Scientia Africana, Vol. 22 (No. 3), December, 2023. Pp 153-164
© Faculty of Science, University of Port Harcourt, Printed in Nigeria *ISSN 1118 – 1931*

A FRAMEWORK FOR THE DETECTION OF DISTRIBUTED DENIAL OF SERVICE ATTACKS ON NETWORK LOGS USING ML AND DL CLASSIFIERS

[1]Musa, M. O. and [2]Odokuma E. E.
[1]Department of Cybersecurity, Faculty of Computing, University of Port Harcourt,Nigeria.
[2]Department of Computer Science, Captain ElechiAmadi Polytechnic, Port Harcourt,Nigeria.
Corresponding author E-mail: martha.musa@uniport edu.ng

Received: 24-09-2023
Accepted: 06-11-2023

https://dx.doi.org/10.4314/sa.v22i3.14

Journal Homepage: http://www.scientia-african.uniportjournal.info
Publisher: *Faculty of Science, University of Port Harcourt.*

ABSTRACT

Despite the promise of machine learning in DDoS mitigation, it is not without its challenges. Attackers can employ adversarial techniques to evade detection by machine learning models. Moreover, machine learning models require large amounts of high-quality data for training and continuous refinement. Security teams must also be vigilant in monitoring and fine-tuning these models to adapt to new attack vectors. Nonetheless, the integration of machine learning into cybersecurity strategies represents a powerful approach to countering the persistent threat of DDoS attacks in an increasingly interconnected world. This paper proposed Machine Learning (ML) models and a Deep Learning (DL) model for the detection of Distributed Denial of Service Attacks (DDOS) on network system. The DDOS dataset is highly imbalanced because the number of instances of the various classes of the dataset are different. To solve the imbalance problem, we performed random under-sampling using under sampling technique in python called random under-sampler. The down sampled dataset was used for the training of the ML and DL classifiers. The trained models are random forest, gradient boosting and recurrent neural network algorithms on the DDOS dataset. The model was trained on the DDOS dataset by fine tuning the hyper parameters. The models was used to make prediction in an unseen dataset to detect the various types of the DDOS attacks. The result of the models were evaluated in terms of accuracy. The results of the models show an accuracy result of 79% for random forest, 82%, for gradient boosting, and 99.47% for recurrent neural network. From the experimental results.

INTRODUCTION

Machine learning can be applied to cyber security for the detection of cyber-attacks such as Distributed Denial of Service (DDoS)and other cyber threats. DDoS attacks involve overwhelming a target system or network with a flood of traffic, it inaccessible to legitimate users. These attacks often employ a network of compromised devices, creating a distributed attack infrastructure. Machine learning, on the other hand, is a subset of artificial intelligence that involves training algorithms to recognize patterns and make predictions based on data. When applied to cybersecurity, machine learning has the potential to mitigate the impact of DDoS attacks (Singh *et al.*, 2020).

Machine learning can play a crucial role in DDoS attack detection and mitigation. By analyzing network traffic patterns, machine learning algorithms can identify abnormal behavior indicative of a DDoS attack in real-time. These algorithms can differentiate between legitimate user traffic and malicious requests, helping security teams respond swiftly to mitigate the attack. Furthermore, machine learning models can adapt to evolving attack techniques, enhancing their accuracy over time. This adaptability is crucial as attackers continually refine their tactics (Aladaileh*et al.*, 2020).

One significant advantage of machine learning in combating DDoS attacks is its ability to automate the response. When an attack is detected, machine learning systems can trigger automated countermeasures, such as rerouting traffic, filtering malicious requests, or deploying additional server resources to absorb the traffic. This rapid response minimizes downtime and ensures that critical online services remain accessible to users, even during a DDoS attack.

Despite the promise of machine learning in DDoS mitigation, it is not without its challenges. Attackers can employ adversarial techniques to evade detection by machine learning models. Moreover, machine learning models require large amounts of high-quality data for training and continuous refinement. Security teams must also be vigilant in monitoring and fine-tuning these models to adapt to new attack vectors. Nonetheless, the integration of machine learning into cybersecurity strategies represents a powerful approach to countering the persistent threat of DDoS attacks in an increasingly interconnected world (Sharafaldin*et al.*, 2019).

RELATED WORKS

Kimmi and Mrunalini (2022) propose a machine learning method for detecting Distributed Denial of Service (DDoS) attacks. They analyze data packet throughput to differentiate between regular and malicious events. They use CAIDA 2007 datasets and use methods like logistic regression and naive Bayes to enhance analysis. The study introduces two models: a mathematical model representing real-world systems and a machine learning model that learns patterns and improves performance. The proposed model aims to establish a correlation between request inter-arrival time and throughput. Additional studies on throughput are also conducted to detect DDoS attacks.

The study conducted by Jiangtao et al. (2019) this study presents a method for detecting Distributed Denial of Service (DDoS) attacks using machine learning. The proposed method consists of two main steps: feature extraction and model identification. During the feature extraction phase, the DDoS attack traffic characteristics, which exhibit a significant fraction, are retrieved through the comparison of data packets that have been categorised based on predefined rules. During the model detection stage, the features that have been extracted are utilised as input features for machine learning purposes. The attack detection model is then trained using the random forest approach. The empirical findings demonstrate that the machine learning-based DDoS attack detection method, as presented, exhibits a commendable level of efficacy in detecting prevalent DDoS attacks.

