Journal of Information Systems Engineering and Management
2025, 10(20s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Anomaly Detection via Rain Optimization Algorithm and Stacked Autoencoder Hoeffding Tree

Shaymaa Abdul Hussein Shnain[1], Zahraa Modher Nabat[2], Mohammed Al Yousef[3]
[1,2,3]*University of Babylon Babylon, Hilla, Iraq*
shaima.almorshedy3@uobabylon.edu.iq ,
zahraa.modher@uobabylon.edu.iq ,
mohammed.alyousef@uobabylon.edu.iq

ARTICLE INFO

Received: 21 Dec 2024

Revised: 27 Jan 2025

Accepted: 15 Feb 2025

ABSTRACT

Networks of Computer are vulnerable to cyberattacks which could affect the mission critical data accessibility, confidentiality, integrity. Anomaly detection became the most basic environment of study because of extend usage range like unusual network traffic manner detection, detection of disease in MRI images, detection of fraud in transactions of credit card. In a lot of real-life anomaly issues of detection, we meet the heterogeneous data including various features' kinds such as categorical and continuous features. Data heterogeneity makes that actually hard for data examples' comparison. In addition, data manners might shift over time in flowing areas. At last, that is difficult foe getting data tags as we get a lot of data every day for classification manually. Autoencoders are a feed-forward neural network kind which could be ordered for performing anomaly detection through learning the stochastic input 'normal' instances' representation and abnormal instances' diagnosis through controlling error of reconstruction in comparison with the predetermined anomaly threshold. This paper concentrates on developing IDS effectiveness by applying proposed Stacked Autoencoder Hoeffding Tree approach (SAE-HT) applying Rain Optimization Algorithm (ROA) for selection of feature. Experiments on the dataset NSL-KDD illustrate that our model multi-classification possesses great performance. In comparison to the other mechanisms of ML in accuracy case, our model performs better than such mechanisms.

Keywords: Anomaly Detection, feature selection, Rain Optimization Algorithm, Autoencoder, Hoeffding Tree.

1- INTRODUCTION

Nowadays, computer system usage is quickly growing and provides details. Offenses of Safety like worms, hackers, viruses, and so on, even raise more rapidly via network. As an Attacker will threaten network credibility, resources, security. Firewall has been the protective initiatives gateway for the past few years. Although, traffic of network which actual valid user/ port has performed is not determined. So, Detection of Intrusion is essential for handling data manipulation by ransom ware. Detection of Intrusion is performed for assigning and supervising tasks done either by network/ personal device [1].

Now, a lot of vulnerabilities recognized are explored through connection of network. That is because of large exposure number to services of Internet. Therefore, different methods and tools are generated and refined for deterring attacks on level of network and not let the attacker in order to get access to that service. Systems of Intrusion detection let containment and response to unauthorized access. In last few years, different tools have been extended and developed. Some heuristics for computer networks detection of intrusion have been tested in recent years. However, with the time passage as well as information systems evolution, several methods lost their effectiveness. Raising network data traffic amount and constant demand for quick responses contributed to that problem [2]. There are two general strategies for that issue: detection of signature (called as detection of misuse), that we search for the patterns signaling anomaly detection, popular attacks, also where we search for deviations from normal behavior [3].

Methods of Anomaly intrusion detection have been improved for dealing with these attacks. Between anomaly detection approaches variety, trees of decision is known to be one of the best algorithms of machine learning for classifying abnormal behaviors.

Trees of Decision are a classifier algorithms kind. Tree of decision is learned top-down with replacing leaves by nodes of test recursively, beginning at root. Whole accessible features are compared and selecting the best one based on several heuristic measure is the way a feature at node is tested. Learners of Classical decision tree are restricted severely in samples' number they can learn from, as they assume that whole samples of training are able to be saved at the same time in memory. Classification of Decision tree applying Hoeffding bound makes tree classification less time consuming. Hoeffding Trees, an incremental, anytime induction algorithm of decision tree is able to learn from extensive streams of data, was improved by Hulten and Domingos. The fact that the small instance is able to be enough to select the optimum feature of splitting is Hoeffding Trees theory. Mathematically, Hoeffding bound supports that opinion with quantifying observations/ samples' amount required for estimating several statistics in prescribed attribute decision/goodness. Hoeffding Trees have sound performance guarantees, the theoretically interesting feature not shared by the other incremental learners of decision tree. Anomaly detection is the basic research topic because of detecting vital information aspect. vital information is able to contain intrusion and the other information of failure [4].

