Detecting Distributed Denial of Service Attacks using Machine Learning Models

Ebtihal Sameer Alghoson, Onytra Abbass
Department of Information Technology
University of Tabuk, KSA

***Abstract*—The Software Defined Networking (SDN) is a vital technology which includes decoupling the control and data planes in the network. The advantages of the separation of the control and data planes including: a dynamic, manageable, flexible, and powerful platform. In addition, a centralized network platform offers situations that challenge security, for instance the Distributed Denial of Service (DDoS) attack on the centralized controller. DDoS attack is a well-known malicious attack attempts to disrupt the normal traffic of targeted server, network, or service, by overwhelming the target's infrastructure with a flood of Internet traffic. This paper involves investigating several machine learning models and employ them with the DDoS detection system. This paper investigates the issue of enhancing the DDoS attacks detection accuracy using a well-known DDoS named as CICDDoS2019 dataset. In addition, the DDoS dataset has been preprocessed using two main approaches to obtain the most relevant features. Four different machine learning models have been selected to work with the DDoS dataset. According to the results obtained from real experiments, the Random Forest machine learning model offered the best detection accuracy with (99.9974%), with an enhancement over the recent developed DDoS detection systems.**

Keywords*—Cybersecurity; distributed denial of service (DDoS); machine learning (ML); Canadian institute cybersecurity - distributed denial of service (CICDDoS2019) dataset

I. Introduction

SDN stands for Software Defined Network Technology, a new technology in the network world, in which the network management and control function is separated from the data routing function, through which engineers attempt to rearrange the parts and roles of all network infrastructure components that have not been modified since the 1980s. It is the transition from NCP to TCP / IP and since then no change has occurred. A change in the level of the network infrastructure to keep pace with the great development that takes place in the field of information technology, especially in virtual computing, which made virtualization of all layers, and the infrastructure is still intractable to this technology, so SDN technology is a successful attempt to separate the data layer from the control layer [1].

Denial of Service (DOS) It is one of the types of electronic attacks, and it is a very powerful technology that has been launched to attack network devices and services, and this type can separate different services from the Internet. Distributed Denial of Service (DDOS) is a more powerful type of DOS and uses multiple distributed attack points [2].

DoS was originally appeared by Gligor in an operating system context [3, 4], where DoS became widely employed. In general, DoS attack tries to reach more than one computer to reach a victim in a coordinated manner is called a DDoS attack.

Software Defined Network (SDN) infrastructure is vulnerable to several security threats. Among the DDoS attacks are the most dominant one. The DDoS attacks are considered as one of the most destructive attacks in the Internet. In general, most website hacking are probably a DDoS attack. The DDoS attack aims to disrupting the normal operation of the system through making services and resources unavailable to legitimate users by overloading the system with unnecessary superfluous traffic from distributed source. In addition, DDoS attack aims to increase in strength and frequency day-by-day. Therefore, the new systems which have been developed should be able to enhance the performance requirements and improve scalability of modern data centers, and provide maximum protection against the DDoS attacks.

This paper aims to mitigate denial of service attacks in software-defined networks through developing an efficient DDoS detection system based on machine learning models. The main contributions of this paper includes the following:

1) Research and analyze the recent developed DDoS detection systems.

2) Adopt several feature selection methods before processing the training stage.

3) Employ various machine learning models in the training process in order to enhance the efficiency of the DDoS detection system.

4) Test the developed machine learning model using real datasets, and real experiments, in order to assess the efficiency of the DDoS detection system.

II. Related Work

This section discusses the recent developed DDoS detection systems employed using the CICDDoS2019 dataset. Authors of [5] proposed a hybrid machine learning-based system to detect DDoS attacks. The proposed system involves combining the Extreme Learning Machine (ELM) algorithm and the black-hole optimization algorithm. Authors conducted several experiments through adopting various datasets to assess the performance of the proposed hybrid machine learning system. The proposed hybrid system has been employed in

detecting the DDoS attacks in cloud computing, and achieves 99.80% detection accuracy using the CICDDoS2019 dataset.

On the other hand, authors of [6] proposed an Intrusion Detection System against DDoS attacks (DDoSNet) in SDN environments. The proposed system is based on the Deep Learning (DL) technique, integrating the Recurrent Neural Network (RNN) with autoencoders. The developed system has been evaluated using the CICDDoS2019 dataset. Authors obtained a significant enhancement in attack detection compared to the existing methods. Therefore, the proposed system offers great confidence in securing SDN environments.

