Hindawi
Security and Communication Networks
Volume 2021, Article ID 5561816, 23 pages
https://doi.org/10.1155/2021/5561816

Review Article

Comparing and Analyzing Applications of Intelligent Techniques in Cyberattack Detection

Priyanka Dixit,[1] **Rashi Kohli**,[2] **Angel Acevedo-Duque**,[3] **Romel Ramon Gonzalez-Diaz**,[4] **and Rutvij H. Jhaveri**[5]

[1]*RGPV University, Bhopal, Madhya Pradesh, India*
[2]*IEEE, New York, NY, USA*
[3]*Faculty of Business and Administration, Public Policy Observatory, Universidad Autónoma de Chile, Santiago, Chile*
[4]*Centro Internacional de Investigacióny Desarrollo (CIID), Monteria 230001, Colombia*
[5]*Pandit Deendayal Energy University, Gandhinagar, Gujarat, India*

Correspondence should be addressed to Angel Acevedo-Duque; angel.acevedo@uautonoma.cl

Received 21 February 2021; Revised 7 May 2021; Accepted 22 May 2021; Published 14 June 2021

Academic Editor: Muhammad Shafiq

Now a day's advancement in technology increases the use of automation, mobility, smart devices, and application over the Internet that can create serious problems for protection and the privacy of digital data and raised the global security issues. Therefore, the necessity of intelligent systems or techniques can prevent and protect the data over the network. Cyberattack is the most prominent problem of cybersecurity and now a challenging area of research for scientists and researchers. These attacks may destroy data, system, and resources and sometimes may damage the whole network. Previously numerous traditional techniques were used for the detection and mitigation of cyberattack, but the techniques are not efficient for new attacks. Today's machine learning and metaheuristic techniques are popularly applied in different areas to achieve efficient computation and fast processing of complex data of the network. This paper is discussing the improvements and enhancement of security models, frameworks for the detection of cyberattacks, and prevention by using different machine learning and optimization techniques in the domain of cybersecurity. This paper is focused on the literature of different metaheuristic algorithms for optimal feature selection and machine learning techniques for the classification of attacks, and some of the prominent algorithms such as GA, evolutionary, PSO, machine learning, and others are discussed in detail. This study provides descriptions and tutorials that can be referred from various literature citations, references, or latest research papers. The techniques discussed are efficiently applied with high performance for detection, mitigation, and identification of cyberattacks and provide a security mechanism over the network. Hence, this survey presents the description of various existing intelligent techniques, attack datasets, different observations, and comparative studies in detail.

1. Introduction

An excessive use of the Internet in various areas are encouraging the researchers and scientists to use intelligent systems that can support the users, and different applications also ensure efficient computation, maintaining the quality of service over the network. The traditional methods were time-consuming, less efficient, and giving average performance and cannot fit for providing the solution for complex, multiobjective, or real-world problems. Hence, the necessity of efficient attack detection systems can reduce the harmful effects of cyber threats [1]. Cybersecurity is the collection of various technologies and security mechanisms that develop for the protection of data, information, network, a program from the different attack activities such as data modification, stealing, unauthorized access, and destruction over the Internet or network. Cybersecurity components concern mainly host protection and network security systems [2]. Currently it is used to protect many areas such as cloud computing [3], wireless sensor network [4], and IoT. There are a lot of security measures which are available for providing security to the systems or networks such as antivirus,

firewall, and IDS. However, still, cyber threats continuously harm and disrupt Internet services every day. This motivates many researchers for providing their extensive contribution to design security systems [5–9].

The following are the popular cyberattacks such as denial of service attack, distributed denial of service attack [10], remote to local attack, probing, user to root attack, adversarial attacks, poisoning and evasion attacks [11], botnet [7], phishing attack [12], spamming [13], and zero-day attack [6].

Many different methods are used for attack detection that broadly categorized into three major categories such as anomaly-, misuse-, and hybrid-based detection. Misuse-based detection can be scanned by prestored attack signatures and mostly used to detect identified attacks. It is useful to detect known attacks with minimum false alarms. It requires a certain modification of the signature and rules of attacks on the database.

The anomaly-based technique is capable to detect both the attack types either known or unknown. It can capture network and host machine behavior and also determines anomalies as deriving from normal behavior. It is the most popular method because it can detect zero-day attacks. There are many merits of using this method, and one of them is the customization of profiling actions due to which attackers get confused about which activity they follow to enter and remain undetected. However, besides the merits, there is a drawback also it evaluates with very high false alarm rates and sometimes the legitimate activity considered as an anomaly.

Another is a hybrid technique, a fusion of anomaly- and misuse-based detection. It supports high performance in the detection phase and a minimum false alarm rate.

Here presents some existing research that is lighting the contribution of machine learning and metaheuristic techniques in cyberattack detection, especially focusing on better classification and optimal feature extraction also with their results. Tu et al. [14] proposed hybridization of PSO and SVM for feature extraction, and in this method, fitness function of PSO is used for classification.

Athari and Borna [15] proposed a hybrid metaheuristic particle swarm intelligence, genetic algorithm (GA), and glowworm which are collectively used for classification and optimal feature extraction in the wireless sensor network. The purpose of using the metaheuristic algorithm is to solve the problems such as low convergence and low local optimality. In this paper, certain parameters were calculated as permittivity against DoS attack, reliability, number of active nodes, and energy consumption. The results shown by PSO have the highest permittivity, reliability, and larger number of active nodes compared to the genetic algorithm and glowworm optimization technique for DoS attack, GA has less permittivity, reliability, and number of active nodes than PSO and GSO, and energy consumption of GA is very low compared with the above two techniques because of its simplicity. The bioinspired algorithms are popularly used for the optimal feature selection and solving optimization problems in different fields compared with data mining techniques that were previously used in classification and feature selection in different applications such as pattern recognition, intrusion detection, clustering, and data classification.

Sagarin and Taylor [16] proposed a biological evolutionary system for providing better approaches in the field of security. Jamali and Shaker [17] proposed a metaheuristic approach for recognizing denial of service attack type on TCP protocol called TCP SYN flood attack that requests for TCP connections in the form of huge flood request to the server. This attack detection framework is designed by particle swarm optimization (PSO) and the queuing model for optimally using the buffer space and solving attack recognition problem over the network.

Tarao and Okamato [18] used an artificial immune algorithm of the metaheuristic family to model the framework for DoS attack detection to overcome the vulnerabilities of the server-side. By this technique, the false alarm rate is minimized and detection performance is simulated by the machine learning approach. Metaheuristic algorithms are very efficiently used in cybersecurity for the implementation of the attack recognition framework with high learning capabilities. Bhattacharya et al. [19] proposed a hybrid principal component analysis and firefly-based model to classify intrusion detection system datasets. The model performs one-hot encoding for the transformation of the attack datasets and then hybrid PCA-firefly algorithm used for dimensional reduction. The another XGBoost algorithm is used for classification of attacks. This hybrid model perform well by achieving high accuracy of 99.9, sensitivity 93.1, and specificity 99.9.

Visumathi and Shunmuganathan [20] proposed intelligent computational techniques such as SOM, SVM, multilayer perceptron (MLP), Bayesian network (BN), and logistic regression for the classification of attack data. Srinoy [21] proposed a hybrid combination of particle swarm intelligence for an optimized feature selection and support vector machine (SVM) that classify attack data. After the result is evaluated, it had been found that that the above-mentioned hybrid technique can easily identify not only known attacks but also detect the early apprehensive activities that cause unknown attack. This method is efficiently solved feature selection problem and achieved detection rate of 96.11% with high classification accuracy.

Mourougan and Aramudhan [22] proposed a computational model for solving classification problems and extracting features by the hybrid combination of the PSO technique and the GA algorithm. The proposed model of attack detection can identify DoS attack with maximum detection accuracy and minimum false alarms by genetic particle swarm intelligence-based binding feature extraction that is mostly used for intrusion feature selection. Results are shown with maximum accuracy and minimum false alarms as compared with the fuzzy clustering technique.

Akyazi and Sima Uyar [23] proposed the model which was built on an anomaly-based intrusion detection method and using the artificial immune system (AIS) to improve multiobjective evolutionary algorithm, to get the better performance of the proposed model for the detection of DDoS attack tested on the DARPA-based LLDOS 1.0 dataset. The proposed model is applied iteratively for computation, and if we find the negative selection, then we redefine the objectives with the same concept. The zero-

percent false positive rate is found by applying the above approach, and hence, results show that the method is successful with better accuracy.

Ben Sujitha and Kavitha [24] proposed a method that can efficiently detect cyberattack and provide better accuracy and efficiency. For performance improvement in the attack detection model, optimized features selection approach was used. The proposed system is built to provide an optimal feature extraction algorithm to construct summarized features applied to the multiobjective PSO algorithm. The anomaly detection method was applied, and the proposed system was tested on the KDDCUP99 intrusion dataset. The result of the proposed system shows that it can successfully deal with real-time attacks worked with high speed.

The rest of the paper is arranged as follows: Section 2 focuses on important steps of classification and machine learning. Section 3 focuses on the detail of different metaheuristic algorithms used in cyberattack detection. Section 4 describes the different machine learning techniques used in cyberattack detection. Section 5 focuses on different datasets. Section 6 presents observations and evaluations. Section 7 presents challenges and future directions, and finally, Section 8 concludes the paper.

2. Important Steps of Classification and Machine Learning

Many techniques were previously used in knowledge discovery of database (KDD), especially data mining techniques such as clustering and data classification techniques. KDD is dealing with extracting useful information from the data source. In Figure 1, the various steps for extracting knowledge are data preparation, data selection, data cleaning, and extracting features or patterns from the data. According to Periyar and Salem [22], data preprocessing is a most essential step of machine learning computation that can remove noisy data such as repeated values, out-of-limit values, irrelevant data logics, checking null values, and missing terms or instances. The data preprocessing has certain steps such as learning, normalization, transformation, feature selection, and extraction. Outcomes of preprocessed data are input or works as the training sets to extract knowledge for the testing phase. The precision of any classifier depends on selecting the optimal feature upsets from the original data [22].

Feature selection is the most essential step of data preprocessing, used before the classification process [24]. This method is useful to reduction of some repeated data patterns and noisy and unnecessary features, which is very useful to achieve accuracy in classification and improves attack detection rate. It is the method of selecting some subset of the actual features and can generate different new features [24]. FS has to perform two basic objectives firstly to provide accuracy in classification performance and reduce the number of features. Complex datasets sometimes degrade classification performance in the attack detection process, and it can create problems such as irrelevant data and repeated features, uncertainty, and ambiguity. These certain problems are obstacles not only in concern of

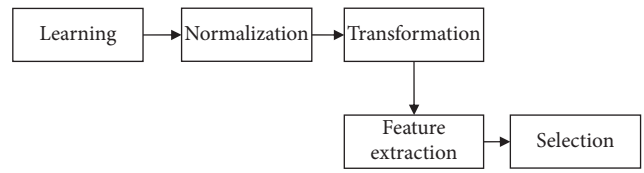

Figure 1: Steps of preprocessing of data.

detection speed but also in the performance of the detection process [24].

The approach comprises two major phases which are training phase and testing phase, and these phases are processed by using the following steps:

(1) Identification of different features, attributes, or classes of data during the phase of reprocessing these attributes which are extracted from the data

(2) Selection of attributes that is useful for the classification

(3) Learning processed by the help of training data

(4) Training the model used for the detection of unknown threats

These abovementioned are the various steps that are followed to process machine learning. In the training phase, signature-based classes are learned by using some training sets. In the testing phase, testing of new data is carried out by the classifier and they are checked whether they match with that class or not.

In another anomaly-based approach, the regular traffic data are defined in the training phase, this trained model is applied to the new data in the testing phase, and finally, testing sets are classified as a normal or malicious one.

In many research papers, machine learning (ML) is broadly categorised into three phases such as the first training phase, secondly testing phase, and the last validation phase. Machine learning has numerous methods for the training and testing process, some of the popular methods are artificial neural network (ANN) methods such as SVM, SOM, and multiple layer perceptron network (MLP), and these techniques have different parameters such as the number of layers, nodes, and processing units. When the training phase is completed, then a number of models are available and the selection of the model depends on its efficiency, accuracy, and error estimation. .

There are the following three types of ML methods which are broadly classified as supervised, unsupervised, and semisupervised [2]. When the model is trained by certain rules (training sets) and the data are well labeled, then it comes under supervised learning. Most of the supervised anomaly techniques were proposed using a support vector machine (SVM), multilevel perceptron network (MLP), and decision tree [25].

When some part of the dataset is labeled by preprocessing of data methods due to which the problems introduce, that comes under semisupervised learning. If the

dataset is unlabeled, then some problems arise in extracting various attributes, classes, structures, and patterns from those data, and such problems come under unsupervised learning [2]. Which is the model or leaning method used depends on the problem that should be solved. Hence, according to the problem, the best-suited learning approach is used.

Once the steps of the classification model are completed such as training, validation, and testing sets, hence the model is able to be preferred in the future for further problem-solving strategies. There are many machine learning methods which are available for solving any classification problem efficiently such as artificial neural network learning methods, both supervised or unsupervised learning techniques such as self-organizing map, linear logistic regression, and other feedforward neural network methods, naïve Bayes, support vector machine (SVM), and multilevel perceptron (MLP) classifiers. These different methods were applied to different benchmarks and popularly used for solving classification problems such as well-known KDDCUP99 dataset for intrusion detection according to Srinoy et al. [21] using the anomaly-based approach with a hybrid form of PSO and SVM for the optimal feature selection and classification tasks.

