A Review on Feature Selection and Ensemble Techniques for Intrusion Detection System

Majid Torabi[1]*, Nur Izura Udzir [2]*, Mohd Taufik Abdullah[3], Razali Yaakob[4]
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia, 43400 Serdang
Selangor, Malaysia

***Abstract*—Intrusion detection has drawn considerable interest as researchers endeavor to produce efficient models that offer high detection accuracy. Nevertheless, the challenge remains in developing reliable and efficient Intrusion Detection System (IDS) that is capable of handling large amounts of data, with trends evolving in real-time circumstances. The design of such a system relies on the detection methods used, particularly the feature selection techniques and machine learning algorithms used. Thus motivated, this paper presents a review on feature selection and ensemble techniques used in anomaly-based IDS research. Dimensionality reduction methods are reviewed, followed by the categorization of feature selection techniques to illustrate their effectiveness on training phase and detection. Selection of the most relevant features in data has been proven to increase the efficiency of detection in terms of accuracy and computational efficiency, hence its important role in the design of an anomaly-based IDS. We then analyze and discuss a variety of IDS-based machine learning techniques with various detection models (single classifier-based or ensemble-based), to illustrate their significance and success in the intrusion detection area. Besides supervised and unsupervised learning methods in machine learning, ensemble methods combine several base models to produce one optimal predictive model and improve accuracy performance of IDS. The review consequently focuses on ensemble techniques employed in anomaly-based IDS models and illustrates how their use improves the performance of the anomaly-based IDS models. Finally, the paper laments on open issues in the area and offers research trends to be considered by researchers in designing efficient anomaly-based IDSs.**

***Keywords*—Intrusion detection system (IDS); anomaly-based IDS; feature selection (FS); ensemble**

I. Introduction

Intrusion detection system (IDS) is one of the widely used security mechanisms intended to protect computers, programs, networks, and information against intrusion, illegitimate access, alteration, or demolition. At a minimum, security systems for computers (host) and networks requires firewalls, antivirus applications and IDSs. Intrusion detection aims to detect acts performed against information systems by intruders, which attempt to gain illegitimate access to a computer asset (data, information, and network). Intruders may involve local or remote intruders: local intruders are network users with some level of legal access which attempts to elevate their levels of access through abuse of unauthorized privileges; while remote intruders refer to users who attempt to obtain illegal access to device data outside the target network [1], [2].

One of the techniques utilized to construct the intrusion detection system in order to track and deter attacks are machine learning (ML) algorithms. These techniques analyze and distinguish between from normal and abnormal packets, are attempting to avoid system harm from the attack.

A. Challenges

Among the main challenges in IDS research are related to the selection of relevant data to be investigated. For training expert machine in IDS datasets, there are some attributes which are irrelevant or may not influence the final results and also increases the execution time. The strain on expert machine is minimized by eliminating these attributes by utilizing dimensionality reduction techniques [3], [4].

The other challenge is to build an appropriate feature subset selection to be used in the process of intrusion detection, which will not only reduce the detection time but also increase the accuracy of detection. In addition, generating proper feature subset helps expert machine by avoiding over fitting issue and enhances predictive performance [5], [6].

Utilizing suitable machine learning algorithms in order to detect the intrusion is another challenge in IDS. Threats and security landscape are become increasingly complex, and strategies based on low-level machine learning are ineffective in coping with rising security issues.

There are various machine learning algorithms and methods but the main challenge is to select the one that yields optimal performance for the IDS model [7], [8].

B. Motivation

The above challenges have inspired the discussion in this review paper, which focuses on application of machine learning algorithms for feature selection and ensemble-based detection in anomaly-based IDSs. The papers are classified base on mentioned issues and addressed their objectives and attributes. This study discusses and compares the models and examines the particularities of each model in order to promote more studies in this area.

C. Previous Study

Numerous existing review articles on IDS concentrated on feature selection or detection mechanism without taking into account the future trends and open topics. Anomaly-based intrusion detection has already been studied in several review articles [9], [10]. They described various elements of IDS but did not discuss articles in-depth, the strengths and weaknesses

*Corresponding Author

in different methods of feature selection and ensemble detection for anomaly-based intrusion detection system. Furthermore, future trends and open issues are also addressed.

D. Contributions

The contributions of this review are summarized below:

- Classification of detection methods in IDS. The methods, feature selection technique used, classification type, evaluation tool and dataset are all mentioned.
- Classification of machine learning techniques used in anomaly-based IDS.
- Identification of feature selection for anomaly-based IDS, as summarized in Tables II and III.
- Identification of ensemble classification for anomaly-based IDS.
- Presentation of future directions on the state-of-the-art anomaly-based feature selection and ensemble classification.

To achieve the mentioned contributions, some research questions are ready for this analysis and the responses are given in the following sections:

RQ1. What are the detection methods utilized for IDS?

RQ2. Which evaluation tools are utilized to assess the effectiveness of the IDS?

RQ3. What are the datasets reported in the review to be used in anomaly-based IDS?

RQ4. Which feature selection methods are used for anomaly-based IDS?

RQ5. What are the machine learning algorithms used for detecting intrusions in anomaly-based detection?

RQ6. What of the ensemble techniques included in the review are reported to be used in anomaly-based IDS?

The remainder of this paper is organized as follows: Section 2 provides an overview of detection methods in IDS. Then, the taxonomy of machine learning algorithms and their methods which are employed in IDS follows in Section 3, while Section 4 reviews and compares different techniques of feature selection. Next, ensemble classification algorithms and their methods which are employed in anomaly-based IDS follows in Section 5. In Section 6, discussions on the open issues and future trends for IDSs are provided, and finally Section 7 concludes this survey.

