Journal of Data Science and Cyber Security, ISSN: 2584-0010, Volume 1 Issue 1 June 2023

Novel Machine Learning model for DDoS Cyber Attacks Threat Detection

Mouli Prasad J[1], B Rajesh Reddy[1], Micheal Olaolu Arowolo[2], Khushboo Tripathi[3,*]

[1] School of Computer Science and Engineering, Vellore Institute of Technology, Chennai, 600127, Tamilnadu, India
[2] Landmark University – Omu-aran, Kwara State, Nigeria
[3]Dept. of Computer Engineering, Amity University Haryana Gurgaon, India

***Corresponding Author:** Khushboo Tripathi, ktripathi@ggn.amity.edu

Abstract

DDoS (Distributed Denial of Service) is a kind of online threat in which an attacker overwhelms a server or network by sending a lot of traffic from several sources, most of which are usually hacked devices. These attacks have the potential to waste a lot of bandwidth and lose sensitive information. As such, accurately and efficiently detecting DDoS threats has become increasingly important in recent years. DDoS detection has previously been studied as a binary classification problem, with the goal of assessing whether a particular network traffic represents a DDoS attack or not. Therefore, it is essential to understand which DDoS attack type is being targeted in order to properly fight against it. The detection problem is transformed into a multilabel classification using an ensemble classifier in a novel method that addresses this limitation. The most effective algorithms from diverse AI and ML techniques are combined in the proposed Ensemble Classifier. This method works well for identifying various DDoS attacks and categorising them into pertinent groups. The multilabel classification approach can determine the presence of multiple types of attacks simultaneously, providing a more comprehensive and accurate detection mechanism. The effectiveness of the proposed approach is presented with various AI and ML algorithms, and it is found that the Ensemble Classifier outperforms the other algorithms in terms of accuracy and efficiency. By utilizing the best-performing algorithms, the proposed approach can effectively identify various types of such attacks and improve the accuracy of detection. DDoS attacks can cause significant damage to organizations and their networks, making their detection and prevention critical. The suggested Ensemble Classifier methodology offers an effective and efficient way to identify various DDoS attacks since it integrates top-performing algorithms and transforms the detection problem into a multilabel classification.

Keywords

DoS, DDoS, Cyber attacks, machine learning, predictive modeling, classification, feature selection, data analysis

1. Introduction

The hacker initiates a UDP threat by sending a huge number of UDP packets. By using a lag switch or other software that consumes excessive amounts of network bandwidth, UDP-Lag slows down the user by severing the client-server connection. The SYN assault repeatedly sends SYN packets until the system fails in order to take advantage of the TCP three-way handshake. DDoS assaults pose a severe cybersecurity risk, and it's essential to identify different DDoS threats in order to stop them. The two primary DDoS threats are reflection-based and exploitation-based attacks, each with a variety of variations. It is important to have an effective DDoS detection and prevention system to protect against these attacks (Figure 1).

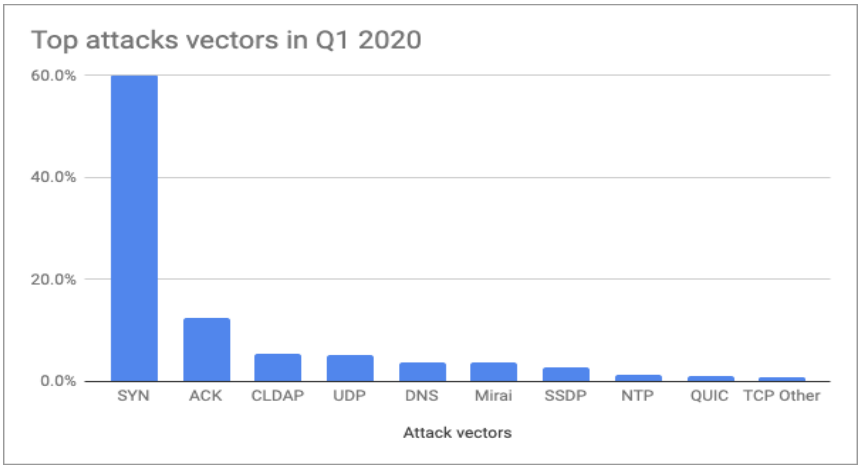

Figure 1: Attack vectors

Attacks on the network layer have increased in bit rate, especially since the COVID-19 incident. A source claims that in March 2020, the maximum assault bit rate was 550 Gbps.A huge DDoS attack against Amazon that was 2.3 Terabits per

second (Tbps) in size and lasted longer than usual (Figure 2).

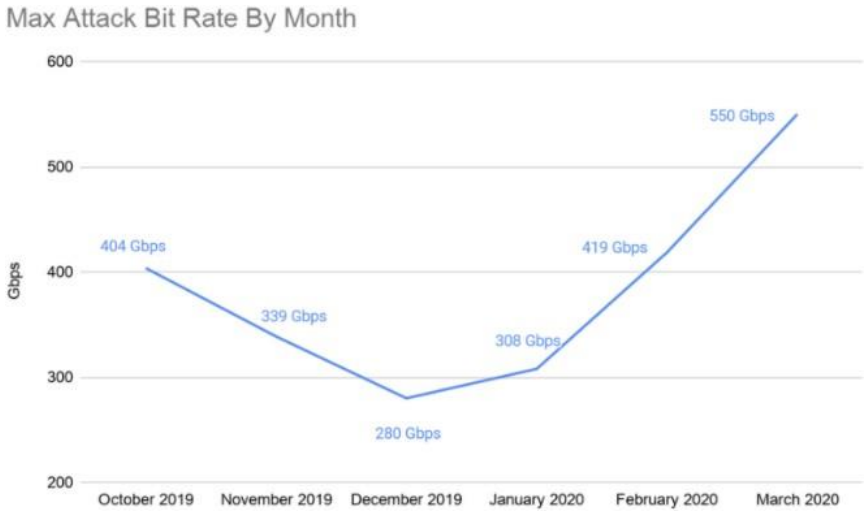

Figure 2: Attack bit rate by month

The need for improved internet security has led to the development of a number of statistical detection techniques, including wavelet-based, port entropy-based, and destination entropy-based approaches. Due to the internet's dynamic nature, the procedures are frequently time-consuming and ineffective. To address this, researchers have turned to AI techniques to detect DDoS attacks, but the lack of classification for the various types of DDoS attacks motivated us to develop efficient ML and AI models that can capture the variability of changing internet domains[8,9]. However, the large dataset with 87 features posed computational complexity and prediction time challenges. To overcome this, we applied feature selection using the Extra Trees classifier technique to select the top 20 relevant features.

