International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Detection Of DDOS Attack Using Machine Learning

Pradeep H K[1], Pavan Kumar[2], Pradeepa A J[3], Prashantha S[4], Saad Faisal Khan[5]

[1]*Information Science and Engineering & J N N College of Engineering*
[2] *Information Science and Engineering & J N N College of Engineering*
[3] *Information Science and Engineering & J N N College of Engineering*
[4] *Information Science and Engineering & J N N College of Engineering*
[5] *Information Science and Engineering & J N N College of Engineering*

Abstract - The Distributed Denial-of-Service (DDoS) attack is one of the most dangerous cyber threats, surpassing traditional Denial-of-Service (DoS) attacks due to its distributed nature, where multiple hosts collectively target a system, rendering its services inaccessible. Addressing this challenge requires an advanced and reliable detection mechanism. This research presents a machine learning-based approach for DDoS attack detection using Logistic Regression, Random Forest, and Neural Network classifiers. The proposed model is trained on a cleaned and pre-processed dataset with feature scaling to enhance model performance. A Flask-based web application deploys these models, enabling real-time prediction through a user-friendly interface. The trained models are evaluated using key metrics such as accuracy, F1-score, precision, recall, and confusion matrix. Comparative analysis reveals the strengths of ensemble-based methods, offering a scalable and robust solution for mitigating DDoS attacks in real-world environments.

Key Words: **Cyber Attack, DDOS Attack, Logistic Regression, Neural Network, Random Forest Algorithm.**

1. INTRODUCTION

In today's digital era, businesses and organizations across the globe face growing concerns over cyber threats. Cybercriminals exploit technological vulnerabilities to launch attacks against individual computers or entire networks, causing significant disruption. Among various cyber threats, Distributed Denial-of-Service (DDoS) attacks stand out as one of the most common and damaging internet security risks. These attacks aim to overwhelm targeted systems, reducing service availability and resulting in financial losses, reputational damage, and operational setbacks.

DDoS attacks can be executed in various ways, making them a persistent challenge for cybersecurity professionals. Their increasing prevalence has driven researchers and organizations to explore advanced detection mechanisms. Machine learning (ML) has emerged as a promising approach, offering scalable solutions capable of processing massive data volumes in real time. Despite extensive research, developing a highly accurate and efficient DDoS detection system remains a critical area of investigation. The primary goal of using machine learning in DDoS detection is to accurately classify incoming requests, distinguishing between legitimate traffic and malicious requests. Achieving high prediction accuracy while minimizing model training time is essential for developing an effective detection system. Key factors influencing the performance of machine learning models include the selection of classification algorithms, dataset size, feature selection methods, and algorithm parameter tuning.

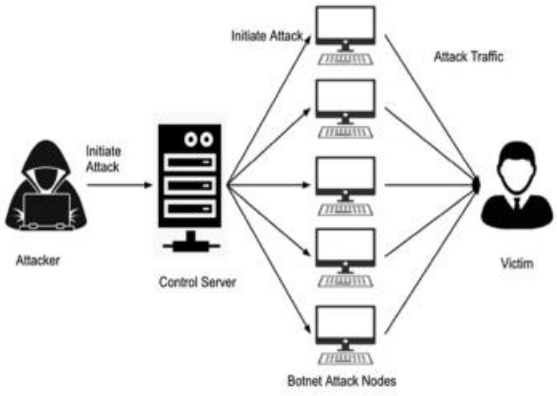

Fig -1: DDOS Attack Scenario

Feature selection plays a crucial role in enhancing detection performance by removing irrelevant or redundant data, thereby reducing training time and improving model accuracy. Proper parameter tuning further optimizes the model's predictive power. In time-sensitive environments, the trade-off between high accuracy and faster model training becomes particularly important. Ensuring precise DDoS detection is critical for minimizing service downtime and mitigating the adverse effects of cyberattacks. This research introduces a DDoS detection framework that utilizes multiple classifiers within a machine learning approach. By systematically analyzing key metrics such as accuracy, precision, recall, and F1-score, we aim to identify the most effective model for real-time DDoS detection and mitigation, contributing to a more secure and resilient cyberspace.