Najafimehr, Zarifzadeh, and Mostafavi (2023) conducted a study. This survey study presents a thorough classification system of machine learning techniques utilised in the detection of Distributed Denial of Service (DDoS) assaults. The paper examines supervised, unsupervised, and hybrid approaches, while also addressing the associated problems. Additionally, the researchers conducted an in-depth analysis of pertinent datasets, emphasising their respective merits and drawbacks. Furthermore, they put forth potential avenues for future investigation in order to bridge the existing research gaps within this field. This work aims to enhance the comprehension of DDoS attack detection techniques, hence assisting researchers and practitioners in the

Scientia Africana, Vol. 22 (No. 3), December, 2023. Pp 153-164
© Faculty of Science, University of Port Harcourt, Printed in Nigeria *ISSN 1118 – 1931*

development of robust cybersecurity strategies to counteract these attacks.

In this study, Subhan et al. (2023) provide a machine learning pipeline designed to effectively tackle the problem of detecting Distributed Denial of Service (DDoS) attacks in Internet of Things (IoT) networks. The utilised methodologies consist of three main components: (i) a processing module responsible for data preparation in order to facilitate further analysis, (ii) a dynamic attribute selection module that identifies the most suitable and efficient features, hence reducing training time, and (iii) a classification module designed to detect Distributed Denial of Service (DDoS) attacks. The success of our technique is assessed by employing the CICI-IDS-2018 dataset and five machine learning classifiers that are known for their power and simplicity: Decision Tree (DT), Gaussian Naive Bayes, Logistic Regression (LR), K-Nearest Neighbour (KNN), and Random Forest (RF).

The paper titled "DDoS Attack Detection in IoT-Based Networks Using Machine Learning Models: A Survey and Research Directions" was proposed by Alahmadi et al. (2023). The objective of this study is to conduct a comprehensive assessment of pertinent studies and publications pertaining to the subject of detecting Distributed Denial of Service (DDoS) attacks in Internet of Things (IoT) networks by leveraging machine learning methodologies.

Ebtihal and Onytra (2021) put forth a proposal for the detection of Distributed Denial of Service (DDoS) attacks through the utilisation of machine learning models. This study entails the examination of various machine learning models and their use within the DDoS detection system. This study examines the matter of improving the accuracy of detecting DDoS assaults by utilising the widely recognised CICDDoS2019 dataset. The researchers employed two primary methodologies to identify the most pertinent aspects. Based on the obtained findings, it was observed that the Random Forest Machine Learning model exhibited the highest level of detection accuracy, achieving a remarkable rate of 99.9974%.

Sumathi and Karthikeyan (2021) put up a novel approach for detecting distributed denial of service attacks through the utilisation of deep learning neural networks. This study assesses the network performance by employing a deep learning neural network classifier alongside a cost reduction technique. The evaluation is conducted on a publicly accessible dataset. The methodology employed in this study incorporates the utilisation of the KDD Cup, DARPA 1999, DARPA 2000, and CONFICKER databases. Performance analysis involves the evaluation of many performance measures, including but not limited to detection accuracy, cost per sample, average latency, packet loss, overhead, packet delivery ratio, and throughput. The simulation results indicate that the DNN Cost minimization algorithm yields superior outcomes in terms of achieving a high detection accuracy of 99% while also reducing false positives. Additionally, this algorithm demonstrates high average delay, minimal packet loss, reduced overhead, high packet delivery ratio, and increased throughput compared to existing algorithms. Pasumponpandia and Smys (2019) introduced a machine learning-based approach for detecting DDoS attacks in telecommunication networks. The telecommunication network, which consists of terminal nodes, facilitates connectivity across the entire system. The rapid advancements in communications networks and information technology have facilitated a seamless connectivity and the ability to store and transmit large amounts of sensitive information in the form of text and speech. Telecommunication networks are susceptible to many cyber-threats, with distributed denial of service (DDoS) attacks being the most prevalent. These cyber-threats result in the denial of services to users. The present study employs a hybrid approach, integrating neural network and support vector machine techniques, to propose a method for detecting and classifying distributed denial-of-service (DDoS) assaults within telecommunication networks. The utilisation of the network simulator-2 for performance evaluation allows for improved detection accuracy in relation to the proposed method.

Muhammad and Syed (2019) put out a methodology that utilises clustering as a basis for semi-supervised machine learning in the classification of DDoS attacks. This research presents a clustering-based methodology for differentiating network traffic data that encompasses both regular and Distributed Denial of Service (DDoS) activity. The attributes are utilised for the purpose of identifying attacks from the perspective of the victim. The study showcases three specific features that can be observed on the targeted system. The clustering techniques encompass agglomerative clustering and K-means clustering, both of which are employed in conjunction with feature extraction using Principal Component Analysis (PCA). A proposed voting mechanism is utilised to assign labels to the data and derive distinct classifications for distinguishing attacks from typical traffic. Following the process of labelling, the application of supervised machine learning methods such as k-Nearest Neighbours (kNN), Support Vector Machine (SVM), and Random Forest (RF) is carried out to acquire the trained models that will be utilised for subsequent classification tasks. The experimental results demonstrate that the k-nearest neighbours (kNN), support vector machine (SVM), and random forest (RF) models achieve accuracy scores of 95%, 92%, and 96.66% correspondingly when optimised parameter tuning is applied within the specified value ranges. Ultimately, the approach is further substantiated through the utilisation of a portion of benchmark dataset containing novel attack vectors.