II. Detection Methods

Intrusion detection methods are classified into four groups based on the detection method used in the system: signature-based, anomaly-based, specification-based, and hybrid. In signature-based detection the IDS identifies threats when the system or network operation matches the threat pattern (called signature) stored in the IDS local databases, and an alert will be activated. Signature-based IDSs are effective and efficient in identifying existing attacks, and their task is simple to comprehend. However, this technique has not been effective in identifying zero day attacks and new variants of previously identified attacks which are still elusive as the associated signature for these attacks [11]. Signature-based schemes offer very strong outcomes for popular, well-known threats. However, they are unable to identify new, unseen attacks, even though they are designed as minimal variants of attacks previously identified. Examples of signature-based IDSs are Artificial Immune System (AIS) [12], Collaborative Block Chained Signature-Based IDS (CBSigIDS) [13], IPFIX-based IDS (FIXIDS) [14].

Anomaly-based detection aims to predict the system's "ordinary" pattern to be covered and produce an anomaly warning whenever the difference between an immediate occurrence and normal pattern reaches a predetermined threshold. The key benefit of anomaly-based detection method is their ability to recognize previously undiscovered attack incidents. Nevertheless, in anomaly-based systems, the rate of false positives (FP), or wrongly defined as attacks is typically higher than that of signature-based method, considering the possible inaccuracy in formal signature specifications. Examples of anomaly-based IDSs are Hybridized Feature Selection Approach (HFSA) [15], Hybrid Anomaly Detection Model (HADM) [16], Unsupervised Heterogeneous Anomaly Based IDS [17].

For specification-based detection method, a human expert manually constructs the desired template which consists of a series of rules (specifications) that aim to evaluate valid behavior of a device. If the parameters are sufficiently accurate, the template may identify unlawful patterns of behavior. In addition, the false positive rate is decreased, primarily because benign behaviors that were not previously observed are not flagged as intrusions in this type of system. Specifications could also be created using some formal tool, for example, with a sequence of states and their transitions, the Finite State Machine (FSM) methodology appears suitable for modelling network protocols [18] [19]. Standard languages of representation such as LOTOS, UML and N-grammars can be considered for this reason.

Hybrid detection aims to benefit from the strengths of each intrusion detection method, minimized their weaknesses and build strong schema to detect the intrusion. A notable aspect in hybrid detection is common uses of a key signature-based detection system in conjunction with an additional anomaly-based model. This integration of the two forms of detection strategies in a "Hybrid NIDS" [20] aims to increase the final accuracy of signature-based models for intrusion detection while eliminating the usual high level of false positives of network-based IDS (NIDS) approaches, hence a hybrid approach is embraced by most existing platforms. Other examples of hybrid are Signature-Based Anomaly Detection Scheme (SADS) [21], Artificial Bee Colony and Artificial Fish Swarm (ABC-AFS) [22], Hybrid Intrusion Detection Approach In Cloud Computing (HIDCC) [23].

Table I shows the type of detection methods utilized by researchers in IDS. RQ1, RQ2, and RQ3 are all answered in detail in the table. It specifies the detection method, the

evaluation tool, dataset used in the articles, and so on. From the table it is apparent that signature-based and specification-based detection methods did not utilize feature selection and ensemble classifier to detect intrusions, in contrast to the anomaly-based detection which utilized both of them. For evaluation tools and dataset, signature-based and specification-based models were deployed and validated using simulation and real data while anomaly-based approaches were evaluated by experiments and standard IDS datasets. The NSL-KDD dataset is the most utilized dataset based on the articles in this review.

TABLE I. COMPARISON OF DIFFERENT DETECTION METHODS FOR IDS

Author	Signature-based	Anomaly-based	Specification-based	Hybrid	Feature selection	Single classifier	Multi classifier/ Ensemble	Evaluation Tool	Dataset
2011 [30]	×	×	√	×	×	×	×	Simulation	Real Data
2012 [31]	×	×	√	×	×	×	×	Simulation	Real Data
2013 [32]	×	√	×	×	√	√	×	Experiment	NSL-KDD
2014 [33]	√	×	×	×	×	√	×	Experiment	DARPA 1999 Real Data
2015 [15]	×	√	×	×	√	×	√	Experiment	NSL-KDD
2015 [21]	×	×	×	√	×	×	√	Experiment	DARPA 1999 ISCX 2012
2016 [34]	×	√	×	×	√	√	×	Experiment	KDD Cup 99 NSL-KDD
2017 [35]	×	×	×	√	√	×	×	Simulation	Real Data
2017 [36]	×	×	√	×	×	×	×	Simulation	Real Data
2018 [37]	√	×	×	×	×	×	×	Simulation	Real Data
2018 [22]	×	×	×	√	√	×	√	Experiment	NSL-KDD UNSW-NB15
2018 [14]	√	×	×	×	×	×	×	Simulation	Real Data
2019 [13]	√	×	×	×	×	×	×	Simulation	Real Data
2019 [38]	×	√	×	×	√	×	√	Experiment	NSL-KDD
2020 [39]	×	√	×	×	×	×	√	Experiment	CICIDS-2017 CSIC-2010v2 UNSW-NB15 NSL-KDD
2020 [40]	×	√	×	×	√	×	√	Experiment	ISCX 2012 NSL-KDD CIC-IDS2017
2020 [41]	×	×	×	√	×	×	√	Experiment	Real Data
2020 [6]	×	√	×	×	√	√	×	Experiment	KDD Cup99 NSLKDD UNSW-NB15
2020 [42]	×	√	×	×	×	×	√	Experiment	ADFA NSL-KDD
2021 [43]	×	√	×	×	√	√	×	Experiment	UNSW-NB15

III. Machine Learning in Anomaly-based IDS

Machine learning (ML) algorithms are classified into unsupervised learning and supervised learning, depending on the availability of training dataset and the successful outcome of learning algorithms. Fig. 1 illustrates the taxonomy of machine learning algorithms in anomaly-based IDS. Regarding RQ5, it has been noted that most studies focus on the following algorithms for IDS.