The study uses baseline, advanced, and an Ensemble Classifier MV-4 that combines the top 4 AI/ML models for better categorization. The paper is divided into several sections, with Section 3 providing an overview of pertinent research and studies on DDoS cyberthreat detection, Section 4 providing a description of the procedures and variables used in this study, and Section 5 providing details on the dataset, software, and metrics used to assess the AI and ML models. In Section 6, testing methods are covered, and in Section 7, findings and model analysis are covered. The report concludes with potential directions for future research.

2. Literature Review

The majority of studies utilising machine learning techniques to attain high accuracy are focused on binary classification, which dominates the literature on DDoS attack detection. In order to detect DDoS attacks, Balkanli et al. used the open-source systems Bro and Corsaro as well as the machine learning classifiers CART and Naive Bayes, achieving accuracy scores of 0.99 and 0.98, respectively. They failed to categorise the sort of assault, instead concentrating primarily on identifying backscatter traffic[5,6,7]. Nowadays, research is frequently using the Integrated Approach of an IoT and Cyber-Physical System for Security and SDN technology with DDoS effect, as presented in [16,17].

Using a Multilayer Perceptron network with complementing passive metrics and ROC curves, Siaterlis et al. categorised DDoS sources, victims, and kinds with success in both normal and attack modes. Nguyen et al. classified network condition into Normal, pre-attack, and DDoS attack stages using k-Nearest Neighbor, while Barati et al. employed Multilayer Perceptron to improve binary DDoS detection with an accuracy of 99.9971%.

In order to obtain high accuracy in binary classification of DoS attacks targeting layers 3 and 4 of the OSI 7-layer model, Zekri et al. investigated algorithms including Naive Bayesian, C4.5, and K-Means. Deeb et al. classified six different types of DDoS attacks using the Improved Multi-Class Support Vector in the KD-DCup 99 dataset. By beating standard classifiers like Random Forest and Decision Tree, KiruthikaDevi et al. introduced a new model called HCF-SVM to binary categorise DDoS attacks as normal or attack with an accuracy of 98.99%.

A study conducted by Arun.et.al proposed the use of Genetic fuzzy and Neuro-fuzzy methods to enhance the accuracy of DDoS attack detection. They integrated NFBoost algorithm as a subsystem of the ensemble to classify DDoS attacks as normal or attack. Their evaluation was based on two metrics: cost per sample and detection accuracy, which showed a high accuracy of 99.2%.

On the KDD-99 dataset, Li.et.al tested the effectiveness of various methods to identify DDoS attacks. They came to the conclusion that PCA-RNN has a better detection efficiency and accuracy. However, the KDD-99 dataset is not specifically designed to detect DDoS attacks and is an older version of NSL-KDD, which is more suitable for detecting a wider range of attacks. Therefore, the proposed method is only suitable for binary classification of DDoS attacks.

Rahman et al. evaluated the binary classification of DDoS attacks using the J48, RF, SVM, and KNN algorithms[10,11,12]. They came to the conclusion that J48 has the best accuracy, sensitivity, specificity, recall, and F1-score of any classifier. They also created the C4.5

decision tree-based system, which had a 98.8% accuracy rate when it came to identifying the DDoS attack signatures. These models did not identify the different types of attacks; instead, they could only predict whether an attack would occur.

Our research focuses on the identification of various DDoS attack types, which is essential for the successful deployment of countermeasures in cyber security.

3. Selected Algorithms:

In this study, the selection of AI and ML algorithms for detecting different types of DDoS cyber threats was based on their computational complexity [13,14]. This factor is crucial in cases where multiple algorithms have similar performance, as it helps in selecting the one with lower computational complexity. The Computational Complexity of each algorithm used in the study is listed in Table 1 to aid in this selection process.

Machine Learning Algorithms:

Support Vector Machines (SVMs): Support Vector Machines (SVMs) are a subclass of supervised learning algorithms that are mostly employed in classification and regression. SVMs generalise effectively because they employ a hyperplane to divide classes and create a classifier that performs well on novel, untried samples. In this study, a multiclass parameter termed "ovr" was applied with Sci-Kit Learn's LinearSVC for multilabel classification. A square-hinge was employed as the loss function, and the hyperparameters were adjusted to prevent overfitting. This function uses straightforward mathematics to provide results that are computationally effective. To more severely penalise the samples inside the margin, the regularisation value, also known as the cost parameter, was set to 1. This facilitates accurate classification with a greater C-value. To avoid overfitting, a large value of C was avoided, and a value lower than 1 resulted in a soft margin. L1 normalisation produced excessively sparse coefficients, thus L2 normalisation was utilised for penalization. In this study, SVMs with the selected parameters proved to be a reliable tool for identifying DDoS cyberthreats.

Decision Tree : A supervised learning technique known as a decision tree sorts through a tree structure from the root node to the leaf node to determine the final classification label. Hyperplanes or axis-parallel rectangles are used by decision trees to categorise the feature space. They require less data processing since they are less sensitive to outliers. Decision Trees are frequently used as a starting point for comparison with other algorithms.

In our investigation, the minimum number of samples needed to split a node was 3, and the Gini Index was chosen as the splitting criterion. The optimal split was chosen as the splitting strategy at each node. In order to keep the cost complexity low, no pruning was done. We took into account all of the features in the dataset, up to the maximum number of features, to identify the optimum split. With the help of this method, we were able to classify data accurately and forecast the future.