Francisco et al. (2019) put out a proposal titled "Smart Detection: An Online Approach for DoS/DDoS Attack Detection Using Machine Learning." This article introduces a method for detecting Denial-of-Service (DoS) attacks using machine learning (ML) techniques. The suggested methodology utilises inferences derived from previously collected signatures obtained from samples of network traffic. The studies were conducted utilising four contemporary benchmark datasets. The findings indicate that the online detection rate (DR) of attacks exceeds 96%, demonstrating a high level of accuracy (PREC) and a low false alarm rate (FAR) when employing a sample rate (SR) equivalent to 20% of the network traffic.

DESIGN METHODOLOGY

This section describes the various components of the proposed system architecture. It also shows how the various components are connected in building a robust framework for the detection of DDoS attacks on network logs.

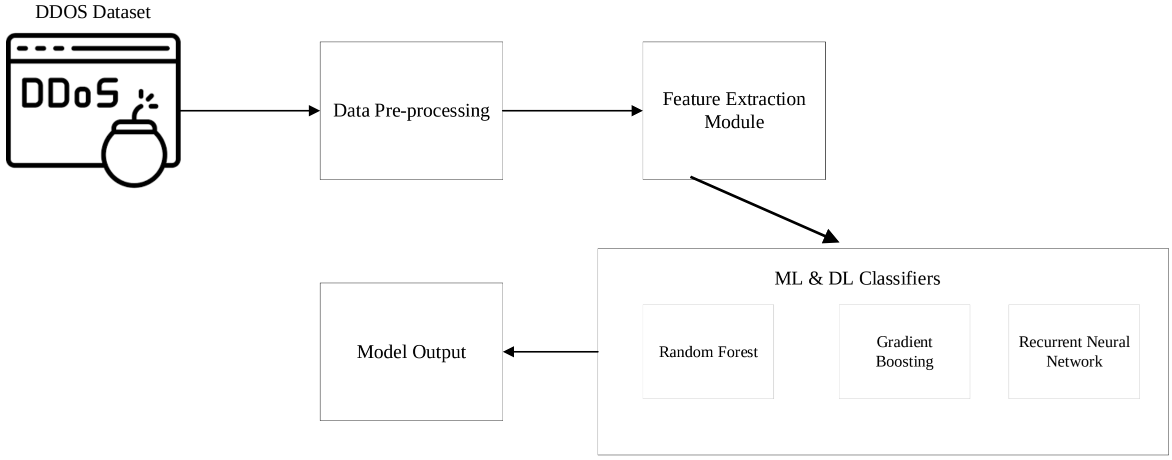
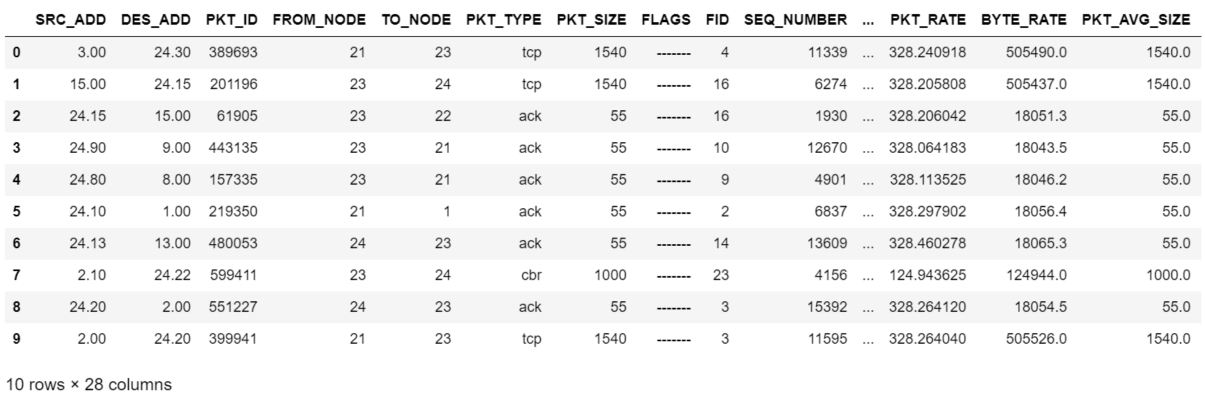


Figure 1: Architectural Design

DDOS Dataset: This dataset contains labelled network logs from various types of DDOS attacks in the ARFF file format. Data is labelled by attributes that describe each component of the data. The dataset contains 29 attributes.



	SRC_ADD	DES_ADD	PKT_ID	FROM_NODE	TO_NODE	PKT_TYPE	PKT_SIZE	FLAGS	FID	SEQ_NUMBER	...	PKT_RATE	BYTE_RATE	PKT_AVG_SIZE
0	3.00	24.30	389693	21	23	tcp	1540	-------	4	11339	...	328.240918	505490.0	1540.0
1	15.00	24.15	201196	23	24	tcp	1540	-------	16	6274	...	328.205808	505437.0	1540.0
2	24.15	15.00	61905	23	22	ack	55	-------	16	1930	...	328.206042	18051.3	55.0
3	24.90	9.00	443135	23	21	ack	55	-------	10	12670	...	328.064183	18043.5	55.0
4	24.80	8.00	157335	23	21	ack	55	-------	9	4901	...	328.113525	18046.2	55.0
5	24.10	1.00	219350	21	1	ack	55	-------	2	6837	...	328.297902	18056.4	55.0
6	24.13	13.00	480053	24	23	ack	55	-------	14	13609	...	328.460278	18065.3	55.0
7	2.10	24.22	599411	23	24	cbr	1000	-------	23	4156	...	124.943625	124944.0	1000.0
8	24.20	2.00	551227	24	23	ack	55	-------	3	15392	...	328.264120	18054.5	55.0
9	2.00	24.20	399941	21	23	tcp	1540	-------	3	11595	...	328.264040	505526.0	1540.0

10 rows × 28 columns

Figure 2: Dataset Sample

Data Pre-Processing: we applied Standard scaler and data cleaning data cleaning process in this module. Data cleaning involves identifying and handling missing values, outliers, and inconsistencies in a dataset to ensure data quality and reliability.