In supervised learning, the training function is provided with input and target output pairs, and an expert model is trained to predict the output of functions at minimal expense. Supervised learnings are classified based on learning algorithms, frameworks and objective functions. Support vector machine (SVM), decision trees, and artificial neural network (ANN) are common categorizations.

For unsupervised learning there is no tag or label in the sample dataset. Unsupervised learning algorithms are proposed to simplify the data's key features and shape clusters of natural input patterns due to a particular cost function. Hierarchical clustering, K-means clustering, and self-organization map are the most common unsupervised learning methods. One of the challenges of unsupervised training is that it is hard to evaluate because it does not have a specific educator and therefore does not have labelled test data.

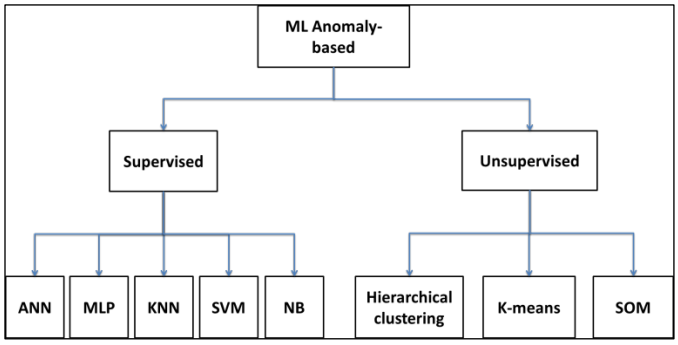

Fig. 1. Taxonomy of Machine Learning Algorithms in Anomaly-based IDS.

A. Supervised learning

1) Artificial Neural Network (ANN): ANN is one of the major algorithms of machine learning which is widely utilized as a detector operator in IDSs in many studies. ANN is used to solve a variety of issues faced by other existing intrusion detection approaches and has been suggested as a substitute for the statistical analysis aspect of detection of anomaly schemes. Initially, the ANN acquires its expertise by training the machine to properly detect pre-selected examples of problems. The neural network result will be checked and the machine configuration will be optimized until the training data neural network response reaches a sufficient level. Besides the initial training phase, the neural network often gains expertise over time as it performs review of the problem-related data [24], [25] . A hypervisor layer anomaly detection system called Hypervisor Detector that employs a combination algorithm that is a hybrid of Fuzzy C-Means clustering algorithm and Artificial Neural Network (FCM-ANN) was introduced to enhance the detection system accuracy [26]. The KDD Cup 99 sample dataset is used to test the design system to test for the reliability of five attack forms. The model was good at finding normal and probe attacks, but did not yield good results for DOS (99.96–5.33), U2R (96.78–3.22) and R2L (93.73–6.27) attacks, even for accuracy and false alarm rate. A reasonable solution using ANN in hierarchical anomaly-based IDS can be pointed to [27], which used neural Self Organization Map (SOM) networks to identify and distinguish normal packet from the attack traffic. The proposed machine was used to configure, train and evaluate the SOM Neural Network for intrusion detection. Detection output was performed to evaluate the SOM efficiency in detecting anomaly intrusion and the findings show that SOM with the KDD Cup 99 dataset can distinguish attack packet from normal one at 92.37%, while with NSL-KDD the detection rate is 75.49%.

The work in [28] tackles detection problems by presenting a simple ANN-based IDS system, utilizing back propagation and feed forward algorithms together with different other optimization methods to minimize the total computing overhead while maintaining a high level of performance. Results of the experiment on the NSL-KDD benchmark dataset showed that the quality of the proposed ANN (accuracy and detection speed) was 98.86% for accuracy and 95.77% for detection rate. An effective method to identify brute force attack in the Secure Shell (SSH) was proposed by [29]. A brute force attack is performed by the implementation into the private cloud of a client-server SSH model and the server captures traffic related to attack and normal. Next, ANN's Multi-Layer Perceptron model extracts indicative traffic characteristics and uses them to distinguish the attack and normal packets. Results obtained from this approach indicate that the suggested framework is able to detect the attack successfully with great accuracy and minimal false alarm.

2) Multi-layer Perceptron Neural Network (MLP): MLP is a supervised learning classifier which utilizes back propagation algorithm in the learning phase to train the model. It can learn a non-linear approximate function for both regression and classification task, by providing a group of features and a target in which one or more non-linear layers called hidden layers between the inputs and outputs are distinguished from logistic regressions [44].

The MLP neurons are positioned in layers with always-flowing outputs toward the output layer, either one layer (called a perceptron) or a multilayer perceptron, if multiple layers exist [45], where every neuron in a single layer has direct connections to the subsequent layer's neurons. The units of those networks apply a sigmoid function as activation function in many applications.

A wrapper-based feature selection is designed by utilizing the Discernibility Function as algorithm for search to construct subsets of feature and the MLP classifier is used to determine the subsets of features. Thus, the C4.5 decision tree and the MLP classifier, which are commonly utilized in the IDS, are

used to illustrate better classification rates. With this hybrid method, the findings for the KDD Cup 99 shows improved accuracy of approximately 12% for U2R, 2% for Probe, and 1% for DOS classes [46].

To build effective IDS, a hybrid multi-layer perceptron (MLP) and Artificial Bee Colony (ABC) algorithm were designed. The MLP classifier was used to distinguish among the attack and normal traffic in network. Training and testing have been conducted using the NSL-KDD dataset. Results of the experiments show that the suggested solution gives a high detection rate of about 87.54 % and error rate of 0.124% [47].

