International Journal of Computational Intelligence Systems
Vol. **14(1)**, 2021, *pp.* 560–571
DOI: https://doi.org/10.2991/ijcis.d.210105.001; ISSN: 1875-6891; eISSN: 1875-6883
https://www.atlantis-press.com/journals/ijcis/

Research Article

Intrusion Detection Systems, Issues, Challenges, and Needs

Mohammad Aljanabi[1, 2,*], Mohd Arfian Ismail[2], Ahmed Hussein Ali[1]

[1]*College of Education, Aliraqia University, Baghdad, Iraq*
[2]*Faculty of Computing, University Malaysia Pahang, Gambang, Malaysia*

ARTICLE INFO

Article History
Received 17 Jun 2020
Accepted 11 Sep 2020

Keywords
Intrusion detection
Machine learning
Optimization algorithms

ABSTRACT

Intrusion detection systems (IDSs) are one of the promising tools for protecting data and networks; many classification algorithms, such as neural network (NN), Naive Bayes (NB), decision tree (DT), and support vector machine (SVM) have been used for IDS in the last decades. However, these classifiers is not working well if they applied alone without any other algorithms that can tune the parameters of these classifiers or choose the best sub set features of the problem. Such parameters are C in SVM and gamma which effect the performance of SVM if not tuned well. Optimization algorithms such as genetic algorithm (GA), particle swarm optimization (PSO) algorithm , ant colony algorithm, and many other algorithms are used along with classifiers to improve the work of these classifiers in detecting intrusion and to increase the performance of these classifiers. However, these algorithms suffer from many lacks especially when apply to detect new type of attacks, and need for new algorithms such as JAYA algorithm, teaching learning-based optimization algorithm (TLBO) algorithm is arise. In this paper, we review the classifiers and optimization algorithms used in IDS, state their strength and weaknesses, and provide the researchers with alternative algorithms that could be use in the field of IDS in future works.

1. INTRODUCTION

The use of computer systems and the Internet in recent times has led to serious security, privacy, and confidentiality issues due to the processes involved in the electronic transformation of data. Much effort has been channeled to the improvement of the security and privacy of computer systems, however, these problems still exist in computer systems. In fact, there is no system in the world that is completely secure. Additionally, there are different types of network attacks [1]; these attacks evolve when there is a new signature with an abnormal behavior in the database of signatures. With the emergence of several types of attacks, several tools are being developed and used in several forms of network attacks. One of these tools is the intrusion detection systems (IDSs). This tool allows the monitoring of a range of network systems, cloud computing system, as well as an information system. The IDS can monitor and detect attacks which aim to compromise the security features (confidentiality, availability, and integrity) of a system. This paper review the existing work, methods, and techniques in IDS, Section 2 give an overview about IDS, then followed by brief description about the main types of IDS and the techniques used in detection explained in 2.1 and 2.2, Section 3 state the existing challenges exist in modern IDS, in Section 4 the most used ML algorithms in the IDS are reviewed in details the main weaknesses and strengths of each algorithm is given in Section 4.6. Section 5 explain two types of optimization algorithms parameters containing and parameters less algorithms. Finally, conclusion on what have been done given in Section 6.

2. INTRUSION DETECTION

The work of an IDS is to monitor a system or network and detect any form of abnormal activities within the system. IDSs can either be host-based IDS or network-based IDS [2,3]. NIDSs identify attacks by analyzing specific network events while HIDS detects intruders in individual hosts. An NIDS scans a packet sniffer which is the program that reads the raw packets off a segment of the local network. It can also track more network objectives in a bid to detect threats that may be missed by HIDS since HIDSs cannot read packet headers and cannot detect certain attack types. For instance, NIDSs can effectively detect most of the IP-based DoS attacks because they can only view packet headers as they pass through the networks. However, NIDSs do not rely on the operating system (OS) of the host as identification sources but HIDSs depend on the OS to function properly. Hybrid IDSs that combine client and network-based technologies have also been developed for intrusion detection (ID) [4]. The ID techniques can also be in the form of misuse detection or anomalies detection [3]. The detection mechanism used in IDS are three main types which is: statistical method, machine learning (ML), and data-mining methods, Figure 1 summarize the IDS.

* *Corresponding author. Email: mohammad.cs88@gmail.com*

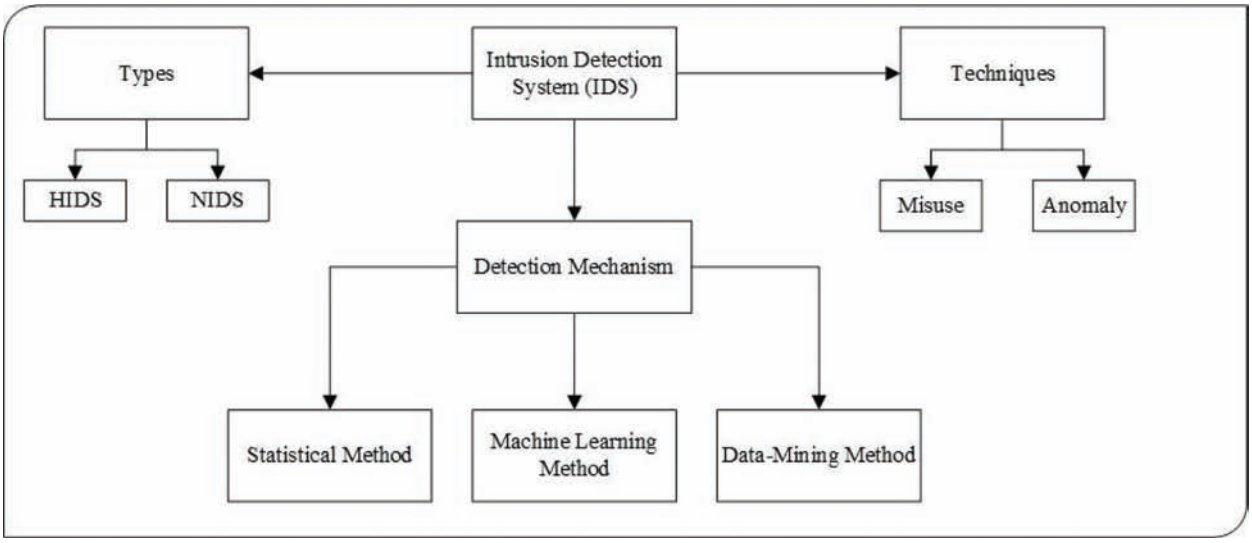

Figure 1 | Intrusion detection system (IDS) overview.

2.1. Misuse Detection

Detection of abuse detects intrusions through looking for known patterns of attacks. This strategy is employed by current commercial NIDSs. A downside to preventing misuse is that it can't spot unknown attacks. Specific methods, such as expert networks, signature analysis, state-transition analysis, and data mining, have been used to identify violations.

To describe intrusions, the expert system uses a set of rules [2]. Audit events are converted into facts in the expert framework which bear their semantic significance. Then, using certain rules and evidence, an inference engine will draw conclusions. Analysis of the state transition will identify attacks with certain goals based on the state-transition diagrams. Any network activity that triggers an intrusion state will be detected by the system as an intrusion. Signature-based analysis relies on the analysis of attack signatures during audit trail [2]; activities that matches already identified attack signatures are recognized as an attack. Data-mining techniques have been recently proposed for ID. Data mining is the process of extracting important but previously unnoticed information or pattern from a given set of data. For other ways, the models or patterns may be defined, such as rules, instance-based instances, decision chains, and neural networks (NNs). Data-mining frameworks are mainly used for misuse identification. Different data-mining processes also rely on the association rules algorithm [4]. Decision tree (DT) and association rules are also used in the detection of intrusion [5]. The IDS performance is enhanced with the NN algorithm [6].

2.2. Anomaly Detection

Because detection of misuse cannot identify unknown threats, detection of abnormalities is used to counter this shortcoming. Different approaches to anomaly detection clustering, classification, etc., were proposed and implemented [7]. Supervised detection of anomalies requires the create of normal network activities using the training dataset; then, intrusions are detected based of deviation profiles of the suspected activities. ADAM [8] builds typical attack profiles based on attack-free training set; the attack profile is described as a set of association rules. ADAM can perform real-time ID based on the attack profile. Other supervised methods, such as GA, fuzzy data mining [9], support vector machine (SVM), and NNs [10,11] are also used for ID. For supervised detection of anomalies, mathematical techniques and specialist structures are often applied [2]. Statistical methods create profile of users through a variety of examples of typical behavior. This also relies on attack profiles to detect abnormal activities. Expert systems rely on a set of rules to define normal user behavior; they depend on these rules to detect abnormal behaviors.

