IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 30, 2022, accepted April 16, 2022, date of publication May 2, 2022, date of current version May 5, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3171660

An Innovative Perceptual Pigeon Galvanized Optimization (PPGO) Based Likelihood Naïve Bayes (LNB) Classification Approach for Network Intrusion Detection System

S. SHITHARTH[1], (Member, IEEE), PRAVIN R. KSHIRSAGAR[2], (Senior Member, IEEE), PRAVEEN KUMAR BALACHANDRAN[3], (Member, IEEE), KHALED H. ALYOUBI[4], AND ALAA O. KHADIDOS[4]

[1]Department of Computer Science and Engineering, Kebri Dehar University, Kebri Dehar, Somali 001, Ethiopia
[2]Department of Artificial Intelligence, G. H. Raisoni College of Engineering, Nagpur, Maharashtra 440016, India
[3]Department of Electrical and Electronics Engineering, Vardhaman College of Engineering, Hyderabad, Telangana 501218, India
[4]Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding author: S. Shitharth (shitharth.s@ieee.org)

ABSTRACT Intrusion detection and classification have gained significant attention recently due to the increased utilization of networks. For this purpose, there are different types of Network Intrusion Detection System (NIDS) approaches developed in the conventional works, which mainly focus on identifying the intrusions from the datasets with the help of classification techniques. Still, it is limited by the significant problems of inefficiency in handling large dimensional datasets, high computational complexity, false detection, and more time consumption for training the models. To solve these problems, this research intends to develop an innovative clustering-based classification methodology to precisely detect intrusions from the different types of IDS datasets. Here, the most recent and extensively used IDS datasets such as NSL-KDD, CICIDS, and Bot-IoT have been employed for detecting intrusions. Data preprocessing has been performed to normalize the dataset to eliminate irrelevant attributes and organize the features. Then, the data separation is applied by forming the clusters by using an intelligent Anticipated Distance-based Clustering (ADC) incorporated with the Density-Based Spatial clustering of applications with noise (DBScan) algorithm. It helps to find the distance and density measures for grouping the attributes into the clusters, which increases the efficiency of classification. Here, the most suitable optimal parameters are selected using the Perpetual Pigeon Galvanized Optimization (PPGO) technique. The extracted features are used for training and testing the dataset samples. Consequently, the Likelihood Naïve Bayes (LNB) classification approach is implemented to accurately predict the classified label as to whether normal or attack. During the evaluation, the performance of the proposed IDS framework is validated and compared using various evaluation metrics. The results show that the proposed ADC-DBScan-LNB model outperforms the other techniques with improved performance outcomes.

INDEX TERMS Network intrusion detection system (NIDS), density-based spatial clustering of applications with noise (DBSCAN), anticipated distance-based clustering (ADC), data preprocessing, Likelihood Naïve Bayes (LNB), and IDS datasets.

I. INTRODUCTION

The internet plays an essential role in our daily part of life, a resource used for learning information in different fields such as education, business, and others. Specifically, most organizations could use the Internet as the technology for accomplishing their management activities [1], [2]. They can utilize the Internet application for profit growth and keep confidential/private information secret. Also, it is used to establish good communication between the customers of

The associate editor coordinating the review of this manuscript and approving it for publication was Jiafeng Xie.

the organization. In addition to that, it supports the organizations to improve the operating efficiency against the network vulnerabilities [3]. In this platform, the data processing and execution are moderately dangerous due to the frequent assaults on the internet.

Hence, it is essential to ensure increased data anonymity and improve public interest in the security [4], [5]. Typically, the attacks are the kind of unwanted/malicious actions performed by the attackers [6] for degrading or affecting the networking system, which is easily detected with the help of anti-attacking systems. To predict harmful attacks, many security approaches are used today. These techniques are employed to detect the attacking activities in the network based on the signature of patterns. In many conventional works, the Network Intrusion Detection System (NIDS) [7] framework has been deployed to detect intrusions based on certain dataset features, such as traffic patterns, the flow of data, and packet information. Still, it faces some difficulties related to the factors of signature dependency, single point of failure, requires an increased amount of time for training the models, and complexity in algorithm design. Hence, various intrusion detection approaches such as clustering, optimization, and classification Field [5], [8], [9] have been utilized in the existing works to identify and classify the intrusions based on their attributes/features.

Typically, the main reason for using clustering approaches Field [10] is to group the attributes in the cluster based on the distance value. Most of the clustering techniques are developed based on the parameters of density and distance, which helps to separate the data. It includes the types [11], [12] of hierarchical clustering, k-means clustering, partition based clustering, spatial clustering, and centroid based clustering. Consequently, the optimization methodologies select the most suitable attributes from the given datasets by computing the optimal fitness value. Generally, the increased number of attributes can degrade the classifier's performance in terms of misclassified labels, high false positives, and more time-consuming training the data samples. Recently, the meta-heuristic optimization techniques [13] have been widely applied to solve classification problems, which identifies the best fitness value based on the weight value. It includes the types [14], [15] of Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Ant Bee Colony (ABC), Whale Optimization (WO), and Firefly (FF) optimization.

Moreover, machine learning and deep learning classification techniques are used to predict the classified labels for the given problems. The classifier's performance highly depends on the optimal number of features used for training the model. There are different types of classification techniques [16] are used for detecting the intrusions from the IDS datasets, which includes [13], [17] Neural Network (NN), Naïve Bayes (NB), Decision Tree (DT), Support Vector Machine (SVM), Relevance Vector Machine (RVM), and Fuzzy Logic (FL). However, the significant limitations of the conventional approaches are as follows: single point of failure, highly dependent on signatures, increased computational cost, high false alarm rate, and difficulty in detection. Hence, this research intends to develop an intelligent IDS using enhanced clustering, optimization, and classification techniques. The main contribution of this paper is to precisely predict the intrusions from the given network IDS datasets by implementing an advanced optimization based classification methodology. In this system, the three different and popular IDS datasets, such as NSL-KDD, CICIDS, and Bot-IoT, have been used to validate the proposed system's performance.

Here, a group of methodologies such as clustering, optimization and classification are utilized for processing the datasets in order, and classification is utilized to process the datasets to predict the intrusions. Typically, data clustering is one of the most suitable techniques used for simplifying the detection process because which helps to group the attributes into the form of a cluster. Here, the distance-based clustering mechanism is utilized for constructing the group of clusters concerning the parameters of cutoff distance, lower and higher density values. Also, the clustering helps to preprocess the datasets for filling the missing values, eliminating the irrelevant and redundant attributes. Specifically, the classifier's performance highly depends on the quality of data and features used for training the model. Hence, it is essential to preprocess the datasets before using them for further processing. After that, a novel Perceptual Pigeon Galvanized Optimization (PPGO) technique is employed for optimally selecting the features from the clustered data. The primary purpose of using this technique is to reduce the classifier's computational complexity and time consumption. Generally, the classification techniques consume more time for training and testing the data attributes to predict the classified label, so the increased amount of features can degrade the classifier's performance with high misclassification outputs, error rate, and increased false positives. Due to these factors, the proposed work objects to utilize an optimization technique for selecting the most suited attributes based on the global best solution, which also helps to obtain increased detection accuracy. Finally, the set optimal numbers of features are fed to the classifier for training the models to predict whether the data is normal or intrusion. For this purpose, an enhanced Likelihood Naïve Bayes (LNB) based machine learning classification approach is utilized in this system, accurately identifying the intrusions from the given datasets with increased accuracy and reduced complexity.

- Algorithm I – ADC with DBScan Clustering.
- Algorithm II – Perpetual Pigeon Galvanized Optimization (PPGO)
- Algorithm III – Likelihood Naïve Bayes (LNB) Classification

The primary objectives behind this work are as follows:

- Anticipated Distance-based Clustering (ADC) incorporated with the Density-Based Spatial clustering of applications with noise DBScan) technique is deployed to form the cluster for separating the dataset.

- An efficient Likelihood Naïve Bayes (LNB) classification technique is employed to improve the detection accuracy and predicted outcomes.
- The Perpetual Pigeon Galvanized Optimization to select the optimal feature such as energy, probability based on density, likelihood, trust, energy, weight function, IP, traffic stamp, port, and average total number of packets (PPGO) technique is utilized.
- The most widely used IDS datasets such as NSL-KDD, CICIDS, and Bot-IoT are utilized to test this system.
- To evaluate the performance of this system, there are various measures such as accuracy, precision, FPR, TPR, F1-score, and similarity coefficients are estimated.

The remaining portions of this paper are structuralized as follows: Section II reviews the existing NIDS approaches used to improve networking systems' security, where each technique's advantages and disadvantages are discussed based on its key features and operating functions. The overall description of the proposed methodology is represented with its clear flow and algorithmic illustrations in Section III. The performance and comparative analysis of both existing and proposed IDS are validated and compared using different datasets in Section IV. Finally, the overall paper is summarized with its obtainment and future scope in Section V.

II. RELATED WORKS

This section reviews some of the conventional works related to the NIDS with its benefits and limitations. The machine learning classification techniques, including supervised and unsupervised learning models, are highly used in many security applications. Which some of the recent methods are studied as shown below:

- *Decision Tree:* In machine learning, a classification tree is also known as a prediction model or decision tree [18]. It is a graph in a tree-like manner with internal nodes that indicates the test properties, branches, and terminal nodes or leaves that reflect the class that belongs to any object. The classification tree algorithms ID3 and C4.5 are the fundamental and extensively utilized. The two techniques in the building of the tree are top-down tree structure and bottom-up pruning. ID3 and C4.5 reflect the top-down tree structure. More techniques to classify trees were found to be more exact than categorizing ship bays than decision-making trees.
- *Fuzzy Logic:* It is based on fuzzy set theories [19], which deal with reasoning that is not precisely inferred but approaches traditional predicate logic. The fuzzy set theory covers well-thought-out real-world expert values for a complex topic. The information in this approach is classified based on several statistical measures. These data parts are utilized to classify them as usual or as malicious with logical standards that have been broken. Several intrusion data extraction techniques expound specific changes in current data mining algorithms to enhance efficiency and precise extraction of intrusion detection patterns.
- *Naïve Bayes Network:* There are various instances in which there are statistical dependencies or causal interactions among system variables. The probabilistic interactions between these variables can be difficult to define accurately. In other words, the system's previous knowledge is just that others could change some variable. A probabilistic graphical model known as the Naive Bayesian Networks (NB) Field [20] might be used to take advantage of this structural link between the problem's random variables. This model responds to questions such as, ''What is the chance of a certain sort of assault if a few observed incidents are presented?'' The conditional probability formula might be utilized. The NB structure is often represented as a DAG, with each node representing one system variable and each link coding the influence of one node over another. A DAG usually represents the NB. While the accuracy of the decision tree is much superior, the computational time of the Bayesian network is minimal when comparing the decision tree and Bayesian techniques. Therefore, it is efficient to utilize NB models if the data set is huge.
- *Genetic Algorithm:* In the subject of computational biology, it was first introduced. These algorithms are part of the broader evolutionary algorithm class (EA). Convolution techniques, such as heritage, selection, mutation, and crossing, find answers to optimization issues. In several fields they have been utilized with highly promising results. For intrusion detection, the genetic algorithm (GA) [21] produces the audit data from a set of categorization rules. The support and trust framework is utilized as a fitness function to determine the quality of each rule. Significant features include GA's noise and self-learning strength. The advantages of GA methods include high attack detection rates and few false positives.
- *Neural Networks:* It is a network of linked nodes that replicate the brain's functioning. Each node has an extensive connection to several other nodes in neighboring layers. Individual nodes can use the weights and a simple function [22] to determine the output values from the linked nodes. Neural networks can be established for supervised or unsupervised learning. The user must specify the number of hidden levels and the number of nodes in a hidden layer. The neural network output layer may include one or more nodes depending on the application. The neural networks [23] of the Multilayer Perceptions (MLP) have been successful in many applications and have produced more precise results. They can approximate any continuous function to the random precision provided they contain sufficient hidden units. Therefore, such models potentially establish every decision boundary for classification within the feature space and behave as a non-linear discriminatory function.
- *Support Vector Machine:* Various supervised learning approaches for classification and regression. The Support Vector Machine (SVM) Field [24] is frequently employed in the pattern recognition business. The invasions are also recognized. The SVM one class is based on several instances of

a particular class and does not employ adverse and favorable examples. SVM exceeded NN in terms of false alarm rates and accuracy in most types of assaults compared to neural networks in the KDD cup data set.

