Survey of Intrusion Detection of Imbalanced Network based on Machine Learning Algorithm

Rachita Kulshrestha
M. Tech Research Scholar
Dept. of CSE
SIRTS, Bhopal, India

Chetan Gupta
Assistant Professor
Dept. of CSE
SIRT, Bhopal, India

Ritu Shrivastava, PhD
Head of Dept.
Dept. of CSE
SIRT, Bhopal, India

ABSTRACT
Intrusion detection is the process of analyzing the network packets to identify if the packet is legitimate or anomalous. The major challenges involved in this domain includes the huge volume of data for training and the fast and streaming data that is to be provided for the prediction process. The accuracy and timely detection should be ensured by Network Intrusion Detection System (NIDS). For intrusion detection in balance and imbalance network traffic, machine learning and deep learning methods can be used. In this paper a survey of different intrusion detection systems based on machine learning and deep learning methods is performed. The proposed system adds on ensemble learning approach to improve accuracy. A review on various intrusion detection system (IDS) using the techniques in machine learning is been put forwarded.

Keywords
IDS, Imbalance, Balance, Machine Learning

1. INTRODUCTION
One of the resources that is used the most is the Internet. People are carrying out digitized transactions as a result of the huge increase in internet usage. The expansion in number of web clients is 4% to 6% consistently. It has been observed that the growth is much higher in developing nations like India. A huge amount of data is being generated online as a result of high levels of computer and internet technology adoption [1, 2]. It is difficult to find a person who does not have an online presence because these technologies have also become essential components of human life. Expansion in on the web presence has likewise leaded to sharing or support of individual data on the web. Albeit this acquires enormous comfort terms of access, they are additionally helpless against assaults or interruptions. These intrusions may result in significant financial losses as well as the disclosure of private information to unintentional users [3].

Currently, more people are entering the digital realm thanks to the use of mobile devices. Additionally, there has been an increase in the number of web pages served and optimized for mobile devices. Numerous users have chosen mobile-based commerce applications due to the widespread use of mobile devices. The high reception levels of Web and cell phones as well as the expansion in web based business an exchange happening by means of organizations has brought about the expansion in huge number of cybercrimes.

Worldwide, cybercrime has cost 1,418 million dollars. (Source: www.statista.com). This demonstrates the growing significance of intrusion detection models in enhancing online user security. It moreover orders the utilization of better interruption discovery frameworks and displays the absence of proficiency in the current interruption discovery frameworks.

The process of identifying unauthorized anomalous activities on computer systems is referred to as "intrusion detection." An Intrusion Detection System (IDS)'s primary function is to classify users' actions as normal or abnormal based on the data they transmit.

Conventional insurance frameworks utilized firewalls, information encryption and verification strategies. However, the current intrusion scenarios are extremely sophisticated and have the ability to easily breach the security mechanisms that are enforced by conventional security measures. This has come about in a colossal expansion in the quantity of specialists working in this space and has too expanded the quantity of examination commitments in this space [4].

In various industry scenarios, IDS models have multiple applications and requirements. The process of intrusion detection in personal systems, or distributed scenarios, is a major application. The architecture of current operating systems includes intrusion detection mechanisms. However, the systems' handling capabilities remain a mystery. As a result, for added security, the majority of users tend to use commercial intrusion detection models. In addition, there is a demand for server-based IDS that can be used in clustered environments.

2. LITERATURE REVIEW
Kezhou Ren et al. [1], network attacks pose significant threats to network services' security. Therefore, it is essential to make use of brand-new technical approaches in order to boost the effectiveness of intrusion detection systems. In recent years, a number of reinforcement learning algorithms for network intrusion systems, such as Markov and others, have been developed to meet the IDS's needs for both intelligent and unmanned systems. A deep feed-forward neural network approach is used in conjunction with a deep Q-learning-based network intrusion detection model that incorporates reinforcement learning to provide continuous automatic learning capability for network environments. To test the model's performance, experiments were carried out with the CSE-CIC-IDS2018 dataset, which contains a comprehensive collection of actual network traffic. The results of the experiments show that the proposed model can successfully identify threats in the network, outperforms standard machine learning methods, and has promising potential as unmanned IDS in complex network settings.