The work presented in [7] includes examining the impact of data balancing algorithm in the network traffic classification problem on several types of DDoS attacks using the CICDDoS2019 dataset, which consists of various information about the reflection-based and exploitation-based attacks. The obtained results showed that the effectiveness of data balancing algorithms such as synthetic minority sampling, naïve random, and adaptive synthetic sampling in classifying network attacks.

Authors of [8] proposed a detection system which was able to detect the different types of DDoS attacks based on several classification algorithms using the CICDDoS 2019 dataset. In addition, authors captured packets from SDK environment, apply preprocessing function for the dataset, and then apply classification algorithm to detect the DDoS attacks. Authors revealed that the decision tree offers the better performance compared to SVM and Naïve Bayes machine learning models.

The work presented in [9] involves analyzing the success rate in the intrusion detection system through adopting several machine learning methods. The CICDDoS2019 dataset was employed, where several machine learning models were investigated, including: the ANN, Support Vector Machine (SVM), Gaussian Naïve Bayes, Multinomial Naïve Bayes, Bernouli aïve Bayes, Logistic Regression, K-nearest neighbor (KNN), Decision Tree, and Random Forest algorithms. Authors showed that the K-nearest neighbor, logistic regression, and Naïve Bayes offers the best prediction accuracy.

Authors of [10] employed the Deep Neural Network (DNN) as a deep learning method to detect the DDoS attacks on the sample of packets captured from network traffic. The DNN model can work rapidly and with high detection accuracy even with small samples, since it contains feature extraction and classification methods. Authors preformed their experiments using the CICDDoS2019 dataset which contains several DDoS attack types created in 2019. The proposed system achieves 94.57% accuracy rate using the deep learning model.

The work presented in [11] surveys the recent developed DDoS detection approaches using the machine learning models. Authors of [12, 13, 14] proposed a DDoS detection system using Naïve Bayes model. On the other hand, the support vector machine model has been adopted in this works [15, 16, 17] to detect the present of DDoS attacks. In addition, Decision Tree algorithm has also been adopted to detect the DDoS attacks, as presented in [18, 19].

TABLE I. A COMPARISION BETWEEN THE EXISTING SYSTEMS THAT EMPLOYED THE CICDDOS2019 DATASET

Research work	Algorithm	Detection Accuracy
[5]	Extreme Learning Machine & blackhole algorithms	99.80%
[6]	Recurrent Neural Network with autoencoders	92.54%
[7]	SMOTE	93.51%
[8]	Decision Tree	92.15%
[9]	Artificial Neural Network, Support Vector Machine, Guassian Naïve Bayes, Random Forest Algorithm & K-Nearest Neighbor	Naïve Bayes offers the best detection accuracy
[10]	Deep Neural Network	94.57%

As presented above, several DDoS detection systems have been developed recently based on the employment of the CICDDoS2019 dataset. Table I presents a comparison between the existing developed systems based on the algorithm used and the detection accuracy.

III. SYSTEM DESIGN

The Distributed Denial of Service (DDoS) attacks include transmitting multiple requests to the attacked web resource, with the goal of exceeding the website's capacity to handle multiple requests, and hence prevent the website from functioning correctly. Several researchers have discussed the DDoS attacks and analyzed the major security threats and the corresponding solutions. This section discusses the main methods which have been employed in order to develop the DDoS system. In addition, this section presents the experimental setup including: the development environment, the selected DoS datasets, and the experimental setup.

A. System Methodology

Fig. 1 shows the development process of the DDoS detection system. As presented below, the first stage includes searching an efficient DDoS dataset, that are being developed recently by several research works. The second stage involves cleaning up the dataset and apply feature extracting methods, in order to pick the most significant features. Next, several machine learning models will be implemented to test the performance of the developed machine learning system, and then obtain the model's accuracy after conducting the training and testing processes.

B. DDoS Dataset

For each single machine learning model, a training and testing processes are needed to be implemented in order to assess the performance of the developed machine learning model. An extensive research has been carried out in order to identify the best DDoS datasets, which will be employed later in the training and testing process. Several datasets are available online such as the CIC-DDoS2019 dataset, where it contains benign and the most up-to-date common DDoS attacks, which resembles the true real-world data.