A novel tool for detecting regularities and irregularities in big datasets has already been implemented in the network security environment. Hybrid learning approaches can attain the greatest possible accuracy and detection rate. A combination of clustering and classification techniques might be employed to develop a hybrid training method. Clustering is an anomaly-based approach to detection, able without prealerting to identify novel threats and recognize natural data groups based on common patterns. It might also be used to quickly uncover a set of similar traffic behaviors using cluster analyses, such as k-means and Db Scan. The classification of Naive Bayes is more efficient and can produce very competitive results in anomaly-based network infiltration because of its fundamental structure.

In normal and consistent distribution circumstances, the effectiveness of k-means and k-medoids clustering algorithms was examined using vast data sets. The average time taken for k-means is higher than the average time used for the two situations by k-medoids. A methodology was created and suggested by [25] that include a three-story rating of a decision tree for boosting the detection rate. This approach detects known assaults more effectively, but its poor detectability rates for new attacks and high false alarms constitute a serious shortcoming. The author in [26] has suggested an IDS model combined with the DT-SVM (the decision tree), which provides a high detectable rate while reducing special attacks from standard behavior [27], [28]. The system uses a hierarchical intellectual hybrid system incorporating the decision tree.

The ADAM is an intrusion detector designed to identify intrusions using data mining techniques (Audit Data Analytic and Mining). An Intrusion Detection using Data Mining (IDDM) is the real-time NIDS for abuse and anomaly detection, which is used as the Data Mining Technique (DMM) [29]. It applies laws of association, Meta rules and rules of character. Data mining is used to describe the network data and to analyze deviations using this information. Authors in [30] offer a method of detecting intruders with an expanding neural fuzzy network. This learning method integrates the artificial neural network (ANN) with FIS systems and evolutionary algorithms. They develop an algorithm using fluffy rules and allow the creation of new neurons. They employ Snort to collect algorithm data and then compare their technology to an enlarged neural network.

Author in [31] develops anomaly detection statistical neural network classifiers to recognize UDP flood attacks. The back propagation neural network (BPN) was demonstrated to be more efficient in developing IDS compared with various neural network classifiers. It employs the background multiplication method for intrusion detection by the sample and attributes queries to analyze and determine the essential training data components. It can reduce the time of processing, storage, etc. The Bayesian rule of conditional probability was written by a well-known article, which shows that the base-rate failure of intrusion detection is involved. Clustering is an anomaly-based approach of detection that is capable without prior notification to detect new attacks and to identify natural data groups based on pattern similitudes. In K-Means, DBScan, and others use cluster analysis to identify a set of traffic behaviors. The Naïve Bayes classifiers provide even this classifier with a simple structure for its experimental investigation with a very competitive result. According to the author, the classification task of Naïve Bayes is more effective. It shows that Naïve Bayes classify network intrusion more efficiently than neural network detection.

Genge et al. [32] suggested an innovative approach for resilient distributed intrusion detection systems. The framework controls the outcomes of the risk assessment method to recognize and rank serious communications flows. These kinds of flows are incorporated in the issues related to the optimization, which lessen the organized detection devices when applying an algorithm of a shortest-path routing to reduce the delay in the communication. This work elaborated the resilient dispersed intrusion finding design algorithm that can detect the devices can fail or cooperate. This algorithm accurately positioned the detection devices to confirm whether the infrastructure was strong for most of the K communications path letdowns. The outcomes from the experiments demonstrated the distributed intrusion detection design framework effectiveness.

Ravale et al. [33] designed a hybrid method which was combined the data mining approaches such as K Means clustering algorithm with the classification module of the RBF kernel function of Support Vector Machine. The main motive of this work was to lessen the quantity of attributes that are associated with every point of data. This work proved the overwhelmed performance in terms of accuracy and detection rate when carried out on the KDDCUP'99 Data Set. *Gupta et al.* [34] performed an intrusion detection by ant colony gives good classification by using the NSL-KDD data set. It does not contain redundant records. The NSL-KDD dataset comprises two parts are 1) average establishment 2) termination. The figure shows a flow chart for intrusion detection technique by ant colony optimization.NSL KDD comprises three kinds 1) Fundamental individual connection features, 2) connection content features 3) traffic features. Table 1 investigates some machine learning classification techniques used to develop the IDS framework.

In the proposed NIDS framework distance-based clustering and classification methods are used for detecting and classifying the types of intrusions from the IDS datasets. For data normalization, the attribute normalization is performed based on its minimum and maximum values along with the real value. For clustering, the Gaussian parameter, and a number of clusters are considered for grouping the data attributes into clusters based on the distance value. Moreover, the number of pigeons, compass factor, and probability

TABLE 1. (A) Conventional IDS frameworks (B) Machine learning classification techniques used for IDS.

Authors & Year	Methodology	Parameters	Advantages/Disadvantages
Mirsky, et al [35]	Kitsune core algorithm (i.e. KitNET) is developed for detecting anomalies in the networking systems.	Here, the mean, standard deviation, and variance of the unbounded data. Also, the covariance, magnitude and correlation coefficient are estimated feature mapping.	Advantages: 1. Light weight model 2. Reduced time consumption 3. Efficient processing in speed
Bovenzi, et al [36]	Multi Modal Deep Auto-Encoder model is developing for improving the security of IoT.	The encoder model is mainly utilized to identify anomalous traffics by computing the measures of minimum and maximum distance values, standard deviation and an average of the input data.	Advantages: 1. Efficient training 2. Intrinsic operating efficiency 3. High detection rate Disadvantages: 1. Increased computational complexity.
Sultana, et al [37]	This work presented a detailed survey on various machine learning models used for developing an IDS framework.	The signature and anomaly-based detection techniques are categorized concerning speed, reliability, false alarm rate, robustness, and flexibility.	Advantages: 1. Reduced misclassification rate due to optimization 2. Increased accuracy by handling the flow of packets Disadvantages: 1. Increased bottleneck 2. Highly dependent on the relevancy of features
Ren, et al [38]	An iForest-GA-RF model is developed for designing an efficient IDS framework.	In this model, the survival probability function is computed based on the fitness value of chromosomes.	Advantages: 1. Increased detection rate 2. Reduced false alarm rate Disadvantages: 1. Data sampling requires increased time consumption 2. Increased computational complexity
Gao, et al [39]	An adaptive ensemble based machine learning models are used for developing an IDS framework.	Here, the split Gini index is computed according to the sum of probability and the number of classes in the dataset. Also, the weight coefficient matrices have been estimated for incorporating the test results.	Advantages: 1. Ensured detection accuracy 2. Reduced time consumption 3. Minimal misclassification outcomes Disadvantages: 1. Inefficient in handling large dimensional datasets. 2. High complexity in algorithm design

TABLE 1. *(Continued.)* (A) Conventional IDS frameworks (B) Machine learning classification techniques used for IDS.

Machine Learning IDS Techniques	Purpose	References
Decision Tree	A classification tree is also known as a prediction model or decision tree, which produces the classified results for the given problems by taking decisions using a tree structure.	[40, 41]
Fuzzy Logic	Fuzzy set theory is used to cover well-thought-out real-world expert values for a difficult topic. The IDS framework is extensively used for anomaly prediction and classification based on rule formation.	[42-44]
Naïve Bayes	The NB is usually represented by a DAG format, which has the ability to handle large dimensional datasets with reduced time consumption.	[45, 46]
Genetic Algorithm	Many IDS applications are mainly used to produce the audit data from a set of categorization rules by computing the fitness function with the crossover, selection, and mutation operations.	[47-49]
Neural Networks	It is a type of network constructed with a set of nodes that replicates the brain's functioning. Here, each node has extensive connections to several other nodes in neighboring layers.	[50-52]
Support Vector Machine	It is a kind of supervised learning technique and extensively used in many multi-class prediction models due to its increased accuracy and efficiency.	[24, 53]

function are computed to improve the classifier's detection accuracy.

III. PROPOSED WORK

This section presents a detailed description of the proposed Anticipated Distance-based Clustering (ADC) incorporated with the DBScan mechanism for accurately identifying the intrusions from the given datasets. The main contribution

of this work is to precisely predict the normal and attacking labels from the given IDS datasets by using intelligent distance-based clustering and classification methodologies. The proposed attack detection framework objects to classify the types of attacks based on its attribute feature vectors with the likelihood function. Here, the ADC-DBScan clustering methodology is implemented to organize the attributes of the normalized dataset by computing the distance value. During data clustering, the preprocessed data is segregated into different chunks and, the attributes of each data unit are extracted based on the minimum distance value. Then, the novel Perpetual Pigeon Galvanized Optimization (PPGO) technique is employed for optimally selecting the attributes based on the global best fitness function. After that, the selected number of features is given to the classifier for the training models. The probability of selected attributes is computed for predicting the normal and attacking labels. For this purpose, the LNB classifier is employed in this work, which uses the training samples as the input and produces the classified label as the output according to the probability and likelihood functions. The novelty of this work is, it identifies the intrusions based on the optimal features by partitioning the normalized dataset into different chunks concerning the minimum distance value. Here, the deviation occurrence has been computed with respect to IP, port, timestamp and type of traffic. The overall flow of the proposed system is shown in Figure 2, which includes the following working modules:

- Dataset obtainment
- Preprocessing
- Clustering
- Optimization
- Intrusion Detection and Classification

In this work, the recent IDS datasets such as NSL-KDD, CICIDS and BotIoT have been utilized for intrusion detection and classification. Initially, data preprocessing is performed for obtaining the normalized data by eliminating the special characters and blank spaces in the raw datasets. Typically, identifying the intrusions or attacks from the large dimensional datasets requires increased time consumption for processing, leading to the system's increased complexity. Hence, the ADC incorporated with the DBScan data clustering technique has been applied to group the data into the form of clusters, where the distance is estimated for grouping the attributes. The PPGO technique is employed for selecting the best suitable features concerning the fitness function based on the likelihood function concerning the weight value of particles. After that, the machine learning classifier named as Likelihood Naïve Bayes (LNB) mechanism is employed for classifying whether the data is normal or intrusion. Then, the proposed intrusion detection and classification system were implemented in two phases of work. The deviation is computed for both regular and unauthenticated users while accessing the cloud applications. Based on this value, the cloud administrator gets alerted during unauthenticated data access from the cloud. The key benefits of using the proposed ADC-DBScan-LNB based intrusion detection and classification system are increased detection accuracy, ensured prediction outcomes, minimal time consumption for training and testing models, and reduced computational complexity.

