e-ISSN: 2614-8404
p-ISSN : 2776-3234

Network Intrusion Detection Based on Machine Learning Classification Algorithms: A Review

Aqeel H.younus[1]*), Adnan Mohsin Abdulazeez[2]
[1]Akre University for Applied Sciences/ Technical College of Informatics -Akre/ Department of Information Technology
[2]Duhok Polytechnic University, Duhok, Kurdistan Region, IRAQ
Email: [1]aqeel.hanash@auas.edu.krd , [2]adnan.mohsin@dpu.edu.krd

Abstract – The worldwide internet continues to spread, presenting numerous escalating hazards with significant potential. Existing static detection systems necessitate frequent updates to signature-based databases and solely detect known malicious threats. Efforts are currently being made to develop network intrusion detection systems that can utilize machine learning techniques to accurately detect and classify hazardous content. This would result in a decrease in the overall workload required. Network Intrusion Detection Systems are created with a diverse range of machine learning algorithms. The objective of the review is to provide a comprehensive overview of the existing machine learning-based intrusion detection systems, with the aim of assisting those involved in the development of network intrusion detection systems..

Keywords: Intrusion Detection Systems, Machine learning, SVM, Random Forest.

I. INTRODUCTION

Currently, the intrusion detection systems provides a key component when it comes to making sure the systems owners are safe against the cyber-threats. IDS (Intrusion Detection System) is a forms of gather and analyze network data to classify types of attacks[1]. For the network traffice, it is the used of many day-to-day features creation in the form of detecting many types of attacks [2]. Due to the rapid increase in the data that is being generated via the internet in daily life, the industry faces a severe challenge[3]. Datasets are sets ofs situation which includemany features and they are relating to the response of the intrusion detection system[4]. Understanding the type of data that is being collected becomes more important because it has attack types and attributes[5]. The KDD'99 cup is the most widely used dataset for intrusion detection systems. It is used to construct predictive models that can distinguish between different types of intrusions or attacks [6]. The intrusion detection system constructs the model using security datasets such as KDD99 and NSL-KDD [7]. The system has many features, akin to a predictor, that differentiate between normal attacks and aberrant ones. These features are the focus of the system [8] .The categorization model divides the data set into two parts: a training stage and a testing stage [9]. The abundance of characteristics with large dimensions results in intricacy during the training process and consumes valuable time. Hence, it is necessary to carefully choose a subset of valuable and pertinent features from the complete set of features in order to enhance the model's performance during the testing phase [10]. Data preparation is a crucial step in enhancing the quality of a classification model's performance, as stated by machine learning algorithms[11]. The process of solving various forms of large data sets is a highly important phase [12].Machine Learning (ML) techniques, which are commonly employed in computer security data sets, have lately gained popularity in the field of security technology [13]. It aids in the examination and management of large volumes of data and identifies the crucial characteristics that are employed in different feature selection strategies [14]. Intrusion Detection System (IDS) is a widely employed machine learning classifier that is utilized to differentiate between different types of attacks inside a given class [15]. Several supervised classification algorithms are commonly used in Intrusion Detection Systems (IDS), including Decision Trees, Naïve Bayes, K-Nearest Neighbor, Tree C4.5, Random Forest, Support Vector Machine, and Logistic Regression [16]. Assessment of different classifiers is based on the list of statistical measures above all, the results of the confusion matrix-dependent diagnoses are considered to distinguish the kind of dangers [17]. The goal of the article is to contribute to the network intrusion detection system development process by providing an exhaustive study of the present machine learning-based intrusion detection systems.

The remainder of the review is arranged as follows: In Section 2, the types of intrusion detection systems are described, as are network attacks and the types of them, additionally, the algorithms of machine learning and classical system architecture are explained. In Section 3, the study provide a review of the literature on intrusion detection systems. Section 4, compares and discusses intrusion detection systems. In Section 5, the conclusion is presented in the final part.

II. INTRUSION DETECTION SYSTEM

Intrusion detection systems (IDS) monitor network traffic data on systems or networks through the use of hardware or software. An Intrusion Detection System (IDS) typically reports any instances of policy violations or security breaches. Figure 1 displays a standard block diagram of an Intrusion Detection System (IDS) [18]. An intrusion detection system includes a static database that contains information about known malicious activity. The input is compared to the records in this database, which encompass system activity or network traffic. If the input is malicious, the

e-ISSN: 2614-8404
p-ISSN : 2776-3234

severity of the threat is determined and a suitable countermeasure is implemented. The countermeasures range from simple notifications to halting the potentially hazardous activity. The two predominant varieties of intrusion detection systems (IDS) are host-based IDS and network-based IDS. In networking contexts, Intrusion Detection Systems (IDS) are employed as a means to detect and locate unauthorized access attempts. It detects instances of malicious utilization of computer network resources. This feature is essential for recognizing internet threats that originate from hosts and networks[19].

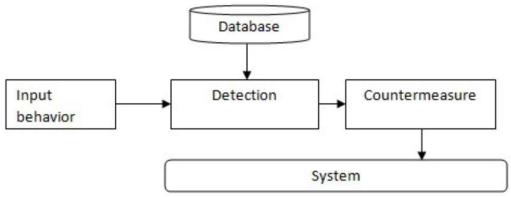

Fig 1: Block Diagram of IDS*[20]*

2.1. NETWORK ASSAULTS AND THEIR KINDS

Network assaults refer to unauthorized actions aimed at governmental or commercial IT assets, with the objective of causing destruction and pilfering confidential information. Attacks can be classified into two distinct categories: aggressive and passive. Presently, hackers are engaged in the act of altering confidential information or fortifying computer systems with excessive security measures. Some examples of cyber threats include Trojan horses, worms, viruses, code injections, network data probing, and login information theft. The prevalent and widely recognized active attacks include denial of service, replay, repudiation, masquerade, and message alteration [19]. A "passive attack" is an attempt to access important data by observing and monitoring sensitive information without causing any disruption to system resources. Two prevalent and widely recognized passive attacks are traffic analysis and message content release.

The attack can manifest in several forms, and it can be either proactive or passive.

- The Definition of Service Denial (DoS) is a DDOS|DDoS| Denial of Service(DOS), its purpose is to starve network and system resources for computer networks or to just send lots of unnecessary data to the network to make its termination.
- Scanning attacks' investigations require the following two steps: identification of network weaknesses and attackers. Next, the victim will get DMCP bypass which leads to all legitimate procedures.
- Remote to Local (R2L): Such a scenario implies that an intruder tries to make a remote login directly, therefore it is a hacking attempt that is likely going to be a brute force attack that pretends to be a genuine user.
- User to Root (U2R): An intruder who has user-level access in an attempt to take over the high-level authority.

2.3. CLASSICAL SYSTEM ARCHITECTURE

The typical IDS architecture consists of five essential modules: talk about data gathering, data experience, categorizing or gunning and invading prevention. The preprocessing module of our system takes on all necessary data from benchmark datasets that find use in wavelet transform and applies a sequence of preprocessing operations. The very first step in any analysis process is data purification, and it is in this module data preparation takes care of it. Data preparation involves a set of fundamental stages such as: data combination, cleansing, standardization, alteration, really the degree of data reduction and the binning of categorical data. The feature selection module employed resilient and intelligent algorithms to identify the crucial traits required for enhanced classification. By utilizing the selected attribute for categorization, the classifications exhibit enhanced precision. The decision manager, who is responsible for each module, uses the rules recorded in a rule base. Usually, the rules are saved in the form of IF-THEN expressions (Fig. 2) [21].

