ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (12), December 2023, pp. 25-37

Article Title: DDOS Attack Prediction using Deep Learning

DDOS Attack Prediction using Deep Learning

R. Siva Nagini[1], N. Janaki[2]

[1]R & D Engineer, AB Technologies, Nagercoil, Tamil Nadu, India.
sivanagini23.abt@gmail.com

[2]Department of Electrical and Electronics Engineering, Vels Institute of Science, Technology and Advanced Studies, Tamil Nadu, janaki.se@velsuniv.ac.in

ABSTRACT

DDoS attacks on distributed networks are another name for distributed network attacks. These attacks exploit specific constraints that are applicable to every arrangement asset, like the authorized organization's website framework. This proposes a deep learning approach for predicting DDoS attacks. Deep learning methods were developed for the purpose of classifying attacks. The deep learning method includes the classification methods Multilayer Perceptron and Long Short-Term Memory Grids. The datasets are pre-processed using StandardScaler. Distributed denial-of-service attack detection and categorization are aided by deep learning techniques. MLP is an artificial neural network feed-forward model that maps input sets to output sets. The LSTM classifier is designed to classify faults based on this proposed project produced a confusion matrix to identify the performance of the model, and it is used for fault diagnosis in that region. This implemented in Python software.

Keywords: Distributed Denial of Service (DDoS), Multilayer Perceptron (MLP), Long Short-Term Memory Networks (LSTM), Convolutional neural network (CNN), Recurrent neural networks (RNNs).

1 Introduction

A smart grid is an electrical supply classification that more effectively responds to energy needs and distribution by fusing contemporary information technology with an established power network. It provides a number of innovative capabilities, including as bi-directional communication, remote management of smart home equipment, consumer behaviour updates, and stability monitoring of the power grid. The most common type of cyberattack is a distributed denial of service attack. These attacks employ several crushed systems to disrupt or terminate the services of hosts related to the Internet. These attacks typically target bank or credit card payment network Web servers. As a result, a single attack could result in significant loss. Predicting and identifying DDoS assaults is a difficult endeavour's typical DDoS finding system's goal is to distinguish between legitimate and mean packet traffic [1]. This is a crucial and time-consuming step in the data analysis process. Here, the data will be filtered to remove unnecessary information and transformed into high-quality information. In this stage, substitute values in the data that are not relevant to our experimental study by cleaning the data using statistical approaches. For the first phase of the examination, this is a requirement for all data analyses [2]. These components consist of data availability, data integrity, and data

confidentiality. Researchers have tested and developed a number of preventative techniques to shield databases and network devices from online intrusions [3]. Direct and indirect cyberattacks were divided into two groups, and the direct cyberattacks were further divided into four subgroups. Data intrusion assaults are widely recognized as the most prevalent sort of cyber-attacks, with denial of service (DoS) attacks being the most significant type of attack among them [4]. In these attacks, the adversary introduces fake loads to the primary service source and disrupts the regular genuine service in an attempt to break the normal trend of services. The majority of DoS attacks that occur nowadays are distributed (DDoS), in which multiple adversaries launch strikes at once [5]. Use the LSTM technique to forecast DDoS attacks enter the three-dimensional matrix into this module's input layer. The output layer outputs the predictions following the unseen layer's operation [6]. To combine these heuristics, we employ a Multilayer Perceptron as a data fusion mechanism. To eliminate magic numbers and site-specific criteria that would need significant installation and customization costs, employ a machine learning technique. An MLP makes a good foundation for a detection mechanism since it can be taught by example and can combine multiple detection metrics [7]. DDoS is a serious network-oriented cyber threat that has been becoming worse over the past ten years. For instance, there was allegedly a peak volume of DDoS assaults against Amazon AWS in the first quarter of 2020 [8]. The majority of DoS attacks that occur nowadays are distributed (DDoS), in which multiple adversaries launch strikes at once [9]. Therefore, if an attack is detected and prevented by one node, it won't stop and will only make it harder to distinguish between real and fake service requests [10]. The list below demonstrates how the paper is organised: Section I gives a description of the introduction. The recent works are described in Section II. The proposed model and description are explained in Part III. The results are existing in Section IV. In Section V, the conclusion is presented.