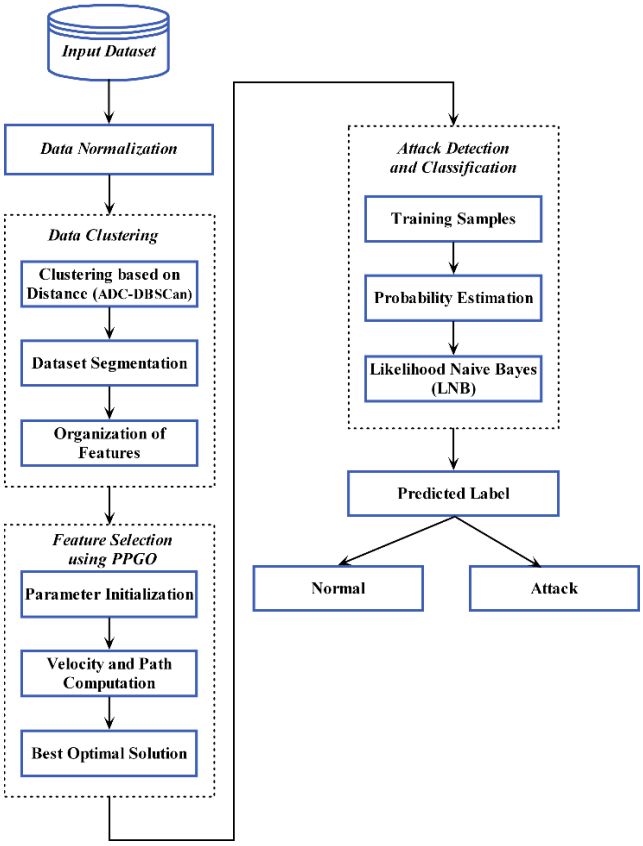

FIGURE 1. The overall flow of the proposed intrusion detection and classification system.

A. DATA PREPROCESSING

Typically, data preprocessing/normalization is one of the essential processes of attacking detection and classification systems. The network intrusion datasets are generally large in dimension, and huge data with noisy contents can degrade the system performance with misclassification results, increased false positives, and high time consumption for training the models. Hence, it is essential to preprocess the dataset before using it for attack/intrusion detection. In this work, the recent IDS datasets such as NSL-KDD, CICIDS, and BotIoT have been utilized for intrusion detection and classification. At first, the raw datasets are preprocessed to eliminate unwanted attributes in the dataset, which helps to improve the efficiency and accuracy of intrusion detection. It involves the processes of replacing missing values, removing irrelevant attribute information, and arranging attributes. Data preprocessing is defined as the extraction of best records from many records by criteria that all attacks are in equilibrium. Here, the normalization is performed by finding the minimum and maximum values of data attributes for transforming it to

the range of 0 to 1 as shown in below:

$$N_i = \frac{A_i - \min(A_i)}{\max(A_i) - \min(A_i)} \quad (1)$$

where, the data attribute N_i is set as 0 *if* (max *value* = min *value*), and A_i indicates the real value of attribute. Typically, the data preprocessing or normalization is more essential for improving the anomaly detection process. Also, the overall performance of the intrusion detection and classification system is highly depends on the normalized dataset without noisy contents. Finally, the normalized data attributes are given to the clustering scheme for grouping the similar data items based on the distance value.

B. ADC BASED DBSCAN CLUSTERING

Clustering is one of the essential processes and plays a vital role in accurately detecting intrusions on the dataset. The group of information helps to improve the overall system performance. For this purpose, different types of distance and density-based clustering techniques are employed in the conventional works, which intends to strengthen group the attributes for identifying the intrusions that exist in the dataset. But, its significant limitations are the inability to handle large datasets, inefficient data separation, and high time complexity. To solve these problems, this work intends to develop a new clustering technique by incorporating the functionalities of both the ADC and DBScan clustering approaches. It helps to improve the overall efficiency and detection accuracy of intrusion detection and classification. Also, it reduces the computational complexity of classification by efficiently grouping the data based probability distance measure. Based on the feature attributes of the known attacks, the frequent attacks have been identified and detected based on its probability value. The mean of entire cluster has been calculated concerning the threshold value, if the mean value is beyond the threshold value, an automatic alert message has been generated to the administrator that helps to find the illegal access of the network with the IP address.

The DBcan is a density-based clustering technique that uses the minimum number of points and density of neighborhood pixels. It forms the new cluster based on the neighborhood points concerning the radius, and the main reason for using this technique is that it efficiently reduces the average time complexity. Also, it utilizes the global density parameters for identifying the clusters with different shapes and densities. The conventional density-based clustering techniques highly depend on the single value parameters, leading to reduced clustering efficiency. But, the proposed DBSCan technique can fine-tune different values of parameters in every group of the cluster. Here, the local density function is computed at each point based on the approximation of the overall density function, which is estimated with respect to the sum of all functions. In this model, the neighborhood information of the data points in a sparse region is dynamically captured for clustering.

Algorithm I – ADC With DBScan Clustering

Input: Preprocessed dataset $P_D = \{\overrightarrow{d_1}, \overrightarrow{d_2}, \ldots \overrightarrow{d_N}\}$, number of clusters NC, and Gaussian parameter φ;
Output: Clustered group data $G_1, G_2 \ldots G_{NC}$;
Step 1: Estimate the distance value $D_{i,j}$ between the data of $\overrightarrow{d_i}$ and $\overrightarrow{d_j}$ by using the following equation:
$D_{i,j} = \left\| \sigma\left(\overrightarrow{d_i}\right) - \sigma\left(\overrightarrow{d_j}\right) \right\|^2$
Step 2: Compute the cutoff distance C_D;
Step 3: Then, the local density α_i function is computed with respect to the cutoff distance value as shown in below:

$$\alpha_i = \sum\nolimits_{j \in i} e^{-\left(\frac{D_{ij}}{C_D}\right)^2}$$

Step 4: After that, the distance β_i is estimated for each data attribute as represented below:

$$\beta_{xi} = \begin{Bmatrix} \min\limits_{j<i} \{D_{xixj}\}, & i \geq 2 \\ \max\limits_{j \geq 2} \{D_{xj}\}, & i = 1 \end{Bmatrix}$$

Step 5: Consequently, estimate the distance $\varepsilon_i = \alpha_i \beta_i, \quad i \epsilon I_{P_D}$;
Step 6: The data points are selected according to the NC maximum values of clusters $\{\varepsilon_i\}_{i=1}^{N}$ as shown in below:
$C_E = \{C_{Ei}\}_{i=1}^{NC}$ // Where, C_E–Cluster center;
Step 7: Finally, the remaining points of same cluster are grouped based on the higher density value of nearest neighbors. The preprocessed dataset P_D is split into NC number of clusters as $G_1, G_2 \ldots G_{NC}$;
Step 8: Return, the subsets of clusters $G_1, G_2 \ldots G_{NC}$;

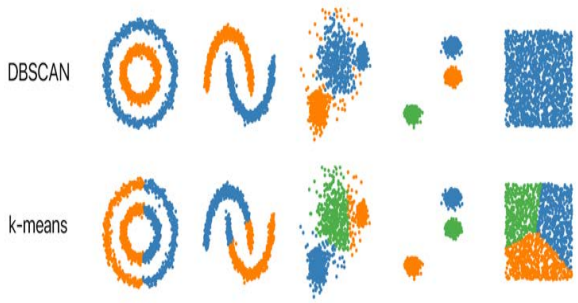

FIGURE 2. Comparison of density-based spatial clustering in DBSCAN & K-Means.

Figure 2 compares the data clustering of conventional K-means and DBScan methods, in which the DBScan model can handle malicious data with noise disturbance. Directly density-reachable: A point p is directly density reachable from a point q w.r.t. Eps, MinPts, if NEps (q): {pbelongs to D | dist(p,q) ≤ Eps} and |N Eps (q)| ≥ MinPts.MaxEps: Maximum radius of the neighborhood

MinEps: Minimum number of points in an Eps-neighborhood of that point

C. PERPETUAL PIGEON GALVANIZED OPTIMIZATION (PPGO)

After normalizing the dataset, the novel Perpetual Pigeon Galvanized Optimization (PPGO) technique is employed to select the best suitable fitness function features. In this stage, the optimal parameters are identified based on the likelihood function with respect to the weight value of particles. The best fitness value is computed at varying iterations based on the maximum likelihood value and weight value. Consequently, the optimal features include energy, probability based on density, likelihood, trust, energy, weight function, IP, traffic stamp, port and average total number of packets are selected and used for training and testing the classifier. The PPGO mechanism is a bio-inspired optimization technique developed based on the swarm intelligence behavior model. Moreover, the proposed PPGO technique has the ability to efficiently handle multi-objective and complex optimization problems with reduced number of iterations. Hence, it is more suitable for improving the accuracy and detection efficiency of NIDS. The significant benefits of the PPGO technique are as follows: reduced computational complexity, best optimal solution with a reduced number of iterations, increased convergence speed, and high efficiency. Here, the number of pigeons in the current iteration No_P, and compass factor C_R are considered as the inputs for this optimization, and then it produced the best optimal solution B_{opt} as the output. At first, the random number of pigeons and number of iterations are initialized as shown in below:

$$P_{xi} = P_{x1}, P_{x2} \ldots N_{No_P} \tag{2}$$

$$No_P(k+1) = \frac{No_P(k)}{2} \tag{3}$$

where, P_{xi} is the number of pigeons, k indicates the current iteration, and No_P denotes the pigeons in the current iteration. Then, the pigeons are computed with respect to the minimum fitness value by using the following model:

$$F_{fun} = wt_1 \times \frac{At_S}{At_T} + wt_2 \times FPR + wt_3 \times \frac{1}{TPR} \tag{4}$$

$$P_{xG} = best\ pigeon(min\_F_{fun}) \tag{5}$$

where, F_{fun} is the fitness function, wt_1, wt_2 and wt_3 are the weight values, FPR is the false positive rate, and TPR is the True Positive Rate. After that, the path and velocity of each pigeon are computed as shown in below:

$$H_i(k+1) = H_i(k) \cdot e^{-C_{Rk}} + rand \cdot (P_{xG} - P_{xi}(k)) \tag{6}$$

$$P_{xi}(k+1) = P_{xi}(k) + H_i(k+1) \tag{7}$$

where, H_i indicates the velocity, C_R denotes the compass factor, P_{xG} is the global solution, $P_{xi}(k)$ defines the present position of pigeon, and $H_i(k)$ is the present velocity. Yet again, the pigeons are computed with respect to the estimated fitness function and the global best solution P_{xG} is updated. Correspondingly, the loop has been executed until reaching the number of iterations of the No_P, where the order of subscript base arranges the pigeons $= \frac{No_P}{2}$. Moreover, the destinations of pigeons are estimated by using the following equation:

$$CP_{pos}(k+1) = \frac{\sum P_x(kk+1) \cdot Fitness(P_x(k+1))}{No_P \sum Fitness(P_x(k+1))} \tag{8}$$

where, center pigeon CP_{pos} at current iteration, P_{xi} indicates the current position of all pigeons, then the updated position is indicated as follows:

$$P_{xi}(k+1) = P_{xi}(k) + rand \cdot \left(CP_{pos}(k+1) - P_{xi}(k)\right) \tag{9}$$

Finally, the global best solution is updated and returned as follows:

$$B_{opt} = P_{xG} \tag{10}$$

Based on this solution, the optimal number of features are selected and used to train the classifier to accurately detect the intrusions.