According to Shinde and Parvat [26], using the NSL-KDD dataset, we apply the hybrid form of PSO and ABC on SVM for solving feature selection and classification problems to achieve high DR and low FAR. Prasad et al. [27] analyzed metaheuristic anomaly-based algorithms for real-time detection of application layer distributed denial of service attack successfully detected by using the hybrid combination of cuckoo search, bat, and firefly algorithm and proved to be an efficient technique by improving the parameters such as accuracy, efficiency, and performance analysis. Jadidi et al. [28] proposed multilevel perceptron (MLP) based on the anomaly attack detection method in a high-speed network. The PSOGSA and cuckoo algorithms based hybrid approach was used that ensures improved accuracy to classify abnormal traffic.

Akyazi and Sima Uyar [23] proposed a model for attack detection against the DoS attack by using the AIS algorithm based on anomaly detection and applied on the DARPA LLDOS 1.0 dataset that provides an efficient result, high TPR, and very low FPR. Hence, a multiobjective evolutionary algorithm is used inspired by AIS that is proved to be very effective for DDoS attack detection. Hence, machine learning methods are very popularly used for cyberattack detection and proved to be very efficient on various benchmarks. Today, for the better computational result, hybrid metaheuristic algorithms and ML approaches are used for optimal feature extraction, and in many classification problems, they specially deal with complex datasets.

In the above section of the paper discussing machine learning approaches with some current research studies, now after preprocessing, feature extraction, and classification of the data model we talk about the computational matrices of classification. There are several classification matrices which are used for machine learning in the attack-detection process. These matrices are discussed below in this part. The evaluation can be done on four main parameters such as false positive (FP) attacks which are wrongly classified as attacks, true positive (TP) which shows that attacks are correctly classified, true negative (Tn) which shows that the system is correct in spotting normal conditions, and false negative (Fn) attacks which are correctly classified as attacks [29]. The following are the attack detection matrices based on the anomaly detection method:

(i) To measure the overall performance, four matrices are broadly used such as accuracy, error rate (ER), miss rate (MR), and false alarm rate (FAR) [28, 30].

(ii) Accuracy can be measured as

$$A = (TP + Tn)/(TP + Tn + FP + Fn)$$

$$\text{Error rate, ER} = (Fn + FP)/(TP + Tn + FP + Fn)$$

$$\text{Miss rate, MR} = (Fn)/(TP + Fn)$$

$$\text{False alarm rate, FAR} = (FP)/(Tn + Fp).$$

(iii) True negative rate, also called specificity, TNR = (Tn)/(Tn + Tp) ratio of items which are correctly classified as negative [2].

(iv) True positive rate, also called as sensitivity or recall or detection rate, TPR (Tp)/(Tp + Fn) [2].

(v) Negative predictive value (NPV) ratio of items which are correctly classified as negative, NPV = (Tn)/(Tn + Fn) [2].

(vi) FP rate or fall out rate ratio of items incorrectly classified as positive fall out = (FP)/(Tn + Fp) [2].

In the attack detection mechanism during the classification of data, certain metrics are evaluated as false alarm rate (FAR) and true positive rate (TPR). Both of the abovementioned matrices are directly proportional to relationships with each other. Both of FAR and TPR are plotted with the help of receiver-operating characteristics with different axes, x-axis for FAR and y-axis for TPR, when FAR increases, then TPR increases, and if FAS falls, the TPR falls [18]. The overall performance is measured by certain matrices such as total detection accuracy (TDA) that can be evaluated as a total sum of correctly classified data items to the sum of samples. The average detection time (ADT) is calculated as the total detection time to the total sum of samples [31], performance, class detection rate, detection rate, or false positive rate.

The performance matrices are also evaluated by recall or precision factors. Recall is measured as TP/TP + Fn and precision factor is measured as TP/(TP + Fp), other matrices measured based on precision and recall are F-measures, and weight mean acts as a tradeoff between the above two. The F-measures was measured as (2 * Recall * Precision)/(Recall + Precision) [32].

3. Metaheuristic Approaches for Optimal Feature Selection

In this section of the paper, we discuss various methods of metaheuristics that are broadly used in many areas for solving different complex optimization problems [5].

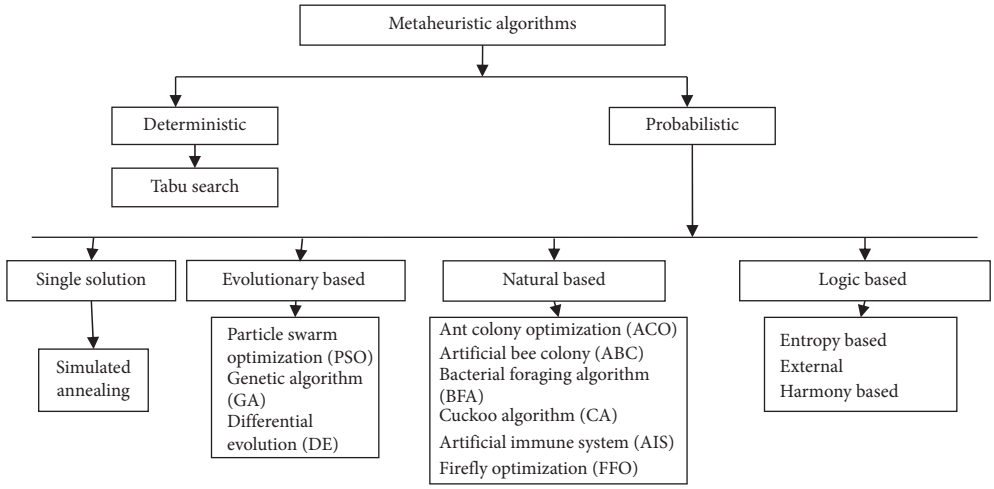

Figure 2: Categories of metaheuristic algorithms.

Figure 2 presents the classification of optimization such as swarm optimization techniques, genetic algorithm (GA), ant colony optimization (ACO), artificial immune algorithm (AIA), cuckoo algorithm (CA), artificial bee colony algorithm (ABC), and bacterial foraging algorithm (BFA), with their results by using the reference of different research papers and latest articles.

3.1. Particle Swarm Optimization. It is a widely used technique and one of the members of the swarm family which was firstly discussed by James Kennedy and Russell C. Eberhart in 1995 [14], and PSO is described in his first research paper "*A New Optimizer Using Particle Swarm Theory*." It is an intelligent optimization technique and a member of a class called metaheuristics. Particle swarm intelligence is stimulated by the socialized behavior of animals such as bird flocking and fish schooling nature.

Particle swarm intelligence is a simple yet powerful optimizing algorithm and also successfully applied to the number of applications of different areas and broadly in fields of science and engineering. According to its concept, each solution in the search space is considered a particle. Taking information from its surrounding particle will be in motion [15]. Its prototype can be implemented with ease programming and economical in terms of both storage and speed. Its computation and working steps are similar to the evolutionary and genetic algorithms [17]. PSO has the strong global best minima, each particle of the population has some randomized positions, and every particle is attached with some velocity. The velocity of the particle is adjusted through some previous behavior of each particle and its neighbors while roaming around the search space. All the particles update their positions and velocity to the best optimal value by communicating with each other [24].

Each particle in the problem space has certain coordinates that are connected along with some fitness value. This value is noted as P best. The other value also used by the PSO is considered as the best value that occurred up to any particle in the neighbors of the particle. Then, the particular location of the particle is noted as L best. As soon as a particle takes the population of its surrounding neighbors, the best value, called global best, is considered as best. Hence around this theory of particle swarm optimization every, time there is a change in velocity of each particle that value near its P best and L best locationsis selected [25, 33].

Let us consider that each particle is categorized into two parameters which are position vector and velocity vector, denoted as $(Y_i(t))$ and $(U_i(t))$. The location and velocity of particle i at the iteration of t times can be presented as

$$Y_i(t) = \left[Y_{i1}(t), Y_{i2}(t), \ldots\ldots, Y_{in}(t)\right], \tag{1}$$

$$U_i(t) = \left[U_{i1}(t), U_{i2}(t), \ldots\ldots, U_{in}(t)\right]. \tag{2}$$

Hence, the performance of the process, each particle having its independent knowledge P best, means that they have their own best value in the position, and collective knowledge G best means best of its best neighbor. The velocity will be updated using formula (3) [33–35]. The velocity of particles can be calculated as

$$\begin{aligned} U(\text{new}) &= w \times u_{\text{pd}}^{\text{old}} + C_1 \times \text{rand}_1 \times \text{pbest}_{\text{pd}} \\ &= y_{\text{pd}}^{\text{old}} + C_2 \times \text{rand} \times \text{gbest} - y^{\text{old}}, \end{aligned} \tag{3}$$

where w is inertia weight and is denoted as random number that may be considered between 0 and 1, and C1 and C2 are the constant value that can change the velocity of a particle towards the P best and G best, and its value can be set to 2 [25, 33, 34]. Hence, equation (4) represents update positions as

$$Y_i(t+1) = Y_i(t) + U_i(t+1). \tag{4}$$

Particle swarm optimization is applicable to various cyberattack detection systems for optimal feature selection problems and providing optimal solution. Here, we discuss some cyberattack detection systems designed in reference papers presented by various authors that use particle optimization for optimal results. Jamali and Shaker [17]

proposed a detection model against denial of service attacks which are very prominent attacks over the Internet that block the legitimate users for gaining the network services. A defensive framework is proposed which had used particle swarm optimization to formulate various optimization problems to optimally solving the problem and improving the performance of an attack detection system with efficient consumption of buffer space.

In Hao et al.'s study [36] for achieving pattern recognition on weblog data for differentiating the normal and abnormal or malicious data, the user sessions are extracted from the log record. *k*-means clustering techniques are applied with a hybrid form of particle swarm optimization to generate an efficient attack detection model. This model successfully recognizes a DDoS attack with improved accuracy.

Momanyi Nyabuga et al. [37] provide defensive and preventive models against the denial of service (DoS) attack by applying the particle swarm intelligence algorithm in the VANET network. After the results, they found that PSO is efficient with high accuracy used for optimization in the attack detection system.

Shinde and Parvat [26] proposed a framework focused on achieving a high rate of attack detection with minimum false alarms by applying a hybrid form of SVM and swarm intelligence for selecting correct parameters. SVM had enabled us to provide an efficient classification of attack data. The attack detection framework uses knowledge gaining for selecting features and combines with support vector machine classifier. The important features will be selected by using an optimization approach called particle swarm intelligence and applied on the NSL-KDD dataset, and outcomes found high DR and low FAR after compared by regular SVM.

Guoli [38] presented the attack detection model based on which PSO is proposed with the Elman neural network. This fusion is used for optimal parameterization which improves performance. The experimented results are better as compared with traditional techniques.

3.2. Genetic Algorithm. **G**enetic algorithm (GA) is another metaheuristic technique that works on the concept of theory of evolution. The genetic algorithm (GA) is also called as population-based algorithms. These bioinspired algorithms are based on the iterative or repeating operations, and its basic ideology is adapted from genetics. Genetic algorithms can be designed as a simulation model in which the population of samples (chromosomes) from solution candidates in optimization problems will lead to an improved solution. One of the main features of the genetic algorithm is that it constantly works on chromosomes and solution space [15].

The population is mainly a collection of chromosomes in which each chromosome represents a certain position in the problem domain and probably a solution to the problem. By applying genetic operators on each population, they result in creating a new population that can have the same number of chromosomes. Then, a fitness function is determined for them, using operators are selection, crossover, and mutation, these are generally used by genetic algorithms, and a new generation can be created. The number of generations such as chromosome population is determined in the algorithm initialization step and methods of setting the parameter.

In the selection phase of GA, several chromosomes are selected from the existing chromosomes in a population to reproduction. Best chromosomes have more chances to be selected for reproduction. Chromosomes that make the next generation are being selected by this operator:

$$p_i = \frac{f_i}{\sum_{j=1}^{\text{pop-size}} f_j}, \tag{5}$$

$$l_i = \pi r^2 * p_i. \tag{6}$$

Equation (5) represents how to calculate the probability of each chromosome selection. p_i represents the probability of selection of ith chromosome, f_i is the fitness function value of the ith chromosome, and the dominator part of the equation shows the total amount of fitness function of all chromosomes. In equation (6), l_i is the length of the ith chromosome. Then, the crossover operator produces the child chromosome. This operator produces two-parent chromosome genes in the new chromosome (child). The chromosomes which were selected from the initial population as a parent for crossover operation are obtained from

$$n_c = 2\left[\frac{p_c \times n_{\text{pop}}}{2}\right]. \tag{7}$$

The mutation operator also selects a gene from the chromosome arbitrarily and then alters the content of that gene. The mutation operator guarantees that the genetic algorithm does not fall in the trap of local minimum point and covers all chromosomes which may be destroyed during the performance of other operations such as selection and crossover.

Genetic algorithms are based on the global search optima, and hence, they are efficiently used in the attack detection system; some of the following researchers use this technique for solving the optimization problem in attack data classification.

