International Journal of Applied Engineering Research

Network Intrusion Detection by using the Random Forest with Extra Tree Classifier

[1]Muhammad Farhan
muhammadfarhan01@gmail.com

[2]Hafiz Waheed Ud Din
waheedqaziksa@gmail.com

[3]Muhammad Ijaz Khan
ijaz171@gmail.com

[4]Waqas Tariq Paracha
waqasparacha125@gmail.com

[5]Saadat Ullah
Ksaadat125@gmail.com

[1,2, 3, 4,5]Institute of Computing and Information Technology, Gomal University, Pakistan.

Submitted: 04th March 2022 Revised: 15th April 2022 Accepted: 25th May 2022

Abstract: The constantly expanding popularity of the World Wide Web and the excessive amount of network traffic are two of the main factors responsible of a continuing increase in the number of security threats that affect networks. In order to obtain sensitive data, cybercriminals take advantage of vulnerabilities in the architecture of networks. Utilizing a Network Intrusion Detection System, also known as an NIDS, allows for more accurate identification of various attacks and more quick protection of network resources. The application of machine learning algorithms allows for the detection of anomalies in network traffic and the resolving of network security concerns. The complexity of recent attacks, on the other hand, makes it difficult for the current generation of NIDS to recognize novel threats. It is necessary to find a solution to this problem in order to build and train NIDS using an up-to- date dataset that is comprehensive and incorporates the most recent attack activity. This research presents a comprehensive classification model that is built on machine learning, and it may be used to identify many kinds of network attacks. This study makes use of a dataset known as UNSW-NB15, which includes a vast amount of information pertaining to network traffic as well as various kinds of network attacks. The accuracy of comprehensive machine learning models with regard to multi-class is 96.1%, respectively, which is greater than the accuracy of previous models. This is achieved by using selected features from the Extra tree classifier, a technique for feature selection that selects high relevant features.

Introduction

The highly developed and interconnected world of today has resulted in an increase in the prevalence of data and network security issues. The primary causes of this issue are the growth in the volume of network traffic and the developments in technology, both of which may have played a significant role in the emergence of a new kind of cyberattack. As a direct consequence of this, the intensity of the attack becomes increased [1] [2]. In today's world, there are many types of threats to the validity of a network's security, and an enormous amount of intrusion detection systems (IDS) are being developed and put into use to identify threats as rapidly as possible. An intrusion detection system, sometimes known as an IDS, is a monitoring tool that looks for potentially malicious activity and notifies users as soon as it finds it. An analyst at the security operations center will investigate the problem after

receiving these alerts and will then take the necessary steps to protect against the threat. Both intrusion detection systems (IDS) and firewalls are concerned with the security of networks; however, an IDS monitors the network for intrusions from both the outside and the inside, while a firewall is designed to prevent intrusions from occurring by checking for them only on the outside of the network. Machine learning technologies are currently being used by security specialists to protect the data of organizations and the reputations of such businesses. This is due to the complexity and severity of security attacks on computer networks. Because it can analyze vast amounts of data and execute a range of computations in an effective manner. The utilization of machine learning models in conjunction with an additional tree classifier is the primary emphasis of the work that has been offered in order to detect intrusion from the dataset that has been provided. Examining machine learning methods as a means of building an intrusion detection system-based model that is capable of automatically identifying a wide variety of network threats is the objective of this research. Its primary goals are the identification of potential threats, the assessment of the effectiveness of the model, and the classification of abnormal and typical attacks. The purpose of the study is to model and improve detection rates while examining unknown attacks in comparison to benchmarks [3].Existing work has been done multi-classification [3] but the accuracy with respect to the multi-classification is not so good. It is a challenging process to detect incursion because it is influenced by a number of different elements, such as the growth of network traffic, technological improvements that have given rise to new forms of attacks, and the utilization of older versions of datasets [4] [3].We are currently looking into the UNSW-NB15 dataset (https://research.unsw.edu.au/projects/unsw-nb15-dataset) as a possible solution to this problem. This is a more recent type of dataset that can effectively detect many kinds of attacks.

