Journal of Science, Computing and Engineering Research (JSCER)
Volume-3, Issue- 5, September –October 2022.
DOI: *https://doi.org/10.46379/jscer.2022.030501*

Efficient Network Threat Detection And Classification Method Using SVM+NB Algorithm On Cloud Computing

[1]Abishek.M, [2]Anil Kumar

[1,2]Asst.Professor, Department of Mechanical Engineering, PERI Institute of Technology,Chennai.

Article Information

Received : 06 Sep 2022
Revised : 22 Sep 2022
Accepted : 10 Oct 2022
Published : 12 Oct 2022

Corresponding Author:

Abishek.M

Abstract— In this work, NSL-KDD dataset is considered for processing the session flow and to analyse threat which is related to files in the cloud. Also, feature extraction is performed based on two essential features like high level features and low level features which are considered for classification purpose. Here, three effectual classifiers like efficient Naive Bayes (NB) classifier, Improved Support Vector Machine (ISVM) and Artificial Neural Networks (ANN) are utilized over the features extracted from the dataset for intrusion detection. Empirical outcomes depict that the selected features will offer better outcome to design the threat detection system which will be more efficient for cloud based network security. Results acquired from this methods show better trade off in comparison with existing systems like decision tree, random forest and so on. Performance metrics like as precision with 100%, accuracy with 99.5%, specificity with 100%, Recall with 100% and F measure with 100% has been evaluated with this machine learning algorithms.

Keywords: ***Artificial Neural Network, Cloud storage, Machine Learning algorithms, NB classifier, SVM.***

Citation: Abishek.M, Anil Kumar. "Efficient Network Threat Detection And Classification Method Using SVM+NB Algorithm On Cloud Computing, Journal of Science, Computing and Engineering Research, 3(5), 01-11, Sep- Oct 2022.

I. INTRODUCTION

With the emergent growth in internet technology, various strategies for hacking information have also been raised. Even though there are numerous deals for protecting methodologies such as encryption and other protecting schemes [1]. However, there are certain types of intrusion that is to be detected.

As well, there is a bigger concerns related to cloud users and enormous hurdles to transfer their in- house business towards public cloud security. The problem might be that cloud vendors have not offer data security whilst being transported towards cloud environment. Furthermore, remote access comes under the influence of clients‟ responsibility and users‟ needs to possess stronger security knowledge associated to internet and networking, before acquiring access to cloud [2]. Anything that is connected with internet is not secured perfectly. Consequently, numerous data breaches occur.

Information stealing is the foremost concern when dealing with cloud computing, based on the Breach level Index website, there is roughly 39891145672 records were stolen in 2013. In 2016, roughly of 554 million records were lost, by reaching highest peak in October (approximately 472 million stolen data). Even technology giants like Microsoft was not left out, with 8 recorded data breaches in the year 2013-2016.

Intrusion Detection (ID) methodologies utilize two ideas: Anomaly Detection and Misuse Detection [3]. The former recognizes pattern sourced attacks departed from normal characteristics whilst the latter recognizes pattern sourced attacks that is inhibited from acknowledge intrusions. Accuracy is essential factor for intrusion detection, extensibility and adaptability as networking is crucial in network computation. Foremost crisis associated with present intrusion detection is the failure to identify 0-day attacks. Anomaly detection are generally adaptive where it cannot be proficient to foresee new attack, still it cannot recognize certain types of attack. From this view point, machine learning approach is cast off to enhance accuracy and other performance related factors.

Numerous cloud based attacks like fraud and phishing, malicious insider, unauthorized access, insecure interfaces etc. causes‟ data breach. An effectual classification model for cloud security threat [4] sourced on machine learning algorithm is used effectively.

Machine learning tools are measured as an effectual tool and utilized in government, financial services, health care, information security, transportation, etc. [5] for diverse prediction problems. These procedures attain computers to work without explicitly programmed to carry something. However, it makes computers to learn from previous experiences to resolve certain future crisis. By measuring the extensive usage, machine learning algorithms for cloud computing based threat detection is analysed and investigated. Intrusion Detection strategy for classification to effectually extract misuse detection to produce general network profile for anomaly detection, and then design an effectual classification algorithm to identify attacks The ID

classification methodology is supervised learning approach to classify traffic caused by certain types of attacks. Those sorts of techniques are time saving based on security aspects and assists in analysing attack data.

1.1 Work organization

This paper is organized as follows: Section 2 outlines review of literature on cloud security threats. Section 3 introduces our proposed model for threat classification based on the feasibility of machine learning methods to detect them. Section 4 explains the attained results of proposed work and section 5 shows the conclusion of the proposed method.

II. RELATED WORKS

In [6], Monjur Ahmed, foreseen that CC gives opportunities as it is progressively installed into computing functions. Dynamically uncover them and challenges to recognize and contextualize confronts. If security in CC is not tended to adequately, results might be extreme. Generalized threat taxonomy classification gives identifying context, recognizing and resolving threats (2014). Such scientific classification guarantees more secure CC practice and those real patterns are not provided for degree when building up coordinated and robust surveillance system.

In [7], Mouna Jouini, exhibited hybrid threat classification concept that facilitates characterizing and threat classification. For sure, it acts as guidelines to monitor attacks impact and helps in understanding its security determination by exhibiting threat procedures and potential effects as well by joining existing threat criteria (2014). Classification model utilization uses Cyber Security Econometric Model and applies it on functional application named CC.

In [8], Sanchika Gupta, distinguished the basic threats in Cloud domain from clients and specialist service providers while delivering Cloud services for guaranteeing abnormal security towards resultant assaults (2013). Besides, investigations distinguish basic hazard from such attacks and threats all through CC.

In [9], Seyyed Mohsen Hashemi, depicted outrageous advantages in utilizing a cloud based framework, yet there are numerous potential dangers which must be managed. We trust that first efficient starting stage for enhancing Cloud Computing security is precise ID of its threats.

In [10], Madhan Kumar Srinivasan addresses cloud computing scientific categorizations for present cloud computing frameworks. This paper can be a superior basement for any cloud service provider (CSP) to fabricate (new CSP) and additionally improve (existing CSP) their present cloud framework to progressively secure and advanced which in turns give shared advantages to both CSP and client.

In [11], Anurag Jain, clarified Cloud processing is a most recent innovation widely considered in present years. It is a model through which IT services are conveyed and charged based on use. Yet at the same time present advances are not sufficiently grown to use its full capacity. There are numerous immaculate issues in this area, including security domain, in co-operate security management, load balancing, stored data management, automatic service provisioning and energy management have quite recently begun to get consideration from academician and industry. Along these lines, we trust that still there is extent of research.