In this paper, data mining is combined with IDS to perform specific task. The task is to distinguish important, successfully covered up info in less amount of time. The paper focuses on improving effectiveness of IDS through using presented Stacked Autoencoder Hoeffding Tree approach (SAE-HT) by using Rain Optimization Algorithm (ROA) [5] for feature selection.

The paper starts with the related work for anomaly detection using machine learning techniques and the advancements in the methods in Section 2. The methodology section describes the details of the models built for anomaly detection in Section 3. The paper follows by details about the dataset used for the analysis in Section 4. The study presents a comparative analysis of multiple machine learning models vs deep learning models in Section 4. The conclusion is placed in Section 5.

2- RELATED WORK

In [6], bad method of traffic detection given support vector data description is presented, that is known as deep Support Vector Data Description (SVDD). Writers integrate convolutional neural network by support vector data description and train model by normal traffic. Attributes of Normal traffic are mapped to the wide dimensions via neural networks, the compact cloud is trained by unsupervised learning having normal features in wide dimensions. Bad traffic is placed outside the cloud, therefore, that varies among bad and normal traffic. Tests illustrates that the model has low level of false alarm and high speed is able to detect new bad traffic effectively.

In [7], hybrid processing model of data is presented for detecting anomaly of network which develops convolutional neural network and Grey Wolf optimization (GWO). For developing proposed model abilities, learning methods of CNN and GWO were increased by: 1) developing basic abilities of population production, exploitation, exploration, 2) modified outlier performance of operation. Such wide variables are known as Improved-GWO (ImGWO) and Improved-CNN (ImCNN). Presented model acts in 2 steps by efficiently detecting network anomalies aim. In first step, ImGWO is utilized for selecting feature for obtaining optimum exchange among 2 objectives like decreasing rate of error and reducing set of feature. In second step, classification of ImCNN is applied for classifying anomalies of network. Proposed model performance is confirmed in artificial datasets and datasets of benchmark (KDD'99, DARPA'98). Achieved outcomes show that anomaly detection model based on cloud is superior to the other improved models (applied for detection of network anomaly) in terms of F-score, positive false rate, accuracy, detection speed. Presented model Averagely illustrates the total development across standard GWO with CNN respectively in terms of detection rate, false positive, and accuracy of 8.25%, 4.08% and 3.62%.

In [8], writers present detection method of unsupervised multivariate anomaly given the Generative Adversarial Networks (GANs), applying Long-Short-Term-Memory Recurrent Neural Networks (LSTM-RNN) as the models of base (like detector, Generator) is done in GAN for recording temporal correlations in time sequences distribution. Instead of independently taking every data stream into consideration, detection framework of multivariate anomaly takes whole variable into consideration at the same time by Multivariate Anomaly Detection with framework of GAN (MAD-GAN) so this is able to record latent interaction among variables. They applying novel anomaly score known

as DR, for taking full advantage of both discriminator and manufacturer generated by GAN to create anomalies with reconstructing and differentiating. They checked our presented MAD-GAN by applying 2 recent sets of data gathered from datasets of Water Distribution (WADI), The Secure Water Treatment (SWaT) real-world CPSs. Experimental outcomes illustrate indicate that presented MAD-GAN is efficient to report anomalies resulted by different cyber-attacks used to such complicated systems of real-world.

In [9], writers present novel method for detection of anomaly in high-performance systems of computing according to method of machine learning (deep), like neural network kind that is known as self-encoder. Basic opinion is training self-encoders set for learning normal (healthy) supercomputer nodes manner, after training, apply them for identifying abnormal conditions. That is hard from last strategies according to learning unusual conditions, due to that they are much smaller sets of data for them (as they are hard in order to be recognized for beginning). They test method on real supercomputer equipped by fine-grained and scalable infrastructure which is able to present wide data number to describe manner of system. Outcomes are very promising: after step of training for learning normal system manner, the technique can detect anomalies which have never been observed by good accuracy before (values among 88% and 96%).