R. Ahsan et al. [2], the issue of anomaly detection in imbalanced datasets is examined in this work within the context of network intrusion detection. To deal with the class-imbalance issue, a novel anomaly detection solution that takes into account both data-level and algorithm-level approaches is proposed. The oversampling capabilities of a Conditional Generative Adversarial Network (CGAN) and the auto-learning capabilities of reinforcement learning are combined in

this approach. To additionally explore the capability of a CGAN, in imbalanced order undertakings, the impact of CGAN-put together oversampling with respect to the accompanying classifiers is inspected: Logistic Regression, Multilayer Perceptron, Random Forest, and Naive Bayes The experimental results show that the proposed method and CGAN-based oversampling in general perform better than other oversampling methods like the Synthetic Minority Oversampling Technique and Adaptive Synthetic. The Authors in 2021 Cyber-Physical Systems from IET: Hypothesis and Applications distributed by John Wiley and Children Ltd for The Establishment of Designing and Innovation.

S. Dong et al. [3], our production life has been greatly improved by the rapid development of Internet technology, but as a result, security issues have become increasingly prevalent. Users' privacy is at risk, and many aspects of society, including politics, the economy, culture, and people's means of subsistence, face significant security risks as a result of these issues. The development of the data transmission rate grows the extent of assaults and gives a more assault climate to interlopers. Strange location is a compelling security assurance innovation that can screen network transmission continuously, really sense outside assaults, and give reaction choices to pertinent chiefs. A technology for detecting abnormal traffic has also emerged as a result of advances in machine learning. The objective has been to use powerful and quick learning algorithms to respond immediately to shifting threats. The vast majority of the flow unusual identification research depends on reproduction, utilizing public and notable datasets. From one viewpoint, the dataset contains high-layered enormous information, which customary AI strategies can't be handled. However, the labeling cost is extremely high because the dataset's labels are all manually labeled and the labeled data scale is far behind the application requirements. This paper proposes a semi-directed Twofold Profound Q-Organization (SSDDQN)- based improvement strategy for network strange traffic location, mostly founded on Twofold Profound Q-Organization (DDQN), a delegate of Profound Support Learning calculation. The current network in SSDDQN uses a deep neural network as a classifier before using the autoencoder to reconstruct the traffic features. K-Means clustering is used by the target network's unsupervised learning algorithm first, followed by deep neural network prediction. For training and testing, the experiment makes extensive use of the NSL-KDD and AWID datasets and compares them to existing machine learning models. The experimental results demonstrate that SSDDQN performed well in a variety of evaluation metrics and has some advantages in terms of time complexity.

Lau lin et al. [4], in imbalanced organization traffic, pernicious digital assaults can regularly stow away in a lot of typical information. It displays a serious level of covertness and jumbling in the internet, making it hard for Network Intrusion Detection System (NIDS) to guarantee the precision and practicality of discovery. This paper explores AI and profound learning for interruption recognition in imbalanced organization traffic. It proposes an original Difficult Set Sampling Technique (DSSTE) calculation to handle the class awkwardness issue. To start with, utilize the Edited Nearest Neighbor (ENN) calculation to separate the imbalanced preparing set into the troublesome set and the simple set. Then, utilize the K-Means calculation to pack the larger part tests in the troublesome set to diminish the larger part. Zoom in and out the minority tests' persistent characteristics in the troublesome set integrate new examples to expand the minority number.

A. Raghavan et al. [5], successful and proficient malware recognition is at the bleeding edge of examination into building secure computerized frameworks. Similarly as with numerous different fields, malware location research has seen a sensational expansion in the utilization of AI calculations. One AI strategy that has been utilized broadly in the field of example matching overall—and malware identification specifically—is covered up Markov models (HMMs). Gee preparing depends on a slope climb, and thus we can frequently work on a model via preparing on numerous occasions with various beginning qualities. In this exploration, we think about helped HMMs (utilizing AdaBoost) to HMMs prepared with different arbitrary restarts, with regards to malware identification. These procedures are applied to an assortment of testing malware datasets. We observe that irregular restarts perform shockingly well in contrast with helping. Just in the most troublesome "cold beginning" situations (where preparing information is seriously restricted) does helping seem to offer adequate improvement to legitimize its higher computational expense in the scoring stage.