In [12], Raneel Kumar, proposed IDS has demonstrated promising outcomes in recognizing the 7 sorts of DoS attacks which are comprehensively arranged as volume, protocol, and application based attacks. The research novelty has been the utilization of time sensitive traffic stream features with one-class SVM calculation in the plan of detection framework. Alongside the proposed IDS component, this paper gives bits of knowledge into conveying test bed on the cloud for experimentation. As well, dataset created in this examination has been made publically accessible to energize further research in this field.

In [13], Zhijiang Chen, proposed CC based monitoring and threat detection framework to protect basic foundation frameworks by anticipated framework such as agents, infrastructure, and operation centre. Chen utilized both Hadoop Map Reduce and Spark to accelerate information by isolating data streams. To assess adequacy of system risk, assessed viability of proposed framework as for system checking, threat detection, and performance of systems.

In [14], Thu Yein Win, present root kit and malware discovery framework to secure hidden virtualization infrastructure against digital assaults in attacks in cloud security approaches, it utilizes framework associated with hashing and also the utilization of SVM with VMI. Utilization of secured in-VM screen guarantees that inner visitor VM is precisely acquired without imperilled, while utilization of a disconnected SVM in remote monitoring takes into consideration rapid attack classification

In [15], Md Tarique Prwez, proposed structure decision stumps are used as powerless classifiers. Because of frail classifiers, anticipated structure has lesser computation complexity. Because of Strong classifier, FAR is limited and location is amplified.

In this investigation, [16] data security in cloud computing is examined with datasets using various cloud computing techniques like Efficient Naive Bayes classifier, Improved Support Vector Machine and Artificial Neural Networks. The purpose to carry out this investigation is to acquire best algorithm in terms of metrics like: Accuracy, Specificity, Sensitivity, Precision, F-Measure, F1 Score and Recall. This investigation demonstrates that this technique works well while using the above mentioned algorithms than the prevailing techniques in cloud environment

III. PROPOSED SYSTEM

3.1 Dataset- NSL-KDD Dataset

The NSL KDD dataset is publicly obtainable for investigation related to intrusion detection. NSL- KDD dataset is suffered from minor crisis and it is not perfectly suited for prevailing networks, it is utilized in benchmark

TABLE 1: ATTRIBUTE OF NSL KDD DATASET

protocol type	Service	src_bytes	dst_bytes	num failed logins	logged in	num root	srv_count	serror_rate	srv_error rate
'udp'	'other'	146	0	0	0	0	1	0	0
'tcp'	'http'	287	2251	0	1	0	7	0	0
'tcp'	'http'	232	8153	0	1	0	5	0.2	0.2
'tcp'	'http'	199	420	0	1	0	32	0	0
'tcp'	'http'	233	616	0	1	0	3	0	0
'tcp'	'http'	343	1178	0	1	0	10	0	0
'tcp'	'http'	300	13788	0	1	0	9	0	0.11
'tcp'	'http'	253	11905	0	1	0	10	0	0
'udp'	'other'	147	105	0	0	0	1	0	0
'tcp'	'telnet'	437	14421	0	1	0	1	0	0
'tcp'	'http'	227	6588	0	1	0	22	0	0
'tcp'	'http'	215	10499	0	1	0	14	0	0
'tcp'	'http'	241	1400	0	1	0	33	0	0
'tcp'	'http'	303	555	0	1	0	9	0	0
'udp'	'domain_u'	45	45	0	0	0	181	0	0
'udp'	'private'	105	147	0	0	0	2	0	0
'udp'	'domain_u'	43	43	0	0	0	202	0	0
'tcp'	'http'	324	2302	0	1	0	28	0	0
'tcp'	'smtp'	1591	372	0	1	0	2	0	0
'udp'	'other'	146	105	0	0	0	2	0	0
'tcp'	'http'	290	3006	0	1	0	11	0	0
'tcp'	'http'	255	861	0	1	0	44	0.02	0.02
'tcp'	'http'	302	498	0	1	0	10	0	0
'tcp'	'http'	220	1398	0	1	0	42	0	0
'udp'	'domain_u'	43	69	0	0	0	120	0	0

dataset to identify intrusions. While using NSL-KDD dataset, attacks lies in categories given below: The NSL KDD dataset is publicly obtainable for investigation related to intrusion detection. NSL- KDD dataset is suffered from minor crisis and it is not perfectly suited for prevailing networks, it is utilized in benchmark dataset to identify intrusions. While using NSL-KDD dataset, attacks lies in categories given below:

Probing: This attack collects information about target system before attack initiation. Certain examples associated with this attack are ipsweep, satan, nmap.

DoS: It yields an outcome by eliminating requests to resource while bandwidth consumption or during overloading resources. For instance, Neptune, Teardrop.

User to Root (U2R): Here, attacker imitates access to normal user over system and capable to exploit vulnerabilities to acquire access to system. For instance eject and Perl belongs to U2R.

Root to Local (R2L): Here, attacker doesn‟t possess account on remote location to machine over network and utilizes

vulnerability to acquire local access to machine. Few instance: guess password, ftp_write and imap attacks.

3.2 Data Pre-processing

Before commencing classification process, primary cleaning phase of dataset used is essential. This phase comprises of four main sub phases Approximately more than 50% of instances are found to be attacks. Therefore, to acquire reduced rate of attack and to balance distribution of attack class and normal classes has to be performed. The sum of attacks in the dataset has been adjusted. For instance, 50% of network connection is considered as attack connection whilst the remaining is measured as normal connection.

The dataset considered in this work, i.e. NSL-KDD dataset base is not balanced properly. It is formally dominated by 11,656 instances of probes. Here, Denial of service (DoS) attacks possesses numerous instances, i.e. 45, 927. In this case, it is essential to eliminate certain types of attacks to acquire appropriate balancing of other sorts of attacks such as U2R (995) and R2L (52 instances).

The data is divided into two parts comprising of normal and attack types. Henceforth, decompose the chosen data into three subsequent sets with proper length: training, validating and testing data. To properly illustrate the anticipated algorithm, this investigation considers 1000 instances for testing and validating. Then, 50% attack rate is imposed to preserve distributions of attack types. The sample for standardized data is shown in the table 1.