Siva Sankari et al. [39] proposed a model for the detection of a DoS attack also to observe the attacks over the Internet and predict the attack is DoS or not. In the proposed model, the genetic algorithm (GA) is used for optimizing features for optimal feature selection and identify DoS attacks. The GA is capable to learn the things itself and initiates the process of selection. It is used to generate optimal resolution for making the proper solution to complex problems. Genetic algorithms are implemented through the following steps such as selection, crossover, and mutation to find the optimal solutions. This approach is very accurate and efficient for identifying a DoS attack. The proposed system results showed better performance, and it is capable of detecting a DoS attack with high accuracy.

Mizukoshi and Munetomo [40] proposed the system which is designed for attack detection by learning the attack

patterns and other anomalous traffic. The proposed system works as a real-time traffic pattern analyzer using GA for detecting abnormal traffic behavior. The system is built on using Hadoop distributed infrastructure, and the result shows the effectiveness of the DDoS defense system.

Lee et al. [41] proposed an approach which provides a defensive mechanism against DDoS attack by using a traffic matrix. In this work, they proposed an improved attack detection model that enhances the traffic matrix construction process and some particular parameters were optimized using the genetic algorithm (GA). The experiments were tested on DARPA 2000 and LBL-PKT-4 datasets, and the results were evaluated which provides better detection accuracy with a high speed as compared with previous techniques.

Bhuyan et al. [42] provide an inclusive survey on detection or prevention against DDoS attack and also its detection techniques with its tools in the different networks. The article also discussed the different issues, various challenges, and feasible solutions in the concerned domain.

In Dimitris et al.'s study [43], the neural network detector was designed against the detection of DDoS attack. For selecting optimal features, a genetic algorithm is used that can extract 44 statistical features from the packet header. The computation is based on a genetic algorithm that creates an error-free neural network-based DDoS detector. The experimental results have shown the improved succeed features for DDoS detection with high accuracy.

Lee et al. [44] proposed an attack detection model by improving some parameters of traffic matrix through GA to achieve optimization that utilizes a high attack detection rate. The traffic matrix construction operation improved by hash function for minimizing the rate of collisions also used the packet-based window size to minimize cost. The evaluation is applied on DARPA 2000 LLDOS 1.0 and LBL-PKT-4 attack datasets. The proposed work has shown high feasibility in concern of attack detection accuracy and speed.

3.3. Ant Colony Optimization Algorithm. It is a commonly used technique to resolve combinational optimization-based problems and belongs as a member of the metaheuristic family. This algorithm works as an agent-based system, simulates the behavior of ants to develop a learning-based system. The ants preferred to move in a straight line for food searching and protecting themselves from different situations, and firstly, they decide to move from left to right randomly. Then, some assumptions are taken such as the moving speed of each ant is the same and also depositing pheromone in the trail evenly. Hence, the ants prefer to move from left to right direction and will reach the food earlier, and pheromone accomplished the fast shortest path around the obstacles. While the other ants preferred to follow the way where they found the excess amount of chemical called pheromone, hence all the ants meet the target (source of food) through the shortest path.

The ant colony optimization is quite different from the traditional ant system in concern with the pheromone trails which can be updated in two phases. Firstly, when ants decide a tour, they can change the quantity of pheromone locally around routed boundaries by a local updation in position. Secondly, when each of the ants decides their tour, a global updation is applied to adjust the pheromone amount in the boundaries which is considered as best ant tour [21]. Hence, this phenomenon of optimization is used by different research studies for solving optimization problems. Along with the various applications of ant colony optimization, it is also used in cyberattack detection models successfully. Here, we discuss some of the research papers of its contribution in attack detection models. Dimitris et al. [43] proposed an ant colony system-based (DDIACS) framework for identification and detection of a low-rate distributed denial of service (LDDoS) attack detection, another well-known attack over the network. The proposed detection model is built with ant colony optimization, which is another strong optimization algorithm used to resolve complex optimization problems. The proposed framework has improved some parameters that are very complex while detecting multisource attacks such as flexibility, fast convergence, and robustness. This framework was tested upon the dataset DARPA and KDD. The outcomes have shown that the proposed method has successfully overcome the problem or errors with high accuracy than existing models. The proposed model found more than 89% of the detection rate and 83% accuracy.

Aldwairi et al. [45] proposed an anomaly-based detection model for the detection of unknown attacks. They proposed a model by using the ant colony optimization technique for selecting optimized features to improve the overall classification accuracy by rejecting unwanted features. In this proposed work, ant colony optimization of three levels of updating feature selection process had been proposed. This method efficiently used the information of each ant in the process of feature extraction and also improved the accuracy of the proposed system and classification of features. The evaluation results have shown that the proposed approach performed well as compared with previously used feature selection techniques.

3.4. Artificial Bee Colony. Swarm intelligence is a kind of self-organized system that can solve different optimization problems. Artificial bee colony (ABC) is another prominent optimization technique that works by the concept of imitating the foraging technique of bee swarms, firstly predicted by Visumathi and Shunmuganathan [20]. The artificial bee colony algorithm worked in the following three basic steps: the first is food source; it is based on some important factors such as the amount and quality of nectar, the total efforts for its extraction, and nearest to the colony. Secondly, foragers, employed foragers grasp information of food sources, and the third one is unemployed foragers continuously looking for food sources and broadly categorised into two types which are scout bees and onlooker bees. The whole process of searching food starts with scout bees sent for searching food sources in the colony in a random distribution manner. While the scout bees return, the food sources are rated by some threshold value and perform waggle dance [21]. The waggle dance is a unique interaction way and also helps to

determine the food source direction through the nectar amount that represents fitness value. Onlooker bees choose the best food source by collecting all the information that is exposed by the waggle dances. This information helps them to reach the best sources without the help of any maps [46].

The following authors used artificial bee colony (ABC) in the attack-detecting system for solving optimization problems. Mahale and Gothawal [47] proposed an ABC algorithm to optimize some attributes of the artificial neural network, improve local optima problem, and also overcome low convergence speed of the neural network. The ABC algorithm can be efficiently used for finding the optimal solutions in minimum time. In this research work, the proposed algorithm was applied for attack detection and the evaluation outcome shows that the proposed method had performed and improved in some parameters such as DR and efficiency.

Priyadharshini and Kuppusamy [48] proposed an attack detection model based on anomaly detection techniques that detect attacks and improve performance by low false alarms. This proposed work is based on the ABC algorithm by anomaly-based attack detection with a feature extraction technique to optimize some attributes for the classification. The experiments were performed on the KDDCUP99 dataset, and results were evaluated by calculating some parameters such as accuracy and speed. After evaluation, the accuracy rate was noted 97.5% for the known attack, and for unknown attack, it was noted as 93.2%.

3.5. Cuckoo Algorithm. A cuckoo algorithm is one of the optimal search algorithms inspired by the holoparasite act of cuckoo birds. The birds of these types are not able to complete their reproduction phase by lacking proper host, and these birds can lay their eggs to the nest of the birds that contain eggs that look like them, which means they place their eggs inside the nest of other similar birds. The searching approach followed by the bird is acceptable in different areas for solving different optimization problems. Cuckoo search is applied with three traditional rules: firstly, randomly search location of host nest for placing eggs; secondly, the nest that contains similar eggs as compared to a cuckoo egg; third, the finite number of nests that is considered as 15 for cuckoo search. Hence, the probability P can be taken for its eggs as an object which is represented as $\{P\ (a)\exists a\in(0, 1)\}$. The following authors used the cuckoo algorithm to achieve optimization in the attack detection model.

Hao et al. [49] proposed security against denial of service attack by using a cross-layer approach as the best solution. The cross-layer approach was the combined form of device-driver packet filter (cuckoo-based filter) and remotely firewall. Packet filter was designed to filter out abnormal network traffic before it utilizes the resource for higher network protocol layers at a server-side. The performance of the proposed technique was checked through wide-ranging simulated by java and performs better for DDoS attack detection.

3.6. Bacterial Foraging Algorithm. BFO technique is stimulated by a collection of forage behavior of bacteria such as *E. coli* and *M. xanthus*. Particularly, the BFOA algorithm based on the chemotaxis behavior of bacteria can determine chemical gradients and move toward or away from particular signals. The information-conveying process of the algorithm is used to allow cells to collect together swarm to optima. This can be implemented by a sequence of three main processes on a population of replicated cells: first, chemotaxis; second, reproduction; and last, elimination-dispersal. These first steps are responsible for the cost of cells is redefined through the closeness of other cells, and they can move along the modified cost surface at once. The second one in which only those cells are preferred performs best in their whole life that allows being the part of next generation, and in the third one, the cells may discard and low probability new random samples are added.

The following optimization algorithm is efficiently used for cyberattack detection mechanism. Damodaram and Valarmathi [50] applied the bacterial foraging algorithm for the detection of phishing attacks. The traditional systems are intelligent, flexible, and efficient based on association and classification of data mining algorithms, but they are not successful to provide the optimal solution. The proposed model introduced a hybrid optimization algorithm BFOA for achieving an optimal solution for identifying phishing websites. Experimental results were compared with the traditional techniques proved to be very efficient by comparison. Table 1 shows the comprehensive analysis including techniques, datasets, description, and outcomes of different articles from the literature. Table 2 presents the brief description of different metaheuristic techniques with their features and application.

4. Machine Learning Methods

In our day-to-day life, artificial intelligence plays an important role to solve many complex problems. It includes many applications such as speech recognition, language processing, machine intelligence, and fog computing [53, 54]. Machine learning is one of the popular fields of artificial intelligence that is successfully used in solving various computational problems of different areas [56, 57]. Now a days it is extended to more deep networks such as deep learning [58], extreme learning [59], deep extreme learning networks etc.

Machine learning algorithms are classified as "classification," "clustering," or "regression."

This section of the paper discusses various methods of machine learning used in an attack detection system. Here, certain details of these techniques with their results are presented by taking the help of different research papers for each method. In Figure 3, classification of machine learning techniques such as decision trees (DTs), artificial neural networks (ANNs), naive Bayes (NB), and fuzzy set-based approach are referred from the previous literature survey.

The paper presents a detailed study of some important intelligent classification techniques are discussed below.

4.1. Artificial Neural Networks (ANNs). ANN is among the efficiently used systems that stimulated its working like the human brain [1]. ANN works like the human brain which

Table 1: Detailed analysis of comprehensive survey and research articles.

Reference	Technique used	Dataset used	Description	Outcomes
Hao et al. [36]	*The hybrid form of k-means + PSO*	*KDDCUP99*	The proposed model can be used to detect the crowd (undetermined session) is normal or an attack.	The proposed model can detect attacks with better performance.
Momanyi Nyabuga et al. [37]	*Particle swarm optimization*	*KDDCUP99*	The proposed model provides a review and discussions of the denial of service attack detection and prevention mechanisms; moreover, it intended to propose the particle swarm algorithm optimally helps to detect DOS attack.	The simulated outcomes have shown that the proposed PSO-based model was efficiently used for attack detection as compared with other methods.
Shinde and Parvat [26]	*The hybrid form of PSO + SVM*	*NSL-KDD*	The attack detection model was designed using a hybrid form of SVM machine with the PSO technique for the selection of optimal features to achieve high accuracy and performance also lower the FAR alarm than normal IDS.	The hybrid approach of machine learning and optimization technique (ABC-SVM) provides better results than the other single approach. The results showed a detection rate with 98.53% and a false alarm rate with 0.0374.
Siva Sankari1 et al. [39]	*Genetic algorithms*	*KDDCUP99*	The proposed model is designed by using the genetic algorithm (GA) for the detection of DoS.	This detection approach was better-performed attack detection but not proved to be very efficient as comparing its performance with the hybrid technique approached model. However, it provides better results than the traditional one.
Mizukoshi and Munetomo [40]	*Genetic algorithms*	*KDDCUP99*	This proposed model is based on real-time traffic pattern analysis using a genetic algorithm (GA) approach for optimal pattern extraction.	The experimental result has shown that the proposed method performed well as compared with other traditional methods.
Lee et al. [41]	*Genetic algorithms*	*DARPA 2000, LBL-PKT-4*	This proposed model is designed for the detection of distributed denial of service attack using a traffic matrix and optimizes some features of the traffic matrix by using GA.	The detection rate and accuracy by using this method were better compared with other traditional techniques.
Dimitris et al. [43]	*Genetic algorithms*	*KDDCUP99*	This proposed work is designed for the detection of DDoS attacks using a genetic algorithm for efficient feature selection and optimizing some parameters. Genetic algorithm (GA) evaluation used designed error-free neural network detector.	The evaluated results have shown that the features that best qualify for DDoS attack detection were optimally selected by the proposed approach and provide better results.
Chen et.al. [51]	*Ant colony optimization*	*DARPA/ LLDOS KDDCUP99*	This proposed work investigated different complexity of the DDIACS framework and also presents its comparison with the swarm technique and other probability-based techniques.	The results have shown that the proposed framework successfully resolved the problems related to processing attributes, and DDIACS framework provides higher performance than existing methods.
Kumar and Walia [52]	*Ant colony optimization*	*KDDCUP99*	The objective of this work was to design and implement OSLR and DSR protocols for the blackhole attack also prevent the system from the threat.	After evaluation, results showed that the proposed approach performed well on various network performance metrics such as bit error rate, throughput, delay, and packet delivery ratio.
Rais and Mehmood [53]	*Ant colony optimization*	*KDDCUP99*	The proposed model used the ACO optimization technique for better feature selection by various stages of pheromones that help ants to find the optimal features.	Evaluation of the result shows that the proposed approach outperformed in optimal feature selection as compared with the traditional techniques.
Bhuyan et al. [42]	*Artificial bee colony*	*KDDCUP99*	This proposed method is applied to ABC algorithm. Anomaly-based attack detection is used by using different feature selection techniques to minimize the number of unwanted features and pick the best one.	Experimental results have shown that the performance of ABC algorithm was better than traditional approaches and also achieved a high accuracy rate.

