[1] Omar Abdulkhaleq Aldabash

[2] Mehmet Fatih Akay

ANN Based Whale Sine Algorithm for Optimal Intrusion Detection

Journal of Electrical Systems

Abstract: - IDS (Intrusion Detection Systems) is extensively used to secure and monitor the networks. An efficient feature selection approach always directly influences on the performance styles such as computational information and integrity. Two techniques are proposed in this study, in order to detect the malicious activities that exists in the network–Optimal Whale Sine Algorithm (OWSA) for feature selection and ANN Weighted Random Forest (AWRF) for classification. The present study improves the IDS which is evaluated by using the NSL-KDD dataset. These proposed model in the present study obtains better results in accordance with performance metrics such as precision, recall, f1-score and accuracy and there by increases the efficiency of IDS. The proposed model produces more attractive outcomes that exposes the performance of the algorithm in order to select the best features and perform optimal intrusion detection.

Keywords: Intrusion Detection, Machine Learning, Sine Cosine Algorithm, Whale Optimization Algorithm.

I. INTRODUCTION

An Intrusion Detection System- (IDS) has been monitoring (checking) system that detects and picks up the suspicious activities and creates alarms when they are detected. When these alarms are raised, to treat the respective problem there is a special team in precise to handle these issues. To take up these problems and build up a solution, there is team called SOC Security Operations Centre. The malicious software (malware) on the run poses a precarious task in designing or building up an Intrusion Detection system (IDS). Malicious software had now become harder in aspects of being detected and identified by an IDS systems so that the malware systems are not being that easy in case of eradication [1].

In 2017, the (ACSC) -Australian Cyber Security Centre which critically examined the various levels of difficulty worked by the attackers. The main objective of the IDS lies in the activity of examining and detecting the malicious activity from the attackers. This cannot be done or less possible using the Traditional Firewall systems in case of more complex intrusion to the respective software or a dataset. With the improving the volume of computer malware, the enlargement of improved IDSs has become really important [2].

An IDS acts efficient by examining and by eradicating the intrusions either by detecting the traffic in the network system. This can be caused by unwanted intrusions, but also ensures in maintaining the surety of confidentiality, integrity, and availability. Since then, there has a need of captivating and exploring many IDS products. This is done in order to maintain the security of each and every data stored in cloud or in any security needed systems because of its large network nosed and connectivity. This results in, a large count of significant data is being produced and shared across changed network nodes. Existing and old IDSs have shown less efficacy in maintaining the security over data's and are not capable of detecting some complicated activities such as the zero-day attacks, False Alarm Rates (FAR). Various DL- and ML-based solutions have been suggested through the researchers to create IDS efficient in identify malicious attacks [3]

The Intrusion detection has been most significant part of the cyber security technology to detect malicious activities, in order to monitor and analyze the network system traffic raised from different resources and identify malicious actions. To detect different malicious attacks – Deep Neural Network is used. Another powerful method is incorporated to detect malicious actions, which is Deep Belief Network (DBN). Several approaches were surveyed to understand wide varieties of deep learning and machine learning technique for IDS related work, which contains DBN approach but it failed to deliver a proper review which related both DBN and IDS models [4].

Deep Belief Network (DBN), Stacked auto encoder (SAE), CNN (convolutional neural network), RNN (Recurrent Neural Network) are some of neural networks which have been implemented by DL techniques. DBN algorithm is one of these DNN techniques which is rained using unlabeled datasets and are made to fine tune using some other specialized algorithms. In IDS, DBN is the most frequently used technology as well as most substantial technology used in DL techniques. SAE is nothing but layers (stacks) of many auto encoders in them. AE involves of a encoder, which has been limited to its dimension, and the important activity of the encoder lies in reconstructing the input data [5].

In Internet of Things (IoT), network security plays a vital role because most manufacturers do not imply more on the security of systems when designing. The main aim and objective of the intrusion detection systems, firewalls,

[1, 2] Department of Computer Engineering, Çukurova University

* Corresponding Author Email: [1]omarabd1970@gmail.com, [2] mfakay@cu.edu.tr

and intrusion systems mainly rely on the concept of lessening the false alarm rates and producing enhanced alarm rates while procuring detection. There has been objective of the firewall is to catch damages and decrease the upcoming traffic to the particular network. Due to innovative practices of attacks it is complicated to distinguish between regular traffic and malicious traffic.