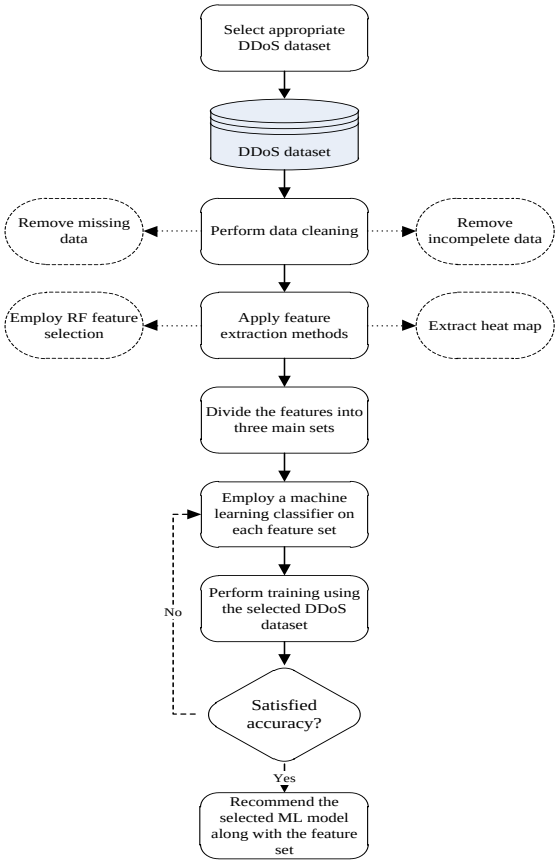

Fig. 1. The Development Process for the DDoS System.

Authors of [20] generated the CICDDoS2019 dataset that remedies several shortcomings and limitation which are presented in the existing datasets. CICDDoS2019 is labeled with 80 network traffic features that were extracted and calculated for all benign and denial of service flows. CIC-DDoS2019 dataset contains the results of the network traffic analysis with labeled flows based on the time stamp, source and destination IPs, source and destination ports, protocols and attack.

Therefore, the CICDDoS2019 dataset will be divided into two subsets: the training subset, and the testing subset. The training subset contains samples of data used to fit the machine learning models, whereas the testing subset is a gold standard employed to assess the performance of the trained machine learning model.

TABLE II. CICDDOS2019 DATASET GENERAL STATISTICS

Parameter name	**Total #**
Total number of records	12,794,627
Total number of features	82
Total number of labels	1
Total number of normal records	6,398,925
Total number of attack records	6,395,702
% of normal records	50.02%
% of attack records	49.98%

Table II shows the general statistics for the CICDDoS2019 dataset. CICDDoS2019 dataset is a large dataset in size and records, it consists of (12,794,627) records with a total memory size (6.3 gigabyte). The CICDDoS2019 is a balanced dataset, where the total number of normal records is (6,398,925) with the percentage of (50.02%), and the total number of fraud records is (6,395,702) with the percentage of (49.89%).

C. Data Preparation

In general, data preparation is considered as the most difficult stage in machine learning, and includes: data cleaning, data pre-processing, data wrangling, and feature engineering. Data preparation involves transforming raw data into a format where the machine learning algorithms can deal with, in order to uncover insights or make predictions. The data preparation process may consist of several steps, however, the most significant one involves processing the missing or incomplete data in the CICDDoS2019 dataset.

Data cleaning includes identifying and correcting errors or mistakes in the CICDDoS2019 dataset. Dropping columns that include missing or incomplete data, since missing and incomplete data affect the efficiency of the machine learning model. Therefore, it is important to process the missing and incomplete data in the dataset. For the CICDDoS2019 dataset, we noticed several attributes (columns) that contain zero values, and this will affect the machine learning model in negative way. For instance, *Fwd Byts/b Avg, Fwd Pkts/b Avg, Fwd Blk Rate Avg, Bwd Byts/b Avg, Bwd Pkts/b Avg, and Bwd Blk Rate Avg* attributes contain zero values in most of the records. Therefore, an important stage is required to remove these attributes from the CICDDoS2019 dataset.

The CICDDoS2019 dataset consists of several categorical data which are unsuitable for machine learning model. Therefore, there is a significant demand to remove these attributes from the CICDDoS2019 dataset in order to be able to train the machine learning model in a proper way. Moreover, the columns (attributes) that contain missing values more than 50% will be dropped from the CICDDoS2019 dataset. In addition, the rows where their columns contain more than 5% missing values are dropped. And finally, the faulty data in the CICDDoS2019 dataset are required to be considered. For instance, all records that contain negative values will be removed from the dataset.

The new shape for the dataset is presented in Fig. 2 after considering several data preparation methods. As noticed, the 19 attributes (columns) have been removed from the dataset, and 48,187 records have been removed from the CICDDoS2019 dataset.

