ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 9, Issue 12, December 2022

A Review on Intrusion Detection Using Machine Learning Techniques

[1] Dhoma Harshavardhan Reddy, [2] Anupriya Elumalai

[1] B.Tech, Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad, India.
[2] B.E, M.Tech, Ph.D, Professor and Head, Computer Science and Engineering, MLR Institute of Technology, Hyderabad, India.
Corresponding Author Email: [1] dhreddy2001@gmail.com, [2] anu.ibrict@gmail.com

Abstract— *An essential tool for monitoring and identifying intrusion threats is the intrusion detection system (IDS). As a result, intrusion detection systems monitor network traffic heading through computer systems to detect for malicious activity and recognized dangers, and send alerts. With a focus on datasets, ML methods, and metrics, this study tries to analyse recent IDS research using a Machine Learning (ML) approach. To make sure the model is suitable for IDS application, dataset selection is crucial. The efficiency of the ML method can also be impacted by the dataset structure. As a result, the choice of ML algorithm depends on the dataset's structure. Metric will then offer a quantitative assessment of ML algorithms for a given dataset. In addition True Positive Rate (TPR), False Positive Rate (FPR) and accuracy, are the three metrics for IDS performance evaluation that are most frequently utilized. This is understandable given that these metrics offer crucial cues that are crucial to IDS performance. A clear path and direction for future study has been provided by the discussion and comparison of the results from various works.*

Index terms—*Classifiers, Intrusion Detection System (IDS), Machine Learning, Metric.*

I. INTRODUCTION

The Intrusion Detection System (IDS) detects malicious or threat activity [1]. The intrusion detection methods could depend on anomaly or signature detection. The network's packet flow is monitored through signature-based detection, which compares and configured known threats signatures. When compared with events that reveal a deviation from the normal user parameters, the anomaly detection method can identify assaults [2]. After malicious behavior is detected in a network, the IDS generates records and a network administrator is notified [3].

The major drawback is due to earlier intrusion detection systems that require constant updating of their databases of known attack signatures because hackers frequently find ways to exploit network activities [4]. Machine learning made it feasible to undertake anomaly detection, which is the process of identifying unknown attacks by contrasting occurrences that deviate from normal user behavior with genuine user characteristics. A number ML approaches have been used over time with the hopes of raising IDS's predicting accuracy, decreasing false positives, and improving detection rates. We will examine the effectiveness of single, hybrid, and ensemble ML approaches in IDS in this review of the literature [5]. Fig 1. depicts the block diagram of IDS.

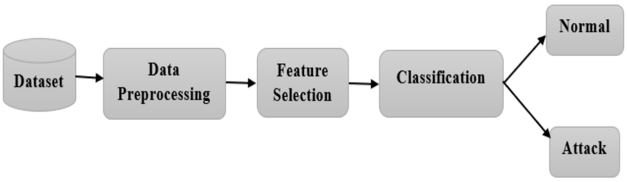

Fig 1: Intrusion Detection System Block Diagram

ML uses a statistical modeling technique to learn historical data patterns before predicting the most likely outcome with new data. As a result, IDS has been subjected to ML algorithm employing anomaly-based methodology. The difficulty lies in developing a system with a low rate of false alarms and high accuracy. As a result, this study's objective is to investigate at current IDS research with an ML strategy, with a focus on datasets, ML algorithms, and metrics. To make sure the model is suitable for IDS application, dataset selection is crucial. The efficiency of the ML method can also be impacted by the dataset structure. As a result, the choice of ML algorithm depends on the dataset's structure. Metric will then offer a quantitative assessment of ML algorithms for a given dataset.

II. RELATED WORKS

Alkasassbeh and Almseidin [6] employed Three categorization strategies to address the problems with IDS that use an fuzzy clustering and artificial neural network commonly experience low accuracy while responding to infrequent attacks. Bhavani et al. [7] constructed an IDS using a single ML classifier employing random forest and decision tree methods on the dataset of KDD-NSL. The superior result is with an accuracy of 95.323% given by the random classifier. The proposed work did not address low detection rates or false positive rates. ML methods were utilized to detect network infiltration. Logistic regression, Support Vector Machine (SVM), Decision tree (DT), and random forest (RF) are the techniques employed in the study [8] used dataset is KDD-NSL. According to the study, a RF classifier gives the IDS its highest performance.