2 Recent Works

Mitali Sinha *et al* [2021] have employed In Network-on-Chip architectures, flooding-based denial-of-service attacks are frequently encountered because all on-chip modules are accessible and shared. An attack of this type is initiated by a Malicious Intellectual Property embedded in a System-on-Chip that causes a significant drop in bandwidth by bombarding the NoC with meaningless messages. Sniffer presents an effective MIP localization framework in this study that employs a low-cost machine learning method to accurately map the attack path and come to a consensus regarding the location of the MIPS. Sniffer can localize MIPs with high accuracy and low overhead, according to the testing results. However, the localization of the source of attack (MIPs) in the context of NoC-based designs has not been thoroughly studied.
Mohammad Tayyab *et al* [2020] have proposed that the address depletion problem was fixed with the release of Internet Protocol version 6. It was also necessary to use Internet Control Message Protocol version six messages in recently released features such as the Neighbour Discovery Protocol. This study uses machine learning techniques as single and hybrid

classifiers to evaluate and categorize DDoS threats. An open challenge regarding the detection of DDoS and DoS attacks using collaborative architectures based on the ensemble framework has also been suggested as a solution: the use of block chains this review study is the first to directly address block chain applicability as a potential remedy for IDSs, based on machine learning techniques in this field.

Hassan A. Alamri *et al* [2020] have introduced a SDN is a novel network design that offers centralized running over a essential regulator that separates the data and control aircrafts, thereby addressing the shortcomings of conventional networks. In order to guarantee precise attack identification and effective use of network resources, this study suggests a DDoS mitigation strategy for SDN. The two stages of the method are the Extreme Gradient Boosting Algorithm and a bandwidth control mechanism. The bandwidth control mechanism initiates the algorithm when the adaptive bandwidth profile-based edge and bandwidth control algorithm are violated. Network traffic current that above a predetermined threshold is classified as either abnormal or normal by the algorithm. The collected results demonstrate that the system effectively uses available resources to protect SDN occurrences with high precision and low error.

Ahamed Aljuhani *et al* [2021] have demonstrate a distributed denial of service attack presents a serious risk. Specifically, DDoS attack aims to interfere with and prevent authorized users from accessing services by overflowing the target with an excessive volume of malicious requests. This paper reviews recent research on single and hybrid machine learning slants for DDoS detection in contemporary networking systems. The study also covers different guard systems based on machine learning approaches that operate in virtualized contexts, including network functions virtualization environments, cloud computing, and software-defined networks. The benefits of this strategy are that the research community may create and develop systems that are effective in thwarting various forms of DDoS attacks.

Rajorshi Biswas *et al* [2021] have presented a critical to keep an eye on every internal datacenter flow to defend against internal spread denial-of-service attacks. This paper presents a hierarchical VM clustering technique that follows a current federation approach based on behavioural similarity among. All flows within a group have a constant sampling rate. It investigates the connection between DDoS result rate and sampling rate. Create an optimization problem that can be solved with mix-integer linear programming to determine the best sampling rate distribution. Sample rate distribution optimization problem, which can only be solved by an optimization solver.

3 Proposed Work Explanation

The proposed approach makes use of a deep learning technique to forecast the various DDoS attack types. This makes use of the SDN Dataset. The input dataset is analysed as part of the first step of data pre-processing. StandardScaler is used for pre-processing the datasets. To deal with the superfluous data in the dataset, data pre-processing techniques are applied. Following

data pre-processing in the data analysis process is researching the observed data. The observed data is a step in the data analysis process that comes after data pre-processing. The following deep learning classifier technique is acceptable for the given data. MLP and LSTM are used in the deep learning classifier approach. The Multilayer Perceptron (MLP) uses training to educate the network, and following training, it generates predictions quickly. The DDOS attacks are classified using the Long Short-term Memory Networks Algorithm classifier in order to identify problems in that area.