Table 2: Comparative study of metaheuristic algorithms.

Features	*PSO*	*GA*	*ACO*	*AIS*	*ABC*	*BFA*
Representation	Dimensional vector for position speed, the best state	Binary, real list of rules, permutation of elements	Undirected graph	Attribute str. (a real-valued vector), integer string, binary string symbolic string	D-dimensional vector (xi = 1, 2, . . ., *D*)	Represents *i*-th bacterium at *j*th chemotactic, *k*-th reproductive, and *l*-th dispersal step
Operators	initializer, update, and evaluator	Crossover, mutation, selection, inversion	Pheromone update and measure, trail evaporation	Immune operators cloning, hypermutation and selection based on elitism	Reproduction, replacement of bee, selection	Reproduction, chemotaxis, dispersion, elimination
Datasets for attack detection	*KDDCUP99, DARPA98*, *NSL-KDD*	*KDDCUP99, DARPA98*	*KDDCUP99, DARPA98*	*KDDCUP99, DARPA98, LLDOS*	*KDDCUP99, DARPA98*	*KDDCUP99, DARPA98*
Permittivity against DoS attack	High	Low	High	Low	Low	Low
Structure and dynamics	Discrete and network components and evolution or learning based	Discrete and network components and evolution based	Discrete components and evolution and learning based	Discrete and network components and evolution or learning based	Discrete and components and evolution based	Discrete and network components and evolution or learning based
Reliability	High	Low	High	High	Low	Low
Time-consuming	Time-consuming because of its complexity but the no. of repetitions is less	Low time-consuming because of its simplicity but the no. of repetitions is high	Time-consuming because of its complexity	Low time-consuming because of its simplicity	Low time-consuming because of its simplicity	Low time-consuming because of its simplicity
Robustness	High	Low as compared to PSO, but more than others	Lower than PSO and GA, more than others	Lower than ACO better	Lowest	Higher than ABS
Parameters	Number of particles, dimension of particles, range of particles, maximum number of iterations, inertia weight	Population size, max generation number, cross-over probability	Number of ants, iterations, pheromone evaporation rate, amount of reinforcement	Population size, no. of antibodies to be selected for hypermutation, number of antibodies to be replaced	No. of food sources which is equal to the no. of employed onlooker bees	The dimension of the search space, number of bacteria, number of steps chemotactic, no. of elimination and dispersal events, no. of reproduction steps, probability
Applications	Power system optimization problems, multimodel problems, multiobjective, dynamic, constrained, and combinatorial optimization problems, anomaly detection, sequential ordering problem, etc.	Pattern recognition, reactive power dispatch, sensor-based robot path planning, multiobjective vehicle routing problem, molecular modeling, web service selection, etc.	Continuous optimization and parallel processing implementations. Vehicle routing problem, graph coloring and set covering, agent-based dynamic scheduling, etc.	Computer security, anomaly detection, clustering/ classification, numeric function optimization, virus detection, pattern recognition, etc.	Solving reliability redundancy allocation problem, training neural networks, XOR, decoder-encoder, and 3-bit parity benchmark problems, pattern classification, etc.	Application for harmonic estimation problem in power systems, the parameters of membership functions, and the weights of rules of a fuzzy rule set are estimated, etc.

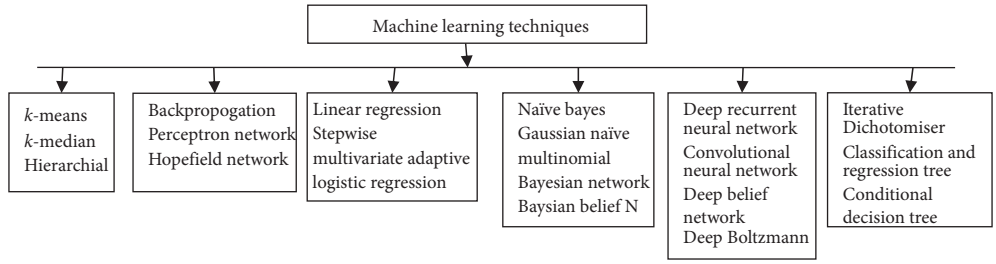

Figure 3: Categories of machine learning techniques.

comprises billions of neurons which are interlinked by different synapses, and its functionality is separated into three major layers which are input, output, and hidden layers in which each connection is associated with some weight. The entire networks are trained and learn from its learning phase and training phase through the weight adjustment, so it enables us to calculate the accurate class to the set of inputs. ANN, as shown in Figure 4, is also defined as a network of numerous computing elements or units that are closely interconnected with each other and also transform a set of inputs to the required outputs. The outcomes are evaluated using the unique weights and elements that are related to each other by interconnectivity between them. The network can generate the desired output by modifying links connecting nodes [60]. Activation function is applied to the set of input nodes, then passed through hidden layer nodes, and finally reaches the output nodes.. ANN works as a well-designed transformation of a set of input to output values. An artificial neural network can work for both the methods of the anomaly- and signature-based attack detection [61].

4.1.1. Anomaly-Based Detection Using Artificial Neural Network. Jadidi et al. [28] proposed a model for attack detection using the hybrid form of ANN for detecting attacks by using a flow-based dataset and also applied metaheuristic optimization algorithms for achieving an optimal solution. In this proposed work, there were two hybrid heuristic algorithms such as PSOGSA and cuckoo, which were used to efficiently use the interconnected weights of an MLP network. The resultant network analyzed by flow-based datasets compared its performance with the previously used techniques and found that the proposed hybrid technique enables us to detect attacks with better accuracy.

Jiang et al. [62] proposed a model designed by using hierarchical neural networks for an attack detection system that worked on RBF. The proposed method used the combination of the anomaly- and signature-based detection methods, also having the benefit of the RBF for low training with better accuracy. The RBF anomaly classifier is used for the identification of normal or attack data. Hence, the proposed method enabled us to analyze real-time network traffic.

In Jadidi et al.'s study [63], the proposed model was built on an anomaly-based detection approach which is a very well-known technique and efficiently used for detecting unknown attacks. This work is based on the anomaly-based attack detection method and MLP neural network with a single hidden layer was used. In this attack detection system,

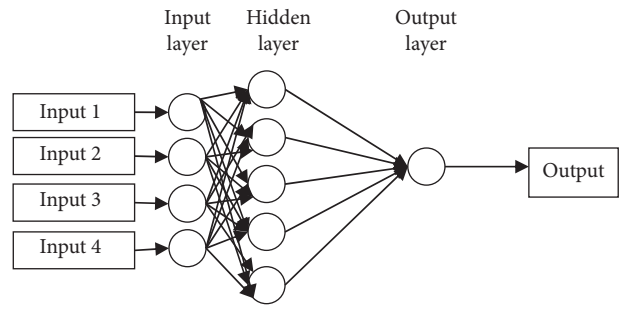

Figure 4: Neural network architecture.

GSA was used for the optimization of interconnected weights of a multilayer perceptron network. Hence, the proposed GSA-based detection system successfully achieved 99.43% accuracy.

Ryan et al. [64] proposed a model of intrusion detection designed by using ANN in which the BPNN algorithm is used to model attack detection systems in which the system of some users was used. The dataset used for training and testing was taken from the logs of the UNIX environment. The evaluation of the result found 96% accuracy and a 7% false alarm rate.

4.1.2. The Signature-Based Detection Approach. Cannady [65] proposed a model for intrusion detection that was built by using artificial neural network designed by a multistage classifier approach to detect signature-based (misuse based) detection. The data created by a real-time secure network consist of attack signatures and analyzed approximately thousands of events in which 3000 were simulated attacks.

Nine different features are selected after the data preprocessing step.. Normal or abnormal traffic is recognized by training the system using an artificial neural network, which enables us to learn the collective signatures. The proposed model resulted in 93% accuracy and found efficient after compared with other algorithms.

4.2. Bayesian Network. A Bayesian network is one of the ML techniques that work on the concept of probabilistic graphical model that is represented by some particular variables and the associations between them [66]. The Bayesian network can easily handle incomplete datasets [33]. The network is generally created in the form of a graph where nodes or vertices (V) are used as the random variables and edges (E) as a connecting association between them, and

a directed acyclic graph (DAG) is set up. The lower-level nodes are called child nodes that depend on parent nodes or upper-level nodes. Every node or vertices are assigned a random variable and conditional probability [16].

Bayesian classifiers based on Bayes' theorem are used for the classification of the new instances of a data sample named Y. Each instance is a set of attribute values that are denoted as $Y = (y_1, y_2, \ldots, y_n)$. Considering n number of classes, the sample Y is assigned to the class C_i if a given condition is satisfied:

$P(Y \mid C_i)\, P(C_i) > P(Y \mid C_j)\, P(C_j)$ for all i and j in $(1, m)$.

The sample is considered to be the class that has a max probability. In the Bayesian network, the attributes are implicitly conditional independent. Instead of that, naive Bayesian classification provides acceptable outcomes as it focused on the identification of the classes for the instances instead of probabilities. Hence, it can be used in various applications such as text data classification and attack data classification [16].

4.2.1. Anomaly-Based Detection. Panda and Patra [61] had proposed a structure for an attack detection system that used the naïve Bayes algorithm, one of the techniques of machine learning. The experiments applied on a 10% KDDCUP99 dataset, and the system is evaluated by tenfold cross-validations. The experimental results show the proposed approach achieved a higher detection rate than other approaches, the detection rate was noted as 95%, the error rate was 5%, and it was fast and cost-effective.

Farid et al. [67] had proposed representation for intrusion detection where data classification can be done by using one of the popular learning algorithms, naive Bayesian technique. The overall working of the proposed algorithm for intrusion detection had been evaluated on 10% of KDDCUP99. The experimental results founded high accuracy with minimum false positives.

Muda et al. [66] proposed a hybrid method in which the hybridization of two machine learning approaches which are naïve Bayes and k-means clustering technique was used for solving a classification problem. The computational evaluation can be performed on the benchmark KDDCUP99. The proposed model worked with two different phases in the first phase, and the grouping of similar data instances was done according to their behaviors by using the k-means clustering technique. In the second phase, the naïve Bayes classifier was used for classification task and the results are achieved by this approach: the accuracy was noted as 99% and false alarm was less than 0.5%.

Ben Amor et al. [33] proposed a model built on the naïve Bayes classifier and built a normal Bayesian network, and the evaluation is applied on the KDD 1999 dataset and collecting the classes of attacks in the following three major stages for performance measurements. In the first stage, calculate single attack in normal data, in the second stage, contain all four attack types of the KDD 1999 dataset, the problems were resolved by using multiclass classification based on the misuse detection technique, and the third stage consists of normal data and all four attack types using anomaly-based attack detection technique. The experimental results found with better accuracy.

4.2.2. The Signature-Based Detection Approach. Panda and Patra [61] proposed the attack detection model designed by naïve Bayes technique using Weka tool [23], and the experiments were applied on the KDD 1999 dataset that is grouped into different attacks of KDD datasets; finally, the results are compared with the neural network classifier and reported as the naïve Bayes classifier has a high accuracy and false alarm rate than NN.

4.3. Support Vector Machine. SVM is capable of resolving various pattern recognition problems, proposed by Vapnik [25]. SVM uses the concept of supervised learning with related learning algorithms which were mostly applied to signature-based detection in the last few years. It transforms the set of inputs into a high-dimensional space and can be creating an optimal divided hyperplane into the high-dimensional feature space [60]. The SVM classifier is applied to provide improved output for binary classification as compared with further classifiers. SVM promises good performance, and hence, it is used in various fields such as pattern recognition, bioinformatics, text categorization, speaker verification, character recognition, engineering and science, and financial market evaluation [32]. SVM is popular for solving various classification problems because its robustness and efficiently dealing with high-dimensional data also remove the nuisance of the dimensionality problem. SVM was initially designed for binary classification for constructing an optimal hyperplane to maximize the division line among the negative and positive datasets [32].

4.3.1. Anomaly-Based Detection Using Artificial Neural Network. Mukkamala et al. [68] proposed a hybrid model that is the collection of techniques ANN and SVM for the attack detection system. The purpose of using SVM is to achieve better speed and scalability in attack detection system. The experimental results were carried on the DARPA 1998 dataset. The result of the proposed method shown as training time for SVMs is significantly minimum, and it is reported as 17.77 sec shorter than neural networks. The performance of SVM showed that the attack detection system had a higher rate of detection than neural networks.

Chen et al. [69] proposed a model used the combination of set theory and SVM for the attack detection system. The experiments are applied on the KDDCUP99 dataset, and the rough set theory concept is applied at the preprocessing phase and to optimized features. The selection of best features was selected and applied to train the SVM model and accordingly tested. The experimental result has shown that the accuracy was noted as 86.79% and FPR was 29.97%. The accuracy was found better with a reduced false positive rate.

Wang et al. [70] proposed a model for attack detection designed by using a two-hybrid combination of algorithms that worked on improved SVM by using the collective form of PSO and PCA. The experiments were performed on the KDDCUP99 dataset, and principle component analysis (PCA) was used as an effective technique for decreasing dimensions of the dataset. The particle swarm intelligence technique was applied with different parameters in SVM.