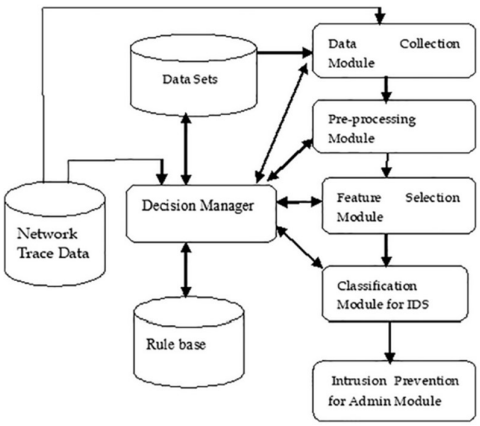

Fig 2. Classical System Architecture of IDS.

2.4. MACHINE LEARNING ALGORITHMS

The learning algorithm extracts the pertinent data from the training sets. Machine learning algorithms can be classified into two categories: semi-supervised and supervised. While an unsupervised learning algorithm navigates through unfamiliar data, a supervised learning system [22] acquires information from tagged samples. Prior to developing a decision model, the classifier goes through a training step. Below, we

e-ISSN: 2614-8404
p-ISSN : 2776-3234

present a detailed description of the pivotal machine learning classifier that is capable of detecting network flow attacks.

2.4.1. SUPPORT VECTOR MACHINE.

Finding boundaries in multidimensional space is accomplished by categorization and guesswork using the Support Vector Machine (SVM) supervised learning technique. It uses a hyperplane to separate data points into two classes, +1 and -1. Therefore, ordinary data is represented by a +1, and dubious data by a -1.The hyperplane can be expressed as follows: WX + b = 0, where b is a scalar and W= {w1, w2,..., wn} is the weight vector for n attribute values{x1, x2, x3,..., xn}. One of SVM's preferred features is its ability to

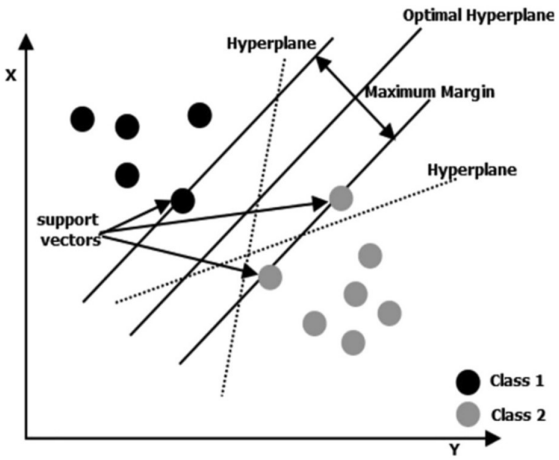

Fig 3. Multiclass Support Vector Machine.

classify using support vectors instead of the complete dataset, which makes it incredibly resilient to outliers and allows for exceptionally accurate guessing. Discovering the linear optimal hyper plane is the goal of the support vector machine in order to amplify the partition boundary between the two classes .It is decided that the hyperplane with the peak margin is the best one[23]. This machine does multiclass classification [24], which is achieved by creating a support vector machine (SVM) for each of the two classes together "Figure. 2".

2.4.2. DECISION TREE ALGORITHM (DT)

DT algorithms generally deduce relevant classifications. The decision tree consists of leaf nodes, edges, and root nodes. The initial node is the root node, which does not have any incoming nodes; the second node is an internal node, and the rest of the nodes are referred to as decision nodes (leaf). We evaluate the internal nodes using a diverse range of criteria and characteristics. When constructing a decision tree based on attribute features and information gain values, we choose the decision node that has the maximum information gain value. Decision trees [25]exhibit a high level of accuracy and efficiency in classifying data.

2.4.3. NEAREST NEIGHBOR ALGORITHM (K-NN)

Comparing the K-NN classifier to other classification algorithms, it is incredibly easy to learn and straightforward. It determines the separation between data points and assigns an unlabeled data point to the student who is closest to it [26]. The data is allocated to the fellow citizen's class if k = 1. When the K value is high, classification and prediction take a long time (lazy learners). The value of k will therefore depend on the classification time. Numerous studies employ various formulas, including Manhattan, Murkowski, and x Euclidean, to calculate the distance between neighboring nodes. When two data points with m quantitative features are separated by a Euclidean distance, with x=(x1, x2,..., xn) and y=(y1, y2,..., yn),

$$d(x,y) = \sqrt{\sum_{i=1}^{m}(x1 - y1)^2} \quad (1)$$

Assume y is the data point that is closest to x. The steps of the K-NN algorithm

1. Keep the labeled NSL-KDD Training data set in storage.
2. K is the quantity of nearby nodes.
3. Determine the distance between the test and training samples (x, y) and the two data points (x', y'). Then, designate the node that is closest to its neighbor in terms of distance.
4. Carry out step 4 for every data point in the testing dataset.
5. Come to an end. These types of indolent learners take longer to categorize and perform the best in predictions.

2.4.4. RANDOM FOREST (RF)

A decision forest classifier, which is randomly generated, is capable of handling both regression and classification tasks. This classification technique generates a multitude of decision trees by employing a random feature selection process. Utilize the voting methodology from different decision trees to allocate each desired value. The upper echelon of voters decides the definitive and more precise forecast [27]. This technique exhibits a higher accuracy rate and prediction capability due to its ability to create fewer classification errors [28].

2.4.5. NEURAL NETWORK ALGORITHM (NN)

The neural network algorithm consists of three layers: the input layer, the hidden layer, and the

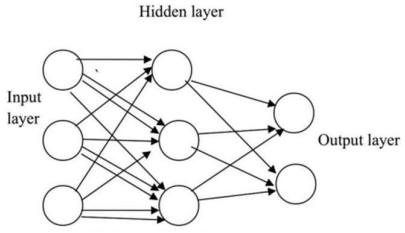

Fig 4. Neural Network

e-ISSN: 2614-8404
p-ISSN : 2776-3234

output layer. Upon completion of data processing, the output layer receives the data from the hidden layer. A multi-layer perceptron algorithm effectively detects and accurately recognizes several threats. This algorithm utilizes the back-propagation technique, which is rooted in the principles of feed-forward and back-propagation. The neural network illustrated in "Figure 3" consists of input, hidden, and output layers.
.

III. LITERATURE REVIEW

This article examined recent studies on machine learning methodologies for detecting intrusions. This study specifically examines recent publications from the years 2020 to 2024 as shown in “Table 1”. In addition, this article examines various machine learning techniques, such as single, hybrid, and ensemble classifiers that are employed in the field of intrusion detection. Further investigation is still necessary to develop machine learning algorithms for intrusion detection systems, using the comparative findings from relevant studies.

Abrar et. al. in 2020 [29], created a highly efficient intrusion detection system (IDS) by employing machine learning classifiers to identify and prevent network intrusions, safeguarding network assets. The study utilized a range of machine learning classifiers, such as Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Logistic Regression (LR), Naive Bayes (NB), Multi-Layer Perceptron (MLP), Random Forest (RF), Extra Trees Classifier (ETC), and Decision Trees (DT), to assess their effectiveness on the NSL-KDD dataset. The study conducted preprocessing on the dataset to eliminate extraneous attributes, and subsequently trained and evaluated the model using various feature subsets. The test results demonstrated that the RF, ETC, and DT classifiers achieved an accuracy rate of 99% for all sorts of attacks, employing various feature sets. This demonstrates the efficacy of the proposed approach in accurately predicting network intrusions while also minimizing the required workload.

Kiran et. al. in 2020 [30], developed an IoT-specific intrusion detection system (IDS) using machine learning to detect future threats. The procedures encompassed constructing a test platform to replicate an Internet of Things (IoT) setting, developing a hostile system to generate malicious assaults, collecting the flow of data within the network, and generating machine learning algorithms to categorize the attacks. The Sensor480 dataset consists of 480 records and includes attributes that represent both normal and attack scenarios. The classifiers, including SVM, Naïve Bayes, Decision Tree, and Adaboost, exhibited accuracy levels ranging from 97.89% to 100%. Among them, the Decision Tree model had the best accuracy, attaining a perfect score of 100%. In summary, the study proved the efficacy of machine learning algorithms in accurately categorizing attacks in IoT networks.