3.3 Features Extraction

In this investigation, two types of features have been considered (higher level features (HLF) and lower level features (LLF)). LLF are hauled out from dataset directly (IP source, port number source, IP destination and port number destination).Higher level features are computed from ANN, SVM and Naive Bayes classifiers decision.

1. Feature Extraction: In this segment, statistical features and additional feature are chosen and extracted to produce input vector that has to be cast off in classification and detection.

2. Data Enrichment: If data is needed for analytical model, enrichment produces computing or merged data collected.

3. Normalization: To examine data accurately, normalize feature value from appropriate range. Else, it can leads to unexpected error as scale of each features.

3.4 Classification

Most of the classification model for security threat is generally restricted in any one or two criteria to categorize threats and other non-exhaustive threats (all will not be considered during classification) and these classification is not mutually exclusive [17].

This is appropriate for constant circumstances like (small organization) where threats are extremely constant, however in continuous varying circumstances, organizations fails to provide security against threats [18]. Usually, they are prone to various sorts of threats as it influences reputations and they recognize threat [19] to mitigate risks.

Classification facilitates threat detection and threat affects resources, therefore secure resources is essential [20]. As well, it assists managers to construct information systems of organizations‟ with least vulnerability. As well, significant problems can be recognized in prevailing threats. Actually, prevailing models do not maintain principles. General solution is to merge classifications and generate hybrid model.

From above discussions, hybrid model for classification in cloud system security is proposed here, which is termed as Multi-dimensional model for classification intended to follow all classification principles. Idea behind this anticipated method is to merge classifications criteria and to illustrate probable impacts. Classification list [21] is given below:

Threat source: Internal or external originating region.

Threat agents: it identifies three significant classes: environmental, human, and technological. Threat motivation: Attackers goal on system which may be non-malicious or malicious.

Security threat intention: Human intent which causes threat is accidental or intentional. This criterion provides full malicious behaviour and reconstructing attack behaviours to recognize its intention. It provides factor to assist investigators to offer high accuracy and therefore diminish risks to speed up decision making for capturing agent.

Threats impacts: It is security violation that outcomes from action. Here, threat impacts are recognized [22 and 23]: Information destruction, corruption, Loss/theft, disclosure, DoS, elevation and illegal usage.

Here, security threat classification is analysed based on system performance, based on five basic criteria leads to numerous threat classes. The following criteria are designed in classification model: Agent, source, intention, motivation and impacts.

The threat is classified, initially based on sources. Here, threat is caused either within organization or external origin of point [24]. Instead, threats may be internal or external which is owing to natural processes that commences from system boundaries.

As well, threats may be natural and provided without mistakes committed or malicious are owing to unintentional functional. Actions are well known with user objective: Malicious/Non malicious threats [25]. Threats are partitioned based on attacker‟s intent.

Intentional threats- Threats are occurred due to its chemical and physical processes over material [26]. These threats are commences without mistakes committed and malicious goals caused owing to unintentional actions.

3.5 Naive Bayes Classifier

NB is an effectual probabilistic classifier that evaluates probabilistic set by evaluating frequency and values in provided NSL-KDD dataset. Here, NB classifier utilizes some set of lower level features to recognize the incoming threats and influence towards the content in cloud storage. The probabilistic occurrence of threat in system storage is also considered as feature for training classifier. These features are extracted from the selected datasets.

Naive Bayes approach makes use of Bayes theorem to describe the probabilistic occurrence of threat (P –which is defined in equation 1). Some features possess higher influence to identify that file received from a source has the probability of possessing threat or not.

For instance, suppose if a packet is incoming to cloud environment, class of the packet, hypothesis of the packet, flow of packet in the network, initial probability of threat occurrence have been learned from the probability of the given network environment. The speed and accuracy of these computations is higher in Naive Bayes classifier.

(i) P(packet model) is probability that hypothesis holds the packet.

(ii) P(threat) is probability of packet being threat.

(iii) P(class) is probability that intrusion occurs will message transmission.

(iv) P(non − threat) is probability that packet is not threat.

(v) P(classnon − redundant) is probability that class appears in any particular class Algorithm.

Algorithm

Input : Input from NSL-KDD dataset

Output: Threat identification

Step 1: Initialize file

Step2: Extract feature with low level and high level representation (Source_Id, destination_Id and so on)

Step 3: Training dataset with NB.

Step 4: Evaluate threat probability and non-threat occurence.

P_threat = (sum(train_matrix(threat_indices,)) + 1) ./ (threat_classes + numtokens) P_nonthreat = (sum(train_matrix(nnon-threat_indices,)) + 1) ./ (non threat_classes + numtokens)

Step 5: Testing data log_a = test_matrix*(log(p_class_threat))' + log(p_threat)

log_b = T_matrix*(log(p_classes_nonthreat))'+ log(1 - p_threat) if output = log_a > log_b then file are threat else files have non-threat

Step 6: Classify malicious and non-malicious threat

Step 7: Identify error rate and evaluate wrongly classified class wrong = num docs_wrongmtest_docs

3.5.1 Description

In this investigation, Bayes theorem is used to carry out threat classification. Initial step is to choose file from dataset and use feature extraction, i.e. both lower and higher level representations are considered to know the intention of attack towards the file. Subsequent step is to train chosen dataset by the extracted features. For training data, compute lower and higher level probability attributes. Final step is to test data using NB classifier for which probability of threat occurrence attributes and non-threat attributes is evaluated. Thus, predictions are made to identify the higher probabilistic value. If low attributes are more than higher attributes while computation, then it is R2L or U2R threat else it is DoS attack.

3.5.2 Support Vector Machine

Here, SVM is discussed on threat classification. SVM has the ability to solve non-linear classification crisis using kernel trick and the schematic diagram is shown in figure 1. Because of this reason, SVM is very popular. The principle behind SVM is to design hyper-plane to guarantee relation between the attributes, i.e. lower level or higher level attribute class. While the probability of occurrence is maximized, SVM reduce the chance of wrong classification prediction to reduce the false positive values. With this, SVM is needed for optimization technique.

A separable training set is provided, which is composed of vectors, SVM has to design an optimal hyper-plane as in equation 2, Where, b vector; If (mi,ni) specifies samples from training set, d and *+, where „d‟ is specified as vector dimension of mi and ni is specified as classification outcome of feature vector bi.