Ensemble Learning Algorithms

a. **Random Forest :** The decision trees in the random forest classifier are drawn at random from a subset of the training set, and their collective votes are then added up to produce the final prediction. Due to its ability to effectively process a high number of input variables without the need to delete any of them, this classifier is a popular choice for processing large datasets. A combination of randomly chosen features and subsets of the training data can also assist prevent overfitting and boost accuracy. The classifier generates objective generalisation error estimates that can be used to assess its performance by creating a forest of decision trees. Several parameters need to be specified correctly for this classifier to get the best results.

b. **Xtreme Gradient Boosting :** A strong algorithm that makes use of the Gradient Boosting method (GBM), Extreme Gradient Boosting (XGBoost), is especially beneficial for handling unstructured data. The GBM technique uses the gradient descent algorithm to minimize errors, but XGBoost goes a step further by optimizing the GBM using various methods, including parallelization, tree pruning, hardware optimization, regularization, sparsity awareness, and cross-validation. This approach improves the scalability and computational speed of the algorithm, making it effective in memory-restricted systems. The algorithm's performance is measured using the "friedman mse" criterion, which evaluates the quality of the split using the Mean Squared Error (MSE) and Friedman's improvement score. In our study, this criterion provides the best approximation of the performance of the algorithm. Furthermore, XGBoost employs the deviance loss function, which is equivalent to logistic regression in classifying probabilistic outputs.

c. **Adaptive Boosting :** Adaboost, also known as adaptive boosting, is an iterative ensemble learning technique that strengthens weak classifiers by learning from their errors and turning them into strong classifiers. This approach makes Adaboost superior to other learning algorithms that rely on random guessing for prediction. The algorithm's base estimator is typically decision trees, which Adaboost can leverage to increase accuracy. However, Adaboost can be computationally slow when compared to XGBoost, and it is highly sensitive to outliers and noise. Moreover, the Gini criterion is used to measure the quality of the split in this algorithm.

d. **Majority Vote Classifier:** By choosing the label that has gotten the most votes from a group of classifiers, the majority voting ensemble learning technique chooses the class label for a given sample. A label is regarded as the chosen class label if it has garnered more than 50% of the votes. In our study, we used the top four classifiers to build a more reliable and accurate model for classifying different DDoS threats. The MV-4 classifier is the name of this group of classifiers, which distinguishes it from other classifiers.

Based on accuracy scores, we chose Random Forest, AdaBoost, Decision Tree, and XGBoost as the top four performing classifiers. To ensure correctness and effectiveness, the Python3.7 code for this ensemble was created from scratch.

In comparison to employing separate classifiers, the MV-4 classifier offers a more dependable and accurate prediction of the class labels for various DDoS threats. We were able to detect and categorise different DDoS attacks with a greater accuracy rate by combining the advantages of each classifier.

Deep Learning Supervised Algorithms:

Multi-Layer Perceptron (MLP) trains on a dataset using the backpropagation learning technique. The capacity of the MLP to learn non-linear functions in real-time, one of its main advantages, makes it appropriate for a variety of applications. However, because the loss function is concave, optimising the MLP necessitates the application of appropriate optimisation methods and hyperparameter adjustment. The MLP is, to put it simply, a machine learning algorithm that can learn from data and categorise it into several groups. To enhance its performance, it makes use of several activation mechanisms and optimisation strategies. The MLP is helpful in numerous real-world applications because it can manage complex interactions between input and output data.

Softmax was utilised as the output layer for prediction and two layers of LSTM with eight units each were employed to get the highest accuracy for multiclass labels.

Apart from the Ensemble Learning method, other algorithms were considered in this study because ensemble-based classifiers are prone to overfitting when the number of features is higher. Additionally, there is no literature on the performance of these algorithms for multilabel classification of different DDoS threats. Hence, a detailed study was conducted to evaluate the performance of different algorithms.

4. Implementation

Dataset:

This study uses the CICDDoS2019 dataset, which is an improvement over its predecessor and addresses. To the best of our knowledge, this dataset has not yet been used in a research publication for multiclass classification for the detection of different forms of DDoS attacks. The dataset is about 26 GB in size, but it is also rather big, making it difficult to process on regular PCs. Consequently, it is essential to decrease the dataset while keeping the integrity of the data distribution in order to train AI and ML models effectively. The dataset is condensed using the Scikit-learn Python Library, maintaining its integrity.

Software used for Implementation:

Together with the use of well-known libraries like NumPy and Pandas Python modules, Python 3.7 serves as the main platform on which all the programmes for each of the Machine Learning algorithms are implemented. Moreover, TensorFlow library [31] and Keras are used as the application layer and backend support, respectively, for Python 3.7 while developing AI models.

Evaluation Metrics:

The following metrics are used to assess artificial intelligence and machine learning (AI and ML) models:

Accuracy Score

The accuracy rating indicates how closely a value is met. The Accuracy Score Formula as Shown in Sci-Kit Learn. The Accuracy score for multilabel types gives us the subset's accuracy.

F1-Score:

The F1-rating's maximum cost is 1, which is also referred to as the dice similarity coefficient.

Receiver Operating Characteristic Curve:

The classifier is in its ideal position in this situation. When the area beneath the curve is greater, the classifier functions more effectively overall.

Testing:

The effectiveness of several AI and ML algorithms utilised in this study for the detection of DDoS attacks is thoroughly examined in this part. Table 2 displays the accuracy score for each algorithm, with SVM having the lowest accuracy score (92.75%) and Decision Tree having the best accuracy score (98.99%) among all ML methods. The AI algorithms, however, fared better than the ML algorithms. LSTM outperformed MLP in terms of accuracy, scoring 98.17% as opposed to 97.33% for MLP. The ensemble learning strategy, which consists of the four algorithms Random Forest, Adaptive Boosting, XGBoost, and Majority Voting, was also assessed (MJV-4). Random Forest had the greatest accuracy rating of all the algorithms, scoring 99.24%, while MJV-4 came in second with a rating of 99.01%. Comparing XGBoost's accuracy score to that of Adaptive Boosting and Random Forest, it was 98.59% lower. The ensemble-based classifiers performed better than the individual AI and ML systems, it is important to note. This is so that the ensemble approach can increase overall accuracy and minimise overfitting by combining the strengths of various algorithms.

In summary, this study provides a comprehensive evaluation of various AI and ML algorithms for detecting DDoS attacks. The results show that ensemble learning methods, especially Random Forest and MJV-4, outperform individual algorithms for multilabel classification on the CICDDoS2019 dataset. These findings can be useful for designing effective DDoS attack detection systems in real-world scenarios.