2. RELATTED WORKS

Recent research in DDoS attack detection has demonstrated significant advancements through machine learning applications, particularly in addressing the growing sophistication of cyber threats and the increasing complexity of network environments. Santhosh et al [1]. (2024) achieved 97% accuracy using a modified XGBoost classifier on the CICDDoS2019 dataset, emphasizing the importance of preprocessing techniques such as label encoding for effective algorithm performance and demonstrating superior scalability in handling large-scale network traffic data. Belide and Gandhi (2024) [2] established the importance of feature selection across multiple algorithms including logistic regression and random forests, while also developing a user-friendly web application for practical implementation, highlighting the critical balance between model complexity and operational efficiency. Nalayini

et al [3]. highlighted Support Vector Machine's superiority over Naive Bayes, despite higher false alarm rates, emphasizing the need for continuous adaptation in detection systems and the significance of real-world applicability in model development, particularly in dynamic network environments where attack patterns constantly evolve. Kumari et al [4].'s examination of deep learning-based strategies showed marked improvements over traditional approaches, particularly noting the effectiveness of semi-supervised ML algorithms in connectivity diversity prediction and demonstrating the potential for reduced reliance on labeled training data. Ismail et al [5]. introduced hybrid models combining CNN and LSTM architectures, achieving 85.14% accuracy on the KDD dataset, complementing Al-Shareeda et al.'s (2022) [6] findings on multi-algorithm integration and adaptive techniques within IoT environments, while also addressing the unique challenges posed by resource-constrained devices. Rustam et al. (2022) [7] demonstrated 100% accuracy in specific attack scenarios using Random Forest with hybrid feature selection through PCA and SVD, particularly excelling in detecting application layer attacks such as SSL and HTTP flood DDoS attacks, while emphasizing the importance of feature dimensionality reduction in improving computational efficiency. Kachavimath et al. [8] achieved 45.6% reduction in computational resources through advanced feature reduction, while maintaining a false alarm rate below 1% across both simulated and real-time datasets from Baidu cloud computing, demonstrating the practical feasibility of implementing sophisticated detection mechanisms in production environments. Amitha and Srivenkatesh (2023) [9] showcased LSTM models achieving 99.947% accuracy in cloud environments, particularly when integrated with RBFNN for temporal dependency analysis, demonstrating superior performance across various attack types including PortMap, NetBIOS, and DNS attacks, while also addressing the challenges of model deployment in distributed cloud architectures. Kumar et al [10]. further validated LSTM's effectiveness, achieving 98% accuracy on the CICDDoS2019 dataset, while highlighting the crucial role of the SoftMax activation function and binary cross-entropy loss function in optimization, and demonstrating the importance of proper model architecture selection in achieving optimal performance. The research collectively emphasizes the importance of hyperparameter tuning, dataset diversity, and real-time monitoring capabilities for effective DDoS detection, with particular attention to the trade-offs between model complexity and detection speed. While these studies demonstrate significant progress, common challenges persist, including data quality dependencies, computational resource requirements, and implementation complexity. Additionally, the studies highlight the importance of model interpretability and the need for balanced datasets to avoid biased training outcomes, while also addressing the challenges of maintaining detection accuracy in the face of zero-day attacks and previously unseen attack patterns. Future research directions suggest focus on hybrid models, transfer learning, and real-time detection capabilities, with particular emphasis on developing more robust datasets and exploring the applicability of incremental learning for adapting to new attack vectors. The integration of these methodologies not only enhances detection accuracy but also contributes to the development of more resilient network infrastructures capable of withstanding sophisticated cyber threats, indicating a promising trajectory for ML-based DDoS detection advancements. The literature also points to emerging trends in federated learning approaches for distributed detection systems and the potential of quantum computing in enhancing computational capabilities for complex model training. Furthermore, the research emphasizes the growing importance of explainable AI techniques in building trust and accountability in automated detection systems, while also highlighting the need for standardized evaluation metrics and benchmarking frameworks to facilitate meaningful comparisons between different detection approaches.

3. PROPOSED SYSTEM

In this research, the CICDoS2019 dataset and machine learning techniques are used to construct a model for the detection of DDoS attacks.

- Using the Linear Regression, Neural network and Random Forest Classifier the attack detection model is trained and tested.
- The Linear Regression, Neural Network and Random Forest Classifier algorithm with the lowest accuracy rate is modified to achieve a higher accuracy rate.
- The obtained results are compared in order to determine which model has the highest accuracy rate for detecting DDoS attacks.