Elmrabit et. al. in 2020 [31], assessed twelve machine learning (ML) algorithms to identify abnormal behaviors that suggest cyber assaults. The evaluation was conducted using three datasets: UNSW-NB15, CICIDS-2017, and ICS cyber-attack. The approach entailed utilizing both traditional machine learning algorithms and deep learning algorithms to train and test the data. Subsequently, the performance was assessed using diverse metrics. The study attained the highest level of accuracy using the Random Forest (RF) approach, with 88.5% accuracy for binary classification in the UNSW-NB15 dataset and 99.9% accuracy for binary classification in the CICIDS-2017 dataset. RF attained an accuracy of 73.6% in the UNSW-NB15 dataset and an accuracy of 99.9% in the CICIDS-2017 dataset for multi-class classification. The study revealed that RF had superior performance compared to the other algorithms in the majority of situations, therefore confirming its efficacy in identifying abnormal behaviors.

Injadat et. al. in 2020 [32], suggested a multi-stage optimized architecture for network intrusion detection system (NIDS) that utilizes machine learning to decrease computational complexity while preserving detection performance. The study employed two contemporary intrusion detection datasets, namely CICIDS 2017 and UNSW-NB 2015, and assessed their performance using diverse measures including accuracy, precision, recall, and false alarm rate. The proposed framework incorporates data pre-processing, feature selection, and hyper-parameter optimization strategies to improve the performance of the NIDS. The findings demonstrated that the BO-TPE-RF optimized random forest classifier, employing Bayesian optimization with Tree Parzen Estimator, achieved a superior detection accuracy of more than 99% for both datasets. Furthermore, it exhibited a higher level of precision compared to alternative optimization techniques and recent scholarly articles, with an improvement of 1% to 2%. Additionally, it demonstrated a reduced rate of false alarms by the same percentage. In addition, the feature selection approaches successfully decreased the size of the feature set by over 60% and further reduced the necessary training sample size by 33-50% compared to the training sample size after implementing the oversampling methodology.

Mebawondu et .al. in 2020 [33], a network-based Intrusion Detection System (NIDS) was created utilizing machine learning algorithms for the purpose of identifying and thwarting network intrusions. The study employed the UNSW-NB15 benchmark network intrusion dataset and applied feature weighting techniques such as information gain and gain ratio to determine the most significant features. The study constructed classification models utilizing the Naive Bayes (NB) and C4.5 algorithms, employing the

e-ISSN: 2614-8404
p-ISSN : 2776-3234

chosen features. The findings demonstrated that the C4.5 algorithm surpassed NB, attaining a 90.44% accuracy compared to NB's 75.09% accuracy in a two-class model simulation. The trials demonstrated that the accuracy of the technique for real-time network intrusion detection rose as the training ratio grew, indicating its feasibility.

Thaseen et.al.in 2020 [34], network breaches were identified using machine learning methods, including Naive Bayes, Support Vector Machine, Random Forest, and KNearest Neighbors, without the need to decrypt the packet contents. The dataset was generated by employing Wireshark to collect packets transported across a network. Subsequently, the study scrutinized their characteristics to categorize them as encrypted, unencrypted, malicious, or normal. The study obtained accuracy scores of 83.63%, 98.23%, 99.81%, and 95.13% for the Naive Bayes, Support Vector Machine, Random Forest, and KNearest Neighbors models, respectively. The Random Forest algorithm was determined to be the optimal classifier, achieving a remarkable accuracy rate of 99.81%. In summary, the study highlighted the efficacy of machine learning methods in categorizing network packets and identifying intrusions without the need to decrypt their contents.

Islam et. al. in 2021 [35], proposed a system based on learning to identify and safeguard IoT infrastructures from assaults. The study examined various intrusion detection systems (IDS) that utilize both shallow and deep machine learning models. The methodologies utilized included data analysis approaches, preprocessing of datasets, and the utilization of several machine learning and deep learning algorithms for training sets. The models' performance was assessed using benchmark datasets including NSL-KDD, IoTDevNet, DS2OS, IoTID20, and the IoT Botnet dataset. The performance of these models was evaluated using multiple performance metrics including accuracy, precision, recall, F1-score, Mathew correlation, and Cohen's Kappa coefficient. The findings demonstrated that deep machine learning (IDS) surpassed shallow machine learning in detecting IoT assaults. The Bi-LSTM model demonstrated superior performance compared to the other four deep learning models (DNN, DBN, LSTM, and stacked LSTM) in terms of both train and test accuracy. The SVM model achieved exceptional performance, exhibiting 99.44% accuracy in both training and testing, which is on par with the performance of NSL-KDD, IoTDevNet, and DS2OS. The stacked LSTM model attained a high accuracy of 98.19%, which was similar to the results obtained using a cascaded RNN-based technique.

Xu.in 2021 [36], machine learning approaches were used to detect intrusion traffic, hence enhancing the security of computer networks. The study examined a subset of data from the KDD99 dataset and implemented both supervised and unsupervised learning algorithms. The effectiveness of different classifiers, including as Naive Bayes, decision trees, support vector machines, and logistic regression, was assessed and compared to identify the most efficient method for identifying network intrusions. The investigation determined that the decision tree classifier exhibited superior performance, with a detection accuracy rate of 0.9207 and an F1-score of 0.91. The study also examined the potential of an enhanced K-means clustering algorithm for detecting changes, and demonstrated its superior performance in identifying network intrusions in probing, U2R, and R2L attacks.

Carneiro et. al.in 2021 [37], the performance of two machine learning models, Random Forest (RF) and K-Nearest Neighbors (KNN), trained with two different labels (class and attack type), was compared in the CIDDS-001 dataset for network-based intrusion detection systems. Initially, the dataset underwent a process of cleaning. Subsequently, the RF and KNN models were trained using both sets of labels. Finally, the models were evaluated using metrics such as accuracy, precision, recall, and F1-score. The CIDDS-001 dataset is a collection of network traffic data that includes both regular network activity and various forms of cyber-attacks, including ping scans, port scans, brute force attacks, and denial of service attacks. The study's near-100% accuracy for the class label was likely influenced by overfitting. The RF model achieved an accuracy of 95.60% and an F1-score of 91.34% for the AttackType label, whereas the KNN model achieved an accuracy of 96.94% and an F1-score of 91.61%. The results suggest that the AttackType label shown favorable performance for intrusion detection.

Kumar and Bhatnagar.in 2021 [38], created an advanced intrusion detection system (IDS) with improved capabilities for detecting network threats. In order to accomplish this, the authors suggested a structure for an Intrusion Detection System (IDS) that is implemented on the KDD Cup99 dataset, utilizing machine learning algorithms including Random Forest, Support Vector Machine (SVM), and Naïve Bayes to enhance the precision, accuracy, and recall value of the detection process. Upon conducting a performance analysis of each method, it was concluded that Random Forest exhibited the highest suitability, with an accuracy of 99.99% and a detection rate of 0.999. The dataset underwent preprocessing using techniques of component analysis and was subsequently divided into separate training and testing datasets. The study determined that Random Forest had the highest precision and detection rate out of all the classifiers that were proposed.