5. Results and Analysis

For the CICDDoS2019 dataset, the effectiveness of various machine learning techniques was assessed in section 5.1. As shown in Table 2, the Support Vector Machine (SVM) algorithm's accuracy score was 92.75%. Unfortunately, this accuracy score is 5.84% and 6.24% lower respectively than XGBoost and the Decision Tree algorithm. Also, among all the algorithms examined, SVM got the lowest accuracy rating.

The F1-Score for each form of DDoS cyber threat is displayed in Table 2. SVM's F1-Score for identifying SYN assaults was a perfect 1.00. For detecting Web DDoS attacks, it had an F1-Score of 0.52, showing that SVM is ineffective at detecting this kind of attack. SVM's F1-Score for benign traffic was 0.90, showing that it does a good job of differentiating between normal and aberrant traffic. Also, the SSDP assault's F1-Score of SVM was 0.80, which is lower than the F1-Scores of other attack types that were greater than 0.89. Figure 3's ROC curve study reveals that SVM has a lower area under the curve for some assaults, like LDAP and SSDP, and a higher area under the curve for others, including DNS, NetBIOS, and SNMP, which is in line with the F1-Score found in Table 2.

The SVM method performs better than the Decision Tree and XGBoost algorithms with a macro-average F1-Score of 0.98. Overall, the data indicates that SVM may not be as good at detecting Web DDoS attacks and some other forms of attacks as it is at identifying SYN attacks and some network protocols.

Table 1: Accuracy Score of Artificial Intelligence and Machine Learning Models

Algorithm	Accuracy Score
SVM	92.75%
Decision Tree	98.99%
Random Forest	99.24%
XGBoost	98.59%
AdaBoost	99.01%
MJV-4	99.01%
MLP	97.33%
LSTM	98.16%

Decision Tree has a multilabel accuracy score of 98.99%, which is around 6% higher than SVM's. Moreover, all attack types save for Web DDoS have substantially superior F1-Scores for Decision Tree. In comparison to SVM, the F1-Score for Web DDoS is 0.38, which is lower. This programme can flawlessly detect the SYN assault and can detect benign traffic with a score of 0.90. The F1-Score of the other attack types is 0.99, which is higher than the F1-Score of the SVM. Decision Tree has a multilabel accuracy score of 98.99%, which is around 6% higher than SVM's. Moreover, all attack types save for Web DDoS have substantially superior F1-Scores for Decision Tree. In comparison to SVM, the F1-Score for Web DDoS is 0.38, which is lower. This programme can flawlessly detect the SYN assault and can detect benign traffic with a score of 0.90. The F1-Score of the other attack types is 0.99, which is higher than the F1-Score of the SVM.

Decision Tree has a multilabel accuracy score of 98.99%, which is around 6% higher than SVM's. Moreover, all attack types save for Web DDoS have substantially superior F1-Scores for Decision Tree. In comparison to SVM, the F1-Score for Web DDoS is 0.38, which is lower. This programme can flawlessly detect the SYN assault and can detect benign traffic with a score of 0.90. The F1-Score of the other attack types is 0.99, which is higher than the F1-Score of the SVM. SVM and Decision Tree, on the other hand, failed to accurately detect the SYN threat. All other threats, excluding this one, have an F1-Score of 0.98 or 0.99. Figure 5 displays the RoC Curve for the XGBoost algorithm. It is clear from the value of 0.45 that benign traffic has a smaller area under the curve. Moreover, the Web DDoS Cyberthreat's area under the curve is 0.00, indicating that it completely misses this form of attack. The macro average's area under the curve is 0.86, which is less than that of both SVM and Decision Tree.

Table 2: F1-Score of Artificial Intelligence and Machine Learning Models

Threats	SVM	Decision Tree	Random Forest	XGBoost	AdaBoost	MJV-4	MLP	LSTM
BENIGN	0.90	0.90	0.95	0.46	0.90	0.90	0.92	0.89
DNS	0.90	0.99	0.99	0.98	0.99	0.99	0.97	0.98
LDAP	0.95	0.99	0.99	0.98	0.99	0.99	0.98	0.98
MSSQL	0.89	0.99	0.99	0.98	0.99	0.99	0.99	0.99
NTP	0.98	0.99	1.00	0.99	0.99	0.99	0.99	1.00
NetBIOS	0.93	0.99	0.99	0.99	0.99	0.99	0.97	0.97

SNMP	0.97	0.99	0.99	0.99	0.99	0.99	0.97	0.97
SSDP	0.80	0.99	0.99	0.98	0.99	0.99	0.97	0.98
UDP	0.89	0.99	0.99	0.99	0.99	0.99	0.95	0.97
Syn	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
UDP-lag	0.95	0.99	1.00	0.99	0.99	0.99	0.97	0.98
WebDDoS	0.52	0.38	0.53	0.00	0.32	0.38	0.00	0.00

The accuracy score of 99.24% for Random Forest is higher than that of all the AI and ML algorithms discussed in this study. This accuracy score outperforms the SVM algorithm by roughly 6.50% and the XGBoost score by 0.65%. Yet, there is only a 0.25% difference in performance between Random Forest and Decision Tree. Shown is the Random Forest's F1-Score, and it is clear that, of all the algorithms listed in Table 2, it has the highest F1-Score. The F1-Score of 1.00 indicates that the three threats, SYN, UDP-Lag, and NTP, are perfectly detected. Also, it outperforms all other algorithms with an F1-Score of 0.95 for benign traffic, demonstrating that it can distinguish between normal and abnormal traffic with accuracy. In comparison to the previously mentioned methods, it has a better F1-Score of 0.53 for Web DDoS attacks. As shown in table 2, the micro and macro averages both have a score of 1.00.

