# Unsw-Nb15 Dataset and Machine Learning Based Intrusion Detection Systems

**AvinashR.Sonule, Mukesh Kalla, Amit Jain, D.S. Chouhan**

***Abstract:*** ***The network attacks become the most important security problems in the today's world. There is a high increase in use of computers, mobiles, sensors,IoTs in networks, Big Data, Web Application/Server,Clouds and other computing resources. With the high increase in network traffic, hackers and malicious users are planning new ways of network intrusions. Many techniques have been developed to detect these intrusions which are based on data mining and machine learning methods. Machine learning algorithms intend to detect anomalies using supervised and unsupervised approaches.Both the detection techniques have been implemented using IDS datasets like DARPA98, KDDCUP99, NSL-KDD, ISCX, ISOT.UNSW-NB15 is the latest dataset. This data set contains nine different modern attack types and wide varieties of real normal activities. In this paper, a detailed survey of various machine learning based techniques applied on UNSW-NB15 data set have been carried out and suggested thatUNSW-NB15 is more complex than other datasets and is assumed as a new benchmark data set for evaluating NIDSs.***

***Keywords: Intrusion Detection System, UNSW-NB15dataset,Network Intrusion Detection System (NIDS).***

## I. INTRODUCTION

Intrusion Detection Systems (IDS)[1][2] is a device or software application that monitors network and the system for suspicious activities and warns the system or network administrator.There are Host based IDS and Network based IDS. A Host based Intrusion Detection System keeps track of individual host machine and givesnoticeto the user if suspicious activities likedeleting or modifying a system file, undesired configuration changes, unnecessary sequence of system calls are detected. Generally, a Networkbased Intrusion Detection System(NIDS)[3] is kept atnetwork points like agateway or routers to detect the intrusions in the network traffic.

A NIDS monitors and detects network-attack patterns over networking environmentsand protect computing resources against malicious activities. At high level, IDS can be categorized by the detection mechanism used by it.These IDSes are :i) misuse detection,ii) anomaly detection and iii) hybrid detection. Misuse detection techniques have been used to detect known attacks while theAnomaly detection techniques have been used to detect unknown attacks.

**Revised Manuscript Received on February 05, 2020.**

**AvinashR.Sonule,** Department of Computer Science & Engineering, Sir Padampat Singhania University(SPSU), Udaipur-313601, Rajasthan, India

**Mukesh Kalla,** Department of Computer Science &Engineering, Sir Padampat Singhania University(SPSU), Udaipur-313601, Rajasthan, India

**Amit Jain,** Department of Computer Science &Engineering, Sir Padampat Singhania University(SPSU), Udaipur-313601, Rajasthan, India

**D.S. Chouhan,** Department of Computer Science &Engineering, Sir Padampat Singhania University(SPSU), Udaipur-313601, Rajasthan, India

Machine Learning(ML) can be used for all the three types of detection techniques. Machine learning is subclass of Artificial Intelligence(AI) that used in computers having the skill to learn without being absolutely computed. A machine learning models have two parts: training and testing. The training data samples are the input in which by making use of a learning algorithm the features are learnedin the training. In the testing, an execution engine is used by the learning algorithm makes prediction for the unknown test data. The classified data is given as the output by the learning model to detect novel attacks.

Machine learning algorithms are applied on different network attack datasets with or without feature selection approaches. Supervised learning algorithms build a mathematical model of a set of data which contains both the inputs and the desired outputs. The data is known as training data, and consists of a set of training examples. Each training example has one or more inputs and a desired output. It is also known as a supervisory signal. Unsupervised learning algorithms take a set of data that contains only inputs, and find pattern in the data, such as grouping or clustering of data points. The algorithms therefore learn from test data that has not been labeled, classified or categorized. Instead of responding to feedback, unsupervised learning algorithms identify commonalities in the data and react based on the presence or absence of such commonalities in each new piece of data.

The rest of the paper is organized as follows: section 2gives related work for survey. Section 3gives in detail of the existing datasets generation and its shortcomings.In section 4, the synthetic environment configuration and UNSW-NB15 dataset generation details are given.Section 5 gives description UNSW-NB15 Dataset Section 6 presents different machine learning based IDS applied on UNSW. Section 7 displays the summary of experimental results by all machines learning applied on UNSW-NB15 dataset. Finally, section VIII gives future direction and section IX concludes the work.

## II. RELATED WORK

Agrawal et al. [4] have carried out a survey on anomaly detection with data mining techniques to detect intrusions. They have classified the anomaly detection techniques with three features: classification based techniques , clustering based techniques and hybrid techniques. Buczak et al. [5] have done survey which describes the application of data mining and machine learning techniques to detect known and unknown attack. They showed clear distinction between data mining (DM) and machine learning (ML).

Mishra.et al[6] done surveys on machine learning based IDS usingmixture of all IDS datasets mostly used DARPA[7], KDD'99[8][9]and NSL-KDD[10] and other datasets.

Many IDS researchers have applied their Intrusion Detection System on one or more datasets.

In our survey, different intrusion detection techniques based on machine learning methods using new benchmark dataset UNSW-NB15[9][11] have been thoroughly analyzed.A detailed analysisof various machine learning methods with or without feature selection have been carried out in detecting intrusive activities. The presentworks shows thatno one specific intrusion detection technique can detect all types of attacks. Therefore to detect a particular set of attacks, the use of specific intrusion detection technique is suggested. A summary of different intrusion detection approaches using UNSW-NB15 dataset is discussed.

## III. IDS DATASETS AND SHORTCOMINGS

A standard of the NIDS dataset have two important characteristics: a comprehensive reflection of contemporary threats and inclusive normal range of traffic. The quality of the dataset have an effect on the reliable outcome of any NIDS. The disadvantages of existing data sets for NIDS are discussed in this section.

DARAP98 Dataset: At MIT University,IST group of Lincoln laboratories carried out a simulation with normal and abnormal traffic in a military network environment. The simulation carried out for nine weeks of raw tcpdump files. Thefour GBs training data and composed of compressed binary tcpdump files from seven weeks of network traffic was used. Approximately five million connection records were processed from it. Two weeks of test data which have two million connection records was provided by simulation. DARAP98 network data features comprehensivenessupgrading, utilizing the same U.S. Air Force LAN environment, the simulation completed which have 41 features for each connection along with the class label using Bro-IDS tool. Several issues are found with DARPA98, including the unrealistic network architecture, overt synthesis of data, questionable evaluation methodology and high tolerance for false alarms.

KDDCUP99 Dataset: KDDCUP99 is the upgraded version of DARAP98. In the KDDCUP99 data set, all extracted features were classified into three groups of intrinsic features, content features and traffic features. Also attack records in this data set are categorized into normal or specific type of attack DoS, R2L, U2R, and Probe. The training set of KDDCUP99 contained 22 attack types and test data had 17 attack types.

Many IDS researchers have made use of these datasets due to their public availability. However, many researchers have reported some important disadvantages of the datasets which can affect the transparency of the IDS evaluation. The success of NIDS is assessed based on their performance to identify attacks which requires a comprehensive data set that contains normal and abnormal behaviors. It is discovered through several studies, evaluating a NIDS using this data set does not reflect realistic output performance due to several reasons. First reason is the KDDCUP 99 data set contains a tremendous number of redundant records in the training set. These redundant records affect the results of detection biases toward the frequent records. Second, there are multiple missing records which is important factor in changing the nature of the data.Third, every attack data packets have a time to live value (TTL) of 126 or 253, while the packets of the traffic mostly have a TTL of 127 or 254. But TTL values 126 and 253 do not occur in the training records of the attack. Fourth, the probability distribution of the training set is different from the probability distribution of the testing set,as there is adding of new attack records in the testing set. This results in skew or bias classification methods for some records rather than the balancing between the types of attack and normal observations. Fifth, the data set is not a comprehensive representation of latest reported low foot print attack projections.Other reasons against the usage of KDD CUP'99 dataset are :Non- consideration of emergence of complex network scenarios, Non-inclusion of rapid surge in attack vectors, missing of network traffic diversity in the created test bed and the presence of semantic gap between experimental results and operational environment. Unfortunately, KDDCUP'99 have several weaknesses which deter its use in the modern context, including: its age, pattern redundancy, non-stationarity between training and test datasets, highly skewed targets, and irrelevant features.