Amanoul et. al. in 2021 [39], assessed the efficacy of different machine learning (ML) and deep learning (DL) algorithms for intrusion detection systems (IDS) by analyzing the KDD Cup 99 dataset. The study utilized machine learning (ML) techniques such as Bayes Net and Random Forest, as well as deep learning

e-ISSN: 2614-8404
p-ISSN : 2776-3234

(DL) algorithms like Neural Network, RNN, and LSTM. The ML algorithms exhibited accuracy levels ranging from 98.7869% to 99.9824%, with Random Forest attaining the best accuracy. Conversely, the DL algorithms demonstrated decreased accuracy, with LSTM surpassing RNN. In summary, the study concluded that the Random Forest algorithm demonstrated the highest level of accuracy when applied to Intrusion Detection Systems (IDS) using the KDD Cup 99 dataset.

Krishnaveni et.al.in 2021 [40], created a very effective intrusion detection system for the cloud environment by employing ensemble-based feature selection and classification approaches. The study employed real-time honeypot datasets, feature selection approaches, and ensemble classifiers to attain optimal accuracy and minimize false alarms in detecting network intrusions. The Univariate Ensemble Feature Selection (UEFFS) method was utilized on three intrusion datasets (Honeypot, NSL-KDD, and Kyoto) and shown superior accuracy rates compared to other feature selection measures. The study employed precision-recall analysis and ROC-AUC analysis to evaluate the efficacy of the suggested strategy in enhancing the accuracy and reliability of intrusion detection systems.

Pise.in 2021 [41], utilized machine learning techniques to detect intrusions by applying them to the KDD99 benchmark dataset. Additionally, I assessed the effectiveness of various classifiers in this task. The tactics employed encompassed feature selection procedures to diminish the quantity of characteristics, alongside the utilization of machine learning algorithms such as ZeroR, J48, Naive Bayes, and Random Forest. The study obtained a precision rate of 99.92% for the Random Forest algorithm and 99.91% for the J48 algorithm when applied to the KDD99 dataset. The findings indicated that tree-based classifiers such as J48 and ensemble approaches like Random Forest demonstrated superior performance, with Random Forest achieving the highest level of accuracy. Furthermore, the study emphasized the significance of feature selection in enhancing the effectiveness of the intrusion detection system.

Aziz and Abdulazeez.in 2021 [42], proposed doing a comparative examination of several Machine Learning (ML) approaches employed in Intrusion Detection Systems (IDS) with the aim of identifying intrusions. The approaches the study prioritized were Support Vector Machine (SVM), J48, and Naive Bayes. The study utilized the KDD CUP 99 dataset and the WEKA tool. The study evaluated the algorithms using several performance metrics. The results revealed that J48 achieved the best accuracy rate of 99.96%, closely followed by SVM with a rate of 99.89%. On the other hand, Naive Bayes had the lowest accuracy rate. The experts have determined that no single learning machine algorithm can successfully handle all forms of attacks with precision. In addition, they emphasized the need for varied strategies to be employed in response to various types of attacks.

Azizan et. al.in 2021 [43], proposed a machine learning-based model for a network intrusion detection system (NIDS) and evaluated the performance of three machine learning methods (decision jungle, random forest, and support vector machine) in detecting anomalous network traffic. The study examined the efficacy of the knowledge discovery in databases (KDD) approach with the intrusion detection assessment dataset (CIC-IDS2017). The mean accuracy findings indicated that the support vector machine (SVM) attained the maximum accuracy of 98.18%, followed by random forest (RF) with 96.76% and decision jungle (DJ) with 96.50%. Similarly, the average precision findings showed that the Support Vector Machine (SVM) had the highest precision rate of 98.74%, followed by Random Forest (RF) at 97.96% and Decision Tree (DJ) at 97.82%. The study determined that the Support Vector Machine (SVM) algorithm exhibited the highest efficacy in identifying intrusions within the system.

Ahmed et. al.in 2022 [44], established a Network Intrusion Detection System (NIDS) utilizing machine learning methods for the purpose of identifying network intrusions. The study utilized the UNSW-NB15 dataset, which comprised a substantial volume of network traffic data and encompassed nine distinct categories of network attacks. The study utilized a range of pre-processing approaches, feature selection strategies, and class balance methods. A total of five classification models were employed, namely Random Forests, Decision Trees, Logistic Regression, K-Nearest Neighbors, and Artificial Neural Networks. The Random Forest algorithm attained the maximum accuracy rate of 89.29%. Subsequently, by implementing the SMOTE methodology, the Random Forest classifier demonstrated an accuracy of 95.1%, utilizing 24 selected features obtained from the Principal Component Analysis method.

Mekala et. al.in 2022 [45] ,provided a machine learning-based network intrusion detection solution specifically designed for virtualized data. The techniques utilized encompassed pre-processing, feature selection, feature reduction, and classification utilizing support vector machines (SVM) and Naïve Bayes algorithms. The classifiers were trained and tested using the NSL-KDD dataset. The study obtained a precision rate of 98.2% for Support Vector Machines (SVM), 64.7% for Naïve Bayes, and 53.3% for random tree classifiers. The results unequivocally showcased the efficacy of the suggested methodology in precisely identifying intrusions in virtualized systems.

Singh et. al.in 2022 [46], a powerful Intrusion Detection System (IDS) was created utilizing machine learning methods to identify rare cyber-attacks in network data. The study employed the CIC-IDS 2017 dataset and implemented supervised machine learning classifiers including Random Forest, Decision Tree, Extra Tree, and K-Nearest Neighbor. The model

e-ISSN: 2614-8404
p-ISSN : 2776-3234

attained an average accuracy of 99% and a recall of 100% for all four classifiers. The findings indicated that the Random Forest classifier surpassed the other classifiers, with an impressive accuracy of 99.61% while maintaining a false positive rate of 0.0%. The study sought to tackle the difficulty of identifying sophisticated cyber threats and showcased the efficacy of the suggested machine learning-based Intrusion Detection System (IDS) in precisely categorizing rare attacks in network data.

Chishakwe et. al.in 2022 [47], created an intrusion detection system (IDS) specifically designed for Internet of Things (IoT) scenarios by utilizing advanced machine learning algorithms. The techniques utilized involved creating a simulated IoT environment using an IoT testbed, detecting abnormalities, classifying attacks, and generating notifications upon the detection of intrusions. The UNSW-NB15 dataset was utilized for the purpose of training and assessing machine learning models. Among these models, the Random Forest classifier demonstrated the highest level of accuracy, reaching a score of 87%. The study effectively created a web-based application that can detect unauthorized access in an Internet of Things (IoT) network. It utilizes the Random Forest classifier to recognize abnormal activities and promptly inform users.

Yilmaz.in 2022 [48], suggested an enhanced approach utilizing machine learning to identify unauthorized access in computer networks. The proposed approach consisted of four distinct stages: preprocessing, feature selection, parameter optimization, and classification. The Correlation-Based Feature Selection technique was utilized to identify relevant features. Particle swarm optimization was employed to optimize the parameters. Four machine learning methods, namely Random Tree, AdaBoost, K-Nearest Neighbor, and Support Vector Machine, were employed for the purpose of classification. The suggested methodology underwent testing on two datasets: NSL-KDD and CIC-DDoS2019. The experimental results demonstrated that the suggested method exhibited a high detection rate and surpassed existing machine learning techniques in classifying intrusions, achieving a detection rate of over 99% for all classifiers. These findings indicate the method's potential for practical applications.

Tahri et. al.in 2022 [49], presented an Intrusion Detection System (IDS) that employs the machine learning algorithms of Naive Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbors (KNN) to identify harmful network traffic. The model's performance was evaluated by conducting experiments on two datasets, namely UNSWNB15 and NSL-KDD. The algorithm that performed the best out of the three was chosen for the second stage of processing the database, resulting in the most efficient algorithm. According to the study, SVM demonstrated superior performance, irrespective of the dataset's attack size or kind. The SVM achieved accuracy rates of 97.78% and 97.29% on the UNSWNB15 and NSL-KDD datasets, respectively.