4. SYSTEM ARCHITECTURE

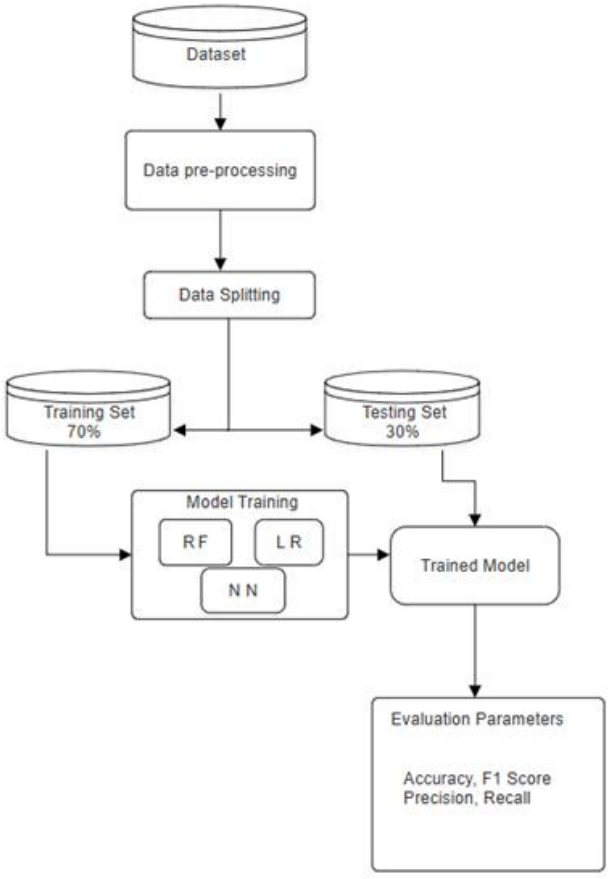

Fig-2: System Architecture of DDoS Attack Detection Model.

5. MATERIAL AND METHODS

After careful analysis the system has been identified to have the following modules:

A. DATASET

We use the CICDDoS2019 dataset, one of the newest and most popular datasets, to verify the effectiveness of algorithms. As a result, it accurately depicts both conventional network traffic and various Botnet-based network attacks. The CICDDoS2019 dataset was used in this investigation (DDoS Evaluation Dataset). This dataset includes 78 features extracted using CICFlowMeter and 11 types of DoS attack s that were produced in a controlled setting. Additionally, the researchers who created this dataset (from the Canadian Institute for Cybersecurity) devised a technique to identify the 22 key components of each DDoS attack.

B. DATA PREPROCESSING

Pre-processing is used to change unstructured data into a machine learning-friendly format. The time-consuming step in data analysis is data pre-processing. Pre-processing data is crucial to getting clear data, despite the fact that it takes time.

C. RANDOM FOREST CLASSIFIER

A random forest combines the decision tree. When compared to other classifiers, it is very quick. After feature scaling, the machine learning classification model is the next stage. We will employ the random forest classification method in the study we have presented. The suggested model uses the random forest, popular and effective machine learning classification techniques, to make a number of decisions.

D. LOGISTIC REGRESSION

Linear Regression is a widely-used statistical model for binary classification problems. It predicts the probability of a target class by applying a logistic function to the linear combination of input features. In the presented study, logistic regression is utilized as one of the machine learning classifiers after feature scaling. Its simplicity and interpretability make it a valuable choice for initial predictions. By optimizing its parameters, the model ensures a balance between computational efficiency and prediction accuracy.

E. NEURAL NETWORK CLASSIFIER

Neural Network Classifier are advanced machine learning models designed to simulate the way human brains process information. They consist of layers of interconnected nodes (neurons) that capture complex relationships within the data. In this study, a neural network classifier is employed to detect DDoS attacks, leveraging its ability to learn non-linear patterns and intricate data structures. After feature scaling, the neural network model undergoes training to refine its predictive capabilities. This method is particularly effective in handling high-dimensional data, offering a sophisticated solution for accurate classification.

E. DATA SPLITTING

We separated the dataset into training and testing datasets for our model.

● Training Each dataset's data is given to pre-processing and feature extraction throughout the training phase. At this stage, classes are assigned to trained features once they have been trained for each unique piece of dataset data. This algorithm has two classes: normal and DDoS attack.

● Testing The test samples are sent into the classifier during testing to categories the test instance with the specified class using the training features. Correct classification of the provided class by the classifier will improve the process's output

F. DATA VISUALIZATION

A larger amount of information represented in graphic form is easier to understand and analyze. So, in this model the classified output from the classifiers is to be visualized and compared.

6. RESULTS AND DISCUSSIONS

All of the outcomes of the models we suggested are presented in this section. An explanation of the conclusions is provided along with each step-by-step presentation of the results in figures. We give a quick overview of the performance of our proposed model and compare it to a few of its rivals, direct rivals, and earlier studies and it detect normal traffic or DDoS attack

● DATA PREPROCESSING

To look at the information and significance of pictorial representation. Here, we utilized a heat map to graphically represent the missing values. The results demonstrate that there are no unnecessary values that should be eliminated. Additionally, during the data pre-processing phase, we noticed and noted that almost-cleanness of our datasets.