Feature Extraction: we applied random forest classifier in selecting important features. The steps for the random forest classifier can be seen as follows:

1. For each tree b in the Random Forest:
   i. Calculate the Gini impurity (or another chosen impurity measure) of the dataset at the root node before the split, denoted as Gini_root.
   ii. Calculate the weighted average Gini impurity of the child nodes after the split, denoted as Gini_children. The weight is the number of data points in each child node divided by the total number of data points in the root node.
2. For each tree b, calculate the decrease in Gini impurity (or other chosen impurity measure) due to splitting on feature X_i:
   i. Decrease in Gini (X_i) = Gini_root - Gini_children
3. Calculate the average decrease in Gini impurity (or impurity measure) for feature X_i over all trees in the Random Forest:
   i. Average Decrease in Gini (X_i) = (1 / B) * Σ Decrease in Gini (X_i) for all trees b
4. Normalize the feature importance values so that they sum up to 1 or 100 (depending on your preference).

ML & DL Classifiers: Here, we applied two ML models and one DL model for the detection of Distributed Denial of Service (DDOS) attacks on network logs. The three classifiers were both applied on the DDOS dataset. The classifiers were evaluated using classification report and confusion matrix.

Model Output: This displays the various types of DDOS attacks on network logs. The output of the model shows the classified result of the model which comprise of the different types of DDoS attacks on network logs.

4. Experimental Setup

We set up an experiment on googlecolab. The experimental phase has to with the analysis phase and the implementation of two Machine Learning (ML) models and a Deep Learning (DL) model for the detection of Distributed Denial of Service Attacks

(DDOS) on network system. The ML models are Random Forest Classifier and Gradient Bosten Classifier. Recurrent Neural Network was used for the Deep Learning Model.

Data Analysis Phase

We made use of pandas, seaborn, and matplotlib library in conducting analysis on the DDOS dataset. The analysis was to enable us get a proper insight on the dataset before training the DL models for the detection DDOS attacks on network system. First, we check if the dataset contains some nan and duplicate values. We used pandas in achieving this. Second, we plotted a bar chart to check if the number of classes (different types of the DDOS attacks) have the same number of instances. The bar chat in Figure 3 shows that the number of instances of each of the different types of DDOS attacks. From the bar chart, we can see that the number of instances of the different classes of the DDOS attacks are not the same. That simply make the dataset imbalance, this simply means that if the data imbalance is not solved, the ML and Dl classifier will produce high rate of false positive and negative. To solve the data imbalance problem, we perform random under sampling using an under-sampling technique called RandomUnderSampler. We used this to down sample the dataset, making all the classes have equal number of instances. The down sampled data can be seen in the bar chart in Figure 4.

Finally, we extracted important features from the dataset by using the Random forest Classifier (RF). The RF classifier was used in ranking the features of the dataset. Table 1 shows the extracted features (The most important features), and Figure 5 shows the visualized plot of the important features.

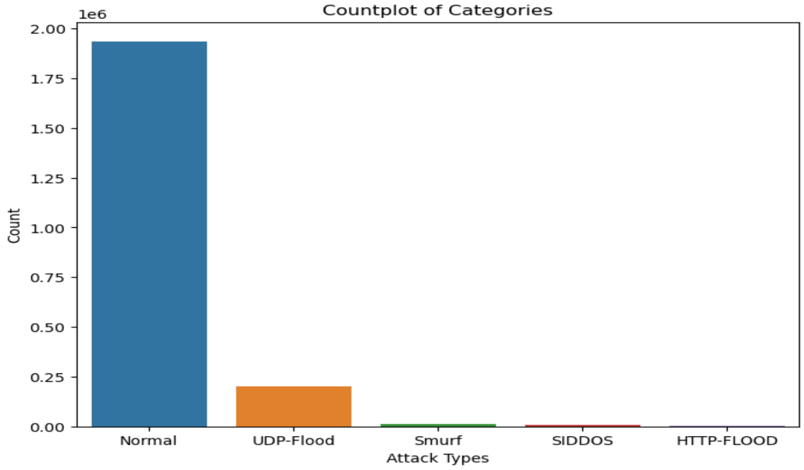


Figure 3: Bar chart of the imbalanced classes.

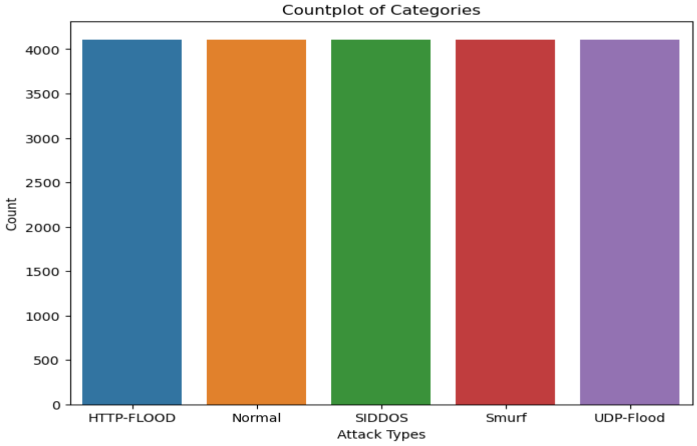


Figure 4: Bar chat of the balanced classes.