If hyper-plane can partition sample set, the following condition has to be satisfied as in equation 3; To maximize the classification rate of the sample set, false positive or true negative has to be minimized as in equation 4 & 5; By resolving the optimization crisis, the following result has been attained;

Where, M is Lagrange multiplier. If training set is not capable for partitioning, relaxation factor and penalty parameter „P‟ is given. Optimal objective function is transformed as in equation 6 and with following constraints as in equation 7. SVM considers sign of {ni} to differentiate sample from diverse categories as in equation 8.

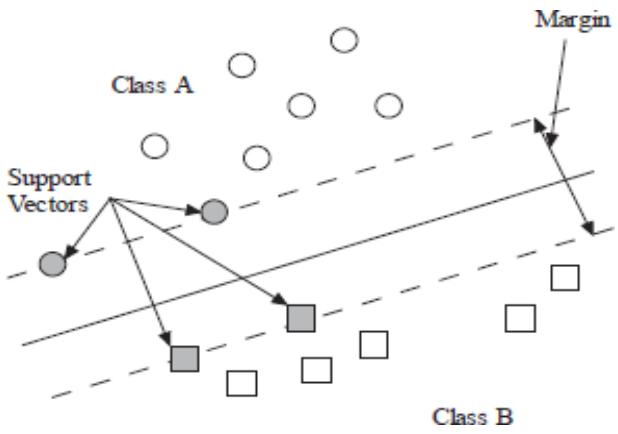

Figure1: Schematic representation of SVM

Where φ(x) is specified as higher dimensional vector. As kernel function K(x, y) provides relationship to „x‟ and „y‟ are „low dimensional vectors, hence no rise in complexity is encountered.

3.6 SVM-NB Algorithm

From above discussion, limitations of Naive Bayes classifier are identified, that is assumption of independent feature vectors extracted from samples. Subsequently, SVM is competent to attain hyperplane to partition samples in two types. SVM is to improve distance amongst boundaries, merging of boundaries is not possible. In general, assumption is not in applications; hence accuracy and recall value are affected. Apply SVM based approach to remove samples that are of irrelevant categories using naive bayes. Feature vectors dependencies are diminished and independence between samples is improved. With the consideration of Naive Bayes classifier speed, SVM based NB algorithms have been anticipated for accuracy and efficiency to classify threats, and thus attaining effectual threat identification.

In the SVM-NB algorithm, the training samples are first processed by the original NB algorithm. For every feature vector extracted from the training set, there will be a corresponding category generated by the NB algorithm. Therefore, we will have a set of results

In SVM based NB algorithm, training samples are processed initially using Naive Bayes algorithm. For each feature vectors extracted for training set, then corresponding categories are produced using Naive Bayes algorithm. Henceforth, set of results attained in equation 10: where ε Rn, {−1, +1}, i = 1, 2, ••• , „i‟ and „j‟ denotes feature vectors dimension. Subsequently, nearest neighbourhood for each feature vector has to be attained. For feature vector, nearest neighbourhood and vectors should belongs to similar category, where vector will be maintained. Else, vectors are dropped from training set (category information). Vectors will be eradicated as it is categorized into irrelevant, dependence amongst nearest neighbourhood and vectors. However, remove samples from training set. This is also stipulating the accuracy of classification. With this equation, NN of feature vector are easily recognized. Based on SVM principle, two nearest feature vectors are extremely belongs to same category. Else, classification outcome has to be modified. SVM-NB algorithm is depicted in algorithm given below: In this procedure, X = {x1, x2, ••• , xm} and Y = {y1, y2, ••• , ym} are used to signify feature vectors and related categories. Suppose V = {v1, v2,•••, vm} specifies classification outcomes produced by original Naive Bayes classifier. The outcome attained using NB classifier is refined using SVM-NB classifier, i.e. vectors are eliminated from training set. „V‟ is used in Naive Bayes algorithm to carry out classification task.

Input: * + * +

V = {v_1, v_2, ⋯ , v_m} Output: V′

∀i=1 to x do

∀ j=1 to x do If i ≠ j then

Dij D (x_i, x_j) End if

End for End for

∀ i=1 to x do Min α ANN i

For all j=1 to x do

If min > D(x_i,x_j) then Min D(x_i,x_j) ANNj

End if End for

If y_i≠y_j then

Eliminate vi for →

End if End for Return →

3.7 Artificial Neural Networks

ANN is termed as Neural Network which is computational model sourced on biological NNs. It comprises of interconnected collection of neurons. ANN is adaptive that modifies structure sourced on information streams through ANN during learning process. There exist two conventional types of neural networks, multi-layer perceptron and perceptron. Here, perceptron algorithm is utilized.

The concept behind perceptron is to acquire linear vector ( ) where ( ) for one class of vectors and ( ) for another vector class. Here, ( ) is vector co- efficient function (weights), and bias is represented as „b‟. The class numbers are denoted by +1 and - 1, hence the decision function is given by () (). Perceptron learning is carried out in iterative. It commences with selected parameters ( ) of decision and revises the iteration. Over the ith algorithm iteration, sample ( ) is selected, in a manner that current decision function is not classified appropriately, i.e Parameters are revised using the equation 12 and 13.

$W_{n+1} = w_n + cx$

$b_{n+1} = b_n + c$

Stage1: Training

Step 1: Initialize „w‟ and „b‟ (values 1-0).

Determine training sample (x,c) for sign ( ). With no sample, training is accomplished

Store final w and terminate. Otherwise go to

Update (w,b): w := w + cx, b := b + c. Move to previous step.

Stage 2: Filtering

Provided message „x", describe class as sign (w^T x + b). The algorithm terminates when decision function is found to be classified accurately with training samples. Above description is provided in algorithm above. In general, three layer NN is utilized as bench mark. In input layer, 50 nodes are equal to system calls. Sum of output nodes was equivalent to 1. Sum of hidden nodes was acquired from W and M; where W is weight of number of interconnects and it has to satisfy the equation 14.

Where,

M □ training samples, I □ input nodes,

O □ output nodes, H □ hidden nodes

Size was managed by assuring proportion of sum of training samples to number of weights equivalent or larger than 3. With above equation, maximal hidden nodes are 10. By altering sum of hidden nodes from 4 to 10, determine ANN with 8 and 6 hidden nodes (frequency based encoding) to attain finest performance. Hidden nodes are utilized to sigmoid transfer function and output node uses transfer function. Training ANN complexity is V2N, where V specifies number of input features (input nodes) and N specifies number of classifiers. This provides complexity for ANN with respect to input features.