3) K-nearest neighbor (KNN): KNN algorithm is a nonparametric technique for classification and is a simple and straightforward machine learning algorithm. It is vast used based on many experiments reported on intrusion detection, pattern recognition, text categorization and countless others [48].

A combination of the Learning Vector Quantization ANN and KNN method for intrusion detection was suggested by [49]. The analysis was performed on the NSL-KDD dataset and the proposed model has a detection rate of 97.2% (five classes) with a false alarm rate of approximately 1%.

4) Support Vector Machines (SVM): SVM is one of the algorithms in machine learning that used labeled instance (packet) to train the model and differentiate the packet to different classes by generating templates that could determine which class a new instance belongs into [50], [51]. SVM's main objective is to discover a linear optimized hyper plane that maximizes the isolation boundary between groups. The SVM then trains the model across sections or portions of the data[52].

A hybrid intrusion detection KPCASVM with GAs design was proposed [53], where KPCA is implemented in the N-KPCA-GA-SVM system to obtain the key data features of intrusion detection, and a multi-layer SVM classifier is used to determine normal or attack behavior. The test was conducted on the KDD Cup 99 dataset and the detection rate was 96%. BIRCH hierarchical clustering SVM-based network intrusion detection framework [54] was proposed for pre-processing of data. Instead of the original large data set, the BIRCH hierarchical clustering could provide the SVM learning with highly qualified, abstracted and reduced data sets. The proposed solution could achieve a 95.72% accuracy with a false positive rate of 0.7% overall, but was not satisfactory with the division accuracy for each attack type (Prob= 97.55%, U2R=19.73% and R2L=28.81%).

A new Combining Support Vectors with Ant Colony (CSVAC) algorithm was proposed to produce cluster classifiers in intrusion detection [55] using two existing machine learning techniques (SVM and CSOACN) to improve overall detection rates and speed. The method is applied and tested using the standard KDD Cup 99 dataset benchmark, and yields a classification rate of 94.86% with a false negative ratio of about 1% and the false-positive ratio of 6.01%.

5) Naive Bayes Network (NB): Naive Bayes (NB) is a simple method of creating classifiers that allocate labels of class to problematic cases identified as values of feature vectors, where class tags are drawn from a restricted set. There is no single algorithm for learning such classifiers but a set of algorithms based on a common concept. A Directed Acyclic Graph (DAG) usually describes the structure of an NB, that each node represents a process variables and each reference encodes one node's control over another [56]. By comparing the decision tree and Bayesian techniques, the decision tree's accuracy is much higher but the processing time of the Bayesian network is low [57]. Therefore, it will be effective to use NB models when the dataset is very large.

A Naive Bayes-based IDS which obtained better findings than neural network IDS while tested on the KDD Cup 99 was proposed by [58]. The average accuracy obtained by utilizing Naive Bayes was 91.52%. While being basic in design, it can produce accurate results. A hybrid intrusion detection system based on Naive Bayes and decision tree was proposed by [59]. The model has been compared and tested using benchmark KDD Cup 99 dataset, the detection rate was 99.63%. A Fuzzy Intrusion Recognition Engine (FIRE) Intrusion Detection System Simple data mining approaches used to process network stimulus data packets and reveal essential anomaly detection indicators was developed by [60]. Such indicators were accessed for each observed value and used afterwards to classify network attacks. An intrusion detection model with information gain for feature selection and SimpleCart algorithm to detect the intrusion was suggested by [61]. First, the features were reduced to 33 and then the SimpleCart algorithm used for detection. The model was applied on NSL-KDD dataset and the detection accuracy was 82.32% and error rate was 17.67%. A hybrid strategy to learning is suggested by integrating Naive Bayes and K-Means clustering classifier. The suggested solution has been compared and tested using the benchmark dataset KDD Cup 99. These combinations learning methodology achieved rather low error rates with an average of less than 0.5% while retaining accuracy and detection rates above 99%. The method is capable of accurately classifying all data except the U2R and R2L attacks. to overcome this limitation, it was recommended to consider the Integrated Intrusion Detection Program which is ideal for identifying R2L and U2R threats [62], [63]. In SSH traffic, a combination of Bayesian Network and Genetic Algorithm was introduced to improve identification of brute force attacks [64]. The proposed method implements brute force attack data obtained in a client-server model. Their findings show that the most effective features were chosen and the final result was better than the benchmark.

B. Unsupervised learning

Unsupervised detection of anomalies (often recognized as outlier detection) employs clustering approaches to classify potentially malicious incidents without previous knowledge in a dataset. Clustering aims to divide a limited unlabeled data into a discrete and finite collection of "natural" unseen structures of data instead of providing a precise non-observed characterization incidents produced within the same distribution probability [65]. In another aspect, the goal of

unsupervised algorithms is to divide the data into categories (clusters) that reach great similar internal and external dissimilarities without previous knowledge.

All clustering approaches are based on the following hypotheses for this reason. First, the volume of normal instances in a database surpasses the volume of anomalies. Next, the anomaly packet themselves vary from normal instances qualitatively [66]. Scores are allocated to the installed clusters after the cluster formation. If a cluster's score reaches the threshold pre-defined or automatically determined, a potential anomaly is considered. When clustering is utilized to identify attacks on the network, respectively, one believes that malicious traffic is less than the normal packet and normal packet is distinguished from the malicious one in some way. In other words, the features that characterize the attacks well enough to be defined must be selected concerning to the process of detection. The aim of clustering is to categorize network packets or flows without prior knowledge, but based solely on their relationships. As a result, large normal packet clusters would be formed when attack packets produce small clusters and cases not belonging to other groups. A static or dynamic threshold may be utilized to determine that clusters are deemed to be attack based on the testing and algorithm adjustment used. The main benefit of clustering models is their capability to identify unseen threats without previous information, thereby eliminating the need for labeled traffic. The main disadvantage is their high false-positive rate.