Scientia Africana, Vol. 22 (No. 3), December, 2023. Pp 153-164
© Faculty of Science, University of Port Harcourt, Printed in Nigeria *ISSN 1118 – 1931*

Table 1: Extracted Features

S/N	Feature	Important_Features
0	PKT_RATE	0.104397
1	LAST_PKT_RESEVED	0.102332
2	BYTE_RATE	0.08825
3	SEQ_NUMBER	0.063824
4	UTILIZATION	0.06279
5	NUMBER_OF_BYTE	0.052854
6	PKT_AVG_SIZE	0.047177
7	PKT_ID	0.042966
8	DES_ADD	0.039965
9	FID	0.039018

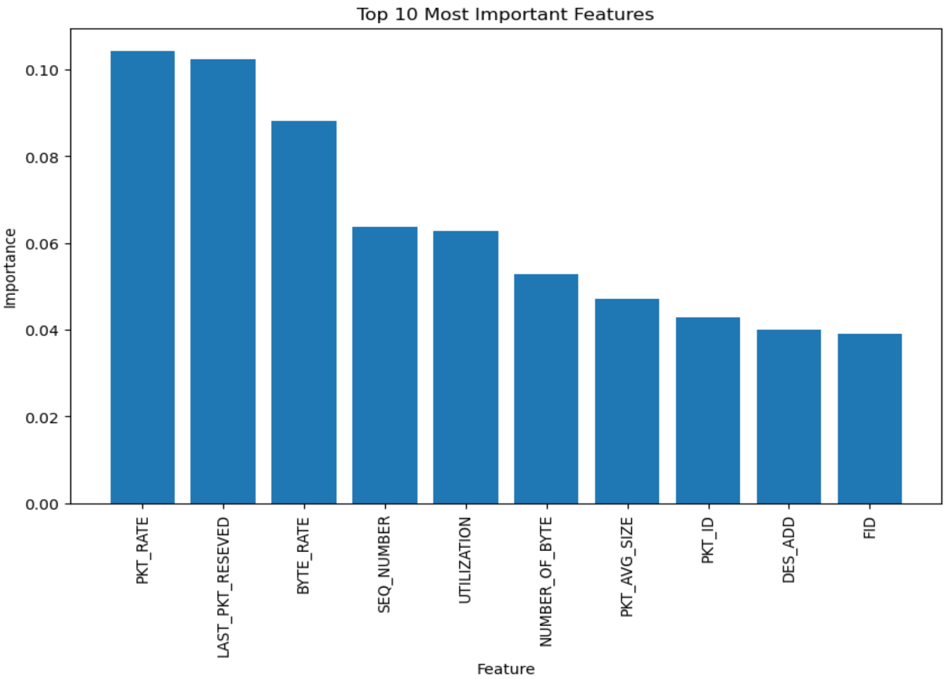
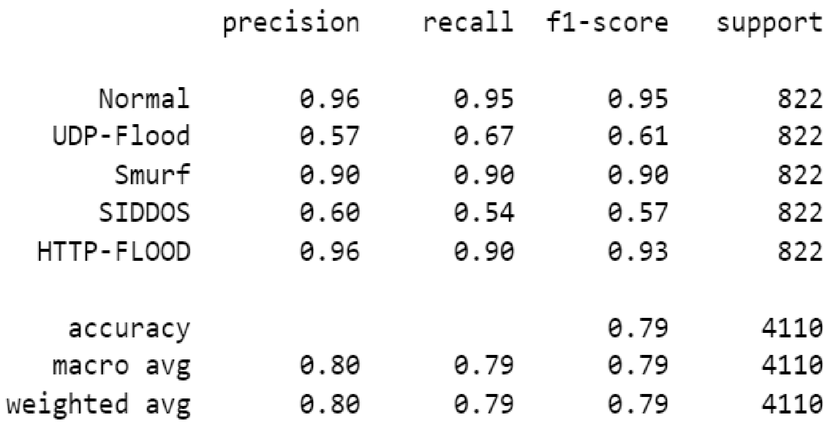


Figure 5: Bar Chart of Top 10 Important Features

Model Training with Random Forest (RF)

For the detection of DDOS attacks, we trained a RF model on the DDOS dataset. The RF model was trained on the DDOS dataset by fine tuning the hyper parameters. The number of estimators of the RF model was fine tuned to 1000. The RF model was used to make prediction in an unseen dataset to detect the various types of the DDOS attacks. The result of the RF model was evaluated using matrix evaluation (Classification matrix, and Confusion matrix). The result of the RF model can be seen in Figure 6 and Figure 7.



```
              precision    recall  f1-score   support

      Normal       0.96      0.95      0.95       822
   UDP-Flood       0.57      0.67      0.61       822
       Smurf       0.90      0.90      0.90       822
      SIDDOS       0.60      0.54      0.57       822
  HTTP-FLOOD       0.96      0.90      0.93       822

    accuracy                           0.79      4110
   macro avg       0.80      0.79      0.79      4110
weighted avg       0.80      0.79      0.79      4110
```

Figure 6: Classification Report of the RF Model.