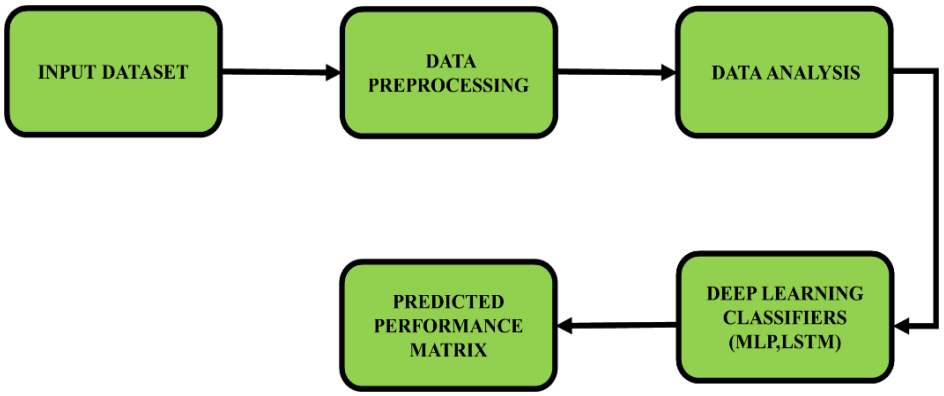

Figure 1: *Proposed Block diagram*

Input Dataset

To guarantee that intrusion detection systems are accurate, the dataset that was utilized has to be scrutinized. Strong network resilience is required in light of today's exponential growth of networks and applications. Choosing the appropriate learning and testing datasets is how this is achieved.

Data Preprocessing

The process of transforming Data pre-processing is the process of converting raw data into a format that computers and machine learning can understand and analyse. It is a phase in the process and mining data. Text, images, video, and other raw, unprocessed data from the real world are jumbled. It frequently lacks a consistent tone in addition to possibly having errors and inconsistencies.

Data preprocessing steps

- Data Transformation
- Data cleaning
- Data quality assessment
- Data reduction

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (12), December 2023, pp. 25-37

Article Title: DDOS Attack Prediction using Deep Learning

Data quality assessment

Almost any data set has a variety of anomalies and underlying matters that should be watched out for, such as:

Mismatched data types

Gather information from a variety of sources; it could arrive in many formats. Even though reformatting your facts for computers is the conclusion goal of this entire technique, you still essential to start with data that is configured identically.

Mixed data values

It's likely that changed sources describe traits differently, using terms like "man" or "male." It is necessary to standardize each of these value descriptions.

Missing data

Check for survey questions that remain unaddressed, blank gaps in text, and missing data fields. Incomplete or human error-related data may be the cause of this. To handle incomplete data and carry out data cleansing.

Data cleaning

The process of filling in missing information and removing, repairing, or restoring erroneous or superfluous information from a data set is known as data cleaning. Dating cleansing is the most important pre-processing step because it ensures that your data is prepared for your needs in the future. Through data cleaning, all of the inconsistent data that you discovered during the data quality calculation will be secure.

Data transformation

Our data has already started to be modified through However, data transformation initiates the process of adapting the information into the appropriate formats.
This generally arises in one or more of the resulting situations:

- Discreditation
- Concept hierarchy generation
- Feature selection
- Aggregation
- Normalization

Aggregation

Data aggregation is the process of joining all of your data into a consistent presentation.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (12), December 2023, pp. 25-37

Article Title: DDOS Attack Prediction using Deep Learning

Normalization

By transforming data into a normalized variety, normalization enables more precise comparisons. For instance, If it is comparing staff loss or gain across several establishments (some with a dozen employees and some with 200+), it must scale the data within a given choice, such as -1.0 to 1.0 or 0.0 to 1.0.

Feature selection

The process of identifying which variables—features, structures, groupings, etc.—are most important to your analysis is known as feature selection. ML models will be trained on these structures. It's crucial to keep in mind that adding more characteristics could make training take longer and occasionally produce less accurate results because some feature skills might not connect to or be as common in the facts.

Discreditization

Discreditiization divides data into more manageable chunks. It is comparable to discarding in certain ways, but it often occurs after the data has been cleaned. For example, instead of using precise minutes and seconds, you could syndicate data to fall into groupings such as 0-15 minutes, 15-30 minutes, and so on when determining the normal regular exercise.

Concept hierarchy generation

The creation of concept hierarchies can introduce a grading between and within your features that in the unique data. You may include the hierarchy for the species Canis if your research includes animals like coyotes and wolves, for instance.