On the other hand, the feature selection methods are considered next. According to [21], there are more than 2.5 quintillion bytes of data is produced every day. However, most of the generated data is required first to be pre-processed before starting any statistically analysis with the selected data, moreover, the produced data needs to be analysed using machine learning techniques in order to provide insights and to create predictions.

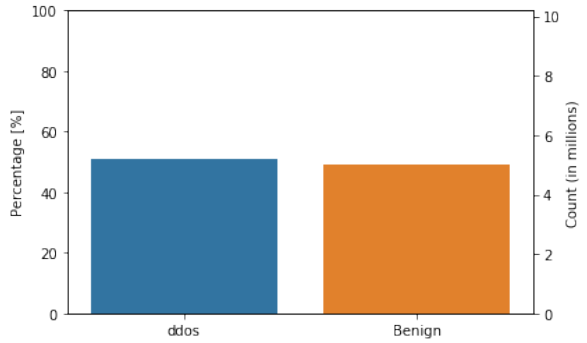

Fig. 2. Distribution of Records in the CICDDoS2019 Dataset.

As presented earlier in Table II, there are 82 features in the CICDDoS2019 dataset, and this makes the training and prediction tasks are very difficult. Therefore, it is important to minimize the number of features in CICDDoS2019 dataset through adopting several feature extraction models. This section discusses several methods which are used in order to extract the most significant features in the CICDDoS2019 dataset.

Minimizing the number of features may lead to several benefits, including: accuracy improvement, speed up in the training process, reducing the overfitting, and improve data visualization. Therefore, there are several different feature selection methods which can be applied to select the most significant feature in a given dataset, some of the most significant methods are: Filter method, and embedded method.

The first feature selection method is the filter method. Filter method involves filtering the dataset and take only a subset containing the most relevant features. This can be done using correlation matrix using Pearson Correlation. In general, the heat map (correlation matrix) is a graphical representation where individual values of matrix are represented as colours in order to display the correlation between attributes in a certain dataset and hence perform better prediction. The heat map for several features are shown below. For instance, Fig. 3 presents the heat map for 16 features, in order to show the relation among them. As seen in below, there is a high correlation between BWD IAT Std feature and FWD IAT Tot feature, and Bwd IAT Tot and FwdIAT Tot.

An embedded method is adopted next in order to enhance the prediction results. Embedded method includes examining the different training iterations of the machine learning model and then ranks the importance of the input features on how much each of the features contributed to the machine learning model through the training process.

For this stage, the Decision Tree model has been selected to rank the importance of CICDDoS2019's features. The Decision Trees models that are based on ensembles, can be used to rank the significance of the input features in the dataset. Since, extruding the most significant features offer vital importance on training the machine learning model, and hence obtaining efficient prediction accuracy. In addition, the features which will not offer any benefits to the machine learning model will be removed from the selected dataset.

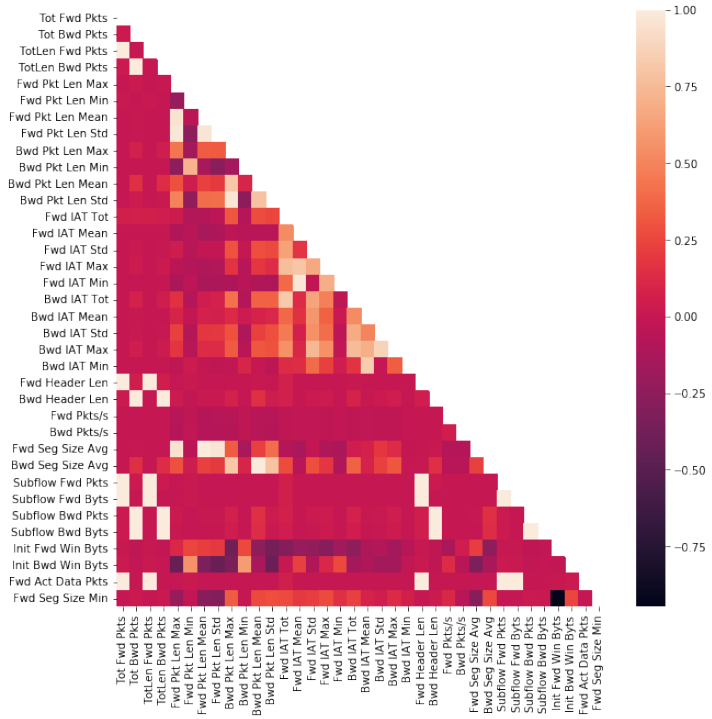

Fig. 3. The Heat Map for 36 Features.