2.3. IDS Challenges

Studies have been ongoing on new systems for an automatic detection of abnormal system usages. Moreover, Denning reported the development of an intrusion detecting model, which he suggested as a framework for a general-purpose IDS [12]. Since then, experts have developed and applied several algorithms for automating the process of network ID. They have also continually pursued more accurate, faster, and scalable methods for this purpose. With the arrival of the "IoT" and Big data era, it is expected that the number of connected devices would exceed 26 billion by the year 2020 [13]. With this trend, the type and number of cybersecurity issues are also expected to increase. Figure 2 summarize the challenges in IDS. These challenges are false alarm rate, low detection rate, unbalanced datasets, and response time.

Some researchers have recently advocated for more categories of IDS. Liao *et al.* [14], for instance, claimed that IDS should be further categorized into 5 sub-categories which may belong to any of the aforementioned classes. The suggested sub-classes are pattern-based, rule-based, statistics-based, state-based, and heuristic-based IDS. Meanwhile, such a classification could result in confusion due to the number of similarities between the strength of the individual techniques, as well as the lack of clear criteria that distinguishes one technique from the other. The signature or rule-based IDS are normally associated with a certain degree of false positive (FP) alarm rates and are unable to detect novel forms of attacks [15]. It

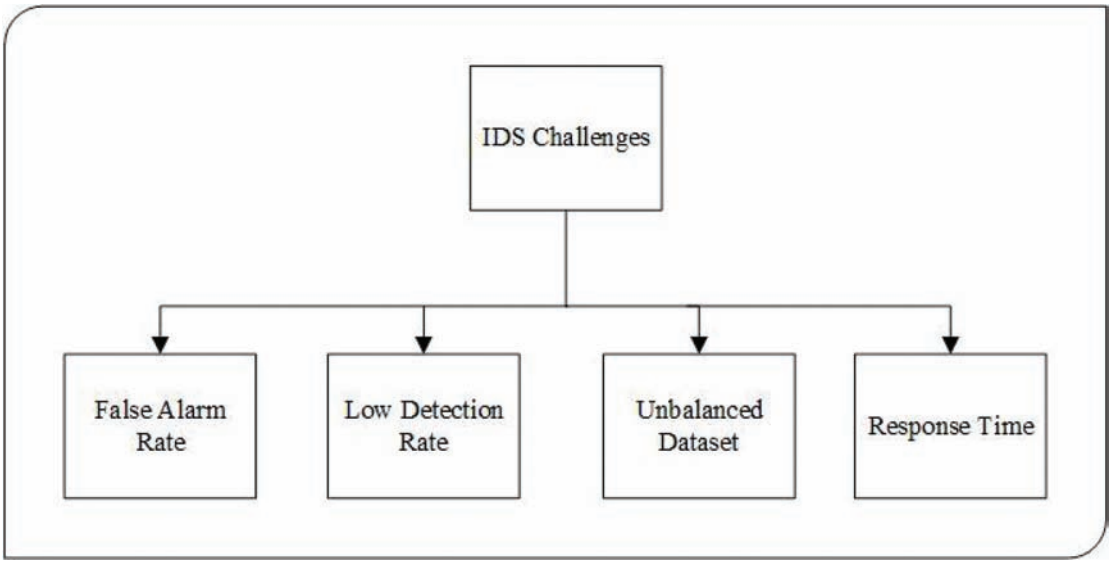

Figure 2 Intrusion detection system (IDS) challenges.

is expected that IDSs should exhibit a high level of FP detection. ID systems that are reliant on stateful protocol analysis exhibit different detection performances based on the level of their profile definition [16]. A major challenge of this approach is keeping an up-to-date profile as new protocols evolve over time.

As earlier discussed, this study is focused on the development of an anomaly-based IDS with a good accuracy and a minimal FP detection. Many studies have been performed on false alarm reduction in IDS. Pietraszek estimated that about 99% of ID alerts do not involve cybersecurity issues due to the observed slight differences between normal and malicious activities [17]. The other challenged pointed by Pietraszek include the development of accurate signatures that can detect abnormal behavior but are not triggered during normal network activities as some activities could be allowed in some instances but detected as intrusion at other instances. The Adaptive Learner for Alert Classification (ALAC) approach proposed by Pietraszek [18] uses ML techniques, coupled with a human-based observatory training to adaptively learn the implicit classification rules. Due to the involvement of human factor during the training, the ALAC system can be incrementally upgraded as the condition changes. As mentioned in the previous chapter, the main challenges of the current anomaly IDS are that the complication of building a system with such features is greater in the case of misuse detection [19]. Furthermore, a higher percentage of false alarms is raised [20], coupled with a low detection rate [21,22]. There is also an issue of the unbalanced dataset which impacts the evaluation of the models [15].

3. MACHINE LEARNING

ML was first defined by Arthur Samuel in 1959 as a field of study that confers computers with the ability to learn without the need of first being programmed [23]. Network security uses ML to make many important calculations and decisions in order to decide which packets to drop and which ones to allow in the system. This section focuses on ML algorithms used to build security tools with an emphasis on describing network ID techniques. According to [24], a key requirement for any technique to work, one needs a sound understanding of the system, with which we concur. To create a useful tool for any environment, the most important requirement is understanding the system as well as its functioning, capabilities, and limitations. In the quest to create secure systems, ML has proven to be one of the most powerful tools yet.

The study by Chandola *et al.* [7] covered most of the existing anomaly detectors and their application area. ML offers both supervised and unsupervised forms of learning for ID. Supervised learning uses identifies misuses via classifiers from training data that is labeled a priori. This property makes it an ideal choice for misuse detection. Whereas, data available using unsupervised learning is classless by grouping similar data. Therefore, it is usually employed in anomaly detection. And like all approaches, they come with certain disadvantages that need to be dealt with carefully when designing IDS. There is a third class called semi-supervised in which a portion of the data is labeled during acquisition of the data sets.

With misuse detection, the definition of misuse is ever evolving. An IDS based on misuse detection may fail easily if the knowledge base of potential misuses is not updated periodically, which is an expensive operation. Using algorithms for detecting anomalies requires a carefully created normal profile of the system. Similar to misuse detection, anomalies require the system to periodically acquire the latest definition of normal, as well as new ways of intrusions. This dramatically raises false alarm rates when previously unseen system behaviors may be categorized as threats. Therefore, anomaly detection problem, in its most general form, is also not easy to solve.

Despite these limitations, ML offers a lot of advantages and is widely used in any industry that uses large data sets. The usages are many and are not limited to modeling predictions, identifying anomalies, detecting threats, malware, frauds, and so on. Since the focus of this work is network-based ID, we present a summary of the research in the area of ML-based NIDS systems. Most ML methods have three phases—training, testing, and validation. They usually involve:

1. use small subset of the data (usually 10%) to identify features from training data