IV. NUMERICAL RESULTS AND DISCUSSION

Here, experimental results of classification techniques with NSL-KDD IDS dataset are provided to identify intrusions and comparison with prevailing approaches is to evaluate efficacy of this model. MATLAB in Pentium-4 with 2.86GHz CPU, 1GB RAM was used. NSL-KDD dataset contains 60438 instances, where 22544 instances for testing with 42 attribute and 38 types for classifications to construct IDS.

1) TP: Instances classified correctly as positive class.

2) False Positive: Instances wrongly classified as positive class.

3) Sensitivity (TPR0: It measures positives proportion that is correctly recognized. It depicts probability of correctly recognized positive class from subset (equation 15).

4) Specificity (TNR): It measures negative proportion that is appropriately recognized. It depicts probability of appropriately identifying negative class from subset (equation 16).

5) Precision: It is depicted as ratio of appropriately classified positive from all elements. It is proportion of instances that are appropriately classified positive class by total instances classified to positive class (equation 17).

6) Accuracy: It is ratio of appropriately identified instances to total instances (equation 18 and 19).

7) F-measure: F-measure is used as single measure of test performance for positive class. It is harmonic mean of precision (equation 20)

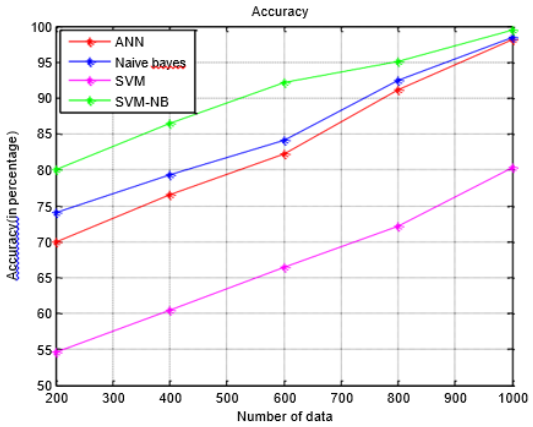

Figure 2: Graphical representation of Number of data Vs Accuracy rate

Figure 2 depicts graphical representation of accuracy computation of anticipated SVM+NB with the existing techniques such as ANN, Naive Bayes and SVM. The accuracy attained due to the proposed approach is better than the existing techniques. SVM-NB attains 80% accuracy while ANN is 70%, Naive Bayes is 74%, SVM is 54.6 for 200 data respectively. With respective to 1000 data, the accuracy attained is 99.5% for the proposed method.

Figure 3 shows the graphical representation of specificity computation of the proposed SVM+NB with the existing techniques such as ANN, Naive Bayes and SVM. Specificity attained due to the proposed approach is better than the existing techniques.

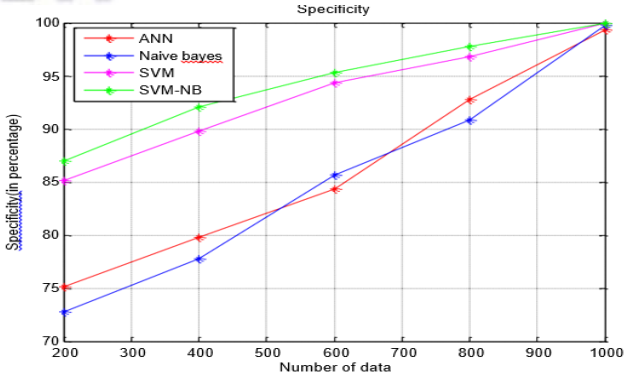

Figure 3: Graphical representation of Number of data Vs Specificity

SVM-NB attains 87% specificity while ANN is 75.12%, Naive Bayes is 72.8%, and SVM is 85.12 for 200 data respectively. With respective to 1000 data, the accuracy attained is 100% for the proposed method.

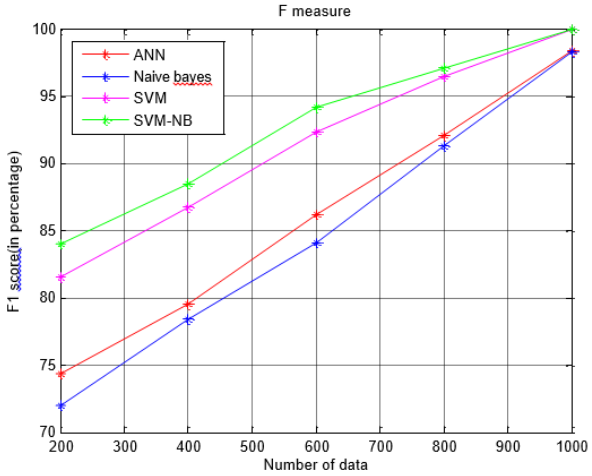

Figure 4: Graphical representation of Number of data Vs F1 score

Figure 4 shows the graphical representation of F1 score computation of the proposed SVM+NB with the existing techniques such as ANN, Naive Bayes and SVM. F1 score attained due to the proposed approach is better than the existing techniques. SVM-NB attains 84% F1 score while ANN is 74.3%, Naive Bayes is 72%, and SVM is 81.5 for 200 data respectively. With respective to 1000 data, the accuracy attained is 100% for the proposed method.

Figure 5 depicts graphical representation of Precision computation of anticipated SVM+NB with the existing techniques such as ANN, Naive Bayes and SVM. Precision attained due to the proposed approach is better than the existing techniques. SVM-NB attains 84% Precision while ANN is 79.44%, Naive Bayes is 81%, and SVM is 83.6 for 200 data respectively. With respective to 1000 data, the accuracy attained is 100% for the proposed method.