Rajput .et al.in 2022[50], analyzed the efficacy of machine learning methods in detecting network attacks by analyzing the KDD Cup99 dataset. The procedures encompassed preprocessing the dataset, training and testing the machine learning models, and assessing their performance based on accuracy, F1-score, and cross-entropy loss. The study revealed that the Random Forest method attained a remarkable accuracy rate of 100%. Additionally, Decision Trees, Support Vector Machine, Linear Regression, Gradient Boosting, and Deep Neural Networks also shown commendable accuracy scores. The findings demonstrated that machine learning algorithms have the capability to precisely categorize both benign and malevolent network data, rendering them highly efficient for detecting unauthorized access in the field of cybersecurity.

Chua and Salam.in 2022 [51], assessed the enduring effectiveness of intrusion detection systems (IDS) based on machine learning by utilizing distinct datasets for training and testing, replicating real-life situations. The study included six machine learning models: decision tree (DT), random forest (RF), support vector machine (SVM), naïve bayes (NB), artificial neural network (ANN), and deep neural network (DNN). The utilized datasets included the CIC dataset and the LUFlow dataset. The models' accuracy was tested using metrics such as accuracy, precision, recall, and F1-score. The findings indicated that Artificial Neural Network (ANN) exhibited superior performance in the long run, whereas Decision Tree (DT) was found to be more appropriate for firms that are less frequently targeted by attacks. The study also emphasized the significance of regularly updating Intrusion Detection Systems (IDS) with more recent data in order to uphold accuracy.

Mehmood .et al.in 2022 [52], presented a novel hybrid approach for intrusion detection and attack classification that effectively tackles the issue of high false positives and low false negatives in intrusion detection. In order to accomplish this goal, the study employed the NSL-KDD dataset and implemented data transformation, feature selection, and classification algorithms including FGSVM and ANFIS. The suggested technique attained a binary class classification accuracy of 99.3% and Mean Square Error (MSE) values of 0.084964 for training data, 0.0855203 for testing data, and 0.084964 for validation in multiclass categorization. In summary, the study conclusively showed that the hybrid approach is highly effective in precisely identifying and classifying network intrusions.

Ahmad .et al.in 2022 [53], created a highly effective network intrusion detection system utilizing the UNSW-NB15 dataset. The approach utilized

e-ISSN: 2614-8404
p-ISSN : 2776-3234

feature selection through the utilization of a correlation matrix and a decision tree classifier based on AdaBoost. The dataset comprised 49 input variables, and the suggested system attained a remarkable accuracy of 99.3% in categorizing regular network traffic and network hazards. The results indicated that the suggested system surpassed other existing methods, highlighting its efficacy in network security applications and research sectors.

Chua and Salam.in 2023 [54], conducted experiments on six machine learning models to evaluate their effectiveness in detecting intrusions. The experiments utilized a dataset that was created specifically for testing purposes, separate from the dataset used for training. The purpose of this was to facilitate a comparative analysis of the long-term performance of these models and to effectively demonstrate the variations in attack types and network infrastructure over time. The six models assessed were decision tree, random forest, support vector machine, naïve Bayes, artificial neural network, and deep neural network. The evaluation of the study was conducted using three datasets: CIC-IDS2017, CSE-CIC-IDS2018, and LUFlow. The trials demonstrated that Support Vector Machines (SVM) and Artificial Neural Networks (ANN) had the highest resistance to overfitting, whilst Decision Trees (DT) and Random Forests (RF) experienced the greatest susceptibility. Nevertheless, all models exhibited satisfactory performance when the disparity between the training and testing datasets was minimal. The precision of all models varied between 93% and 100%, with the exception of the UNSW-NB15 dataset. The study determined that the suggested approach for evaluating intrusion detection systems based on machine learning, utilizing a progressive dataset, may more effectively evaluate their performance over a lengthy period of time.

ANAND et. al.in 2023[55], suggested an Intrusion Detection System (IDS) that employs eBPF and machine learning algorithms to identify Denial of Service (DoS) and Distributed DoS (DDoS) threats. The study employed the CIC-IDS-2017 dataset and conducted preprocessing, feature extraction using the ANOVA F-test, and analysis utilizing machine learning techniques such as Decision Tree, Random Forest, Support Vector Machine (SVM), and TwinSVM. The experimental results demonstrated that the machine learning algorithms suggested in this study surpassed the performance of previous relevant work. The Decision Tree algorithm achieved an accuracy of 99.38%, the Random Forest algorithm achieved an accuracy of 99.44%, the SVM algorithm achieved an accuracy of 88.74%, and the TwinSVM algorithm achieved an accuracy of 93.82%. Upon evaluation, it was determined that the eBPF implementation exhibited superior performance compared to the userspace implementation, achieving a greater packet processing rate per second.The beginning or uppermost part

Bacevicius and Paulauskaite-Taraseviciene.in 2023 [56], assessed the efficacy of machine learning models in categorizing network intrusions by utilizing imbalanced raw data from the CIC-IDS2017 and CSE-CIC-IDS2018 datasets. A range of machine learning models, such as Logistic Regression, Random Forest, Decision Trees, CNNs, and Artificial Neural Networks, were utilized. The findings revealed that decision trees implemented with the CART algorithm had superior performance, attaining an average macro F1-score of 0.96878. The study also examined the potential of explainable AI (XAI) techniques such as LIME and SHAP to interpret the results and identify the significant elements of the dataset that greatly influence the classification outcomes.

Paricherla et. al.in 2023 [57], created a machine learning framework to precisely classify and detect intrusions in computer networks. The approach utilized a hybridization of ant colony optimization (ACO) and the firefly algorithm for feature selection, in conjunction with machine learning algorithms including AdaBoost, gradient boost, and Bayesian networks for classification. The study included three datasets: NSL-KDD, UNSW-NB15, and CICIDS 2017. When the ACO (Ant Colony Optimization) algorithm and the firefly method for feature selection were combined, the experimental results demonstrated a significant increase in classification accuracy. The gradient boost method demonstrated superior performance in detecting and categorizing intrusions. The performance of the classification techniques was evaluated using accuracy, precision, recall, and F1 score. The findings showed that the accuracy and precision were high.

Somashekar and Boraiah.in 2023 [58], created a fusion model at the prediction level to identify and classify intrusions using machine learning techniques. The main focus was on improving the performance of the intrusion detection system (IDS) by retraining the model to handle unexpected threats. The researchers employed machine learning techniques, including tree ensemble, gradient-boosted tree, and random forest, to perform experiments on the NSL-KDD dataset. The classification accuracy varied from 90.03% for a basic model to 96.31% for the combined and retrained models, demonstrating a notable enhancement in IDS performance. The proposed model showcased the capability of integrating machine learning techniques with a fusion model to enhance the accuracy and security of IDS.

Alotaibi.in 2023 [59], developed an advanced model that utilizes integrated machine learning approaches to identify and prevent early-stage network intrusions, safeguarding networks against malicious attacks. The methodology consisted of two phases: training and validation. In both phases, a fused machine learning model was employed for intrusion detection, utilizing both Naive Bayes and SVM algorithms. The

e-ISSN: 2614-8404
p-ISSN : 2776-3234

simulation results demonstrated the efficacy of the suggested intrusion detection model, achieving an accuracy of 0.909 and a miss rate of 0.091 in identifying early-stage network intrusions. The study utilized a dataset sourced from the UCI Machine Learning Data Repository, which was partitioned into training and validation sets at a ratio of 7:3. The system's performance was evaluated using statistical measures of precision, sensitivity, specificity, and accuracy.