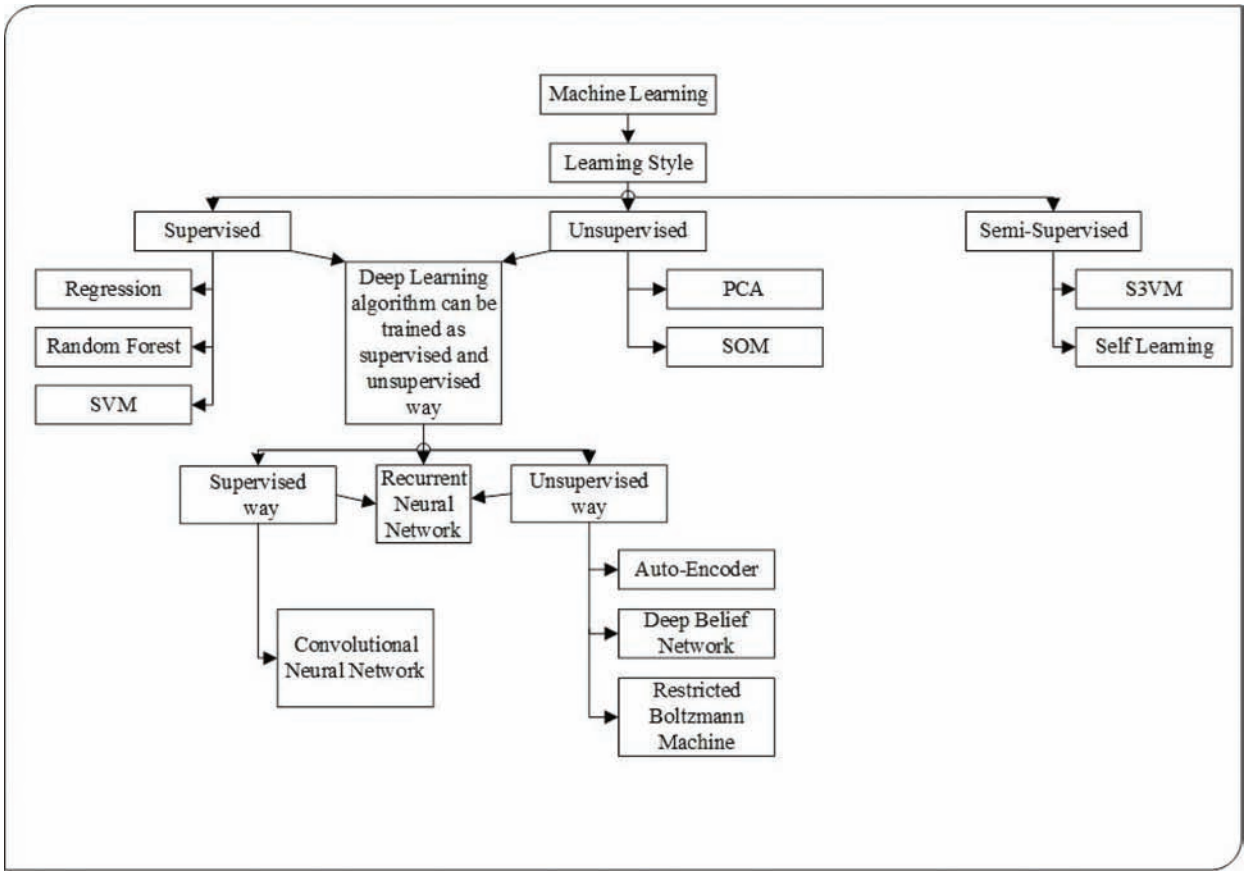

Figure 3 | Machine learning overview.

2. if the number of features is high, reduce the number of features through the process of dimensionality reduction
3. using the training dataset, learn the model and then classify another small subset of the data
4. using small subset of the data, validate the model, the one that performs the best on validation data is the model finally used

ML can be classified into 3 categories based on the learning style: supervised, unsupervised, and semi-supervised, each one of these categorize has many algorithms, as shown in Figure 3, which give an overview of ML approaches.

The next sub-section, provide a detailed refresher of some of the most widely used ML algorithms in detecting intrusions in a network,

3.1. Kk-Nearest Neighbor

Known as one the easiest to implement, and the most fundamental classifications approach, k-nearest neighbor (k-NN) is a widely used algorithm. This is due to its simplicity of use, and ease of learning complex functions. To classify a given dataset, k-NN looks at the k-NN and votes them based on the measurement calculated. The votes given to the majority value of the nearest neighbor determines what class to classify the dataset too. For instance, if the value of the variable k is 2, it will look at the two nearest neighbors (2-NN) and determine what to classify the dataset to by measuring it with respect to the 2-NNs. This measurement is commonly based on the Euclidean distance measured between a given test sample and a specified training sample of data [25]. Lixiang *et al.* [26] propose a novel method for IDS that combines KNN and density peaks clustering. This method relies on density peaks clustering for training purposes while KNN is for classification tasks. This novel method is validated using KDDCUP99 dataset and chose 15 features from the whole 41 features, this method succeed to improve the accuracy of the Probe attack in 15% compared to others method, but the main drawback of this method is the low accuracy of detection when detect U2R attack [27]. Another work by Yu Xue [28] which propose the self-adaptive differential evolution (SaDE) algorithm for feature selection in IDS and the k-NN is used to assess the chosen features by SaDE, KDDCUP99 intrusion dataset have been used in this work, also the original KDDCUP99 dataset is subdivided into four sub-datasets and on each one the applied SaDE and k-NN separately, 3-fold cross validation is employed to measure the classification accuracy of k-NN, this method divide the dataset into smaller four dataset which may not give the true performance if the method applied on the whole dataset [28]. Two-step hybrid IDS approach proposed by LONGJIE LI, this approach is based on binary classification and K-NN, in this approach more than one binary classifiers and one aggregation model used to identify the abnormal connections, C4.5 algorithm is used to train the binary classifiers, in first step, for the connection that are not certainly identify belongs to which class (normal or attack) in step one, they classified further in step 2 using the K-NN algorithm, this method shows a good result on a small dataset like NSL-KDD which is small part of

KDDCUP99, however two-step classification could be inefficient in the matter of time when applied on larger datasets like KDDCUP99 and CICIDS2017 because need to classify the data two times [29].

3.2. Artificial Neural Networks (ANN)

ANN models are developed to mimic the functionality of the biological NN [30]; they normally contain several layers in which the output of one layer serves as the input for the succeeding layer. During classification tasks, the final output of the output layer generates the final classification category. NNs were the earliest methods used for ID. Further proposals have been made for new intelligent IDS which depends on few features for ID. Such systems extract data features based on the correlation and information gain concept. Firstly, the extracted data features were ranked using correlation and information gain and then were combined by using suitable method. Redundant and irrelevant data are eliminated through a preprocessing phase to optimize resource commitment and minimize time complexity. ANN is used to construct the classification system and then train and test the proposed system on 5 different KDD99 dataset [31]. Investigation on deep neural networks (DNNs) ensemble techniques done by Simone A. Ludwig [32] on IDS; the study used the NSL-KDD dataset while the DNN was trained with the backpropagation NN. The proposed method was evaluated based on the false alarm rate, and detection rate; the evaluation showed that DNN achieved low false alarm rate and high detection rate for Dos and Probe but failed to efficiently capture U2R and R2L attacks. A scheme for ID based on RNN and restricted Boltzmann machines (RBMs) has been proposed by Chaopeng Li *et al.* [33]. This method makes use of the traffic data as input to obtain a fine-grained model at the byte level. The RBM is used to construct the network packet representation, and the features of these packets are captured as vectors that represent the packet feature. Then, an RNN model is used to process the temporal information between neighboring packets and the micro-flow representation is constructed as flow feature vectors. Two datasets were used in this study to evaluate the proposed model, ISCX-2012 dataset and DARPA-1998 dataset, three evaluation metrics are used in this study, accuracy, recall, and FP rate, the method shows a good performance regarding these three metrics [33]. Chuanlong yin *et al.* [34] propose a deep learning approach for IDS, based on recurrent neural network (RNN), preprocessing phase is done on the original NSL-KDD dataset to convert nonnumeric data to numeric data, the performance of this approach is measured in binary and multi class classification, three metrics accuracy, detection ate, and FP rate used in evaluation of this approach, the approach shoed good result in these three metrics for all attacks type, however in U2R and R2L attacks it shows low detection rate, also the training time still need some improvements [34]. Another model for IDS based on convolutional neural network (CNN) proposed by Yihan xiao [35], the proposed model consist of three steps, step 1 the symbolic data converted to numeric data and the redundant data is removed, in step 2, the CNN model is trained and the optimal features is obtained from the dataset, step 3, in this step revers fine tuning improve the model performance, back propagation algorithm fine-tune the network model parameters, after the optimal parameters are determined, the evaluation process of the model is start by the classification results of the test dataset, KDDCUP99 dataset is used, the performance evaluation of the model shows there are many works reached better results than this model. ANN also applied on the latest intrusion datasets such as CICIDS2017 [36]. Yong zhang *et al.* [37], propose a new model for IDS. The new model is comprised of two NN layers; the first layer is based on improved CNN and used for extraction of the spatial features of the traffic; the second layer serves in the extraction of the temporal features, these two layers (networks) are trained simultaneously to extract the spatial and temporal features of the flow, the method shows high accuracy results [37].