Research Questions

RQ. 1: How can the proposed model, which utilizes Random Forest with Extra Tree Classifier, be used to identify intrusions within a given dataset?
RQ. 2: What is the accuracy exhibited by the intrusion classification models proposed in similar research studies when it comes to predicting model predictions?

Research Objectives

1. Predicting Network intrusion detection by using the usage of the Random Forest with extra Tree Classifier.
2. Evaluating the advised model's performance to that of other benchmark works.

Related Work

Several authors have presented forth ideas for preventing and detecting network attacks.
In the research published in 2011, presents three different Machine Learning algorithms through an evaluation using the KDD Cup 1999 dataset and the WEKA data mining tool. The author discusses NIDS and the two primary forms of classification that it employs, namely detection of abnormalities and detection of misuse. An investigation was carried out in 2013 by researchers, [5] to compare and analyze a variety of tree-based classification strategies using the NSL-KDD 99 dataset as the basis for their work. The following sets of algorithms were selected in order to carry out the research: AD Tree, C4.5, LAD Tree, NB Tree, Random Tree, Random Forest, and REP Tree. The results of the trials showed that the Random Tree model, the Random Forest model, and the REP Tree model all achieved the best successful accuracy scores.[6][7][8]have provided an overview of how machine learning technologies are being utilized in IDS to discover attacks and construct efficient IDS systems.[9][10][11][12]In order to develop a deep learning method for discovering anomalies, you should first apply a deep learning strategy to the KDD-99 dataset. The article [13] provides a highly in-depth analysis of the machine learning models as well as the datasets that can be utilized to locate network intrusions. [14] illustrates how difficult the UNSW-NB15 data set is to work with. The findings of the experiments indicate that UNSW-NB15 was a more difficult data set than KDD99; hence, it is being regarded as a new benchmark data set for assessing NIDS. [15] uses a total of four distinct supervised machine learning approaches in order to investigate the KDD99 dataset and look for unexpected patterns. According to the findings of the research, it appears that each machine learning algorithm produces distinctive outcomes for the attack classes represented in each KDD 99 dataset. The researchers believe that the application of feature selection algorithms will result in the development of better results for work that will be done in the future. An ensemble-based artificial neural network cascade is presented in reference [16] for the purpose of performing multi-class intrusion detection in computer network traffic. Both the KDD CUP 1999 dataset and UNSW-NB15, a more recent synthetic attack activity dataset, were utilized in the evaluation of the method that was suggested. The results of our experiments point to the possibility that our method can effectively detect a variety attacks in computer networks. In [17], a comparison of five different machine learning approaches is carried out using the KDD Cup 99 dataset. The authors provide a concise overview of each of the five machine learning methods used in this study, which are as follows: Naive Bayes, Bayes NET, Random Forest, Multilayer Perception, and Sequential Minimal Optimization. Additionally, IDS devices are discussed. [18] presents a two-stage classifier for use in network intrusion detection systems. The author's classifier is founded on the RepTree algorithm as well as a subset of protocols. The authors utilized two distinct data sets—the UNSW-NB15 data

