International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 06 Issue: 06 | June - 2022 Impact Factor: 7.185 ISSN: 2582-3930

# MACHINE LEARNING BASED DDOS ATTACK DETECTION USING LSTM ALGORITHM

**MANOJ KUMAR M[1], HARIHARAN S[2], MURALI SHANKAR S[3] UMARANI C[4]**

[1]*UG Scholar, Department of CSE, Kingston College, Vellore-59*
[2]*UG Scholar, Department of CSE, Kingston College, Vellore-59*
[3]*UG Scholar, Department of CSE, Kingston College, Vellore-59*
[4]Asst.*Professor, Department of CSE, Kingston College, Vellore-59*

---

**Abstract - With the emergence of network-based computing technologies like Cloud Computing ,Fog Computing and IoT(Internet of Things),the context of digitizing the confidential data over the network is being adopted by various organizations where the security of that sensitive data is considered as a major concern. Over a decade there is a massive growth in the usage of internet along with the technological advancements that demand the need for the development of efficient security algorithms that could withstand various patterns of the security breaches These attacks take advantage of specific limitations that apply to any arrangement asset, such as the framework of the authorized organization's site. In the existing research study, the author worked on an old KDD dataset. It is necessary to work with the latest dataset to identify the current state of DDoS attacks. This paper, used a machine learning approach for DDoS attack types classification and prediction For this purpose, used LSTM and CNN classification algorithms. To access the research proposed dataset UNWS-np-15 was extracted complete framework for DDoS attacks prediction to get the better accuracy**

***Key Words***: DDOS, CNN,LSTM,KDD DATASET

## 1.INTRODUCTION

Now a day's with the advent of 4G,5G networks and economic smart devices there is a massive growth in the usage of the internet that has become a part of daily life. A vast range of services provided over the internet in diverse application areas such as business, entertainment, and education ,etc .made it a vital component in framing various business models. This context made security over wireless networks as the most important factor while using the internet from unsecured connections Different security algorithms and frameworks are developed to enable protection from Internet-based attacks while devising high performance IDS(Intrusion detection systems)which act as a defensive wall while confronting the attacks over internet based devices. Distributed architecture based computing environments like cloud computing and IoT are more prone towards DDoS attacks in which multiple devices are coordinated to launch attacks over distributed targets. DDOS attacks are primarily launched in the context of exhausting the connectivity and the processing of the target server resources in which it enables the access constraints to the legitimate users to utilize the services provided by the target server that leads towards the partial unavailability or total unavailability of the services. The phenomenon of distributed computing is based on the one-to-many dimension in which these types of attacks may cause a possible amount of damage to the server resources.

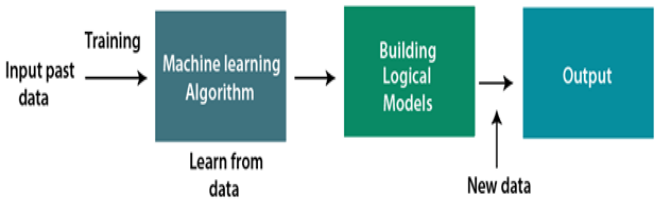



**Figure 1:Machine Learning**

## 2. RELATED WORKS

In the literature review section we briefly explained all the related model and the closest rival to our proposed study. We studied the latest research papers of the past two years for this research work and also Gozde Karatas et al. [2] proposed a machine learning approach for attacks classification. They used different machine learning algorithms and found that the KNN model is best for classification as compared to other research work. Nuno Martins et al. [1] proposed intrusion detection using machine learning approaches. They used the KDD dataset which is available on the UCI repository. They performed different supervised models to balance un classification algorithm for better performance. In this work, a comparative study was proposed by the use of different classification algorithms and found good results in their work. Laurens D'hooge et al. [6] proposed a systematic review for malware detection using machine learning models. They compared different malware datasets from online resources as well as approaches for the dataset. They found that machine learning supervised models are

very effective for malware detection to make a better decision in less time. Xianwei Gao et al. [7] proposed a comparative work for network traffic classification. They used machine learning classifiers for intrusion detection. The dataset is taken is CICIDS and KDD from the UCI repository. They found support vector machine SVM one of the best algorithms as compare to others. Tongtong Suet al. [3] proposed adaptive learning for intrusion detection. They used the KDD dataset from an online repository. These models are Dtree, R-forest, and KNN classifiers. In this study, the authors found that tree and ensemble models are good for classification results. The overall accuracy of the proposed work is 85%. Kaiyuan Jiang et al. [4] proposed deep learning models for intrusion detection. The dataset is KDD and the models are Convention neural network (CNN), BAT-MC, BAT, and Recurrent neural network. The overall model's performance was very good. They found CNN as best for learning. The accuracy is improved from 82% to 85%. Arun Nagaraja et al.

