ISSN 2347 - 3983
Volume 9. No.12, December 2021
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter059122021.pdf
https://doi.org/10.30534/ijeter/2021/059122021

Machine Learning Techniques for Network Intrusion Detection System (NIDS): A Survey

Rajesh Dhakad[1], Shivani Katare[2]
[1]Shri G. S. Institute of Tech. and Science, Indore, India, rajesh_dhakad@rediffmail.com
[1]Shri G. S. Institute of Tech. and Science, Indore, India, shivanikatare.glg@gmail.com

Received Date : November 06, 2021 Accepted Date : November 28, 2021 Published Date : December 07, 2021

ABSTRACT

In computer network, security of the network is a major issue and intrusion is the most common threats to security. Cyber attacks detection is becoming more enlightened challenge in detecting these threats accurately. In network security, intrusion detection system (IDS) has played a vital role to detect intrusion. In recent years, numerous methods have been proposed for intrusion detection to detect these security threats. This survey paper study examines recent work in the topic of network security, machine learning based techniques as well as a discussion of the many datasets that are commonly used to evaluate IDS. It also explains how researchers employ Machine Learning Based Techniques to detect intrusions.

Key words: Network Security, Machine Learning, Data Mining, Intrusion Detection System, Classifiers, Dataset.

1. INTRODUCTION

The rapid usage of networking has made the world a tiny place and things get easily accessible. Due to the spite growth of tricks and attacking tools, efficient and effective method for intrusion detection has become the priority to protect the network [1]. In recent years, researchers were doing lots of experiment to analyzing, processing and arranging the data in to a systematic and orderly manner by employing a number of different data mining techniques. The problem of attack detection can be reduced by applying various data mining algorithms for classifying data [2]. The major goal of this study is to design a Network Intrusion Detection System (NIDS). NIDS, in comparison to other forms of IDSs, can identify the broadest range of harmful activity. The traffic is inspected by a NIDS to spot incoming and ongoing attacks on a network.

Intrusions are unauthorized and unusual operations that attempt to harm the system or data. Intrusion is defined as a series of interconnected acts taken by an attacker that leads to the compromise of a victim system [3].

Intrusion detection is a monitoring process to monitor the activity or action occurring in a network or in a system and monitoring them for signs of intrusions [3].

Intrusion Detection Systems are inspection devices that have been installed to a system's security wall to identify anomalous activity. Intrusion detection systems are typically designed to protect a particular network or device [3].

1.1 TERMINOLOGY OF INTRUSION DETECTION

The goal of IDS is to detect intruders and keep the system safe. As a result, an examination and classification of intrusiondetection techniques is required.

A. ***Intrusion Detection System Classification:***

1) Based on various data sources IDS are classified as follows:

Host-based detection: The architecture of the network for Host-based IDS is agent-based. It works for a single network device or host. It keeps track of the device's activities and sends an alert to the administrator if any changes or modifications to the system file are discovered [4].

Network-based detection: Network-based intrusion detection is a type of IDS that monitors data transfer between machines via a network. It keeps track of the network's incoming and outgoing traffic from a strategic point or locations. It analyses and monitors all incoming and outgoing traffic on the entire subnet and compares it to a database of known attacks to see if there are any similarities [4].

2) Based on distinct analysis methods IDS are classified as follows:

Misuse or Signature-Based Detection: In signature based detection approach, previously known intrusions are acknowledged and stored in the database. Security analyst takes decision to identify the pattern of intrusion based on their past experience. For incoming and out-going traffic on the network, IDS search for the patterns and signature. If any pattern match with the previously stored pattern is considered as the intrusion. Misuse based detection have a

main disadvantage that is, it can only detect the attack whose attack patterns are already known or store in the database. Unknown attack pattern cannot be detected by this method [4].