set and the NSL-KDD data set—in order to evaluate the efficacy of the technique. The study were able to attain a detection accuracy of 88.95% throughout the full UNSW-NB15 data set as well as 89.85% across the entire NSL-KDD data set, respectively.[19]Using Deep Learning, create an anomaly-based NIDS that will be applied to the KDDCup99 dataset. Experiments performed on the KDDCup99 dataset demonstrate that the work was able to effectively identify abnormalities in network-based intrusion detection systems and classified intrusions into five groups utilizing network data sources. Additionally, the experiments show that the work was successful in classifying intrusions. [20] presented a system in their research that identifies botnets and the traces of their use through the application of machine learning techniques. They used network flow identifiers on some of the data that was collected for the UNSW-NB15 project. Methods of categorization including ARM, ANN, NB, and DT were utilized. The Decision Tree technique achieved the best levels of accuracy (93.23%), as well as the lowest levels of the False Positive Rate (FPR), at 6.77%.In a study that was carried out by[21]in the year 2018, the KDDCup99 Test datasets were investigated with the help of particular machine learning algorithms (Bayes Net, J48, Random Forest, and Random Tree). The purpose of this investigation was to evaluate the degree to which these algorithms were accurate in classifying various attacks into their appropriate categories. The research utilized a constructive research technique, and the findings of the experimental analysis reveal that the Random Forest and Random Tree algorithms demonstrate a high level of efficiency in accurately identifying the attacks that are contained within the Test dataset using the NSL-KDD dataset [22] examined and compared three different machine learning algorithms by using the NSL-KDD dataset. The researchers discussed the importance of the dataset for their investigation. They took the dataset that was provided by NSL-KDD, preprocessed the data, selected ML classifiers (SVM, RF, and ELM), and then determined accuracy, precision, and recall. The researchers developed a decision tree-based IDS framework for usage over Big Data in a fog environment, and they utilized a dataset that was obtained from KDDCUP99 [23].The results of the study showed that the strategy that was presented was one that was not only effective but also reliable. A Convolutional Neural Network (CNN) served as the basis for the method that was proposed in [24], which made use of it as its model. This CNN was built up of a number of layers of perceptron units. They trained their model by maximizing the effectiveness of the hyper parameters. The information presented in [3] provides a framework for determining the many types of attacks that can be carried out against a network. Random Forest, Decision Tree, Logistic Regression, K-Nearest Neighbors, and Artificial Neural Networks are the five various methodologies that were utilized in order to accomplish the task of attack detection. Throughout the course of our research, we make use of a dataset known as UNSW-NB15 that was produced and distributed by the University of New South Wales. The result with respect to different methods, Radom Forest result showed an accuracy of 95.1% after it was applied.

Research Methodology

Research methodology is discussed in this section. The UNSW-NB15 dataset is analyzed in an effort to train machine learning algorithms like Random Forest to identify attacks.

Research Model

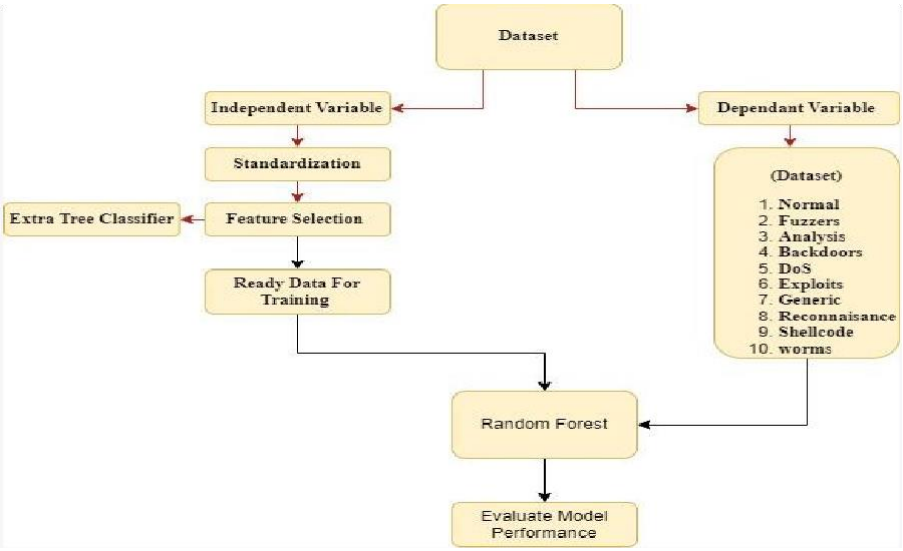

Figure 1: A research model for predicting network attack categories

Sequence of Steps

The study was done in a structured way by following a set of steps in order to get the results. At first, StandardScaler was used on the UNSW-NB15 dataset to take the value in some common range. Then, feature selection was carried out to get rid of features that are irrelevant. After that, the chosen attributes from the feature engineering methods were used to train classifiers. Finally, The Random forest classifier will be used to evaluate the result.

Experiment Results and Discussion

The results from each phase of the study will be outlined in this section, along with an explanation of how the various tests were conducted. The UNSW-NB15 dataset, which comprises 80,000 records, is the one taken into consideration for this study (https://www.kaggle.com/datasets/mrwellsdavid/unsw-nb15). On this dataset, a multi classification model will be applied.