In [10], writers present novel neural network for anomaly detection (known as anomaly network/ AnomalyNet) by deep achievement for learning of dictionary, learning of feature, scattered display in 3 typical neural blocks of processing. Specifically, for learning better attributes, they design the hybrid motion block by feature transfer block to take advantage of noise elimination of background, data defect reduction, motion record. In addition, for solving several existing scattered coding optimizers' demerits (such as non-adaptive updating) and having neural network competencies (such as parallel computing), they design the recurrent neural network for dictionary learning and scattered representation proposing the Iterative Shrinkage/Thresholding Algorithms algorithm (adaptive ISTA) and adaptive ISTA reformulating as novel long short-term memory (LSTM). Based on our knowledge, that is able to be one of the first researches for making connection among s1-solver to LSTM, also might provide novel insight based on model-based optimization and understanding LSTM (or as distinctive programming), also present scattered coding-based anomaly detection. Extensive tests illustrate improved method performance in detecting abnormal events task.

In [11], writers examine by recognizing system of anomaly in desired networks in interactive environment letting system to actively communicate with human expert in producing questionnaires' restricted amount on anomalies based on truth. Aim is providing max certain anomalies to specialist of human after applying particular budget. In addition, they formulate an issue via framing by multi-armed bandit and improve new bandit collaborative environmental algorithm known as GraphUCB. Particularly, produced algorithm: (1) explicitly models features of node and dependencies in typical framework. (2) controls exploration-exploitation dilemma while querying on different anomalies. Wide tests on datasets of real-world illustrate developments to presented algorithm in comparison to the improved algorithms.

In [12], some machine learning models' performance is compared carefully to predict IoT systems anomalies and attacks. Algorithms of machine learning utilized here are: support vector machine (SVM), logistic regression (LR), artificial neural network (ANN), random forest (RF), decision tree (DT). Criteria of Evaluation that are applied for comparing performance such as F1 score, correctness, accuracy, recall circumstances under feature curve of receiver operating. system obtained test accuracy of 99.4% for ANN, tree of decision tree, random forest. Although, such methods have similar accuracy, the other criteria prove that random forest relatively act better.

In [13], that provides hidden Markov model algorithm (HMM-D) and describes Gaussian process regression (HMM-GP) based algorithm. In addition, they provide novel and more complete 8 algorithms evaluation for anomaly detection by signals of sound, motion, force gathered from robot which feeding yogurt to participants, picking up yogurt with a spoon, closes doors of microwave, closing the toolbox, that is physically capable. In general, HMM-GP had the highest performance in terms of the circumstances under curve for such tasks in real world, developed some methods of performance which several anomalies are able to be better detected by specific techniques. By artificial anomalies, HMM-D illustrates less latency in detection, HMM-GP outperforms by anomalies with high magnitude. Generally, artificial anomalies are more rapidly diagnosed with higher values.

In [14], DP algorithm benefits are applied for hyperspectral abnormalities detection so that overcome 2 negative natures which impact performance of detection: undeniable assumption on background statistics contamination and

Gaussian distribution which happen because of abnormalities. Particularly, spectral anomaly detection method based on DP works in this way: For the first time, for displaying hidden Markov model algorithm (HSI), expensive calculation density calculations, an image is segmented in local windows. In every local window, DP is done for calculating each pixel density. At last, they apply achieved density map for identifying anomalies, according to reality that anomalies have less probability generally in order to be exist in image and therefore, that has less density. Results of tests achieved in 4 real hyperspectral sets of data show that proposed method detection performance is superior to several methods of anomaly detection that are widely utilized.

In [15], that produces machine learning-based anomaly detection (MLAD). At first, prediction of load presented by neural networks is applied for reconstructing scaling of data applying clustering of k-meaning. Secondly, pattern of cyber-attack is estimated by classification of Bayesian naive given performance of pooling distribution as well as scaling data statistical features. At last, dynamic programming is applied for computing cyber-attack parameter and occurrence on data of load prediction. Broadly utilized symbolic method of pooling is compared to improved method of MLAD. Generally, Numerical simulation load data illustrates that method of MLAD is able to detect cyber-attacks effectively for data of load prediction with relatively high accuracy. So, MLAD strength is validated by thousands of scenarios of attack given the simulations of Monte Carlo.