[5] proposed a hybrid model deep learning model for intrusion detection. They combined two deep learning models for the classification of CNN+ LSTM from the RNN model. The dataset was used in this work is KDD. They found an 85.14% average accuracy for the proposed. Yanqing Yang et al. [8] proposed a similarity-based approach for anomaly detection using machine learning. They used k mean cluster model for feature similarity detection and naïve Bayes model used for classification. Hui Jiang et al. [4] used an auto-encoder for labels and performed deep learning classification models on the KDD dataset. They found an 85% average accuracy for the proposed model [9]. SANA ULLAH JAN et al. [10] proposed a PSO-Xgboost model because it is higher than the overall classification accuracy alternative models, e.g. Xgboost, Random-Forest, Bagging, and Adaboost. First, establish a classification model based on Xgboost, and then use the adaptive search PSO optimal structure Xgboost. NSL-KDD, reference dataset used for the proposed model evaluation. Our results show that, PSO-Xgboost model of precision, recall, and macro-average average accuracy, especially in determining the U2R and R2L attacks. This work also provides an experimental basis for the application group NIDS in intelligence. Maede Zolanvari et al. [11] proposed a recurrent neural
network model for classification intrusion detection. They compared other deep learning models with RNN. Finally, they found RNN is the best model for intrusion detection by using the KDD dataset. Yijing Chen et al. [12] proposed a domain that generates an algorithm for botnet classification. It was a multiple classification problem. They used advanced deep learning LSTM for multiple classification problems. They found good results with 89% average accuracy for the proposed work.

Larriva-Novo et al. [13] proposed two benchmark datasets, especially UGR16 and UNSW-NB15, and the most used dataset KDD99 were used for evaluation. The pre-processing strategy is evaluated based on scalar and standardization capabilities. These pre-processing models are applied through various attribute arrangements. These attributes depend on the classification of the four sets of highlights: basic associated highlights, content quality, fact attributes, and finally the creation of highlights based on traffic and traffic quality based on associated titles Collection. The goal of this inspection is to evaluate this arrangement by using different information pre-processing methods to obtain the most accurate model. Our proposition shows that by applying the order of organizing traffic and some preprocessing strategies, the accuracy can be improved by up to 45%. The pre-processing of a specific quality set takes into account more prominent accuracy, allowing AI calculations to effectively group these boundaries identified as potential attacks. Zeeshan Ahmad et al. [14] proposed a scientific classification approach, which depends on the well-known ML and DL processes included in the planning network-based 21446 IDS (NIDS) framework. By examining the quality and certain limitations of the proposed arrangements, an extensive review of the new clauses based on NIDS was conducted. By then, regarding the proposed technology, evaluation measurement, and dataset selection, the ongoing patterns and progress of NIDS based on ML and DL are given. Taking advantage of the deficiencies of the proposed technology, in this paper, we put forward different exploration challenges and give suggestions. Muhammad Aamir et al. [15] proposed AI calculations were prepared and tried on the latest distributed benchmark dataset (CICIDS2017) to distinguish the best performance calculations on information, which contains the latest vectors of port checks and DDoS attacks. The permutation results show that every variation of isolation check and support vector machine (SVM) can provide high test accuracy, for example, more than 90%. According to the abstract scoring criteria cited in this article, 9 calculations from a bunch of AI tests received the most noteworthy score (highest) because they gave more than 85% representation (test) accuracy in 22 absolute calculations. In addition, this related investigation was also conducted to note that through the k-fold cross approval, the area under the curve (AUC) check of the receiver operating characteristic (ROC) curve, and the use of principal component analysis (PCA) for size reduction in preparation for AI execution model. When considering such late attacks, it was found that many checks on different AI calculations of the CICIDS2017 datasets were not sufficient for port checks and DDoS attacks. Kwak et al. [16], proposed a video steganography bot net model. In addition, they plan to use another video steganography technology based on the payload method (DECM: Frequency Division Embedded Component Method), which can use two open devices Virtual Dub and Steg no to implant significantly more privileges than existing tools information. They show that proposed model can be performed in the Telegram SNS

courier, and compared proposed model and DECM with the current image steganography-based botnets and methods in terms of the effectiveness and imperceptibility [17]. Zahid Akhtar et al. [18] proposed a concise overview of malware, followed by a summary of different inspection challenges. This is a hypothetical point of view article that needs to be improved. Duy-Cat and Can. et al [19] became familiar with a model that can identify and arrange distributed denial of service attacks that rely on the use of the proposed program including selected segments of neural tissue. The experimental results of the CIC-DDoS 2019 dataset show that our proposed model beats other AI-based models to a large extent. We also studied the selection of weighted misfortune and the choice of pivotal misfortune in taking care of class embarrassment [20]. Qiumei Cheng et al. [21] proposed a novel in-depth binding review (OFDPI) method with OpenFlow function in SDN using AI computing. OFDPI supports in-depth bundling inspection of the two decoded packages.