In the Decision Tree, the CICDDoS2019 dataset was divided into two subsets: training subset, and testing subset, with 80% for training and 20% for testing. After completing the training process of the Random Forest Classifier, a set of feature importance plot is established according to the results obtained from the training stage. Fig. 4 shows the most 30 significant features in the CICDDoS2019 dataset.

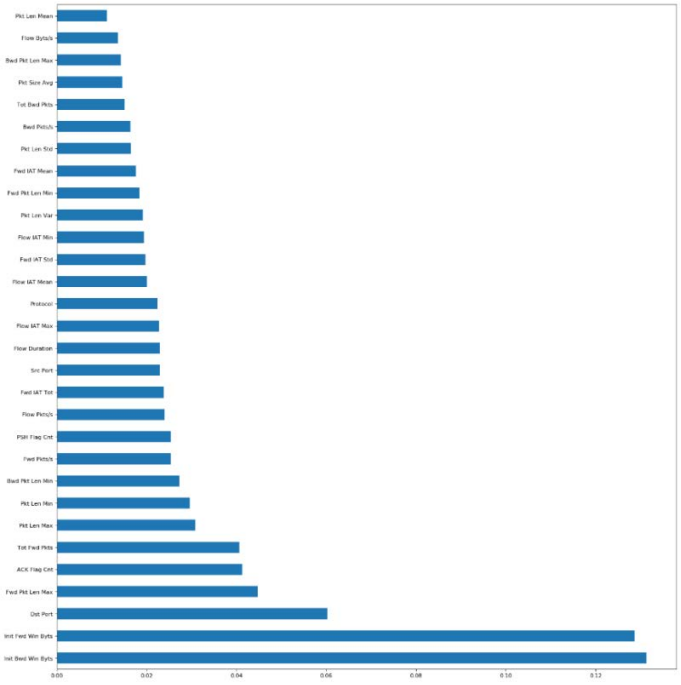

Fig. 4. The Feature Importance Plot for the 30 Most Significant Features.

In this paper, we investigate the efficiency of different features sets (10, 20, and 30), where each set is employed with every machine learning model and then assess and analyze the system's performance. The next section discusses the machine learning models which will be employed for the detection purposes.

D. DDoS Detection Models

In this project, several supervised machine learning models will be investigated, implemented, and tested, including: random forest, Light Gradient Boosting, CatBoost, and Convolutional Neural Networks.

- Random Forest (RF): it is also known as random decision forests that are ensemble learning method for classification and regression. RF operates through constructing multitude of decision trees at the training time, and producing the class which is the mode of the classes (classification) or average prediction (regression) of the individual trees.
- Light Gradient Boosting: is a light, fast, distributed, and high-performance gradient boosting framework, which is based on the decision tree algorithm, used for classification, ranking, and several machine learning tasks. It works by splitting up the tree leaf wise with the best fit, however, other boosting algorithms split the tree depth wise rather than the leaf-wise.
- CatBoost: CatBoost is an algorithm for gradient boosting on decision trees. CatBoost can be easily integrated with deep learning architectures. In addition, it can work with several data types to help solving a wide range of problem. CatBoost provides the best-in-class accuracy.
- Convolutional Neural Network (CNN): CNN is a deep neural networks, and is multilayer perceptron, which means that the CNN network is a fully connected. In any layer, each neuron is connected to all neurons in the next layer. CNN employs a mathematical operation named as convolution, where convolution is a specialized kind of linear operation.

IV. Expreimental Results

This section discusses the results obtained from several experiments conducted to assess the efficiency of different machine learning models. Several experiments have been conducted using the developed environment discussed earlier, in order to assess the DDoS systems' efficiency. Moreover, this section includes analyzing the obtained results and compares the system's efficiency with the recent developed systems.

A. Perforamcne Analysis

Several parameters are considered in order to assess the performance of the implemented DoS detection system; the parameters include:

- Average Training Time: this refers to the total time required to train the machine learning model.
- Accuracy: is the total of transactions that were correctly predicted over the total number of transactions.
- Precision: this indicates the total number of cases that were correctly classified among that class. Precision is the percentage of correctly predicted cases over the total predicted.
- Recall: is the ability of the classifier to correctly find all the positive instances. Recall is the ratio of true positives to the sum (total) of true positives and false negatives.
- Misclassification rate (error rate): this refers to how often the classifier is wrong.