The extraction or selection of features is among the most critical stages of unsupervised detection. The use of clustering techniques to identify a range of attacks by checking alarm records from heterogeneous database was proposed [17], instead of utilizing the attributes of abnormalities that carry specific actions to suit instances or the standard approach of testing and training currently utilized in abnormal detection. Even though it required less time for the three clustering algorithms tested in the system to forecast and build clusters, the clusters' accuracy produced by one algorithm was not consistent across various logs and subsets. The obtained result indicates the way or route to develop abnormal detectors that could use pure activity logs obtained from heterogeneous databases on the tracked network and compare instances through alarm records to identify intrusion.

IV. Feature Selection Techniques

Feature Selection (FS) is a method for removing unnecessary and redundant features and choosing the most suitable feature subset that will result in a better classification of patterns which belong to various classes of attack. So, from researchers' view there are reasons why feature selection needs to be performed:

1) A single selection strategy is not adequate to obtain consistency across multiple datasets, as network traffic activity is changing [67]–[69].

2) An appropriate subset for each attack types should be identified, since one general subset of features is insufficient to properly represent all the various attacks[69]–[71].

3) FS can significantly improve not only the accuracy of detection but also the computational efficiency, where:

a) features which are irrelevant or redundant can result in poor detection rate and overfitting, therefore, reducing them can increase the detection accuracy; and

b) more features for each data point would cause higher computational costs and complexity—reducing irrelevant features will increase the computational efficiency [67], [69]–[74].

4) Ultimately, R2L (Remote-to-Local) and U2R (User-to-Root) attack groups are known to become the most challenging to identify since they are too isolated and could be mislabeled as normal packet. Studies and experiments have shown that FS can solve this issue by defining a feature subset adapted to the behavior of each attack type classes [70], [71], [74].

Methods of FS are generally classified into filter, wrapper and optimization-based FS methods for selecting features. Table II illustrates the advantages and disadvantages of the mentioned features selection methods and Table III summarized the reviewed feature selection for anomaly-based IDS. RQ3 and RQ4 are all answered in detail in the table. It specifies the feature selection methods, the algorithm's origin, subset size, strength, weakness, dataset used in the articles, and so on.

TABLE II. Comparison of different Feature Selection Methods

Method	Advantages	Disadvantages	Examples
filter	• Faster than wrapper • Not dependent on classifier • Less computational complexity than wrapper • Less over-fitting issues • Use statistical methods for evaluation of the attributes	• Lack of interaction to classifiers • Lack of dependency among attributes • Less detection rate compared to wrapper	Euclidean distance, information gain, correlation-based, etc.
Wrapper	• Interact with classifier • Consider attributes dependency • Better detection rate • Use cross validation for evaluation attributes	• Longer execution time • More risky for over-fitting issues	Sequential forward selection, Sequential backward selection, Hill climbing, Stepwise selection, etc.
Optimization-based	• Interact with classifier • Less over-fitting issues • Better detection of global optima • Better attribute selection • Simple to implement	• Difficult to be adjusted to a new situation • Complexity to adjusted different parameters	Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Cuckoo Search Algorithm (CSA), Genetic Algorithm (GA), etc.

TABLE III. SUMMARY OF THE REVIEWED FEATURE SELECTION FOR ANOMALY-BASED INTRUSION DETECTION SYSTEM

Author	Method	Algorithm	Type of feature subset	Subset size	Strengths	Weaknesses	Dataset
2011 [79]	Filter	Mutual information-based feature selection	Class-based Subset	15 out of 41	Improve relevancy and reduce redundancy for feature selection.	Performance metric is low.	KDD Cup 99
2013 [32]	Wrapper	Wrapper based on Bayesian Network classifier	Single Subset	11 out of 41	Improve the performance metric.	Accuracy of U2R attack was not satisfactory.	NSL-KDD
2015 [102]	Wrapper	Layered wrapper feature selection approach (LAWRA)	Single Subset	16 out of 41	LAWRA utilized external cluster validity indices, F-measure, and Fowlkes–Mallows index, for feature selection.	Overall accuracy (around 83%) and false parameter was not satisfactory. Lack of class-based feature selection.	NSL-KDD
2016 [34]	Optimization-based	Ant Colony Optimization (ACO) and K-Nearest Neighbor (KNN)	Class-based Subset	4 to 8 out of 41	Needs no prior knowledge of features.	False positive was not satisfactory (2.59).	KDD Cup 99 NSL-KDD
2016 [103]	Filter	Flexible Mutual Information Feature Selection (FMIFS) and Least Square SVM	Single Subset exempt Class-based Subset for KDD Cup 99	KDD Cup 99 (12 to 23 features) NSL-KDD (18 features) Kyoto2006+ (-)	Proved the generality of their model by utilizing different datasets.	Accuracy of U2R attack was not satisfactory. Lack of class-based feature selection two dataset.	KDD Cup 99 NSL-KDD Kyoto2006+
2017 [92]	Optimization-based	Hypergraph based Genetic Algorithm (HG-GA)	Single Subset	35 out of 41	HG-GA utilized a weighted objective function to improve performance metrics.	Overall accuracy 97.14%. Lack of class-based feature selection.	NSL-KDD
2019 [104]	Filter/ Optimization-based	Linear correlation coefficient algorithm and cuttlefish algorithm (CFA) and Decision tree	Single Subset	10 out of 41	Integration of filter method with cuttlefish algorithm optimization helps the model to detect the attack with less false alarm.	Overall accuracy 95.03%. Lack of class-based feature selection.	KDD Cup 99
2019 [38]	Filter/Wrapper	CART Algorithm	Single Subset	17 out of 41	Utilizing Gini and CART algorithm.	Overall accuracy 79.7%. Lack of class-based feature selection.	NSL-KDD
2020 [105]	Wrapper / Optimization-based	Genetic Algorithm with Multi-parent Crossover and Multi-Parent Mutation (MGA)	Single Subset	4 out of 41	Propose a new operator, called multi-parent-crossover-mutation to enhance the GA performance.	Lack of class-based feature selection.	NSL-KDD
2020 [94]	Optimization-based	Mutation Cuckoo Fuzzy (MCF) and ANN classifier	Single Subset	22 out of 41	Integrates mutation operator with cuckoo search and Fuzzy C Means (FCM).	Lack of class-based feature selection.	NSL-KDD
2020 [106]	Optimization-based	Pigeon Inspired Optimizer	Single Subset	KDDCUPP99 (7 features), NSL-KDD (5 features), and UNSW-NB15 (5 features)	Proved the generality of their model by utilizing different datasets.	Performance metric was not satisfactory. Lack of class-based feature selection.	KDD CUPP 99 NSL-KDD UNSW-NB15
2020 [40]	Filter/ Optimization-based	Ensemble of (mRMR, JMI CMIM) and Chaotic Adaptive Grasshopper Optimization Algorithm (CAGOA)	Single Subset	ISCX 2012 (20 features), NSL-KDD (19 features) and CIC-IDS2017 (12 features)	This feature selection combination gives good accuracy and less false alarm.	Lack of class-based feature selection.	ISCX 2012 NSL-KDD CIC-IDS2017
2020 [107]	Optimization-based	Multi-objective method (NSGAII) and ANN	Single Subset	NSL-KDD (24 features) UNSW-NB15 (19 features)	Proved the generality of their model by utilizing different datasets.	Performance metric was not satisfactory. Lack of class-based feature selection.	NSL-KDD UNSW-NB15