The results have shown that the attack detection rate was found to be improved by PCA and PSO combination as compared with PSO-SVM.

Srinoy et al. [21] proposed an attack detection model that is built on a hybrid combination of algorithms which are PSO for SVMs and optimal feature selection, which act as a fitness function of PSO. The evaluation of the result showed that the proposed method successfully recognized both the unknown and known attacks. This proposed technique achieved better classification accuracy by comparing it with different traditional methods.

Shinde and Parvat [26] proposed a model built by the use of PSO and SVM for choosing optimal parameters for gaining a high detection rate and low false alarm rate. Support vector machine (SVM) has the potential for achieving better classification for the attack detection system. The working of SVM is depending on choosing the better parameters. This proposed work used the SVM classifier with the knowledge gain for feature selection. The classified parameters of SVM will be optimally chosen by a PSO algorithm. The experiments were applied on the NSL-KDD dataset, and results show that the proposed method can attain a low false alarm rate and higher detection rate.

Saxena and Richaariya [46] proposed a model which was a hybrid form of the PSO method and SVM. The experiments are applied to the KDDCUP99 dataset. The selection of optimized parameters by binary PSO and classification problem was solved by the support vector machine. The binary PSO provides the finest promising feature subset for creating a better intrusion detection system. The proposed method had to complete that task by following certain major steps which are preprocessing, feature reduction using information gain, and training using hybrid SVM-PSO. Finally, the evaluation has shown that the hybrid combination of PSO and SVM achieved a high detection rate than the simple SVM method.

Wang et al. [71] proposed a model designed by using the SVM-based feature selection algorithm for reducing the dimension of sample data. The anomaly-based intrusion detection technique is used, and MSVM is used with highly optimized parameters by the use of particle swarm optimization (PSO) in the collective form detecting anomalous connections. The experiments were performed on the KDDCUP dataset to measure the efficiency of proposed algorithms (FS-SVM and MSVM-PSO) and the detection precision of MSVM-PSO, also comparing the MSVM-PSO with three different algorithms: these were Bayesian algorithm, k-means, and multiclass support vector machine with optimized parameters of the method (MSVM-grid). The experimental results have shown that the hybrid form of MSVM-PSO outperforms than three algorithms in terms of different parameters such as detection rate and accuracy.

4.4. Decision Trees. DT is one of the prominent methods of ML that represented in form of a tree structure where each lower-level node is used to represent a decision or test on the data item which is taken into consideration [1]. The outcome of the test decides the selection of any branch. Classification of any data item is based on a process in which the decision tree algorithms start with its root node and follow the assertions, and the process is carried out until reaching a terminal leaf or node. After reaching the terminal node, a decision is being made. A decision tree is also represented as a unique form of a rule set, categorized by the hierarchical association of rules. A decision tree comprises of following essentials [37]. A decision node which represents a condition or a test on a data item, one of the possible attribute values, or test attribute outcomes is given by branch or edge, and the object belongs to which class is given by the leaf. It starts from the root node of the decision tree and follows the branch indicated by the outcome of each test until a leaf node is reached which is the procedure to classify an object. The leaf node level class is called as unknown object, and the information gaining of the attributes provides the best attribute on the division of subsets.

Ben Amor et al.[28] proposed an attack detection model that was built on the combination of two ML techniques which were DT and Bayesian network. The experiments were applied to the KDDCUP99 dataset and also compared the performance of both the techniques. After the evaluation of the proposed technique, that DT-based method provides much better results than naïve Bayes. If the comparison is based on the computation, then the building of DT is much slower than the Bayesian network. The decision tree selects the optimal features for each node during the creation of the tree based on some defined criteria. The advantage of DT is that it has a good speed of operation and high attack detection accuracy.

Stein et al. [72] proposed the attack detection model which was based on the GA approach for the optimal selection of features used with DT, to achieve maximum detection rate and minimum false alarm rate. The experiments were performed on different attack datasets with different attacks separately. The GA improved some of the categories such as in performance gaining on probe. Hence, it was found that the performance improvement on other attack types is much higher in the testing data.

Karthik et al. [73] proposed the model used three techniques which are used for hybridization, i.e., chi-square, information gain, and relief, and compared their performance using the decision tree classifier. Evaluation can be done by using the KDDCUP99 dataset, and the results have shown that the decision tree which classifies the performance is improved with high accuracy.

4.5. Random Forests. RF is one of the classification techniques that consist of a set of tree-based classifiers [20]. RF is a collection of classification and regression which is invincible in terms of accuracy among the DM techniques. The RF algorithm has numerous applications, used in prediction, probability estimation, and pattern analysis [74]. Random forest classifiers consist of a collection of a huge number of DTs. It is a collection of tree hierarchy in which each tree depends on the values of a random vector that is sampled individually and with the same distribution for all trees in

the forest. If new records were taken as input, then random forest makes trees for that records and keeps them in the forest [74].

Malik et al. [75] proposed a hybrid combination of the binary PSO and RF algorithm for the optimal feature selection and classification of attacks in a network. Particle swarm optimization is one of the popular algorithms of the swarm family that has the capability of robust global search and is used for optimal feature extraction, and random forest (RF) is a highly accurate classifier and used for classification. The experiments of the proposed technique are applied to the KDDCUP99 dataset. We also compared the evaluated results of the proposed system with other classifiers, and the final results show that performance achieved by the proposed classifier is much better than the traditional approaches.

Malik and Khan [74] proposed a model designed by using the BMOPSO approach to detect PROBE attacks in the network. The proposed technique focused on two basic parameters to be achieved as first attack DR and second FAR to follow the procedure of feature selection. The experiments were applied to the KDDCUP99 dataset. The proposed approach is used for optimal feature selection from a set of different features and RF techniques used for highly accurate and fast classification. The results have shown that the proposed method performed well for the classification.

4.6. Association Rule and Clustering. Hao et al. [36] proposed an attack detection model for the application level DDoS attack. The weblog record was used for user session extraction and recognizing patterns of the data that are normal or abnormal, and also evaluating the similarities between different sessions. The traditional k-means clustering algorithm fails into local optimality. The hybrid collective form of particle swarm optimization k-means clustering algorithm (PSO-KMC) was used for constructing an attack detection model. The proposed model enables to detect whether the undetermined sessions are part of DDoS attack or not. The experimental results have shown that the proposed technique can detect attacks effectively with high performance.

Srinoy and Kurutach [76] proposed a model designed by using a data mining-based hybrid approach for attack detection. The algorithm is hybridized by k-means and artificial ant clustering algorithm primarily used to generate raw clusters, and they were refined further by k-means particle swarm optimization (KPSO). The proposed model had been developed as the evolutionary-based clustering technique. We hybridized the k-means algorithm and PSO to find good partitions of the data. The proposed approach allows recognizing not only known attacks but also unknown attacks. After evaluation, the results have shown that the proposed hybrid algorithm performed well with high accuracy.

Ensafi et al. [77] proposed a model designed by using a hybrid combination of algorithms for the attack detection model which were fuzzy logic-based approach and swarm-based approach. The proposed technique is efficiently applied to solve local minima and complex classification problems. The proposed SFK-means approach used the benefits of k-means, fuzzy k-means, and swarm k-means, and all together successfully resolved most of the problems. The importance of the SFK-means algorithm was to overcome local convergence problems in fuzzy k-means and the sharp boundary problem in swarm k-means. The experiment is applied to the KDDCUP99 dataset and found that the proposed approach was effective in detecting various attacks.

4.7. Hidden Markov Models. Markov chains or hidden Markov models (HMMs) are members of the class of the Markov model. A Markov model consists of set of states that experience some transitions from one state to another by the transition probabilities which decide the logical structure of the model [31]. HMM is one of the machine learning models used in various applications such as speech, pattern, gesture recognition, bioinformatics, and language processing domain. It is also named as a statistical or sequence model in which the system modeling can be done by the Markov process for unknown parameters. The challenging task in the HMM model is to recognize different hidden parameters from the visible parameters [31]. The states of a hidden Markov model represent undetermined conditions that should be modeled and have dissimilar output probability distributions at each state.

4.7.1. Anomaly Detection and Hybrid Detection. Joshi and Phoha [78] proposed the hidden Markov model for intrusion detection. The HMM was used with the following five states and six symbols per state. The states were interconnected in a manner that anyone of state can reach other states. Baum-Welch et al. were applied to evaluate the hidden Markov parameters. The experiments were applied to the KDD 1999 dataset, also evaluates other parameters such as FP and FN rate. The results showed that the accuracy is significantly improved by using more than five features.

4.7.2. Misuse-Based Detection. Ariu et al. [31] proposed an attack detection model for the web applications and used hidden Markov models for the attack signature extraction. The proposed method competitively modeled the classifiers. The experiments were applied to the DARPA 1999 and HTTP dataset. The experiments were applied, and the detection ratewas evaluated higher than 0.8.

4.8. Deep Learning. Deep learning is another field of machine learning that deals with algorithms based on structure and function that resemble the working of the human brain which is called artificial neural networks. It belongs to both the categories of supervised and unsupervised learning and dealing with multilevel representation and features of hierarchical architecture in classification and pattern recognition. Deep learning is popularly used in different applications such as image processing and audio, text, and speech recognition. These techniques are targeted to learn the best feature representation from the bulk amount of unstructured data. Deep learning-based different methods are used to overcome different problems of modeling an efficient attack detection system. There are various deep learning methods which are

available such as recurrent neural network (RNN) [79], Boltzmann machine (BM), restricted Boltzmann machine (RBM), deep Boltzmann machine (DBM), deep neural network (DNN) [80], autoencoder, deep/stacked autoencoder, stacked denoising autoencoder, distributed representation, and convolution neural network (CNN). Self-taught learning (STL) [81] is also one of the DL approaches that consist of two stages for the classification. First, the best feature is selected from bulky data, called unsupervised feature learning (UFL). In the second stage, it learned representation of labeled data and used for the classification. DL architectures are similar to a neural network of multiple layers of architecture and various linear or nonlinear functions. The deep learning algorithm that works with high speed and fast learning capability and provides an efficient solution is said to be a successful method.

Niyaz et al. [82] proposed an attack detection model against DDoS attack, a DL-based approach of the multivector DDoS detection system in an SDN environment. SDN provides facility to program network devices for accomplished different tasks. The proposed system is designed as a network application over the software-defined network environment controller. The feature selection or classification is done by using deep learning. The result showed high accuracy with a low FP rate for attack detection in the proposed system.

Yin et al. [79] proposed a model designed by using deep learning technique, called the RNN-based intrusion detection system. The performance can be evaluated in the form of binary and multiclass classification, and the number of layers, neurons, and different learning rate was included. Hence, comparison was carried out with different techniques such as ANN, RF, SVM, and other machine learning methods that were previously used by the researchers. The experimental results have shown that the proposed intrusion detection model is suitable for modeling complex classification models that have high accuracy and performance is also superior to the traditional machine learning classification methods in both binary and multiclass classification. The proposed model ensures that improved accuracy of the intrusion detection also provides a better method for intrusion detection.

Javaid et al. [81] proposed a deep learning-based IDS model, using self-taught learning (STL), a DL technique was applied, and the experiment was evaluated on the NSL-KDD dataset for network intrusion. The performance of the proposed approach is better than other traditional approaches of previous work. Comparison can be done on certain parameters which are accuracy, precision, recall, and F-measure values.

Table 3 presents comprehensive study among the machine learning techniques. Table 4 and Table 5 present the performance comparison among machine learning techniques.

5. Datasets Used in Cyberattack Detection

Using machine learning and optimization algorithms for classification and feature selection problems, the dataset is considered to be a very important element. Since these techniques are working with the learning and testing phase in which they learn from the existing data, hence it is essential to have proper knowledge of the dataset that should be used to be aware of how the various authors and scientists may apply different machine learning and optimization algorithms. In this section, we describe different types of datasets used for attack detection in detail by applying machine learning and optimization algorithms. Here, we discuss three broad categories of the dataset which are public datasets, net flow datasets, and packet flow-based datasets.

5.1. Public Datasets. For the attack detection system, there are the following public datasets are discussed in detail. The different public datasets are given in the following sections.

5.1.1. Defense Advanced Research Projects Agency (DARPA 1998). DARPA 1998 is firstly created through the Cyber Systems and Technology Group of MIT Lincoln Laboratory, under Defense Advanced Research Projects Agency and Air Force Research, a laboratory for the assessment of network attack detection systems. DARPA 1998 and 1999 are widely used in various experiments and repeatedly cited in many publications. The DARPA 1998 dataset is formed by the MIT/LL. It creates an interest in the various researchers that may work on different issues of network, base station, and network attack detection system. The evaluation is designed to concentrate on core issues of technology and to motivate much participation to work on security and privacy concerns.

The DARPA 1999 dataset significantly has many attack types as compared to the DARPA 1998 dataset. There were two parts of intrusion detection evaluation of 1999 DARPA: firstly offline evaluation and secondly by a real-time evaluation. The attack detection systems are tested in the offline evaluation mode using network traffic and various audit logs together on network simulators. Some batch mode is applied for processing these data.

5.1.2. Knowledge Discovery Dataset (KDD). The most commonly used benchmark for attack detection is named as the KDD 1999 dataset [66] which was formed for the KDDCUP challenge in 1999. The KDD dataset consists of three major terms which are basic, content, and traffic features and creates 41 attributes (Table 6). The KDD 1999 dataset has some resemblance with the NetFlow dataset, but it is more complicated and has detailed features since the attacks were evaluated.