NSLKDD Dataset:An improved version of the KDD dataset is referred to as NSL-KDD[8]. Its first aim was, to remove the duplicaterecords in the training and testing sets of the KDDCUP99 data set to eliminate classifiers biased to more repeated records. Secondly, to choose a variety of the records from different parts of the original KDD data set to achieve reliable results from classifier systems. Third, to eliminate the unbalancing problem among the number of records in the training and testing phase to reduce the False Alarm Rates (FARs). Each traffic sample has 41 features. Attacks in the dataset are dividedinto four categories: DoS, R2L, U2R, and Probe attacks. The training dataset includes 24 attack types, while the testing dataset contains 38 attack types. The main disadvantage of NSLKDD is that it does not show the modern low foot print attack scenarios. A significant trend is the poor performance of classifiers on minority classes of KDDCUP-99, an obstacle which NSL-KDD is unable to eliminate.

ISCX-2012 Dataset[12]: DARPA98, KDDCUP99, NSL-KDDareverypopulardatasetsusedintheclassification in the IDS domain; however, they have been thoroughly denounced for being unable to provide a realistic scenario. TheInformationSecurity Centre of Excellence of the University of New Brunswick developed ISCX Dataset. This dataset is the result of capturing seven days of network traffic in a controlled testbed made of a subnetwork placed behind a firewall. Normal traffic was generated with the help of agents that simulated normal requests of human users following some probability distributions extrapolated from real traffic.

Attack were generated with the help of humanoperators.
Theresultisafully labeled dataset with realistic traffic scenarios. Indeed, the dataset consists of standards pcap (packet capture) files one for each day containingtherelativenetworktraffic.All different days contain different attack scenarios, ranging from HTTP Denial of Service, DDoS, Brute Force SSH and attempts of infiltrating the subnetwork from the inside. The drawback of this dataset is that there very few types of attacks viz flood and Privilege escalation(priesc)/(Probe).
There are many IDS datasets like HTTP CSIC dataset, CISDA 2009, CAIDA2011,ISOT and other datasets which are not so popular.

## IV. UNSW-NB15 DATASET

The existing datasets do not represent the comprehensive representation of the modern orientation of network traffic and attack scenarios.These reasons have instigated a serious challenge for the cyber security research group at the Australian Centre for Cyber Security (ACCS) and other researchers of this domain around the globe. The raw network packets of the UNSW-NB15 dataset[6] was created by the IXIA PerfectStorm tool in the Cyber Range Lab of ACCS for generating a hybrid of real modern normal activities and synthetic contemporary attack behaviors.
The IXIA tool simulates nine families of attacks. The IXIA tool has all information about latest attacks that are updated continuously from a CVE site. This site acts as a dictionary of publicly known information security vulnerabilities and exposures. The tcpdump tool is used to capture network traffic in the form of packets. To capture 100 GBs , the simulation period was 16 hours on Jan 22, 2015 and 15 hours on Feb 17, 2015. Each pcap file is divided into 1000 MB using the tcpdump tool. To create reliable features from the pcap files, Argus6 andBro-IDS7 tools are used. Twelve algorithms were developed using a C# language to analyze in-depth the flows of the connection packets.

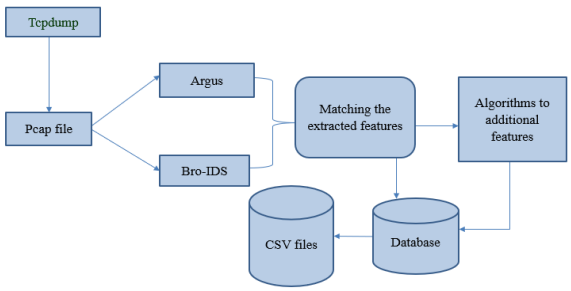


**Fig.1. Framework Architecture togenerate UNSW-NB15 Dataset**

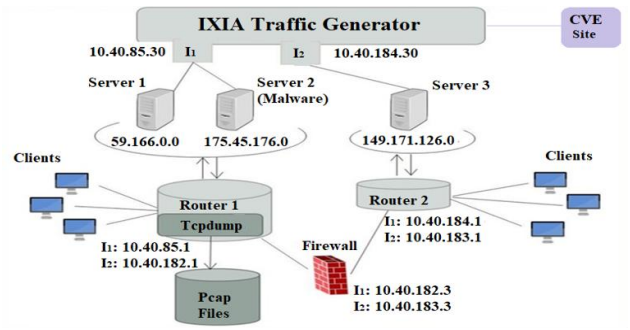


**Fig.2: UNSW-NB15 Testbed**

Fig 2 shows the configuration details oftestbed and all processes involved in generating UNSW-NB15 dataset. The IXIA traffic generator is configured which had the three virtual servers. These servers 1 and 3 are configured for normal spread of the traffic. The server 2 generated the malicious activities in the network traffic. To establish the intercommunication between the servers and to acquire public and private network traffic, they have configured two virtual interfaces with IP addresses, 10.40.184.30 and10.40.85.30. The servers are connected to hosts through two routers. The router 1 is configured with 10.40.85.1 and 10.40.182.1 IP addresses and router 2 is configured with 10.40.184.1 and 10.40.183.1 IP addresses. All routers are connected to the firewall device and configured to pass all the normal and abnormal traffic. The tcpdump tool is installed on the router 1 for capturing the Pcap files of the simulation uptime. The central idea of this whole testbed was to capture the normal or abnormal trafficoriginating from the IXIA tool and spread among network nodes. The IXIA tool is used as an attack traffic generator along with as normal traffic.The attack behavior is nourished from the CVE site for a real representation of a modern threat environment.
This dataset is divided into nine types of attacks. The Argus, Bro-IDS tools are used with twelve algorithms to generate total 49 features with the class label. The total number of records is two million and 540,044 which are stored in the four CSV files, namely, UNSW-NB15_1.csv, UNSW-NB15_2.csv, UNSW-NB15_3.csv and UNSW-NB15_4.csv. The number of records in the training set is 175,341 records and the testing set is 82,332 records from the different types, attack and normal.
In contrast to the datasets such as DARPA98, KDDCUP99, NSL-KDD and ISCX,realized a limited number of attacks and information of packets which are outdated. It is expected thatin future, the UNSW-NB15 data set can be useful to the NIDS research community and considered as a modern NIDS benchmark dataset.

## V. DESCRIPTIONOF THE UNSW-NB15 DATASET

There are nine attackstypes discovered in UNSW-NB15 Dataset.
(1) Fuzzers: an attack in which the attacker tries to discover security loopholes in the Operating System, program or network and make these resources suspended for some time period and can even crash them.
(2) Analysis: a type intrusions that penetrate the web applications through port scanning, malicious web scripting and dispatching spam emails etc.
(3) Backdoor: a technique in which attacker can bypass the usual authentication and can get unauthorized remote access to a system.
(4) DoS: an intrusion in which attacker tries to disrupt the computing resources, by making them extremely busy in order to prevent the authorized access to the resources.
(5) Exploit: the intrusions which utilize the software vulnerabilities, error or glitch within the operating systems(OS) or software.