In [16], novel tracking method of hyper personal anomaly by kernel Isolation Forest (iForest) is presented. The technique is according to assumption that anomalies are more prone to isolation in space of kernel instead of background. Given the opinion, presented technique diagnoses anomalies as below. At first, data of polar are plotted in space of kernel and the first main elements of K. after that, instances of separation are recognized in image with iForest, that is generated by applying random instances in basic elements. Finally, basic anomaly detection map is refined frequently with local iForest, that is generated in joined areas and with wide areas. results of tests in some real datasets of superpower illustrate that presented technique is better rather than the other new techniques.

3- PROPOSED METHOD

Presented work utilizes benchmark dataset of NSL_KDD for analysis of intrusion. first step concentrates on selection of feature by applying Rain Optimization Algorithm and chooses main features which contribute to intrusion. Second step focuses on using presented method of Stacked Encoding Hoeffding Tree for classifying data given the metrics of performance such as sensitivity, accuracy, false-negative rate, F1 score, specificity, false-positive rate.

Benchmark dataset of NSL_KDD is considered for analysis. Set of data that we have considered is standalone dataset for incorporating data of stream, set of data is streamed by applying methods in Matlab tool. Method of system object simplifies process of streaming in Matlab. Now, data is continuous and that has streaming data features. attacks are detected applying presented classification strategy of Stacked Autoencoder Hoeffding Tree (SAE-HT). bio-inspired method known as Rain Optimization Algorithm (ROA) increases SAE-HT classification method performance. distracting variance is eliminated from data by applying method of ROA feature selection which makes classifier able to outperform, particularly while coping with high dimensional features. Fig. 1 illustrates SAE-HT classification technique flow method.

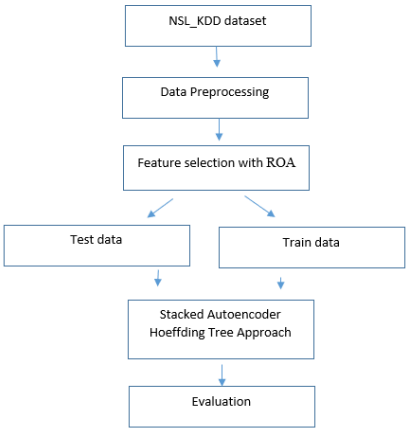

Figure 1. Flowchart of SAE-HT classification technique.

3.1. Data Preprocessing

Data Preprocessing is the important step in ML. raw data collected is made ready to be applied by ML techniques to extract meaningful insights from data. NSL_KDD Dataset taken from Canadian Institute for Cybersecurity is analyzed. While downloaded dataset is in gz /Tcpdump form, shift that to CSV file format and dataset load in environment. Presented method of classification just supports numeric data. As most of the methods of machine learning (ML) apply mathematical equations that just support numeric data use, categorical data conversion in numerical data applying functions of data conversion must happen. one-hot method of encoding changes categorical data to numerical data, hence making the easy for using methods of machine learning to set of data.

3.1.1. One-hot encoding technique

One-hot encoding technique is the most effective encoding technology to deal with several conversion to categorical attributes. This could shift categorical features to binary vector. vector holds Zeroes, One's as amounts. vector holds only 1 element with amount 1 and the other amounts associated with Zero. Element with amount 1 illustrates possible amounts happening against categorical attributes. Dataset of NSL_KDD includes 3 categorical features like flag, protocol_type, service. For instance, protocol_type includes 3 features: UDP, ICMP, TCP. By using one hot encoding technique, ICMP could be encoded as (1,0,0), TCP could be encoded as (0,1,0), UDP could be encoded as (0,0,1). Similarly, categorical features' service and flag are encoded in one-hot encoding vectors.

3.2. Feature selection with Rain Optimization Algorithm

Basically, selection of Feature decreases features with eliminating less significant/ insignificant cases. A lot of feature selection methods exist. The study applies method of Bio-Inspired feature selection known as Rain Optimization Algorithm (ROA) for choosing optimum features. Basic ROA aim is finding non-redundant and highly correlated features, therefore removes the least correlated features.