Anomaly-Based Detection: In the Anomaly Based Detection Approach, the user's regular behavior is first saved in the database, and then the user's present behavior is compared to the data stored in the database. When a big divergence is discovered, it is assumed that an intrusion has taken place [5]. Unknown threats and anomalous activity of a network or host are detected via anomaly-based detection. Anomalies can be discovered in a variety of methods. Researchers have created a variety of Machine Learning-based approaches to detect these anomalies. Although anomaly-based IDS detects new attacks, it has the disadvantage of producing a large numbers of false negative and false positives.

B. ***Limitation of IDS****:* IDS have some limitations which describe as follows:

- Traditional IDS can only detect known attacks. So the new unknown and original attacks cannot be identifiedby these traditional IDS.
- **Unstructured Data:** There are lots of incoming and on- going traffic on the network. This traffic data have no valid format. So there is a big problem to bring this traffic data into a valid format or in systematic manner.
- **False Negative:** Most of the IDS produced false negativestate. In this state an IDS mistakenly detect an activity as normal when the activity actually an attack.
- **False Positive:** Various IDS suffered with false positive state. In this state IDS identify normal activity as anattack.

By addressing all these mentioned problems, computational power can be reduced and detection rate of IDS can be improved with the help of machine learning Techniques. By using Machine learning based techniques network traffic datacan be analyzed properly and organized into a systematic manner. The learning process follows a data-centric approach.It is assumed that in the audit data every authorized and unauthorized activity have their footprint [3]. In Machine learning field classification is one of the methods. In thismethod a model is constructed from the pre-classified dataset. Many academics and researchers have worked on develop- ing IDS over the years. This paper summarizes recent researchin IDS. This paper listed the various Machine Learning based classification techniques and algorithms.

2. RELATED WORK

This section gives an overview of recent work that has been proposed to ensure a higher detection accuracy rate of IDS. There are three types of IDS methods: knowledge-based, statistics-based, and machine learning-based. The knowledge- based approach uses existing system data such as network traffic instances and protocol specifications to determine the requested actions [7]. In contrast, a statistics-based technique builds a statistical model of user behavior by collecting and analyzing every data record in a group of objects [8], whereas machine-learning algorithms identify deep hidden pattern-matching from training data [9]. This survey focuses mostly on machine learning-based approaches.

2.1 Machine Learning Approach

There are three main types of machine learning methodologies i. e. supervised learning, unsupervised learning and reinforcement learning.

A. ***Supervised Learning****:* It is named as supervised, because it used labeled data instances in training phase to train algorithm that classify the data instances and predicts the outcomes. The algorithm is trained until it can detect the hidden pattern and co-relation between the input data and output labels. Classification is one of the methods of supervised learning [4] [6] [10].

Various algorithms for supervised learning exist. Nearest Neighbor Algorithm, Artificial Neural Network, Support Vector Machine, Decision Trees (ID3, CART, C4.5, Random Forrest), Bayesian Statistics, K-nearest neighbour, Hidden Markov Model Boosting, Ensembles classifiers, Naive Bayes classifier, Bayesian Networks Logistic regression, Fisher Linear Discriminant, Perceptron). Quadratic classifiers are all examples of linear classifiers.

B. **Unsupervised Learning**: As the name indicates, unsupervised learning is a machine learning approach in which models are not trained with labeled data instances in training phase. Instead, model discover hidden pattern from the given data by itself. Clustering is one of the method of unsupervised learning. Fuzzy clustering, K-means clustering, Apriori algorithm, Eclat algorithm, Hierarchical clustering and Outliers detection (Local outlier factor), Self-organizing map are some of the unsupervised learners [5] [10].

C. **Reinforcement Learning:** Reinforcement learning refers to a computer system that interacts with a dynamic environment in order to achieve a specific goal [10].

2.2 *Supervised Learning Based Classifiers to Design Ids.*

A Machine learning technique or algorithm can be employed as a standalone classifier in the development of an IDS. In this section some machine learning-based techniques are discussed that have been utilized to construct IDS.

1) ***Naive Bayes****:* The Naive Bayes classifier considers that the training dataset's attributes are conditionally independent and so attempts to estimate the class-conditional probability using the given class label [11]. When only simpler relationships exist, the Naive Bayes classifier delivers the best results. The Naive Bayes classifier just needs to scan the training dataset once, which saves time and effort.