To employee and to distinguish between these problems, a two-stage hybrid approach has been suggested. To increase the precision of the recommended framework, genetic algorithm has been implemented which is used to pick up the apt features. Later, ML algorithms such as SVM and classifiers such as Random forest tree are worked together. Using 10-fold cross validation accuracy rate of 99.8% is reached using NSL-KDD database. There are more forms of intrusion or a disturbance commonly known for the attack to the information or the network systems some of them are include [6].

This in turn results in need of building a strong system against intrusion and propose new requirements for the dataset to be safe. By means of the newly proposed techniques on these software systems, it results in building up datasets which can safeguard themselves from the already tagged instances and newly upcoming network traffics in form of intrusions. This will allow using collected dataset in building a strong and durable next-generation IDS. One such example can be finding the characteristics of malicious file or a network (malware) using sandbox. Malicious (malware) files can be readily downloaded from web resources, where it is already pre classified by the researchers. The files which are downloaded from the web sources can run smoothly on the sandbox on its own characteristics where its results can also be checked here.

Hybrid analysis can be used to detect or to check the activity of the application programing interface. This results in assembling a large dataset of ML algorithms in detecting the harness of the file. The drawback of this approach could be is, when this methodology is launched in the virtual environment to act as intrusion, these files will act wisely as if it is a non- malicious software and as genuine file of convenience. [7].

II. Significance And Types Of Ids

Rapid development and growth of the network systems has triggered or made a speed in ensuring the security of the data. This has taken a crucial part because data are being transferred from one edge to the other. IDS ensures in preserving its structure to be safe and secured throughout the transmission phase. This should be preserved in view that, the data has much of complexity and the heterogeneous pack of information's in it. IDS has also ensured in storing and keeping up the computational power of the devices in addition to maintaining the steady progression while in need of transmitting information, local storing of data and also drives the computation of drive network in the direction of edge devices. IDS system plays an important role in safeguarding privacy and safety of devices [8].

IDS has shown its implication in many areas and in aspects of data's security and data de- attacking. Some of its main application could serve on computing, IoT, smart vehicles, smart and operation systems by remote gadgets. The most vital and a crucial part of these IoT s would be ensuring the connectivity from edge-to-edge since the main backgrounds such as independency and some aspects such as electrification highly depends on the connectivity of the information's. So the data's shared between one among another should be ensured with not ruining up the confidentiality, the integrity of the message of data shared in-between and to be free from denial of service to the data or the network systems. Computerized, secure, nonstop cloud service accessibility framework is very much needed for connected vehicles. This is usually done to maintain the QoS (Quality of Services) and also the measure the experience of the customers. HIDS based monitoring system uses Deep Belief Network which is D2H-IDS, which identify attacks suggested and calculated by the suggested algorithms. Furthermore, one such technique is used in the three-phase data traffic analysis, to detect helpful reliable service appeals alongside false requests that occurs in the course of intrusion in order to detect and classify the forms of intrusions to the detection system [9].

III. Probem Identification

False alarm rate is still one of problems in IDS, which increases the burden for security analyst. Therefore, many researchers across the world have taken a step to implement the ideas of IDS along with decreasing the production of false alarm rates and also implementing accurate and higher prediction rates. Inability to detect unidentified attacks is still considered as one of the major issues with existing IDSs. Therefore, to deal with this issues researchers have begun to concentrate on construction of IDS using Machine learning algorithms. However, traditional machine learning models like KNN (K-Nearest Neighbor), SVM, cannot be used since they are ineffective and futile. Therefore, to showcase abundant ML algorithms used in this sector, IDS taxonomy choose to consider data sources as their primary objective. Since IDS's purpose is to detect attacks, it is important to choose appropriate data source [4].