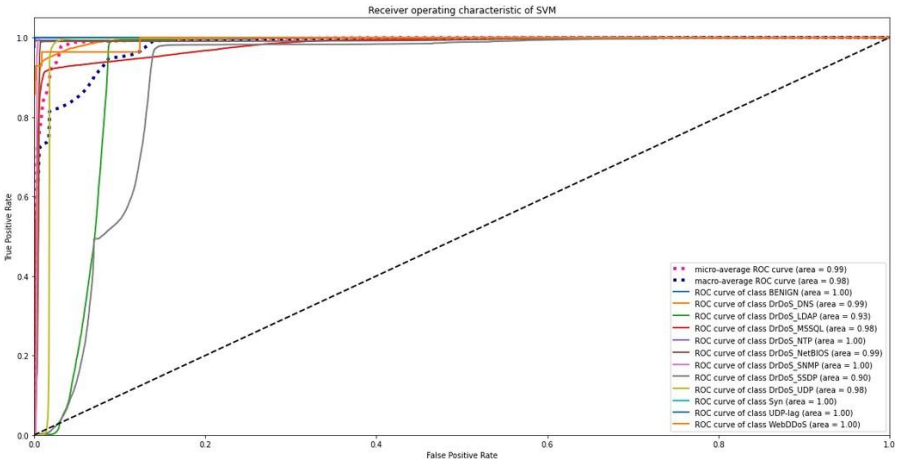

Figure 3: RoC Curve for Support Vector Machine (SVM)

In this study, the ensemble learning method called Adaptive Boosting uses Decision Tree as its basis estimator. This algorithm's accuracy score is 99.01 percent, which is greater than that of SVM, Decision Tree, and XGBoost but slightly behind that of the Random Forest approach. Table 2 also provides the F1-Score for each of the various threats.

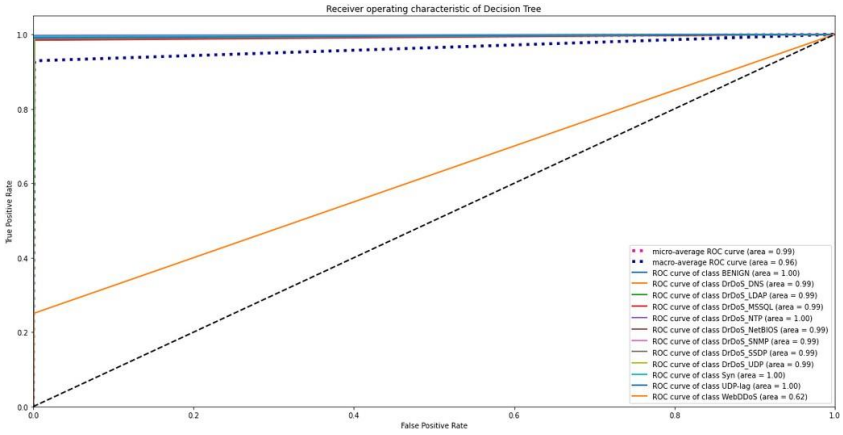

Figure 4: RoC Curve for Decision Tree

The F1-Score for benign traffic is 0.90, which is well detected by this technique and performs 0.44 better than XGBoost. Also, the bulk of threats have an F1-Score of 0.99, whereas SYN Cyberthreat has a perfect F1-Score of 1.00, indicating that this system detects them very well. The area under the curve for the Web DDoS Cyberthreat is 0.62, which is less than that of Random

Forest, as seen in Figure 7's RoC Curve for this technique. Moreover, the micro and macro averages fall short of Random Forest's. According to this RoC, this method outperforms SVM, Decision Tree, and XGBoost while slightly underperforming Random Forest.

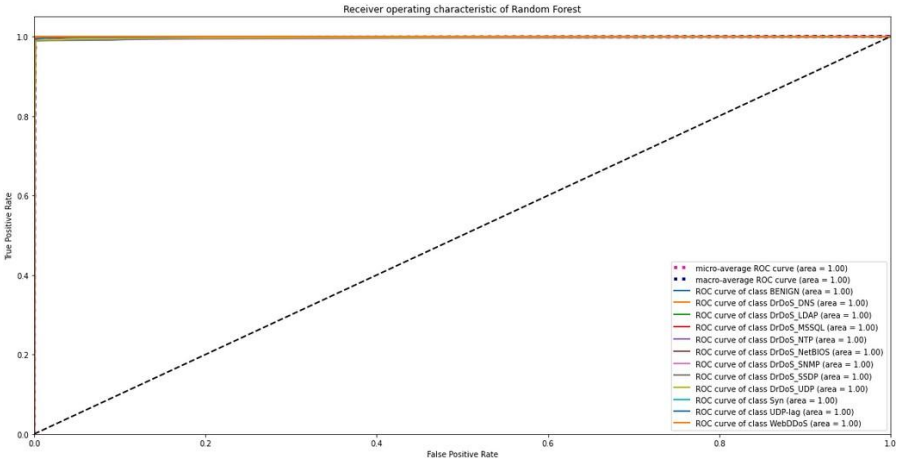

Figure 5: RoC Curve for Random Forest

The Majority Voting Classifier (MV-4) combines the performance of the Adaboost, Decision Tree, Random Forest, and XGBoost algorithms, and as a result, achieves an accuracy score of 99.01%, which is identical to Adaboost's and only marginally higher than Random Forest's. The F1-Score in Table 2 makes it clear that SYN has a score of 1.00 while the majority of the threats have an F1-Score of 0.99. Moreover, Benign has a score of 0.90, which is lower than Random Forest's 0.95, while Web DDoS threat has a lower score of 0.38. The area under the curve in Figure 8's RoC Curve for the MV-4 classifier is 1.00 [2].