Data analysis

While many individuals, institutions, and authorities handle data investigation in various ways, most of them may be added up in a single, universal definition. The method of cleaning, converting, and analysing rare information to produce relevant and useful data that helps businesses make informed decisions is known as data analysis. By providing relevant information, the method reduces the risks associated with decision making. Facts and insights, many of which are presented as tables, graphs, charts, and graphics.

Data analysis process

The data examination process, also known as the data analysis phases, contains meeting all of the material, dispensation it, examining it, and applying it to identify decorations and other insights.

- **Data Requirement Gathering:** Consider your objectives for showing this examination, the kinds of data you hope to employ, and the figures you aim to examine.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (12), December 2023, pp. 25-37

Article Title: DDOS Attack Prediction using Deep Learning

- **Data Collection:** It's time to start gathering facts from your foundations based on the needs you've identified. Case studies, questionnaires, interviews, surveys, focus groups, and direct observation are some examples of sources. Make a point of organizing the data you've collected for analysis.
- **Data Cleaning:** Since not all of the data you collect will be useful, it is time to purge it. This is the method for getting rid of white spaces, matching records, and basic errors. Before sending the data to be analysed, it must be prepared.
- **Data Analysis:** Here, you'll Use data analysis software and additional resources to assist you in deciphering, comprehending, and drawing conclusions from the data. Among the tools available are Excel, Rapid Miner, Python, R, Metabases Looker, Rewash, Charito, and Microsoft Power BI are a few tools for data analysis.
- **Data Interpretation:** you must now interpret your results and regulate the best next steps based on your results.
- **Data Visualization:** Data visualization is an elaborate way of saying "graphically show your information in a way that people read and understand it." Graphs, maps, bullet points, and charts and other methods all be used. Visualization can help you gain valuable insights by allowing you to comparison datasets and observe associations.

Deep learning

Artificial intelligence (AI) and machine learning that mimics how people learn certain things is called deep learning.

Long short-term memory (LSTM)

Although it is an unconfirmed learning method, it is officially trained using supervised knowledge systems known as self-supervised learning. It is a class of regular neural RNNs can learn long-term dependencies, which is specifically useful in classification prediction difficulties.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (12), December 2023, pp. 25-37

4 Results and Discussion

Input dataset

	dt	switch	src	dst	pktcount	bytecount	dur	dur_nsec	tot_dur	flows
0	11425	1	10.0.0.1	10.0.0.8	45304	48294064	100	716000000	1.010000e+11	3
1	11605	1	10.0.0.1	10.0.0.8	126395	134737070	280	734000000	2.810000e+11	2
2	11425	1	10.0.0.2	10.0.0.8	90333	96294978	200	744000000	2.010000e+11	3
3	11425	1	10.0.0.2	10.0.0.8	90333	96294978	200	744000000	2.010000e+11	3
4	11425	1	10.0.0.2	10.0.0.8	90333	96294978	200	744000000	2.010000e+11	3
5	11425	1	10.0.0.2	10.0.0.8	90333	96294978	200	744000000	2.010000e+11	3
6	11425	1	10.0.0.1	10.0.0.8	45304	48294064	100	716000000	1.010000e+11	3
7	11425	1	10.0.0.1	10.0.0.8	45304	48294064	100	716000000	1.010000e+11	3
8	11425	1	10.0.0.1	10.0.0.8	45304	48294064	100	716000000	1.010000e+11	3
9	11425	1	10.0.0.2	10.0.0.8	90333	96294978	200	744000000	2.010000e+11	3

pktrate	Pairflow	Protocol	port_no	tx_bytes	rx_bytes	tx_kbps	rx_kbps	tot_kbps	label
451	0	UDP	3	143928631	3917	0	0.0	0.0	0
451	0	UDP	4	3842	3520	0	0.0	0.0	0
451	0	UDP	1	3795	1242	0	0.0	0.0	0
451	0	UDP	2	3688	1492	0	0.0	0.0	0
451	0	UDP	3	3413	3665	0	0.0	0.0	0
451	0	UDP	1	3795	1402	0	0.0	0.0	0
451	0	UDP	4	3665	3413	0	0.0	0.0	0
451	0	UDP	1	3775	1492	0	0.0	0.0	0
451	0	UDP	2	3845	1402	0	0.0	0.0	0
451	0	UDP	4	354583059	4295	16578	0.0	16578.0	0