2020 [6]	Wrapper / Optimization-based	Hybrid of Fruit Fly Algorithm (FFA) and Ant Lion Optimizer (ALO) Algorithm.	Single Subset	KDD Cup99 (12 features), NSLKDD (16 features), UNSW-NB15 (15 features)	This hybrid algorithms increase the diversity of populations, which yields better detection.	Lack of class-based feature selection.	KDD Cup99 NSLKDD UNSW-NB15
2020 [108]	Optimization-based	Many Objective Evolutionary Algorithm and Artificial Bee Colony (MaOEA-ABC)	Single Subset	11 out of 41	Propose an adaptive selection probability approach that will adjust the selection probability and enhance the algorithm's ability to find the best solution.	Lack of class-based feature selection.	NSLKDD
2020 [43]	Wrapper / Optimization-based	Tabu Search and Random Forest (TS-RF)	Single Subset	16 features out of 41	Reducing feature vector by more than 60%. This reduces computational complexity of the proposed solution.	Lack of solving class imbalance problem present in UNSW-NB15. Performance metric was not satisfactory.	UNSW-NB15

According to the reviewed articles in Table III and the result from Fig. 2, it shows that optimization-based methods were mostly utilized for feature selection in the recent years. This method has undergone a significant improvement in terms of feature numbers. Based on the review, NSL-KDD dataset was mostly used by researchers to prove their models. In addition, some research utilized different datasets to highlight the generality of their solutions, like Kyoto2006+, ISCX 2012, UNSW-NB15, and CIC-IDS2017.

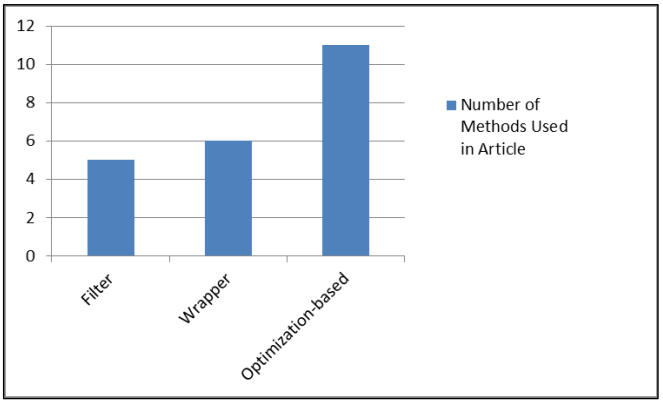

Fig. 2. Number of Studied Feature Selection Methods.

5) Filter: Filter methods use different information theory and mathematical formula for feature selection. Due to their simplicity, ranking methods are used and had good performance for practical applications. The rating of variables is based on an accepted ranking criterion and the threshold is being utilized to eliminate variables just below the value of threshold. The methods of ranking are filtering approaches as less relevant variables are extracted before the classification. A fundamental characteristic of a distinctive feature is the provision of useful information on the various classes of data. This characteristic could be described as a feature relevance[75] which defines a measure of the efficacy of the feature in order to distinguish among various classes. There are different ways to calculate a feature's relevance to the data point or outcome. Different publications [75]–[77] proposed different understandings and measurements for a variable's importance and relevancy. One description that can be listed that would be valuable is ‘‘A feature can be regarded as irrelevant if it is conditionally independent of the class labels.’’ [78]. This clearly stipulates that the data could be distinct but not separate from labels of the class, if the feature is to be relevant. The feature that does not impact class labels can be omitted. As noted above, for assessing specific features, correlation of features plays a key role. The underlying distribution of practical uses is unclear: it is calculated by the accuracy of the classifier. Because of this, an ideal subset of features may not be special because using different feature sets it may be possible to reach the similar accuracy of classifier. An improved feature selection algorithm has been proposed [79] to efficiently classify the attacks behaviors by measuring mutual information. Correlation can also be extended to evaluate of the efficiency of a feature subset, where a subset of features is perfect if the correlation among the classification and the feature subset is significant, but the correlation among the specific feature and the other features within the subset of features is poor. In addition, distance calculation for the selection of features can also be utilized [80]. Widely utilized distance calculation includes Euclidean distance, Martensitic distance and standardized Euclidean distance.