Abeshek et. al. in 2023 [60], explored the efficacy and feasibility of utilizing machine learning methods for detecting network intrusions. The dataset underwent preprocessing to guarantee its integrity and appropriateness. The applicability and efficacy of three distinct machine learning models, specifically the XGBoost classifier, the Extra Trees classifier, and the Artificial Neural Network (ANN), were assessed for their application in network intrusion detection classification systems. The XGBoost Classifier demonstrated good performance in spotting abnormalities, achieving an accuracy of 0.988, precision of 0.982, recall of 0.995, and an F1-score of 0.989. Both the XGBoost classifier and the Extra Trees classifier are suitable options for network intrusion detection, as they yield comparable outcomes according to the comparison analysis.

Güney.in 2023 [61], conducted an analysis to evaluate the effects of various data preprocessing methods, such as normalization, feature selection, and classifier optimization, on the classification accuracy of support vector machines (SVM) for intrusion detection datasets. The performance of various approaches was evaluated using three benchmark datasets: NSL-KDD, UNSW-NB15, and CICIDS2017. The log-scaling normalization technique was determined to be the most effective method. By employing SVM with feature selection and classifier optimization, accuracy rates of 81.51% and 85.27% were achieved for the NSLKDD and UNSW-NB15 datasets, respectively, using 2 and 32 features. Similarly, an accuracy of 99.43% was attained for the CICIDS2017 dataset using 16 features. This work offered valuable insights into the process of data preprocessing in machine learning (ML) applications and demonstrated the critical importance of data pretreatment in constructing IDSs (Intrusion Detection Systems) that are both precise and efficient.

Sulaiman and Abdulazeez.in 2024 [62], explored the utilization of machine learning techniques for identifying anomalies and detecting misuse in intrusion detection systems. Additionally, investigated the effectiveness of ensemble learning models including AdaBoost, LightGBM, and XGBoost. The study employed the KDD Cup 99 dataset as a standard to evaluate and contrast the efficacy of various models, with a specific emphasis on detecting smurf attacks. According to the study, XGBoost demonstrated superior performance compared to the other two models in terms of accuracy. XGBoost achieved an accuracy of 0.99985, whereas AdaBoost earned an accuracy of 0.99076 and LightGBM achieved an accuracy of 0.99925. The study determined that the combination of machine learning approaches and a thorough comprehension of cybersecurity threats is crucial for developing efficient and robust intrusion detection systems.

Table 1Literature Review

Ref.	***Years***	***Dataset***	***Technique***	***Classifier***	***Accuracy***
[29]	2020	NSL-KDD	Various ML	SVM, KNN, LR, NB, MLP, RF, ETC, DT	RF, ETC, DT: 99%
[30]	2020	Sensor480	ML	SVM, NB, DT, Adaboost	DT: 100%
[31]	2020	UNSW-NB15, CICIDS-2017, ICS cyber-attack	ML & DL	RF	UNSW-NB15: 88.5% (binary), 73.6% (multi-class); CICIDS-2017: 99.9% (binary), 99.9% (multi-class)
[32]	2020	CICIDS 2017, UNSW-NB 2015	ML	BO-TPE-RF (optimized RF)	>99% (both datasets)
[33]	2020	UNSW-NB15	ML	NB, C4.5	C4.5: 90.44%
[34]	2020	Custom generated dataset	ML	NB, SVM, RF, KNN	RF: 99.81%

e-ISSN: 2614-8404
p-ISSN : 2776-3234

Ref.	*Years*	*Dataset*	*Technique*	*Classifier*	*Accuracy*
[35]	2021	NSL-KDD, IoTDevNet, DS2OS, IoTID20, IoT Botnet	ML & DL	Bi-LSTM, SVM	SVM: 99.44%
[36]	2021	KDD99	ML	Decision trees	92.07%
[37]	2021	CIDDS-001	ML	RF, KNN	RF: 95.60% (AttackType)
[38]	2021	KDD Cup99	ML	Random Forest	99.99%
[39]	2021	KDD Cup 99	ML & DL	Random Forest	99.9824%
[40]	2021	Honeypot, NSL-KDD, Kyoto	ML	UEFFS	Superior to other feature selection methods
[41]	2021	KDD99	ML	J48, Random Forest	Random Forest: 99.92%
[42]	2021	KDD CUP 99	ML	J48, SVM, NB	J48: 99.96%
[43]	2021	CIC-IDS2017	ML	SVM, RF, DJ	SVM: 98.18%
[44]	2022	UNSW-NB15	ML	Random Forest	SMOTE-RF: 95.1%
[45]	2022	NSL-KDD	ML	SVM, Naïve Bayes, Random Tree	SVM: 98.2%
[46]	2022	CIC-IDS 2017	ML	Random Forest	99%
[47]	2022	UNSW-NB15	ML	Random Forest	87%
[48]	2022	NSL-KDD, CIC-DDoS2019	Feature selection, Parameter optimization, Classification	Random Tree, AdaBoost, K-Nearest Neighbor, Support Vector Machine	>99% for all classifiers
[49]	2022	UNSWNB15, NSL-KDD	Intrusion Detection System with Naive Bayes, Support Vector Machine, K-Nearest Neighbors	Support Vector Machine	97.78% (UNSWNB15), 97.29% (NSL-KDD)
[50]	2022	KDD Cup99	Preprocessing, Training, Testing, Various ML models	Random Forest, Decision Trees, SVM, Linear Regression, Gradient Boost, Deep Neural Networks	100% (Random Forest)
[51]	2022	CIC dataset, LUFlow	Comparison of 6 ML models for IDS	ANN	Varies, but ANN showed superior long-term performance
[52]	2022	NSL-KDD	Hybrid approach with FGSVM and ANFIS	FGSVM, ANFIS	99.3% (binary class), 0.084964 (MSE)
[53]	2022	UNSW-NB15	Feature selection, Decision Tree Classifier based on AdaBoost	Decision Tree (AdaBoost)	99.3%

e-ISSN: 2614-8404
p-ISSN : 2776-3234

Ref.	*Years*	*Dataset*	*Technique*	*Classifier*	*Accuracy*
[54]	2023	CIC-IDS2017, CSE-CIC-IDS2018, LUFlow	Evaluation of 6 ML models over time	SVM, ANN	93% to 100% precision
[55]	2023	CIC-IDS-2017	IDS using eBPF and ML algorithms	Decision Tree, Random Forest, SVM, TwinSVM	88.74% to 99.44%
[56]	2023	CIC-IDS2017, CSE-CIC-IDS2018	ML models for imbalanced data classification	Decision Trees (CART)	Avg. macro F1-score: 0.96878
[57]	2023	NSL-KDD, UNSW-NB15, CICIDS 2017	ACO, Firefly Algorithm, Machine Learning Algorithms	Gradient Boost	High accuracy and precision
[58]	2023	NSL-KDD	Fusion model for IDS with ML techniques	Tree ensemble, Gradient-boosted tree, Random forest	90.03% to 96.31%
[59]	2023	UCI Machine Learning Data Repository	Fused ML model for intrusion detection using Naive Bayes and SVM	Naive Bayes, SVM	0.909
[60]	2023	-	XGBoost, Extra Trees, ANN for network intrusion detection	XGBoost, Extra Trees, ANN	0.988
[61]	2023	NSL-KDD, UNSW-NB15, CICIDS2017	SVM with preprocessing methods	SVM	81.51% to 99.43%
[62]	2024	KDD Cup 99	Ensemble learning models for IDS	AdaBoost, LightGBM, XGBoost	Accuracy: 0.99985 (XGBoost)

IV. COMPARISON AND DISCUSSION

The review paper thoroughly analyzes the effectiveness of different machine learning classification methods in network intrusion detection across multiple investigations. [29] Demonstrated the efficacy of machine learning classifiers such as Random Forest (RF), Extra Trees Classifier (ETC), and Decision Trees (DT) in detecting network intrusions. Their proposed method had an impressive 99% accuracy rate, highlighting its effectiveness. [30] Conducted a study on an intrusion detection system specifically designed for IoT. They achieved accuracy levels ranging from 97.89% to 100%, with the Decision Tree model performing better than other models. [31] Evaluated twelve machine learning algorithms and emphasized the better performance of Random Forest in detecting deviant behaviors with an accuracy of up to 99.9%. [32] Proposed a multi-stage optimized architecture that utilized the BO-TPE-RF optimized random forest classifier to achieve a detection rate of above 99%. The results indicate that machine learning techniques, particularly Random Forest, are highly effective in predicting and preventing network intrusions in various datasets and scenarios.