IV. Motivation

Since cyber-attacks are becoming more frequent, it is important detect these attacks and eliminate from the network as early as possible. However, Firewall alone cannot offer sufficient protection against modern cyber threats. Legal and valid types of traffic such as web traffic and email are used to deliver certain malware and malicious content, which is not appropriate. Hence IDS has the ability to detect the content of these communications and recognize any malware that they might contain. Motivated by this factor, the study proposes optimal whale sine algorithm for effective feature selection and ANN weighted Random forest method for classification to enhance the accuracy of IDS system.

Optimal whale sine algorithm: For Feature selection, Optimal Whale Sine algorithm is proposed because it has the capability to solve real life optimization applications. Due to its feasible execution time, advantages of few control parameters, simple calculation, solid capability to examine optimal solution, decent convergence and acceleration rate, high effectiveness associated to several well-regarded optimization algorithms, and its ability to be readily hybridized with other optimization algorithms. Optimal whale sine algorithm - It updates the location and of the obtained best solution and also update the positions using sine cosine algorithm. At last the best solution will be returned as the global optimum. Motivated by these merits, this study considers optimal whale sine algorithm.

ANN weighted Random Forest Classification: For classification, ANN weighted Random forest method is introduced for classifying intrusion detection into attack and non-attack. This classification is proposed because it decreases over fitting in decision trees which subsequently helps to improve the accuracy. It automates missing values present in the data. It also has the numerical strength that can perform more than one job at the same time.

V. Aim And Objective

The existence of the IDS to the system is checked using this proposed methodology. Looking up to this objective two techniques are proposed in this study to detect the presence and absence of malicious activities existing in the network – Optimal Whale Sine Algorithm (OWSA) for feature selection and ANN Weighted Random Forest (AWRF).

- To perform feature selection using Optimal Whale Sine (OWS) Algorithm for improvising classification algorithm.
- To classify the presence and absence of intrusions/attacks using the proposed AWRF (ANN weighted Random Forest) classifier for better prediction.
- To evaluate the proposed method using performance metrics for confirming the efficiency of the proposed method.

VI. RESULTS

6.1 Performance metrics

The Performance metrics have become a part of each ML. Performance metrics are utilized in analyzing the classification models for the provided balanced datasets. The performance of the proposed IDS design is determined with the performance metrics like F1-score, precision, accuracy and recall.

The performance analysis for the performance metrics NSL-KDD of dataset has been resulted 99.92 percentage of accuracy, 99.97 percent of precision, 99.38 percentage of recall and 99.54 percentage of F1-score Fig 1 also provided for the performance metrics of NSL-KDD dataset.

Table I: Performance Analysis for Nsl-Kdd

Accuracy	**Precision**	**Recall**	**F1-Score**
99.92	99.97	99.38	99.54

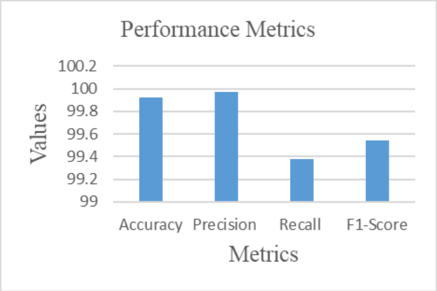

Fig. 1. Performance analysis for NSL-KDD

The NSL-KDD dataset includes both numeric and non-numeric attribute. The non-numeric attribute such as flag attribute, service and protocol type wanted to transform as numeric attribute due to the testing input and training input.

The training set includes 22 types of training attacks and the testing set contains additional 17 attack types and remove 2 types such as wareclient and spy attacks from training set. Hence, there are 37 types of attacks that contains in testing set. The fake attack drop in anyone of four classification, they are DoS (Denial of service) attack, Probing attack, U2R (User to Root) attack and R2L (Remote to Local) attack.

Table II: Accuracy score for various weight in NSL-KDD

Weight	**Accuracy**
1	99.92
2	99.92
3	42.5

4	99.92
5	42.5
6	99.92
7	99.92
8	42.5
9	42.5
10	99.92

The table 2 shows the accuracy for various weights using NSL-KDD dataset. 10 different weight has been analysed with NSL-KDD dataset. If the weight increases the accuracy also increases, even though it oscillates in some points, (for weight 5 – accuracy 42.5) and finally it gives the best accuracy result at weight 10 with 99.92 accuracy rate.