Marzia Z. and Chung-Horng L. [9] implemented an ensemble-based approach IDS in which the outcomes are combined of various supervised and unsupervised ML algorithms using voting classifiers. The work improves the current IDS performance and accuracy. They used the Kyoto2006+ dataset, which is more promising than the KDDCup '99 dataset, which is the most practical given its age. This enables their job to reach a given level of accuracy, but in a small number of circumstances, the Recall of the result is quite poor, indicating large values of false negative rate (FPR). Dutt I. et al. [10] proposed a hybrid IDS approach where the high detection rate in this study was made possible by the anomaly detection technique's ability to identify patterns of intrusions as attacks when they managed to evade abuse detection that the accuracy reached a substantial value of 92.65%.

A study by Verma et al. [11] demonstrates that there is opportunity for improvement in anomaly-based IDS, particularly with regard to the false positive rate. Adaptive boosting (AdaBoost) and the extreme gradient boosting (XGBoost) learning techniques of the NSL-KDD were used. Although an accuracy of 84.253 was achieved, it is still necessary to increase performance by using ensemble or hybrid ML classifiers. Using various ML methods, and feature selection was carried out using the wrapper technique in [12]. In comparison to earlier works that used the same dataset, this accuracy improvement was comparatively superior. The suggested work by Zhou et al. [13] introduced a unique IDS that benefits from the combination of ensemble classifier and feature selection, which improves efficiency and high accuracy intrusion detection. NSL-KDD dataset as well as the two most current datasets, AWID and CIC-IDS2017, were used in the research. The CFS-BA based approach was employed for feature selection. Disha and Waheed [14] proposed Gini Impurity-based Weighted Random Forest (GIWRF) feature selection. Saif et al. [15] established hybrid IDS with ML for IoT based healthcare. Dang [16] studied advanced ML for IDS. Zhang et al. [17] applied IDS with ML in a UAS/RADAR System.

The research's findings demonstrated that, when employing the UNSW NB-15and NSL-KDD datasets, respectively, a machine learning technique in terms of misclassification gaps of 1.19% and 1.62%. Rajagopal et al. suggested a stacking ensemble method using heterogeneous datasets. The most recent dataset from UNSW NB-15 and UGR '16 was used for the study [18]. A hybrid network-based IDS was proposed by Perez D. et al. employing several hybrid machine learning approaches that operate on the NSL-KDD dataset [19]. Neural networks, a supervised ML technique, were integrated with feature selection and K-Means clustering, an unsupervised ML technique. K-means clustering and SVM were combined in another arrangement. The outcomes unmistakably demonstrated how combining such unsupervised and supervised ML techniques enhances IDS performance. The most accurate results are obtained when feature selection is combined with K-means and SVM. To decrease the false positive rate, more hybrid-based models must be created.

III. RESEARCH METHODOLOGY

We selected certain crucial criteria to ensure that we only review studies of interest. First and foremost, the articles selected have been published in 2015 till present. This is to make sure that we only receive the current studies, ensuring that our research is current and not outdated. Second, the article needs to be published in a conference or scientific journal. This is done to ensure that the content is accurate and has undergone peer review and approval. Thirdly, ML for IDS must be used in the article. Fig. 2 depicts the total number of articles published in the year 2015-2022.