3.2.1. Rain Optimization Algorithm (ROA)

Each solution concern could be modeled by raindrop. Given the concern, some answer points in space could be randomly selected as raindrops fall on ground. Main attribute of rain drop is it's radius. Every radius of raindrop could be reduced as time passes by and which could be increased as raindrop is linked to the other drops. When the main crowd of answers is created, each radius of droplet could be assigned randomly in proper range. In each iteration, every droplet investigates its' neighbor given its' amount. Every droplet which are not integrated still to every other droplet, just investigate for end place restriction that was covered. When we are solving the problem in n-dimensional space, every droplet contains n criteria. Thus, at the first step, upper and lower criteria restriction would be investigated as these restrictions will be determined by radius of droplet. At the next step, two endpoints of criteria will be investigated which goes on until the final criterion. In that stage, the first price of droplet will be updated through moving that downward. It is not the last action for droplet while price task is reducing which would move downward in the same direction. The action would be performed for entire droplets, price and location of droplets would be assigned. Every radius of droplet would be transformed in two manners [5]:

1- While two droplets with radius r1 and r2 are so close to every other that has normal area with each other; they could join to shape the greater droplet of radius R:

$$R = (r_1^n + r_2^n)^{1/n} \tag{1}$$

2- that n is amount of criteria in every droplet. 2- while droplet with radius r1 does not transmit, given the soil attributes which is shown by α, some amount percentage could be attracted.

$$R = (\alpha\, r_1^n)^{1/n} \tag{2}$$

In fact, α shows percent of droplet amount which could be attracted in every iteration and is the number between 0 -100 percentage. We define minimum for droplets radius rmin, which droplets by that rmin smaller radius will not appear.

3.3. Stacked strategy of Autoencoder Hoeffding Tree

Classification is supervised method of machine learning which shares data set in various levels. Since generation of data is continuous, that is not possible to save extensive data amount. Therefore, data requires in order to be analyzed as that comes in. methods of classification are too much [17].

$$h_n = f(W_1 x_n + b_1) \tag{3}$$

That h_n shows encoder, vector assigned from x_n. function of Encoding f, encoder weight matrix W_1, vector of bias b_1. Eq. 4 shows process of decoder. That g shows function of decoding, W_2 shows matrix of decoders weight, b_2 shows vector of bias.

$$\widehat{X_n} = g(W_2 h_n + b_2) \tag{4}$$

That decoder reconstructs data of input, a possibility exists that leads in error of reconstruction. Eq. (5) reduces error of reconstruction.

$$\phi(\Theta) = \underset{\theta,\theta'}{\arg\min} \frac{1}{n} \sum_{i=1}^{n} L(X^i, \hat{X}^i) \tag{5}$$

That L shows function of loss, $L(X, \hat{X})$ shows function of loss. Todays, Hoeffeding tree is applied for classifying labels of class. Hoeffding tree is the decision tree type which includes root node, leaf node, test node. leaf node keeps prediction of class. Basic need in streaming data is classifying data in one pass. data provided as structure of tree applying method of Hoeffding tree while model incrementally generated. Basic Hoeffding tree classification demerit is that it fails to classify data in tree while tie happens. Eq. (6) presents Hoeffding bound computation formula.

$$\epsilon = \sqrt{\frac{\left(R^2 \log\left(\frac{1}{\delta}\right)\right)}{2n}} \tag{6}$$

That R shows random variable range, δ shows desired probability not in expected value ε, N presents samples' number gathered at node.

4- RESULT ANALYSIS

4.1. Data set description

Presented study applies benchmark dataset of NSL_KDD for analysis of intrusion. Dataset of KDD CUP'99 is well-known dataset of benchmark applied for system of network intrusion detection. Basic KDD CUP'99 dataset restriction is that includes high extra records number which have impact on evaluated system effectiveness. Developed KDD CUP'99 version refers to dataset of NSL_KDD where extra records are eliminated. Dataset of NSL_KDD have nearly 125,973 data of training and 22,544 data of testing. Like KDD CUP'99, records in dataset of NSL KDD are singly and labelled as anomaly and normal. That has 41 attributes which address 4 various attacks' groups [18].

4.2 Experimental Setup

The ability of a test to correctly distinguish fluctuations and normal events from other is called accuracy. For computing test accuracy, one should get true positive and negative instances' total ratio to sum cases' number. Mathematically, the rate could be assessed as:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \tag{7}$$

Sensitivity is called as rate of true positive. Sensitivity is described as:

$$\text{sensitivity} = \frac{TP}{TP+FN} \tag{8}$$

Specificity is described as:

$$\text{specificity} = \frac{TN}{TN+FP} \tag{9}$$

Variable of F-score integrates 2 two precision and sensitivity variables and is described as:

$$\text{F-score}=\frac{2\times\text{precision}\times\text{sensitivity}}{\text{precision}+\text{sensitivity}}) \quad (10)$$

In addition to the mentioned criteria, the terms N, P, TP, TN, FP, FN are summarized in a matrix called the confusion matrix, which is shown in the figure below.