RQ. 1: How can the proposed model, which utilizes Random Forest with Extra Tree Classifier, be used to identify intrusions within a given dataset?

Based on the study question, the UNSW-NB15 dataset has been processed through the Random Forest model using 80,000 samples. 70% of the data is used for training and 30% is used for testing, with a focus on multi-classification. The first four and the last four independent attributes are presented in Table 1, together with the actual classes that correspond to each one. In particular, the samples chosen are the first five and the last four rows, as shown in Table 1.

Table 1: Dataset Samples for Random Forest classification model

id	dur	proto	service	state	…..	ct_flw_http_mthd	ct_src_ltm	ct_srv_dst	is_sm_ips_ports	Attack_Cat
64516	1.10E-05	udp	dns	INT		0	3	4	0	Generic
13291	4.00E-06	udp	dns	INT		0	30	30	0	Generic
57823	1.00E-05	udp	dns	INT		0	4	11	0	Generic
35019	23.412241	tcp	http	FIN		1	2	3	0	Normal
42367	0.643499	tcp	-	FIN		0	2	3	0	Normal
…….										
74520	1.00E-05	udp	-	INT		0	4	3	0	Normal
48773	0.218185	tcp	http	FIN		0	1	2	0	Generic
68879	2.00E-05	tcp	-	REQ		0	6	14	0	Normal
16278	6.00E-06	udp	dns	INT		0	26	39	0	Generic

After gathering the samples, a method of feature selection known as the Extra tree classifier was utilized on the various attributes. The dataset had a total of 43 different attributes. The extra tree classifier will only choose the 13 attributes that are most highly relevant shown in Table 2.

Table 2: Selected attributes taken from Extra Tree Classifier

Sr.No	Selected Attributes Through Extra Tree classifier	High Relevance values selecting For Attribute
1	proto	(0.203618)
2	spkts	(0.074687)
3	sload	(0.069601)
4	dload	(0.043578)

5	ct_ftp_cmd	(0.043030)
6	is_ftp_login	(0.041661)
7	dtcpb	(0.041405)
8	attack_cat	(0.037940)
9	ct_src_dport_ltm	(0.034086)
10	trans_depth	(0.033082)
11	ct_dst_src_ltm	(0.028974)
12	ct_dst_ltm	(0.027750)
13	dpkts	(0.023387)

After choosing a highly relevant attribute, a standard scaler process is used to scale the values of the selected attributes. **Table 3** displays the first five and the last four of the standardized data.

Table 3: Standardization data of Random Forest Model

	proto	**spkts**	**sload**	**dload**	**ct_ftp_cmd**	**is_ftp_login**	**dtcpb**	**attack_cat**	**ct_src_dport_ltm**	**trans_depth**	**ct_dst_src_ltm**	**ct_dst_ltm**	**dpkts**
0	1.60	-0.67	-0.56	0.72	1.35	0.96	-0.42	-0.41	-0.35	-0.47	-0.45	-0.39	-0.56
1	1.57	-0.67	0.93	0.72	-0.81	-1.04	-0.48	-0.50	0.58	-0.47	-0.45	-0.39	-0.56
2	1.03	-0.67	-0.56	-1.16	1.35	0.96	4.86	-0.41	-0.35	-0.47	-0.45	-0.57	-0.74
3	1.22	-0.67	-2.04	-1.16	1.35	0.96	0.14	-0.50	1.51	-0.35	-0.45	-0.31	-0.47
4	-0.12	0.28	-2.04	-1.47	-0.56	-1.04	-0.32	-0.23	-1.27	-0.35	-0.45	-0.48	-0.65
……													
79996	0.08	-0.67	0.93	0.72	-0.81	-1.04	1.96	1.22	0.58	0.25	-0.45	1.36	1.25
79997	0.71	0.28	0.93	0.72	-0.81	-1.04	-0.40	-0.05	0.58	-0.11	0.05	-0.31	-0.02
79998	1.49	-0.67	-0.56	0.72	1.35	0.96	-0.39	-0.50	-0.35	-0.47	-0.45	-0.48	-0.65
79999	-0.77	0.28	0.93	0.72	-0.81	-1.04	-0.40	0.94	0.58	1.80	2.75	1.18	0.97

Table 4 displays the results of the Random Forest model, including both the predicted value and the actual value after standardization.