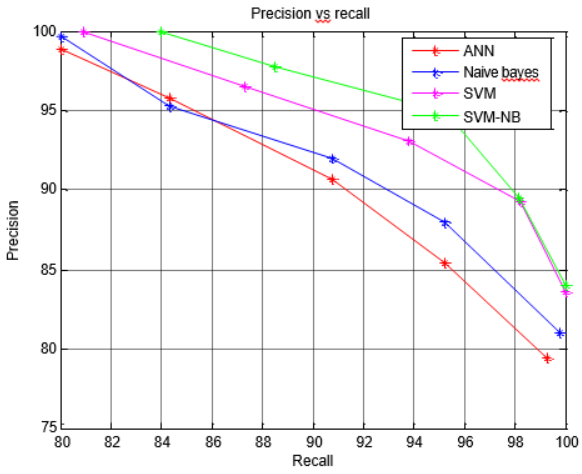

Figure 5: Graphical representation of Number of data Vs Precision

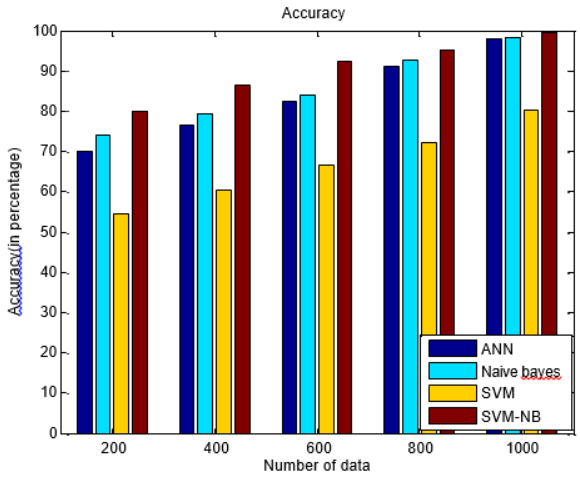

Figure 6: Comparison chart of Number of data Vs Accuracy rate

Figure 6 shows the graphical representation of Accuracy of the proposed SVM+NB with the existing techniques such as ANN, Naive Bayes and SVM. Accuracy attained due to the proposed approach is better than the existing techniques. With respective to 1000 data, the accuracy attained is 100% for the proposed method.

Table 2: Accuracy comparison of SVM-NB with existing techniques

Number of data	200	400	600	800	1000
ANN	70	76.5	82.2	91.1	98.2
Naive Bayes	74	79.3	84.1	92.5	98.5
SVM	54.6	60.5	66.4	72.2	80.4
SVM-NB	80	86.5	92.2	95.1	99.5

Table 2 depicts the Accuracy comparison of proposed SVM+NB with the existing techniques like ANN, Naive bayes, SVM. The accuracy of the existing techniques with 1000 data is 98.2%, 98.5%, 80.4% correspondingly. In addition, the proposed SVM+NB provide 99.5% accuracy respectively. The proposed approach shows better trade off than the prevailing techniques.

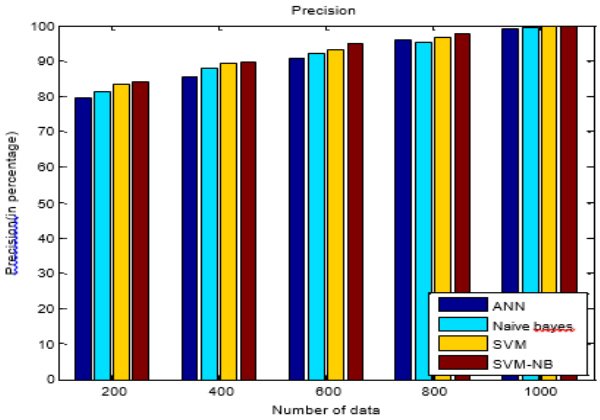

Figure 7: Comparison chart of Number of data Vs Precision

Figure 7 depicts graphical representation of Precision of anticipated SVM+NB with the existing techniques such as ANN, Naive Bayes and SVM. Precision attained due to the proposed approach is better than the existing techniques. With respective to 1000 data, precision attained is 100% for the proposed method.

Table 3: Precision comparison of SVM-NB with existing techniques

Number of data	200	400	600	800	1000
ANN	79.44	85.43	90.7	95.8	98.9
Naive Bayes	81	88	92	95.3	99.7
SVM	83.6	89.3	93.1	96.5	100
SVM-NB	84	89.5	94.9	97.8	100

Table 3 depicts the Precision comparison of proposed SVM+NB with the existing techniques like ANN, Naive bayes, SVM. Precision rate of the existing techniques with 1000 data is 98.9%, 99.7%, 100% correspondingly. In addition, the proposed SVM+NB provide 100% precision rate respectively. The proposed approach shows better trade off than the prevailing techniques.

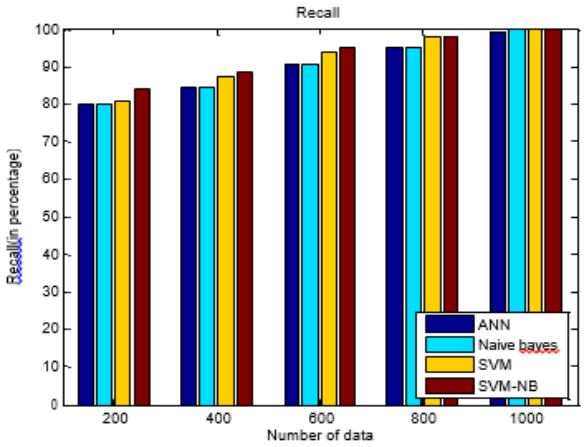

Figure 8: Comparison chart of Number of data Vs Recall

Figure 8 shows the graphical representation of Recall of the proposed SVM+NB with the existing techniques such as ANN, Naive Bayes and SVM. Recall attained due to the proposed approach is better than the existing techniques. With respective to 1000 data, Recall attained is 100% for the proposed method.

Table 4: Recall comparison of SVM-NB with existing techniques

Number of data	200	400	600	800	1000
ANN	80	84.3	90.78	95.21	99.3
Naive Bayes	80	84.3	90.78	95.21	99.8
SVM	80.9	87.30	93.78	98.21	100
SVM-NB	84	88.5	95.28	98.2	100

Table 4 depicts Recall comparison of proposed SVM+NB with the existing techniques like ANN, Naive bayes, SVM. Recall rate of the existing techniques with 1000 data is 99.3%, 99.8%, 100% correspondingly. In addition, the proposed SVM+NB provide 100% recall rate respectively. The proposed approach shows better trade off than the prevailing techniques.

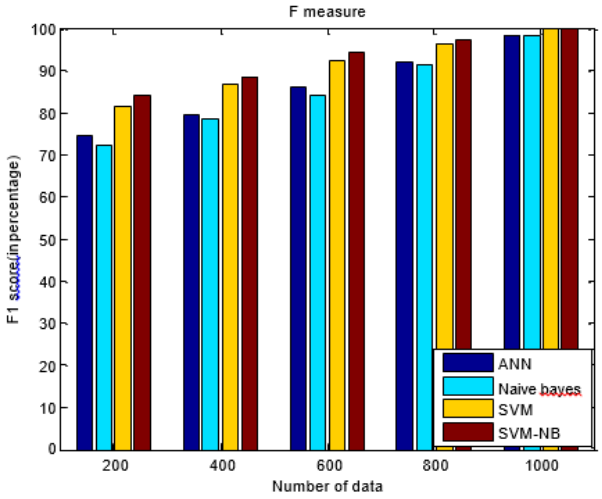

Figure 9: Comparison chart of Number of data Vs F1 score

Figure 9 shows the graphical representation of F1 score of the proposed SVM+NB with the existing techniques such as ANN, Naive Bayes and SVM. F1 score attained due to the proposed approach is better than the existing techniques. With respective to 1000 data, F1 score attained is 100% for the proposed method.

