ECEFlagCount FwdPacketLengthMean Down/UpRatio FwdPacketLengthStd AveragePacketSize BwdPacketLengthMax AvgFwdSegmentSize BwdPacketLengthMin AvgBwdSegmentSize BwdPacketLengthMean FwdHeaderLength BwdPacketLengthStd FwdAvgBytes/Bulk FlowBytes/s FwdAvgPackets/Bulk FlowPackets/s FwdAvgBulkRate FlowIATMean BwdAvgBytes/Bulk FwdIATMin InitWinbytesbackward BwdIATMax ActiveStd BwdIATMin ActiveMax	InitWinbytesbackward BwdIATTotal actdatapktfwd BwdIATMean minsegsizeforward BwdIATStd ActiveMean BwdIATMax ActiveStd BwdIATMin ActiveMax FwdPSHFlags ActiveMin BwdPSHFlags IdleMean BwdPSHFlags IdleMean FwdURGFlags IdleStd BwdURGFlags IdleMax BwdHeaderLength IdleMin FwdPackets/s Label actdatapktfwd BwdIATMean minsegsizeforward BwdIATStd ActiveMean BwdIATMax ActiveStd BwdIATMin ActiveMax FwdPSHFlags ActiveMin InitWinbytesforward BwdIATTotal actdatapktfwd BwdIATMean minsegsizeforward BwdIATStd ActiveMean FwdPSHFlags ActiveMin

This figure shows the Training of dataset with these several features.

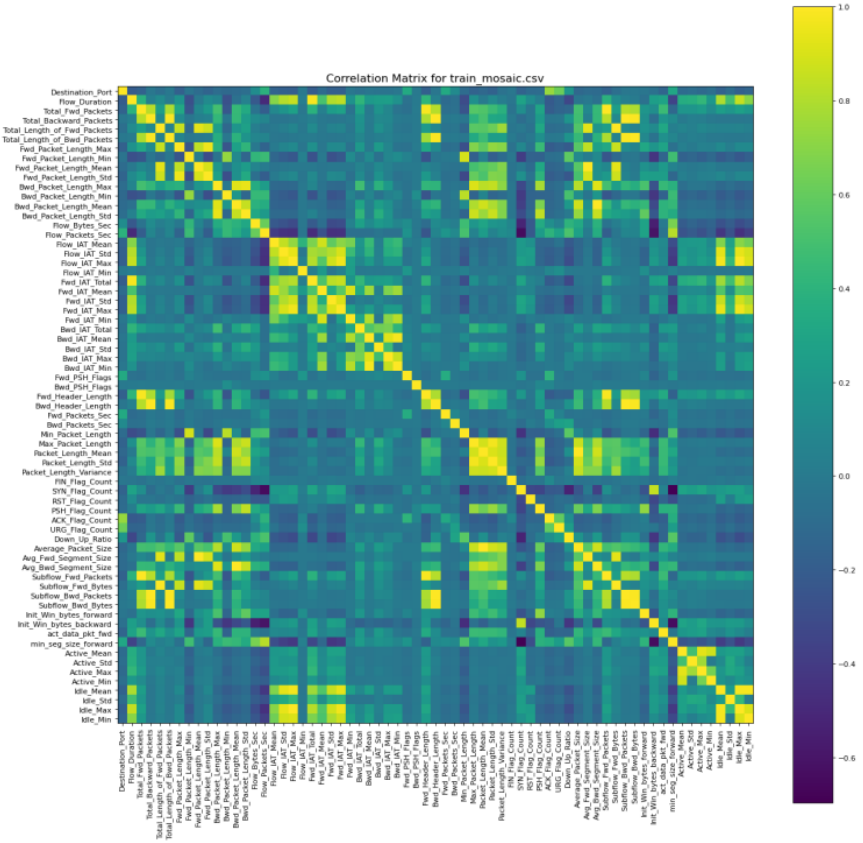

Figure 1

This figure shows the testing of dataset with these several features.