(6) Generic: This attack act against a cryptographical system and it tries to break the key of the security system.
(7) Reconnaissance: It can be defined as a probe; an attack that gathers information about the target computer network in order to bypass its security control.
(8) Shellcode: a malware attack in which the attacker penetrates a slight piece of code starting from a shell to control the compromised machine.
(9) Worm: malware that replicate themselves and spread to other computers by using the network to spread the attack, depending on the security failures on the target computer which it want to access.
The UNSW-NB15 data set features are classified into six groups as follows:
1) Flow features: Thesefeatures have the identifier attributes between hosts, such as client-to-serve or server-to-client.
2) Basic features: These featuresinclude the attributes that represent protocols connections.
3) Content features: These featurescontain the attributes of TCP/IP; also they contain some attributes of http services.
4) Time features: This group contains the attributes of time, for example, arrival time between packets, start/end packet time and round trip time of TCP protocol.
5) Additional generated features: This group can be further divided into two groups: (1) General purpose features which each feature has its own purpose, in order to protect the service of protocols. (2) Connection features are built from the flow of 100 record connections based on the sequential order of the last time feature.
6) LabelledFeatures:Thiscategoryrepresentsthelabelofeach record.

## VI. MACHINE LEARNING BASED IDS

The some of the benefits of IDSbased onMachine learning are as follows:
• IDSbased on Machine learning which usesupervised techniques can easilyidentify theattackvariantsastheygainthebehavior of the traffic flow.
• IDSbased on Machine learning which use unsupervised learning algorithms can detect new attacks.
•In the IDSbased on Machine learning, the CPU load is low to moderate.
• IDSbased on Machine learning can find the complex properties of the attack behavior.It also improve the detection accuracy and speed.
• Different types of attacks keep on evolving. IDSbased on Machine learning which use clustering and outlier detection do not require updatesin attack's database.
In this paper, we have mainly discussed machine learning based IDS with UNSW NB-15 dataset for misuse anomaly and hybrid detection. A detailed study of different machine learning approaches is useful to find solutions for detecting advanced cyber intrusion. The machine learning based IDSesare using : (i)Single classifiers using all features(SCAF) of data set (ii)Multiple classifiers using all features (MCAF)of data set (iii) Single classifiers using limited features(SCLF) of data set and (iv) Multiple classifiers using limited features(MCLF) of dataset.
The major contributions of our paper are as follows:
• Discussion of variousIDS datasets, their shortcoming, benefit of using UNSW-NB15 dataset.
• The attacks classification based on their characteristics of UNSW NB-15 datasets is presented.
• The discussion of different existing literature for intrusion detection is provided, highlighting the key characteristics, feature selection employed, the detection mechanism, attacks detection capability.
• The critical performance analysis of different intrusion detection techniques is given with respect to their attackdetectioncapability.Thelimitationsandcomparison with other approaches are also discussed.
• Future directions to use the machine learning for intrusion detection applicationsare provided.
In this Section, we have discussed different machine learning techniques applied on UNSW dataset by various researchers to detect intrusions. Their proposed techniques have different characteristics and give different results for detecting intrusions.
Several performance measures, i.e. accuracy, precision, recall and false alarm rate as calculated as follows.
Accuracy =(TP+TN)/ (TP+FP+TN+FN )…….. (1)
Precision =TP /(TP+FP)……..……(2)
Recall =TP /(TP+FN )……..…… (3)
Sensitivity or True Positive Rate (TPR)=TP /(TP + FN)… (4)
Specificity or True Negative Rate (TNR)=TN /(FP + TN )… (5)
FPR =FP /(FP+TN)……
(6)
FAR =(FPR + FNR) /2 …..……(7)
FNR = FN / (FN+TP)……..……..(8)
F1 Score = 2(Precision xRecall)/(Precision +Recall)…….(9)
where
True positive(TP) means the correct intrusion detection
False Positive(FP) means to assume the normal traffic as the cyberattack.
True negative (TN) refers to normal traffic correctlylabeled as normal.
False Negative (FN) means to failintrusion disclosure.
FPR is the false positive rate.
FNR is the false negative rate.
False Alarm Rate (FAR) means the average ratio of the misclassified to classified records either normal or abnormal.
The $F_1$score refers the harmonic average of the precision and recall.
Machine Learning methodshave training and testing steps. In training step, the mathematical calculations are carried on the training dataset to learn the behavior of traffic over a period. In the testing step, a test instance is classified as normal or attack based on the learned behavior. In this section, we have discussed the working of different ML based techniques on with their characteristics using UNSW-NB15 dataset.

Moustafa et al. [13] suggested a hybrid feature selection approach which reduce the irrelevant features set. These reduced features used with machine learning algorithms to detect intrusion. The proposed NIDS architecture is then used for anomaly intrusion detection and misuse intrusion detection. NIDS takes the input from theUNSW-NB15 dataset and then computes the center points for attribute values. A center point means the most frequent value of the attribute. All these center points for the attributes are given to the association rule mining algorithm (Apriori) as an inputthereby reducing its processing time. This association rule mining finds out the highly ranked attributes /features using the correlation of the two or more attributes. The filtered dataset which consists of the selected features feed as an input to the detection engine. They applied three ML algorithms: Expectation Maximization (EM) clustering, Naive Bayes (NB) and Logistic Regression (LR) on UNSW-NB15.EM gives an accuracy of 77.2% and 13.1% FAR. LR gives accuracy of 83.0% and 14.2% FAR and NB gives 79.5% accuracy and 23.5% FAR.

Gharaee et al. [14] presented the feature selection based intrusion detection system (GF-SVM) which detect intrusions in the network. The SVM and a Genetic algorithm (GA) are combined to give an optimal set of features. They modified the fitness function of the GA slightly. They have used TPR, FPR and NumFas parameters for fitness function. These parametersare multiplied by certain weight as per user requirements. Every chromosome is determined for each iteration of GA and chromosomes with the highest classification accuracy by SVM are selected. The filtered dataset is obtained by using optimal features from UNSW-NB15 dataset . Least Squared Support Vector Machine (LSSVM) is used to learn the training dataset with selected features and also to test dataset. Authors have used seven different features for normal attacks.They used 6-14 features for different attacks types. Their system provides an accuracy of 97.45% , 98.47% TPR and 0.04% FPR to detect the normal traffic. It provides anaccuracy of 79.19%-99.45% , TPR 67.31%-100% and FPR 0.01%-0.09% to detect the various attacks types.

Chowdhury et al.[15] combined simulated annealing (SA) and Support Vector Machine (SVM) for network intrusion detection. They have applied this combination to increase the detection accuracy and decrease the false alarms. In this proposed misuse detection algorithms they can classify the normal and abnormal classes. In this algorithm,SA algorithm is used to select first n features from a set of K features using UNSW-NB15 dataset. Then dataset with n selected features is applied to train the SVM. The trained model is applied to detect the future test instances. From the dataset, at random 150, 000 samples are selected which have 75,000 normal and 75,000 anomaly samples. They used 70% of the total dataset for trainingand30%for testing. They achieved88.03% accuracy with normalSVM. The proposed scheme gives an accuracy of 98.76% with a randomly selected three features with SA approach. They achieved FPR 0.09% and FNR 1.35% which is reasonably low.

Bhamare et al. [16] proposed the machine learning approach to detectattacks in the cyber network. They have executed different machine learning algorithms using UNSW-NB15 dataset. This has comprehensive representation of modern attack which provide real attack scenarios. Misuse detection algorithms such as NB, DT, LR and SVM usethree different kernels, which are Polynomial, Linear, RBF are applied on Dataset. NB gives an accuracy of 73.8%, DT givesan accuracy of 88.67%, SVM with polynomial kernel gives 68.06% accuracy, SVM with linear kernel gives 69.54% accuracy, SVM with RBF kernel gives 70.15% accuracy, and LR gives 89.26% accuracy. DT gives 6.9% FPR, SVM with RBF function p gives 4.1% FPR, SVM with poly function gives 53.3% FPR, SVM with linear function gives 50.7% FPR, NB gives 7.3% FPR, LR gives 4.3% FPR. Among all Logistic regression is giving better results with low FPR. They used simple methods of Machine Learning that are not giving good result.