Figure 2: *Input Dataset*

The SDN dataset contains There are DoS attacks that can be launched at various layers of the OSI model. DDoS attack scenarios such as, UDP Flood, and ICMP Flood are also included in the SDN dataset. The input dataset is shown in figure 2.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (12), December 2023, pp. 25-37

Data preprocessing

dt	0
switch	0
src	0
dst	0
pktcount	0
bytecount	0
dur	0
dur_nsec	0
tot_dur	0
flows	0
packetins	0
pktperflow	0
byteperflow	0
pktrate	0
Pairflow	0
Protocol	0
port_no	0
tx_bytes	0
rx_bytes	0
tx_kbps	0
rx_kbps	506
tot_kbps	506
label	0
dtype: int64	

Figure 3: *Null Dataset*

The null values are shown in Figure 3. Data preprocessing is a subset of data grounding that encompasses any type of processing that is completed on raw data to get it ready for additional processing. Preprocessing is critical for improving data quality. During preprocessing, the collected SDN datasets are fed into the machine, which is then trained.

Data analysis

Data analysis is the regular application of arithmetical techniques to define and illustrate, abbreviate and review, and evaluate data. Data analysis is used to extract useful information from data and make a decision based on the analysis.

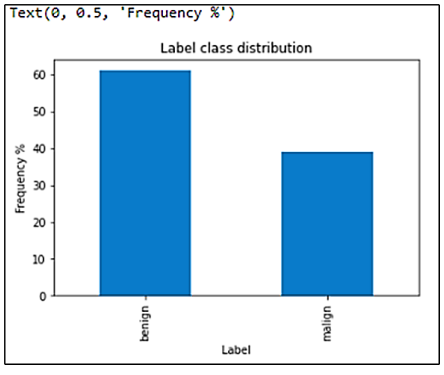

Figure 4: *Label Class Distribution*

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (12), December 2023, pp. 25-37

The label class distribution is shown in figure 4.

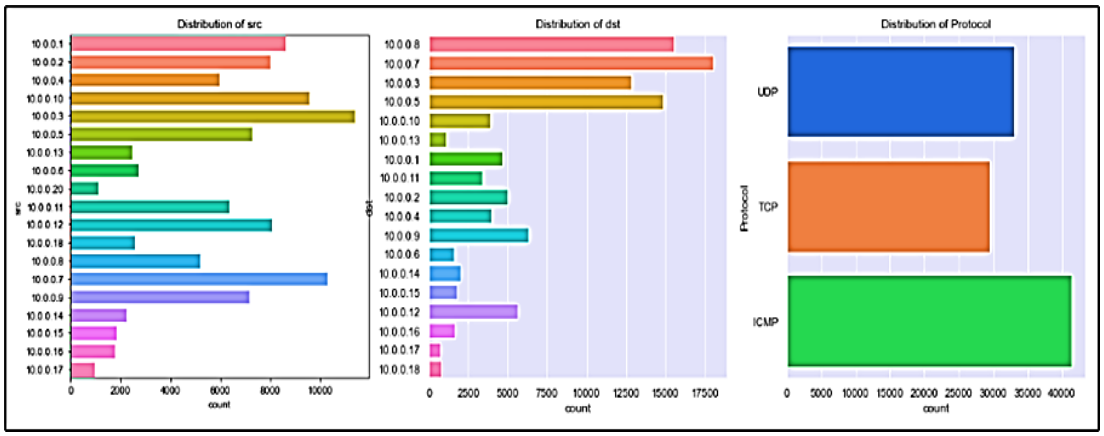

Figure 5: *Distributions of SRC, DST and Protocol*

Distributions of SRC, DST, and Protocol are illustrated in figure 5.

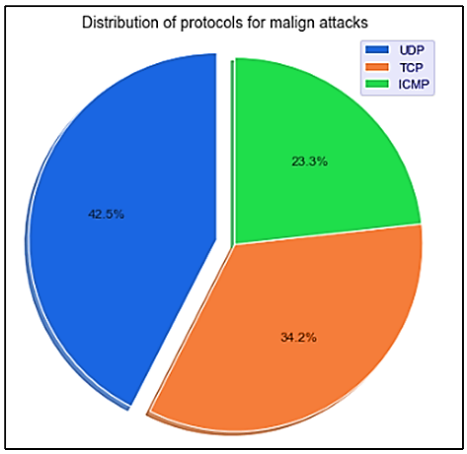

Figure 6: *Distribution of protocol for malign attacks*

Figure 6 demonstrates the spatial distribution of the protocol for malicious attacks.