The RF model achieved an accuracy of 79% on the test data (unseen) data.

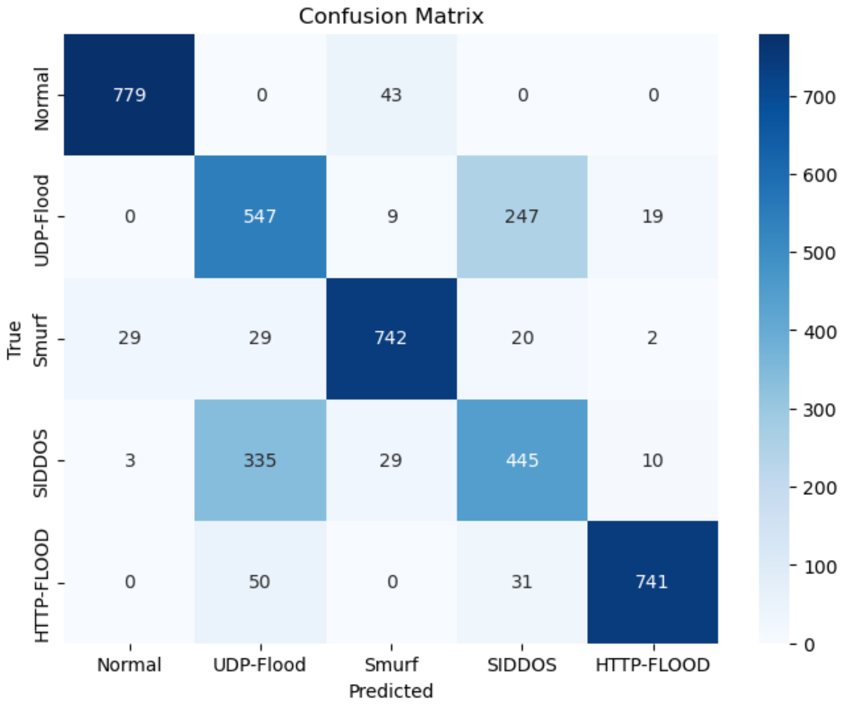
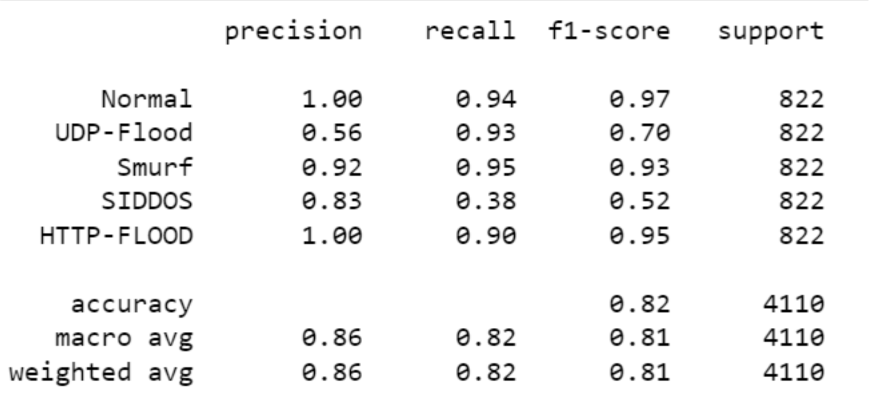


Figure 7: Confusion Matrix of the RF Model.

Model Training with Gradient Boost Classifier (GBC)

The GBC model on the DDOS dataset. We fine-tuned the hyper parameters. The fine-tuned parameters are n_estimators=100, learning_rate=0.1, max_depth=3, random_state=42. The GBC model was used to make prediction in an unseen dataset to detect the various types of the DDOS attacks. The result of the GBC model was also evaluated using matrix evaluation (Classification matrix, and Confusion matrix). The result of the GBC model can be seen in Figure 8 and Figure 9.



```
              precision    recall  f1-score   support

      Normal       1.00      0.94      0.97       822
   UDP-Flood       0.56      0.93      0.70       822
       Smurf       0.92      0.95      0.93       822
      SIDDOS       0.83      0.38      0.52       822
  HTTP-FLOOD       1.00      0.90      0.95       822

    accuracy                           0.82      4110
   macro avg       0.86      0.82      0.81      4110
weighted avg       0.86      0.82      0.81      4110
```

Figure 8: Classification Report GBC model.

The GBC model had an accuracy of 82% on the test data.

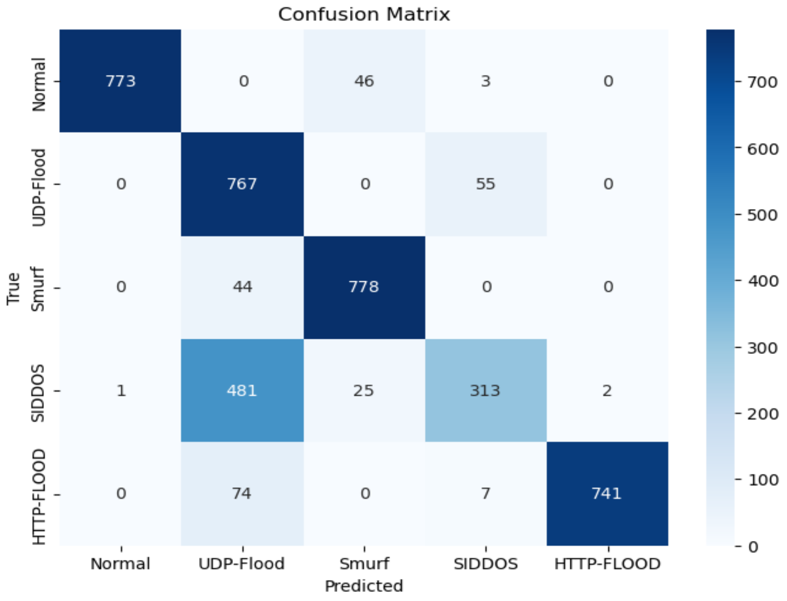


Figure 9: Confusion Matrix of the GBC Model.