Table 1- The confusion matrix

		Predicted class		
		yes	no	total
Real class	yes	TP	FN	P
	no	FP	TN	N
	total	P'	N'	P+N

However, the mentioned confusion matrix is designed for datasets with two classes, it can easily and similarly be generalized for data with more than two class labels.

We compared our presented model performance with other alike models (article [6] and [17]). We compared performance applying 4 metrics, known as F1-score, accuracy, sensitivity, specificity. table 2 shows that our presented technique could get the accuracy higher than 98.85% and the highest F1-score 99.36%. From Table 1, we could observe that our outcomes shown that our strategy suggests high classes of f1-score, precision, accuracy, recall particularly when we apply ROA for selection of Feature.

Table 2. Performance comparison with other approaches on NSLKDD

Method	Accuracy	Sensitivity	Specificity	F1-score
Paper [6]	96	-	-	-
Paper [17]	98	97	98.75	98
Proposed method	98.849	99.28	99.342	99.335

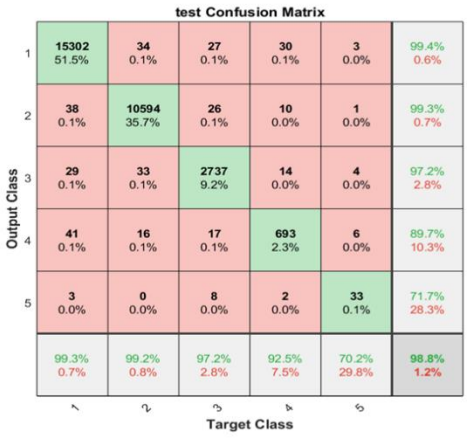

Fig. 2: (a) Confusion Matrix of NSL-KDD Data set using proposed method.

Fig. 2 illustrates NSL dataset confusion matrix by applying DNN. That illustrates that entire IDS accuracy is 98.8 percent.

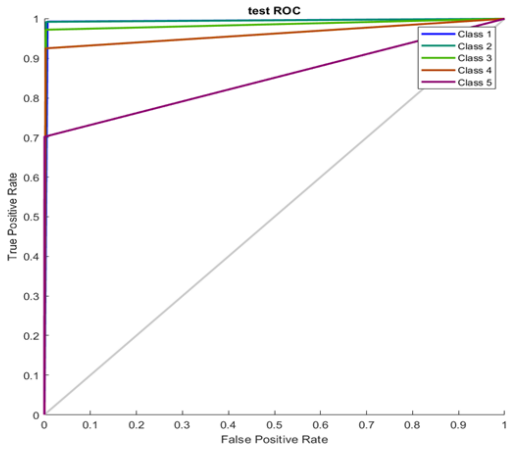

Fig. 3: ROC Curve of NSL-KDD Data set using proposed method

Figure 3 shows our proposed model curve of ROC (receiver operating feature curve). Our model adopts the great performance through generating score of AUC (area under curve of ROC) with high rate of true positive rate with rate of low false-positive.

5. CONCLUSION

Unusual detection of Network plays the vital role as that presents the efficient algorithm for preventing cyberattacks. With new Artificial Intelligence (AI) improvement, a number of deep learning strategies based on Autoencoder (AE) exist for network anomaly detection for developing our posture to the safety of network. Present new AE models' performance applied for anomaly detection of network differs with no holistic strategy propose for comprehending crucial AE models and detection accuracy essential performance indicators core set effects. This paper concentrates on developing IDS effectiveness by applying proposed Stacked Autoencoder Hoeffding Tree approach (SAE-HT) applying Rain Optimization Algorithm (ROA) for selection of feature. Experiments on the dataset NSL-KDD illustrate that our model multi-classification possesses great performance. In comparison to the other mechanisms of ML in accuracy case, our model performs better that such algorithms. Our future study would be directed to checking deep learning as a device of feature extraction for learning effective presentation of data for the issue of anomaly detection.