V. CONCLUSION

The literature analysis on network intrusion detection systems (NIDS) utilizing machine learning classification algorithms uncovers a wide range of methodology and approaches designed to accurately detect and thwart network invasions. The research have used different machine learning classifiers, such as Support Vector Machines (SVM), Decision Trees (DT), Random Forest (RF), Naive Bayes (NB), and ensemble approaches, to evaluate how well they can detect aberrant behaviors and classify network attacks. The results routinely show high accuracy rates, frequently over 90%, with certain algorithms obtaining almost flawless accuracy in particular situations. The key findings highlight the superiority of specific classifiers, such as Random Forest, in accurately predicting incursions while limiting computational complexity. Moreover, research emphasizes the importance of preprocessing methods, feature selection approaches, and model tuning in improving the effectiveness of NIDS. In summary, the paper

e-ISSN: 2614-8404
p-ISSN : 2776-3234

highlights the effectiveness of machine learning methods in enhancing network security by offering strong intrusion detection capabilities, thus protecting important network assets from unwanted activity.

REFERENCES

[1] Q.-V. Dang, "Active learning for intrusion detection systems," presented at the 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), IEEE, 2020, pp. 1–3.

[2] R. Singh, M. Kalra, and S. Solanki, "A hybrid approach for intrusion detection based on machine learning," Int. J. Secur. Netw., vol. 15, no. 4, pp. 233–242, 2020.

[3] J. Lee, J. Kim, I. Kim, and K. Han, "Cyber threat detection based on artificial neural networks using event profiles," Ieee Access, vol. 7, pp. 165607–165626, 2019.

[4] A. Abdulazeez, B. Salim, D. Zeebaree, and D. Doghramachi, "Comparison of VPN Protocols at Network Layer Focusing on Wire Guard Protocol," 2020.

[5] C. J. Ugochukwu, E. Bennett, and P. Harcourt, An intrusion detection system using machine learning algorithm. LAP LAMBERT Academic Publishing, 2019.

[6] A. A. Salih and M. B. Abdulrazaq, "Combining best features selection using three classifiers in intrusion detection system," presented at the 2019 International Conference on Advanced Science and Engineering (ICOASE), IEEE, 2019, pp. 94–99.

[7] W. A. H. Ghanem, A. Jantan, S. A. A. Ghaleb, and A. B. Nasser, "An efficient intrusion detection model based on hybridization of artificial bee colony and dragonfly algorithms for training multilayer perceptrons," IEEE Access, vol. 8, pp. 130452–130475, 2020.

[8] T. A. Alamiedy, M. Anbar, Z. N. Alqattan, and Q. M. Alzubi, "Anomaly-based intrusion detection system using multi-objective grey wolf optimisation algorithm," J. Ambient Intell. Humaniz. Comput., vol. 11, no. 9, pp. 3735–3756, 2020.

[9] A. Bhumgara and A. Pitale, "Detection of network intrusions using hybrid intelligent systems," presented at the 2019 1st International Conference on Advances in Information Technology (ICAIT), IEEE, 2019, pp. 500–506.

[10] A. Rai, "Optimizing a new intrusion detection system using ensemble methods and deep neural network," presented at the 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), IEEE, 2020, pp. 527–532.

[11] A. H. Mirza, "Computer network intrusion detection using various classifiers and ensemble learning," presented at the 2018 26th Signal processing and communications applications conference (SIU), IEEE, 2018, pp. 1–4.

[12] S. Ahmad, F. Arif, Z. Zabeehullah, and N. Iltaf, "Novel approach using deep learning for intrusion detection and classification of the network traffic," presented at the 2020 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA), IEEE, 2020, pp. 1–6.

[13] D. A. Hasan and A. M. Abdulazeez, "A modified convolutional neural networks model for medical image segmentation," learning, vol. 20, p. 22, 2020.

[14] A. Phadke, M. Kulkarni, P. Bhawalkar, and R. Bhattad, "A review of machine learning methodologies for network intrusion detection," presented at the 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 2019, pp. 272–275.

[15] A. Golrang, A. M. Golrang, S. Yildirim Yayilgan, and O. Elezaj, "A novel hybrid IDS based on modified NSGAII-ANN and random forest," electronics, vol. 9, no. 4, p. 577, 2020.

[16] K. Shashank and M. Balachandra, "Review on network intrusion detection techniques using machine learning," presented at the 2018 IEEE Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), IEEE, 2018, pp. 104–109.

[17] F. Yihunie, E. Abdelfattah, and A. Regmi, "Applying machine learning to anomaly-based intrusion detection systems," presented at the 2019 IEEE Long Island Systems, Applications and Technology Conference (LISAT), IEEE, 2019, pp. 1–5.

[18] A. Ghosal and S. Halder, "A survey on energy efficient intrusion detection in wireless sensor networks," J. Ambient Intell. Smart Environ., vol. 9, no. 2, pp. 239–261, 2017.

[19] C. Kalimuthan and J. A. Renjit, "Review on intrusion detection using feature selection with machine learning techniques," Mater. Today Proc., vol. 33, pp. 3794–3802, 2020.

[20] A. Phadke, M. Kulkarni, P. Bhawalkar, and R. Bhattad, "A review of machine learning methodologies for network intrusion detection," presented at the 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 2019, pp. 272–275.

[21] C. Kalimuthan and J. A. Renjit, "Review on intrusion detection using feature selection with machine learning techniques," Mater. Today Proc., vol. 33, pp. 3794–3802, 2020.

[22] M. Almseidin, M. Alzubi, S. Kovacs, and M. Alkasassbeh, "Evaluation of machine learning algorithms for intrusion detection system," presented at the 2017 IEEE 15th

e-ISSN: 2614-8404
p-ISSN : 2776-3234

international symposium on intelligent systems and informatics (SISY), IEEE, 2017, pp. 000277–000282.

[23] I. S. Thaseen and C. A. Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," J. King Saud Univ.-Comput. Inf. Sci., vol. 29, no. 4, pp. 462–472, 2017.

[24] S. Ganapathy, N. Jaisankar, P. Yogesh, and A. Kannan, "An intelligent intrusion detection system using outlier detection and multiclass SVM," Int. J. Recent Trends Eng. Technol., vol. 5, no. 01, 2011.

[25] B. Gupta, A. Rawat, A. Jain, A. Arora, and N. Dhami, "Analysis of various decision tree algorithms for classification in data mining," Int. J. Comput. Appl., vol. 163, no. 8, pp. 15–19, 2017.

[26] S. Latha and S. J. Prakash, "A survey on network attacks and Intrusion detection systems," presented at the 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), IEEE, 2017, pp. 1–7.

[27] J. Zhang and M. Zulkernine, "Network Intrusion Detection using Random Forests.," presented at the Pst, Citeseer, 2005.

[28] S. Aljawarneh, M. B. Yassein, and M. Aljundi, "An enhanced J48 classification algorithm for the anomaly intrusion detection systems," Clust. Comput., vol. 22, pp. 10549–10565, 2019.

[29] I. Abrar, Z. Ayub, F. Masoodi, and A. M. Bamhdi, "A machine learning approach for intrusion detection system on NSL-KDD dataset," presented at the 2020 international conference on smart electronics and communication (ICOSEC), IEEE, 2020, pp. 919–924.