3.3. Decision Trees

It is mainly a classifier of attributes as prespecified in the tree. It has three essential components—a decision node that specifies the test attribute, an edge that characterizes the possible attribute values, and a leaf that represents the class of the object. Based on the DT concept, new instances can be classified starting from the root, and proceeding to the branch based on the related value of the considered object. An instance is considered classified upon reaching a leaf [38]. A rule injection method has been proposed by Dimitrios *et al.* [39] for IDS; this method is based on GA and DT. The method comprises of specific GA steps for the enhancement of the classification capability of the DT. The GA is applied on a set of DTs with the aim of searching the solution space of the problem by expanding them. The aim is to search all the possible solution paths that will later form the final detection rules. Each step introduces heuristic techniques to address the issues that are mainly caused by the nature of the task. The proposed method was evaluated on three datasets (NSL-KDD, UNSW-NB15, and KDDCUP99) and the results for this method shows that it still need for more improvements [40]. Kai Peng *et al.* [26], propose an IDS for big data, the proposed IDS based on DT and consist of three steps, step 1 is the preprocessing step, dataset contains both numeric and string data, the string convert to numeric data to reduce the complexity of computation, step 2, is the normalization process over the dataset, step 3 is the final step and DT classify the data and produce the final result, the method evaluated using KDDCUP99 dataset, and shows a good results [41]. Another work based on DT proposed by Gisung Kim *et al.* [42], the proposed method integrating misuse detection with anomaly detection, the model consist of DT and multiple 1-class SVM; the C4.5 DT was first used for the building of the misuse detection model for the decomposition of the normal training data into smaller subsets. Next, the one-class SVM (1-class SVM) was deployed for the building of the anomaly detection model in each of the decomposed region. KDDCUP99 dataset used to evaluate this method [42]. Hassan [43] presented a comprehensive analysis of several ML techniques in IDS, methods like DT, ANN, and random forest used in this study, and datasets like KDDCUP99, ISC2012, and the latest intrusion dataset CICIDS2017 are used in the evaluation process, results of the studied methods presented at the end with some statistical test regarding these results [27].

3.4. Bayesian

The Bayesian classifier (BC) relies on the idea that a (natural) class is expected to predict the feature values for its members. Many IDS researchers have used BC for the classification of network activities into normal and abnormal classes. Wajdi *et al.* [44] developed an anomaly ID method that consist of infinite bounded mixture model and Bayesian, infinite bounded mixture us used for feature selection

with the Bayesian, Bayesian work as a model selection, feature selection and parameter estimation for the infinite mixture model, three datasets used to evaluate this model, KDDCUP99, Kyoto2006+, and ISC2012, the model is designed for the IOT environment security [45]. Another work by Abdulhammed *et al.* [46] proposed feature reduction in IDS; in this work, auto encoder and principle component analysis (PCA) is used for feature reduction. Many classifiers have been used to design efficient IDS and the evaluations showed the relevance of feature reduction techniques in achieving good results based on some performance evaluation metrics. These results portray the possible usefulness of PCA and auto encoder in feature dimensionality reduction for IDS [46]. Most studies have focused on improving the detection rate of the four major types of intrusion using BC method. For instance, Altwaijry [47] succeeded in using BC as a data classification tool. The results of the study highlighted the ability of the approach to achieve better performance on R2L detection (DR = 85.35%) using 3 features (23, 24, and 31) and 0.6 threshold value [47]. Bayesian network (BN) classifiers with robust reasoning potentials have consistently been employed for ID and they have achieved reasonable efficiency and accuracy levels in such applications. The study by Xiao *et al.* [36] used BC that was built from Bayesian model averaging (BMA) over the k-best BN classifiers. The model was called Bayesian network model averaging (BNMA) classifier; firstly, subset feature from NSL-KDD dataset selected to build the classifier, random sample from the whole dataset used for training process and then the whole dataset is employed for model testing. Data discretization was also performed on the training and testing data features; the next step is finding the k-best BN structures based on the training set, followed by conditional probability distributions (CPDs) estimation for each network using Bayesian Estimation to generate the k-independent BN classifiers that will be applied to the testing dataset. Then, the accuracy of the four experimental groups is calculated by repeating the same steps on different sets of training data. Then, the average of the results for each classifier and each configuration of different sizes is calculated. The outcome of the study showed a good accuracy level of the classification process on NSL-KDD dataset [36].

3.5. Support Vector Machine

SVM is a supervised ML technique mainly used for classification and prediction [48,49]. To classify, it maps an input vector into a predefined high-dimensional feature space through some nonlinear mapping [50]. The feature space is constructed using an optimal separating hyperplane. As described in [50], if one can construct optimally separating hyperplanes with a small expectation of the number of support vectors, the generalization ability of the constructed hyper planes is high. Thus, SVM's biggest advantage is that it is effective in high-dimensional spaces. SVM separates hyperplanes into two classes: +1 and −1, where the positive value represents normal data and the negative one is for anomalous data.

Wenying Feng *et al.* [51], propose a new approach for IDS based on ant colony and SVM, the ant colony used to cluster the data before the classification process start, many classes may be under one cluster or vice versa, after that the SVM classify the clustered data into normal or 4 abnormal classes, KDDCUP99 dataset used to evaluate the proposed approach, the combined approach showed better results than original SVM and ant colony [51]. Another work by Enamul Kabir *et al.* [52] propose a novel statistical technique for IDS, the proposed approach based on sampling with least square SVM (LS-SVM). Decision-making in this method proceeds in 2 stages; the first stage involves partitioning of the whole dataset into a set of arbitrarily predetermined subgroups. Then, the proposed approach selects the sample representative that will reflect the entire dataset from these subgroups. The study also developed an optimum allocation scheme based on the observed variabilities within the subgroups. Then, the second stage is the application of the LS-SVM to the extracted samples for ID. The model was tested on the KDDCUP99 dataset and the approach showed good result for both static and incremental datasets [52]. Vijayanand *et al.* [53] propose an IDS for mesh networks based on SVM and genetic algorithm (GA), the proposed model consist of two steps, in step 1 GA wrapper feature selection method is used to select the optimal features for each class of attacks, the next step, step 2 is a multiple SVM classifier, each classifiers detect one type of attacks. A different classifier is assigned to each attack group and trained with the specific features of each attack data as selected by the proposed FS technique. The classifiers are linearly arranged, with each being placed based on the attack severity. For each classifier, the output is either belongs to the attack group or to the nonattack group except the output of the last classifier which belongs to a new category. If the data is classified as belonging to the attack category, the classifier will report the user for more processing; else, it will forward the input data to the next classifier to determine the category. This process is continued until the input data category is determined. The performance of the proposed system was evaluated on intrusion datasets generated by a wireless mesh network simulation process in NS3 based on the delay, packet delivery ratio, etc., as the metrics [53]. A novel SVM has been proposed by Kuang *et al.* [54] for IDS which combines kernel PCA (KPCA) with improved chaotic particle swarm optimization (ICPSO). This method uses multi-layer SVM to detect attack or normal activities while the KPCA is for feature dimensionality reduction. The ICPSO is used to optimize the SVM parameters; the aim of this study is to shorten the computational time and increase the predictive accuracy, small sample of KDDCUP99 dataset used to evaluate this model, the model shows shorten training and testing time, and increase the accuracy comparing to single SVM [54]. Another work does the feature selection and SVM parameters optimization in the field of IDS is presented by Bamakan *et al.* [49]. In this study, IDS frame work presented by using precise optimization techniques such as time varying chaos PSO (TVCPSO) to perform feature selection and parameter setting simultaneously for MCLP and SVM. The study aims to maximize the detection rate and reduce the rate of false alarm using the least number of features. This method was evaluated on the NSL-KDD and KDDCUP99 datasets and the results showed reduced false alarm rate and good detection rate for Probe and Dos attacks while failing to effectively capture R2L and U2R attacks [49]. The study by Raman *et al.* [55] proposed an adaptive IDS that relied on hypergraph-based GA (HG-GA) for feature selection and parameter setting in SVM. The parameters optimized in this study are C, γ for SVM each gen in the chromosome of GA is represent a feature in the feature space of the dataset, the dataset used in this study is NSL-KDD dataset, the study focus on improving the detection ate and false alarm rate, and it gain a good result in this domain compare to other work [55]. Another work based on subset feature selection and parameters optimization is presented by Tao *et al.* [56], GA and SVM is proposed in this work, GA is used to select optimal sub set of features first and optimize the SVM

Table 1 Comparative analysis of ML algorithms.