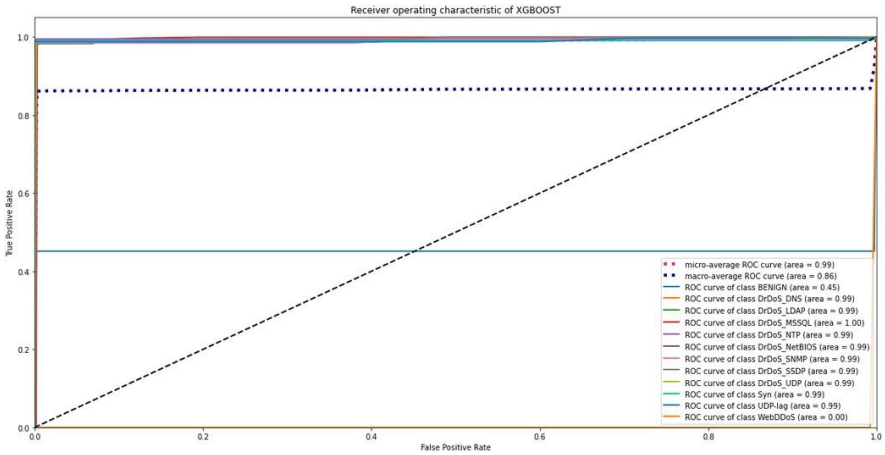

Figure 6: RoC Curve for XGBoost Classifie

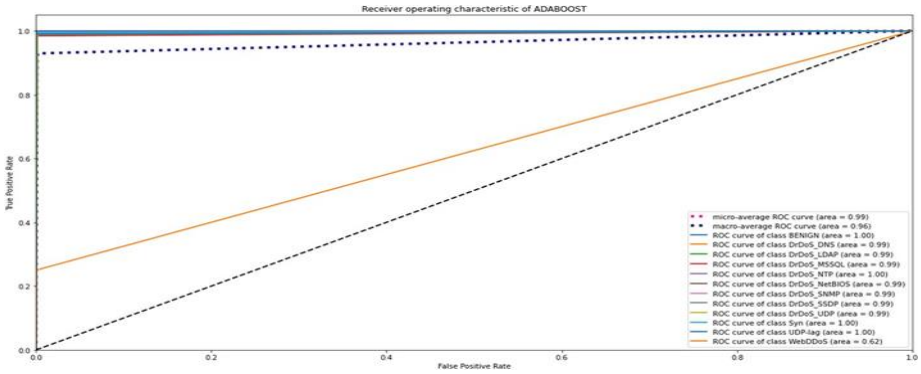

Figure 7: RoC Curve for Adaptive Boosting Classifier (Adaboost)

When compared to device learning algorithms, the performance of AI algorithms is a little bit worse. The MLP accuracy rating is 97.33%, which is higher than SVM's accuracy rating of 4.58% and lower than Random Forest's accuracy rating of approximately 2%. Table 3 displays the F1 rating for the MLP set of rules for exceptional DDoS cyberthreats. Net DDoS Cyberthreat's F1-score is 0.00, which indicates that it is unable to detect this threat. Benign website visitors also have a 0 F1 rating. 92, which is greater than XGBoost and somewhat lower than Random forest. Each threat has an appropriate F1 score. In Distinct 9, the RoC Curve for MLP is demonstrated. The area under the curve for WebDDoS on this parent is zero. fifty two, which shifts the location beneath the macro average curve to 0.96.

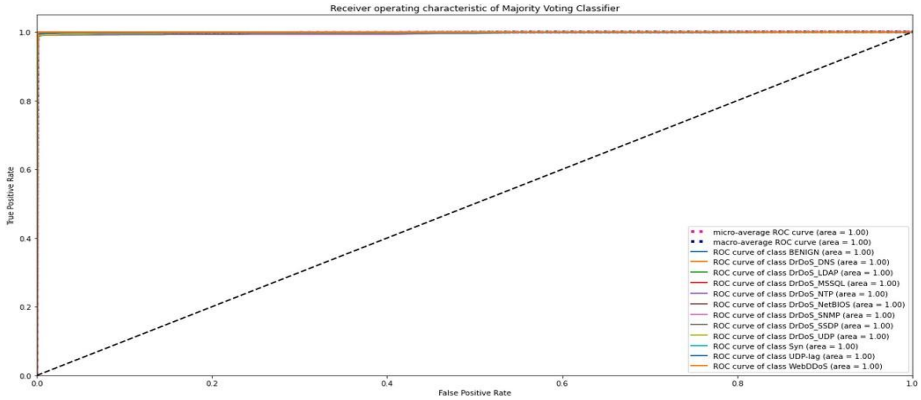

Figure 8: RoC Curve for Majority Voting Classifier (MV-4)

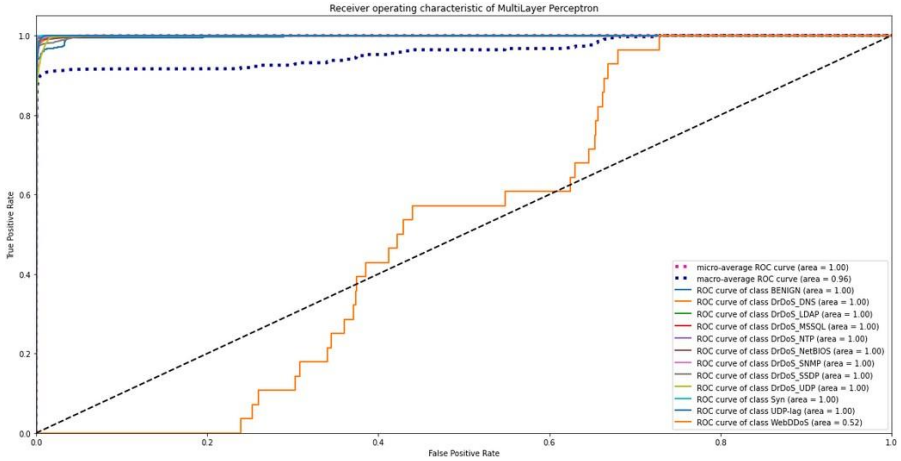

Figure 9: RoC Curve for Multilayer Perceptron (MLP)

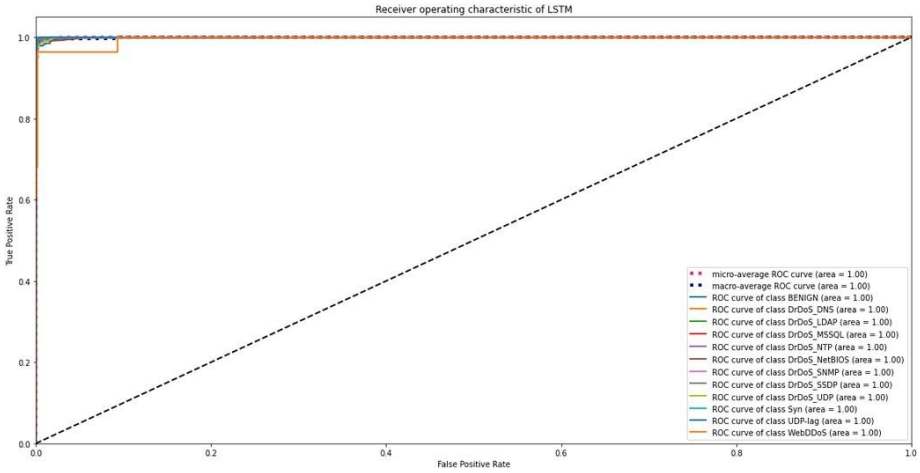

Figure 10: RoC Curve for Long Short Term Memory (LSTM)