6) Wrapper: Wrapper feature selection use machine learning as a fitness function and determine the best feature subset across all subsets of features. This problem formulation allows generic optimization techniques to be used with the machine learning to rank subsets of feature based on their prediction. Therefore, in the aspect of a machine learning final predictive accuracy, the wrapper method typically surpasses the filter approach. The wrapper technique was widely popularized by [75], and provides an easy but efficient way to tackle the issue of selection of features. However, the wrapper method incurs more computation cost and need more execution time compared to filtering methods. A feature selection method using machine learning algorithms was proposed [81] for efficient intrusion detection, which blends the characteristics of distributed denial of service (DDoS) characteristic-based features (DCF) and consistency set evaluation (CSE). To identify the most relevant features, the NSL-KDD dataset is utilized as an attack dataset and is built

on a few selections of feature methods, along with consistency-based evaluation of subsets and DDoS characteristic-based features (DCF). The experimental result shows that their proposed system has greater accuracy and efficiency compared to other approaches.

7) Optimization-based methods: Classic wrapper and filter strategies are independently evaluated and subset chosen. However, some features are not independent, but they are really successful when they work together. Therefore, the classic strategies in this respect are not very successful. Metaheuristic-based methods were already used to select and classify the selected features as a result of its vast improvement capability of in detection [82], [83]. Examples of optimization-based methods are Particle Swarm Optimization (PSO) [84]–[86] entropy of network features [87], Genetic Algorithm [88], [89], ant colony optimization [34], [90] and Kernel Principal Component Analysis (KPCA) [91]. With the increase in the dataset dimension, the space of the problem of selection of feature rises significantly. This leads to a large solution space with additional features. Furthermore, in a wide solution space, a huge proportion of duplicate or uncorrelated features generate several local optima.

A new anomaly based detection model of Hypergraph based Genetic Algorithm (HG - GA) was proposed by [92]. The Hypergraph's attribute was used to generate initial population in order to speed up the quest for the optimum solution and avoid trapping at local minima. HG-GA utilized a weighted objective function to achieve the balance among maximum detection rate and reducing false positive, as well as reducing features number. HG-GA SVM performance was assessed by NSL-KDD dataset.

An Ant Colony Optimization (ACO) for selection of feature method was proposed [34] using K-Nearest Neighbor (KNN) for the classification process and the accuracy was utilized as the assessment function for the model. The studies were performed using the KDD Cup 99 dataset, giving 98.9 % for accuracy and 2.59% for false positive rate.

A learning model for fast learning network (FLN) based on PSO was proposed by [93]. The PSO-based optimized FLN was trained using particle swarm optimization to pick weights. For evaluation, the research utilized KDD Cup 99 dataset to explore the effects of PSO-FLN model. The findings indicated that the model had good impact on intrusion detection.

An enhancement of Cuckoo Search Algorithm (CSA), named Mutation Cuckoo Fuzzy (MCF) was proposed by [94] for feature selection method and multiverse optimization ANN for classification at anomaly-based IDS. For feature selection phase, MCF that integrates mutation operator with cuckoo search and Fuzzy C Means (FCM) clustering was utilized. Through this method, the cuckoo Search efficiency to detect the optimal features was increased. The proposed feature selection choses 22 out of 41 features and for evaluation part well known dataset, called NSL-KDD was used to illustrate the effectiveness of their anomaly-based IDS.

A. Limitations of the Related Works

After analyzing the data collected from the literature related to feature selection, some limitations and shortcomings of the works are identified:

1) The optimal detection methods or strategies for various datasets have yet to be established.

2) There is a lack of proper feature subset to train faster with minimal computation and optimal performance in detecting intrusion with high accuracy and less false alarms.

V. Ensemble

The idea of merging results from a collection of learners into one is known as ensemble [95]. To obtain reliable and more accurate predictions, an ensemble can integrate multiple learners. It is possible to use a variety of techniques to generate and incorporate learners. Various datasets could be utilized to train the same training frameworks or the similar dataset could be utilized to learn various frameworks [96]. The biggest issue on the learning of the ensemble is to choose the algorithms that construct the ensemble and the function of decision or fusion that incorporates these algorithms' results. Of course, it is easy to use more algorithms to enhance the fusion results, but bearing in mind the computing cost of adding a new algorithm, it needs careful consideration. Dietterich [95] offered three key explanations for the use of an ensemble-based system. First, the empirical justification is related to the absence of sufficient knowledge to accurately classify the quest space's best hypothesis. Second, the computational description is to resolve the issue that most machine learning methods might be trapped in the local optima when looking for the perfect solution. Finally, the rationale for representation is to resolve the problem of the failure of several machine learning methods to accurately depict the border of the searched decision. Creating an ensemble takes two main parts: creating and combining [97].

The creation process has to construct a collection of base classifiers. The decision on how to integrate the results of the base classifiers into one is taken in the combining process. Many of the well-known modern ML algorithms were constructed around the idea of the ensemble. The three widely used ensemble model are bagging, boosting, and stacking [98]. Such techniques combine various models of learning into a single model so that bagging (variance), boosting (bias) or stacking (predictions) can be minimized. Fig. 3 demonstrates the general design methods of the ensemble.

A. Bagging

Among the first ensemble algorithms, one of the simplest and easiest way to accomplish a better efficacy was bagging [99]. When bootstrapped copies were used, varieties of results are generated in bagging, which is to say, various data subsets are randomly selected from the complete dataset of training. A different same type of classifier is designed by utilizing the learning data portion. Using a majority vote on their lists, the fusion of different classifiers is accomplished. Therefore, the decision of the ensemble is the category chosen by the largest number of classifiers for any instance data.