2) ***Decision tree****:* A decision tree is a tree-based classification technique that predicts the target class value for unknown test data instances based on a set of conditions and previously known data examples. On the basis of some decision rules, a decision tree classifier classifies unseen incoming test data examples [12]. Because of the easier implementation and simplicity, decision tree is one of the most popular as a single classifier [13]. There are two forms of decision trees: classification trees and regression trees [12].

3) ***K-nearest neighbor****:* In k-nearest neighbor (K-NN), a variety of distance measurement methods are used. TheK-NN technique finds K examples in the training dataset that are closest to the test examples, and then assigns the most common class among the training examples to the test example. The k-NN approach can be used for both classification and regression, but it is most commonly utilized for classification problems. K-NN algorithm is one of themost simple and non-parametric algorithm [14] that does not make any assumption on underlying data.

4) ***Support Vector Machine****:* Support Vector Machine (SVM) was first introduced in the mid-to-late 1990s [15]. The basic concept of using the SVM algorithm to develop an IDS is that it uses the training dataset to describe only the normal class objects or those that are not anomalous in the IDS, while the rest of the class objects are assumed to be anomalous [16]. SVM algorithm is extremely popular because it has the ability to handle multiple categorical and continuous variables.

5) ***Random Forest Classifier****:* It names as Random forest (RF) because it consists a large number of independent decision tree that operate as an ensemble. It generates decision trees from randomly picked data points, receives predictions from each independent tree, and votes on the one that produces the best outcome. The authors [17], used a random forest classifier to create a NIDS. RF algorithm is used to perform effective classification of attacks for IDS.

6) ***Deep Learning****:* Deep learning is a machine learning and artificial intelligence (AI) technique that mimics how people acquire certain types of knowledge. Data science, which covers statistics and predictive modeling, includes deep learning as a key component. It makes the task of data scientists easy and faster to collect, analyze, and understand massive amounts of data. Deep learning algorithm achieves this by employing a multi-layered structure of algorithms known as neural networks [18].

7) ***Artificial neural network (ANN)****:* One of the most unique fields of artificial intelligence is ANN. The underlying concept of an ANN is inspired by the biological neural networks that make up the operation of the human brain [19]. ANN is collection of large number of units that are interconnected in some pattern to provide communication between nodes or neurons. An ANN typically organized in layer manner in which each layer have some interconnected nodes which contain the activation function. The input layer, hidden layer, and output layer are the three layers that make up an Artificial Neural Network. The input layer accepts data in a variety of formats that the programmer provides. Between the input and output layers, the hidden layer acts as an intermediary. To uncover hidden patterns and features, it executes all of the transformations and calculations. After that, the hidden layers are linked to an output layer, which provides the detection result.

Table 1 : Research Papers Based On Supervised Learning Algorithms

Classifier	Paper Name	Author
K-NN	• Anomaly detection techniques for a web defacement monitoring service.	(G. Davanzo, 2011) [22]
Naive Bayes	• A NIDS by using a hidden naive bayes multiclass classifier. • Malicious web content detection by using naive bayes classifier.	(Levent Koc, 2012) [20] (YungTsung, ,2010) [21]
SVM	• Design Network traffic anomaly dete ction system by using an autonomous labeling approach to SVM. • Machine learning based malicious web content detection system. • A web defacement monitoring service for anomaly detection.	(Carlos A. Catania) [15] (Yung-Tsung, 2010) [21] (G. Davanzo, 2011) [22]
Decision Tree	• Malicious web content detection by Using Decision Tree Algorithm. • IDS based on Data Mining Approach. • A Machine learning approach for IDS.	(Yung-Tsung, 2010) [23] (Su-Yun Wua, 2009) [24]
ANN	• A stepping-stone IDS by using NN. • An IDS using NN classifier.	(HanChing Wu) [25] (S.Devaraju, 2013) [26]
Fuzzy Logic	• IDS using Data Mining based Approach.	(Su-Yun Wua, 2009) [24]
ID3	• An Efficient algorithm for NIDS.	(V. Jaiganesh, (2014) [27]

2.3 *Unsupervised Learning Based Algorithm to Design Ids.*

1) ***K-mean cluster****:* The K-mean algorithm is a popular clustering technique that aims to divide 'N' data points into 'K' clusters, with each data point selected by the clusters nearest mean. The approach of K-mean clustering is based on distance. The distance between the data points is calculated using the Euclidean metric approach. The number of clusters

defined by user at the execution time of algorithm. A numbersof solutions will be tested until the most suitable one is chosen[28].