Algorithm II– Perpetual Pigeon Galvanized Optimization (PPGO)

Input: Number of pigeons in the present iteration (o_P), and compass factor C_R;
Output: Best optimal solution B_{opt};
Step 1: At first, randomly initialize the number of pigeons as $P_{x1}, P_{x2} \ldots N_{No_P}$;
Step 2: Initialize the number of iterations as ay_1, ay_2, where $ay_1 > ay_2$;
Step 3: Evaluate the pigeons according to the fitness function by using equ (4) and (5);
Step 4: While $(ay_1 \geq 1)$ do
  Update the path and velocity of each pigeon by using equ (6) and (7) respectively;
Step 5: Consequently, compute the pigeons $P_{x1}, P_{x2} \ldots N_{No_P}$ according to its fitness values;
Step 6: Then, update the best global solution of P_{xG};
Step 7: End while;
Step 8: While ($No_P \geq 1$) do
Step 9: Arrange the pigeons with respect to the fitness value;
Step 10: $No_P = \frac{No_P}{2}$
Step 11: Compute the destination of pigeons by using equ (8) and (9);
Step 12: Update the global best solution as shown in equ (10);
Step 13: End while;

D. LIKELIHOOD NAïVE BAYES (LNB) CLASSIFICATION

In this work, the LNB based machine learning classification model is employed to classify whether the data is average or attack. This technique is developed based on the conventional

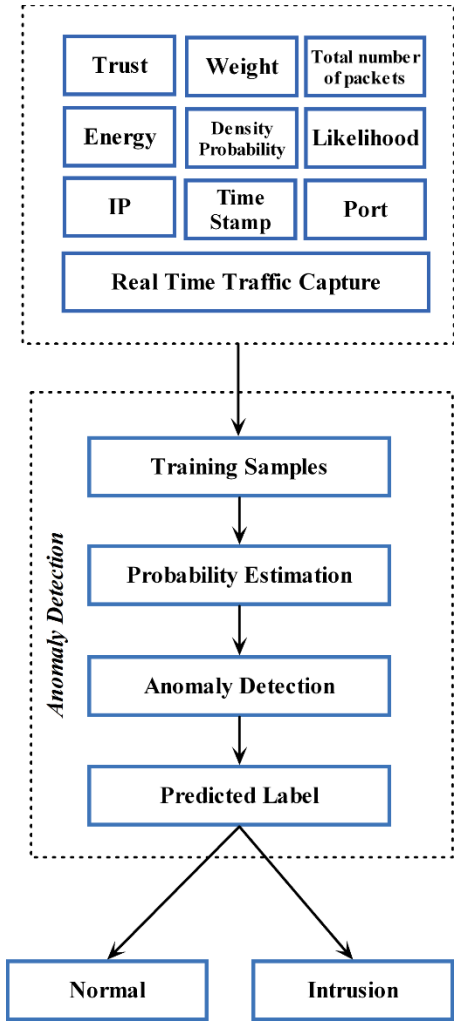

FIGURE 3. Attack analysis and anomaly detection processing.

NB classification technique. But in the proposed mechanism, the optimal parameters are identified and incorporated with the classifier for improving the overall accuracy and efficiency of the intrusion detection and classification system. The main advantage of using this technique is that it can handle the large dimensional datasets with reduced computational and time complexity by splitting and working with the blocks of information. Figure 3 shows the attack analysis and anomaly detection process flow using the LNB based classification technique.

IV. PERFORMANCE ANALYSIS

This section presents the performance analysis of both existing and proposed mechanisms concerning varying evaluation metrics. To validate the results of this work, three different types of datasets have been utilized, including NSL-KDD, CICIDS-2017, and Bot-IoT. The different types of measures used in this analysis are sensitivity, specificity, accuracy, precision, recall, F1 score, error, and False Negative Rate (FNR), delay and detection rate. Figures 4 (a) and (b) show the attacking information and features of the NSL-KDD dataset.

Algorithm III– Likelihood Naïve Bayes (LNB) Classification

Input: Selected attributes s_i based on the optimal solution;
Output: Classified label C_L;
Step 1: At first, the set of attributes are initialized as follows: $s_1, s_2 \ldots s_n$, and its probability of hypothesis is computed in below:

$$p\left(\frac{H}{s_1}, s_2 \ldots s_n\right) = \frac{p(s_1, s_2 \ldots s_n/H)p(H)}{p(s_1, s_2 \ldots s_n)}$$

//Where, $i \in 1, 2 \ldots n$ and $p\left(\frac{s_i}{H}\right)$

Step 2: Then, the prior probability of the training data is estimated as follows:

$$p\left(\frac{H}{s_1, s_2 \ldots s_n}\right) = \frac{p(H)\prod_{i=1}^{n} p(s_i/H)}{p(s_1, s_2 \ldots s_n)}$$

//Where, $p(s_1, s_2 \ldots s_{n-1}, si + 1 \ldots sn)$

Step 3: The classified result is predicted as follows:

$$p\left(\frac{H}{s_1, s_2 \ldots s_n}\right) \alpha P(H) \prod_{i=1}^{n} P(s_i/H)$$

$$P\left(\frac{s}{H}\right) = argmax$$

$$\left\{P\left(\frac{s_1}{H}\right) P(H), P\left(\frac{s_2}{H}\right) P(H) \ldots P\left(\frac{s_n}{H}\right) P(H)\right\}$$

//Where, $P\left(\frac{s}{H}\right)$ indicates the probability function, $P(H)$ denotes the probability of hypothesis, and $P(s)$ defines the probability of training data s.

Step 4: Based on this probability function, the classified label C_L is produced as whether normal or intrusion;

It includes the types of attacks such as DOS, U2R, Probe, and R2L are examined in data pre-processing. Data separation was performed depending on whether the connection value is average establishment or termination. Refining attack data depending on the connection received, initial attack data was made ready. Then, the CICIDS dataset [54] comprises different types of attacks such as DDoS, DoS, web attack, and brute force, including many samples. This dataset is also one of the recent and extensively used IDS datasets in many network application systems, where the training and testing samples are more essential.

Consequently, the Bot-IoT dataset comprises around more than 72,000 records related to different types of attacks. Also, it is one of the new datasets compared to the other IDS datasets but is highly complicated to process. Table 1 and Table 2 depict the detected number of attacks and the original data set in the NSL-KDD dataset for both U2R and R2L, respectively. These results show that the proposed ADC-DBScan-LNB based IDS accurately detect the number of attacks from the given dataset with better training and testing models.

Figure 5 shows the best fitness plot of both existing BAT and proposed LNB techniques concerning the varying number of iterations. The analysis shows that the proposed technique finds the optimal best fitness value with a reduced number of iterations compared to the conventional

S.No	Feature Name	S.No	Feature Name
1	Duration	23	Count
2	Protocol_type	24	Srv_count
3	Service	25	Serror_rate
4	Flag	26	Src_serror_rate
5	Src_bytes	27	Rerror_rate
6	Dst_bytes	28	Srv_rerror_rate
7	Land	29	Same_srv_rate
8	Wrong_fragment	30	Diff_srv_rate
9	Urgent	31	Srv_diff_host_rate
10	Host	32	Dst_host_name
11	Num_failed_logins	33	Dst_host_srv_count
12	Logged_in	34	Dst_host_same_srv_rate
13	Num_compromised	35	Dst_host_diff_srv_rate
14	Root_shell	36	Dst_host_same_src_port_rate
15	Su_attempted	37	Dst_host_srv_diff_host_rate
16	Num_root	38	Dst_host_serror_rate
17	Num_file_creations	39	Dst_host_srv_serror_rate
18	Num_shells	40	Dst_host_rerror_rate
19	Num_access_files	41	Dst_host_rerror_rate
20	Num_outbound_cmds		
21	Is_hot_login		
22	Is_guest_login		

1 to 10 - Basic Features
11 to 22 - Content Features
23 to 41 - Traffic Features

(a)

Attack Class

Attack Type	DoS	Probe	R2L	U2R
	Back	Satan	Guess_password	Buffer_overflow
	Land	Ipsweep	Ftp_write	Load module Rootkit
	Neptune	Nmap	Imap	Perl
	Pod	Portsweep Mscan	Phf	SQL attack
	Smurf	Saint	Multihop Warezmaster	X term
	Teadrop		Warezclient	Ps
	Apache 2		Spy	
	Udp Storm		Xlock	
	Process Table		Xsnoop	
	Worm		Snmpguess	
			Snmpgetattack	
			Httptunnel	
			Sendmail	
			Named	

(b)

FIGURE 4. (a). Features of NSL-KDD dataset. (b). Attack information of NSL-KDD dataset.

approach based on the likelihood function. Consequently, Figure 6 shows the transmission delay of both existing and proposed IDS concerning the varying number of transmitted packets. Typically, the delay is calculated based on the number of packets that are successfully transmitted with reduced time consumption and without any loss of information. This analysis shows that the proposed ADC-DBScan-LNB technique outperforms the other technique with reduced delay of transmission.

In the existing work [55], intrusion detection is performed with the help of radial basis function integrated with BAT algorithm. Figure 7 depicts the confusion matrix of the proposed scheme for the NSL-KDD dataset, where the actual/predicted numbers of classes are determined with the number of sequences. The confusion matrix is mainly

TABLE 2. Examining the dataset for U2R.

Attack	Original in NSL KDD data set (1025973)	Selected
Load module	9	4
Root kit	10	05
Buffer overflow	30	15

TABLE 3. Examining dataset for R2L.

Attack	Original in data set(1,025,973)	Selected
warezmaster	20	10
warezclient	890	445

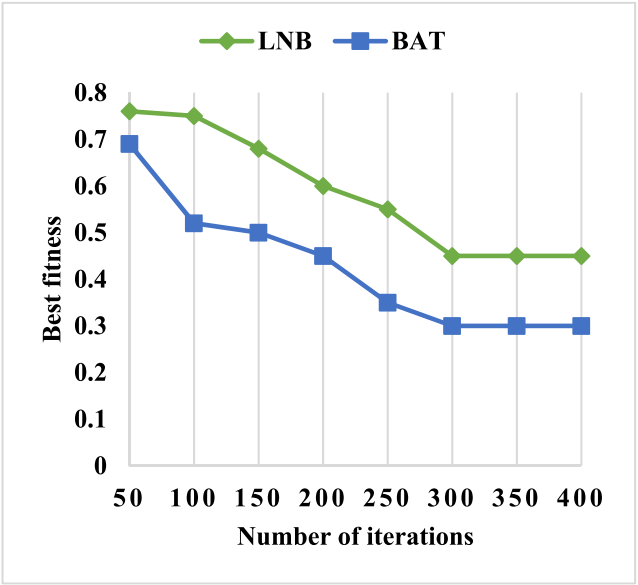

FIGURE 5. Best fitness vs number of iterations.

constructed for determining the detection accuracy of intrusion detection and classification with respect to the differential ratio of the sum of values as shown in the diagonal matrix.

In the work [56], Ensemble Learning depending upon Wi-Fi Network Intrusion Detection System. The effectiveness of Different base learners was increased by using the prediction model with the help of ensemble learning. Wifi network intrusion detection comprises three sections in it namely 1) AWID data set 2) data preparation 3) Ensemble algorithms. Two versions used in AWID dataset are 1) attack-class 2) attack-specific. Flooding, impersonation, injection are the attack classes. Noisy, missing values in dataset are eliminated in dataset and made ready in data preparation by reducing number of features. Ensemble algorithms utilized in this proposed method are Bagging, Random forest, Extra-trees, XGBoost. Training, Prediction is the two phases in bagging. Improving bagging method by decision trees was called as Random Forest. Extra-trees described random forests characteristics. XGBoost performed gradient Boosting.