Technique		Notes	Ref.
k-NN	Weakness	Suffer from high dimensionality and overfitting	[57]
	Strength	Easy to implement	[58]
ANN	Weakness	Need more time for training The detection precision for low frequent attacks is low	[59]
	Strength	Better pattern recognition than other algorithms Self-organizing maps do not require label at input	[60]
DT	Strength	Better in accuracy and false positive Rate compared to the static models Require less training time	[42,61]
BN	Strength	Best algorithm when training data is scare Simple and high accuracy Reduce time complexity	[62]
SVM	Weakness	Dimensionality and size of dataset affect the training complexity Higher FP than other algorithms make it difficult to use Parameters like (c, γ) affect its performance if not tuned well	[56,63]
	Strength	Most use and overall best classifier Overfitting can be solved easy by chaining with other algorithms Process feature vector of high dimension	[63,64]

parameters (C, γ), and the SVM classify the data into normal or four intrusions type, new fitness function also has been proposed in this study, KDDCUP99 dataset used to evaluate the proposed algorithm, the result shows that the proposed algorithm succeed to reduce the classification time, FNR and FPR, and increase the detection rate [56].

3.6. Comparative Study of ML in IDS

Many ML frameworks have found application in IDS; however, each of these ML frameworks has weaknesses and strengths, and the effective of the algorithm applied in the IDS depends on many things, feature selection used or not, number of instance used to train the model, number of instances used to test the model, optimize the parameters of the algorithm (if any), etc., and choosing one of these algorithms as it is the best algorithm for IDS is not straight forward algorithm, many researchers work in the IDS explained, and Table 1 give an abstract for the five ML algorithms studied in the literature, and their weaknesses and strengths.

One of the best ML algorithms used in the IDS is SVM [65], however, SVM like other algorithms has weakness and strength, the strength can be summarized in three points [63], firstly, SVM is the most used ML algorithm in the IDS field and also is give the best overall accuracy among ML algorithms, overfitting which is a general problem in the ML algorithms can be easily solved in SVM by combining SVM with other algorithms like GA, PSO, teaching learning-based optimization algorithm (TLBO), JAYA, etc. [56], another strength in the SVM is its ability to process high dimension feature vector, which make it the best choice to bullied IDS for large intrusion dataset like KDDCUP99 (approximately 5 million records) and CICIDS2017 dataset (approximately 3 million records) [66].

However, even all these strengths in the SVM, still there are many weaknesses need to be solved before using SVM in the IDS field, one of SVM weaknesses is, dimensionality and size of dataset affect the training complexity, that is why feature selection mechanism need to be apply with the SVM before being used in the IDS [63,64], feature selection reduce the dimension of processed data and the size which yield to reduce the complexity and time needed to process the data, another weakness with the SVM is, higher FP than other algorithms make it difficult to use [63], parameters like (C, γ) affect the performance of SVM if not tuned well, these two weaknesses can be solved by tuning the parameters of SVM, that is way optimization algorithm needed to optimize the SVM parameters [56].

4. OPTIMIZATION ALGORITHMS WITH SVM IN IDS

To overcome the weaknesses in the SVM, researchers combine SVM with many optimization algorithms to improve the performance of SVM. This section will explain the most used optimization algorithms in the IDS, the optimization algorithm can be classified into two types weather it has or it hasn't parameters to tune during the execution.

4.1. Parameters Containing Algorithms

4.1.1. Genetic algorithm

This is an optimization method that was based on evolutionary computation and relies on the survival of the fittest concept. It relies on two operators—mutation and crossover to achieve optimality (Holland, [67]). There are many works uses GA with SVM in the field of IDS, Senthilnayaki *et al.* [68], use GA with SVM and propos a model for IDS, GA in this study used as feature selection to select the optimal feature, out of 41 feature in KDDCUP99, GA chose 10 only and these 10 features give better result than the 41 features [68]. Another study in the field of IDS uses GA with SVM is presented by Tao *et al.* [56], the proposed method uses GA along with SVM to choose the optimal subset of features and simultaneously the GA optimize the parameters of SVM and weighting the features, GA accelerate the algorithm performance and increase the accuracy reduce the classification time [56]. Kannan *et al.* [69] use GA with SVM to propose a method for IDS, in this work GA used for two purpose, firstly to remove the duplicate features and secondly, chose the best subset of features, this work succeed to reduce number of features and obtain better result [69].

4.1.2. Particle swarm optimization algorithm (PSO)

This metaheuristic was first proposed in 1995 by Eberhart and Kennedy [70] based on inspiration from social behaviors, such as flocking of birds and schooling of fish. PSO has various similarities with evolutionary computation just that it lacks the evolution operators. Various studies have attempted the combination of SVM and PSO to improve performance in IDS. Wang *et al.* [71] propose a method for IDS combining PSO with SVM, PSO in this study used

to optimize to important parameters of SVM which are C and λ. The result shows that the performance of the SVM is accelerated when using PSO, the classification time is reduced and the accuracy and detection rate are increased [71]. Another work by Wang *et al.* [72] try to combine PSO with SVM. In this study, the original PSO was used for the determination of the free parameters of SVM while the binary PSO was used to derive the optimum feature subset for building the IDS. The outcome of the evaluation showed better performance of the PSO-SVM method compared to regular SVM [72]. A model for FS, ID and parameter optimization has been proposed by Ma *et al.* [73] using binary PSO algorithm and SVM. This study aimed to find the better SVM parameters and a feature subset that represents the key features of network intrusion using the improved BPSO-SVM. From the evaluation report, the proposed approach performed better than the original SVM and some of the other existing IDS [73].

4.1.3. Ant colony algorithm

The ant colony algorithm (ACO) was developed by Dorigo *et al.* [74] as a novel heuristic for solving difficult combinatorial optimization tasks. The ACO was developed based on inspiration from the level of organization in natural ant colonies. Ants are endowed with the ability to find the shortest path to their colonies through deposition of pheromone along their path to guide the subsequent ants toward their food source or to their colony. The pheromone deposited by the ants along their path evaporates with time; hence, paths with less pheromone will become less popular while the path mostly chosen by the ants will keep having higher pheromone concentration as many ants will keep choosing it as the shortest path to their colony (this results to convergence in the ACO). Many works in the field of IDS have tried to combine ACO with SVM; for instance, Gao *et al.* [75] suggested a novel ID approach that combined ACO (for FS) and SVM (for ID). In this approach, the intrusion features are represented as graphere nodes where the edges denote the addition of new features. The ants travel through the graph to add nodes until the termination condition is met. The fisher discrimination rate is adopted as the heuristic information for the ants' traversal. To ensure than several SVM classifiers are not trained, the method succeed to minimize the number of features, thereby improved SVM performance. Another study by Wang *et al.* [71] combined ACO with SVM; the ACO was used for selection of features with a feature weighting SVM. The accuracy of SVM classification and feature subset dimension was used to establish a comprehensive fitness weighting index. Then, ACO was used for global optimization as its multiple search capabilities helped to reach optimality. The selected key network data features and the estimated IG access were used to access various features weights for building the SVM based on the features of network attacks. Lastly, the final design of the local search methods is refined to ensure no redundant features in the selected features while improving the convergence resistance. The performance of the algorithm was evaluated on KDD1999 dataset and the results show the ability of the approach to achieve effective features dimensionality reduction, thereby improving the accuracy and speed of network ID [76].

4.1.4. Limitations of parameters containing algorithm

Combining the optimization algorithm with the classification algorithm could improve the performance of the classifiers, Sections 5.1, 5.2, and 5.3 explain some of the work in the IDS that combine optimization algorithms with SVM, these three algorithms (PSO, GA, and ACO) are no longer can survive against the new types of attacks, all these algorithms suffer from many lacks and problems make it inefficient when facing new types of attacks, these algorithms are probabilistic algorithms and require specific parameters that need to be controlled when it make the experiment take long time to be performed [77] as different frameworks require specific control parameters. For instance, GA relies on mutation and crossover parameters as the selection operator while PSO relies on inertia weight, cognitive and social parameters; similarly, ACO requires tuning of the specific algorithmic parameters. It is important that the algorithm specific parameters are well tuned as they determine the performance of the specific algorithms. Improper parameter tuning could either increase the algorithmic computational effort or cause local optima entrapment. Hence, the need to use new designed optimization algorithms that overcome these lacks and problems has become the new way for the researchers in the field of IDS. In the next section we explain two newly proposed optimization algorithms that could be used in the field of IDS.