*2) **Hierarchical Clustering**:* It is a technique of clustering whose aims to create a hierarchy of clusters. Agglomerative and Divisive are the two basic approaches for hierarchical clustering [28].

- **Agglomerative cluster:** It's a clustering method that uses a bottom-up approach. Clusters have sub-clusters, which have pairs of clusters, and sub-clusters are joined as one travelsup the hierarchy in this clustering process.
- **Divisive Cluster:** It's also a clustering method that uses an iterative clustering strategy. The cluster with the biggest diameter in feature space is chosen and divided into binary sub-clusters with a lower range using this method.

Table 2: Comparative Analysis of Machine Learning Techniques for Ids

Classifier	Method	Advantages	Disadvantages
Support Vector Machine	A SVM algorithm is a regression and classification based machine learning technique, it constructs set of hyper planes or a hyper plane in a high or infinite dimensional space [16].	• It gives high accuracy [15]. • This model is able to handle complex and nonlinear decision boundaries [21]. • Compare to other model this model have less prone to over fitting [4].	• Algorithm complexity is high [16]. • It required extensive memory space [15]. • Choice of the kernel is difficult [4]. • SVM speed for training and testingis slow [16].
K-NN Neighbor	In K-NN, a variety of distance measurement methods are used. The K-NN neighbor technique finds K examples in the training dataset that are closest to the test examples, and then assigns the most common class amongthe training examples to the test example [14].	• Task of implementation is simple [22]. • The model has high adaptive behavior. • Parallel implementation is easy by using this algorithm [4].	• Storage requirement is high [4]. • Classification and testing speedis slow [14].
Artificial Neural Network (ANN)	ANN is a computational processing unit Whose base theme is inspired by the biological neural networks that constitute functionality of human brain [19].	• It can tolerate noise data [19]. • Complex nonlinear relationships b/w independent and dependent variables implicitly detected by this model [25]. • It requires less statistical training [4].	• It suffer by over fitting [19]. • Computational burden is high. • Training time requirement is high [26].
Decision Tree (DT)	• Decision tree is a tree based classification technique that predicts the target class value for unknown test data instances based on a set of conditions and multiple previously known data examples. On the basis of some decision rules, a decision tree classifier classifies unseen incoming test data examples [12].	• Domain Knowledge not required to Constructs DT [4]. • High dimensional data can be handledby this algorithm [2] • Representation of the tree is easyto understand [23].	• It is unstable algorithm [1]. • it complex to create a tree With numeric dataset [4]. • Output attribute must be categorical [2]. • This model is limited to only one output attribute [1].
Na¨ıve Bayes	• The Naive Bayes classifier considers that the training dataset's attributes are conditionally independent and so attempts to estimate the class -conditional probability using the given class label [11]. When only simpler relationships exist, the Naive Bayes classifier delivers the best results.	• Construction of model is easy [6]. • Training is fast an easy [21]. • it can handle noisy data [4]. • This model is highly scalable [21].	• Attributes are conditionally autonomous, which is not right all the time [20].

2.4 *Ensemble Learning*

In simple, the process of training multiple machine learning models and combining their outputs together to get better prediction is known as ensemble learning [29]. Various machine learning models are used as a base to create an efficient model. Averaging the outputs of different models is one of the simple ensemble learning techniques. Bagging and booting are also common methods for combining various learners [30] and there are many more complex algorithms and techniques developed to combine the prediction of many base machine learning models together.