Table 5: F-measure comparison of SVM-NB with existing techniques

Number of data	200	400	600	800	1000
ANN	74.3	79.5	86.2	92.1	98.4
Naive Bayes	72	78.4	84.1	91.3	98.3
SVM	81.5	86.7	92.35	96.44	100
SVM-NB	84	88.5	94.2	97.1	100

Table 5 depicts F measure comparison of proposed SVM+NB with the existing techniques like ANN, Naive bayes, SVM. F measure of the existing techniques with 1000 data is 98.4%, 98.3%, 100% correspondingly. In addition, the proposed SVM+NB provide 100% F measure rate respectively. Anticipated approach demonstrates better trade off than prevailing techniques.

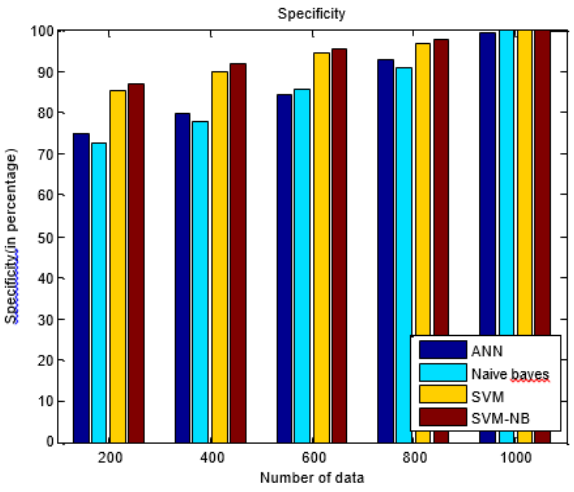

Figure 10: Comparison chart of Number of data Vs specificity

Figure 10 shows the graphical representation of specificity of the proposed SVM+NB with the existing techniques such as ANN, Naive Bayes and SVM. Specificity attained due to the proposed approach is better than the existing techniques. With respective to 1000 data, specificity attained is 100% for the proposed method.

Table 6: Specificity comparison of SVM-NB with existing techniques

Number of data	200	400	600	800	1000
ANN	75.12	79.8	84.33	92.76	99.3
Naive Bayes	72.8	77.8	85.63	90.86	99.8
SVM	85.12	89.8	94.33	96.76	100
SVM-NB	87	92.1	95.33	97.76	100

Table 6 depicts specificity comparison of proposed SVM+NB with the existing techniques like ANN, Naive bayes, SVM. Specificity of the existing techniques with 1000 data is 99.3%, 99.8%, 100% correspondingly. In addition, the proposed SVM+NB provide 100% Specificity rate respectively. The proposed approach shows better trade off than the prevailing techniques.

Table 7: Execution time of SVM-NB with existing techniques

Algorithms	Execution time(in secs)
ANN	20.644
Naive Bayes	5.177
SVM	29.07
SVM-NB	29.18

Table 7 shows the execution time of SVM-NB with the existing techniques. The execution time of ANN, Naive bayes, SVM techniques are 20.644s, 5.177s, 29.07s correspondingly. The execution time of SVM-NB is 29.18s respectively

V. CONCLUSION

Security in cloud storage is more important and a key factor for establishing degree of trust. Cloud user community needs to guarantee security of data stored and performance through Quality of service. Enormous amount of models has been investigated to assure security. Identification and prediction of threats that influence the cloud data has to be tackled effectively. Network threat detection and classification with the base of hybrid learning algorithm is executed in this paper. To enhance performs by identifying the legitimate cloud users (legitimate, non-legitimate and partially legitimate) and to detect the threat, this work cast off hybrid algorithm such as Efficient NB classifier, ISVM and ANN. Initially, for the purpose of monitoring the incoming data, effectual feature selection method has to be identified. After recognizing the sensitive data (i.e. threat aware sensitive information), classification is performed. By doing this, the degree of trust level has been increased amongst the cloud user community. Here, threat detection model is experimented in certain Medical applications. Experimental has been carried out in MATLAB simulation environment and the performance metrics such as precision with 100%, accuracy with 99.5%, specificity with 100%, Recall with 100% and F measure with 100% is computed. The numerical results obtained shows better trade off when compared to existing methods.

REFERENCES

[1]. Chraibi, “Classification of Security Issues and Solutions in Cloud Environments”, ACM Proc of Int Conf on Information Integration and Web-based Applications & Services, 2013, pp 560.

[2]. Rajan, M.S., Arunkumar, J.R., Anusuya, R., Mesfin, A. (2021). Earliest-Arrival Route: A Global Optimized Communication for Networked Control Systems. vol 384. Springer, Cham. https://doi.org/10.1007/978-3-030-80621-7_10.

[3]. Chow, “Authentication in the Clouds: A Framework and its Application to Mobile Users”, ACM CCSW‟10, 2010, pp. 1-6.

[4]. S.Rajkumar, M.Ramkumar Prabhu and A.Sivabalan, 2012. Relaxation Based Electrical Simulation for VLSI Circuits in Research Journal of Applied Sciences, Engineering and Technology*, 4(12):1629-1632, 2012. ISSN:20407459,eISSN:2040-7467.

[5]. Emeakaroha, and Rose, C.A.F, “Towards Autonomic Detection of SLA Violations in Cloud Infrastructures”, Future Generation Computer Systems, Volume 28, Issue 7, July 2012..

[6]. P. Nirmala, T. Manimegalai, J. R. Arunkumar, S. Vimala, G. Vinoth Rajkumar, Raja Raju, "A Mechanism for Detecting the Intruder in the Network through a Stacking Dilated CNN Model", Wireless Communications and Mobile Computing, vol. 2022, Article ID 1955009, 13 pages, 2022. https://doi.org/10.1155/2022/1955009.