Zhiyou Zhang et al. [6], in this paper, aimed at detection of internal intruders in HIDS. Commonly used login ids and passwords may be shared along with co-workers for professional purposes, which can be tampered or used by the attackers as a means of intrusion into the system details. The user was monitored and System Calls (SC) was extracted and the habitual SC pattern based on the habits of the user was taken into account and the profile of the user was stabilized. The forensic technique and other data mining techniques were applied at SC level host IDS to spot the internal attacks. Along with the user login credentials the forensic technique was applied to investigate the computer usage fashion against the collected user profile pattern and thereby check the identity of the user.

Afreen Bhumgara et al. [7], in this paper, With the decision rate threshold of 0.9, the system was able to perform with an accuracy rate of 94%. Nokia Research Center researchers modeled HIDS for mobile devices. The limitation include that each protocol state consume resources for tracing and testing, and its inability to guess the attacks resembling benign protocol. Access control fills in as the cutting edge of resistance against interruptions, bolstering both confidentiality and integrity parameters. Intrusion detection is the process of progressively observing the events occurring in a PC or network, examining them for indications of conceivable episodes and often interdicting the unapproved access. A state transition diagram can be constructed for the sequence of events, but not for the complex forms and hence the attacks having complex behavior which cannot be modeled as the state transition diagram will go unnoticed by the system.

Ritumbhira Uikey et al. [8], in this paper, along with various protection mechanisms accompanied with mobiles they felt an urge for attack monitor methods as a second line of defense. The framework was designed, taking into consideration the privacy of the mobile user in creating the user profile. The framework had a major share with the host-based intrusion detection in line with the network-based detection system, as researchers felt that mobile requires the monitoring system at both ends. The framework included data collection and IDS modules, the former entrusted with responsibility of monitoring the operating system activities, calculating the system measurements and the data collection at the application level

and the later feeding on the collected and pre-processed data performs the actual intrusion detection.

Aditya Phadke et al. [9], targeted Advanced Persistent Threat attacks, by analyzing the 30 behavioral pattern of the host user through a 83-dimensional vector, each attribute representing one manner of the user. In order to form the database, they collected 8.7 million features from 4000 malicious and normal programs through the Virtual Machine (VM) environment. The system was designed in such way that frequency of occurrence of each behavior is calculated for each process. C4.5 decision tree was used to build a classifier for the collected information, and each new instance was analyzed against the tree to be segregated as malicious or normal instance. The model had a false positive rate of 5.8% and a false negative rate of 2.0%.

S. Sivantham et al. [10], in this paper, represented a novel HIDS aimed at discovering unknown malware codes. The collection of previous malware codes was taken as repository and each new sequence of behavior was compared with the repository to identify new malware code. Applied rule-based IDS to tackle the DDoS attacks in which the resources are made unavailable for the user when they are required. The utmost capacity of each of the middle-ware layer was fixed and set of rules were formed to detect the DDoS attacks. The system produced an alert when the count of the requests to a particular resource exceeded a particular threshold and concepts from learning automata were employed to avoid further attacks.

Problem formulation:-
Following are the problems which is to be consider in a IDS based on machine learning approach (base paper) are as follows-

- Inferior detection accuracy in actual environments- Machine learning methods has a certain ability to detect intrusions, but they do not often perform well on completely unfamiliar data. When data set does not cover all typical real world samples it would decrease the accuracy.
- Low efficiency- most studies emphasizes the detection results; therefore they usually employ complicated models and extensive data processing methods, leading to low efficiency.
- Lacking of available data sets is the another major problem because the main task of machine learning is to extract the valuable information of data set. So if data sets are not available then it would be problem in detection.

3. INTRUSION DETECTION SYSTEM

Like other security measures like antivirus software, firewalls, and access control plans, Intrusion Detection Systems (IDS) are designed to improve the security of information and Internet of Things communication systems. The firewall's primary function is to sort packets according to allow/deny rules based on information in the header fields. The filtering of packets that pass through particular hosts or network ports, which are typically open on the majority of computer systems, is the firewall's primary function. It doesn't do deep analysis, which is like finding malicious code in a packet, and it treats each packet as a separate thing. An anti-virus program is a running process that, rather than monitoring network traffic, examines executables, worms, and viruses in the memory of protected computer/network systems [6].