Another form of NSL-KDD dataset that has 42 attributes (Table 6) was used in this study. NSL-KDD is an enhanced form of the KDD99 dataset on which repeated instances were removed. The NSL-KDD dataset has many different versions in which only 20 percent of training data are used which are determined as KDD train 20% along with 25192 instances. The tested dataset is determined as KDD test that has 22544 instances. Table 6 describes the KDD dataset attributes with class labels. Hence, from these 42 attributes, the 41 can be classified into four different classes mentioned as follows: basic (B) features are of individual TCP connections, content (C) features are inside a connection recommended by domain knowledge, traffic (T) features are processed by two-second

Table 3: Comparative study among different machine learning techniques.

Technique	Principle	Parameters	Advantages	Limitations
k-means	Find out k points called centers that are evaluated as the sum of the distances of all points to their respective cluster centers	Cluster center location	High computation, produce closet clusters	Calculation of K is a very tough task for a fixed number of clusters. The dissimilarity, the initial and final cluster partition
K-nearest neighbor	The input consists of k-closest training of the feature space by using instance-based learning	Class of nearest neighbor	It is easy to implement, less complex	Difficult to deal with arbitrary attributes
Support vector machine	The mapping of input data to the high-dimension space and also dealing with linearly separated data for classification	Features of high dimension	Having high accuracy, flexible and robust in dealing with errors	Takes large time for training, complex to handle learned function (weights)
Hidden Markov model	It is a statistical or sequence-based model that consists set of states, transitions represent the set of possible positions	Pixels in a vision-based input	High-scalable model and easy to understand	Many assumptions about the data. A large number of parameters required to be set. Highly needed training data

Table 4: Performance comparison of different machine learning techniques.

Parameters	ANN	NB	SVM	KNN	DT	RF	DL
Accuracy	High	High	High	Low	Low	High	High
Training time	High training time due to complexity	Low training time due to simplicity	High training	High training time	High training time due to complex structuring	High training time	High training time complex structuring
Execution time	Average	High	High	Low	Low	Low	High
Large attributes	Dealing well with large attributes	Dealing well with large attributes	Dealing good with large attributes but speed will be very slow	Dealing well with large attributes	Average dealing with larger attributes	Dealing well with large attributes	Dealing good with large attributes but speed will be very slow
Lots of missing attributes	Contradictory	Good performed	Good performed	Low performed	Good performed	Good performed	Good performed
Lots of noisy data	Contradictory	Better dealing with noise	Better dealing with noise	Low capability dealing with noisy data	Average dealing with noise	Average dealing with noise	Better dealing with noise but the overall process is time-consuming
Large datasets	Cannot handle large dataset and speed of processing will be very slow	Better while handling large dataset	Average performed while handling large dataset but processing speed will be very slow	Better while handling large dataset	Average performed	Average performed	Better while handling large dataset but processing speed will be very slow due to complex structure
Detection rate	High	High	High	Low	High	Low	High
Datasets suitable	KDD99 and NSL-KDD	KDD99 and NSL-KDD	NSL-KDD, KDDCUP99, and DARPA	NSL-KDD, KDDCUP99, and DARPA	KDDCUP99	KDDCUP99	KDD99 and NSL-KDD

time window, and host (H) features are planned to assess attacks that last for more than two seconds.

5.2. Packet Flow-Based Dataset. Over the Internet, there are the following broadly used protocols, for example, IP, ICMP, IGMP, TCP, and UDP. Hence, due to the client programs, running these protocols can create huge traffic over the network. The overall incoming and outgoing packets use the physical interfaces such as Ethernet port for transmission and reception of the packets. Because at the network layer, none of the abovementioned protocols is directly transferred to its lower layer, hence these protocols are encapsulated inside the data field of the IP packet format, and then it can be transferable to its lower layer. In the data link layer or MAC

Table 5: Complexities of various machine learning techniques.

Algorithm	Complexity	Capability	References
ANN	O (emnk)	Low	e = no. of epoch, k = no. of neuron
Naive Bayesian	O (mn)	High	m = no. of training, n = no. of feature
SVM	R^3 & nS	High	R = no. of free support vector
Decision tree	O (mn^2)	Medium	M = no. of features
Clustering k-means	O (kmni)	High	I = no. of iteration, k = no. of cluster
Clustering hierarchical	O (n^3)	High	N = no. of data points
Association rules	O (n^2)	Low	N = no. of input transactions

Table 6: Detail of KDD dataset attributes with different classes.

S. no.	Dataset feature name	Classes
1	Duration	Basic
2	protocol_type	Basic
3	Service	Basic
4	src_bytes	Basic
5	dst_bytes	Basic
6	Flag	Basic
7	Land	Basic
8	wrong_fragme nt	Basic
9	Urgent	Basic
10	Hot	Content
11	num_failed_lo gins	Content
12	logged_in	Content
13	num_compro missed	Content
14	root_shell	Content
15	su_attempted	Content
16	num_root	Content
17	num_file_crea tions	Content
18	num_shells	Content
19	num_access_fi less	Content
20	num_outboun d_cmds	Content
21	is_hot_login	Content
22	is_guest_login	Content
23	Count	Traffic
24	serror_rate	Traffic
25	rerror_rate	Traffic
26	same_srv_ rate	Traffic
27	diff_srv_r ate	Traffic
28	srv_count	Traffic
29	srv_serror_rate	Traffic
30	srv_rerror_rate	Traffic
31	srv_diff_h ost_rate	Traffic
32	dst_host_count	Host
33	dst_host_srv_co unt	Host
34	dst_host_same_ srv_rate	Host
35	dst_host_diff_sr v_rate	Host
36	dst_host_same_src_port_rate	Host
37	dst_host_count	Host
38	dst_host_srv_co unt	Host
39	dst_host_same_ srv_rate	Host
40	dst_host_diff_sr v_rate	Host
41	dst_host_srv_re rror_rate	Host
42	Class	Host

layer, an Ethernet frame consists of about 1500 bytes of the payload, and this consists of encapsulated IP payload with IP header and IP packet contains itself its header and in its data section higher-level protocols such as HTTP, NFS, POP, telnet, and TFTP.

5.3. NetFlow Data. It contains the detail of the router and its features, and the routers have the capability to collect IP packet traffic as these packets are in and out from the router. The NetFlow generally Cisco's version 5 can introduce the network flowing as the sequential flow of packets in the same direction and defines seven types of attributes which are as follows: source and destination IP address, IP protocol, source and destination port, interface, and IP type of service.

6. Observations and Evaluations

The overall performance of any attack detection system is done by evaluating the performance of the various techniques that are applied to it and by the help of some parameters. In this literature survey, it is observed that the KDDCUP and DARPA benchmarks are very suitable and highly used for the evaluation of the performance of the system by applying different metaheuristic and machine learning techniques. In this study, different machine learning and metaheuristic techniques were not applied to build any IDS system but applied to certain cyber data for evaluating their performance. The major parameters that were calculated by applying different techniques on the cyber data are accuracy, detection rate, false alarms, detection rate, false positive, false negative, etc. These parameters show that the ability of a particular technique is suitable for attack detection or not. By the study of different research papers, a certain comparison is shown based on parameters and different techniques of metaheuristic and machine learning on the various cyber data. Figures 5–7 show a comparison among various ML and MH techniques applied on different datasets by different researchers and detection rate, accuracy, and FAR parameter evaluated.

In Figures 5–8, the comparison on the basis of different criteria among previous research studies that used ML and MH techniques in cyberattack detection is shown. It is found that ML and MH techniques improve performance in cyberattack detection models. In Figure 5, we have used previous papers and showed improved detection rate of the different models by ML and MH techniques. In Figure 6, we highlight another important parameter accuracy that is also seen to be improved by using ML and MH techniques. In Figure 7, we highlight the performance of algorithm on the FAR parameter that is also seen to be improved, and finally, in Figure 8, we have shown the usability of benchmarks that are popular in cyberattack detection. Hence, in this research, we present the successful usability of ML and MH

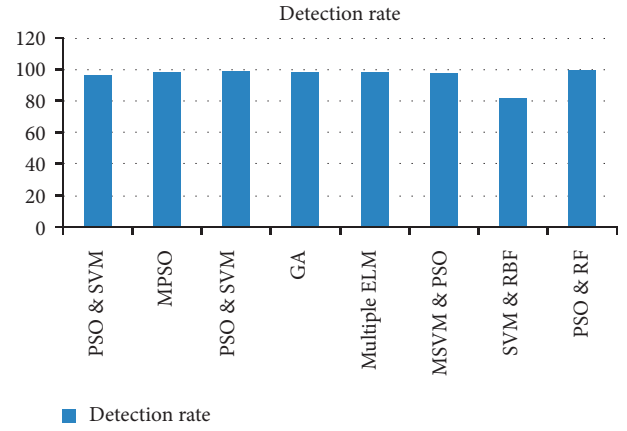

Figure 5: Comparison among different ML and MH techniques on the basis of detection rate from Table 7.

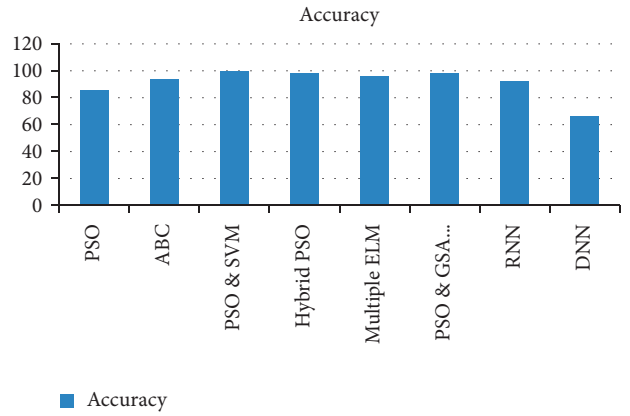

Figure 6: Comparison among different ML and MH techniques on the basis of accuracy from Table 7.

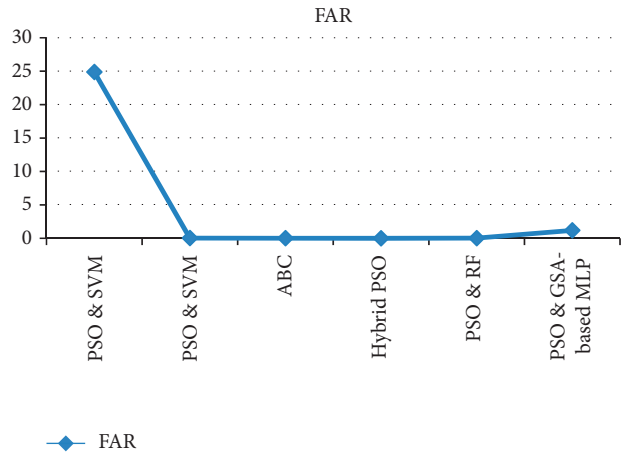

Figure 7: Comparison among different ML and MH techniques on the basis of FAR from Table 7.

techniques in cybersecurity domain, and furthermore, these techniques will be explored to perform well in this domain.

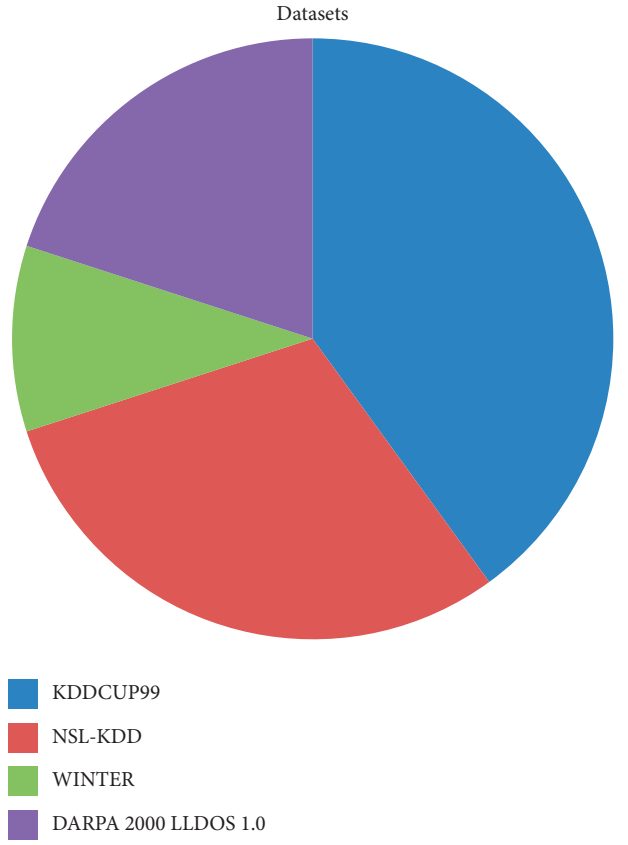

Figure 8: Comparison among different ML and MH techniques of previous research studies based on datasets from Table 7.

Table 7 presents the comparative study on the basis of performance among various literature articles and techniques.

7. Challenging Issues and Future Directions

Here, we discuss the different challenges of machine leaning as well as optimization algorithms as follows.