Model Training with Recurrent Neural Network (RNN)

Finally, we trained an RNN model for the detection of DDOS attacks on network system. The RNN model was trained by fine tuning it's hyper parameters. The fine tuned parameters. The RNN model has three layers, one input layer with input neuron of 256, a hidden layer with an input neural of 256, and finally the output layer with dense layer 5. The hyper parameters used here are relu and softmax for activation functions, optimizer = 'adam', and loss ='categorical_consentropy', batch_size=64, and epoch =7. The result of the RNN model for both training and evaluation can be seen in Table 2, 10, 11.

Table 2: Training Steps of RNN model

```
Epoch 1/5
100000/100000 [==============================] - 32s 318us/step - loss: 0.1963 - accuracy: 0.9308 -
val_loss: 0.1937 - val_accuracy: 0.9334
Epoch 2/5
100000/100000 [==============================] - 28s 281us/step - loss: 0.0757 - accuracy: 0.9742 -
val_loss: 0.1265 - val_accuracy: 0.9609
Epoch 3/5
100000/100000 [==============================] - 27s 271us/step - loss: 0.0519 - accuracy: 0.9828 -
val_loss: 0.1504 - val_accuracy: 0.9549
Epoch 4/5
100000/100000 [==============================] - 28s 281us/step - loss: 0.0395 - accuracy: 0.9869 -
val_loss: 0.1420 - val_accuracy: 0.9616
Epoch 5/5
100000/100000 [==============================] - 27s 271us/step - loss: 0.0323 - accuracy: 0.9897 -
val_loss: 0.1616 - val_accuracy: 0.9567
```

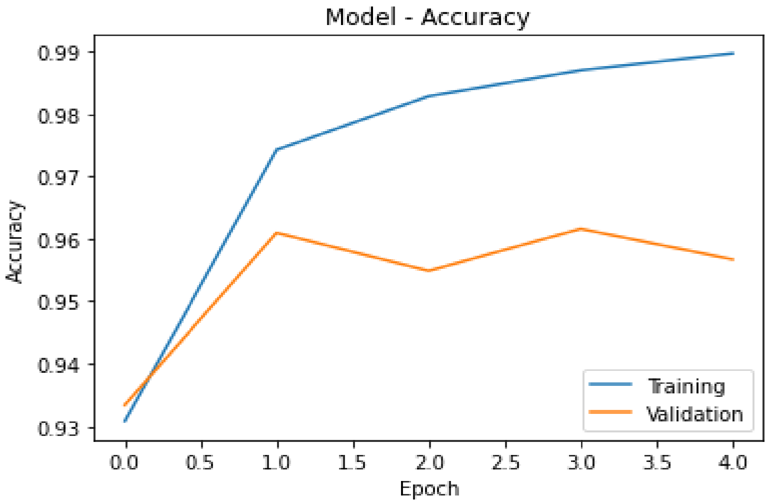


Figure 10: Training Accuracy For Both Training and Validation.

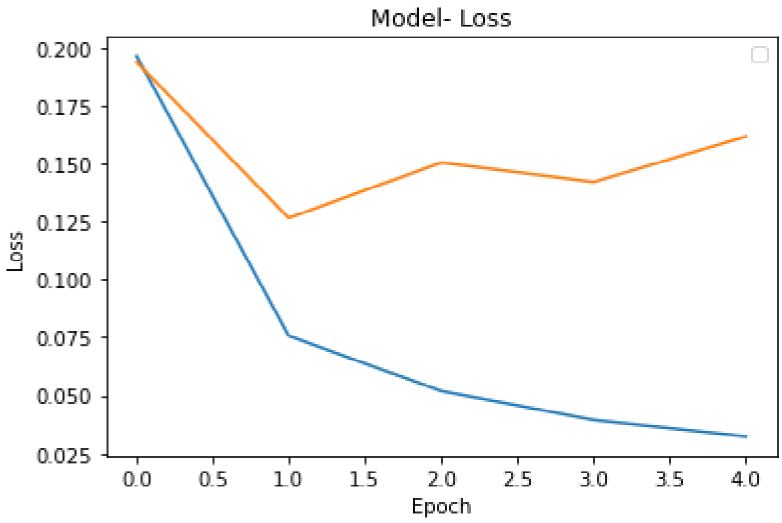


Figure 11: Loss values for training and Validation.