REFERENCES

[1] Deepa, M., and P. Sumitra. "An enhanced classification approach for network intrusion detection using Hoeffiding induction tree algorithm." *European Journal of Molecular & Clinical Medicine* 7, no. 09 (2020): 2020.

[2] Corrêa, Diego Guarnieri, Fabrício Enembreck, and Carlos N. Silla. "An investigation of the hoeffding adaptive tree for the problem of network intrusion detecti

[3] Pecht, Michael G., and Myeongsu Kang. "Machine learning: Anomaly detection." (2019): 131-162. on." In *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 4065-4072. IEEE, 2017.

[4] Muallem, Asmah, Sachin Shetty, Jan W. Pan, Juan Zhao, and Biswajit Biswal. "Hoeffding tree algorithms for anomaly detection in streaming datasets: A survey." *Journal of Information Security* 8, no. 4 (2017).

[5] Moazzeni, Ali Reza, and Ehsan Khamehchi. "Rain optimization algorithm (ROA): A new metaheuristic method for drilling optimization solutions." Journal of Petroleum Science and Engineering 195 (2020): 107512.

[6] Chen, Xiaoqing, Chunjie Cao, and Jianbin Mai. "Network Anomaly Detection Based on Deep Support Vector Data Description." In 2020 5th IEEE International Conference on Big Data Analytics (ICBDA), pp. 251-255. IEEE, 2020.

[7] Garg, Sahil, Kuljeet Kaur, Neeraj Kumar, Georges Kaddoum, Albert Y. Zomaya, and Rajiv Ranjan. "A hybrid deep learning-based model for anomaly detection in cloud datacenter networks." IEEE Transactions on Network and Service Management 16, no. 3 (2019): 924-935.

[8] Li, Dan, Dacheng Chen, Baihong Jin, Lei Shi, Jonathan Goh, and See-Kiong Ng. "MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks." In International Conference on Artificial Neural Networks, pp. 703-716. Springer, Cham, 2019.

[9] Borghesi, Andrea, Andrea Bartolini, Michele Lombardi, Michela Milano, and Luca Benini. "Anomaly detection using autoencoders in high performance computing systems." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 9428-9433. 2019.

[10] Zhou, Joey Tianyi, Jiawei Du, Hongyuan Zhu, Xi Peng, Yong Liu, and Rick Siow Mong Goh. "Anomalynet: An anomaly detection network for video surveillance." IEEE Transactions on Information Forensics and Security 14, no. 10 (2019): 2537-2550.

[11] Ding, Kaize, Jundong Li, and Huan Liu. "Interactive anomaly detection on attributed networks." In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining, pp. 357-365. 2019.

[12] Hasan, Mahmudul, Md Milon Islam, Md Ishrak Islam Zarif, and M. M. A. Hashem. "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches." Internet of Things 7 (2019): 100059.

[13] Park, Daehyung, Hokeun Kim, and Charles C. Kemp. "Multimodal anomaly detection for assistive robots." Autonomous Robots 43, no. 3 (2019): 611-629.

[14] Tu, Bing, Xianchang Yang, Nanying Li, Chengle Zhou, and Danbing He. "Hyperspectral anomaly detection via density peak clustering." Pattern Recognition Letters 129 (2020): 144-149.

[15] Cui, Mingjian, Jianhui Wang, and Meng Yue. "Machine learning-based anomaly detection for load forecasting under cyberattacks." IEEE Transactions on Smart Grid 10, no. 5 (2019): 5724-5734.

[16] Li, Shutao, Kunzhong Zhang, Puhong Duan, and Xudong Kang. "Hyperspectral anomaly detection with kernel isolation forest." IEEE Transactions on Geoscience and Remote Sensing 58, no. 1 (2019): 319-329.

[17] Seraphim, B. Ida, E. Poovammal, Kadiyala Ramana, Natalia Kryvinska, and N. Penchalaiah. "A hybrid network intrusion detection using darwinian particle swarm optimization and stacked autoencoder hoeffding tree." *Mathematical Biosciences and Engineering* 18, no. 6 (2021): 8024-8044.

[18] Krishnaveni, S., S. Sivamohan, S. S. Sridhar, and S. Prabakaran. "Efficient feature selection and classification through ensemble method for network intrusion detection on cloud computing." Cluster Computing 24, no. 3 (2021): 1761-1779.