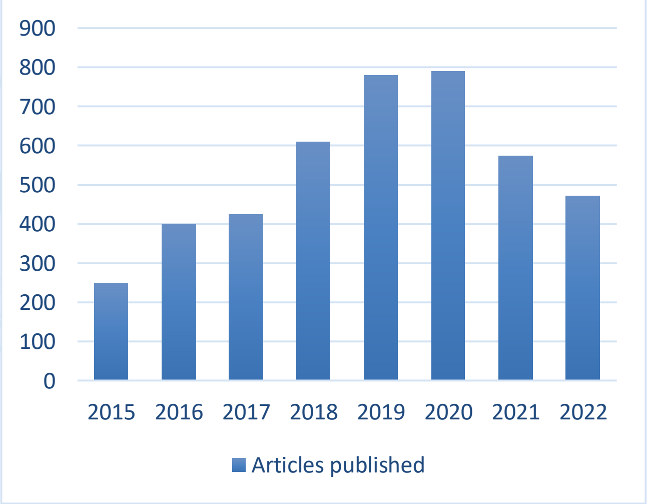

Fig.2 : Published articles in the year 2015-2022

A. Intrusion Detection Systems(IDS)

There are many different types of intrusion detection systems since network settings might vary. In terms of price, setup, and detection rate, each form of IDS has unique benefits and drawbacks. The main function of an IDS is to analyze and monitor network traffic and determine whether it is normal or abnormal. IDS are divided into distributed, host-based, hypervisor-based, network-based, and IDS based on where they are deployed. Fig 3 displays the flow diagram of IDS types.

Host based IDS

In host-based IDS, the intrusive events are discovered by gathering data from a specific host and examining it with system logs and operating system audit records. If the system behaves differently, the network manager is notified and informed that the system or network is being attacked [20]. The main disadvantage of HIDS is that greater storage capacity is needed to accommodate audit records and historical events.

Network based IDS

To identify any potential intrusions, the Network Based IDS records and analyses all network traffic. The transport layer header of the captured packets' Internet Protocol packets are examined as part of the intrusion detection analysis. Both anomaly and signature-based detections will be used by NIDS [21]. The main flaw is that it is unable to decrypt encrypted packets, making it impossible to identify an assault.

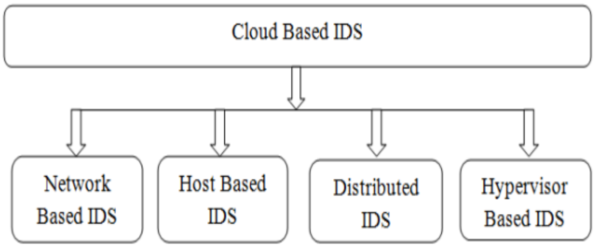

Fig. 3: Types of IDS

Hypervisor based IDS

Hypervisors are used to facilitate communication between virtual machines in a distributed cloud environment. The system installed at the hypervisor layer is an IDS based on a hypervisor. It aids in the detection of irregularities between the hypervisor and virtual machine.

Distributed IDS

Numerous IDS can be found in distributed IDS (DIDS). To identify anomalies or intrusive behavior in such networks, DIDS is implemented in large-scale networks. Communication with the full distributed server is possible through DIDS. The detection component and correlation manager are the two main parts of DIDS. Both the processing server and the host system have DIDS installed.

IV. MACHINE LEARNING

Eleven categories can be used to group ML algorithms. Naive Bayes is the most often used algorithm in this area. The decision tree root node, which is the best predictor, is the beginning of a tree-like structure. continues moving forward. Finding traits that are crucial to the outcome is the goal of dimensional reduction. This will eliminate extraneous and pointless features. The majority of the work is done in the pre-processing stage. Principal Component Analysis is the algorithm that is most widely used (PCA) Memory-based learning is another name for instance-based learning. This group of algorithms looks for the examples, or training data, that are the most comparable to the new information.

The k-NN technique is the most used in this category (kNN). Data points that are near to one another are grouped together to form a cluster. The unsupervised learning strategy, which does not require labelled data, benefits greatly from this class of algorithm. Among these algorithms, logistic regression is the most widely used. The neuron, a type of brain cell that makes up the biological neural network, is the model for neural networks. In order to make its forecast, this category analyses the data for patterns. Typically, making a decent prediction would require a lot of data. Perceptron is the most often used algorithm in this group. The two most common techniques are bagging and boosting. The research articles that were found in this investigation are displayed in Table 1.