Table 4: Actual and Predicted value of Random Forest Model

Id	**Actual value**	**Predicted value in array form**	**Predicted value**
10933	7	[0. 0. 0. 0. 0. 0. 0. 1. 0. 0.]	7
20317	5	[0. 0. 0. 0. 0. 1. 0. 0. 0. 0.]	5
3276	1	[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]	3
6258	3	[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]	3

5072	3	[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]	3
3280	3	[0. 0. 0. 1. 0. 0. 0. 0. 0. 0.]	3
59056	1	[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]	1
13534	5	[0. 0. 0. 0. 0. 1. 0. 0. 0. 0.]	5
12700	5	[0. 0. 0. 0. 0. 1. 0. 0. 0. 0.]	5
77555	6	[0. 0. 0. 0. 0. 0. 1. 0. 0. 0.]	6

In the final analysis, the end result of the multi-classification process in terms of accuracy, precision, recall, and f1-score is presented in the table that can be found below.

Table 5: Classification report of Random Forest Model

Accuracy: 0.9624583333333333

	precision	recall	f1-score	support
0	0.70	0.66	0.68	181
1	0.72	0.71	0.72	176
2	0.84	0.77	0.81	1235
3	0.88	0.92	0.90	3214
4	0.91	0.94	0.93	1746
5	1.00	0.99	0.99	5492
6	1.00	1.00	1.00	10832
7	0.95	0.93	0.94	987
8	0.90	0.78	0.84	125
9	0.77	0.83	0.80	12
accuracy			0.96	24000
macro avg	0.87	0.85	0.86	24000
weighted avg	0.96	0.96	0.96	24000

The confusion matrix of Random Forest Model is shown in **Figure 2.**

Predicted / True	0	1	2	3	4	5	6	7	8	9	All
0	119	12	9	15	26	0	0	0	0	0	181
1	6	125	8	24	13	0	0	0	0	0	176
2	7	12	956	212	24	1	3	19	1	0	1235
3	15	15	121	2964	56	9	0	25	6	3	3214
4	23	6	20	53	1640	1	0	2	1	0	1746
5	0	0	6	31	10	5441	0	2	2	0	5492
6	0	0	0	1	2	0	10829	0	0	0	10832
7	0	3	16	38	10	2	0	917	1	0	987
8	0	0	0	12	13	0	0	2	98	0	125
9	0	0	0	0	1	0	0	1	0	10	12
All	170	173	1136	3350	1795	5454	10832	968	109	13	24000

Figure 2: Confusion matrix report of Random Forest Model

Comparison with Benchmarks

In order to get an answer to the Second research question, the performance of the suggested model was evaluated against a benchmark study.

Study-1 [25]

In order to create an effective Intrusion Detection System (IDS), the research proposes a novel hybrid data optimization approach that is called DO IDS. This approach comprises two critical processes, data sampling and feature selection, to construct an effective IDS. During the process of data sampling, Isolation Forest (iForest) is used in order to get rid of outliers, genetic algorithm (GA) is used to improve the sampling ratio, and Random Forest (RF) classifier is used as the assessment criteria to get the best training dataset possible. During the process of feature selection, both GA and RF are utilized once more in order to identify the best possible feature subset. In the end, an RF-based intrusion detection system is built utilizing the optimal training dataset obtained via data sampling and the features chosen by feature selection. The experiment is carried out using the UNSW- NB15 dataset, and the results show that the suggested model performs better than other methods when it comes to identifying prevalent behaviors. In this particular research endeavor, the random forest model had an accuracy of 86.5%.

Study-2 [3]

This study uses the recently released UNSW-NB15 dataset from the University of New South Wales, which includes network traffic data from nine types of network intrusions. The model used in this study was Random Forest. The accuracy for Random Forest model was 85.7%.