[30] K. S. Kiran, R. K. Devisetty, N. P. Kalyan, K. Mukundini, and R. Karthi, "Building a intrusion detection system for IoT environment using machine learning techniques," Procedia Comput. Sci., vol. 171, pp. 2372–2379, 2020.

[31] N. Elmrabit, F. Zhou, F. Li, and H. Zhou, "Evaluation of machine learning algorithms for anomaly detection," presented at the 2020 international conference on cyber security and protection of digital services (cyber security), IEEE, 2020, pp. 1–8.

[32] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Multi-stage optimized machine learning framework for network intrusion detection," IEEE Trans. Netw. Serv. Manag., vol. 18, no. 2, pp. 1803–1816, 2020.

[33] O. J. Mebawondu, A. O. Adetunmbi, J. O. Mebawondu, and O. D. Alowolodu, "Feature Weighting and Classification Modeling for Network Intrusion Detection Using Machine Learning Algorithms," presented at the International Conference on Information and Communication Technology and Applications, Springer, 2020, pp. 315–327.

[34] I. S. Thaseen, B. Poorva, and P. S. Ushasree, "Network intrusion detection using machine learning techniques," presented at the 2020 International conference on emerging trends in information technology and engineering (IC-ETITE), IEEE, 2020, pp. 1–7.

[35] N. Islam et al., "Towards Machine Learning Based Intrusion Detection in IoT Networks.," Comput. Mater. Contin., vol. 69, no. 2, 2021.

[36] G. Xu, "Research on network intrusion detection method based on machine learning," presented at the Journal of Physics: Conference Series, IOP Publishing, 2021, p. 012034.

[37] J. Carneiro, N. Oliveira, N. Sousa, E. Maia, and I. Praça, "Machine learning for network-based intrusion detection systems: an analysis of the CIDDS-001 dataset," presented at the International Symposium on Distributed Computing and Artificial Intelligence, Springer, 2021, pp. 148–158.

[38] K. Kumar and V. Bhatnagar, "Machine Learning Algorithms Performance Evaluation for Intrusion Detection," J. Inf. Technol. Manag., vol. 13, no. 1, pp. 42–61, 2021.

[39] S. V. Amanoul, A. M. Abdulazeez, D. Q. Zeebare, and F. Y. Ahmed, "Intrusion detection systems based on machine learning algorithms," presented at the 2021 IEEE international conference on automatic control & intelligent systems (I2CACIS), IEEE, 2021, pp. 282–287.

[40] S. Krishnaveni, S. Sivamohan, S. Sridhar, and S. Prabakaran, "Efficient feature selection and classification through ensemble method for network intrusion detection on cloud computing," Clust. Comput., vol. 24, no. 3, pp. 1761–1779, 2021.

[41] N. Pise, "Application of machine learning for intrusion detection system," Inf. Technol. Ind., vol. 9, no. 1, pp. 314–323, 2021.

[42] Z. A. Aziz and A. M. Abdulazeez, "Application of Machine Learning Approaches in Intrusion Detection System," J. Soft Comput. Data Min., vol. 2, no. 2, pp. 1–13, 2021.

[43] A. H. Azizan et al., "A machine learning approach for improving the performance of network intrusion detection systems," Ann. Emerg. Technol. Comput. AETiC, vol. 5, no. 5, pp. 201–208, 2021.

[44] H. A. Ahmed, A. Hameed, and N. Z. Bawany, "Network intrusion detection using oversampling technique and machine learning algorithms," PeerJ Comput. Sci., vol. 8, p. e820, 2022.

[45] S. Mekala, R. Jatothu, S. Kodati, K. Pradeep Reddy, and N. Sreekanth, "Network Intrusion Detection Using Machine Learning for Virtualized Data," in Innovations in Signal

e-ISSN: 2614-8404
p-ISSN : 2776-3234

Processing and Embedded Systems: Proceedings of ICISPES 2021, Springer, 2022, pp. 235–244.

[46] A. P. Singh, S. Kumar, A. Kumar, and M. Usama, "Machine learning based intrusion detection system for minority attacks classification," presented at the 2022 international conference on computational intelligence and sustainable engineering solutions (CISES), IEEE, 2022, pp. 256–261.

[47] S. Chishakwe, N. Moyo, B. M. Ndlovu, and S. Dube, "Intrusion Detection System for IoT environments using Machine Learning Techniques," presented at the 2022 1st Zimbabwe Conference of Information and Communication Technologies (ZCICT), IEEE, 2022, pp. 1–7.

[48] A. A. Yilmaz, "Intrusion detection in computer networks using optimized machine learning algorithms," presented at the 2022 3rd International Informatics and Software Engineering Conference (IISEC), IEEE, 2022, pp. 1–5.

[49] R. Tahri, Y. Balouki, A. Jarrar, and A. Lasbahani, "Intrusion detection system using machine learning algorithms," presented at the ITM Web of Conferences, EDP Sciences, 2022, p. 02003.

[50] M. A. Rajput, M. Umar, A. Ahmed, A. R. Bhangwar, and K. S. Memon, "Evaluation of Machine Learning based Network Attack Detection," Sukkur IBA J. Emerg. Technol., vol. 5, no. 2, pp. 57–66, 2022.

[51] T.-H. Chua and I. Salam, "Evaluation of machine learning algorithms in network-based intrusion detection system," ArXiv Prepr. ArXiv220305232, 2022.

[52] M. Mehmood et al., "A hybrid approach for network intrusion detection," CMC-Comput Mater Contin, vol. 70, pp. 91–107, 2022.

[53] I. Ahmad, Q. E. Ul Haq, M. Imran, M. O. Alassafi, and R. A. AlGhamdi, "An efficient network intrusion detection and classification system," Mathematics, vol. 10, no. 3, p. 530, 2022.

[54] T.-H. Chua and I. Salam, "Evaluation of Machine Learning Algorithms in Network-Based Intrusion Detection Using Progressive Dataset," Symmetry, vol. 15, no. 6, p. 1251, 2023.

[55] N. ANAND, M. SAIFULLA, and P. K. Aakula, "High-performance Intrusion Detection Systemusing eBPF with Machine Learning algorithms," 2023.

[56] M. Bacevicius and A. Paulauskaite-Taraseviciene, "Machine Learning Algorithms for Raw and Unbalanced Intrusion Detection Data in a Multi-Class Classification Problem," Appl. Sci., vol. 13, no. 12, p. 7328, 2023.

[57] M. Paricherla, M. Ritonga, S. R. Shinde, S. M. Chaudhari, R. Linur, and A. Raghuvanshi, "Machine learning techniques for accurate classification and detection of intrusions in computer network," Bull. Electr. Eng. Inform., vol. 12, no. 4, pp. 2340–2347, 2023.

[58] H. Somashekar and R. Boraiah, "Network intrusion detection and classification using machine learning predictions fusion," Indones. J. Electr. Eng. Comput. Sci., vol. 31, no. 2, pp. 1147–1153, 2023.

[59] F. M. Alotaibi, "Network Intrusion Detection Model Using Fused Machine Learning Technique.," Comput. Mater. Contin., vol. 75, no. 2, 2023.

[60] A. Abeshek, S. Venkatraman, S. Aravintakshan, V. Santhosh, and R. Manoharan, "Network Intrusion Detection Using Machine Learning Approach," EasyChair, 2516–2314, 2023.

[61] H. Güney, "Preprocessing Impact Analysis for Machine Learning-Based Network Intrusion Detection," Sak. Univ. J. Comput. Inf. Sci., vol. 6, no. 1, pp. 67–79, 2023.

[62] S. M. Sulaiman and A. M. Abdulazeez, "Leveraging of Gradient Boosting Algorithm in Misuse Intrusion Detection using KDD Cup 99 Dataset," Indones. J. Comput. Sci., vol. 13, no. 1, 2024.