Baig et al.[17] proposed a cascade of ensemble-based artificial neural network for multi-class intrusion detection (CANID) in computer network traffic. The boostingbased ANN learning used to learnweights of a given neural network using AdaBoost. The cascade structure and an associated example filtering mechanism used to learn an effective multi-class classifier by combining several binary classifiers connected as a decision tree or cascade. The cascade structure is a generalization of one-vs-remaining encoding strategy of building a multi-class classifier by combining several binary classifiers in the form of a tree structure. The Boostron algorithm has been extended further to learn parameters of an ANN with a single hidden layer and a single output neuron. Using UNSW-NB15 They achieved Accuracy - 86.40% ,Precision- 0.8674,Recall-0.9338, F1 Score- 0.8994.

Belouch et al. [18]proposed a two-stage classifier based on Reduced Error Pruning Tree (REPTree) algorithm and protocols subset for network intrusion detection system. The combination of information gain and consistency through an evolutionary search method is used for the proposed feature selection.A ranker algorithm ranks the features in the data set to select the appropriate number of features based on user's requirements. To evaluate the performance, they used the UNSW-NB15 data set. In first phase this approach divides the incoming network traffics into three type of protocols TCP, UDP or Other, then classifies into normal or anomaly. In next stage a multiclass algorithm classify the anomaly detected in the first phase to identify the attacks class in order to choose the appropriate intervention. The number of features is reduced from over 40 to less than 20 features, according to the protocol, using feature selection techniques. They achieved the detection accuracy of 88,95% on the complete UNSW-NB15 data set.

Al-Zewairi et al.[19] proposed a deep learning binomial classifier for Network Intrusion Detection System evaluated using the UNSW-NB15 dataset. The proposed DL model is built using the H2O platform. It is a multilayer feedforward artificial neural network(MFFANN) using backpropagation and stochastic gradient descent method.Three different experiments were executed in order to determine the optimal activation function, then to select the most important features and finally to test the proposed model on unseen data.The most important features are identified using the Gedeon method.

The evaluation results found that the proposed classifier outperforms other models with 98.99% accuracy and 0.56% false alarm rate on unknown data.

Anwer et al.[20] proposed framework for efficient network anomaly detection using different machine learning classifiers. The feature selection framework applies five different strategies for features selection. The aim of this framework is to select the minimum number of features that gives the highest accuracy. UNSW-NB15 dataset is used in the experimental results to evaluate the proposed framework. J48 and Naïve Bayes algorithms are used as classifiers. The experimental results obtained show that, the best strategy is by using 18 features from the GR ranking method and applying J48 as a classifier getting an accuracy of 88% and a speedup factor of 2.

Mithun et al.[21] proposed an Intrusion Detection System(IDS), which detects the family of attack in a dataset. In this proposed work, the data is extracted from UNSW_NB15 dataset. The K- means algorithm is used to identify the data cluster centers. A new and one dimensional distance based feature is used to represent each data sample. Using reduced data, an ensemble classifier is used to classify the data. An Algorithm classify five families of attack. It is found that the k means clustering algorithm efficiently identifies the cluster centers and the nearest neighbors. Using the feature selection algorithm an one dimensional data set with distance as its only feature is obtained. The various classes of attacks are identified by training and testing the ensemble classifier. Their system classifies the attack with 90% accuracy.

Idhammad et al.[22] proposed a detection method of the DoS attack based on ANN, named ADDM. A multi-layer perceptron was optimized to improves the detection accuracy and the detection time. In the proposed work, a Feed-forward Neural Network (FNN)is optimized to detect DoS attack with minimum resources usage. The proposed method consists of three major steps: First, Collection of the incoming network traffic.Second, selection of relevant features for DoS detection using an unsupervised Correlation-based Feature Selection (CFS) method. Third, classification of the incoming network traffic into DoS traffic or normal traffic. Using UNSW-NB15, various experiments were conducted to evaluate the performance of the proposed method. The obtained testing results are compared with the findings in the related works . The ADDM has the highest testing accuracy rates 97.1% in the shortest period of time 0.46s on UNSW-NB15. The DoS detection accuracy rates of DDMA, HSV-ANN, NSL-ANN and ANN are respectively 98%, 92%, 81.2% and 81.34%. The applied optimizations techniques on the ADDM have improved significantly the DoS detection accuracy rate. The feature selection phase has enabled the ADDM to reduce the DoS detection time. The shortest DoS detection time intervals are 0.46s and 0.35s which correspond to the ADDM.

Hajisalem et al.[23] proposed a new hybrid classification method based on Artificial Bee Colony (ABC) and Artificial Fish Swarm (AFS) algorithms. The Fuzzy C-Means Clustering (FCM) is applied to divide the training dataset and Correlation-based Feature Selection (CFS) techniques is used to remove the irrelevant features. In addition, If-Then rules are generated through the CART technique according to the selected features in order to distinguish the normal and anomaly records. The proposed hybrid method is trained via the generated rules. The simulation results on UNSW-NB15Datasets shows that the proposed method achieves in terms of different performance metrics and can achieve 98.6% detection rate, 98.9 accuracy and 0.13% false Positive rate.

Guha et al.[24] proposed an effective approach to detect cyber-attacks in cloud infrastructures, including those through remote computing devices. In this approach an artificial neural network (ANN) is trained using the network traffic data on the connecting links of the cloud infrastructures. In their approach a genetic algorithm is used to reduce the number of features extracted from the network traffic data. This approach is illustrated by using UNSW-NB 15 dataset of network traffic, and shown that the results are better than those of existing methods for detecting cyber-attacks in cloud infrastructures.They achieved 95.46 % accuracy.

Kamarudin et al.[25] proposed an anomaly-based intrusion detection system using an ensemble classification approach to detect unknown attacks on web servers. The process involves removing irrelevant and redundant features utilizing a filter and wrapper selection procedure. Logitboost is then employed together with random forests as a weak classifier. The proposed ensemble technique was evaluated using UNSW-NB15 data set. They achieved false alarm rate = 0.18%, detection rate = 99.10 % and accuracy rate = 99.45%.This algorithm is more suitable for handling noisy and outlier data.

Moustafa et al.[26]proposed a novel Geometric Area Analysis (GAA) technique based on Trapezoidal Area Estimation (TAE) for each observation computed from the parameters of the Beta Mixture Model (BMM) for features and the distances between observations. GAA-based detection depends on the methodology of anomaly-based detection. It constructs the areas of normal observations in a normal profile with those of the testing set estimated from the same parameters to recognize abnormal patterns. They also designed a scalable framework for handling large-scale networks. Their GAA technique considers a decision engine module in this framework. Using UNSW-NB15 datasets the performance GAA technique is evaluated. Principal Component Analysis (PCA) is applied to reduce the high-dimensional data of network connectionsand then evaluated its influence on the GAA technique.They achieved the overall DR and accuracy 77.4% and 91.8%, respectively, but the overall FPR decreases from 8.3% to 5.8%.

Nguyen, et al.[27] proposed a novel framework that uses a deep learning approach to detect cyberattacks in mobile cloud environment. The proposed framework applied on UNSW-NB15 dataset to recognize diverse cyberattacks. The learning model detects cyberattacks in the cloud system. There are two phases in the learning model,i.e.,feature analysis and learning process. The learning process includes three main steps, i.e., pre-learning, deep learning, and softmax regression steps.

Deep learning algorithm achieves high accuracy up to 95.84% and TPR 79.19% respectively.

Primartha et al.[28] proposed an effective random forest classifier with parameter setting for improving the performance of anomaly detection in IoT network. Ten classifiers were built and evaluated on the basis of the number of tree in the ensemble, with UNSW-NB15 data sets, involved in the experiment. Their study revealed that RF800 was statistically significant compared to other classifiers. Furthermore, the proposed model outperformed other methods with respect to accuracy and FAR metric. It shows an excellent result so far using 10fold cross validation technique. They achieved Accuracy 95.5 % and 7.22 % FAR .

Siddiqui et al.[29]proposed a wavelet based multiscale Hebbian learning approach in neural networks to address the challenge of class overlap. The proposed methodology is able to distinguish non-linear and overlapping classification boundaries sufficiently well. Empirical results on simulated and real-world UNSW-NB15 dataset have been presented.The classification performance results for gradient descent(GD), single scale Hebbian and multiscale Hebbian based neural network(HNN) have been shown. The proposed NN-Multi Scale Hebbian has successfully improved true negative rate of the dataset to 95% .It is found that mean detection accuracy for NN- Multi Scale Hebbian is 93.56%.