Table I. Displays measurement measures for feature selection algorithms used in machine learning classification approaches

Ref No.	Dataset	Classification and Feature selection	Accuracy (%)
[22]	NSL-KDD; KDD cup 99	C4.5, Random Forest, correlation-based Bat-Algorithm	99.14, 97.56
[23]	KDD Cup99	PSO and Correlation; SVM, k-NN and Naive Bayes	99.93
[24]	KDD Cup 99	Bayesian, C4.5 Decision Tree, PCA based feature selection	DoS=99.98

[25]	KDD Cup 99	Genetic Algorithms and Particle Swarm Optimizations;
		Rule Induction, k- NN, Decision Tree, Naive Bayes
			PSO =99.26%.,
			GA= 99.70%

[8]	KDD-NSL	Random Forest (RF), Decision Tree (DT), Support Vector Machine (SVM), Logistic Regression (LR),	DT=72=303,
			RF=73.784,
			LR=68.674,
			SVM=71.779

[7]	NSL-KDD	Decision Tree (DT), Random Forest (RF)	DT=81.868,
			RF=95.323

[26]	NSL-KDD	Random-Forest, Information gain; J48 and Partial Decision List (PART)	RF= 99.78

[9]	Kyoto2006+	K-Nearest Neighbor (KNN), K-Means, Ensemble, Fuzzy C-Means (FCM), Support Vector Machine (SVM), Radial Basis function (RBF), Naïve bayes (NB),	KNN=97.54.
			RBF=97.54,
			NB=96.72,
			Ensemble=96.72,
			FCM=83.60,
			SVM=94.26,
			K-Means=83.60

[27]	UNSW NB-15,
	NSL-KDD
		(K- Nearest Neighbor (KNN)),
		Single Machine Learning Classifier Ensemble Technique (Random Committee (RC))
			UNSW NB-15-RC=98.955,
			KNN=97.3346;
			NSL-KDD-RC=99.696,
			KNN=98.727

[28]	NSLKDD	SVM, Recursive Feature Elimination (RFE) with SVM, Random Forest	U2R=99.9
[29]	NSL KDD	Combination of Binary PSO, The binary-based PSO; Standard PSO and SVM	99.10
[30]	CICIDS2017	Based Feature selection, Using Fisher Score Algorithm Payload Classifier and MLP	95.2
[31]	NSL-KDD	Random forest (RF) based ensemble classifier	99.67
[32]	NSL-KDD	SVM and Rule Based Classification, Optimal Feature Selection algorithm (IG);	99.25
[18]	UNSW NB-15, UGR '16	Stacking Ensemble: LR, KNN, SVM and RF	UGR '16=98.71 UNSW NB=15-94.00
[33]	NSL-KDD	RNN-LSTM, Bayes Classifier, Decision Tree, Ensemble, Random Forest	Ensemble=85.20

[34]	KDDCup '99	Support Vector Machine (SVM), Genetic Algorithm (GA),
		Hybrid Model
			SVM=94.8000,
			GA=84.0333,
			Hybrid (GA+SVM)=98.333%

[35]	KDD Cup 99	Cuttlefish algorithm and linear correlation coefficient algorithm; ID3, LSSVM	95.03

V. DATASET

The major constituent in training ML to recognize anomaly threats is a dataset. The study's analysis reveals, however, that many researchers continue to use the out-of-date datasets NSL-KDD and KDDCup99, they are widely criticized for being outdated and unnecessary given the state of the network infrastructure. The landscape of network infrastructure is evolving as a result of information technology's rapid development and innovations, including social media, the Internet of Things and the cloud computing. These adjustments are the driving force behind the threat attack's own change. Because the dataset used does not reflect the current danger or infrastructure, many research findings that show high accuracy are considered to be overblown. Every occurrence is classified as regular or certain kind of attack. These attacks can be divided into one of the four groups outlined below: DoS, probe, R2L and U2R [36].

- Denial of Service (DoS): This sort of attack prevents the lawful use of network resources by overloading computational resources or using all available bandwidth.
- Probe: Before beginning an actual attack, this kind of attack probes the target system to gather information.
- Remote to Local (R2L): In this scenario, an attacker sends a packet to a remote machine across a network without having an account there, then uses the

machine's weaknesses to gain access locally as the system's user.