Table III: k-fold cross validated results for NSL-KDD

Folds	Accuracy
Fold 1	0.9375
Fold 2	0.956
Fold 3	0.9375
Fold 4	0.94
Fold 5	0.9375

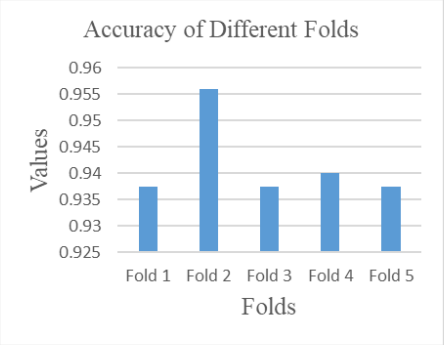

Fig. 2. k-fold cross validation using NSL-KDD dataset

The table 3 accuracy for k-folds cross validation using NSL-KDD dataset. and also the Fig 2 denotes the graphical representation of the accuracy of k-folds cross validation using NSL-KDD dataset. 5 different k-fold cross validation has been analysed with NSL-KDD dataset. If the K-fold increases the accuracy also increases, even though it oscillates in some points, (for fold 2 – accuracy 0.956) and finally it gives the best accuracy result at fold 5 with 0.9375 accuracy rate.

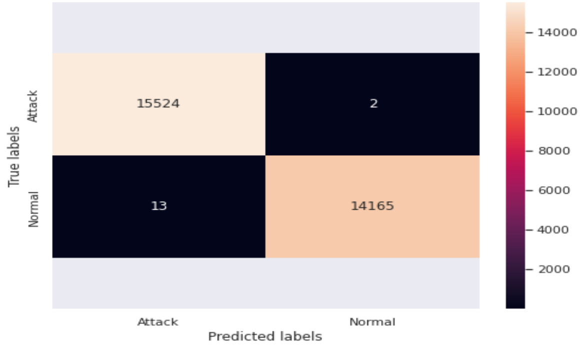

Fig. 3. confusion matrix for NSL-KDD dataset

The Fig 3 represents the confusion matrix of NSL-KDD dataset. While using NSL-KDD dataset in MLP classifier, 15524 attacks have been predicted as attacks and 2 attacks have been misclassified as normal. 14165 normal have been properly sorted as normal and 13 attacks have been misclassified as normal.

6.2 Comparative Analysis

Many algorithms are available to carry out the process of Network Intrusion Detection System (NIDS). The suggested work has been analyzed with various existing methods which used different techniques and algorithms to perform the process of IDS. The outcomes of comparative analysis have shown that the proposed system that used OWSA and AWRF (ANN Weighted Random Forest classification) shows better results in accordance with performance metrics.

The performance metrics of suggested method has been analyzed with the existing study. The existing paper [10] has used Focal Loss (FL-NIDS) algorithm by using DNN and CNN for the dataset UNSW-Bot. It is clearly found that, the recommended model which used OWSA has showed better outcomes in all performance metrics. The recall value of existing model has been given as 0.9433 but in the proposed method it is given as 0.9975 which is one of the prominent increased value. The values of accuracy and precision has been increased by 0.0002 and 0.027 respectively. F1-score has showed noticeable increase from 0.9566 to 0.9971 by using the proposed model.

For NSL-KDD dataset, the proposed algorithm accuracy has been compared with the existing system [11] which has used the technique of XGBoost–DNN and, LR, NB, SVM, in that, XGBoost–DNN has high accuracy value of

0.976 when compared with LR, NB and SVM. But the proposed algorithm accuracy is 0.99. Hence, the proposed algorithm has produced good outcomes in accuracy.

For NSL-KDD dataset, the proposed algorithm has been compared precision value with the existing model [11] which has used the technique of XGBoost–DNN, LR, NB and SVM. The precision value of proposed system is mentioned as 0.999 which the highest value when compared with existing method which has used various algorithms. Hence, the proposed model stands with good results in terms of precision which has been shown in the Fig 4 and table 4.