4.2. Parameters Less Algorithms

4.2.1. Teaching learning-based optimization algorithm (TLBO)

This algorithm was developed by [78] as an optimization strategy for solving mechanical design-related problems; it requires no user-defined parameter and has been evaluated on different benchmark functions where it has performed excellently compared to other algorithms. The feasibility of using the novel TLBO in the selection of optimal free SVM parameters has been studied by [79]. The results showed the success of the hybrid SVM–TLBO in finding the optimal parameters and achieved good predictions in comparison to the normal SVM. Another study by [80] extended the SVM-TLBO via introduction of a dimension reduction approach for minimizing the number of input variables using PCA, KPCA, and independent component analysis (ICA). The study by [81] reported the better performance of TLBO compared to some of the existing population-based optimization algorithms. The effect of sample size and number of iterations on TLBO performance has also been evaluated by Rao and Patel [82]; the study confirmed the possibility of applying TLBO on several optimization tasks. Kiziloz *et al.* [83] recommended 3 multi-objective TLBO frameworks as feature subset selectors in binary classification FSS-BCP. Among the 3 methods, MTLBO-ST presented as the fastest algorithms despite providing few numbers of non-dominated solutions. However, MTLBO-NS explored the solution space and achieved a set of nondominated solutions at reduced execution time. MTLBO-MD achieved similar solutions with MTLBO-NS within a significantly reduce period. Hence, the proposed MTLBO were evaluated for their performance using ELM, SVM, and LR.

4.2.2. JAYA optimization algorithm

This is a novel algorithm put together by Rao [77] that has found application in various intractable optimization problems. JAYA is different from most of the existing optimization algorithms by not requiring any form of parameters tuning [77]. It has been used as a benchmarking function for both constrained and unconstrained tasks. Despite requiring no parameters tuning (parameter-free), it differed from TLBO as it requires no form of learning by working only with the teacher phase [84]. JAYA works on the concept of finding the solution to problems by moving toward the best solution only while actively avoiding the low-quality ones. It relies on few control parameters, such as the number of design variables, the maximum number of generations, as well as the population size; it requires no specific control parameter that may require tuning before the computational phase. The study by [85] developed a new computer vision-based alcoholism identification method from healthy controls. There are three basic components of this method which are the proposed wavelet Renyi entropy, feedforward NN, and a three-segment encoded Jaya algorithm. The evaluations showed the method to reach a level of sensitivity though its accuracy requires some improvements. Parallel algorithms and their detailed analysis were presented by Hector *et al.* [86]. The study presented a hybrid framework that relies on inherent parallelism at 2 different levels. It exploited the lower and upper level via parallel and distributed shared memory platforms, respectively. It was reported that both algorithms achieved scalability; hence, the hybrid algorithm was found applicable to several processes with almost perfect levels of efficiency. The experiments showed that about 60 processes were required to achieve the ideal level of efficiency. The system was also analyzed with 30 unconstrained functions and found to obtain similar level of efficiencies for all the considered test functions. A novel E-JAYA algorithm has been proposed by [87] for improving the functionality of the traditional JAY. The E-JAYA relies on the average of the better and worse groups to establish optimality. The analysis showed that E-JAYA achieved better accuracy in comparison to the traditional JAYA. In the proposed E-JAYA, the behavior of the swarm was considered rather than just considering the behaviors of the best and worst individuals. Furthermore, the E-JAYA was deployed on 12 benchmark functions of different levels of dimensionality. An effective scheme for demand-side management has been presented by [88] for preventing peak creation while lowering electricity bill. This report employed 3 frameworks (JAYA, SBA, and EDE) to meet some set objectives; the TOU pricing scheme was also used for computing electricity bill. The outcome of the study showed that JAYA succeeded in achieving low electricity bill and keeping the PAR low while keeping the users happy. The SBA was also found to outperform JAYA and EDE in meeting user demands as it related inversely with the electricity bill. The study by [85] presented an improved JAYA (IJAYA) for accurate and steady PV model parameters estimation. The study introduced a self-adaptive weight in the IJAYA for adjusting the chances of reaching the best solution while avoiding the bad ones during active search. The weight was incorporated to aid the framework reach the possible search area early, as well as to perform the local search later. A learning strategy derived from the experience of other individuals was incorporated in the algorithm; this learning strategy is randomly employed for improving the population diversity.

The new proposed optimization algorithms (TLBO and JAYA) show their superiority to other optimization algorithms [77,81,83], these two algorithms are new proposed and not yet been applied in the field of IDS, applying these algorithms in IDS could give better result than other optimization algorithms, these two algorithms can be apply in feature selection, parameter optimization, and many other optimization fields in IDS. However, even these newly proposed algorithms still need some improvements to improve and enhance their work [87,89,57,58].

5. CONCLUSION

This paper represents literature review with several sections, begin with introduction to IDS, next section explain the IDS types and techniques specially based on ML. Moreover, next section includes analysis of ML algorithms such as ANN, BN, DT, KNN, and SVM, and state weaknesses and strengths for each one that lead to choose SVM. Optimization algorithms and their advantage when used in IDS explain in the next section, also the limitation of existing algorithms and the need for new parameter less algorithms in IDS explain that lead to choose TLBO and JAYA.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHORS' CONTRIBUTIONS

Methodology, Mohammad Aljanabi; validation, Mohammad Aljanabi , Mohd Arfian Ismail; formal analysis, Ahmed Hussein Ali; investigation, Mohammad Aljanabi , Mohd Arfian Ismail; data curation, Mohammad Aljanabi , Mohd Arfian Ismail; writing-original draft preparation Mohammad Aljanabi; writing-review and editing, Mohammad Aljanabi , Mohd Arfian Ismail; visualization, Ahmed Hussein Ali; supervision, Mohd Arfian Ismail; project administration, Mohd Arfian Ismail; funding acquisition, Mohd Arfian Ismail.

ACKNOWLEDGMENTS

This study was supported by Fundamental Research Grant Scheme (FRGS) with Vot No. FRGS/1/2018/ICT02/UMP/02/2: RDU190113 from Ministry of Higher Education (MOHE) and managed by University Malaysia Pahang.

REFERENCES

[1] A. Wespi, M. Dacier, H. Debar, "Intrusion detection using variable-length audit trail patterns," in *International Workshop on Recent Advances in Intrusion Detection*, Springer 2000, pp. 110–129.

[2] D. Barbará, J. Couto, S. Jajodia, and N. Wu, "An architecture for anomaly detection," in *Applications of Data Mining in Computer Security*, 2002, pp. 6376. https://books.google.iq/books?hl=en&lr=&id=QXNj15Lp1OsC&oi=fnd&pg=PR1&dq=D.+Barbar%C3%A1,+S.+Jajodia,+(2002).&ots=AttxJR5mXr&sig=pa6Gx83Nr5PpTiY_Oc7Lhh_2s4I&redir_esc=y#v=onepage&q=D.%20Barbar%C3%A1%2C%20S.%20Jajodia%2C%20(2002).&f=false

[3] D. Anderson, T. Frivold, and A. Valdes, "Next-generation intrusion detection expert system (NIDES): A summary," 1995. http://merlot.usc.edu/cs530-s08/papers/Anderson95a.pdf

[4] W. Lee, S. J. Stolfo, A framework for constructing features and models for intrusion detection systems, ACM Trans. Inf. Syst. Secur. 3 (2000), 227–261.

[5] S.M. Bridges, R.B. Vaughn, Fuzzy data mining and genetic algorithms applied to intrusion detection, in Proceedings of 12th Annual Canadian Information Technology Security Symposium, USA, 2000. https://www.csee.umbc.edu/csee/research/cadip/readings/DMID/005slide.pdf

[6] W. Lee, and S. Stolfo, "Data mining approaches for intrusion detection," 1998. https://www.usenix.org/legacy/publications/library/proceedings/sec98/full_papers/lee/lee.pdf

[7] V. Chandola, A. Banerjee, V. Kumar, Anomaly detection, ACM Comput. Surv. 41 (2009), 1–58.

[8] D. Barbara, J. Couto, S. Jajodia, L. Popyack, N. Wu, ADAM:detecting intrusions by data mining, in Proceedings of the IEEE Workshop on Information Assurance and Security, West Point, NY, USA, 2001. https://www.semanticscholar.org/paper/ADAM%3A-Detecting-Intrusions-by-Data-Mining-Barbará-Couto/d69ae114a54a0295fe0a882d205611a121f981e1?p2df

[9] J. Luo, S.M. Bridges, Mining fuzzy association rules and fuzzy frequency episodes for intrusion detection, Int. J. Intell. Syst. 15 (2000), 687–703.

[10] A. Bivens, C. Palagiri, R. Smith, B. Szymanski, M. Embrechts, Network-based intrusion detection using neural networks, Intell. Eng. Syst. Artif. Neural Netw. 12 (2002), 579–584.