- User to Root (U2R): In this instance, an attacker initially takes control of a regular user account on the system and then uses that account to exploit system vulnerabilities to become the system's root user.

The KDDCup99 dataset was developed in 1999, but the NSL-KDD dataset was created in 2009. In addition to the imbalanced occurrences and the diversity of assault types, by eliminating unnecessary entries, NSL-KDD attempts to enhance the KDDCup99 dataset [2]. It still carries over the dataset's underlying flaw, though. KDDCup99 contains a lot of flaws [37]. New threats will emerge as a result of these changes. The other two well-liked datasets are UNSW-NB15 and ISCX 2012. ISCX 2012. This dataset includes data from 7 days with the labels "normal" or "attack" (two). The dataset only provides binary categorization because it lacks a classification of the different types of attacks. This dataset, however, is no longer accessible. This is as a result of the center's creation of the CICIDS2017 dataset [38]. The Canadian Institute for Cybersecurity is now the name of the facility (CIC). Unfortunately, at the time of this investigation, no article utilizing this additional dataset could be located. The UNSW-NB15 dataset, utilizing IXIA PerfectStorm to produce 9 different types of attacks, is another well-known one. Analysis, fuzzers, DoS, backdoors, generic, exploits, shellcode, reconnaissance, and worms are the 9 different forms of attacks mentioned above. There are two labels for each of the dataset's 47 features [39].

VI. METRIC

Metrics are used to evaluate how well a ML performs on a certain dataset. It offers a means of comparison, allowing one to ascertain which method is more effective overall. A confusion matrix table can be used to derive the majority of metrics, as illustrated in Table 2 below.

Table II. Confusion matrix

		Predicted Class	
		Positive (Attack)	Negative (Normal)
Actual Class	Positive (Attack)	True Positive(TP)	False Negative (FN)
	Negative (Normal)	False Positive(FP)	True Negative (TN)

The most often used metric is accuracy. This statistic gives the proportion of accurately anticipated results to all observed results [40]. As a result, in this study, it serves as the main metric for comparison. Equation 1 displays the formula:

$$\frac{TN+TP}{TN+TP+FP+FN}$$

There are three additional names for True Positive Rate (TPR), but they all refer to the same metric namely Sensitivity, recall, and detection rate. This statistic measures the proportion of accurately predicted positive outcomes to observations that were in fact positive [40]. The following equation 2 displays the formula:

$$\frac{TP}{FN+TP}$$

False Alarm Rate (FAR), fall-out, and False Positive Rate (FPR) are other names for the FPR. This metric measures the proportion of incorrectly projected positive results to actual negative findings [40]. The following equation 3 illustrates the formula:

$$\frac{FP}{FP+TN}$$

Specificity is another name for True Negative Rate (TNR). The ratio of accurately predicted negative outcomes to actual negative observations is this metric [40]. The following equation 4 illustrates the formula:

$$\frac{TN}{TN+FP}$$

Miss rate is another name for false negative rate (FNR). This metric measures the proportion of incorrectly projected negative results to actual positive findings [41]. Below, in equation 5, is the formula:

$$\frac{FN}{TP+FN}$$

Precision is defined as the ratio of successfully predicted positive outcomes to properly predicted positive outcomes [40]. The following equation 6 illustrates the formula:

$$\frac{TP}{TP+FP}$$

F-score is another name for F-measure. This statistic offers performance assessment based on recall and precision [42]. The following equation 7 illustrates the formula:

$$\frac{2TP}{2TP+FN+FP}$$

Time is the unit of measurement for efficiency. There are two possible measuring phases. Throughout the training phase, one measurement is taken, and when testing is in progress, another. Other metrics were also detected in this study, but they were less prevalent, so we won't talk about them here.

According to this survey, more than 80% of studies used two metrics. Accuracy and TPR are these. The degree of accuracy gives a decent indicator of how accurately the algorithm can forecast the right result. This is significant since it demonstrates how much confidence may be placed in the result's accuracy. The TPR or detection rate, gives a measure of how well the algorithm can identify an infiltration attempt. This measure is significant because An IDS's objective is to recognize attacks.