2.5 *Hybrid Classifiers*

The term hybrid classifier refers to a classification system that uses multiple Machine Learning approaches or algorithms. A hybrid method is used to improve the detection rate of IDS. There are some papers in which author uses hybrid classifiers to build an efficient intrusion detection system [31].

• An IDS is implemented by using a combination of fuzzy clustering, K-NN and Damper-shafer theory to improve the performance of NIDS in [32] Chan TS, Yen KK and Luo J, using an high and low dimensional feature spaces with correlation analysis.

• Arif jamal malik et.al, proposed"Hybrid binary PSO and random forest (RF) based NIDS". To select most appropriate features for classification author used BINARY- PSO. Random forest classifier algorithm is used to classify the classes of attacks [33].

• Mohammad abu alsheikh et al in [34] have proposed a multilevel based classification model by using Neural Network and K-NN classification algorithm. For anomaly detection K- NN classifier was used to classify the given data into one of thetwo category either normal or anomaly. In next step a neural network is used to detect a specific type of attack in anomaly.

Table 3 : Research Paper Based On Hybrid Classifier

Classifier	Paper Name	Author
ANN Fuzzy Clustering	• An approach to Intrusion Detection using ANN and fuzzy clustering.	Gang Wang, 2010 [35]
Genetic – algorithm	• Real-time Anomaly Detection systems for DOS attacks by weighted k-NN classifier.	Su, 2011 [36]
C4.5 Decision Tree Support Vector Machine. K-means clustering	• A new clustering based appr--oach for Anomaly IDS. – –	Ravi Ranjan, 2014 [37] – –
Artificial Neural Network Fuzzy Clustering.	• Anomaly Based Intrusion Detection System Using Artificial Neural Network and Fuzzy Clustering. –	Prof. D.P.Gaikwad, 2012 [38]
Random Forest Rough Set Theory	• Hybrid Approach for Network Intrusion Detection System.	B.Nanda,Ajay Parikh 2019 [39]
J48 classifier Random Tree	• Hybrid Intrusion Detection System by combining Data Mining Techniques.	Elekar,Kailas Shivsankar (2015) [40] – .

3. DATASETS USED IN RESEARCH

Dataset has an important role in the task of designing machine learning based algorithm for e.g., Regression, Classification, Clustering and Function learning [41]. Datasets reviewed in this paper is for the purpose of classification. There are various dataset available for NIDS. This paper show the datasets which are frequently used in previous years to develop IDS.

A. DARPA 1998/99

The most commonly used datasets for intrusion detection are DARPA 1998/99. Within an imitated network environment, DARPA datasets were developed at the MIT LINCOLN lab. The DARPA 1998 and DARPA 1999 datasets are packet-based datasets that comprise network traffic for five weeks. Various types of attacks such as buffer overflow, DoS, rootkits, and port scan are included in this packet-based network dataset [42].

B. KDDCUP_99

The MIT LINCOLNS lab introduced the KDDCUP_99 dataset, which is based on the DARPA dataset, for evaluating and surveying intrusion detection studies. The KDDCUP 99 dataset has 41 attributes, 38 of which are discrete numeric or continuous, and 3 of which are categorical. KDDCUP 99 has four different groupings of attributes: 1. Basic attributes, 2. Content attributes and 3. Time- based attributes and 4. Host–based traffic attributes [24]. By studying the KDDCUP 99 dataset, a total of 23 levels were discovered, one of which falls under the normal data category while the other 22 levels are attacks. These 22 attacks are categorized into four groups: DOS, U2R, PROBE, and R2L [43]. This dataset can be obtained from 3 files known as KDD full dataset, corrected KDD dataset and 10% KDD.

C. NSL KDD

There is a lot of redundancy in the data of KDDCUP 99 dataset, so the NSL KDD dataset was created to overcome the problem of KDD dataset. The NSL KDD dataset, as a result, has around 150,000 data points. The NSL KDD training set contains 125973 data points, whereas the testing set contains 22544 data points. The number of attributes in the NSL KDD dataset is the same as in the KDDCUP 99 dataset [36] [42].