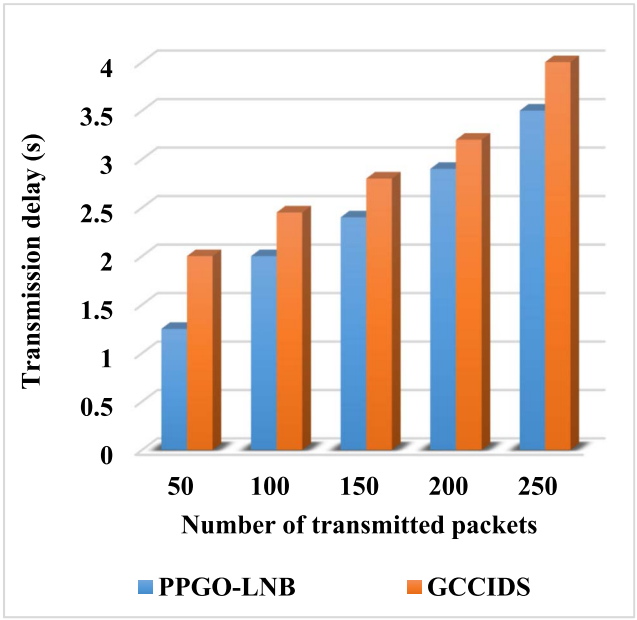

FIGURE 6. Transmission delay vs number of transmitted packets.

	Normal	DoS	Probe	R2L	U2R
Normal	99.2% (66826)	0.9% (392)	0.8% (93)	0.8% (8)	0.0% (0)
DoS	0.6% (405)	99.0% (45468)	0.6% (71)	0.3% (3)	0.0% (0)
Probe	0.1% (101)	0.1% (61)	98.6% (11493)	0.4% (4)	0.0% (0)
R2L	0.0% (10)	0.0% (7)	0.0% (0)	98.5% (981)	0.0% (0)
U2R	0.0% (2)	0.0% (0)	0.0% (0)	0.0% (0)	100.0% (53)

FIGURE 7. Confusion matrix with the predicted labels for NSL-KDD dataset.

Figures 8 (a) and (b) show the proposed scheme's confusion matrix for both CIC-IDS 2017 and Bot-IoT datasets, respectively. The results show that the proposed technique accurately detects the normal and attacking information with increased true positives.

Typically, the performance of IDS can be assessed by using the key measures of accuracy, precision, recall, sensitivity, specificity, and false alarm rate. By using above performance measures the values are calculated and graph has been plotted.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{11}$$

$$False\ alarm\ rate = \frac{FP}{(FP + TP)} \tag{12}$$

	Benign	DoS GoldenEye	DoS Hulk	DoS slowhttpest	DoS Slowloris	Heart Bleed
Benign	99% (435423)	1.9% (199)	1.8% (4268)	2.0% (112)	2.0% (118)	0.0% (0)
DoS GoldenEye	0.0% (192)	97.1% (9994)	0.0% (105)	0.1% (3)	0.0% (2)	0.0% (0)
DoS Hulk	1.0% (4185)	0.9% (96)	98.1% (226606)	0.9% (48)	1.1% (61)	0.0% (0)
DoS slowhttpest	0.0% (117)	0.0% (2)	0.0% (52)	97.0% (5335)	0.0% (2)	0.0% (0)
DoS Slowloris	0.0% (115)	0.0% (3)	0.0% (42)	0.0% (2)	96.8% (5614)	0.0% (0)
Heart Bleed	0.0% (0)	0.0% (0)	0.0% (1)	0.0% (0)	0.0% (0)	100.0% (12)

(a)

	Benign	Information Gathering	DDoS Attack	DoS Attack	Information Theft
Benign	96.6% (7377)	0.0% (11)	0.0% (124)	0.0% (123)	0.0% (0)
Information Gathering	0.1% (6)	96.4% (140454)	0.1% (2833)	0.1% (2450)	0.0% (0)
DDoS Attack	1.9% (148)	1.9% (2812)	98.2% (3027894)	2.0% (51714)	2.0% (26)
DoS Attack	1.3% (103)	1.7% (2453)	1.7% (51721)	97.9% (2586108)	1.7% (21)
Information Theft	0.0% (0)	0.0% (1)	0.0% (26)	0.0% (20)	96.3% (1222)

(b)

FIGURE 8. (a). Confusion matrix with the predicted labels for CIC-IDS 2017 dataset. (b). Confusion matrix with the predicted labels for BoT-IoT dataset.

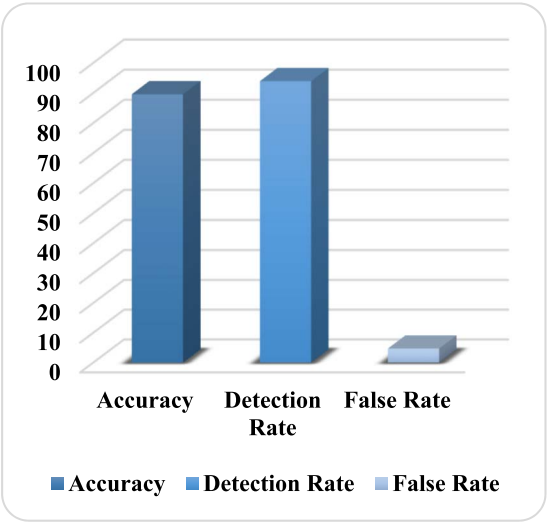

FIGURE 9. Performance analysis of proposed work.

The Figure 9 shows the various performance measures such as Accuracy, Detection rate and the false rate. All the values are measure in terms of Percentages. The calculated accuracy value of the proposed work is 89.56%. The detection

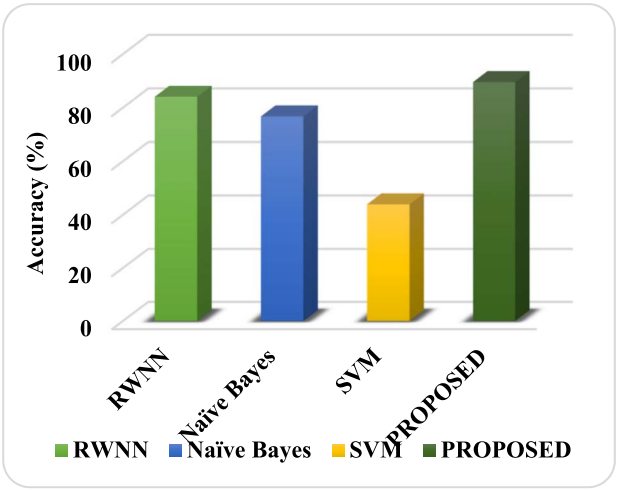

FIGURE 10. Accuracy comparison with existing classifier.

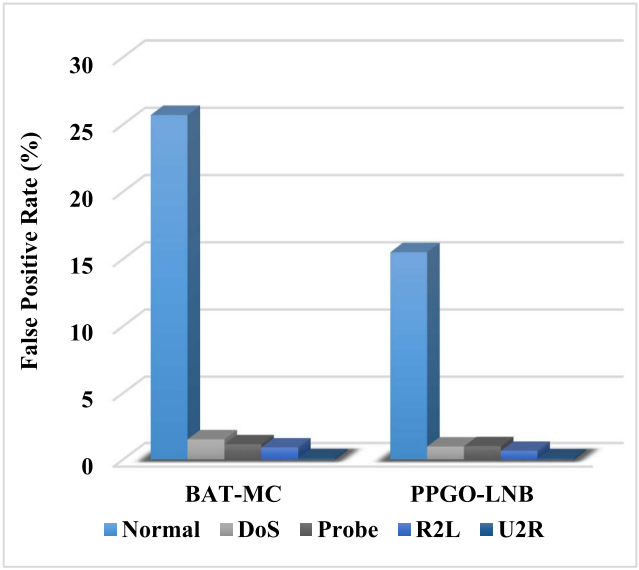

FIGURE 11. False rate comparisons with existing models.

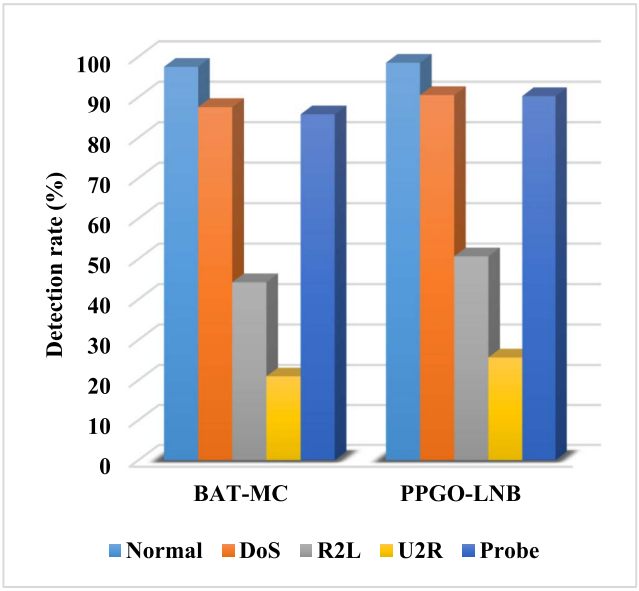

FIGURE 12. Detection rate comparisons with existing models.

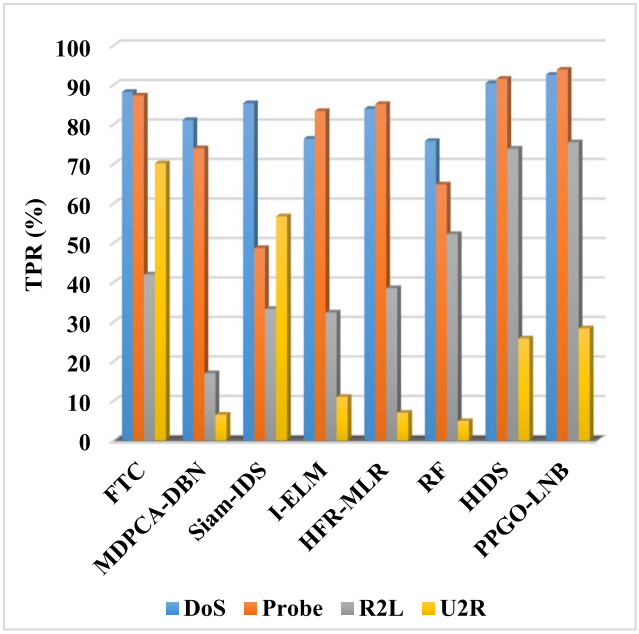

FIGURE 13. TPR of existing and proposed IDS techniques.

rate is evaluated as 93.89%. The proposed work is having minimum false rate as 4.7%. If the minimum value of false rate implies the superior performance of the proposed work.

Figure 10 shows the comparison of accuracy values with the existing classifier. The conserved classifiers [57] are the Random Weight Neural Network (RWNN), Naïve Bayes, and Support Vector Machine (SVM). The RWNN classifiers shows the accuracy value of 84.08 %, Naïve Bayes classifiers are giving 76.74%, SVM classifier produces 43.9% of accuracy. The proposed classifier is overwhelmed among all the existing classifiers.