Table IV: Existing and proposed model comparison of precision for NSL-KDD

Existing and proposed model comparison of precision				
Precision				
Proposed system	**LR**	**SVM**	**NB**	**XGBoost–DNN**
0.999	0.87	0.97	0.28	0.976

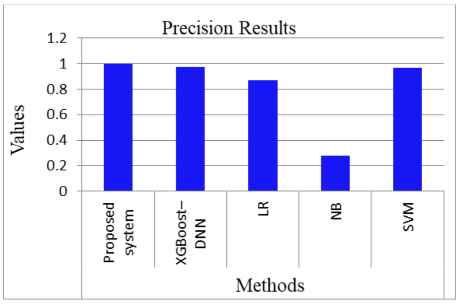

Fig. 4. Existing and proposed model comparison of precision for NSL-KDD

For NSL-KDD dataset, the proposed algorithm has been compared the F1 score with existing model [11] which has used the technique of XGBoost–DNN, LR, NB and SVM. The proposed model stands with high value F1 score i.e. 0.995 in compare with the F1 score gets from other techniques like XGBoost–DNNand, LR, NB and SVM in existing model which has been shown in the Fig 5 and table 5. Hence, it is evident that the proposed model which used OWSA and AWRF classifiers has showed better results and more efficient in accordance with F1 score, accuracy and precision.

Table V: Existing and proposed model comparison of F1 score for NSL-KDD

Existing and proposed model comparison of F1 score				
F1 score				
Proposed system	**LR**	**SVM**	**NB**	**XGBoost–DNN**
0.995	0.87	0.97	0.36	0.976

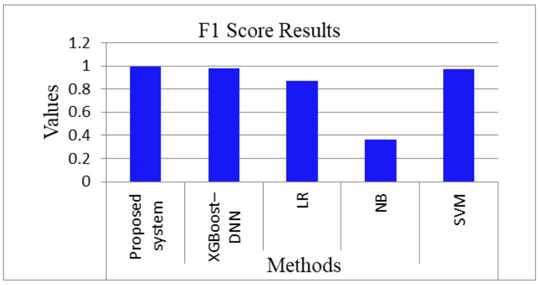

Fig. 5. Existing and proposed model comparison of F1 score for NSL-KDD

Experimental analysis made on NSL-KDD, in which the proposed algorithm has been compared with the existing system [11] that used CFS (Correlation Feature Selection) which has been included with neural network for finding abnormalities. Table 6 shows that the proposed model has displayed better outcome in compare with existing system on specificity and accuracy. Hence, it has been concluded that the proposed model shows efficient and improved values as 99.97 and 99.92 on specificity and accuracy in compare with the existing model.

Table VI: Specificity and Accuracy comparison of different techniques used for IDS on NSL-KDD dataset

Specificity and Accuracy comparison of different techniques used for IDS on NSL-KDD dataset		
Learning techniques	**Specificity (%)**	**Accuracy (%)**
SVM	71.41	69.52
Bi-LSTM	79.64	76.37
CNN	80.75	95.01
Multi-Layer Perceptron	79.57	77.41

C4.5	83.44	81
Naïve Bayes	83.21	81.47
Random Forest	82.35	80.67
CART	82.71	80.3
Ensemble35	89.41	87.28
CNN-BiLSTM34	80.83	80.05
SAE-SVM-RBF21	98.35	95.27
SVM-RBF	91.84	92.55
(CFS+ANN)	99.31	97.49
Proposed System	99.97	99.92

Table 6 and has shown Specificity and Accuracy comparison of different techniques used for IDS on NSL-KDD dataset. Performance metrics of proposed algorithm has been compared with the existing paper. The existing paper has used Focal Loss (FL-NIDS) algorithm by using DNN and CNN for the dataset NSL-KDD. The accuracy of the existing study has used various methods like CNN CE, CNN SMOTE, FL-NIDS. From Fig 6 and table 7 , it is clearly found, the proposed work which used OWSA has shown better accuracy [9].