[11] E. Tombini, H. Debar, L. Mé, and M. Ducassé, "A serial combination of anomaly and misuse IDSes applied to HTTP traffic," in *20th annual computer security applications conference,* (2004). pp. 428–437.

[12] D.E. Denning, An intrusion-detection model, IEEE Trans. Softw. Eng. SE-13 (1987), 222–232.

[13] A. R. Bauer, K. T. Burns, M. V. Esposito, P. L. O'malley, B. J. Olexa, and R. Mcmillan , "Monitoring system for determining and communicating a cost of insurance," ed: Google Patents, (2013).

[14] H.J. Liao, C.-H. R. Lin, Y.C. Lin, K.Y. Tung, Intrusion detection system: a comprehensive review, J. Netw. Comput. Appl. 36 (2013), 16–24.

[15] J.M. Fossaceca, T.A. Mazzuchi, S. Sarkani, MARK-ELM: application of a novel multiple kernel learning framework for improving the robustness of network intrusion detection, Expert Syst. Appl. 42 (2015), 4062–4080.

[16] M. Tavallaee, E. Bagheri, W. Lu, A.A. Ghorbani, A detailed analysis of the KDD CUP 99 data set, in 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, Canada, 2009.

[17] M. Moradi, M. Zulkernine, A neural network based system for intrusion detection and classification of attacks, in IEEE International Conference on Advances in Intelligent Systems-Theory and Applications, Luxembourg, 2004. https://www.researchgate.net/profile/Mehdi_Moradi2/publication/236030027_A_Neural_Network_Based_System_for_Intrusion_Detection_and_Classification_of_Attacks/links/00b7d515e02e36c76b000000.pdf

[18] T. Pietraszek, "Using adaptive alert classification to reduce false positives in intrusion detection," in *International Workshop on Recent Advances in Intrusion Detection*, Springer, 2004. pp. 102–124. https://link.springer.com/chapter/10.1007/978-3-540-30143-1_6

[19] S. Elhag, A. Fernández, A. Bawakid, S. Alshomrani, F. Herrera, On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on intrusion detection systems, Expert Syst. Appl. 42 (2015), 193–202.

[20] S.A.R. Shah, B. Issac, Performance comparison of intrusion detection systems and application of machine learning to Snort system, Future Gener. Comput. Syst. 80 (2018), 157–170.

[21] I. Raghav, S. Chhikara, N. Hasteer, Intrusion detection and prevention in cloud environment: a systematic review, Int. J. Comput. Appl. 68 (2013), 7–11.

[22] R. Singh, H. Kumar, R.K. Singla, An intrusion detection system using network traffic profiling and online sequential extreme learning machine, Expert Syst. Appl. 42 (2015), 8609–8624.

[23] A.L. Samuel, Some studies in machine learning using the game of checkers. II—recent progress, IBM J. Res. Dev. 11 (1967), 601–617.

[24] R. Sommer, V. Paxson, Outside the closed world: on using machine learning for network intrusion detection, in the 2010 IEEE Symposium on Security and Privacy, Berkeley/Oakland, CA, USA, 2010.

[25] L.E. Peterson, "K-nearest neighbor," *Scholarpedia,* (2009). Vol. 4, no. 2, 1883,

[26] K. Peng, V.C.M. Leung, L. Zheng, S. Wang, C. Huang, T. Lin, Intrusion detection system based on decision tree over big data in fog environment, Wireless Commun. Mob. Comput. 2018 (2018), 1–10.

[27] H. Azwar, M. Murtaz, M. Siddique, S. Rehman, Intrusion detection in secure network for cybersecurity systems using machine learning and data mining, in IEEE 5th International Conference on Engineering Technologies and Applied Sciences (ICETAS), Bangkok, Thailand, 2018.

[28] Y. Xue, W. Jia, X. Zhao, W. Pang, "An evolutionary computation based feature selection method for intrusion detection," *Security and Communication Networks*, 2018, (2018).

[29] L. Li, Y. Yu, S. Bai, Y. Hou, X. Chen, An effective two-step intrusion detection approach based on binary classification and *k*-NN, IEEE Access. 6 (2017), 12060–12073.

[30] B. Yegnanarayana, *Artificial neural networks*. PHI Learning Pvt. Ltd., 2009. https://books.google.iq/books?hl=en&lr=&id=RTtvUVU_xL4C&oi=fnd&pg=PR9&dq=B.+Yegnanarayana,+Artificial+neural+networks.+PHI+Learning+Pvt.+Ltd.,+2009&ots=Gd7YBmvEOE&sig=gs_raH5LW86f1r5-zBLi55Jp1ks&redir_esc=y#v=onepage&q=B.%20Yegnanarayana%2C%20Artificial%20neural%20networks.%20PHI%20Learning%20Pvt.%20Ltd.%2C%202009&f=false

[31] Akashdeep, I. Manzoor, N. Kumar, A feature reduced intrusion detection system using ANN classifier, Expert Syst. Appl. 88 (2017), 249–257.

[32] S.A. Ludwig, Applying a neural network ensemble to intrusion detection, J. Artif. Intell. Soft Comput. Res. 9 (2019), 177–188.

[33] C. Li, J. Wang, X. Ye, Using a recurrent neural network and restricted Boltzmann machines for malicious traffic detection, NeuroQuantology. 16 (2018), 823–831.

[34] C. Yin, Y. Zhu, J. Fei, X. He, A deep learning approach for intrusion detection using recurrent neural networks, IEEE Access. 5 (2017), 21954–21961.