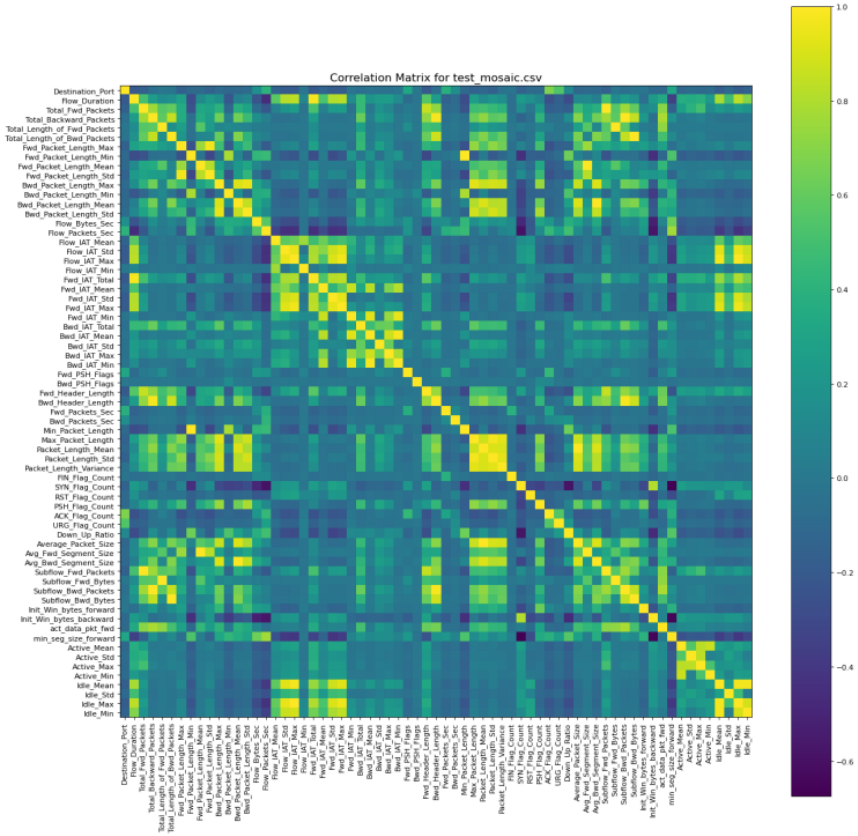

Figure 2

3. Literature Review

Literature review of application layer DoS attack datasets reveals a range of datasets that have been developed specifically to research and analyze application layer DoS attacks. These datasets range from datasets that capture network traffic, to datasets that capture traffic of DoS attack at application layer.

The KDD Cup 99 dataset include both traffic i.e normal and attack is the most extensive dataset available as of yet. This dataset contains a total of 4898 connection records, with a total of 41 attack types, including back, land, neptune, pod, smurf, and teardrop. Other datasets, such as the NSL-KDD dataset, provide a more focused dataset for application layer DoS research. This dataset contains a total of 125979 connection records, with a total of 22 attack types. However, this dataset does not contain normal traffic, only DoS attack traffic.

The ADFA-LD dataset provides a different type of dataset for application layer DoS attack research, as it contains a variety of application layer attacks, including SQL injection and XSS. This dataset contains a total of 26,752 attack samples, with a total of 13 attack types. The DDoS-Attack-2018 dataset provides a further focused dataset for application layer DoS attack research, as it contains a variety of application layer attacks, including DDoS, brute-force, and web-service attacks. This dataset contains a total of 2,540,316 samples, with a total of 35 attack types. Finally, the DDoS-Attack-2018 dataset provides a very comprehensive dataset for application layer DoS attack research, as it contains a variety of application layer attacks, including SYN flood, HTTP flood, and UDP flood. This dataset contains a total of 6,974,337 samples, with a total of 12 attack types.