Table VII: Comparative analysis of NSL-KDD

Method	**Accuracy**
CNN CE	0.8369
CNN SMOTE	0.8334
FL-NIDS	0.8489
Proposed	0.9992

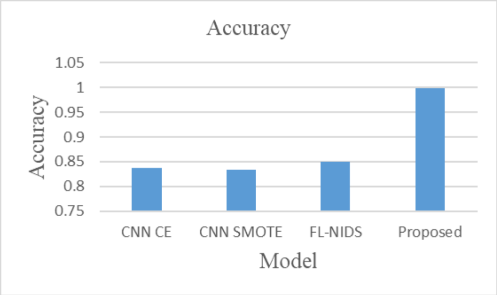

Fig. 6. Comparative analysis of NSL-KDD

Fig 6 and table 7 has shown the Comparative analysis of NSL-KDD. For the dataset NSL-KDD, the proposed model has used the OSWA and it results with optimal feature selection. The existing study [9] has used various ML methods like SVM, RF, NB and ANN. It is observed that the proposed method has reached improved outcomes in accordance with precision, F1-score and recall which is 0.999, 0.995 and 0.998 respectively**.**

For the dataset NSL-KDD, the proposed model has used the OSWA and it results with optimal feature selection. In the existing study [12], the classification analysis of IDS has used different supervised learning algorithms like NB, SVM, KNN, RF, LR and DT on the NSL-KDD dataset. It is observed that the proposed system has attained improved accuracy which is 99.92. Hence, the proposed algorithm is more efficient in comparison with the existing system present in the suggested study.

Table VIII: Comparative analysis of NSL-KDD

Model	**Accuracy**
KNN	98.59656
SVM	96.69778
DT	95.83413
RF	98.70452
GNB	84.32717
LR	96.97085
Proposed	99.92

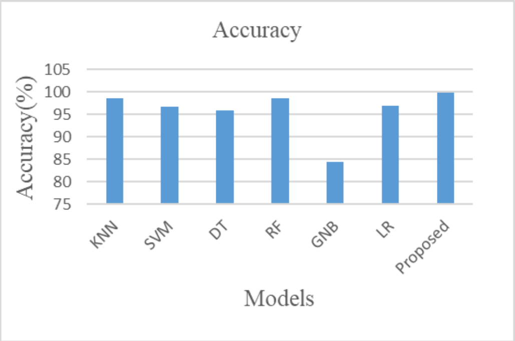

Fig. 7. Comparative analysis of NSL-KDD

Table 8 and Fig 7 has shown the Comparative analysis of NSL-KDD.

Similarly, in the existing study the performance has been evaluated for NSL-KDD dataset. The recommended study [9] has used FL-NIDS (focal loss network intrusion detection system) in which it has been suggested to rectify the problems of imbalanced data. By reshaping and modifying the loss function that is based on the standard cross-entropy, it has addressed and obtained improved classification. Thereby, it resolved the issues in imbalanced data and results in improved performance. In NSL-KDD dataset, by using FL-NIDS with DNN and CNN had helped to overcome the imbalanced distribution issues. In the existing study, the value of accuracy by using the method of FL-NIDS is given as 0.8489 for NSL-KDD dataset. But, the proposed method which has used OWSA and AWRF classifier has produced better values in terms of accuracy which is given as 0. 9992.From the result, it is clearly noted that the proposed algorithm has improved values in all performances metrics in compare with the existing method in the recommended study.

For the dataset NSL-KDD, the existing study [10] has used various ML methods such as SVM, RF, NB and ANN. By using RF classifier, the outcomes of the performance metrics like precision, recall and F1-score has been given as 0.9683, 0.6158 and 0.7528 respectively. The proposed model has used the OSWA and it results with optimal feature selection. It is observed that the proposed method has achieved enhanced outcomes on precision, recall and F1-score which is 0.999, 0.998, and 0.995 respectively**.** Thereby, it is concluded that the present study has obtained optimal results than the existing study.