While IDS requires more embedded intelligence than other security products like antivirus programs, it analyzes the information it collects and derives useful results [7]. This is the difference between IDS and other security products like antivirus programs. DARPA established the CIDF (Common Intrusion Detection Framework) working group in 1998 with the primary goal of coordinating and defining a common framework in the IDS field. This group has produced noteworthy work [8]. A general IDS architecture based on the consideration of the four kinds of functional modules depicted in Figure 1 was developed by the group, which was incorporated into the IETF in the year 2000 and adopted the brand-new acronym IDWG ('Intrusion Detection Working Group').

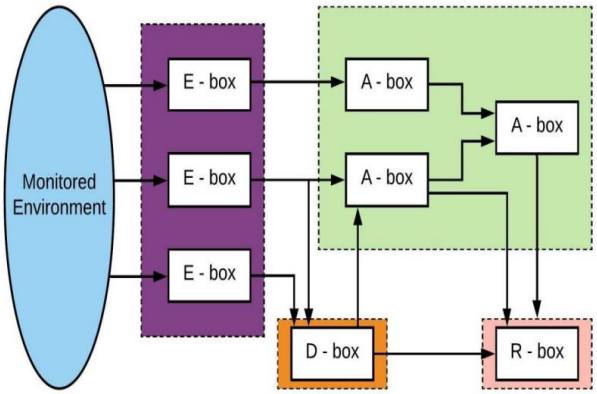

Figure 1: General CIDF architecture for IDS

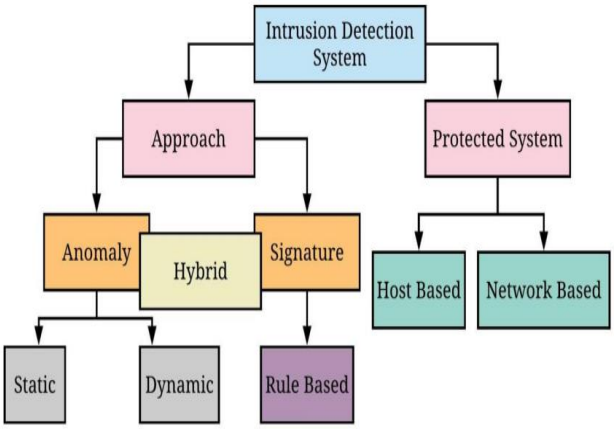

Figure 2: IDS Classifications

Contingent upon the sort of examination did, interruption location frameworks are delegated by the same token signature-based or abnormality based displayed in Figure 2. Signature-based plans (additionally indicated as abuse based) look for characterized examples, or marks, inside the dissected information. A signature database that corresponds to known attacks is specified a priori for this purpose. Anomaly-based detectors, on the other hand, attempt to estimate the "normal" behavior of the system that needs to be protected and issue an anomaly alarm whenever the difference between a specific observation and the normal behavior exceeds a predetermined threshold. Modeling the system's "abnormal" behavior and sending an alert when the difference between what is seen and what is expected falls below a certain threshold is another option.

For specific, well-known attacks, signature-based schemes provide excellent detection results. Even if they are designed as minimal variants of attacks that are already known, they are unable to detect new, unknown intrusions. Contrarily, the main advantage of anomaly-based detection methods [5] is that they

can pick up on intrusions that haven't been seen before. Anomaly-based Intrusion Detection Systems (A-IDS) are currently the primary focus of intrusion detection research and development due to their promising capabilities. Numerous novel plans are being considered, and numerous new systems with A-IDS capabilities are becoming available. Although there are a variety of A-IDS approaches, the fundamental modules or stages depicted in Figure 3 are common to all of them.