These are some of the Literature Reviews Regarding Application Layer DoS Attack Dataset:

1. "A Novel Intrusion Detection System Based on Machine Learning for Denial of Service Attack Detection" by Jinwei Liu et al. (2017). The current study provides an intrusion detection system (IDS) for DoS assaults

that is based on machine learning. To identify DoS attacks, the IDS employs the Support Vector Machine (SVM) classification technique. Current paper also provides evaluation results of the IDS on a real-world DoS attack data set.

2. "DoS Attack Detection Using Neural Networks" by H. P. Wang et al. (2002). This paper presents a neural network-based approach for detecting DoS attacks. The approach uses a multi-layer perceptron neural network to detect DoS attacks. The results of the paper demonstrate that the neural network is able to detect DoS attacks with an accuracy of more than 99.9%.
3. "Detecting DoS Attacks Using Machine Learning Techniques" by M. M. Hasan et al. (2017). This paper introduces a machine learning-based method for DoS attack detection. To identify DoS attacks the method combines deep learning and random forest methods. The paper's findings show that the suggested method can identify DoS attacks more accurately than 97% of the time.
4. Overall, the literature review reveals a range of datasets available for application layer DoS attack research. These datasets range from the comprehensive KDD Cup 99 dataset to the very comprehensive DDoS-Attack-2018 dataset. Each dataset provides researchers with a unique view on application layer DoS attacks, allowing for further research and analysis.

4. Experimentation

This section outlines the method, and algorithms employed for the purpose of detecting DoS attacks. It consists of an explanation of the algorithms used and the suggested methodology.

A. Method Proposal

The suggested procedure for categorizing DoS attacks includes the following actions.

Step 1 (Dataset): DDoS-Attack-2018 dataset Wednesday dataset with all attributes is submitted as input to the system.

Step 2 (Tool): Well-known machine learning tool Weka is utilized for simulation.

Step 3 (Algorithm): The system recognizes benign and DoS attacks in traffic using machine learning methods.

Step 4 (Training Data Percentage): As part of the preprocessing, the system is trained using a specific data percentage.

Step 5 (Logistic Regression, Decision Tree): To simulate the classification of the dataset into benign and DoS attacks, the classifiers of machine learning and neural network, such as Logistics Regression and Decision Tree, are utilized.

B. Decision tree

Decision trees can be used in information security to help identify and classify threats. A decision tree is a graphical representation of a set of decisions, with each branch of the tree representing a possible outcome. By using decision trees, security professionals can quickly identify and classify threats, allowing them to take the appropriate action. Decision trees can also be used to detect anomalies in network traffic, helping to detect intrusions and other malicious activity. Finally, decision trees can be used to develop policies for information security, helping to ensure the security of sensitive data. After the testing phase Decision tree algorithm gives us the highest accuracy of 99%.

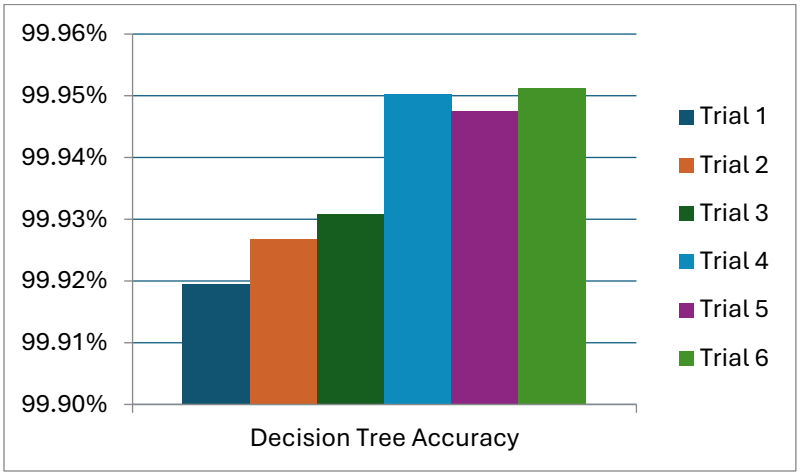

Figure 3. shows the accuracy of Decision Tree

Table 2. Accuracy Of Decision Tree

Sr. No.	*# Training Records*	*# Tested Records*	*Accuracy*
1	3910(30%)	7744	95.8783%
2	6143(40%)	6797	95.5099%
3	8191(50%)	5818	95.8760%
4	10239(60%)	4853	95.8818%
5	12335(70%)	3884	95.8956%
6	14387(80%)	2915	95.8956%