Scientia Africana, Vol. 22 (No. 3), December, 2023. Pp 153-164
© Faculty of Science, University of Port Harcourt, Printed in Nigeria *ISSN 1118 – 1931*

Table 3 Performance Evaluation

Classifier	Accuracy (%)
RF	79
GBC	82
RNN	99

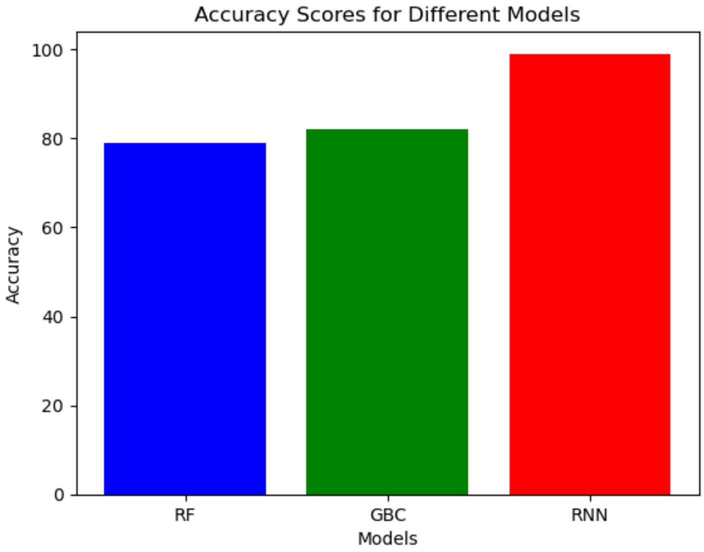


Figure 12: Barchat of Evaluated Parameters.

CONCLUSION

The application of machine learning techniques, including Random Forest, Gradient Boosting, and Recurrent Neural Network, in the detection of Distributed Denial of Service (DDoS) attacks is a promising step forward in the realm of cybersecurity. While this paper acknowledged the challenges associated with machine learning in DDoS mitigation, such as adversarial evasion and the need for high-quality data and continuous model refinement, it also demonstrated the potential effectiveness of these models in addressing the persistent threat of DDoS attacks. One notable aspect of this research was the recognition of the imbalanced nature of DDoS datasets and the proactive approach of employing random under-sampling to mitigate this issue. The results of the models, with a noteworthy accuracy of 99.47% for the recurrent neural network, indicate that these machine learning methods hold great promise for accurately detecting various types of DDoS attacks in real-world scenarios.

However, it is crucial to recognize that cybersecurity is an ever-evolving field, and attackers are continually developing new techniques to bypass defenses. Therefore, security teams must remain vigilant in adapting and fine-tuning machine learning models to address emerging threats effectively. Nevertheless, this research serves as a testament to the potential of machine learning in bolstering our defenses against DDoS attacks and highlights the importance of ongoing research and development in this area to ensure the security of our interconnected world.

REFERENCES

Singh, R., Tanwar, S., & Sharma, T. P. (2020). Utilization of blockchain for mitigating the distributed denial of service attacks. *Security and Privacy*, *3*(3), e96.

Aladaileh, M. A., Anbar, M., Hasbullah, I. H., Chong, Y. W., &Sanjalawe, Y. K. (2020). Detection techniques of distributed denial of service attacks on software-defined networking controller–a review. *IEEE Access*, *8*, 143985-143995.

Sharafaldin, I., Lashkari, A. H., Hakak, S., &Ghorbani, A. A. (2019, October). Developing realistic distributed denial of service (DDoS) attack dataset and taxonomy. In *2019 International Carnahan Conference on Security Technology (ICCST)* (pp. 1-8). IEEE.

Garcia, J. F. C., & Blandon, G. E. T. (2022). A deep learning-based intrusion detection and preventation system for detecting and preventing denial-of-service attacks. *IEEE Access*, *10*, 83043-83060.

Kumari, K., Mrunalini, M. Detecting Denial of Service attacks using machine learning algorithms. *J Big Data* **9**, 56 (2022). https://doi.org/10.1186/s40537-022-00616-0.

Ullah, S., Mahmood, Z., Ali, N., Ahmad, T., &Buriro, A. (2023). Machine Learning-Based Dynamic Attribute Selection Technique for DDoS Attack Classification in IoT Networks. *Computers*, *12*(6), 115.

Alahmadi, A. A., Aljabri, M., Alhaidari, F., Alharthi, D. J., Rayani, G. E., Marghalani, L. A., ...&Bajandouh, S. A. (2023). DDoS Attack Detection in IoT-Based Networks Using Machine Learning Models: A Survey and Research Directions. *Electronics*, *12*(14), 3103.

Ebtihal S. A., &, QnytraA. (2021). Detecting Distributed Denial of Service Attacks using Machine Learning Models Department of Information Technology University of Tabuk, KSA.

S. Sumathi & N. Karthikeyan (2021). Detection of distributed denial of service using deep learning neural network. Journal of Ambient Intelligence and Humanized Computing volume 12, pages5943–5953 (2021) Cite this article 718 Accesses 18 Citations Metrics.

Smys, S. (2019). DDOS attack detection in telecommunication network using machine learning. *Journal of Ubiquitous Computing and Communication Technologies (UCCT)*, *1*(01), 33-44.

Muhammad Aamir & Syed Mustafa Ali Zaidi (2019) Clustering based semi-supervised machine learning for DDoS attack classification. Journal of King Saud University - Computer and Information Sciences Volume 33, Issue 4, May 2021, Pages 436-446.

Lima Filho, F. S. D., Silveira, F. A., de Medeiros Brito Junior, A., Vargas-Solar, G., & Silveira, L. F. (2019). Smart detection: an online approach for DoS/DDoS attack detection using machine learning. *Security and Communication Networks, 2019*, 1-15.