The accuracy score for LSTM is 98.16 percent, which is just a little bit higher than that of MLP. However, this result is 5.41 percent better than SVM and about 1% lower than Random Forest, as shown in Desk 2. The F1-rating for LSTM on exceptional Cyberthreats is also demonstrated in Table 2. From the table, it's clear that the ranks are marginally higher in all cases when compared to the MLP set of standards, with the exception of Benign Traffic where it falls from 0.92 to 0.89 from MLP to LSTM. Similar to MLP, Net DDoS risk has an F1-rating of zero, or 0.00. The parent 10 denotes the RoC for the usage of the LSTM algorithm in uncommon assault types. The figure makes it clear that the area under the curve for all unique cyberthreats is 1.00, along with a common rating of 1.00 for micro and macro threats that is similar to random forest. The RoC curve makes it abundantly clear that the LSTM's overall performance, which is based on the RoC curve, is quite similar to that of a Random Forest-style ideal classifier. As previously said, tables 1 and 2 are given and Figure 3-10 RoC curves are presented according to data sets. Further readers are requested to refer articles [17-35] to know about emerging technologies and their role in solving real world problems.

6. Conclusion and Future Work

Multiclass categorization for DDoS cyberthreats was performed in this work using a variety of AI and ML methods. Several indicators were used to identify and validate each danger. For the purpose of detecting DDoS cyberthreats, the Ensemble Classifier MV-4 was introduced, and it scored an amazing accuracy of 99.01%. The Random Forest Classifier earned the greatest accuracy score of 99.24%, followed by MV-4 and AdaBoost Classifier with an accuracy score of 99.01%, according to a thorough review of numerous AI algorithms. AdaBoost, however, fell short of Random Forest and MV-4 in its ability to identify various threats, as evidenced by the F1-Score and RoC curve data.

The F1-Score for Random Forest and MV-4 is equal, with Random Forest slightly outperforming MV-4 because it correctly identified three threats. All of the algorithms' RoC Curves were shown, offering a thorough evaluation of each algorithm's performance. The MV-4 and Random Forest classifiers performed as ideally as possible, with classifier scores of 1.00 for all varieties of Cyberthreats. The micro and macro averages for both classifiers were a perfect 1.00, behaving as an ideal classifier on this dataset.

The work in the future will concentrate on implementing various solutions for every type of attack to protect the network from such attacks using AI and ML algorithms. This work will be further extended to develop a system that can detect DDoS Cyberthreats and deploy countermeasures to prevent critical Cybersecurity threats.

References

[1]. Carlos A. Moreno C., Jairo R., Montoya Tores, Anicia J., Natacha Gondran, "Sustainability Metrics for Real Case Applications of the Supply Chain Network Design Problem: A Systematic Literature Review, Elsevier, pp.1-50,2019.

[2]. C. Douligeris and A. Mitrokotsa, "DDoS attacks and defense mechanisms: a classification," Proceedings of the 3rd IEEE International Symposium on Signal Processing and Information Technology (IEEE Cat. No.03EX795), Darmstadt, Germany, 2003, pp. 190-193, doi: 10.1109/ISSPIT.2003.1341092.

[3]. B. Nagpal, P. Sharma, N. Chauhan and A. Panesar, "DDoS tools: Classification, analysis and comparison," 2015 2nd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2015, pp. 342-346.

[4]. T. Radivilova, L. Kirichenko, D. Ageiev and V. Bulakh, "Classification Methods of Machine Learning to Detect DDoS Attacks," 2019 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Metz, France, 2019, pp. 207-210, doi: 10.1109/IDAACS.2019.8924406.

[5]. Muhammad Aamir, Syed Mustafa Ali Zaidi, "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.

[6]. ABBASS ASOSHEH, NAGHMEH RAMEZANI, "A Comprehensive Taxonomy of DDoS Attacks and Defense Mechanism Applying in a Smart Classification", WSEAS TRANSACTIONS on COMPUTERS,, vol.4 issue 7, pp.281-291, 2008.

[7]. Jelena Mirkovic and Peter Reiher. 2004. A taxonomy of DDoS attack and DDoS defense mechanisms. SIGCOMM Comput. Commun. Rev. 34, 2 (April 2004), 39–53.

[8]. Alefiya Hussain, John Heidemann, and Christos Papadopoulos. 2003. A framework for classifying denial of service attacks. In Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications (SIGCOMM '03). Association for Computing Machinery, New York, NY, USA, 99–110.

[9]. S. Nandi, S. Phadikar and K. Majumder, "Detection of DDoS Attack and Classification Using a Hybrid Approach," 2020 Third ISEA Conference on Security and Privacy (ISEA-ISAP), Guwahati, India, 2020, pp. 41-47.

[10]. R. F. Fouladi, C. E. Kayatas and E. Anarim, "Frequency based DDoS attack detection approach using naive Bayes classification," 2016 39th International Conference on Telecommunications and Signal Processing (TSP), Vienna, Austria, 2016, pp. 104-107.

[11]. P. S. Saini, S. Behal and S. Bhatia, "Detection of DDoS Attacks using Machine Learning Algorithms," 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2020, pp. 16-21.

[12]. Bayu Adhi, T., Baltimore County and Kyung H.R., " Data mining techniques in DoS/ DDoS attack detection: A Literature review", International Journal on Information,pp.4-10, 2015.

[13]. Lu Zhou, Ye Zhu, Tianrui Zong, Yong Xiang, "A feature selection-based method for DDoS attack flow classification," Future Generation Computer Systems, Volume 132,2022,Pages 67-79.

[14]. A. Alsirhani, S. Sampalli and P. Bodorik, "DDoS Attack Detection System: Utilizing Classification Algorithms with Apache Spark," 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Paris, France, 2018, pp. 1-7.