Nahiyanet al.[30] proposed an automated, agent-based, unsupervised, relatively less complicated cognitive approach. This approach segregates attacks from normal events within the large search space with reduced computational demands. The proposed algorithm collects features from statistical analysis of the observed attributes over each time-step and uses machine learning to isolate the attack events from normal attack using an unsupervised k-means clustering algorithm over the reduced dataset. The agent based architecture is used to optimize the computational load for central processingwhere The agent based architecture deploysagents in hosts, and some processing is done at the host and the rest is performed by the node that performs the classification. They achieved total recall, precision and f1 -score 92%,91% and 91% respectively for time 8 seconds using UNSW-NB15 dataset.

Roy et al.[31] proposed a novel deep learning technique for detecting attacks within the IoT network using Bi-directional Long Short-Term Memory Recurrent Neural Network (BLSTM RNN). A multi-layer Deep Learning Neural Network is trained using a novel benchmark data set: UNSWNB15. They focused on the binary classification of normal and attack patterns on the IoT network. The proposed BLSTM model is able to detect attacks using the reduced UNSW_NB15 dataset, with more than 95% accuracy with 100% precision. The model generates a zero false alarm rates and a very low wrong detection rate of 0.04% with an impressive recall and f1score value of 98%.

Moustafa et al.[32] proposed an ensemble intrusion detection technique to reduce malicious events particularly botnet attacks against DNS, HTTP and MQTT protocols utilized in IoT networks. From these protocols new statistical flow features are obtained based on an analysis of their potential properties. Then, ensemble learning method namedAdaBoost is developed using Decision Tree (DT), Naive Bayes (NB) and Artificial Neural Network (ANN) machine learning techniques.AdaBoost evaluates the effect of these features and detect malicious events effectively. The UNSW-NB15 with simulated IoT sensors' data are used to extract the proposed features and evaluate the ensemble technique. The proposed ensemble technique provides a higher detection rate and a lower false positive rate compared with each classification technique included in the framework. The simplest feature selection method Correlation Coefficient (CC) is used to compute the strength degree between some features.Using the DNS data source of the UNSW-NB15 dataset, the accuracy and DR of the ensemble method achieved 99.54% and 98.93%, respectively, while the FPR produces 1.38%, which outperforms the performance of the DT, NB and ANN techniques. HTTP data source of the UNSW-NB15 dataset, the accuracy and DR of the ensemble method achievedis 98.97%, 97.02% and FPR 2.58%.The DT technique produces a 95.32% accuracy, 94.15% DR and 5.22% FPR, and then the ANN technique achieves a 92.61% accuracy, 91.48% DR and 7.87% FPR. Lastly, the NB technique achieves an accuracy rate of 91.17%, 90.78% DR and 8.25% FPR.

Tama et al.[33]proposed deep neural network for classifying attacks in IoT network. The performance of the proposed method is evaluated on the UNSW-NB15 benchmark datasets in wired and wireless network environment. Deep neural network combined with grid search strategy are utilized to obtain the best parameter settings for dataset. They employed three different resampling strategies i) Cross-validation ii) Repeated cross-validation (RepCV) and iii)Subsampling. The performance of DNN is assessed using these three validation methods. They achieved accuracy94%, precision95%, recall 96% using deep neural network for each resampling strategies.

Beloucha et al.[34] proposed a framework which evaluates the performance of four classification algorithms; SVM, Naive Bayes, Decision Tree and Random Forest using Apache Spark for intrusion detection in network traffic.Apache Spark a big data processing tool. Using UNSW-NB15 dataset it is observed that Random Forest(RF) classifier perform better than all the remaining classifiers in terms of sensitivity. RF gets 93.53% sensitivity followed by Decision Tree with 92.52%. Naive Bayes and SVM have almost same sensitivity with values 92.46% and 92.13% .They found that specificity for the Random Forest and Decision Tree based schemes are almost same with 97.75% and 97.10% respectively. However, specificity for SVM based scheme is about 91.15%. Naive Bayes provides lowest Specificity. Random Forest perform better among the all in terms of accuracy with 97.49% and the accuracy of the Naive Bayes based scheme is lower among the all schemes with 74.19%.

Zhou et al.[35] proposed a framework known as Deep Feature Embedding Learning (DFEL) to detect the internet intrusion in the IoT environment. DFEL boosts classifiers' accuracy to predict cyberattack.

DFEL used to balance the detection performance and speed. The UNSW-NB15 dataset is randomly split using the same rule. 80% of data was used to fit DFEL and get the pre-trained model. The remaining 20% of the data was randomly split into 70%/30% as training/testing data for classifiers. Next, the 20% rest data was transferred to latent attributes using DFEL and the embedding features are split into 70%/30% for embedding training/embedding test. Finally, the performances from traditional machine learning algorithms are compared on embedding data and original data. The machine learning algorithms used for boosting include gradient-boosted trees{GBT), k nearest neighbor(KNN), decision tree(DT), logistic regression(LR), gaussian naive bayes(GNB) and support vector machine(SVM). The DFEL approach boosts most classifiers accuracy and significantly saves the cyber detection time. The performances are evaluated for these algorithms with and without DFEL. The GNB classifier's accuracy increased from 50.45% to 92.52% The KNN's accuracy increased to 91.90%. The DT classifier's accuracy increased to 92.29%. The LR classifier's accuracy increased to 92.35%. The SVM classifier's accuracy increased to 92.32% The GBT achieved higher classifier's accuracy to 93.13%.There is increase in Precision and Recall of all algorithms with DFEL.

Moustafa et al.[36] proposed a Collaborative Anomaly Detection Framework (CADF) for detecting cyber-attacks on big data of cloud computing environments. They provided the technical functions and the way of deployment of this proposed framework for these environments. The technical framework comprises three modules: capturing and logging network data, pre-processing these data and a new Decision Engine (DE) using a Gaussian Mixture Model (GMM) and lower-upper Interquartile Range (IQR) threshold for detecting attacks. The CADF is evaluated by taking the features selected from the UNSW-NB15 dataset.The Receiver Operating Characteristics (ROC) curves displays the relationship between the DRs and FPRs using the w values. It is found that the stable increase in the w value between 1.5 and 3 increased the overall DR and accuracy while decreasing the overall FPR. The overall DR and accuracy increased from 86.3% to 95.6 % and 88.2% to 96.7%, respectively, however the overall FPR decreasedfrom 8.4 % to 3.5% when the w value of ROC curve increased from 1.5 to 3.

AL-Hawawreh et al.[37] presented an anomaly detection technique for Industrial Internet of Things(IIoT) or Internet Industrial Control Systems (IICS) based on deep learning models that can learn and validate using information collected from TCP/IP packets. It includes a consecutive training process executed using a Deep Auto-Encoder (DAE) and Deep Feed Forward Neural Network (DFFNN) architecture which is evaluated using UNSW-NB15 network dataset. In the training phase, a DAE algorithm learns using normal network observations to create the initialization parameters like weights and biases and learn a deep representation of normal behaviors. These parameters are used as an initialization stage for training a standard DFFNN to discover existing and new attack instances. In the testing phase, the DFFNN is used to recognize malicious vectors. They achieved accuracy 92.4%, DR 93%, and FPR 8.2% on UNSW-NB15 dataset. The detection rates for the attack types Analysis, Backdoor, DoS, Exploits, Fuzzer, Generic, Normal, Reconnaissance, Shellcode and Worms are 83.3%, 91.8%, 95.1%, 96%, 60%, 99.5%, 98.9%, 96.8%, 81.1% and 76% respectively.