In the existing study [12], the classification analysis of IDS has used different supervised learning algorithms like NB, SVM, KNN, RF, LR and DT on the NSL-KDD dataset. When using ML methods, the detection accuracy has been improved for security attacks. The accuracy value for ML models such as KNN, LR, SVM, GNB, DT, and RF is given as 98.59656, 96.97085, 96.69778, 84.32717, 95.83413, and 98.70452 respectively. And it is identified that the proposed work has attained better accuracy which is 99.92. Hence, the proposed work is more effective in compare with existing method.

VII. CONCLUSION

With the development of technology, the cyber security threats that risk the integrity, privacy, and accessibility of information in the system have also increased. Cyber security methods mostly include antivirus software, firewalls and IDS that protects the system from hacks. Among them, IDS is a type of detection system that plays an important role in protecting cyberspace by observing the software and hardware conditions in the network.

The present study has used OWSA and AWRF classifiers in order to improve the NIDS which has been evaluated by using the datasets such as NSL-KDD dataset. The results were compared in an internal comparison in terms of F1-score, accuracy, precision and recall. The proposed model has been compared with existing studies in the discussion section. The proposed model which used OWSA and AWRF classifier has performed better. Hence the proposed method is more efficient in compare with the discussed existing methods.

REFERENCES

[1] L. Chhaya, P. Sharma, A. Kumar, and G. Bhagwatikar, "Cybersecurity for smart grid: threats, solutions and standardization," in Advances in Greener Energy Technologies, ed: Springer, 2020, pp. 17-29.

[2] E. H. H. A. S. Saleh, M. F. B. Abd Kadir, and Y. A. El-Ebiary, "Bottleneck Bandwidth And Round-Trip Propagation Time Algorithm Using TFRC Protocol By Artificial Intelligence Algorithm," Journal of Pharmaceutical Negative Results, pp. 841-849, 2022.

[3] A. K. Shukla, "An efficient hybrid evolutionary approach for identification of zero-day attacks on wired/wireless network system," Wireless Personal Communications, pp. 1-29, 2020.

[4] I. Sohn, "Deep belief network based intrusion detection techniques: A survey," Expert Systems with Applications, vol. 167, p. 114170, 2021.

[5] M. Aloqaily, S. Otoum, I. Al Ridhawi, and Y. Jararweh, "An intrusion detection system for connected vehicles in smart cities," Ad Hoc Networks, vol. 90, p. 101842, 2019.

[6] T. Saba, T. Sadad, A. Rehman, Z. Mehmood, and Q. Javaid, "Intrusion detection system through advance machine learning for the internet of things networks," IT Professional, vol. 23, pp. 58-64, 2021.

[7] I. Sumaiya Thaseen, J. Saira Banu, K. Lavanya, M. Rukunuddin Ghalib, and K. Abhishek, "An integrated intrusion detection system using correlation-based attribute selection and artificial neural network," Transactions on Emerging Telecommunications Technologies, vol. 32, p. e4014, 2021.

[8] S. Rastogi, A. Shrotriya, M. K. Singh, and R. V. Potukuchi, "An analysis of intrusion detection classification using supervised machine learning algorithms on NSL-KDD dataset," Journal of Computing Research and Innovation (JCRINN), vol. 7, pp. 124-137, 2022

[9] S. Sapre, P. Ahmadi, and K. Islam, "A robust comparison of the KDDCup99 and NSL-KDD IoT network intrusion detection datasets through various machine learning algorithms," arXiv preprint arXiv:1912.13204, 2019.

[10] M. Mulyanto, M. Faisal, S. W. Prakosa, and J.-S. Leu, "Effectiveness of focal loss for minority classification in network intrusion detection systems," Symmetry, vol. 13, p. 4, 2020.

[11] P. Devan and N. Khare, "An efficient XGBoost–DNN-based classification model for network intrusion detection system," Neural Computing and Applications, vol. 32, pp. 12499-12514, 2020.

[12] S. Rastogi, A. Shrotriya, M. K. Singh, and R. V. Potukuchi, "An analysis of intrusion detection classification using supervised machine learning algorithms on NSL-KDD dataset," Journal of Computing Research and Innovation (JCRINN), vol. 7, pp. 124-137, 2022.