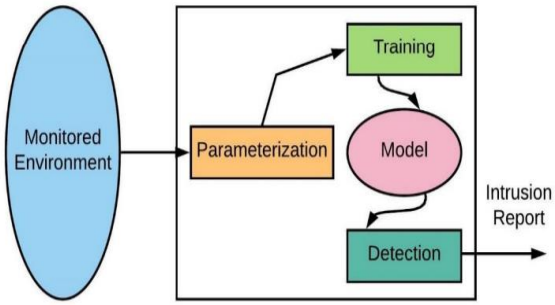

Figure 3: Generic Anomaly based IDS Functional Architecture

4. ML ALGORITHM

Supervised learning is two stage forms, in the initial step: a model is fabricate depicting a foreordained arrangement of information classes or ideas. The model developed by investigating database tuples portrayed by traits. Each tuple is expected to have a place with a predefined class, as dictated by one of the qualities, called to have a place with a reclassified class, as controlled by one of the traits called the class name characteristic. The information tuple are dissected to fabricate the model all things considered from the preparation dataset.

Learning

The main property of an ML is its capability to learn. Learning or preparing is a procedure by methods for which a neural system adjusts to a boost by making legitimate parameter modifications, bringing about the generation of wanted reaction. Learning in an ML is chiefly ordered into two classes as [9].

- Supervised learning
- Unsupervised learning

Supervised Learning

Regulated learning is two stage forms, in the initial step: a model is fabricated depicting a foreordained arrangement of information classes or ideas. The model developed by investigating database tuples portrayed by traits. Each tuple is expected to have a place with a predefined class, as dictated by one of the qualities, called to have a place with a reclassified class, as controlled by one of the traits called the class name characteristic. The information tuple are dissected to fabricate the model all things considered from the preparation dataset.

Unsupervised learning

It is the kind of learning in which the class mark of each preparation test isn't knows, and the number or set of classes to be scholarly may not be known ahead of time. The prerequisite for having a named reaction variable in preparing information from the administered learning system may not be fulfilled in a few circumstances.

Data mining field is a highly efficient techniques like association rule learning. Data mining performs the interesting machine-learning algorithms like inductive-rule learning with the construction of decision trees to development of large databases process. Data mining techniques are employed in large interesting organizations and data investigations. Many data mining approaches use classification related methods for identification of useful information from continuous data streams.

Nearest Neighbors Algorithm

The Nearest Neighbor (NN) rule differentiates the classification of unknown data point because of closest neighbor whose class is known. The nearest neighbor is calculated based on estimation of k that represents how many nearest neighbors are taken to characterize the data point class. It utilizes more than one closest neighbor to find out the class where the given data point belong termed as KNN. The data samples are required in memory at run time called as memory-based technique. The training points are allocated weights based on their distances from the sample data point. However, the computational complexity and memory requirements remained key issue. For addressing the memory utilization problem, size of data gets minimized. The repeated patterns without additional data are removed from the training data set.

Naive Bayes Classifier

Naive Bayes Classifier technique is functioned based on Bayesian theorem. The designed technique is used when dimensionality of input is high. Bayesian Classifier is used for computing the possible output depending on the input. It is feasible to add new raw data at runtime. A Naive Bayes classifier represents presence (or absence) of a feature (attribute) of class that is unrelated to presence (or absence) of any other feature when class variable is known. Naïve Bayesian Classification Algorithm was introduced by Shinde S.B and Amrit Priyadarshi (2015) that denotes statistical method and supervised learning method for classification. Naive Bayesian Algorithm is used to predict the heart disease. Raw hospital dataset is employed. After that, the data gets preprocessed and transformed. Finally by using the designed data mining algorithm, heart disease was predicted and accuracy was computed.

Support Vector Machine

SVM are used in many applications like medical, military for classification purpose. SVM are employed for classification, regression or ranking function. SVM depends on statistical learning theory and structural risk minimization principal. SVM determines the location of decision boundaries called hyper plane for optimal separation of classes as described in figure 1.4. Margin maximization through creating largest distance between separating hyper plane and instances on either side are employed to minimize upper bound on expected generalization error. Classification accuracy of SVM not depends on dimension of classified entities. The data analysis in SVM is based on convex quadratic programming. It is expensive as quadratic programming methods need large matrix operations and time consuming numerical computations.