Moustafa et al.[38] proposed a new threat intelligence scheme. It models the dynamic interactions of industry 4.0 component including physical and network systems. Industry4.0 includes the integration of Cyber-Physical systems (CPS), Internet of Things (IoT), Cloud and Fog computing paradigms for developing smart systems, smart homes, and smart cities. The scheme consists of two components: a smart data management module, and a threat intelligence module. The smart data management module handles heterogeneous data sources. This includes data to and from sensors, actuators, in addition to other forms of network traffic. The proposed threat intelligence technique is designed based on Beta Mixture-Hidden Markov Models (MHMM) for discovering anomalous activities against both physical and network systems. The scheme is evaluated on the UNSW-NB15 dataset of network traffic. The results shows that the proposed technique outperforms five peer mechanisms:Cart, KNN, SVM, RFand OGM. Using the UNSW-NB15 dataset, the proposed MHMM mechanism gives95.89% DR, 96.32% accuracy and 3.82% FPR which is better than others.

Timenko et al.[39] proposed several ensemble classifiers from the supervised learning category to detect network intrusion. They have evaluated Bagged trees, AdaBoost, RUSBoost, LogitBoost and GentleBoost algorithms on UNSW-NB15 dataset. All evaluated classifiers have a C4.5 decision tree and kNN classifier as a base learner. The learning procedure is based on 200 learners, with 0.1 as learning rate value, while settings for the subspace dimension are left on the default value, 1. They achieved overall accuracy of the classifiers as well as the ROC and AUC values for some of the traffic categories like Normal, Exploits DoS, Fuzzers and Reconnaissance. For Normal Traffic classification, Bagged tree and GentleBoost give Detection rate 100% and ROC value0.999. For Exploits Attack Traffic Classification , Bagged tree and GentleBoost give Detection rate 92.2% and 91.7% respectively. For DOS Attack Traffic Classification , AdaBoost and LogitBoost give Detection rate 92.7%. For Reconnaissance Attack Traffic Classification, Bagged tree and GentleBoost give Detection rate 98.5%. For Fuzzers Attack Traffic Classification Bagged tree and GentleBoost give Detection rate 99.1%. It is found that GentleBoost performs with highest accuracy and ROC values.

Moustafa et al.[40] proposed an architectural scheme for designing a threat intelligence technique for web attacks through a four-step methodology: First by collecting web attack data by crawling websites and accumulating network traffic for representing this data as feature vectors; second by dynamically extracting important features using the Association Rule Mining (ARM) algorithm; third by using these extracted features to simulate web attack data; and last by using a new Outlier Gaussian Mixture (OGM) technique for detecting known as well as zero-day attacks based on the

anomaly detection methodology. The OGM technique compared with four competing techniques, namely Cart, KNN, SVM and RF .The Receiver Operating Characteristics (ROC) curves signify the relationship between the DR and FAR in order to effectively show the potential process of running these techniques using the original data in the UNSW-NB15 dataset. Empirical results shows that the OGM outperforms others, producing a 95.68% DR and 4.32% FAR, while the others achieve in an average of 89%-93% DR and 6.4%-10.5% FAR.

Tian et al.[41] proposed a methodology for anomaly detection by introducing Ramp loss function to the original One-class SVM, called "Ramp-OCSVM".The Concave–Convex Procedure (CCCP) is utilized to solve the obtained model that is a non-differentiable non-convex optimization problem. They performed comprehensive experiments and parameters sensitivity analysis on UNSW-NB15 data sets. Ramp-OCSVM outperforms the OC-SVM, ROCSVM and eta OCSVM on UNSW-NB15 data sets. Using Ramp-OCSVM, they achieved values of 97.24%, 93.07% and 2.25% for the total accuracy, detection rate, false alarm rate respectively.

Nawir et al.[42] proposed Network Intrusion Detection System using machine learning algorithms for binary classification. They used three types of ML algorithms from Bayesian's family in WEKA tools. They are Average One Dependence Estimator (AODE), Bayesian Network (BN), and Naive Bayes (NB). The performance these classifiers measured in term of classification rate and processing time for classifier model to classify the data instances of UNSW-NB15 dataset. The parameters of these classifiers set to default as in WEKA and using tenfold cross validation to validate the training set before the model been tested. It is found that AODE is processing fast for network anomaly detection system compared to other two classifiers with accuracy 94.37% with training time 4.13s. BN algorithm gives the accuracy 92.70% and time taken is 4.17s. Naive Bayes algorithm required small amount of time but its accuracy is not comparable to AODE and BN algorithms.

Viet et al. [43] proposed a new network scanning techniques using a Deep Belief Network(DBN). They used both supervised and unsupervised machine learning methods with DBN for port scanning attacks detection. The port scanning attacks detection is the task of probing enterprise networks or Internet wide services, searching for vulnerabilities or ways to infiltrate IT assets. For the UNSW-NB15 dataset, the scanning types are labelled together, they only apply a binary classification to determine whether the data is an attack or not. They also used the ''normal" data to train and test the model. They compared the results of the DBN with SVM and Random Forest. They achieved TPR99.74%, 99.80% and 99.86% for SVM, Random and DBN respectively. They achieved FAR3.20%, 3.31% and 2.76% for SVM, Random and DBN respectively. Experiments with the UNSW-NB15 dataset found that the DBN algorithm gives high detection rates for network scanning, while ensuring a lower false alarm rate.

## VII. SUMMARY OF MACHINE LEARNING BASED IDSS WITH SINGLE OR MULTIPLE CLASSIFIER

Machine learning approaches have been used in different ways to detect intrusions using UNSW-NB15 dataset. Table 1. gives the summary of survey which includes only best performances by ML approaches.

## VIII. FUTURE SCOPES

Deep learning is the betterment of the neural network. It became popular in recent years. The current IDS can be improved by using this new technique. The deep learning methods are classified as per their architecture into threetypes: generative (unsupervised), discriminative (supervised) and hybrid.

To improve efficiencyand minimize the training time we need high computing resources which are very costly and require more power. Reinforcement learning (RL) is one of the emerging field and the research is still going towards attacks detection. Also Deep Reinforcement Learning can be applied as the next step for intrusion detection applications.

Future scopes are provided to help researchers for finding more efficient solutions to detect the attacks. Existing literature is described which have similar techniques with UNSW NB-15 dataset to generalize our observations. All the ML based techniques discussed have not been implemented to check the performance for ensuring the results are reproducible. This is a limitation of our paper and we are very hopeful to improve this as a future work. As a future scope, we would also like to propose an attack detection system to improve the performance of low-frequency attacks. Future directions insists the usage of deep learning and reinforcement learning techniques and Subspace ML for intrusion detection.

## IX. CONCLUSION

The use of the computers, mobiles, sensors, IoTs, Big Data, Web Application/Server, Clouds and other computing resources are increased. All these are prone to intrusions.Researchers have worked on various solutions to detect intrusions. The machine learning based intrusion detection approaches using UNSW-NB15 dataset have beenconsidered in our paper. The analysis performed shows that no one particular intrusion detection technique can help in detecting all types of attacks. Then we have seen how features selection and multiple classifier approaches affect