B. Results of Average Training Time

The average training time is estimated for each machine learning model. As shown in Fig. 5, the RF model requires the largest training time (19,078 seconds), this is because the RF builds multiple decision trees and combines them together to obtain more accurate and stable prediction, and this makes the RF is a slow algorithm compares to others. Next, the average training time for the CNN model is (13,785 seconds), since the CNN training time depends on the training subset, batch size, and number of epochs.

On the other hand, the Light GB offers the minimum training time (150 seconds), since the Light GB is considered as a fast machine learning model for three main reasons: First, it splits the data based on their histogram, Second, it is gradient-based one-side sampling, and Third, the Light GB is used to deal with sparse features. Therefore, the Light GB machine learning model is best in terms of training time.

C. Evaluation of Essential ML Metrics

According to [22], there are three main metrics used to assess the machine learning classification model which are: accuracy, precision, and recall. As discussed earlier in the previous section, four different machine learning models were evaluated, where each machine learning model was evaluated through employing three sets of features. The best RF model was with 20-features set which offers (99.99740%) accuracy. On the other hand, the best LGB machine learning model was with the 20-features set which offers (99.99146%). The best accuracy result for the CatBoost model was with the 30-features set with accuracy (99.98592%). And finally, the best accuracy results for the CNN model was with 20-features set. Table III shows the detection accuracy for the best 4 machine learning models.

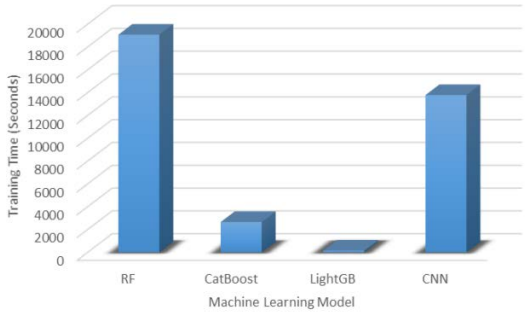

Fig. 5. Average Training Time (in Seconds) for 4 Machine Learning Models.

TABLE III. EVALUATION OF ACCURACY METRIC FOR 4 MACHINE LEARNING MODELS

	Accuracy
RF 20-features set	99.99740%
Light GB 20-features set	99.99146%
CatBoost 30-featues set	99.98592%
CNN 30-features set	98.29388%

TABLE IV. EVALUATION OF PRECISION METRIC FOR 4 MACHINE LEARNING MODELS

	Precision
RF 20-features set	99.99681%
Light GB 20-features set	99.98889%
CatBoost 30-featues set	99.97837%
CNN 20-features set	98.38997%

TABLE V. EVALUATION OF RRECALL METRIC FOR 4 MACHINE LEARNING MODELS

	Recall
RF 30-features set	99.99816%
Light GB 20-features set	99.99430%
CatBoost 30-featues set	99.99391%
CNN 20-features set	99.52368%

TABLE VI. EVALUATION THE FALSE NEGATIVE RATE FOR 4 MACHINE LEARNING MODELS

	False Negative Rate
RF 30-features set	19
Light GB 20-features set	59
CatBoost 30-featues set	63
CNN 30-features set	4,932

The Precision metric is discussed next, where the precision metric was assessed for every machine learning model. Precision refers to how often the machine learning model is able to predict the correct answer. The RF model with 20-feature set offers the best precision result (99.99681%). However, the best precision result using the Light GB model was through adopting 20-feature set with (99.9889%) result, whereas the CatBoost model achieves the best precision result with 30-feature set (99.97837%), and finally, the CNN model offers the best precision result with 20-feature set with (98.3899%) result. Consequently, as presented in Table IV, the machine learning model with best precision result was the Random Forest with 20-feature set.

Finally, the Recall metric is studied in this section. As discussed earlier in the previous section, the RF model with 30-feature set offers the best recall accuracy (99.99816%) among all the RF models (the three trained RF models using different number of features), whereas the Light GB model with 20-features set achieves the best recall accuracy (99.99430%) between all the LGB models. The CatBoost 30-features set offers the best recall result (99.99391%) amongst all the CatBoost models. And finally, the CNN 20-features set offers the best recall results among all the trained models with various number of features. Table V shows the recall results for 4 different machine learning models, and presents that the RF model with 30-features set offers the best recall results (99.99816%).