FPR is a different metric that was employed in more than 50% of studies. False Alarm Rate is another term for this measurement (FAR). This rating reveals how many false alarms the algorithm will generate. This is crucial because it illustrates how much more work must be done after the IDS to more sort out these false alarm findings. Percentage of metric measures of 2015-2022 is depicted in Fig. 4.

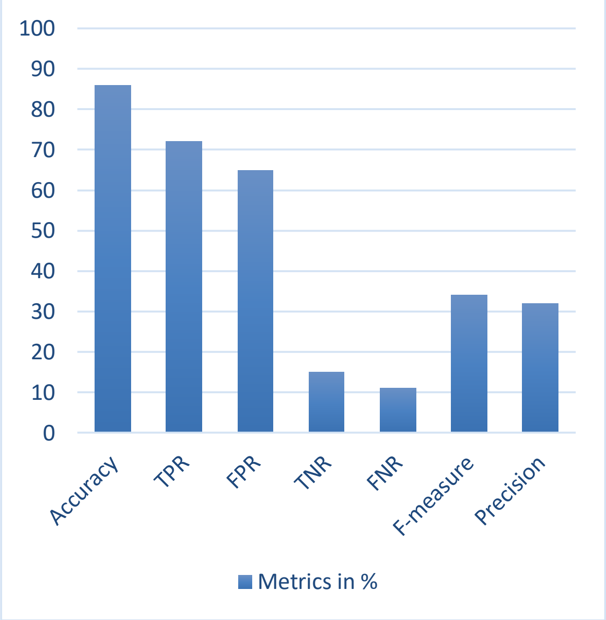

Fig. 4 : Percentage of metrics in IDS

VII. CONCLUSION

Researchers in IDS are paying close attention to soft computing strategies. This is due to the fact that this approach is simple to use and frequently yields better results than a single program. The majority of academics are concentrating on the IDS classification since it helps identify known intrusion attacks. It could be difficult to detect abnormal intrusions, such as new or modified intrusion attacks. Therefore, clustering method should be taken into consideration for future development in order to create a more reliable IDS. Despite being over 20 years old, the two most popular datasets are KDDCup99 and its variation NSL-KDD. While intrusion threats continue to develop alongside new technology and human behaviors, this ongoing process may cause IDS to make static progress. IDS will eventually be rendered obsolete as a cyber security technology as a result of this circumstance. It is crucial to create fresh datasets that accurately reflect the software and hardware configuration of the existing environment. CICIDS2017, the most recent dataset that is openly accessible, should be investigated. Accuracy, TPR, and FPR are the three metrics for IDS performance evaluation that are most frequently utilized. This is understandable given that these metrics offer crucial cues that are crucial to IDS performance. It is possible to merge all three measures into one statistic in order to speed up the evaluation procedure.

REFERENCES

[1] K. A. Tait et al., "Intrusion Detection using Machine Learning Techniques: An Experimental Comparison," 2021 International Congress of Advanced Technology and Engineering (ICOTEN), 2021, pp. 1-10, doi: 10.1109/ICOTEN52080.2021.9493543.

[2] E. K. Viegas, A. O. Santin, and L. S. Oliveira, "Toward a reliable anomaly-based intrusion detection in real-world environments," Comput. Networks, vol. 127, pp. 200-216, 2017.

[3] Y. Zhu, J. Liang, J. Chen, and Z. Ming, "An improved NSGA-III algorithm for feature selection used in intrusion detection, " Knowledge-Based Syst., vol. 116, pp. 74 – 85,2017.

[4] A. H. Hamamoto, L. F. Carvalho, L. D. H. Sampaio, T. Abrão, and M. L. Proença, "Network Anomaly Detection System using Genetic Algorithm and Fuzzy Logic," Expert Syst. Appl., vol. 92, pp. 390402,2018.

[5] W. L. Al-Yaseen, Z. A. Othman, and M. Z. A. Nazri, "Multi-level hybrid support vector machine and extreme learning machine based on modified Kmeans for intrusion detection system, " Expert Syst. Appl., vol. 67, pp. 296-303, 2017.