● LABEL ENCODING

Computers can understand on and off, so they do not work with letter information. Additionally, in this instance, our information's letter form is incomprehensible to our computer algorithms. Consequently, it is crucial to digitize this data so that our suggested model can comprehend it. We can change the machine learning technique used by the tag encoder into the desired shape.

LOGISTIC REGRESSION CLASSIFIER

- CLASSIFICATION RESULT:

Logistic Regression achieves competitive performance in classifying benign and DDoS traffic. Key metrics indicate that the recall (RE) factor is 99.3%, while the precision (PR) factor reaches 91.3%. The F1 score stands at 95.1%, and the model demonstrates an average accuracy (AC) of 94.6%. Despite its simplicity, Logistic Regression proves effective in binary classification tasks.

- PREDICTION RESULT:

Logistic Regression is capable of providing reliable predictions with an accuracy of approximately 94.6% when tested against unseen data. While slightly less robust than ensemble methods, it remains a viable choice for scenarios requiring interpretable and efficient classification models.

NEURAL NETWORK CLASSIFIER

- CLASSIFICATION RESULT:

Neural Networks, with their ability to model non-linear relationships, achieve outstanding results. The recall (RE) factor is 99.1%, and the precision (PR) factor is 97.6%. The average accuracy (AC) is 98.3%, with an F1 score of 98.4%. These results showcase the Neural Network's ability to accurately distinguish between benign and malicious traffic, even in complex datasets.

- PREDICTION RESULT:

Neural Networks deliver predictions with an accuracy of 98.3% on test datasets. This classifier excels in handling high-dimensional and intricate datasets, making it ideal for identifying DDoS attacks with high precision.

RANDOM FOREST CLASSIFIER

- CLASSIFICATION RESULT:

The Random Forest classifier excels in distinguishing between benign and DDoS traffic. Key metrics include a recall (RE) factor of 99.9%, precision (PR) of 100%, and an F1 score of 99.95%. The model achieves an average accuracy (AC) of 99.95%, demonstrating its robustness and reliability. Its ensemble nature allows it to effectively handle high-dimensional data while reducing overfitting.

- PREDICTION RESULT:

Random Forest provides highly accurate predictions, achieving near-perfect accuracy of 99.95% when tested on unseen data. This makes it a standout choice for real-world DDoS detection systems. Its ability to identify patterns and relationships in features such as packet size, flow duration, and protocol type ensures exceptional detection capabilities.

Table -1: Performance Measures

	CLASSIFIER		
MEASURES	LR	NN	RF
ACCURACY	0.9463	0.9827	0.9995
PRECISION	0.9128	0.9760	1.0000
F1-SCORE	0.9512	0.9837	0.9995
RECALL	0.9930	0.9915	0.9990

7. COMPARISON OF WORK

In previous research, the CICDDoS2019 dataset was used alongside various machine learning and deep learning models, such as CNNs and LSTMs, achieving accuracies of 79% and 85%, respectively. These studies highlight the potential of advanced algorithms but also reveal the need for improved performance and reliability in DDoS detection.

For our proposed work, we utilized the CICDDoS2019 dataset and machine learning classifiers such as Random Forest, Logistic Regression, and Neural Networks. Through rigorous experimentation and hyperparameter optimization, we achieved accuracy rates of 99.95%, 94.63%, and 98.27%, respectively. Among these models, Random Forest consistently demonstrated superior performance, with a perfect precision score of 100% and an F1 score of 99.95%, making it the most effective classifier in our study.

Our findings indicate that supervised learning methods outperform unsupervised approaches in detecting DDoS attacks. However, it is essential to note that the dataset's quality and composition significantly impact model performance. The CICDDoS2019 dataset provided a robust foundation for training and testing, enabling high precision and reliability in our results.