D. Results of False Negative Rates

This section evaluates the False Negative Rate (FNR) for each machine learning model employed above. FNR is a significant factor and refers to incorrectly predict the absence of DDoS attack when it is actually present, and this is the most significant metric in DDoS attack detection systems. Therefore, it is important to deal with machine learning model with the minimum FNR.

In this section, 14-different experiments were conducted to assess the efficiency of various machine learning models using 3 different DDoS subsets. Table VI presents the best FNR for 4 machine learning models. As presented in the Table below, the Random Forest classifier with 30-features set offers the best FNR with only 19 DDoS records which were misclassified and predicted as normal DDoS packets, whereas a large difference arises when adopting the CNN model. Fig. 4 depicts the false negative records for each machine learning model.

V. DISCUSSION

This section discusses the results obtained in this report with the results obtained from the previous research works, considering the CICDDoS2019 dataset. Most of the existing works evaluated the efficiency of the DDoS prediction model using the accuracy metric. Therefore, this section compares the accuracy metric obtained in this paper, with the existing works developed recently.

In this work, the detection accuracy that was achieved equal to 99.99740% using the random forest machine learning model with 20-features set. The high detection accuracy refers to the pre-processing methods which have been employed on the CICDDoS2019 dataset before applying the machine learning model. Two different feature selection methods were employed in this paper: filter method and feature extraction methods, in order to extract the most important feature which affect the machine learning model.

Therefore, in this paper, as shown in the previous section, the obtained detection accuracy results are greater than the results obtained from the recent developed works. Table VII presents the overall detection accuracy for various machine learning models used to detect the DDoS attacks. Fig. 6 shows the detection accuracy for several machine learning models, with different classification accuracy.

TABLE VII. DETECTION ACCURACY FOR SEVERAL MACHINE LEARNING MODELS

	Detection Accuracy
[5]	99.80
[6]	92.54
[7]	93.51
[8]	92.15
[10]	94.57
This system	99.99

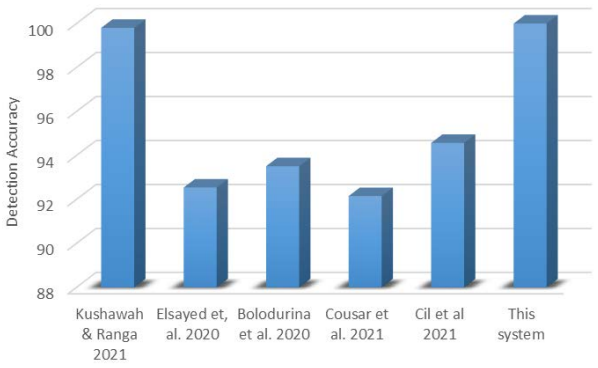

Fig. 6. The Detection Accuracy for Several Machine Learning Models.

VI. Conclusion and Future Work

Recently, the DDoS attack is considered as one of the most significant attack, which is a very powerful technology that has been launched to attack network devices and services. Therefore, in this paper, we consider the DDoS attack to be studied, analysed, and develop a machine learning model to detect such attacks. In this paper, we employed several feature selection methods in order to select the most significant features that can be used to predict the DDoS attacks in an efficient way. Three sets of features have been chosen from the selected dataset, and employed with four machine learning models. According to the obtained results, the RF-machine learning model with 20-features set offers the best precision, accuracy, recall, and false negative rate. For future work, we aim to work with real-time DDoS detection systems which will be able to detect the DDoS attack in real-time situations. Therefore, in this paper, we offered significant improvement in the detection of DDoS attacks using the CICDDoS2019 dataset.

References

[1] Faujdar, N., Sinha, A., Sharma, H., & Verma, E. (2020, October). Network Security in Software defined Networks (SDN). In *2020 International Conference on Smart Technologies in Computing, Electrical and Electronics (ICSTCEE)* (pp. 377-380). IEEE.

[2] Iqbal, M., Iqbal, F., Mohsin, F., Rizwan, M., & Ahmad, F. (2019). Security Issues in Software Defined Networking (SDN): Risks, Challenges and Potential Solutions. *International Journal of Advanced Computer Science and Applications*, *10*(10).

[3] K. Lakshminarayanan, D. Adkins, A. Perrig, and I. Stoica, "Taming ip packet flooding attacks," ACM SIGCOMM Computer Communication Review, vol. 34, no. 1, pp. 45–50, 2004.

[4] V. D. Gligor, "A note on denial-of-service in operating systems," IEEE Transactions on Software Engineering, no. 3, pp. 320–324, 1984.