There are two types of problem are categorized while talking about machine learning: one is the regression and the other is classification. The basic difference between both the approaches is about its output value (either continuous or discrete). In cybersecurity, mostly problems are associated with classification which provides categorical output. In order to design the model that can classify unknown data, it is important to first train the network using representative examples. . This phase is usually called either as training. To achieve the same, robust technique of learning will be take into consideration. For examples, the commonly used techniques are SVM [20], decision tree [72, 73], naïve Bayes [20], etc. The input data are an important factor for the learning techniques because they need preprocessing to meet the specification of techniques [87]. The important solutions for betterment in leaning algorithms in performance prospective are as follows:

(i) Dataset that is used for training should be labeled in case of output-based learning

Table 7: Comparative study of the performance of different research citations.

References	Detection rate (%)	Accuracy (%)	Far	Technique	Datasets
[20]		98.70	95.60%	SVM, naïve Bayesian	MIT Lincoln Lab, IDS
[21]	96.11	—	24.88	PSO, SVM	KDDCUP99
[23]	—	—	—	AIS	nit DARPA LLDOS 1.0
[24]	98	—	—	MPSO	KDDCUP
[26]	98.53	—	0.0374	PSO, SVM	NSL-KDD
[27]	95	95	—	BARTD, cuckoo	KDDCUP
[28]	—	98.2/99.55	1.19/ 0.21	MLP, PSO, GSA, cuckoo	DARPA and WINTER
[72]	98.4/96	—	—	GA	DARPA and WINTER
[25]	96.7	—	—	PSO	KDDCUP
[69]	—	85.13	—	PSO	LLDOS
[83]	—	93.25	0.02	ABC	KDDCUP
[46]	—	99.4	—	PSO, SVM	KDDCUP
[32]	—	99.25	0.75	PSO, SVM	PMU2015
[55]	—	84.29	—	Multiple ELM	NSL-KDD
[71]	97.64	—	—	MSVM, PSO	KDDCUP
[21]	—	97.7	0.002	Hybrid PSO	KDDCUP
[77]	96.11	—	3.89	PSO-SVM	KDDCUP99
[74]	99.92	—	0.029	PSO RF	KDDCUP99
[84]	—	89.6	—	RNN	NSL-KDD
[84]		92		LSTM	NSL-KDD
[85]		93.20, 78.1, 66, 96.6		DNN	KDDCUP99, NSL-KDD, UNSW-NB15, WSN-DS
[86]	81.8	—	—	SVM RBF	KDDCUP

(ii) The sample instances while training must represent all classes of the model

(iii) Identify the algorithm with better learning function, and train the model and regulate the parameters using separate data

(iv) Evaluate the model on dissimilar data such as test data

Now, selection of features is a most significant step for achieving better input representation. There are numerous methods, such as filter based, correlation based, wrapping, and heuristic, which were used. Here, we have discussed about metaheuristic techniques that are popularly used in cybersecurity basically for attack datasets. Metaheuristic methods provide two types of solutions such as single-based or population-based solutions. The population-based solution is mostly used in cyberattack problems due to providing multiple solutions. The most commonly used algorithm are already discussed in this paper such as PSO [21, 26, 38, 88], GA [68], ACO [21, 45], ABC [46], etc.

Generally, in cybersecurity, the attack detection problem requires high-level solution methods (metaheuristic methods) that enable us to escape from local optima and execute a robust search of a solution space. However, these methods are unable to perform with large-size or multidimensional datasets and sometimes suffer with convergence problems. Cybersecurity deals with the high-dimensional data as attack datasets are too large to handle. Hence, the advanced technique of learning comes into picture to deal with high-dimensional data. The advanced learning techniques such as deep learning methods [79, 81, 82] and latest artificial intelligence techniques can provide better solutions for cyberattack detection problem by efficiently covering the classification as well as feature selection problems [89–92].

8. Conclusion and Discussion

Cyberattack is one of the challenging areas of research. This study provides imminent research studies in the field of cyberattack detection, a summary of the different techniques related and work done in recent years.

In this paper, more than eighty recent related and fine publications of different conferences and journals were used, which highlights the previous study of different metaheuristic algorithms (MHs) and machine learning (ML) techniques in the attack detection system. ML and MH performance presents in comparative study in Table 1 and Table 2. In Table 1, we found out that these techniques provide better outcomes; hence there is need of exploring application of such techniques in other fields of cybersecurity.. In Table 2, we discuss about the internal properties of algorithms for their better use in computation.

Initially, the most important step is finding example papers that provide a detailed explanation of different machine learning (ML) and metaheuristic (MH) methods in the cybersecurity environment, for both the signature and anomaly-based detection. So, the analysis of the detailed survey presented in the paper states the fact that the machine learning and optimization techniques are more preferred, but regrettably, techniques that provide maximum efficiency and performances had not been invented yet, it is very difficult to provide one recommendation for each technique, and depending on the type of attack, the system is made-up to detect. For evaluating the performance and effectiveness

of any techniques, there are several criteria which are available, and it cannot be decided by taking some of them into account. The parameters that were evaluated are listed as accuracy, detection rate, the time complexity for classifying an unknown instance with a trained model, and understanding of the final solution of each machine learning and metaheuristic technique. One more critical characteristic of machine learning and metaheuristic algorithms is a type of dataset for the training and testing of systems in the attack detection process that should be carefully selected and preceded. Hence, the potential use of ML and MH techniques for the computation of the attack detection system is inspiring the advances required to realize the reliable, efficient, accurate, and robust attack detection systems.

Abbreviations

ML: Machine learning
MH: Metaheuristic
DoS: Denial of service attack
DDoS: Distributed denial of service attack
GA: Genetic algorithm
ACO: Ant colony optimization
PSO: Particle swarm optimization
AIS: Artificial immune system
ABC: Artificial bee colony
BFA: Bacterial foraging algorithm
ANN: Artificial neural network
NB: Naïve Bayes
KNN: K-nearest neighbor algorithm
SVM: Support vector machine
DT: Decision tree
RF: Random forest
DL: Deep learning
HMM: Hidden Markov model
KDD: Knowledge discovery database.

Data Availability

No data were used to support the findings of this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. Ganapathy, K. kulothungan, S. Muthurajkumar, M. Vijayalaxami, P. Yogesh, and A. kannan, "Intelligent feature selection and classification techniques for intrusion detection in networks: a survey," *Journal of Wireless Networking and Communications*, vol. 271, 2013.

[2] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, 2016.

[3] O. Adil Mahdi, Y. R. Bahar Al-Mayouf, A. Basil Ghazi, A. W. Abdul Wahab, and M. Y. I. B. I. Idna Bin Idris, "An energy-aware and load-balancing routing scheme for wireless sensor networks," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 12, no. 3, pp. 1312–1319, 2018.

[4] S. Deol and L. Kaur, "Review on detection and prevention schemes for flooding attack in WSNS," *International Journal of Advance Research Ideas and Innovations in Technology*, vol. 3, 2017.

[5] C. Borrego, M. Amadeo, A. Molinaro, and R. H. Jhaveri, "Privacy-preserving forwarding using Homomorphic encryption for information-centric wireless Ad Hoc networks," *IEEE Communications Letters*, vol. 23, no. 10, pp. 1708–1711, 2019.

[6] S. Venkatraman and M. Alazab, "Use of data visualisation for zero-day Malware detection," *Security and Communication Networks*, vol. 2018, Article ID 1728303, 13 pages, 2018.

[7] R. U. Khan, X. Zhang, R. Kumar et al., "An adaptive multilayer botnet detection technique using machine learning classifiers," *Applied Sciences*, vol. 9, no. 11, p. 2375, 2019.

[8] A. Karim, S. Azam, B. Shanmugam, K. Kannoorpatti, and M. Alazab, "A comprehensive survey for intelligent spam email detection," *IEEE Access*, vol. 7, pp. 168–295, 2019.

[9] R. M., S. P. Maddikunta, P. K. R. Parimala et al., "An effective feature engineering for DNN using hybrid PCA-GWO for intrusion detection in IoMT architecture," *Computer Communications*, vol. 160, pp. 139–149, 2020.

[10] B. A. Khalaf, S. A. Mostafa, A. Mustapha, M. A. Mohammed, and W. M. Abduallah, "Comprehensive review of artificial intelligence and statistical approaches in distributed denial of service attack and defense methods," *IEEE Access*, vol. 7, pp. 51691–51713, 2019.

[11] P. Dixit and S. Silakari, "Deep learning algorithms for cybersecurity applications: a technological and status review," *Computer Science Review*, vol. 39, p. 100317, 2021.

[12] A. Basit, M. Zafar, X. Liu, A. R. Javed, Z. Jalil, and K. Kifayat, "A comprehensive survey of AI-enabled phishing attacks detection techniques," *Telecommunication Systems*, vol. 76, no. 1, pp. 139–154, 2020.

[13] M. A. Mohammed, S. S. Gunasekaran, S. A Mostafa, A. Mustafa, and M. K. Abd Ghani, "Implementing an agent-based multi-natural language anti-spam mode," in *Proceedings of the International Symposium on Agents, Multi-Agent Systems and Robotics (ISAMSR)*, pp. 1–5, Putrajaya, Malaysia, August 2016.

[14] C. J. Tu, L. Y. Chuang, J. Y. Chang, and C.-H. Yang, "Feature selection using PSO-SVM," *IAENG International Journal Of Computer Science (IJCS)*, vol. 33, pp. 138–143, 2006.

[15] M. Athari and K. Borna, "Using meta heuristic algorithms of genetic, particle swarm optimization and glowworm in the intrusion detection system," *International Journal of Computer Science and Network Security*, vol. 16, no. 10, 2016.

[16] R. D. Sagarin and T. Taylor, "Natural Security: how biological systems use the information to adapt in an unpredictable world," *Information security*, vol. 14, 2012.

[17] S. Jamali and G. Shaker, "PSO-SFDD: defense against SYN flooding DoS attacks by employing PSO algorithm," *Computers & Mathematics with Applications*, vol. 63, no. 1, pp. 214–221, 2012.

[18] M. Tarao and T. Okamoto, "Toward an artificial immune server against cyber attacks: enhancement of protection against DoS attacks," *Procedia Computer Science*, vol. 96, pp. 1137–1146, 2016.

[19] S. Bhattacharya, S. R. K. S, P. K. R. Maddikunta et al., "A Novel PCA-Firefly based XGBoost classification model for Intrusion Detection in Networks using GPU," *Electronics*, vol. 9, no. 2, p. 219, 2020.

[20] J. Visumathi and K. L. Shunmuganathan, "Improved detection of dos attacks using intelligent computation techniques," *SRIMCA*, vol. 3, no. 2, 2010.

[21] S. Srinoy, "Intrusion detection model based on particle swarm optimization and support vector machine," in *Proceedings of the 2007 IEEE Symposium on Computational Intelligence in Security and Defense Applications*, Honolulu, HI, USA, April 2007.

[22] S. Mourougan and M. Aramudhan, "Hybrid evolutionary algorithm based intrusion detection system for denial of service attacks," *Indian Journal of Science and Technology*, vol. 8, no. 35, 2016.

[23] U. Akyazı, A. Şima Uyar, *Detection of DDoS Attacks via an Artificial Immune System-Inspired Multiobjective Evolutionary Algorithm*, pp. 1–10, Springer, Berlin, Germany, 2010.

[24] B. Ben Sujitha and V. Kavitha, "Layered approach for intrusion detection using multiobjective particle swarm optimization," *International Journal of Applied Engineering Research*, vol. 10, no. 12, pp. 31999–32014, 2015.

[25] H. Zheng and M. Hou, ""Yu Wang," an efficient hybrid clustering-PSO algorithm for anomaly intrusion detection," *Journal of Software*, vol. 6, no. 12, 2011.

[26] P. Shinde and T. Parvat, "Analysis on intrusions detection based on support vector machine optimized with swarm intelligence," *International Journal of Computer Science and Mobile Computing IJCSMC*, vol. 3, pp. 559–566, 2015.

[27] K. Munivara Prasad, A. Rama Mohan Reddy, and K. Venugopal Rao, "BARTD: Bio-inspired anomaly-based real-time detection of under rated App-DDoS attack on web," *Journal of King Saud University Computer and Information Sciences*, vol. 32, 2017.

[28] Z. Jadidi, V. Muthukkumarasamy, and E. Sithirasenan, "Metaheuristic algorithms based flow AnomalyDetector," in *Proceedings of the APCC*, Bali, Indonesia, August 2013.

[29] K. P. Mohan Kumar and M. Aramuthan, "Hybrid Network Intrusion Detection for DoS Attacks," *IJCTA*, vol. 9, pp. 15–22, 2016.

[30] Y. Wang and C. Wang, "Based on the ant colony algorithm is a distributed intrusion detection method," *International Journal of Security and Its Applications*, vol. 9, no. 4, pp. 141–152, 2015.

[31] D. Ariu, R. Tronci, and G. Giacinto, "HMMPayl: an intrusion detection system based on Hidden Markov Models," *Computers & Security*, vol. 30, no. 4, pp. 221–241, 2011.

[32] C. Kolias, G. Kambourakis, and M. Maragoudakis, "Swarm intelligence in intrusion detection: a survey," *Computers & Security*, vol. 30, 2011.

[33] N. Ben Amor, S. Benferhat, and Z. Elouedi, "Naive bayesian networks in intrusion detection systems," in *Proceedings of the ACM Symposium on Applied Computing (SAC)*, Nicosia, Cyprus, March 2004.

[34] S. M. H. Bamakan, B. Amiri, and M. M. Yong Shi, "A new intrusion detection approach using PSO based multiple criteria linear programming," *Information Technology and Quantitative Management*, vol. 55, pp. 231–237, 2015.