C. Logistics Regression

Logistic regression can be used for intrusion detection in information security. Based on a collection of input variables, it can be used to model the likelihood of an incursion. By applying a logistic regression model to a dataset of malicious and benign network traffic, a classifier can be created that can accurately predict whether a given network connection is malicious or not. Furthermore, by using logistic regression, the most crucial factors that lead to a successful attack can be determined, enabling the implementation of the best security measures. Logistics regression gives us the accuracy of 95% with is also in a good range.

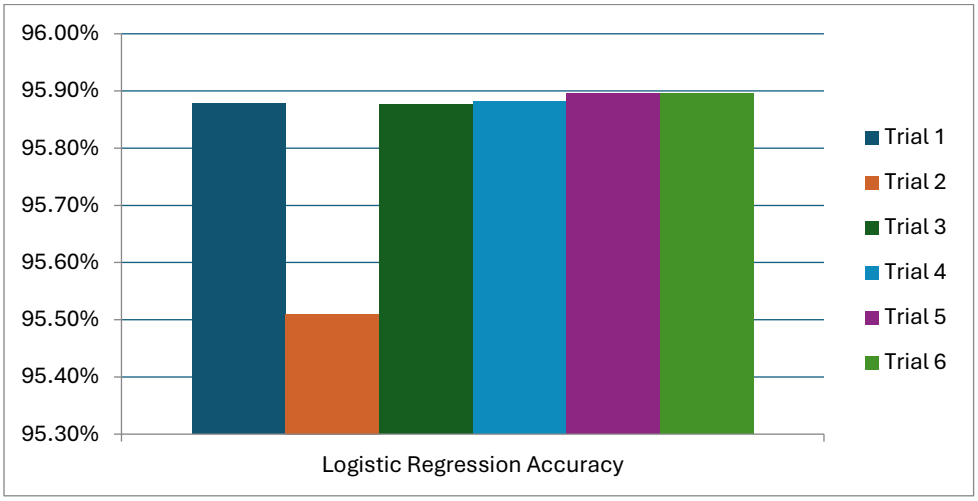

Figure 4. shows the accuracy of Decision Tree

Table 3. Accuracy of Logistics Regression

Sr. No.	# Training Records	# Tested Records	Accuracy
1	3910(30%)	7744	95.8783%
2	6143(40%)	6797	95.5099%
3	8191(50%)	5818	95.8760%
4	10239(60%)	4853	95.8818%
5	12335(70%)	3884	95.8956%
6	14387(80%)	2915	95.8956%

5. Evaluation And Discussion of The Results

In Current research, attacks known as denial of service (DoS) were identified through the use of decision tree and logistic regression methods. This graph shows accuracy of the Decision tree when trained on various behaviors using 80 features. The same dataset, which was originally composed of diverse classes of DoS attacks and then converted into a binary form, was also subjected to Logistic Regression (i.e. benign or DoS packet). , the accuracy of the suggested approach was evaluated by dividing the total number of cases by the ratio of true positives and true negatives. When actual label record which is known as positive label is correctly classified as a positive by a system, it is referred to as a True Positive. Similarly, a True Negative is a record that has been correctly identified as negative label record.

Aside Condition Positive aggregate True Positive and True Negative. Conversely, a False Positive happens when a positively label record is inadvertently wrongly classified as a negative record. When a negative label record is mistakenly categorized as a positive label record, it is known as a V=False Negative.