[35] Y. Xiao, C. Xing, T. Zhang, Z. Zhao, An intrusion detection model based on feature reduction and convolutional neural networks, *IEEE Access,* , 7 (2019), pp. 42210–42219.
[36] L. Xiao, Y. Chen, C. K. Chang, Bayesian model averaging of Bayesian network classifiers for intrusion detection, in IEEE 38th International Computer Software and Applications Conference Workshops, Vasteras, Sweden, 2014.
[37] Y. Zhang, X. Chen, L. Jin, X. Wang, D. Guo, Network intrusion detection: based on deep hierarchical network and original flow data, IEEE Access. 7 (2019), 37004–37016.
[38] N.B. Amor, S. Benferhat, Z. Elouedi, Naive bayes vs decision trees in intrusion detection systems, in ACM Symposium on Applied Computing, Nicosia, Cyprus, 2004.
[39] D. Papamartzivanos, F.G. Mármol, G. Kambourakis, Dendron: Genetic trees driven rule induction for network intrusion detection systems, Future Generation Computer Systems, 79 (2018), pp. 558–574.
[40] D. Papamartzivanos, F.G. Mármol, G. Kambourakis, Dendron: genetic trees driven rule induction for network intrusion detection systems, Future Gener. Comput. Syst. 79 (2018), 558–574.
[41] L. Li, H. Zhang, H. Peng, Y. Yang, Nearest neighbors based density peaks approach to intrusion detection, Chaos Solitons Fract. 110 (2018), 33–40.
[42] G. Kim, S. Lee, S. Kim, A novel hybrid intrusion detection method integrating anomaly detection with misuse detection, Expert Syst. Appl. 41 (2014), 1690–1700.
[43] H.I. Ahmed, N.A. Elfeshawy, S.F. Elzoghdy, H.S. El-Sayed, O.S. Faragallah, A neural network-based learning algorithm for intrusion detection systems, Wireless Personal Communications, 97 (2017), pp. 3097–3112.
[44] W, Alhakami, A. ALharbi, S. Bourouis, R. Alroobaea, N. Bouguila, Network anomaly intrusion detection using a nonparametric Bayesian approach and feature selection, IEEE Access, 7 (2019), pp. 52181–52190.
[45] W. Alhakami, A. ALharbi, S. Bourouis, R. Alroobaea, N. Bouguila, Network anomaly intrusion detection using a nonparametric bayesian approach and feature selection, IEEE Access. 7 (2019), 52181–52190.
[46] R. Abdulhammed, H. Musafer, A. Alessa, M. Faezipour, A. Abuzneid, Features dimensionality reduction approaches for machine learning based network intrusion detection, Electronics. 8 (2019), 322.
[47] H. Altwaijry, Bayesian based intrusion detection system, in: H. Kim, S.I. Ao, B. Rieger (Eds.), IAENG Transactions on Engineering Technologies, Springer, Dordrecht, Netherlands, 2013, pp. 29–44.
[48] H. Tianfield, Data mining based cyber-attack detection, System Simul. Technol. 13 (2017), 3.
[49] S.M.H. Bamakan, H. Wang, T. Yingjie, Y. Shi, An effective intrusion detection framework based on MCLP/SVM optimized by time-varying chaos particle swarm optimization, Neurocomputing. 199 (2016), 90–102.
[50] V. Vapnik, *The nature of statistical learning theory*. Springer science & business media, 2013. https://www.dais.unive.it/~pelillo/Didattica/Artificial%20Intelligence/Old%20Stuff/2015-2016/Slides/SLT.pdf
[51] W. Feng, Q. Zhang, G. Hu, J.X. Huang, Mining network data for intrusion detection through combining SVMs with ant colony networks, Future Gener. Comput. Syst. 37 (2014), 127–140.
[52] E. Kabir, J. Hu, H. Wang, G. Zhuo, A novel statistical technique for intrusion detection systems, Future Gener. Comput. Syst. 79 (2018), 303–318.
[53] R. Vijayanand, D. Devaraj, B. Kannapiran, Intrusion detection system for wireless mesh network using multiple support vector machine classifiers with genetic-algorithm-based feature selection, Comput. Secur. 77 (2018), 304–314.
[54] F. Kuang, S. Zhang, Z. Jin, W. Xu, A novel SVM by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection, Soft Comput. 19 (2015), 1187–1199.
[55] M.R.G. Raman, N. Somu, K. Kirthivasan, R. Liscano, V.S.S. Sriram, An efficient intrusion detection system based on hypergraph - genetic algorithm for parameter optimization and feature selection in support vector machine, Knowl. Based Syst. 134 (2017), 1–12.
[56] P. Tao, Z. Sun, Z. Sun, An improved intrusion detection algorithm based on GA and SVM, IEEE Access. 6 (2018), 13624–13631.
[57] Z. Zhang, W.-C. Hong, J. Li, Electric load forecasting by hybrid self-recurrent support vector regression model with variational mode decomposition and improved cuckoo search algorithm, IEEE Access. 8 (2020), 14642–14658.
[58] Z. Zhang, W.-C. Hong, Electric load forecasting by complete ensemble empirical mode decomposition adaptive noise and support vector regression with quantum-based dragonfly algorithm, Nonlinear Dyn. 98 (2019), 1107–1136.
[59] H. Tang, Z. Cao, Machine learning-based intrusion detection algorithms, *Journal of Computational Information Systems 5* (2009) 1825–1831.
[60] Z. Yu, J. J. P. Tsai, T. Weigert, An Automatically Tuning Intrusion Detection System, *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics) 37* (2) (2007) 373–384.
[61] N. Wattanapongsakorn, S. Srakaew, E. Wonghirunsombat, C. Sribavonmongkol, T. Junhom, P. Jong-subsook, C. Charnsripinyo, A practical network-based intrusion detection and prevention system, *IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications* 2012.
[62] D. M. Farid, M. Z. Rahman, Anomaly Network Intrusion Detection Based on Improved Self Adaptive Bayesian Algorithm, *JCP 5* (2010) 23–31.
[63] F. Salo, M. Injadat, A.B. Nassif, A. Shami, A. Essex, Data mining techniques in intrusion detection systems: a systematic literature review, IEEE Access. 6 (2018), 56046–56058.
[64] T. Shon, J. Moon, A hybrid machine learning approach to network anomaly detection, Inf. Sci. 177 (2007), 3799–3821.
[65] F. Kuang, W. Xu, S. Zhang, A novel hybrid KPCA and SVM with GA model for intrusion detection, Appl. Soft Comput. 18 (2014), 178–184.
[66] E.A. Shams, A. Rizaner, A novel support vector machine based intrusion detection system for mobile ad hoc networks, Wireless Netw. 24 (2018), 1821–1829.
[67] J.H. Holland, Genetic algorithms, Scientific american, 267 (1992), pp. 66–73.
[68] B. Senthilnayaki, K. Venkatalakshmi, A. Kannan, Intrusion detection using optimal genetic feature selection and SVM based classifier, in 2015 3rd International Conference on Signal Processing, Communication and Networking (ICSCN), Chennai, India, 2015.
[69] A. Kannan, G.Q. Maguire, A. Sharma, P. Schoo, Genetic algorithm based feature selection algorithm for effective intrusion

detection in cloud networks, in IEEE 12th International Conference on DataMining Workshops, Brussels, Belgium, 2012.
[70] J. Kennedy, R. Eberhart, Particle Swarm Optimization (PSO), in Proceeding of IEEE International Conference on Neural Networks, Perth, Australia, 1995.
[71] G. Wang, S. Chen, J. Liu, Anomaly-based intrusion detection using multiclass-SVM with parameters optimized by PSO, Int. J. Secur. Appl. 9 (2015), 227–242.
[72] J. Wang, X. Hong, R.R. Ren, T. Li, A real-time intrusion detection system based on PSO-SVM, in The 2009 International Workshop on Information Security and Application, Qingdao, China, 2009.
[73] J. Ma, X. Liu, S. Liu, A new intrusion detection method based on BPSO-SVM, in 2008 International Symposium on Computational Intelligence and Design, Wuhan, China, 2008.
[74] M. Dorigo, V. Maniezzo, A. Colorni, Ant system: optimization by a colony of cooperating agents, IEEE Trans. Syst. Man Cybern. Part B Cybern. 26 (1996), 29–41.
[75] H.H. Gao, H.H. Yang, X.Y. Wang, Ant colony optimization based network intrusion feature selection and detection, in 2005 international conference on machine learning and cybernetics, 6 (2005), pp. 3871–3875.
[76] X. Wang, ACO and SVM selection feature weighting of network intrusion detection method, Int. J. Secur. Appl. 9 (2015), 129–270.
[77] R. Rao, Jaya: a simple and new optimization algorithm for solving constrained and unconstrained optimization problems, Int. J. Ind. Eng. Comput. 7 (2016), 19–34.
[78] R.V. Rao, V.J. Savsani, D.P. Vakharia, Teaching–learning-based optimization: a novel method for constrained mechanical design optimization problems, Comput. Aided Design. 43 (2011), 303–315.
[79] S.P. Das, S. Padhy, A novel hybrid model using teaching-learning-based optimization and a support vector machine for commodity futures index forecasting, Int. J. Mach. Learn. Cybern. 9 (2018), 97–111.
[80] S.P. Das, N.S. Achary, S. Padhy, Novel hybrid SVM-TLBO forecasting model incorporating dimensionality reduction techniques, Appl. Intell. 45 (2016), 1148–1165.
[81] R.V. Rao, V.J. Savsani, J. Balic, Teaching–learning-based optimization algorithm for unconstrained and constrained real-parameter optimization problems, Eng. Optim. 44 (2012), 1447–1462.
[82] R.V. Rao, V. Patel, An improved teaching-learning-based optimization algorithm for solving unconstrained optimization problems, Scientia Iranica, 20 (2013), pp. 710–720.
[83] H.E. Kiziloz, A. Deniz, T. Dokeroglu, A. Cosar, Novel multiobjective TLBO algorithms for the feature subset selection problem, Neurocomputing. 306 (2018), 94–107.
[84] M. Alsajri, M.A. Ismail, S. Baqi, A review on the recent application of Jaya optimization algorithm, in 1st Annual International Conference on Information and Sciences (AiCIS), Fallujah, Iraq, 2018.
[85] K. Yu, J.J. Liang, B.Y. Qu, X. Chen, H. Wang, Parameters identification of photovoltaic models using an improved JAYA optimization algorithm, Energy Conv. Manag. 105 (2017), 742–753.
[86] H. Migallón, A. Jimeno-Morenilla, J.-L. Sanchez-Romero, Parallel improvements of the Jaya optimization algorithm, Appl. Sci. 8 (2018), 819.
[87] C. Gong, An enhanced Jaya algorithm with a two group adaption, Int. J. Comput. Intell. Syst. 10 (2017), 1102–1115.
[88] O. Samuel, N. Javaid, S. Aslam, M.H. Rahim, JAYA optimization based energy management controller for smart grid: JAYA optimization based energy management controller, in 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2018.
[89] K.Z. Zamli, F. Din, S. Baharom, B.S. Ahmed, Fuzzy adaptive teaching learning-based optimization strategy for the problem of generating mixed strength t-way test suites, Eng. Appl. Artif. Intell. 59 (2017), 35–50.