## 3.METHODOLOGY PROPOSED

**Methodologies used:**

A Convolutional Neural Network is a Machine Learning calculation which can take in an info picture, relegate significance (learnable loads and predispositions) to different perspectives/objects in the picture and have the option to separate one from the other. The pre-handling expected in a CNN is a lot of lower when contrasted with other characterization calculations. There are different structures of CNNs accessible which have been key in building calculations which power and will control AI all in all within a reasonable time-frame.

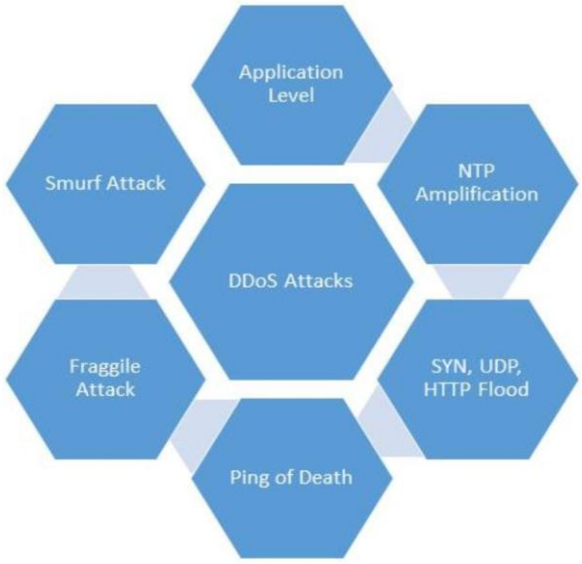



**Figure 2: various types of DDos**

## 3.ARCHITECTURE AND USE CASE DIAGRAM

Diagram is one of the most used words in contemporary architecture and urban design. It's almost a common understanding of using diagrams to explain the design concept. However, it is more than simply showing the audience to help them understand the idea. Diagramming is an element of design in itself.

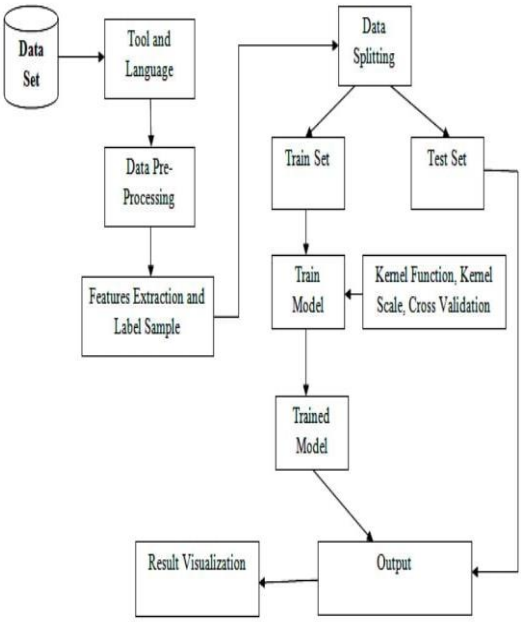



**Figure 3:Architecture Diagram**

**USE CASE DIAGRAM:**

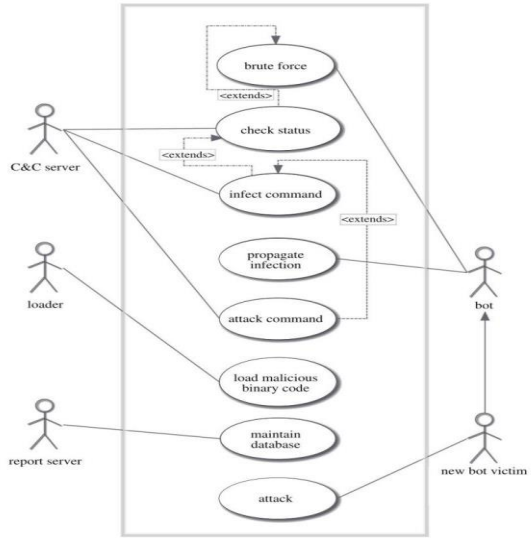



**Figure 4:Use case Diagram**

A use case diagram is a graphical picture of a user's possible interactions with a system. A utilization case outline shows different use cases and various sorts of clients the framework has and will frequently be joined by different kinds of charts too. The use cases are represented by either circles or ellipses. The entertainers are frequently displayed as stick figures. While a utilization case itself could penetrate into a ton of insight concerning each chance, a utilization case outline can assist with giving a more elevated level perspective on the framework It has been said before that "Use case diagrams are the blueprints for your system.