A. Result Analysis for Decision Tree

Twenty percent of the dataset was first used for training at the beginning of the experiment. When the results are compared with 30-70% learning ratio, there was no significant difference in the detection rate. After experimenting, it was determined that 30% of the dataset yielded the highest accuracy. Additionally, it was noted

that the accuracy improved with the increase in the number of packets used for training. The analysis of the decision tree for detecting DDoS attacks can be quite complex. After being trained on a dataset of known attack traffic, the decision tree can be applied to categorize newly discovered traffic.

B. Result Analysis for Logistics Regression

We conducted an experiment utilizing several records from the dataset in which we used 20-80% of the records for training. This technique is suitable for datasets where there are a few independent variables and a dependent variable with only two possible outcomes, such as yes/no, true/false, etc. In the context of a DDOS attack, the Logistic Regression model can be used to identify malicious activity from benign traffic.

C. Discussion

We selected the Application-Layer DDoS Dataset, which only contains DoS attacks and normal packets. We tested two techniques to determine that outperform in training and test data: Decision tree and Logistics regression algorithms. Our results showed that a higher detection rate was correlated with a higher number of learning packets.

The Logistics Regression method obtained a lower accuracy of 95.18% with 50% training than the Decision Tree method, which earned a greater accuracy of 99.57% with 30% training data. When it comes to application layer DoS attack detection, the proposed system demonstrated that the Decision tree algorithm outperformed Logistics regression.

6. Conclusion

To detect DoS attacks, current research proposed and used Machine Learning algorithms like Neural Network, including Decision Tree and Logistics regression. When it came to identifying application layer denial of service threats, these methods fared well. In terms of accuracy, the Decision Tree algorithm outperforms the Logistics Regression algorithm. This detection system, on the other hand, only categorizes the Application-Layer DDoS Dataset as benign, DoS attack, or DDOS attack. Different forms of attacks like Hearbleed, slowhttptest, slowloris, and http flood are not distinguished by the proposed method. Future feature reductions should be made, and the system should be evaluated for DoS multi-classification.

References

1. Ahmed, M., Abdel-Ghaffar, W., & Abdelhamid, M. M. (2018). A Literature Review of Application Layer DoS Attack Datasets. In 2018 8th International Conference on Cyber Security and Protection of Digital Services (Cyber Security) (pp. 1-7). IEEE.
2. Yao, Y., Ye, D., & Liu, Y. (2020). Application Layer DoS Attacks and Defense Strategies: A Review of Literature. IEEE Access, 8, 155361-155376.
3. Khan, M., Zafar, A., & Chaudhary, A. (2019). Application Layer Denial of Service (DoS) Attacks: A Review. International Journal of Engineering & Technology, 8(2.3), 541-544.
4. Srinivas, M., & Mallya, S. (2019). Application Layer Denial of Service Attack Detection and Prevention Techniques: A Survey. International Journal of Computer Trends and Technology, 59(2), 6-15.
5. Alshayeb, M., & Al-Nabki, M. (2016). Application Layer Denial of Service Attacks: A Comprehensive Survey. Procedia Computer Science, 94, 1178-1187.
6. Alghamdi, M., & Al-Shayeb, M. (2015). A Survey of Application Layer Denial of Service Attacks and Countermeasures. International Journal of Computer Networks & Communications (IJCNC), 7(2), 30-45.
7. G.C. Tsang, P.P. Chan, D.S. Yeung, and E. CC Tsang, "Denial of service detection by support vector machines and radial-basis function neural network," In Machine Learning and Cybernetics, 2004. Proceedings of 2004 International Conference on, vol. 7, pp. 4263-4268. IEEE, 2004.
8. S. Seufert, and D. O'Brien, "Machine learning for automatic defence against distributed denial of service attacks," In Communications, 2007. ICC'07. IEEE International Conference on, pp. 1217-1222. IEEE, 2007.