[6] Alkasassbeh and Almseidin, “Machine Learning Methods for Network Intrusions,” International Conference on Computing, Communication (ICCCNT). Arxiv, 2018.

[7] T. T. Bhavani, M. R. Kameswara and A. R. Manohar, “Network Intrusion Detection System using Random Forest and Decision Tree Machine Learning Techniques,” International Conference on Sustainable Technologies for Computational Intelligence (ICSTCI) , Springer, pp. 637-643, 2020.

[8] R. Ponthapalli, et al., “Implementation of Machine Learning Algorithms for Detection of Network Intrusion,” International Journal of Computer Science Trends and Technology (IJCST), pp. 163-169, 2020.

[9] Z. Marzia and L. Chung-Horng, “Evaluation of Machine Learning Techniques for Network Intrusion Detection,” IEEE, pp. 1- 5, 2018.

[10] I. Dutt, et al., “Real Time Hybrid Intrusion Detection System,” International Conference on Communication, Devices and Networking (ICCDN), Springer, pp. 885-894, 2018.

[11] P. Verma, k. Shadab, A. Shayan and B. Sunil, “Network Intrusion Detection using Clustering and Gradient Boosting,” International Conference on Computing, Communication and Networking Technologies (ICCCNT), IEEE, pp. 1-7, 2018.

[12] A. M. Kazi and R. Mahbubur, “Network Intrusion Detection using Supervised Machine Learning Technique with feature selection. International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST). (pp. 643-646), 2019. IEEE.

[13] Z. Yuyang, C. Guang, J. Shanqing and D. Mian, “Building an Efficient Intrusion Detection System Based on Feature Selection and Ensemble Classifier,” Computer Networks, 2019 Doi: https://doi.org/10.1016/j.comnet.2020.107247

[14] R. A. Disha and S. Waheed, “Performance analysis of machine learning models for intrusion detection system using Gini Impurity-based Weighted Random Forest (GIWRF) feature selection technique,” Cybersecurity, vol. 5(1), pp. 1-22, 2022.

[15] S. Saif, P. Das, S. Biswas, M. Khari and V. Shanmuganathan, “HIIDS: Hybrid intelligent intrusion detection system

empowered with machine learning and metaheuristic algorithms for application in IoT based healthcare," Microprocessors and Microsystems, pp. 104622, 2022.

[16] Q. V. Dang, "Using Machine Learning for Intrusion Detection Systems. Computing and Informatics, vol. 41(1), pp. 12-33, 2022.

[17] R. Zhang, J. P. Condomines and E. Lochin, "A Multifractal Analysis and Machine Learning Based Intrusion Detection System with an Application in a UAS/RADAR System," Drones, vol. 6(1), pp. 21, 2022.

[18] S. Rajagopal, P. K. Poornima and S. H. Katiganere, "A Stacking Ensemble for Network Intrusion Detection using Heterogeneous Datasets," Journal of Security and Communication Networks. Hindawi, pp. 1-9, 2020.

[19] P. Deyban, A. A. Miguel, P. A. David and S. Eugenio, "Intrusion detection in computer networks using hybrid, machine learning techniques," XLIII Latin American Computer Conference (CLEI), IEEE, (pp. 1-10), 2017.

[20] M. Yasir, H. Umme, A. Muhammad and R. M. Shibli, "Intrusion Detection System in Cloud Computing: Challenges and Opportunities", 2nd National Conference on Information Assurance (NCIA), pp. 59-66, 2013.

[21] M. Preeti, S. P. Emmanuel, V. Vijay and T. Udaya, "Intrusion detection techniques in cloud environment: A survey", Journal of Network and Computer Applications, Vol.77, pp.18–47, 2017.

[22] Shailendra Sahu, B.M. Mehtre, Network Intrusion Detection System Using J48Decision Tree, in: International Conference on Advances in Computing, Communications and Informatics (ICACCI), IEEE, 2015.

[23] Tohari Ahmad, Mohammad Nasrul Aziz, Data preprocessing and feature selection for machine learning intrusion detection systems, ICIC Int. Express Lett. 13 (2019) ISSN 1881-803X.