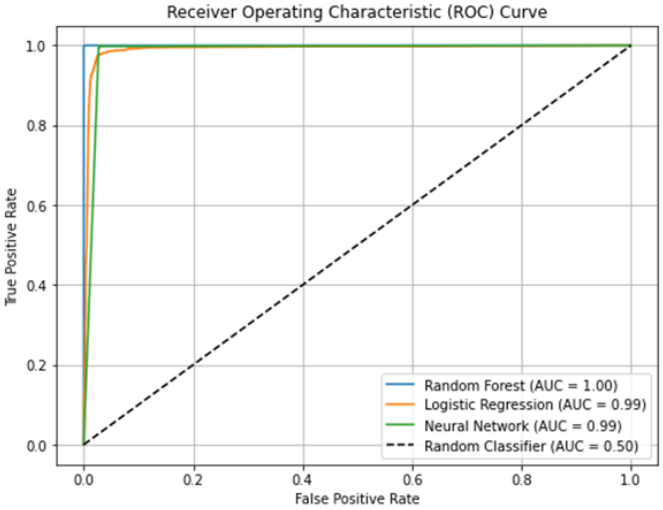

Fig -3: Comparison of Model

8. CONCLUSION

This project presented a systematic and comprehensive approach to detecting Distributed Denial-of-Service (DDoS) attacks. The CICDDoS2019 dataset, contributed by the Canadian Institute for Cybersecurity, was utilized as the foundation for model training and evaluation. Data preprocessing and feature extraction were performed using Python and Visual studio to ensure the dataset's quality and readiness for analysis.

Supervised machine learning techniques were applied, including Random Forest, Logistic Regression, and Neural Networks. Each model generated predictions and classification results, demonstrating strong performance across key metrics. Among these, the Random Forest classifier achieved the highest reliability with an accuracy of 99.95%. Furthermore, a comparative analysis highlighted the effectiveness of ensemble methods like Random Forest and advanced architectures like Neural Networks in distinguishing between benign and malicious traffic.

By comparing the proposed models to previous research, which achieved accuracies of 79% and 85% using CNN and LSTM-based approaches, this study demonstrated a significant improvement. With accuracies of 99.95%, 94.63%, and 98.27% for Random Forest, Logistic Regression, and Neural Networks respectively, this work underscores the potential of supervised learning in enhancing DDoS detection. These results validate

the robustness of the proposed models and their ability to adapt to real-world network security challenges effectively.

REFERENCES

[1] S. Santhosh, M. Sambath, and J. Thangakumar, "Detection of DDoS Attack using Machine Learning Models", In IEEE Access, Vol. 12, No. 1, pp. 1-15, 2023.
[2] Belide, M., and Gandhi, S., "DDoS Attack Classification with Machine Learning", In International Journal of Novel Research and Development, Vol. 9, No. 3, pp. 420-434, 2024.
[3] C. M. Nalayini, and Dr. Jeevaa Katiravan, "Detection of DDoS Attack using Machine Learning Algorithms", In Journal of Emerging Technologies and Innovative Research, Vol. 9, No. 7, pp. 231-231, 2022.
[4] Kimmi Kumari, and M. Mrunalini, "Detecting Denial of Service attacks using machine learning algorithms", SpringerOpen, Vol. 9, No. 56, pp. 1-17, 2022.
[5] Ismail, M., Mohmand, M. I., Hussain, H., Khan, A. A., Ullah, U., Zakarya, M., Ahmed, A., Raza, M., Ur Rahman, I., and Haleem, M., "A Machine Learning-Based Classification and Prediction Technique for DDoS Attacks", In Journal of Network and Computer Applications, Vol. 10, No. 2022, pp. 21454, 2022.
[6] Al-Shareeda, M. A., Manickam, S., and Saare, M. A., "DDoS attacks detection using machine learning and deep learning techniques," In Bulletin of Electrical Engineering and Informatics, Vol. 13, No. 4, pp. 939-950, 2022.
[7] Rustam, F., Mushtaq, M. F., Hamza, A., Farooq, M. S., Jurcut, A. D., and Ashraf, I., "Denial of Service Attack Classification Using Machine Learning with Multi-Features," In Electronics, Vol. 11, No. 20, pp. 3817, 2022.
[8] Amit V Kachavimath, and Narayan D G, "Distributed Denial of Service Attacks Detection using Deep Learning in Software Defined Network", In IEEE, Vol. 13, No. 1, pp. 1-10, 2022.
[9] Marram Amitha, and Dr. Muktevi Srivenkatesh, "DDoS Attack Detection in Cloud Computing Using Deep Learning Algorithms," In International Journal of Intelligent Systems and Applications in Engineering, Vol. 11, No. 4, pp. 82-90, 2023.
[10] Deepak Kumar, R.K. Pateriya, Rajeev Kumar Gupta, Vasudev Dehalwar, and Ashutosh Sharma, "DDoS Detection using Deep Learning", ScienceDirect, Vol. 218, pp. 2420-2429, 2023.