Figure 11 and 12 shows the false positive rate and detection rate of the existing [58] and proposed intrusion detection techniques. Here, the comparative analysis is taken for the NSL-KDD dataset. The obtained results show that the proposed PPGO-LNB technique increases detection rate and reduces FPR for all types of attacks in the NSL-KDD dataset because the proposed scheme detects intrusions based on the optimal selection of attributes from the clustered data, which helps to enhance the overall detection efficiency of the IDS.

Figure 13 compares the TPR of the existing [9] and proposed IDS techniques for the four different types of classes in the NSL-KDD dataset. This analysis shows that the proposed PPGO-LNB scheme provides an increased TPR compared to the other models by predicting the classified labels based on efficient clustering and classification processes.

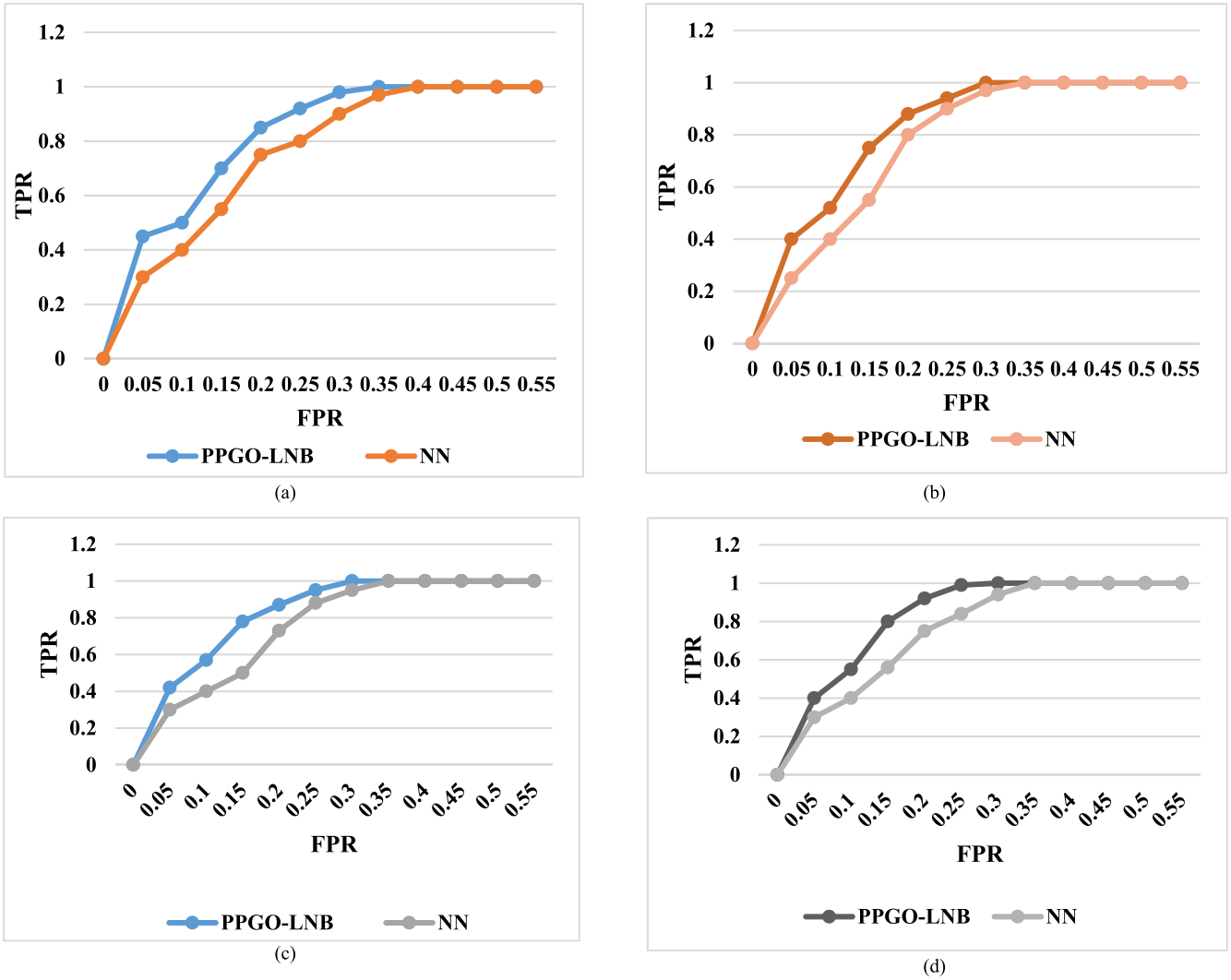

FIGURE 14. (a). ROC for Dos Attacks. (b). ROC for Botnet Attacks. (c). ROC for Web attacks. (d). ROC for Brute-force.

Figure 14 (a) to (d) depicts the ROC analysis of both existing and proposed intrusion detection techniques with respect to different types of attacks such as DoS, Botnet, web attacks, and Brute-force. The analysis shows that the proposed ADC-DBScan-LNB technique outperforms the existing NN technique with improved performance outcomes for all types of attacks. The ROC of the classifier is mainly evaluated for estimating the TPR and FPR of the detection system. Also, the increased value of TPR indicates an improved performance of the system. Here, Figure 15 evaluates the AUC of both existing and proposed techniques, and the results show that the ADC-DBScan-LNB technique improves performance compared to the other methods.

Table 4 and Figure 16 compare the existing and proposed intrusion detection and classification approaches based on sensitivity, specificity, accuracy, precision, recall, F1-score, Matthews Correlation Coefficient (MCC), and kappa coefficient. These measures are calculated as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{13}$$

$$Sensitivity = \frac{TN}{TN + FP} \tag{14}$$

$$Precision = \frac{TP}{TP + FP} \tag{15}$$

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \tag{16}$$

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{17}$$

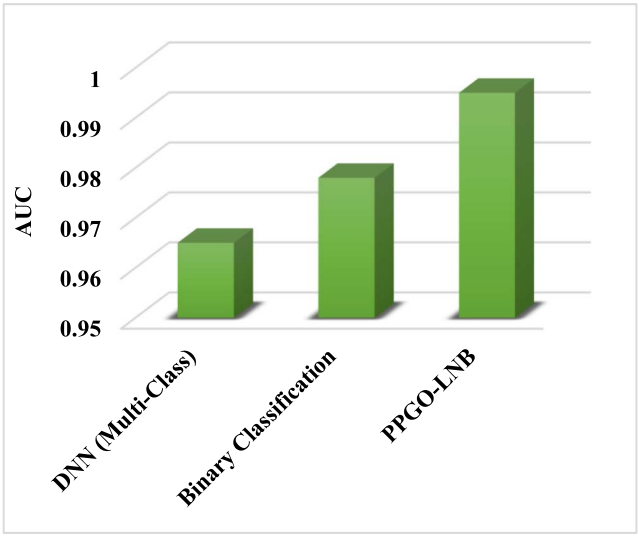

FIGURE 15. Analysis of AUC for both existing and proposed intrusion detection and classification techniques.

TABLE 4. Comparative analysis between existing and proposed intrusion detection and classification techniques.

Parameters	Methods	
	NN	ADC-DBScan-LNB
Sensitivity	99.8	99.9
Specificity	99	99.2
Precision	98	98.99
F1_Score	97	99.5
MCC	98	99.4
Kappa coefficient	97	98.5

$$Kappa = \frac{OA - RA}{1 - (RA)} \tag{18}$$

$$OA = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$RA = \frac{((TN + FP) \times (TN + FN) + (FN + TP) \times (FP + TP))}{OA^2} \tag{20}$$

where, TP indicates the True Positives, TN defines the True Negatives, FP represents the False Positives, and FN indicates the False Negatives. From this analysis, it is proved that the proposed ADC-DBScan-LNB technique provides an improved performance result, when compared to the classification technique. Because in the proposed scheme, the training and testing of classifier is performed based on the optimal number of features obtained from the given datasets. Also, the classifier accurately predicts the classified label based on the likelihood function of optimization, which helps to improve the overall system performance.

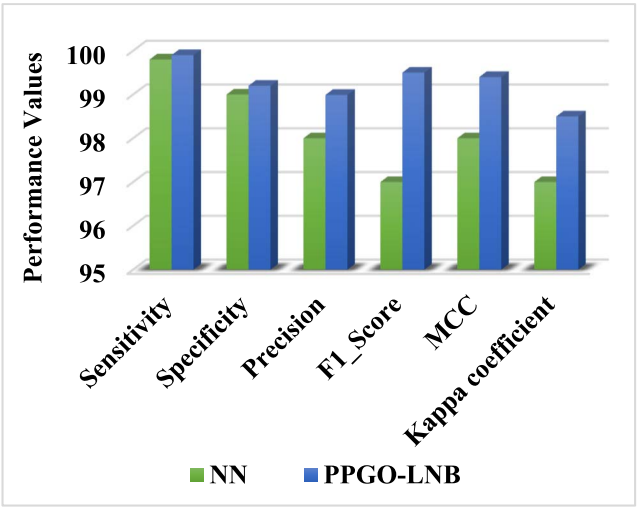

FIGURE 16. Overall comparative analysis.

TABLE 5. Time complexity analysis of existing and proposed techniques for NSL-KDD dataset.

Techniques	Time consumption	
	Training time (s)	Testing time (s)
DT	562	88
LR	1809	132
NB	1394	129
KNN	970	483
ANN	7622	891
SVM	2940	376
RF	201	53
GBDT	176	32
Adaboost	246	61
LightGBM	54	26
PPGO-LNB	35	20

TABLE 6. Time complexity analysis of existing and proposed techniques for CICIDS 2017 dataset.

Techniques	Time consumption	
	Training time (s)	Testing time (s)
DT	1231	62
LR	1338	103
NB	334	146
KNN	11,034	6499
ANN	4026	721
SVM	5034	144
RF	1194	22
GBDT	6459	43
Adaboost	2260	48
LightGBM	198	12
PPGO-LNB	150	10

Table 5 and Table 6 compares the training and testing consumption of both existing [59] and proposed techniques for NSL-KDD and CIC-IDS2017 datasets respectively.

TABLE 7. (A). Performance analysis of existing and proposed techniques for BoT-IoT Dataset. (B). Time complexity analysis of existing and proposed techniques for NSL-KDD Dataset. (C). Performance analysis of existing and proposed techniques for CIC-IDS 2017 Dataset.