D. Kyoto 2006+:

Kyoto 2006+ dataset comprises real network traffic data obtained from several honeypots. The Kyoto 2006+ dataset has 24 features, 14 of which are identical to KDDCUP 99 and the remaining 10 are flow-based, such as IP address, ports or time. The type of attack is indicated by a label feature. Three years of network traffic were collected in the Kyoto 2006+ dataset [44].

Table 4: Dataset Analysis

Dataset	Year of traffic creation	Normal traffic	Attack traffic	Count	labeled
DARPA	1998 1999	Yes	Yes	N.S.	Yes
KDDCUP_ 99	1998	Yes	Yes	5M points	Yes
NSL KDD	1998	Yes	Yes	150k points	Yes
KYOTO 2006+	2006 -2009	Yes	Yes	93M points	Yes

4. CONCLUSION

Now-a-days in the era of high speed network, Machine learning based methods are extensively used to analyzing large volume of traffic data on network. In designing of IDS a significant challenge is reducing false negative and false positive state and achieving high intrusion detection rate. To address these issues, this paper presented a study of detailed overview of IDS and explanation of machine learning based approaches for network intrusion detection. These machine learning based methods are working well to detect intrusion. Many scholars and researchers have worked on the design of IDS over the years. But still more researches and study needs to be done to improve the detection rate and to reduce the false state.

5. FUTURE SCOPE

In future going to design a NIDS by using Machine leaning based Decision Tree Model to improve detection rate by detecting anomalies and reducing false negative state. And

then try to reduce the computation power requirement of traditional signature based IDS based on the output of anomaly based IDS.