[24] B. Selvakumar, K. Muneeswaran, Firefly algorithm based feature selection for network intrusion detection, Computer Securit, 81 (2019) 148–155. Trends in Engineering & Technology [ISSN: 2158-5555, March 2011]

[25] Iwan Syarif, Feature selection of network intrusion data using genetic algorithm and particle swarm optimization, EMITTER Int. J. Eng. Tech. 4 (2) (2016) ISSN: 2443-1168.

[26] Manal Abdullah, Arwa Alshannaq, Asmaa Balamash, Enhanced intrusion detection system using feature selection method and ensemble learning algorithms, Int. J. Computer Sci. Information Security, 16 (2) (2018).

[27] Maniriho et al. (2020). Detecting Intrusions in Computer Network Traffic with Machine Learning Approaches. International Journal of Intelligent Engineering and Systems. INASS. (433-445).

[28] Ripon Patgiri, Udit Varshney, Tanya Akutota, Rakesh Kunde, An investigation on intrusion detection system using machine learning, in: IEEE Symposium Series on Computational Intelligence SSCI, 2018.

[29] Mahmoud M. Sakr, Medhat A. Tawfeeq, Ashraf B. El-Sisi, Network Intrusion Detection System based PSO-SVM for Cloud Computing, I. J. Computer Network Information Security 3 (2019) 22–29.

[30] Ustebay, Z. Turgut, M.A. Aydin, Intrusion detection system with recursive feature elimination by using random Forest and deep learning classifier, in: 2018 international congress on big data, deep learning and fighting cyber terrorism (IBIGDELFT), 2018, pp. 71–76.

[31] Nabila Farnaaz and M.A Jabbar. (2016). Random Forest Modeling for Network Intrusion Detection System. International Multi-conference on information processing (IMCIP) 12 (pp. 213-217). Elsevier.

[32] S. Balakrishnan, K. Venkatalakshmi, A. Kannan, A intrusion detection system using feature selection and classification technique, Int. J. Computer Sci. Appl. (IJCSA), vol. 3(4), pp. 145–151, 2016 .

[33] Y. Vinoth and K. Kamatchi, "Anomaly Based Network Intrusion Detection using Ensemble Machine Learning Technique," International Journal of Research in Engineering, Science and Management. IJRESM. pp. 290-296, 2020.

[34] A. Kayvan, Y. Saadiah, R. Amirali and S. Hazyanti, "Anomaly Detection Based on Profile Signature in Network using Machine Learning Techniques," IEEE TENSYMP. pp. 71-76, IEEE, 2016.

[35] Sara Mohammadi, Hamid Mirvaziri, Mostafa Ghazizadeh Ahsaeea, Hadis Karimipour, "Cyber intrusion detection by combined feature selection algorithm," J. Information Security Appl., vol. 44, pp. 80–88, 2019.

[36] I. Sumaiya Thaseen and C. Aswani Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM, " J. King Saud Univ. - Comput. Inf. Sci., vol. 29, no. 4, pp. 462-472, 2017.

[37] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," J. Comput. Sci., vol. 25, pp. 152160, 2018.

[38] A. H. L. and A. A. G. Iman Sharafaldin, "Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization," Proc. 4th Int. Conf. Inf. Syst. Secur. Priv., no. Cic, pp. 108-116, 2018.

[39] N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set), ^{n} 2015 Mil. Commun. Inf. Syst. Conf., no. November, pp. 1 – 6,2015.

[40] S. Shitharth and D. Prince Winston, "An enhanced optimization based algorithm for intrusion detection in SCADA network, " Comput. Secur., vol. 70, pp. 16 – 26,2017.

[41] A. A. Aburomman and M. Bin Ibne Reaz, "A novel weighted support vector machines multiclass classifier based on differential evolution for intrusion detection systems," Inf. Sci. (Ny)., vol. 414, pp. 225246,2017.

[42] A. S. Amira, S. E. O. Hanafi, and A. E. Hassanien, "Comparison of classification techniques applied for network intrusion detection and classification," J. Appl. Log., vol. 24, pp. 109-118, 2017.