**Table 1.Summary of Machine Learning based IDSs using UNSW-NB15 Dataset**

| Reference | ML Approach | Features | Feature Selection approach | Classifier and Feature Selection | Attacks Detected | Performance | Problem Domain |
|---|---|---|---|---|---|---|---|
| Moustafa et al. [13] | EM, NB, LR | - | CP, Apriori, ARM | MCLF | Normal, DoS, Fuzzer Analysis, etc | Accuracy(LR) 83.0% FAR(EM) 13.1% | Hybrid Detection |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Ghaaee et al. [14] | LSSVM | 6-14 | GA and SVM | SCLF | Normal, DoS,Fuzzer Analysis, etc | Accuracy99.45%, TPR 100%, FPR0.01% | Misuse Detection |
| Chowdhury et al. [15] | SVM | 3 | Simulated Annealing | MCLF | Normal, DoS, Fuzzer Analysis, etc | Accuracy98.76%, FPR0.09%, FNR1.35% | Misuse Detection |
| Bhamare et al. [16] | Logistic Regression | All | NA | SCAF | Normal, DoS, Fuzzer Analysis, etc | Accuracy89.26%, FPR4.3% | Misuse Detection |
| Baig et al.[17] | boosting based ANN AdaBoost | All | | MCAF | Normal, DoS, Fuzzer Analysis, etc | Accuracy 86.40% Precision- 0.8674 Recall 0.9338 F1 Score 0.8994 | Anomaly Detection |
| Belouch et al. [18] | RepTree | 20 | Ranker algorithm | MCLF | Normal, DoS, Fuzzer Analysis, etc | Accuracy 88,95% | Anomaly Detection |
| Al-Zewairi et al.[19] | MFFANN | Gedeon method | | MCLF | DoS, Fuzzer Analysis, etc | Accuracy 98.99% ,FAR 0.56% | Anomaly Detection |
| Anwer et al.[20] | J48 and Naïve Bayes | 18 | Filter and Wrapper | SCLF | Normal, DoS, Fuzzer Analysis, etc | Accuracy 88% | Anomaly Detection |
| Mithun et al.[21] | Ensemble classifier | - | K means | MCLF | Normal, DoS etc | Accuracy 90% | Anomaly Detection |
| Idhammad et al.[22] | FNN ADDM | Relevant features | CFS, CNF | SCAF | DoS | Accuracy 97.1% in 0.46s. | Misuse Detection |
| Hajisalem et al.[23] | ABC AFS | -- | CFS, FCM | MCLF | Normal, DoS, Fuzzer Analysis, etc | DR 98.6%,Accuracy 98.9%, FPR 0.13% | Hybrid Detection |
| Guha et al.[24] | ANN | - | GA | SCLF | Normal, DoS, Fuzzer Analysis, etc | Accuracy.95.46 % | Misuse Detection |
| Kamaruddin et al.[25] | ensemble classifierLogitboost, Random Forests | 5 | Filter and Wrapper | MCLF | Normal, DoS, Fuzzer Analysis, etc | DR 99.10%, Accuracy 99.45%,FAR 0.18% | Anomaly Detection |
| Moustafa et al.[26] | GAA, TAE | - | BMM, PCA | MCAF | Normal, DoS, Fuzzer Analysis, etc | Overall DR77.4%,Accuracy 91.8%, Overall FPR 5.8%. | Anomaly Detection |
| Nguyen1, et al.[27] | Deep learning | PCA | - | SCAF | Normal, DoS, Fuzzer Analysis, etc | Accuracy 95.84% TPR 79.19% | Misuse Detection |
| Primartha et al.[28] | random forest RF800, Esamble Classifier | All | - | MCAF | Normal, DoS, Fuzzer Analysis, etc | Accuracy 95.5% , 7.22% FAR for RF800 . | Anomaly Detection |
| Siddiqui et al.[29] | NN-GD, NN-Multi Scale Hebbian | All | - | MCAF | Normal, DoS, Fuzzer Analysis, etc | TNR 95%, Mean DR 93.56% for NN -MSH | Misuse Detection |
| Nahiyan[30] | k-means | - | Statistical techniques | SCLF | Normal, DoS, Fuzzer Analysis, etc | Total Recall 92%, Precision91%,F1 –score 91% | Anomaly Detection |
| Roy et al.[31] | BLSTM RNN | - | - | SCLF | Norrnal, Attacks | Accuracy 95%, Precision 100%, Recall and F1score 98%. | Anomaly Detection |
| Moustafa et al.[32] | AdaBoost, DT, NB, ANN | - | CC | MCLF | Normal, DoS, Fuzzer Analysis, etc | Accuracy 99.54%, DR 98.93%, FPR 1.38% for AdaBoost | Misuse Detection |
| Tama et al. [33] | DNN | All | - | SCAF | Normal, DoS, Fuzzer Analysis, etc | Accuracy94%, Precision95%, Recall 96% | Anomaly Detection |
| Belouchа et al.[34] | NB, DT, RF, SVM | All | - | SCAF | Normal, DoS, Fuzzer Analysis, etc | Accuracy97.49%, Sensitivity93.53% for RF | Misuse Detection |
| Zhou et al.[35] | DFEL with GBT,GNB ,DT, LR,SVM | - | PCA | SCLF | Normal,DoS, Fuzzer Analysis, etc | Accuracy 93.13%, Precision 92.38%for GBT. | Misuse Detection |
| Moustafa et al.[36] | DE using GMM | All | - | SCAF | Normal, DoS, Fuzzer Analysis, etc | DR 95.6 ,Accuracy 96.7% , FPR 3.5% | Anomaly Detection |
| AL-Hawawreh et al.[37] | DFFNN, DAE | All | - | SCAF | Normal, DoS, Fuzzer Analysis, etc | DR of DoS 95.1%,Exploits 96%,Generic99 .5% ,Normal 98.9%,Reconnaissance96.8% | Anomaly Detection |
| Moustafa et al[38] | MHMM Cart, KNN, SVM, RF and OGM | All | - | SCAF | Normal, DoS, Fuzzer Analysis, etc | DR 95.89%, Accuracy 96.32%, FPR 3.82% for MHMM | Anomaly Detection |
| Timenko et al.[39] | Bagged trees, AdaBoost, RUSBoost , LogitBoost and GentleBoost, C4.5 , kNN | - | - | MCAF | Normal,DoS, Fuzzer Analysis, etc | DR100, ROCvalue0.999 for GentleBoost | Misuse Detection |
| Moustafa et al[40] | OGM Cart, KNN, SVM RF | - | ARM | SCLF | Normal,DoS, Fuzzer Analysis, et | DR 95.68% 4.32% FAR for OGM | Anomaly Detection |
| Tian et al[41] | Ramp-OCSVM, OC-SVM, ROCSVM , eta OCSVM | - | - | SCAF | Normal,DoS, Fuzzer Analysis, etc | Total Accuracy 97.24%, DR 93.07%, FAR 2.25% For Ramp-OCSVM | Anomaly Detection |
| Nawir et al[42] | AODE, BN, NB | - | - | SCAF | Norrnal, Attacks | Accuracy 94.37% for AODE | Misuse Detection |
| Viet et al.[43] | DBN, SVM RF | - | - | SCAF | Normal, Attacks | TPR 99.86%, FAR 2.76% For DBN | Hybrid Detection |

the performance of IDS. We have discussed various datasets with their shortcomings. Then we have insisted useof new benchmark dataset UNSW-NB15 which have all new attacks. We have described various types of attacks in the UNSW-NB15 dataset with their features. Future research directions are rendered to help researchers exploring more efficient solutions for attack detection.All the IDS discussed have not been implemented to find the performance to to make sure that results are reproducible. This endures a limitation of our paper and we are very eager to improve this as a future work.

## REFERENCES


1. Heady R., Luger G., MaccabeA.,Servilla M.:The architecture of a network level intrusion detection system, Tech. rep., Computer Science Department, University of New Mexico, New Mexico, (1990)
2. Stefan A.: Intrusion detection systems: A survey and taxonomy, Technical report, Vol. 99,( 2000)
3. Vigna G., KemmererR. A. : Netstat: A network-based intrusion detection system, in Journal of Computer Security. Citeseer, (1999)
4. Agrawal S.,Agrawal J. : Survey on anomaly detection using data mining techniques, Procedia Computer Science, vol. 60, pp. 708–713, (2015)
5. Buczak A. L., Guven E.:A survey of data mining and machine learning methods for cyber security intrusion detection, IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, (2015)
6. Mishra P., Varadharajan V., Tupakula U., PilliE.S.:A Detailed Investigation and Analysis of using Machine Learning Techniques for Intrusion Detection, IEEE Communications Surveys & Tutorials(2018).