Performance of Proposed Model

Among the above studies, the proposed model outperforms all of them in term of accuracy, precision and recall metrics. Therefore the proposed model is the best among these studies for intrusion detection. Whole comparison is shown in **Table 6.**

Table 6: Comparison of proposed model with similar Studies

Works	Model	Multi Classification Accuracy
"Building an Effective Intrusion Detection System by Using Hybrid Data Optimization Based on Machine Learning Algorithms" [25]	Random Forest	86.5%
"Network intrusion detection using oversampling technique and machine learning algorithms" [3]	Random Forest	95.1%
Proposed Model result	Random Forest	96.2%

Conclusions

In order to identify possible flaws in network security, this research makes use of Random Forest model. The performance of the model that was suggested has been analyzed and compared using the UNSW-NB15 dataset. The models that have been presented use a variety of pre-processing approaches, such as standardizing the data and selecting the features that are most significant by employing the procedure known as the feature selection method. We are able to evaluate the usefulness of the selected features and data standardization procedures by applying them to classification models that are based on Random Forest. The findings demonstrated that the choice of feature selection approach utilized by the Extra Tree Classifier contributed significantly to the overall improvement in accuracy. According to the results of the evaluations, it is possible to conclude that the classification model performed satisfactorily on the UNSW-NB15 dataset in terms of accuracy, precision, recall, and F1-score metrics.

References

[1] A. A. Olayemi, Alesa b, "A Machine Learning Approach for Information System Security," *Int. J. Inf. Comput. Secur.* vol. 16, no. 12, pp. 91–101, 2018.

[2] M. I. Alghamdi, "Survey on Applications of Deep Learning and Machine Learning Techniques for Cyber Security," *Int. J. Interact. Mob. Technol.*, pp. 210–224, 2020.

[3] A. Hameed and N. Z. Bawany, "Network intrusion detection using oversampling technique and machine

learning algorithms," *PeerJ Comput. Sci. 8e820*, pp. 1–19, 2022, doi: 10.7717/peerj-cs.820.

[4] J. W. Mikhail, J. M. Fossaceca, and R. Iammartino, "A Semi-Boosted Nested Model With Sensitivity-Based Weighted Binarization for Multi-Domain Network Intrusion Detection," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 3, 2019.

[5] S. Thaseen and C. A. Kumar, "An analysis of supervised tree based classifiers for intrusion detection system," *Proc. 2013 Int. Conf. Pattern Recognition, Informatics Mob. Eng. PRIME 2013*, pp. 294–299, 2013, doi: 10.1109/ICPRIME.2013.6496489.

[6] S. Wagh, A. ali shah, S. Kishor Wagh, V. K. Pachghare, and S. R. Kolhe, "Survey on Intrusion Detection System using Machine Learning Techniques Cite this paper Analysis of Machine Learning Techniques for Int rusion Det ect ion Syst em: A Review Survey on Intrusion Detection System using Machine Learning Techniques," *Int. J. Comput. Appl.*, vol. 78, no. 16, pp. 975–8887, 2013.

[7] O. Y. Al-Jarrah, A. Siddiqui, M. Elsalamouny, P. D. Yoo, S. Muhaidat, and K. Kim, "Machine-learning-based feature selection techniques for large-scale network intrusion detection," *Proc. - Int. Conf. Distrib. Comput. Syst.*, pp. 177–181, 2014, doi: 10.1109/ICDCSW.2014.14.

[8] S. Choudhury and A. Bhowal, "Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection," *2015 Int. Conf. Smart Technol. Manag. Comput. Commun. Control. Energy Mater. ICSTM 2015 - Proc.*, pp. 89–95, 2015, doi: 10.1109/ICSTM.2015.7225395.

[9] N. Gao, L. Gao, Q. Gao, and H. Wang, "An Intrusion Detection Model Based on Deep Belief Networks," *Proc. - 2014 2nd Int. Conf. Adv. Cloud Big Data, CBD 2014*, pp. 247–252, 2015, doi: 10.1109/CBD.2014.41.

[10] Y. Dong, R. Wang, and J. He, "Real-time network intrusion detection system based on deep learning," *Proc. IEEE Int. Conf. Softw. Eng. Serv. Sci. ICSESS*, pp. 1–4, 2019, doi: 10.1109/ICSESS47205.2019.9040718.