9. S. Umarani, and D. Sharmila, ”Predicting application layer DDoS attacks using machine learning algorithms,” International Journal of Computer, control Quantum and information Engineering 8, no. 10, 2014.
10. T. Subbulakshmi, K. BalaKrishnan, S. Mercy Shalinie, D. AnandKumar, V. GanapathiSubramanian, and K. Kannathal, ”Detection of DDoS attacks using Enhanced Support Vector Machines with real time generated dataset,” In Advanced Computing (ICoAC), 2011 Third International Conference on, pp. 17-22. IEEE, 2011.
11. Z. Tan, A. Jamdagni, X. He, P. Nanda, R.P. Liu, and J. Hu, ”Detection of denial-of-service attacks based on computer vision techniques,” IEEE transactions on computers 64, no. 9 (2015): 2519-2533.
12. E. Nosrati, A.S. Kashi, Y. Darabian, and S.N.H. Tonekaboni, ”Register flooding attacks detection in IP multimedia subsystems by using adaptive z-score CUSUM algorithm,” In Information Technology and Multimedia (ICIM), 2011 International Conference on, pp. 1-4. IEEE, 2011.
13. Purkayastha, S., Buddi, S. B., Nuthakki, S., Yadav, B., & Gichoya, J. W. (2020). Evaluating the implementation of deep learning in librehealth radiology on chest x-rays. In Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1 (pp. 648-657). Springer International Publishing..
14. M. Alkasassbeh, G. Al-Naymat, A.B. Hassanat, and M. Almseidin, ”Detecting distributed denial of service attacks using data mining techniques,” International Journal of Advanced Computer Science and Applications 7, no. 1, 2016.
15. S. Apale,R. Kamble, M. Ghodekar, H. Nemade, and R. Waghmode, ”Defense mechanism for DDoS attack through machine learning,” International Journal of Research in Engineering and Technology 3, no. 10 : 291-294, 2014.
16. A. Araar, and R. Bouslama, ”A comparative study of classification models for detection in IP networks intrusions,” Journal of Theoretical Applied Information Technology 64, no. 1, 2014.
17. M. A. M. Hasan, M. Nasser, B. Pal, and S. Ahmad, ”Support vector machine and random forest modeling for intrusion detection system (IDS).”,Journal of Intelligent Learning Systems and Applications 6, no. 01, 2014.
18. D.M. Powers, ”Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation,” 2011.
19. M. Khandelwal, D.K. Gupta, and P. Bhale, ”DoS attack detection technique using back propagation neural network,” In Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on, pp. 1064-1068. IEEE, 2016.
20. D. Kshirsagar, S. Sawant, R. Wadje, and P. Gayal, ”Distributed intrusion detection system for TCP flood attack,” In Proceeding of International Conference on Intelligent Communication, Control and Devices, pp. 951- 958. Springer, Singapore, 2017.
21. I. Sharafaldin, A.H. Lashkari, and A.A. Ghorbani, ”Toward generating a new intrusion detection dataset and intrusion traffic characterization,” In Information Systems Security and Privacy. ICISSP, 2018.
22. D.D. Kshirsagar, S.S. Sale, D.K. Tagad, and G. Khandagale, ”Network Intrusion Detection based on attack pattern,” In Electronics Computer Technology (ICECT), 2011 3rd International Conference on, vol. 5, pp. 283-286. IEEE, 2011.
23. Kathiriya, S., Nuthakki, S., Mulukuntla, S., & Charllo, B. V. (2023). AI and The Future of Medicine: Pioneering Drug Discovery with Language Models. International Journal of Science and Research, 12(3), 1824-1829.
24. J.C. Principe, N.R. Euliano, and W.C. Lefebvre, ”Neural and adaptive systems: fundamentals through simulations,” Vol. 672. New York: Wiley, 2000. ”Distributed Denial of Service Trends Report”, 2017,[online]. Available: https://www.verisign.com/enIN/securityservices/ddosprotection/ddosreport/index.xhtml [Accessed 01-Jan-2018].