7. McHugh J.:Testing Intrusion detection systems: a critique ofthe 1998 and 1999 DARPA intrusion detection system evaluations as performed by Lincoln Laboratory, ACM Transactions on Information and System Security, vol. 3, no. 4, pp. 262–294, (2000)
8. Tavallaee M., Bagheri E., Lu W., Ghorbani A.:A Detailed Analysis of the KDD CUP 99 Data Set, IEEE Symposium on Computational Intelligence in Security and Defence Applications (CISDA), (2009)
9. Moustaf N, Slay J.:The evaluation of network anomaly detection systems: Statistical analysis of the unsw-nb15 data set and the comparison with the kdd99 data set, Information Security Journal: A Global Perspective, in press. ids(2015)
10. NSLKDD. Available on: http://nsl.cs.unb.ca/NSLKDD/, (2009)
11. Moustafa N., Slay J.:Unsw-nb15: A comprehensive data set for network intrusion detection systems (unsw-nb15 network data set), in Military Communications and Information Systems Conference (MilCIS), Canberra, Australia, pp. 1–6,(2015)
12. Shiravi A., Shiravi H., Tavallaee M., Ghorbani A.A.:Toward developing a systematic approach to generate benchmark datasets for intrusion detection, computers & security, vol. 31, no. 3, pp. 357–374, (2012)
13. Moustafa N., Slay J.:A hybrid feature selection for network intrusion detection systems: Central points, pp. 1–10,( 2015).
14. H. Gharaee H., Hosseinvand H.:A new feature selection ids based on genetic algorithm and svm, in 8th International Symposium on Telecommunications (IST). IEEE, pp. 139–144,(2016)
15. Chowdhury M.N., Ferens K., Ferens M.:Network intrusion detection using machine learning, in Int. Conf. on Security and Management (SAM). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), pp. 1–7(2016)
16. BhamareD., SalmanT., SamakaM., ErbadA., R. Jain:Feasibility of supervised machine learning for cloud security, in International Conference on Information Science and Security (ICISS). IEEE, 1–5(2016)
17. Baig M.M., Awaisa M.M., El-Alfyb E. M.,:A multiclass cascade of artificial neural network for network intrusion detection", Journal of Intelligent & Fuzzy Systems 32 (2017) 2875–2883,(2017)
18. Belouch M.,Hadaj S.E, Idhammad M.:A Two-Stage Classifier Approach using RepTree Algorithm for Network Intrusion Detection, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 6, (2017)
19. Al-Zewairi M., Almajali S., Awajan A.: Experimental Evaluation of a Multi-Layer Feed-Forward Artificial Neural Network Classifier for Network Intrusion Detection System, IEEE International Conference on New Trends in Computing Sciences, (2017)
20. Anwer H.M., Farouk M., Abdel-Hamid A.: A Framework for Efficient Network Anomaly Intrusion Detection with Features Selection, 9th International Conference on Information and Communication Systems (ICICS) 2018, IEEE page no.157,(2018)
21. Mithun A.,Kalaiselvi V.K.G: Design of an Intrusion Detection System Based on Distance Feature Using Ensemble Classifier, IEEE-2017 4th International Conference on Signal Processing, Communications and Networking (ICSCN -2017), March 16 – 18, 2017, Chennai, INDIA,(2017)
22. Idhammad M., Afdel K., Belouch M: DoS Detection Method based on Artificial Neural Networks, (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 8, No. 4, (2017)
23. Hajisalem V., Babaie S.: A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection, Science Direct ,Computer Network, vol 136,page no,37 Elsevier, (2018)
24 .Guha S., Yau S.S., Buduru A.B.: Attack Detection in Cloud Infrastructures Using Artificial Neural Network with Genetic Feature Selection", 2016 IEEE 14th Intl Conf on Dependable, Autonomic and Secure Computing, 14th Intl Conf on Pervasive Intelligence and Computing, 2nd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress, IEEE Computer Society, (2016).
25. Kamarudin M.H. , Maple C., Watson T., Safa N.S.: A LogitBoost-Based Algorithm for Detecting Known and Unknown Web Attacks,IEEE Access Open Access Journal, vol 5 , page no. 26190, (2017)
26. Moustafa N., Slay J., Creech G.:Novel Geometric Area Analysis Technique for Anomaly Detection using Trapezoidal Area Estimation on Large-scale Networks, Journal Of IEEE Transactions on Big Data,(2017)
27. Nguyen K.K., Hoang D.H., Niyato D., Wang P., Nguyen D., Dutkiewicz E.: Cyberattack Detection in Mobile Cloud Computing: A Deep Learning Approach, IEEE Wireless Communications and Networking Conference (WCNC),(2018)
28. Primartha R., Tama B.A.:Anomaly Detection using Random Forest: A Performance Revisited, International Conference on Data and Software Engineering (ICoDSE), 2017
29. Siddiqui S., Khan M.S.,Ferens K.:Multiscale Hebbian Neural Network for Cyber Threat Detection, IEEE,(2017)
30. Nahiyan K., Kaiser S., Ferens K.,McLeod R.: A Multi-agent Based Cognitive Approach to Unsupervised Feature Extraction and Classification for Network Intrusion Detection, Int'l Conf. on Advances on Applied Cognitive Computing| ACC'17 page no. 25,CSERA Press,(2017)
31. Roy B.,Cheung H.: A Deep Learning Approach for Intrusion Detection in Internet of Things using Bi-Directional Long Short-Term Memory Recurrent Neural Network,28th International Telecommunication Network and Applications Conference (ITNAC) IEEE 2018
32. Moustafa N., Turnbull B., Choo K.R.: An Ensemble Intrusion Detection Technique based on proposed Statistical Flow Features for Protecting Network Traffic of Internet of Things, IEEE Internet of Things Journal (2018)
33. Tama B. A., Rhee K.H.: Attack Classification Analysis of IoT Network via Deep Learning Approach, Research Briefs on Information & Communication Technology Evolution (ReBICTE), Vol. 3, Article No. 15 (November 15, 2017)
34. Beloucha M., HadajaS.E.. Idhammadb M.: Performance evaluation of intrusion detection based on machine learning using Apache Spark, The First International Conference On Intelligent Computing in Data Sciences Performance, Procedia Computer Science 127 (2018) 1-6, Elsevier(2018)
35. Zhou Y., Han M., Liu L., He J., Wang Y.:Deep Learning Approach for Cyberattack Detection, 2018 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS): 2018 IEEE Infocom MiseNet Workshop(2018)
36.. Moustafa N., Creech G., Sitnikova E., Keshk M.: Collaborative Anomaly Detection Framework for handling Big Data of Cloud Computing
37. AL-Hawawreh M., Moustafa N.,Sitnikova E.: Identification of malicious activities in industrial internet of things based on deep learning models,Journal of Information Security and Application,41, 1-11 ELSEVIER(2018)
38. Moustafa N.,Adi E.,Turnbull B., Hu J.: A New Threat Intelligence Scheme for Safeguarding Industry 4.0 Systems, IEEE Access Open Access Journal, vol 4 , page no. 1, (2018)
39. Tim_enko V.,Gajin S.: Ensemble classifiers for supervised anomaly based network intrusion detection, IEEE (2017)
40. Moustafa N., Misra G., Slay J :Generalized Outlier Gaussian Mixture technique based on Automated Association Features for Simulating and Detecting Web Application Attacks, Journalof IEEE TransactionsonSustainable Computing,IEEE (2018)
41. Tian Y. , Mirzabagheri M., Mojtaba S., Bamakan H., Wang H., Qu Q.: Ramp loss one-class support vector machine; A robust and effective approach to anomaly detection problems,Journal neorocomputing , Elsevier,(2018)
42. Nawir M., Amir A., Lynn O.B., Yaakob N.,Ahmad R.B.:Performances of Machine Learning Algorithms for Binary Classification of Network Anomaly Detection System,1st International Conference on Big Data and Cloud Computing (ICoBiC) 2017 IOP Publishing,IOP Conf. Series: Journal of Physics(2017)
43. Viet H.N., Van Q.N. , Trang L.L.T.,Nathan S. :Using Deep Learning Model for Network Scanning Detection, ICFET '18, June 25–27, page no, 117,ACM (2018)