Multilayer perceptron (MLP)

A multilayer perceptron is a type of neural network that has multiple layers connected in a directed graph, where there is only one path for signals to travel through each node. Every node, excluding the input nodes, has a nonlinear activation function. This figure represents MLP's performance. Figure 5.6 depicts MLP performance.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (12), December 2023, pp. 25-37

Article Title: DDOS Attack Prediction using Deep Learning

```
Accuracy: 98.49%
Recall: 99.19%
Precision: 96.98%
F1-Score: 98.07%
time to train: 34.74 s
time to predict: 3.80 s
total: 38.54 s
```

Figure 7: *Performance of MLP*

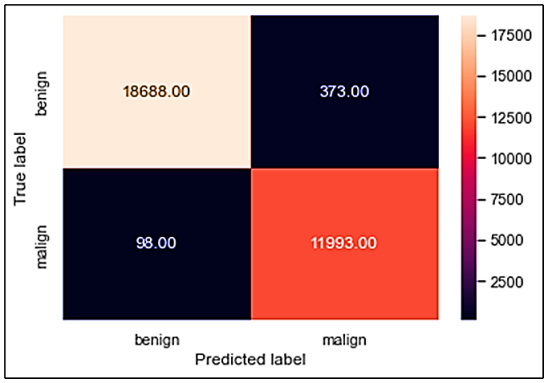

Figure 8: *Confusion matrix for MLP*

Figure 8 depicts the MLP confusion matrix. This graph depicts the value for malignant and benign tumours. The x axis represents the true positive rate, and the y axis represents the false positive rate.

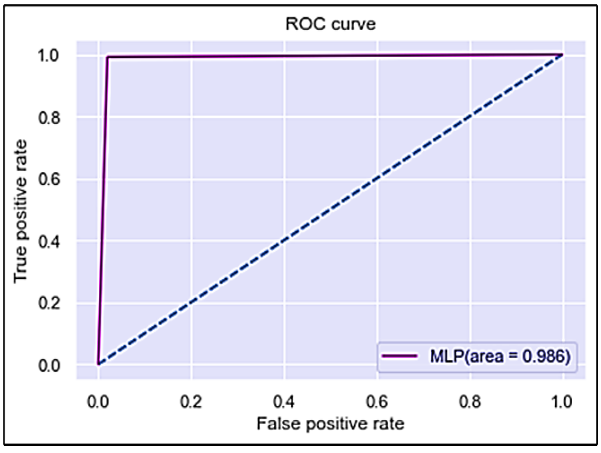

Figure 9: *ROC Curve for MLP*

The ROC curve for MLP is depicted in this figure. The true positive rate is displayed on the x axis, and the false positive rate is displayed on the y axis. MLP is used to determine accuracy, recall precision, and the F1-score.

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 02 (12), December 2023, pp. 25-37

Article Title: DDOS Attack Prediction using Deep Learning

	Accuracy	Recall	Precision	F1-Score	time to train	time to predict	total time
MLP	98.49%	99.19%	96.98%	98.07%	34.7	3.8	38.5
LSTM	98.80%	98.81%	98.11%	98.46%	109.4	6.9	116.3

Figure 10: *Comparison of Classification*

Figure 5.12 depicts how MLP and LSTM predicted Time to train, time to predict, recall, accuracy, F1-score, and total time are all factors to consider.

5 Conclusion

Distributed denial of facility attacks is one of the most dangerous and complex security pressures to computer networks. This proposes a thorough simulation of a deep learning technique for predicting various types of attacks. The arrangement algorithms Multilayer Perceptron and Long Short Term Memory Networks are part of the deep learning technique.it is used to pre-process the datasets. Deep learning techniques are used to detect and organize spread denial of service attacks. MLP is a feed-forward reproduction neural network model that maps input groups to productivity sets. The LSTM classifier is designed to categorize defects based on their type and is used to diagnose faults in that area. This proposed developed a confusion matrix to determine the performance of the model.