[11] K. Alrawashdeh and C. Purdy, "Toward an online anomaly intrusion detection system based on deep learning," *Proc. - 2016 15th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2016*, pp. 195–200, 2017, doi: 10.1109/ICMLA.2016.167.

[12] V. K. Rahul, R. Vinayakumar, K. Soman, and P. Poornachandran, "Evaluating Shallow and Deep Neural Networks for Network Intrusion Detection Systems in Cyber Security," *2018 9th Int. Conf. Comput. Commun. Netw. Technol. ICCCNT 2018*, no. November, pp. 1–6, 2018, doi: 10.1109/ICCCNT.2018.8494096.

[13] A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," *IEEE Commun. Surv. Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016, doi: 10.1109/COMST.2015.2494502.

[14] N. Moustafa and J. Slay, "The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set," *Inf. Secur. J.*, vol. 25, no. 1–3, pp. 18–31, 2016, doi: 10.1080/19393555.2015.1125974.

[15] T. Mehmood and H. B. M. Rais, "Machine Learning Allgorriitthms IIn Conttextt Off Inttrusiion Dettecttiion," *Comput. Inf. Sci. (ICCOINS), 2016 3rd Int. Conf.*, pp. 369–373, 2016.

[16] M. M. Baig, M. M. Awais, and E. S. M. El-Alfy, "A multiclass cascade of artificial neural network for network intrusion detection," *J. Intell. Fuzzy Syst.*, vol. 32, no. 4, pp. 2875–2883, 2017, doi: 10.3233/JIFS-169230.

[17] F. Ertam, I. F. Kilinçer, and O. Yaman, "Intrusion detection in computer networks via machine learning algorithms," *IDAP 2017 - Int. Artif. Intell. Data Process. Symp.*, 2017, doi: 10.1109/IDAP.2017.8090165.

[18] M. Belouch, S. El, and M. Idhammad, "A Two-Stage Classifier Approach using RepTree Algorithm for Network Intrusion Detection," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 6, pp. 389–394, 2017, doi: 10.14569/ijacsa.2017.080651.

[19] N. T. Van, T. N. Thinh, and L. T. Sach, "An anomaly-based network intrusion detection system using Deep learning," *Proc. - 2017 Int. Conf. Syst. Sci. Eng. ICSSE 2017*, pp. 210–214, 2017, doi: 10.1109/ICSSE.2017.8030867.

[20] N. Koroniotis, N. Moustafa, E. Sitnikova, and J. Slay, "towards developing network forensic mechanism

for botnet activities in the IoT based on machine learning techniques," *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng. LNICST*, vol. 235, pp. 30–44, 2018, doi: 10.1007/978-3-319-90775-8_3.

[21] & E. O. B. Chibuzor John Ugochukwu, "An Intrusion Detection System Using Machine Learning Algorithm," *Int. J. Comput. Sci. Math. Theory*, vol. 4, no. 1, pp. 39–47, 2018, doi: 10.1007/978-981-19-2069-1_10.

[22] I. Ahmad, M. Basheri, M. J. Iqbal, and A. Rahim, "Performance Comparison of Support Vector Machine, Random Forest, and Extreme Learning Machine for Intrusion Detection," *IEEE Access*, vol. 6, pp. 33789–33795, 2018, doi: 10.1109/ACCESS.2018.2841987.

[23] K. Peng, V. C. M. Leung, L. Zheng, S. Wang, C. Huang, and T. Lin, "Intrusion detection system based on decision tree over big data in fog environment," *Wirel. Commun. Mob. Comput.*, vol. 2018, pp. 1–11, 2018, doi: 10.1155/2018/4680867.

[24] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in *ICISSP 2018 - Proceedings of the 4th International Conference on Information Systems Security and Privacy*, 2018, pp. 108–116. doi: 10.5220/0006639801080116.

[25] J. Ren, J. Guo, W. Qian, H. Yuan, X. Hao, and H. Jingjing, "Building an Effective Intrusion Detection System by Using Hybrid Data Optimization Based on Machine Learning Algorithms," *Secur. Commun. Networks*, vol. 2019, pp. 1–11, 2019, doi: 10.1155/2019/7130868.