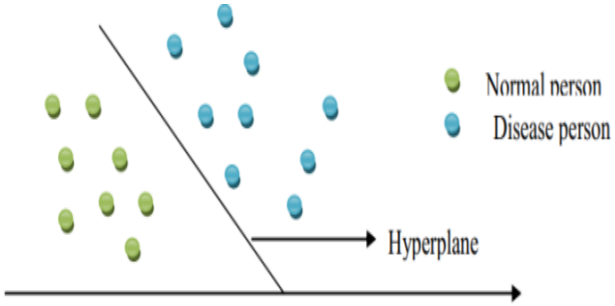

Figure 4: Support Vector Classification

5. PROPOSED METHODOLOGY

The proposed technique is based on LSTM technique. In this paper the explain only LSTM technique and further algorithm will explain result paper.

Long Short Term Memory is a kind of recurrent neural network. In RNN output from the last step is fed as input in the current step. LSTM was designed by Hochreiter & Schmidhuber. It tackled the problem of long-term dependencies of RNN in which the RNN cannot predict the word stored in the long-term memory but can give more accurate predictions from the recent information. As the gap length increases RNN does not give an efficient performance. LSTM can by default retain the information for a long period of time. It is used for processing, predicting, and classifying on the basis of time-series data.

LSTM has a chain structure that contains four neural networks and different memory blocks called cells**.**

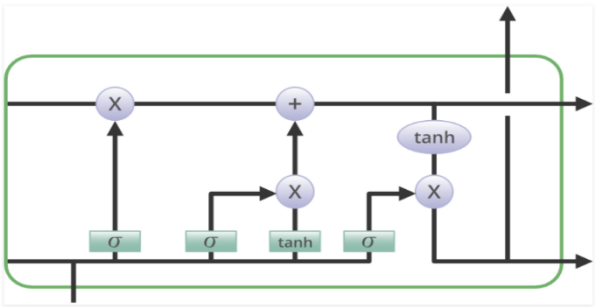

Figure 5: LSTM

Information is retained by the cells and the memory manipulations are done by the **gates.** There are three gates –

1. Forget Gate: The information that is no longer useful in the cell state is removed with the forget gate. Two inputs *x_t* (input at the particular time) and *h_t-1* (previous cell output) are fed to the gate and multiplied with weight matrices followed by the addition of bias. The resultant is passed through an activation function which gives a binary output. If for a particular cell state the output is 0, the piece of information is forgotten and for output 1, the information is retained for future use.

2. Input gate: The addition of useful information to the cell state is done by the input gate. First, the information is regulated using the sigmoid function and filter the values to be remembered similar to the forget gate using inputs *h_t-1* and *x_t*. Then, a vector is created using *tanh* function that gives an output from -1 to +1, which contains all the possible values from h_t-1 and *x_t*. At last, the values of the vector and the regulated values are multiplied to obtain the useful information

3. Output gate: The task of extracting useful information from the current cell state to be presented as output is done by the output gate. First, a vector is generated by applying tanh function on the cell. Then, the information is regulated using the sigmoid function and filter by the values to be remembered using inputs *h_t-1* and *x_t*. At last, the values of the vector and the regulated values are multiplied to be sent as an output and input to the next cell.

6. CONCLUSION

Intrusion detection systems are an important part of today's information technology-based enterprises' security. It's a difficult task to provide an efficient and high-performance IDS approach to deal with a wide range of security assaults. Deep learning approaches have recently been shown to be effective at solving intrusion detection challenges, and various deep learning-based IDS strategies have been published. Deep learning is a subset of machine learning techniques that employ multiple layers to do nonlinear processing and learn multiple levels of data representation. From this experiment found that deep learning is better than machine learning techniques. Malicious cyber-attacks can lurk in enormous amounts of legitimate data in unbalanced network traffic. In cyberspace, it uses a high level of stealth and obfuscation, making it difficult for Network Intrusion Detection Systems (NIDS) to ensure detection accuracy and timeliness. To address the problem of class imbalance, we will offer an enhanced long short term memory (LSTM) algorithm.

7. REFERENCES

[1] Kezhou Ren, Maohuan Wang, Yifan Zeng and Yingchao Zhang, *"*An Unmanned Network Intrusion Detection Model Based on Deep Reinforcement Learning*"*, IEEE International Conference on Unmanned Systems (ICUS), IEEE 2022.