Reference

1. Rasim M. Alguliyev; Ramiz M. Aliguliyev; Fargana J. Abdullayeva, Year: 2019, “Deep learning method for prediction of DDoS attacks on social media”, Advances in Data Science and Adaptive Analysis, Vol: 11, no: 01n02, pp. 1950002.

2. Muhammad Ismail Mohmand; Hameed Hussain; Ayaz Ali Khan; Ubaid Ullah; Muhammad Zakarya; Aftab Ahmed; Mushtaq Raza; Izaz Ur Rahman; Muhammad Haleem, Year: 2022, “A machine learning-based classification and prediction technique for DDoS attacks”, IEEE Access, Vol: 10, pp. 21443 – 21454.

3. Wu Zhijun; Xu Qing; Wang Jingjie; Yue Meng; Liu Liang, Year: 2020, “Low-Rate DDoS Attack Detection Based on Factorization Machine in Software Defined Network”, IEEE Access, Vol: 8, pp. 17404 – 17418, 2020.

4. Beny Nugraha; Rathan Narasimha Murthy, Year: 2020, “Deep Learning-based Slow DDoS Attack Detection in SDN-based Networks”, 2020 IEEE Conference on Network Function Virtualization and Software Defined Networks (NFV-SDN), Leganes, Spain.

5. Jesús Arturo Pérez-Díaz; Ismael Amezcua Valdovinos; Kim-Kwang Raymond Choo; Dakai Zhu, Year: 2020, “A Flexible SDN-Based Architecture for Identifying and Mitigating Low-Rate DDoS

Attacks Using Machine Learning", IEEE Access, Vol: 8, pp. 155859 – 155872.

6. Yan Li; Yifei Lu, Year: 2019, "LSTM-BA: DDoS detection approach combining LSTM and Bayes", in 2019 seventh international conference on advanced cloud and big data (CBD), pp. 180 – 185. IEEE.

7. Christos Siaterlis; Basil Maglaris, Year: 2004, "Detecting DDoS attacks using a multilayer Perceptron classifier", In Proceedings of the ICANN, Vol: 10, pp. 118 – 123.

8. Ivan Cvitić; Dragan Perakovic; Brij B. Gupta; Kim-Kwang Raymond Choo, Year: 2021, "Boosting-based DDoS detection in internet of things systems", IEEE Internet of Things Journal, Vol: 9, no: 3, pp. 2109 – 2123.

9. Kiwon Hong; Youngjun Kim; Hyungoo Choi; Jinwoo Park, Year: 2018, "SDN-Assisted Slow HTTP DDoS Attack Defense Method", IEEE Communications Letters, Vol: 22, no: 4, pp. 688 – 691.

10. Shi Dong; Mudar Sarem, Year: 2020, "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.

11. Mitali Sinha; Setu Gupta; Sidhartha Sankar Rout; Sujay Deb, Year: 2021, "Sniffer: A Machine Learning Approach for DoS Attack Localization in NoC-Based SoCs", IEEE Journal on Emerging and Selected Topics in Circuits and Systems, Vol: 11, no: 2, pp. 278 – 291.

12. Mohammad Tayyab; Bahari Belaton; Mohammed Anbar, Year: 2020, "ICMPv6-Based DoS and DDoS Attacks Detection Using Machine Learning Techniques, Open Challenges, and Block chain Applicability: A Review", IEEE Access, Vol: 8, pp. 170529 – 170547.

13. Hassan A. Alamri; Vijey Thayananthan, Year: 2020, "Bandwidth Control Mechanism and Extreme Gradient Boosting Algorithm for Protecting Software-Defined Networks Against DDoS Attacks", IEEE Access, Vol: 8, pp. 194269 – 194288.

14. Ahamed Aljuhani, Year: 2021, "Machine Learning Approaches for Combating Distributed Denial of Service Attacks in Modern Networking Environments", IEEE Access, Vol: 9, pp. 42236 – 42264.

15. Rajorshi Biswas; Sungji Kim; Jie Wu, Year: 2021, "Sampling Rate Distribution for Flow Monitoring and DDoS Detection in Datacenter", IEEE Transactions on Information Forensics and Security, Vol: 16, pp. 2524 – 2534.