**TRAINING OUTPUT:**

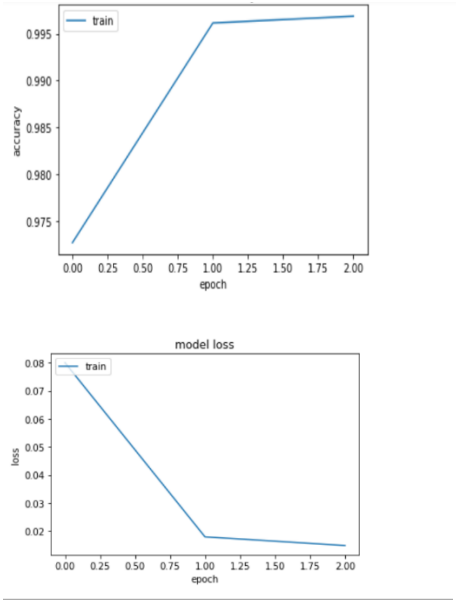
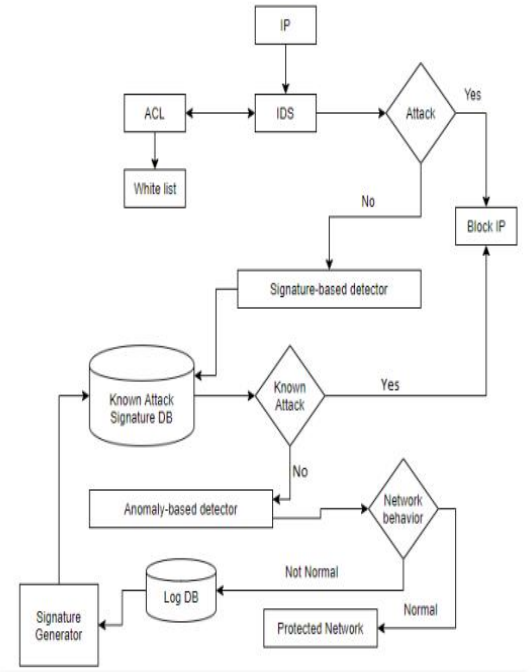



**Figure 5:Training Output**

**ACTIVITY DIAGRAM:**

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.




**3. CONCLUSIONS**:

DDoS attacks analysis and detection were performed using machine learning methods.In this work,a subset of the CICIDS2017 dataset was utilized,which included 10K samples of DDoS and Benign classes.The data contained 84 categorical and numerical features in total,where one feature was dropped,so that feature engineering and machine learning model development were completed with 83 features.A correlation analysis and feature importance exploration using a decision tree were employed in feature engineering..Also,the results of machine learning models,which included decision tree and linear support vector machine models,demonstrated that DDoS and Benign attacks were classified where the accuracy rates of around 100%were achieved.The replication of the original paper was completed,and other machine learning models can be considered for future work.

**ACKNOWLEDGEMENT**

The authors would like to thank MRS.C.UMARANI for her suggestions and excellent guidance throughout the project period.

**REFERENCES**

[1] N. Martins, J.M. Cruz, T.Cruz, and P.H. Abreu,‘‘Adversarial machine learning applied to intrusion and malware scenarios: A systematic review,’’IEEE Access, vol.8, pp. 35403–35419, 2020.

[2] G. Karatas, O. Demir, and O. K. Sahingoz, ‘‘Increasing the performance of machine learning-based IDSs on an imbalanced and up-to-date dataset,’’ IEEE Access, vol. 8, pp. 32150–32162, 2020.

[3] T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, ‘‘BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset,’’ IEEE Access, vol. 8, pp. 29575–29585, 2020
.
[4] H. Jiang, Z. He, G. Ye, and H. Zhang, ‘‘Network intrusion detection based on PSO-xgboost model,’’ IEEE Access, vol. 8, pp. 58392–58401, 2020.