[5] Kushwah, G.S. and Ranga, V., 2021. Optimized extreme learning machine for detecting DDoS attacks in cloud computing. *Computers & Security*, *105*, p.102260.

[6] Elsayed, M.S., Le-Khac, N.A., Dev, S. and Jurcut, A.D., 2020, August. Ddosnet: A deep-learning model for detecting network attacks. In *2020 IEEE 21st International Symposium on" A World of Wireless, Mobile and Multimedia Networks"(WoWMoM)* (pp. 391-396). IEEE.

[7] Bolodurina, I., Shukhman, A., Parfenov, D., Zhigalov, A. and Zabrodina, L., 2020, November. Investigation of the problem of classifying unbalanced datasets in identifying distributed denial of service attacks. In *Journal of Physics: Conference Series* (Vol. 1679, No. 4, p. 042020). IOP Publishing.

[8] Kousar, H., Mulla, M.M., Shettar, P. and Narayan, D.G., 2021, June. Detection of DDoS Attacks in Software Defined Network using Decision Tree. In *2021 10th IEEE International Conference on Communication Systems and Network Technologies (CSNT)* (pp. 783-788). IEEE.

[9] Aytaç, T., Aydın, M.A. and Zaim, A.H., 2020. Detection DDOS Attacks Using Machine Learning Methods.

[10] Cil, A.E., Yildiz, K. and Buldu, A., 2021. Detection of DDoS attacks with feed forward based deep neural network model. *Expert Systems with Applications*, *169*, p.114520.

[11] Arshi, M., M. D. Nasreen, and Karanam Madhavi. "A Survey of DDOS Attacks Using Machine Learning Techniques." In *E3S Web of Conferences*, vol. 184, p. 01052. EDP Sciences, 2020.

[12] A. Bivens, C. Palagiri, R. Smith, B. Szymanski, M. Embrechts, et al, "Networkbased intrusion detection using neural networks," Intelligent Engineering Systems through Artificial Neural Networks, vol. 12, no. 1 , pp. 579–584, 2002.

[13] Jasreena Kaur Bains ,Kiran Kumar Kaki ,Kapil Sharma,‟ Intrusion Detection System with Multi-Layer using Bayesian Networks‟, International Journal of Computer Applications (0975 – 8887) Volume 67– No.5, April 2013.

[14] M. Alkasassbeh, G. Al-Naymat et.al," Detecting Distributed Denial of Service Attacks Using Data Mining Technique," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, pp. 436-445, 2016. Science and Information Technologies, Vol. 6 (2), pp. 1096-1099, 2015.

[15] Mangesh Salunke, RuhiKabra, Ashish Kumar.‟ Layered architecture for DoS attack detection system by combine approach of Naive Bayes and Improved Kmeans Clustering Algorithm‟, International Research Journal of Engineering and Technology (IRJET), Volume: 02 Issue: 03, June-2015.

[16] T. Subbulakshmi et.al, ‟A Unified Approach for Detection and Prevention of DDoS Attacks Using Enhanced Support Vector Machine and Filtering Mechanisms‟, ICTACT Journal on Communication Technology, June 2013.

[17] Yogeswara Reddy B, Srinivas Rao J, Suresh Kumar T, Nagarjuna A, International Journal of Innovative Technology and Exploring Engineering, Vol.8, No. 11, 2019, pp: 1194- 1198.

[18] HodaWaguih, "A Data Mining Approach for the Detection of Denial of Service Attack", International Journal of Artificial Intelligence, vol. 2 pp. 99106(2013).

[19] Dewan Md. Farid, Nouria Harbi, EmnaBahri, Mohammad Zahid ur Rahman, Chowdhury Mofizur Rahman,‟ Attacks Classification in Adaptive Intrusion Detection using Decision Tree "International Journal of Computer, Electrical, Automation, Control and Information Engineering, Vol:4, No:3, 2010.

[20] Sharafaldin, I., Lashkari, A.H., Hakak, S. and 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.

[21] Subramaniam, A. What is Big Data? — A Beginner's Guide to the World of Big Data. Accessed at: https://www.edureka.co/blog/what-is-big-data/.

[22] Handelman, G.S., Kok, H.K., Chandra, R.V., Razavi, A.H., Huang, S., Brooks, M., Lee, M.J. and Asadi, H., 2019. Peering into the black box of artificial intelligence: evaluation metrics of machine learning methods. *American Journal of Roentgenology*, *212*(1), pp.38-43.