REFERENCES

1. **Decision Tree: A Machine Learning for Intrusion Detection** Shilpashree. S, S. C. Lingareddy, Nayana G Bhat, Sunil Kumar G.
2. Neha G.Relan, prof. Dharmaraj R. Patil, **"Implementation of Network Intrusion Detection System using Varient of Decision Tree algorithm"** IsCNTE-2015
3. **"AN IMPROVED METHOD TO DETECT INTRUSION USING MA- CHINE LEARNING ALGORITHMS"**. Urvashi Modi1 and Anurag Jain2. CSE departments, Radharaman inst. of Tech and Science, Bhopal, India
4. "**Network Intrusion Detection using Selected Data Mining Approaches: A Review".** Munawara Saiyara Munia, Samira Samrose, Pranab Dey.
5. Rachna kulhare, and Dr. Divakar **Singh "Survey paper on intrusion detection techniques"** October 2007.
6. Harshna (M.Tech C.S.E), NavneetKaur.**'Survey paper on Data Mining techniques of Intrusion Detection**' April 2013.
7. S. Elhag, A. Ferna´ndez, A. Bawakid, S. Alshomrani, and F. Herrera, **"On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on intrusion detection systems,**" 2015.
8. Chao, S. Wen, and C. Fong, "CANN**: an intrusion detection system based on combining cluster centers and nearest neighbors, Knowl-edge Based System"**2015
9. Meshram A, Haas C (2017) **Anomaly detection in industrial networks using machine learning: a roadmap.** In: Beyerer J, Niggemann O, Ku¨hnert C (eds) Machine learning for cyber physical systems.
10. Anderson, J. (1995). **An introduction to neural networks.** Cambridge: MIT Press.
11. Dewan Md. Farid, M. Z. (2011). **Adaptive Intrusion Detection based on Boosting** .
12. Chih-Fong Tsai, Y.-F. H.-Y.-Y. (2009). **Intrusion detection by machine learning: A review**. expert systems with applications,ELSEVIER .
13. Dewan Md. Farid, L. Z. (2013). **An Adaptive Ensemble Classifier for Mining Concept-Drifting Data Streams**.
14. C.M.Bishop. (1995). **Neural networks for pattern recognition.** Eng- land:Oxford University.
15. Carlos A. Catania, F. B. (2012). **An autonomous labeling approach to support vector machines algorithms for network traffic anomaly detection.**
16. Tax, D. (1999). **Data domain description using support vectors.Proceedings of the european symposium on artificial neural networks**,251-256.
17. Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016) **Random Forest Modeling for Network Intrusion Detec- tion System** Nabila Farnaaz and M. A. Jabbar MJCET Hyderabad, India.
18. Tich Phuoc Tran, L. C. (2009). **Novel Intrusion Detection using Proba- bilistic Neural.**
19. Haykin, S. (1999). **Neural networks: A comprehensive foundation (2nd Edition).** New Jersey: Prentice Hall.
20. Levent Koc, T. A. (2012). **A network intrusion detection system based on a Hidden Na¨ıve Bayes multiclass classifier.**
21. Yung-Tsung Hou, Y. C.-S.-M. (2010). **Malicious web content detection by machine learning.**
22. G. Davanzo, E. M. (2011). **Anomaly detection techniques for a web defacement monitoring service.**
23. Yung-Tsung Hou, Y. C.-S.-M. (2010). **Malicious web content detection by machine learning.**
24. Su-Yun Wua, E. Y. (2009). **Data mining-based intrusion detectors.**
25. Han-Ching Wu, S.-H. S. (2010). **Neural networks-based detection of stepping-stone intrusion.**
26. **"DETECTION OF ACCURACY FOR INTRUSION DETECTION SYSTEM USING NEURAL NETWORK CLASSIFIER.**"S. Devaraju, S. R. (2013).
27. **'An Efficient Algorithm for Network Intrusion Detection System'.** V. Jaiganesh, P. Rutravigneshwaran,P. Sumathi, Ph.D.
28. Li Tian1, Wang **"Research on Network Intrusion Detection System Based on Improved K-means Clustering Algorithm"**,2009.
29. Chih-Fong Tsai, Y.-F. H.-Y.-Y. (2009). **Intrusion detection by machine learning: A review.**
30. Dewan Md. Farid, M. Z. (2011). **Adaptive Intrusion Detection based on Boosting .**
31. Ravi Ranjan, G. S. (2014). **A NEW CLUSTERING APPROACH FOR ANOMALY INTRUSION DETECTION.**
32. Chan TS, Yen KK and Luo J., **"Network intrusion detection design using feature selection of soft omputing paradigms"**.
33. Arif Jamal Malik, Waseem Shahzad, Farrukh Aslam Khan, **"Networ lD using hybrid binary PSO and RF algorithm".**
34. Mohammad Abu Alsheikh, Shaowei Lin, Dusit Niyato1 and Hwee-Pink Tan**, "Machine Learning in Wireless Sensor Networks".**
35. Gang Wang, J. H. (2010). **A new approach to intrusion detection usingArtificial Neural Network.**
36. Su, M.-Y. (2011). **Real-time anomaly detection systems for Denial-ofService attacks by weighted.**
37. Ravi Ranjan, G. S. (2014). **A NEW CLUSTERING APPROACH FOR ANOMALY INTRUSION DETECTION .**
38. Prof. D.P. Gaikwad, S. J. (2012) **Anomaly Based**

Intrusion Detection System Using Artificial Neural Network and Fuzzy clustering.

39. **'Hybrid Approach for Network Intrusion Detection System' using Ran- don forest and Rough SetS"** B.Nanda Ajay Parikh 2019.
40. Elekar, Kailas shuvsankar, **combination of Data Mining methods for intrusion detection system.**
41. Sahu, Shailendra, and B. M. Mehtre, **Network intrusion detection system using J48 Decision Tree, Advances in Computing.**
42. **'A Survey of Network-based Intrusion Detection Data Sets'**.Markus Ring, Sarah Wunderlich, Deniz Scheuring, Dieter Landes and Andreas Hotho.
43. Kayacik, H. Gnes, A. Nur Zincir-Heywood, and Malcolm I. Heywood. **Selecting features for intrusion detection: A feature relevance analysis on KDD 99 intrusion detection datasets**(2005).
44. **'Intrusion Detection System using Data Mining A Review'**. varsga singh, Subha Puthran(2016).