(A)

Techniques	Training Accuracy	Sensitivity	Specificity	F1-score	Testing Accuracy	Sensitivity	Specificity	F1-score
PSO	99.899	99.929	99.898	99.898	99.898	99.928	99.896	99.896
WOA	99.918	99.926	99.919	99.967	99.916	99.924	99.916	99.965
BAT	99.975	99.977	99.9743	99.968	99.973	99.975	99.941	99.966
TSO	99.949	99.944	99.905	99.969	99.947	99.942	99.903	99.967
GWO	99.950	99.919	99.935	99.979	99.948	99.917	99.933	99.977
FFA	99.915	99.928	99.968	99.910	99.913	99.927	99.966	99.908
MVO	99.990	99.923	99.959	99.939	99.989	99.922	99.958	99.937
MFO	99.956	99.966	99.971	99.978	99.954	99.964	99.969	99.976
AQU	99.995	99.994	99.993	99.995	99.994	99.993	99.992	99.992
Proposed PPGO	99.996	99.995	99.995	99.996	99.995	99.995	99.994	99.993

(B)

Techniques	Training Accuracy	Sensitivity	Specificity	F1-score	Testing Accuracy	Sensitivity	Specificity	F1-score
PSO	90.133	93.143	90.043	90.043	67.575	70.585	73.882	67.163
WOA	91.959	92.809	92.099	96.989	69.409	70.259	75.972	74.115
BAT	97.693	97.933	94.533	97.023	75.192	75.432	78.473	74.197
TSO	95.091	94.571	90.681	97.091	72.078	71.558	73.656	73.786
GWO	98.202	92.072	93.753	98.172	72.944	69.814	77.801	75.609
FFA	91.673	93.053	97.013	91.223	69.218	70.598	80.944	68.451
MVO	99.197	92.517	96.167	94.117	76.466	69.786	79.835	71.059
MFO	95.760	96.810	97.320	98.060	73.187	74.237	81.176	75.162
AQU	99.348	99.348	99.350	99.348	77.382	77.382	83.692	77.077
Proposed PPGO	99.452	99.452	99.548	99.425	96.568	96.568	96.456	97.584

(C)

Techniques	Training Accuracy	Sensitivity	Specificity	F1-score	Testing Accuracy	Sensitivity	Specificity	F1-score
PSO	99.687	99.407	99.627	99.387	99.687	99.407	99.627	99.787
WOA	99.730	99.531	99.537	99.470	99.737	99.737	99.537	99.497
BAT	99.537	99.647	99.667	99.472	99.537	99.687	99.667	99.487
TSO	99.724	99.654	99.744	99.436	99.725	99.755	99.785	99.725
GWO	99.417	99.607	99.477	99.427	99.417	99.607	99.477	99.427
FFA	99.497	99.601	99.517	99.470	99.497	99.787	99.517	99.647
MVO	99.577	99.417	99.427	99.457	99.577	99.417	99.427	99.457
MFO	99.407	99.477	99.417	99.427	99.407	99.477	99.417	99.527
AQU	99.996	99.996	99.996	99.996	99.997	99.997	99.997	99.997
Proposed PPGO	99.997	99.997	99.997	99.99	99.998	99.998	99.998	99.998

Based on this evaluation, it is analyzed that the proposed PPGO-LNB technique requires reduced time consumption for both training and testing the models, when compared to the other models. Because, the proposed scheme utilizes the selected number of attributes for training the models, which helps to reduce the time consumption with ensured accuracy.

Table 7 (a) to (c) compares the training and testing accuracy, sensitivity, specificity, and F1-score of existing [60],

TABLE 8. Accuracy of various classification methods for cicids-2017 dataset.

Methods	Accuracy (%)
Adaboost	81.83
Deep Neural Network (DNN)	96.3
Ensemble classifier	96.8
GRU-RNN	89.94
K-Neighbors	94
GWO-PSO-RF	99.88
Proposed	99.99

and proposed intrusion detection methodologies by using the datasets of BoT-IoT, CIC-IDS 2017, and NSL-KDD datasets respectively. The obtained results show that the proposed PPGO technique outperforms the other techniques with improved performance values of these measures. In addition to that, Table 8 compares the accuracy of existing and proposed classification methodologies by using the CICIDS-2017 dataset. Various optimization-based machine learning and deep learning techniques have been considered during this evaluation. The obtained results show that the proposed PPGO-LNB technique provides improved performance outcomes compared to the other approaches. Because the clustering-based optimization mechanism could help to reduce the error rate of classification, which supports obtaining an increased accuracy of prediction.

V. CONCLUSION

Intrusion detection model identifies unauthorized access, abuse of data. Security was enhanced in intrusion detection systems. The intrusion detection system identifies network traffic not noticed by the firewall. Security was enhanced in the intrusion detection systems by deep neural network. Blocking malicious attacks, maintaining normal performance was the advantages of intrusion detection system. The intrusion detection system examined information such as source, destination number, and application version number for attack identification. Audit data identified intruders, log files. The intrusion detection system determined unauthorized access in the absence of confidentiality, integrity, and authentication. Naïve Bayes follows Bayes therm. The proposed work rectifies all the existing issues such as High Computational cost, High false alarm rate and Issues in detection of DOS, SQL injection, Buffer over flow, Login attempt and Apache struts attacks. All the above issues are rectified by using the proposed PPGO-LNB mechanism. Density based clustering identifies nonlinear structure depending upon density. It describes the cluster's density reachability, connectivity. The proposed work performance is verified by using various performance measures such as accuracy, detection rate and false rate. Intrusion detection using supervised, unsupervised algorithms are described in this paper. These measures are compared with the existing classifiers and proven the proposed approach's outperformance. The graphs clearly show the overwhelmed performance of proposed work. There will be a future work to extent this concept to various huge datasets.

In future, this work can be extended by applying an innovative Explainable Artificial Intelligence (EAI) models for designing an IDS architecture. This technique can be used to detect intrusions based on the mobile traffic classification. The multi-modal deep learning technique will be used to improve the overall efficacy of IDS.

ACRONYMS AND DEFINITIONS

Acronyms	Definitions
ABC	Ant Bee Colony
ACO	Ant Colony Optimization
ADC	Anticipated Distance based Clustering
ANN	Artificial Neural Network
DBScan	Density-based spatial clustering of applications with noise
DT	Decision Tree
FF	Firefly Optimization
FL	Fuzzy Logic
GA	Genetic Algorithm
GBDT	Gradient Boosting Decision Tree
GBM	Gradient Boost Modeling
IIDS	Improved Intrusion Detection System
KNN	K-Nearest Neighbor
LNB	Likelihood Naïve Bayes
MCC	Matthews Correlation Coefficient
NB	Naïve Bayes
NB	Naïve Bayes
NIDS	Network Intrusion Detection System
NN	Neural Network
PPGO	Perpetual Pigeon Galvanized Optimization
PSO	Particle Swarm Optimization
RF	Random Forest
RVM	Relevance Vector Machine
RWNN	Random Weight Neural Network
SVM	Support Vector Machine
WO	Whale Optimization

REFERENCES

[1] P. Negandhi, Y. Trivedi, and R. Mangrulkar, "Intrusion detection system using random forest on the NSL-KDD dataset," in *Emerging Research in Computing, Information, Communication and Applications*. Singapore: Springer, 2019, pp. 519–531.

[2] P. Verma, S. Anwar, S. Khan, and S. B. Mane, "Network intrusion detection using clustering and gradient boosting," in *Proc. 9th Int. Conf. Comput., Commun. Netw. Technol.*, 2018, pp. 1–7.

[3] K. Peng, V. C. M. Leung, and Q. Huang, "Clustering approach based on mini batch Kmeans for intrusion detection system over big data," *IEEE Access*, vol. 6, pp. 11897–11906, 2018.

[4] R. Alshamy and M. Ghurab, "A review of big data in network intrusion detection system: Challenges, approaches, datasets, and tools," *J. Comput. Sci. Eng.*, vol. 8, no. 7, pp. 62–74, 2020.

[5] H. Zhang, L. Huang, C. Q. Wu, and Z. Li, "An effective convolutional neural network based on SMOTE and Gaussian mixture model for intrusion detection in imbalanced dataset," *Comput. Netw.*, vol. 177, Aug. 2020, Art. no. 107315.