[2] R. Ahsan, W. Shi, X. Ma, and W. L. Croft, "A comparative analysis of CGAN-based oversampling for anomaly detection," *IET* Cyberphysical Systems: Theory & Applications, vol. 7, no. 1, pp. 40–50, Mar. 2022.

[3] S. Dong, Y. Xia, and T. Peng, "Network Abnormal Traffic Detection Model Based on Semi-Supervised Deep Reinforcement Learning," IEEE Transactions On Network And Service Management, vol. 18, no. 4, pp. 4197–4212, Dec. 2021.

[4] Lan Liu, Pengcheng Wang , Jun Lin, and Langzhou Liu, "Intrusion Detection of Imbalanced Network Traffic Based on Machine Learning and Deep Learning", IEEE Access 2020.

[5] A. Raghavan, F. D. Troia, and M. Stamp, ``Hidden Markov models with random restarts versus boosting for malware detection," *J. Comput. Virol. Hacking Techn.*, vol. 15, no. 2, pp. 97107, Jun. 2019.

[6] Zhiyou Zhang and Peishang Pan "A hybrid intrusion detection method based on improved fuzzy C-Means and SVM", IEEE International Conference on Communication Information System and Computer Engineer (CISCE), pp. no. 210-214, Haikou, China 2019.

[7] Afreen Bhumgara and Anand Pitale, "Detection of Network Intrusion Using Hybrid Intelligent System", IEEE International Conferences on Advances in Information Technology, pp. no. 167-172, Chikmagalur, India 2019.

[8] Ritumbhira Uikey and Dr. Manari Cyanchandani "Survey on Classification Techniques Applied to Intrusion Detection System and its Comparative Analysis", IEEE 4th International Conference on Communication $ Electronics System (ICCES), pp. no. 459-466, Coimbatore, India 2019.

[9] Aditya Phadke, Mohit Kulkarni, Pranav Bhawalkar and Rashmi Bhattad "A Review of Machine Learning Methodologies for Network Intrusion Detection", IEEE 3rd National Conference on Computing Methodologies and Communication (ICCMC), pp. no. 703-709, Erode, India 2019.

[10] S. Sivantham, R.Abirami and R.Gowsalya "Comapring in Anomaly Based Intrusion Detection System for Networks", IEEE International conference on Vision towards Emerging Trends in Communication and Networking (ViTECon), pp. no. 289-293, Coimbatore, India 2019.

[11] Azar Abid Salih and Maiwan Bahjat Abdulrazaq "Combining Best Features selection Using Three Classifiers in Intrusion Detection System", IEEE International Conference on Advanced science and Engineering (ICOASE), pp. no. 453-459, Zakho - Duhok, Iraq 2019.

[12] Lukman Hakim and Rahilla Fatma Novriandi "Influence Analysis of Feature Selection to Network Intrusion Detection System Performance Using NSL-KDD Dataset", IEEE International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), pp. no. 330-336, Jember, Indonesia 2019.

[13] T. Sree Kala and A. Christy, "An Intrusion Detection System Using Opposition Based Particle Swayam Optimization Algorithm and PNN", IEEE International Conference on Machine Learning, Big Data, Cloud and Parallel Computing, pp. no. 564-569, Coimbatore, India 2019.

[14] Xiaoyan Wang and Hanwen Wang "A High Performance Intrusion Detection Method Based on Combining Supervised and Unsupervised Learning", IEEE Smart World, Ubiquitous Intelligence $ Computing Advanced $ Trusted Computing, Scalable Computing, Internet of People and Smart City Innovations, pp. no. 889-897, Guangzhou, China 2018.

[15] P. Singh and M. Venkatesan, "Hybrid Approach for Intrusion Detection System", IEEE International Conference on Current Trends Towards Converging Technologies (ICCTCT), pp. no. 654-659, Coimbatore, India 2018.

[16] M. Tavallaee, E. Bagheri, W, Lu and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 dataset", IEEE Symposium on Computational Intelligence for Security and Defense Applications, pp. no. 892-899, Ottawa, India 2018.

[17] Karuna S. Bhosale and Assoc. prof. Maria, "Data Mining Based Advanced Algorithm for Intrusion Detection in Communication Networks", IEEE International Conference on Computational Techniques, Electronics & Mechanical System (CTEMS), Belgaum, India 2018.