[35] S. Khajouei Nejad, S. Jabbehdari, and M. H. Moattar, "A hybrid intrusion detection system using particle swarm optimization for feature selection," *International Journal of Soft Computing and Artificial Intelligence*, vol. 3, no. 2, 2015.

[36] X. Hao, B. Meng, and K. Gu, "Detecting DDoS attack based on PSO Clustering algorithm," in *Proceedings of the 3rd International Conference on Materials Engineering, Manufacturing Technology and Control (ICMEMTC 2016)*, Taiyuan, China, February 2016.

[37] S. Momanyi Nyabuga, W. Cheruiyot, and M. Kimwele, "Using particle swarm optimization (PSO) algorithm to protect vehicular Ad Hoc networks (VANETS) from denial of service (DOS) attack," *IJARCET*, vol. 5, 2016.

[38] W. Guoli, "Traffic prediction and approach based on PSO optimized Elman neural network," in *Proceedings of the 2019 11th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, Qiqihar, China, April 2019.

[39] L. K. Siva Sankari, D. C. Joy Winnie Wise, and B. Priya, "An efficient method for denial of service attack detection using genetic algorithm," *IJARSE*, vol. 4, 2015.

[40] M. Mizukoshi and M. Munetomo, "Distributed denial of services attack protection system with genetic algorithms on hadoop cluster computing framework," in *Proceedings of the 2015 IEEE Congress on Evolutionary Computation (CEC) IEEE*, Sendai, Japan, May 2015.

[41] J. H. Lee, D. S. Kim, S. M. Lee, and J. S. Park, "Ddos attacks detection using GA based optimized traffic matrix," in *Proceedings of the Fifth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing*, Washington, DC, USA, June 2015.

[42] M. H. Bhuyan, H. J. Kashyap, D. K. Bhattacharyya, and J. K. Kalita, "Detecting distributed denial of service attacks: methods, tools and future directions," *The Computer Journal*, vol. 57, no. 4, pp. 537–556, 2014.

[43] G. Dimitris, T. Ioannis, and D. Evangelos, "Feature selection for robust detection of distributed denial-of-Service attacks using genetic algorithms," in *SETN, Lecture Notes in Computer Science*, vol. 3025, Springer, Berlin, Germany, 2004.

[44] S. M. Lee, D. S. Kim, J. H. Lee, and J. S. Park, "Detection of DDoS attacks using optimized traffic matrix," *Computers & Mathematics with Applications*, vol. 63, no. 2, pp. 501–510, 2012.

[45] M. Aldwairi, Y. Khamayseh, and Mh. A. Masri, "Application of artificial bee colony for intrusion detection systems," *Security and Communication Networks*, vol. 8, 2012.

[46] H. Saxena and V. Richaariya, "Intrusion detection in KDD99 dataset using SVM-PSO and feature reduction with information gain," *International Journal of Computer Applications*, vol. 98, no. 6, 2014.

[47] V. V. Mahale and D. B. Gothawal, "Cuckoo filter & Remote firewall: a mechanism for mitigation of distributed denial of service attacks," *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 6, 2016.

[48] V. Priyadharshini and K. Kuppusamy, "Prevention of DDOS attacks using new cracking algorithm," *Engineering Research and Applications (IJERA)*, vol. 2, no. 3, pp. 2263–2267, 2012.

[49] X. Hao, B. Meng, and K. Gu, "Detecting DDoS attack based on PSO Clustering algorithm," in *Proceedings of the ICMEMTC*, Taiyuan, China, February 2016.

[50] R. Damodaram and M. L. Valarmathi, "Bacterial foraging optimization for fake website detection," *International Journal of Computer Science & Applications (TIJCSA)*, vol. 1, no. 11, 2013.

[51] H. H. Chen and S. K. Huang, "LDDoS attack detection by using ant colony optimization algorithms," *Journal of Information Science & Engineering*, vol. 32, no. 4, 2016.

[52] N. Kumar and L. Walia, "Ant colony optimization based approach for the detection of black Hole attack in WANET," *IJCSMC*, vol. 4, no. 6, pp. 469–480, 2015.

[53] H. M. Rais and T. Mehmood, "Dynamic ant colony system with three-level update feature selection for intrusion

detection," *International Journal of Network Security*, vol. 20, no. 1, pp. 184–192, 2018.

[54] K. H. Abdulkareem, M. A. Mohammed, S. S. Gunasekaran et al., "A review of Fog computing and machine learning: concepts, applications, challenges, and open issues," *IEEE Access*, vol. 7, pp. 153123–153140, 2019.

[55] M. Latah and L. Toker, *An Efficient Flow-Based Multi-Level Hybrid Intrusion Detection System for Software-Defined Networks*, Springer, Berlin, Germany, 2018.

[56] K. Chandra, G. Kapoor, R. Kohli, and A. Gupta, "Improving software quality using machine learning," in *Proceedings of the 2016 International Conference on Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*, pp. 115–118, Greater Noida, India, February 2016.

[57] A. R Javed, M. Usman, S. U. Rehman, M. U. Khan, and M. U. haghighi, "Anomaly detection in automated vehicles using multistage attention-based convolutional neural network," in *Proceedings of the IEEE Transactions on Intelligent Transportation Systems (Early Access)*, IEEE, New York , NY, USA, December 2020.

[58] M. Mittal, C. Iwendi, S. Khan, and A. R Javed, "Analysis of security and energy efficiency for shortest route discovery in low-energy adaptive clustering hierarchy protocol using Levenberg-Marquardt neural network and gated recurrent unit for intrusion detection system," *Transactions on Emerging Telecommunications Technologies*, 2020.

[59] H. Guang-Bin, D. H. Wang, and Y. Lan, "Extreme learning machines: a survey," *International Journal of Machine Learning and Cybernetics*, vol. 2, no. 2, pp. 107–122, 2011.

[60] D. Chavan, C. Francis, E. M. Thomas, and P. Moraye, "Comparative study of preventive algorithms of Ddos attack," *International Journal of Scientific & Engineering Research*, vol. 7, no. 2, 2016.

[61] M. Panda and M. R. Patra, "network intrusion detection using naïve bayes," *International Journal of Computer Science and Network Security*, vol. 7, no. 12, 2007.

[62] J. Jiang, C. Zhang, and M. Kame, "RBF-based real-time hierarchical intrusion detection systems," *International Joint Conference on Neural Networks*, vol. 2, pp. 1512–1516, 2003.

[63] Z. Jadidi, V. Muthukkumarasamy, and M. Sheikhan, "Flow-based anomaly detection using neural network optimized with GSA algorithm," in *Proceedings of the International Conference on Distributed Computing Systems Workshops*, Mesa, Arizona, April 2013.

[64] J. Ryan, M.-J. Lin, and R. Miikkulainen, "Intrusion detection with neural networks," *Advances in Neural Information Processing Systems*, pp. 943–949, 1998.

[65] J. Cannady, "Artificial neural networks for misuse detection," in *Proceedings of the National Information Systems Security Conference (NISSC'98)*, pp. 443–456, Arlington: Virginia Press, October 1998.

[66] Z. Muda, W. Yassin, M. N. Sulaiman, and N. I. Udzir, "K-means clustering and naive bayes classification for intrusion detection," *Journal of Information Technology and Applications*, vol. 4, no. 1, pp. no. 13–25, 2014.

[67] D. M. Farid, M. Zahidur Rahman, and C. Mofizur Rahman, "Adaptive intrusion detection based on boosting and naïve bysian classifier," *International Journal of Computer Applications*, vol. 24, no. 3, 2011.

[68] S. Mukkamala, G. Janoski, and A. Sung, "Intrusion detection: support vector machines and neural networks," *IJCNN*, vol. 2, 2002.

[69] R.-C. Chen, K.-F. Cheng, Y.-H. Chen, and C.-F. Hsieh, "Using Rough set and support vector machine for network intrusion detection system," in *Proceedings of the Asian Conference on Intelligent Information and Database Systems*, Quang binh, Vietnam, April 2009.

[70] H. wang, G. zhang, E. Mingjie, and N. Sun, "A novel Intrusion detection method based on improved SVM by Combaining PCA and PSO," *Wuhan University Journal of Natural Science*, vol. 16, 2011.

[71] G. Wang, S. Y. Chen, and J. Liu, "Anomaly-based intrusion detection using multiclass-SVM with parameters optimized by PSO," *International Journal of Security and Its Applications*, vol. 9, no. 6, pp. 227–242, 2015.

[72] G. stein, B. chen, and A. S. Wu, ""Kien A hua" Decision tree classifier for network intrusion detection with GA based feature selection," in *Proceedings of the 43rd Annual Southeast Regional Conference*, Melbourne, FL, USA, March 2005.

[73] R. Karthik, S. Veni, and B. L. Shivakumar, "Network intrusion detection using feature selection and decision tree classifier," in *Proceedings of the TENCON*, Osaka, Japan, November 2008.

[74] A. J. Malik and F. A. Khan, "A hybrid technique using multi-objective particle swarm optimization and random forests for PROBE attacks detection in a network," in *Proceedings of the International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 2473–2478, IEEE, Bari, Italy, October 2013.

[75] A. J. Malik, W. Shahzad, and F. A. Khan, "Binary PSO and Random Forests Algorithm for PROBE Attacks Detection in a Network," in *Proceedings of the 2011 IEEE Congress of Evolutionary Computation (CEC)*, New Orleans, LA, USA, June 2011.

[76] S. Srinoy and W. Kurutach, "Combination artificial ant clustering and K-PSO clustering approach to network security model," in *Proceedings of the International Conference on Hybrid Information Technology (ICHIT'06)*, Cheju Island, Korea, November 2006.

[77] R. Ensafi, S. Dehghanzadeh, and M. R. Akbarzadeh, "Optimizing Fuzzy K-means for network anomaly detection using PSO," in *2008 IEEE/ACS International Conference on Computer Systems and Applications*, Doha, Qatar, April 2008.

[78] S. S. Joshi and V. V. Phoha, "Investigating hidden Markov models capabilities in anomaly detection," in *Proceedings of the 43rd Annual Association For Computing Machinery Southeast Conference, ACMSE*, New York, NY, USA, June 2005.

[79] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," *IEEE Access*, vol. 5, 2017.

[80] A. Rehman, S. U. Rehman, M. Khan, M. Alazab, and T. Reddy G, "CANintelliIDS: Detecting In-vehicle intrusion attacks on a controller area network using CNN and attention-based GRU," in *Proceedings of the IEEE Transactions on Intelligent Transportation Systems (Early Access)*, IEEE, New York , NY, USA, December 2021.

[81] A. Javaid, Q. Niyaz, W. Sun, and M. Alam, "A deep learning Approach for network intrusion detection system," in *Proceedings of the 9th EAI International Conference on Bio-Inspired Information and Communications Technologies*, New York, NY, USA, December 2016.

[82] Q. Niyaz, W. Sun, and Y Ahmad, "A deep learning based DDoS detection system in Software Defined Networking (SDN)," *EAI Endorsed Transactions on Security and Safety*, vol. 4, 2017.

[83] Q. Qian, J. Cai, and R. Zhang, "Intrusion detection based on neural networks and artificial bee colony algorithm," in *Proceedings of the 2014 IEEE/ACIS 13th International*

Conference on Computer and Information Science (ICIS), IEEE, Taiyuan, China, June 2014.

[84] L. Thi-Thu-Huong, Y. Kim, and H. Kim, "Network intrusion detection based on novel feature selection model and various recurrent neural networks," *Applied Science*, vol. 9, p. 1392, 2020.

[85] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," *IEEE Access*, vol. 7, 2019.

[86] U. Ravale, N. Marathe, and P. Padiya, "Feature selection based hybrid anomaly intrusion detection system using K means and RBF Kernel function," *Procedia Computer Science*, vol. 45, 2015.

[87] A. Q. Lima and B. Keegan, "Challenges of using machine learning algorithms for cybersecurity: a study of threat classification models applied to social media communication data, Cyber Influence and Cognitive Threats," *Cyber influence and Cognitive Threats*, vol. 2020, pp. 33–52, 2019.

[88] B. Tan, Y. Tan, and Y. Li, "Research on intrusion detection system based on improved PSO-SVM algorithm," *Chemical Engineering Transactions*, vol. 51, 2016.

[89] Y. F. Hernandez-Julio, I. Merino-Fuentes, R. R. Gonzalez-Diaz, A. Guerrero-Avendano, L. V. O. Toledo, and W. N. Bernal, "Fuzzy knowledge discovery and decision-making through clustering and Dynamic tables: application in Colombian business Finance," in *Proceedings of the 2020 15th Iberian Conference on Information Systems and Technologies (CISTI)*, Seville, Spain, June 2020.

[90] J. M. Saiz-Álvarez, A. Vega-Muñoz, Á. Acevedo-Duque, and D. Castillo, "B corps: a socioeconomic approach for the COVID-19 post-crisis," *Frontiers in Psychology*, vol. 11, no. 1867, 2020.

[91] R. H. Jhaveri, A. Desai, A. Patel, and Y. Zhong, "A sequence number prediction based bait detection scheme to mitigate sequence number attacks in MANETs," *Security and Communication Networks*, vol. 2018, Article ID 3210207, 13 pages, 2018.

[92] S. Ramani, R. Jhaveri, and C. Borrego, "Applications in security and evasions in machine learning: A survey," *Electronics*, vol. 9, no. 1, p. 97, 2020.