[15]. A. R. Wani, Q. P. Rana, U. Saxena and N. Pandey, "Analysis and Detection of DDoS Attacks on Cloud Computing Environment using Machine Learning Techniques," 2019 Amity International Conference on Artificial Intelligence (AICAI), Dubai, United Arab Emirates, 2019, pp. 870-875.

[16]. Tripathi, Khushboo, Deepshikha Agarwal, and Kumar Krishen. "An Integration Approach of an IoT and Cyber-Physical System for Security Perspective." In Handbook of Research of Internet of Things and Cyber-Physical Systems, pp. 187-218. Apple Academic Press, 2022.

[17]. Midha, Sugandhi, and Khushboo Tripathi. "Remotely triggered blackhole routing in SDN for Handling DoS." In Proceedings of International Conference on IoT Inclusive Life (ICIIL 2019), NITTTR Chandigarh, India, pp. 3-10. Springer Singapore, 2020.

[18]. Rekha, G., Tyagi, A.K., Anuradha, N. (2020). Integration of Fog Computing and Internet of Things: An Useful Overview. In: Singh, P., Kar, A., Singh, Y., Kolekar, M., Tanwar, S. (eds) Proceedings of ICRIC 2019 . Lecture Notes in Electrical Engineering, vol 597. Springer, Cham. https://doi.org/10.1007/978-3-030-29407-6_8

[19]. Tibrewal, I., Srivastava, M., Tyagi, A.K. (2022). Blockchain Technology for Securing Cyber-Infrastructure and Internet of Things Networks. In: Tyagi, A.K., Abraham, A., Kaklauskas, A. (eds) Intelligent Interactive Multimedia Systems for e-Healthcare Applications. Springer, Singapore. https://doi.org/10.1007/978-981-16-6542-4_17

[20]. Tyagi, Amit Kumar, Building a Smart and Sustainable Environment using Internet of Things (February 22, 2019). Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur - India, February 26-28, 2019. http://dx.doi.org/10.2139/ssrn.3356500

[21]. Tyagi, Amit Kumar and M, Shamila, Spy in the Crowd: How User's Privacy Is Getting Affected with the Integration of

Internet of Thing's Devices (March 20, 2019). Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM), Amity University Rajasthan, Jaipur - India, February 26-28, 2019.

[22]. Amit Kumar Tyagi, N. Sreenath, "Preserving Location Privacy in Location Based Services against Sybil Attacks", International Journal of Security and Its Applications (ISSN: 1738-9976 (Print), ISSN: 2207-9629 (Online)), Volume 9, No.12, pp.189-210, December 2015.

[23]. Gillala Rekha, Amit Kumar Tyagi, and V. Krishna Reddy, "A Wide Scale Classification of Class Imbalance Problem and its Solutions: A Systematic Literature Review", Journal of Computer Science, Vol.15, No. 7, 2019, ISSN Print: 1549-3636, pp. 886-929.

[24]. A. K. Tyagi, T. F. Fernandez and S. U. Aswathy, "Blockchain and Aadhaar based Electronic Voting System," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2020, pp. 498-504, doi: 10.1109/ICECA49313.2020.9297655.

[25]. S. U. Aswathy, Amit Kumar Tyagi, Shabnam Kumari, "The Future of Edge Computing with Blockchain Technology: Possibility of Threats, Opportunities and Challenges", in the Book "Recent Trends in Blockchain for Information Systems Security and Privacy", CRC Press, 2021.

[26]. Rekha, G., Tyagi, Amit Kumar, and Krishna Reddy, V. 'Solving Class Imbalance Problem Using Bagging, Boosting Techniques, with and Without Using Noise Filtering Method'. 1 Jan. 2019 : 67 – 76.

[27]. B. Gudeti, S. Mishra, S. Malik, T. F. Fernandez, A. K. Tyagi and S. Kumari, "A Novel Approach to Predict Chronic Kidney Disease using Machine Learning Algorithms," 2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India, 2020, pp. 1630-1635, doi: 10.1109/ICECA49313.2020.9297392.

[28]. Amit Kumar Tyagi, Meenu Gupta, Aswathy SU, Chetanya Ved, "Healthcare Solutions for Smart Era: An Useful Explanation from User's Perspective", in the Book "Recent Trends in Blockchain for Information Systems Security and Privacy", CRC Press, 2021.

[29]. Tyagi, A.K., Kumari, S., Fernandez, T.F., Aravindan, C. (2020). P3 Block: Privacy Preserved, Trusted Smart Parking Allotment for Future Vehicles of Tomorrow. In: , et al. Computational Science and Its Applications – ICCSA 2020. ICCSA 2020. Lecture Notes in Computer Science(), vol 12254. Springer, Cham. https://doi.org/10.1007/978-3-030-58817-5_56

[30]. Kumar, A., Tyagi, A.K., & Tyagi, S.K. (2014). Data Mining: Various Issues and Challenges for Future A Short discussion on Data Mining issues for future work.

[31]. Amit Kumar Tyagi, G. Rekha, "Challenges of Applying Deep Learning in Real-World Applications", Book: Challenges and Applications for Implementing Machine Learning in Computer Vision, IGI Global 2020, p. 92-118. DOI: 10.4018/978-1-7998-0182-5.ch004

[32]. Amit Kumar Tyagi, N. Sreenath, Cyber Physical Systems: Analyses, challenges and possible solutions, Internet of Things and Cyber-Physical Systems, Volume 1, 2021,Pages 22-33,ISSN 2667-3452,https://doi.org/10.1016/j.iotcps.2021.12.002.

[33]. Nair, Meghna Manoj; Tyagi, Amit Kumar "Privacy: History, Statistics, Policy, Laws, Preservation and Threat Analysis", Journal of Information Assurance & Security . 2021, Vol. 16 Issue 1, p24-34. 11p.
S. Mishra and A. K. Tyagi, "Intrusion Detection in Internet of Things (IoTs) Based Applications using Blockchain Technolgy," 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, 2019, pp. 123-128, doi: 10.1109/I-SMAC47947.2019.9032557.

[34]. M. Shamila, K. Vinuthna and A. K. Tyagi, "A Review on Several Critical Issues and Challenges in IoT based e-Healthcare System," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 1036-1043, doi: 10.1109/ICCS45141.2019.9065831.