[7]. M.Ramkumar Prabhu , V.Reji and A.Sivabalan, 2012. Improved Radiation and Bandwidth of Triangular and Star Patch Antenna in Research Journal of Applied Sciences, Engineering and Technology*,4(12) :1740-1748, 2012. ISSN:20407459,eISSN:2040-7467.

[8]. Arunkumar, J. R., Anusuya, R., Rajan, M. S., & Prabhu, M. R. (2020). Underwater wireless information transfer with compressive sensing for energy efficiency. Wireless Personal Communications, 113(2), 715–725.

[9]. Raneel Kumar, ,“Detecting Denial of Service Attacks in Cloud”, IEEE Int Conf on Dependable, Autonomic and Secure Computing, 2016

[10]. Dr.M.Ramkumar Prabhu,Dr.A.Rajalingam ,K.Venkateswara Rao & U.T.Sasikala, “ Design of Rectangular Microstrip Patch Antenna with High Gain for Ku Band” in International Journal of Applied Engineering Research, ISSN 0973-4562 Vol. 10 No.75 (2015), Page 212-215J.

[11]. J.R.Arunkumar, Dr.E.Muthukumar,” A Novel Method to Improve AODV Protocol for WSN” in Journal of Engineering Sciences” ISSN NO: 0377-9254Volume 3, Issue 1, Jul 2012.

[12]. J. R. Arunkumar, Tagele berihun Mengist, 2020” Developing Ethiopian Yirgacheffe Coffee Grading Model using a Deep Learning Classifier” International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-4, February 2020. DOI: 10.35940/ijitee.D1823.029420.

[13]. R.Prabha, M.Ramkumar Prabhu, SU.Suganthi, S.Sridevi, G.A.Senthil, D.Vijendra Babu, “Design of Hybrid Deep Learning Approach for Covid-19 Infected Lung Image Segmentation” in Journal of Physics: Conference Series 2040 (2021).

[14]. Zhijiang Chen, “A Cloud Computing Based Network Monitoring and Threat Detection System for Critical Infrastructures”, Elsevier, 2016

[15]. L. Saravanan, W. Nancy, K. P. Chandran, D. Vijayanandh, J. R. Arunkumar and R. T. Prabhu, "A Novel Approach for a

Smart Early Flood Detection and Awareness System using IoT," 2022 8th International Conference on Smart Structures and Systems (ICSSS), 2022, pp. 1-4, doi: 10.1109/ICSSS54381.2022.9782286.

[16]. L Shylaja, Shaik Fairooz, J. Venkatesh, D. Sunitha, R. Prakash Rao, M.Ramkumar Prabhu, "IoT based crop monitoring scheme using smart device with machine learning methodology", Journal of Physics: Conference Series, 2027 , (2021).

[17]. M. S. Rajan, J. R. Arunkumar, A. Ramasamy and B. Sisay, "A comprehensive study of the Design and Security of the IoT layer Attacks," 2021 6th International Conference on Communication and Electronics Systems (ICCES), 2021, pp. 538-543, doi: 10.1109/ICCES51350.2021.9489235.

[18]. Anusuya Ramasamy, J. R. Arunkumar, and M. Sundar Rajan, "A Secure and Energy Efficient Sensor Nodes in Wireless Sensor Networks using Improved Ant Lion Optimization." International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-9 Issue-1, May 2020. DOI:10.35940/ijrte.A2858.059120.

[19]. K. K. Baseer, M. Jahir Pasha, D. William Albert and V. Sujatha, "Navigation And Obstacle Detection For Visually Impaired People," 2021 Fourth International Conference on Microelectronics, Signals & Systems (ICMSS), 2021, pp. 1-3, doi: 10.1109/ICMSS53060.2021.9673618.

[20]. K.K. Baseer, Neerugatti, V. ., M. Jahir Pasha, & V. D. Satish Kumar. (2020). Internet of Things: A Product Development Cycle for the Entrepreneurs. Helix - The Scientific Explorer | Peer Reviewed Bimonthly International Journal, 10(02), 155-160.

[21]. Prathima Chilukuri, R.Anusuya, M.Ramkumar Prabhu, Comprehensive Design Analysis Of Digital Marketing In Agriculture Sector, International Journal of Early Childhood Special Education (INT-JECSE) DOI:10.9756/INTJECSE/V14I5.81 ISSN: 1308-5581 Vol 14, Issue 05 2022.

[22]. Bakshi, A., & Yogesh, B. (2010, February). "Securing Cloud from DDOSAttacks using Intrusion Detection System in Virtual Machine", IEEE Int.Conf on Communication Software and Networks, 2010, pp. 260-264.

[23]. Houmansadr, "A Cloud Based Intrusion Detection and Response System for Mobile Phones", IEEE/IFIP Int. Conf on Dependable Systems and Networks Workshops, 2011.

[24]. Prathima, C., Muppalaneni, N.B., Kharade, K.G. (2022). Deduplication of IoT Data in Cloud Storage. In: Satyanarayana, C., Gao, XZ., Ting, CY., Muppalaneni, N.B. (eds) Machine Learning and Internet of Things for Societal Issues. Advanced Technologies and Societal Change. Springer, Singapore. https://doi.org/10.1007/978-981-16-5090-1_13.

[25]. Muppalaneni, Naresh Babu, Ch Prathima, and Akula Chandra Sekhar. "Ensemble Deep Learning for Brazil Currency Coin Prediction." IOP Conference Series: Materials Science and Engineering. Vol. 1074. No. 1. IOP Publishing, 2021.

[26]. Ibrahim, Laheeb Mohammad. "Anomaly Network Intrusion Detection System based on Distributed Time-Delay Neural Network (DTDNN)", J. of Eng Science and Technology, Volume 5, No. 4, 2010.

[27]. Prathima, C., Muppalaneni, N.B. (2021). Deep Learning Approach for Prediction of Handwritten Telugu Vowels. In: Reddy, A., Marla, D., Favorskaya, M.N., Satapathy, S.C. (eds) Intelligent Manufacturing and Energy Sustainability. Smart Innovation, Systems and Technologies, vol 213. Springer, Singapore. https://doi.org/10.1007/978-981-33-4443-3_35

[28]. Li, Lei, and Fang-Cheng Shen. "A Novel Rule-Based Intrusion Detection System using Data Mining", IEEE Int. Conf onComputer Science and Information Technology, 2010.

[29]. Anurag Jain, "A Taxonomy of Cloud Computing",Int J of Scientific and Research Publications, 2014.