[6] R. Vinayakumar, M. Alazab, K. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, pp. 41525–41550, 2019.
[7] Amarudin, R. Ferdiana, and Widyawan, "A systematic literature review of intrusion detection system for network security: Research trends, datasets and methods," in *Proc. 4th Int. Conf. Informat. Comput. Sci.*, 2020, pp. 1–6.
[8] L. A. H. Ahmed and Y. A. M. Hamad, "Machine learning techniques for network-based intrusion detection system: A survey paper," in *Proc. Nat. Comput. Colleges Conf.*, 2021, pp. 1–7.
[9] C. Liu, Z. Gu, and J. Wang, "A hybrid intrusion detection system based on scalable K-means+ random forest and deep learning," *IEEE Access*, vol. 9, pp. 75729–75740, 2021.
[10] C. Zhang, M. Ni, H. Yin, and K. Qiu, "Developed density peak clustering with support vector data description for access network intrusion detection," *IEEE Access*, vol. 6, pp. 46356–46362, 2018.
[11] W. Alhakami, A. ALharbi, S. Bourouis, R. Alroobaea, and N. Bouguila, "Network anomaly intrusion detection using a nonparametric Bayesian approach and feature selection," *IEEE Access*, vol. 7, pp. 52181–52190, 2019.
[12] F. Salo, M. Injadat, A. B. Nassif, A. Shami, and A. Essex, "Data mining techniques in intrusion detection systems: A systematic literature review," *IEEE Access*, vol. 6, pp. 56046–56058, 2018.
[13] W. Wei, S. Chen, Q. Lin, J. Ji, and J. Chen, "A multi-objective immune algorithm for intrusion feature selection," *Appl. Soft Comput.*, vol. 95, Oct. 2020, Art. no. 106522.
[14] S. Manimurugan, A.-Q. Majdi, M. Mohmmed, C. Narmatha, and R. Varatharajan, "Intrusion detection in networks using crow search optimization algorithm with adaptive neuro-fuzzy inference system," *Microprocessors Microsyst.*, vol. 79, Nov. 2020, Art. no. 103261.
[15] A. Dickson and C. Thomas, "Improved PSO for optimizing the performance of intrusion detection systems," *J. Intell. Fuzzy Syst.*, vol. 38, no. 5, pp. 6537–6547, May 2020.
[16] J. Lee and K. Park, "GAN-based imbalanced data intrusion detection system," *Pers. Ubiquitous Comput.*, vol. 25, no. 1, pp. 121–128, Feb. 2021.
[17] N. Moustafa, B. Turnbull, and K.-K.-R. Choo, "An ensemble intrusion detection technique based on proposed statistical flow features for protecting network traffic of Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 3, pp. 4815–4830, Jun. 2019.
[18] S. El-Sappagh, A. S. Mohammed, and T. A. AlSheshtawy, "Classification procedures for intrusion detection based on KDD cup 99 data set," *Int. J. Netw. Secur. Appl.*, vol. 11, no. 3, pp. 21–29, May 2019.
[19] S. Elhag, A. Fernández, S. Alshomrani, and F. Herrera, "Evolutionary fuzzy systems: A case study for intrusion detection systems," in *Evolutionary and Swarm Intelligence Algorithms*. Cham, Switzerland: Springer, 2019, pp. 169–190.
[20] S. M. Kasongo and Y. Sun, "A deep learning method with filter based feature engineering for wireless intrusion detection system," *IEEE Access*, vol. 7, pp. 38597–38607, 2019.
[21] M. T. Nguyen and K. Kim, "Genetic convolutional neural network for intrusion detection systems," *Future Gener. Comput. Syst.*, vol. 113, pp. 418–427, Dec. 2020.
[22] S. Shitharth, N. Satheesh, B. P. Kumar, and K. Sangeetha, "IDS detection based on optimization based on WI-CS and GNN algorithm in SCADA network," in *Architectural Wireless Networks Solutions and Security Issues*. Singapore: Springer, 2021, pp. 247–265.
[23] C. Xu, J. Shen, X. Du, and F. Zhang, "An intrusion detection system using a deep neural network with gated recurrent units," *IEEE Access*, vol. 6, pp. 48697–48707, 2018.
[24] B. S. Bhati and C. S. Rai, "Analysis of support vector machine-based intrusion detection techniques," *Arabian J. Sci. Eng.*, vol. 45, no. 4, pp. 2371–2383, Apr. 2020.
[25] M. Khurana, R. Yadav, and M. Kumari, "Buffer overflow and SQL injection: To remotely attack and access information," in *Cyber Security*. Singapore: Springer, 2018, pp. 301–313.
[26] M. N. Ghuge and P. N. Chatur, "Collaborative key management in ciphertext policy attribute based encryption for cloud," in *Proc. 2nd Int. Conf. Inventive Commun. Comput. Technol. (ICICCT)*, Apr. 2018, pp. 156–158.
[27] S. Selvarajan, M. Shaik, S. Ameerjohn, and S. Kannan, "Mining of intrusion attack in SCADA network using clustering and genetically seeded flora-based optimal classification algorithm," *IET Inf. Secur.*, vol. 14, no. 1, pp. 1–11, Jan. 2020.
[28] S. Shitharth, "An enhanced optimization based algorithm for intrusion detection in SCADA network," *Comput. Secur.*, vol. 70, pp. 16–26, Sep. 2017.
[29] S. Basu, A. Bardhan, K. Gupta, P. Saha, M. Pal, M. Bose, K. Basu, S. Chaudhury, and P. Sarkar, "Cloud computing security challenges & solutions—A survey," in *Proc. IEEE 8th Annu. Comput. Commun. Workshop Conf.*, Jan. 2018, pp. 347–356.
[30] B. L. Dalmazo, J. P. Vilela, and M. Curado, "Triple-similarity mechanism for alarm management in the cloud," *Comput. Secur.*, vol. 78, pp. 33–42, Sep. 2018.
[31] J. Cheng, M. Li, X. Tang, V. S. Sheng, Y. Liu, and W. Guo, "Flow correlation degree optimization driven random forest for detecting DDoS attacks in cloud computing," *Secur. Commun. Netw.*, vol. 2018, pp. 1–14, Nov. 2018.
[32] B. Genge, P. Haller, and I. Kiss, "A framework for designing resilient distributed intrusion detection systems for critical infrastructures," *Int. J. Crit. Infrastruct. Protection*, vol. 15, pp. 3–11, Dec. 2016.
[33] U. Ravale, N. Marathe, and P. Padiya, "Feature selection based hybrid anomaly intrusion detection system using K means and RBF kernel function," *Proc. Comput. Sci.*, vol. 45, pp. 428–435, Jan. 2015.
[34] C. Gupta, A. Sinhal, and R. Kamble, "An enhanced associative ant colony optimization technique-based intrusion detection system," in *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*. New Delhi, India: Springer, 2015, pp. 541–553.
[35] Y. Mirsky, T. Doitshman, Y. Elovici, and A. Shabtai, "Kitsune: An ensemble of autoencoders for online network intrusion detection," 2018, *arXiv:1802.09089*.
[36] G. Bovenzi, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescapé, "A hierarchical hybrid intrusion detection approach in IoT scenarios," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2020, pp. 1–7.
[37] N. Sultana, N. Chilamkurti, W. Peng, and R. Alhadad, "Survey on SDN based network intrusion detection system using machine learning approaches," *Peer-to-Peer Netw. Appl.*, vol. 12, no. 2, pp. 493–501, Jan. 2019.
[38] J. Ren, J. Guo, W. Qian, H. Yuan, X. Hao, and H. Jingjing, "Building an effective intrusion detection system by using hybrid data optimization based on machine learning algorithms," *Secur. Commun. Netw.*, vol. 2019, pp. 1–11, Jun. 2019.
[39] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, "An adaptive ensemble machine learning model for intrusion detection," *IEEE Access*, vol. 7, pp. 82512–82521, 2019.
[40] M. A. Ferrag, L. Maglaras, A. Ahmim, M. Derdour, and H. Janicke, "RDTIDS: Rules and decision tree-based intrusion detection system for Internet-of-Things networks," *Future Internet*, vol. 12, no. 3, p. 44, Mar. 2020.
[41] W. Lian, G. Nie, B. Jia, D. Shi, Q. Fan, and Y. Liang, "An intrusion detection method based on decision tree-recursive feature elimination in ensemble learning," *Math. Problems Eng.*, vol. 2020, pp. 1–15, Nov. 2020.
[42] K. P. M. Kumar, M. Saravanan, M. Thenmozhi, and K. Vijayakumar, "Intrusion detection system based on GA-fuzzy classifier for detecting malicious attacks," *Concurrency Comput., Pract. Exp.*, vol. 33, no. 3, Feb. 2021, Art. no. e5242.
[43] A. N. Jaber and S. U. Rehman, "FCM–SVM based intrusion detection system for cloud computing environment," *Cluster Comput.*, vol. 23, no. 4, pp. 3221–3231, Dec. 2020.
[44] M. Almseidin, J. Al-Sawwa, and M. Alkasassbeh, "Anomaly-based intrusion detection system using fuzzy logic," in *Proc. Int. Conf. Inf. Technol.*, 2021, pp. 290–295.
[45] T. Wisanwanichthan and M. Thammawichai, "A double-layered hybrid approach for network intrusion detection system using combined Naive Bayes and SVM," *IEEE Access*, vol. 9, pp. 138432–138450, 2021.
[46] M. Artur, "Review the performance of the Bernoulli Naïve Bayes classifier in intrusion detection systems using recursive feature elimination with cross-validated selection of the best number of features," *Proc. Comput. Sci.*, vol. 190, pp. 564–570, Jan. 2021.
[47] S. Shah, P. S. Muhuri, X. Yuan, K. Roy, and P. Chatterjee, "Implementing a network intrusion detection system using semi-supervised support vector machine and random forest," in *Proc. ACM Southeast Conf.*, 2021, pp. 180–184.
[48] O. Almomani, "A feature selection model for network intrusion detection system based on PSO, GWO, FFA and GA algorithms," *Symmetry*, vol. 12, no. 6, p. 1046, Jun. 2020.

[49] I. S. Thaseen, J. S. Banu, K. Lavanya, M. R. Ghalib, and K. Abhishek, "An integrated intrusion detection system using correlation-based attribute selection and artificial neural network," *Trans. Emerg. Telecommun. Technol.*, vol. 32, no. 2, 2021, Art. no. e4014.
[50] F. Z. Belgrana, N. Benamrane, M. A. Hamaida, A. M. Chaabani, and A. Taleb-Ahmed, "Network intrusion detection system using neural network and condensed nearest neighbors with selection of NSL-KDD influencing features," in *Proc. IEEE Int. Conf. Internet Things Intell. Syst.*, Jan. 2021, pp. 23–29.
[51] H. Wang, Z. Cao, and B. Hong, "A network intrusion detection system based on convolutional neural network," *J. Intell. Fuzzy Syst.*, vol. 38, no. 6, pp. 7623–7637, Jun. 2020.
[52] K. Jiang, W. Wang, A. Wang, and H. Wu, "Network intrusion detection combined hybrid sampling with deep hierarchical network," *IEEE Access*, vol. 8, pp. 32464–32476, 2020.
[53] M. Safaldin, M. Otair, and L. Abualigah, "Improved binary gray wolf optimizer and SVM for intrusion detection system in wireless sensor networks," *J. Ambient Intell. Hum. Comput.*, vol. 12, no. 2, pp. 1559–1576, Feb. 2021.
[54] Z. K. Maseer, R. Yusof, N. Bahaman, S. A. Mostafa, and C. F. M. Foozy, "Benchmarking of machine learning for anomaly based intrusion detection systems in the CICIDS2017 dataset," *IEEE Access*, vol. 9, pp. 22351–22370, 2021.
[55] S. Ghribi, A. M. Makhlouf, and F. Zarai, "C-DIDS: A cooperative and distributed intrusion detection system in cloud environment," in *Proc. 14th Int. Wireless Commun. Mobile Comput. Conf.*, 2018, pp. 267–272.
[56] F. D. Vaca and Q. Niyaz, "An ensemble learning based Wi-Fi network intrusion detection system (WNIDS)," in *Proc. IEEE 17th Int. Symp. Netw. Comput. Appl. (NCA)*, Nov. 2018, pp. 1–5.
[57] R. A. R. Ashfaq, Y.-L. He, and D.-G. Chen, "Toward an efficient fuzziness based instance selection methodology for intrusion detection system," *Int. J. Mach. Learn. Cybern.*, vol. 8, no. 6, pp. 1767–1776, Dec. 2017.
[58] T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, "BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset," *IEEE Access*, vol. 8, pp. 29575–29585, 2020.
[59] J. Liu, Y. Gao, and F. Hu, "A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM," *Comput. Secur.*, vol. 106, Jul. 2021, Art. no. 102289.
[60] A. Fatani, A. Dahou, M. A. A. Al-Qaness, S. Lu, and M. A. A. Elaziz, "Advanced feature extraction and selection approach using deep learning and Aquila optimizer for IoT intrusion detection system," *Sensors*, vol. 22, no. 1, p. 140, Dec. 2021.

S. SHITHARTH (Member, IEEE) received the B.Tech. degree in information technology from the Kgisl Institute of Technology, Coimbatore, India, in affiliation with Anna University, Chennai, India, in 2012, and the M.E. and Ph.D. degrees in computer science and engineering from Anna University, Chennai. He is currently working as an Associate Professor with Kebri Dahar University, Ethiopia. He has published more than ten international journals along with 12 international and national conferences. He has even published three patents in IPR. He is also an Active Member of IEEE Computer Society and in five more professional bodies. His current research interests include cyber security, critical infrastructure and systems, network security, and ethical hacking. He is an active researcher, a reviewer, and an editor for many international journals.

PRAVIN R. KSHIRSAGAR (Senior Member, IEEE) is currently a Professor of artificial intelligence with the G. H. Raisoni College of Engineering, Nagpur, India. Previously, he served as the Vice Principal, the Dean of research and development, and the Head-ETC in the prominent Institute of India. He has vast experience of 18 years in teaching. He has served as a reviewer in many international journals, such as Inderscience, Spinger, Elsevier, and IEEE TRANSACTION. He has also delivered special talks in national and international conferences and chaired various technical sessions in international conferences. He has published various research articles in reputed journals. He has published more than 60 patents in national and international levels. He has six books in his credential. His research interests include data science, machine learning, artificial intelligence, and computer networks.

PRAVEEN KUMAR BALACHANDRAN (Member, IEEE) received the B.E. degree in electrical and electronics engineering and the M.E. and Ph.D. degrees in power systems engineering from Anna University, Chennai, India, in 2014, 2016, and 2019, respectively. He is currently working as an Associate Professor with the Department of Electrical and Electronics Engineering, Vardhaman College of Engineering, Hyderabad, India. He has published various research papers in reputed journals. He has filed seven patents and he has three book chapters in his credential. He is also an Active Member of IEEE Power and Energy Society and in five more professional bodies. His current research interests include solar photovoltaics, solar still, and renewable energy systems.

KHALED H. ALYOUBI received the Ph.D. degree in computer science from the Birkbeck University of London, U.K. He is currently an Associate Professor of computer science with the Faculty of Computing and Information Technology, King Abdulaziz University. His research interests include data sciences, data management, IR, data analytics, and data mining.

ALAA O. KHADIDOS received the B.S. degree in computer science from King Abdulaziz University, Jeddah, Saudi Arabia, in 2006, the M.Sc. degree in computer science from the University of Birmingham, Birmingham, U.K., in 2011, and the Ph.D. degree in computer science from the University of Warwick, Coventry, U.K., in 2017. He is currently an Assistant Professor with the Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University. His main research interests include the areas of computer vision, machine learning, optimization, and medical image analysis. He has authored several technical papers in these areas.