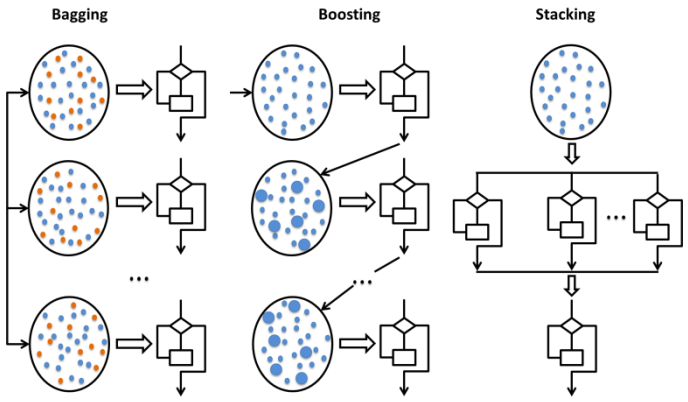

Fig. 3. Three General Ensemble Designs.

Random Forests is a method which is produced from bagging [100]. Training of multiple decision trees and randomly changing parameters relevant to training is a way to create this sort of classifier. As in bagging, copies of the training data could be bootstrapped from those parameters; but, unlike bagging, they can also be unique subsets of features, which is the case in the random subspace process.

From bagging, another method was generated, named "pasting small votes." It was a technique designed to run on huge datasets, unlike bagging [101]. Large size datasets are divided to the small size portions called "bites," used for learning various classifiers.

Small votes have resulted in the design of two combinations: first, named as Rvotes, randomly produces the subsets of data; second, named as Ivotes, creates consecutive datasets, taking into account the importance of the instances. Ivotes has been shown to deliver better results similar to the approach in boosting methods where the classifier advises the most suitable instances for the ensemble component used [109].

New method of ensemble classification [110] are proposed using bagging classifiers and their performance is evaluated with accuracy in mind. A classifier ensemble is built as a base classifier using the Support Vector Machine (SVM) and Radial Basis Function (RBF). The effectiveness and advantages of the approaches proposed are demonstrated through NSL-KDD datasets. The accuracy for bagged RBF was 86.40% and bagged SVM was 93.92%.

B. Boosting

In 1990, Schapire [111] demonstrated a weak learner (algorithm) which produces classifiers that can moderately surpass random guessing, can be converted into a powerful learner that can properly classify all instances except an extremely small fraction. The boosting created a group of classifiers by resampling the data and integrating results by majority voting. Re-sampling in boosting is designed to provide the most detailed training data for successive classifiers. In general three classifiers are created by boosting: a randomized subset of available training data is utilized to construct the first one. For training of the second classifier, knowledgeable subset of data provided to the first classifier is utilized where the knowledgeable data portion includes instances of training dataset, so the first classifier correctly identified half of them and the other half was misidentified. Ultimately, learning information for the third classifier is made up of cases where there was a conflict between the first and second classifiers. The results of the three classifiers would be combined with a majority vote.

A simplified edition of the initial boosting algorithm called "adaptive boosting" or "AdaBoost" was proposed in 1997 by Freund and Schapire [112]. Two algorithms of this group, AdaBoostM1 and AdaBoostR are the most commonly used variants, as they are perfect to cope for problems of regression and multiclass. AdaBoost generates some assumptions and the same assumptions apply to aggregate decisions by weighted majority voting of the groups decided. By extracting instances from a successively updated distribution of training data, a weak classifier is trained to build the assumptions. Updating the distribution ensures that the following classifier examples that were incorrectly identified by the prior classifier are return back to dataset to train other classifiers. Therefore, training data from various classifiers continue to move into instances that are becoming increasingly difficult to classify.

C. Stacking

Many cases are very likely to be miscategorized because they may happen to be in the near neighboring of the decision line and thus are typically located on the incorrect side of the line identified by the machine learning classifier. On the other hand, since it is on the right side and far from the boundaries of the appropriate decision, there may be instances that are likely to be well defined. If a group of classifiers performs with a dataset from an undefined source, could we create a relationship among the classifiers' results and correctly detect groups? The concept motivating generalization of Wolpert's is that the results of a classifying ensemble serve as sources to the next meta-classifier at second level with the goal of learning the manner in which the ensemble's findings are related to the correct label instances [113].

Stacking is the term used for Stacked Generalization [113], which is to find the ideal composition of a base learner set. Stacking is an algorithm class that requires training a "meta-learner" second level to find the combination. Stacking aims to combine solid, different sets of learners, unlike bagging and boosting. Besides, ensemble methods such as boosting and bagging are often utilized to construct alike ensembles, while stacking could be utilized to create diverse ensembles.

D. Other Work

New ensemble methods [114] proposed are Net-GR based ANN-Bayesian approach that implies ensemble of Bayesian Net with Gain Ratio (GR) feature selection approach and ANN. They have applied a variety of single classification methods and their proposed ensemble on NSL-KDD and KDD Cup 99 datasets to evaluate for model's robustness. With 29 features which were selected, a 97.78% and 99.38% accuracy detection were achieved when the model was applied to the NSL KDD and KDD Cup 99 datasets to detect intrusions.

A hybrid approach that combines the synthetic minority oversampling technique (SMOTE) and cluster center and nearest neighbor (CANN) was proposed [115]. Significant

features were selected by utilizing the leave one out (LOO) approach. In addition, the research utilized the NSL-KDD dataset and the results illustrate that the proposed approach increases the accuracy of the R2L and U2R attacks as opposed to the benchmark paper by 50% and 94%, respectively.

A Hybrid RBF-SVM ensemble classification was proposed by [110] utilizing Support Vector Machine (SVM) and Radial Basis Function (RBF) as base classification. The efficacy and advantages of the proposed model are presented using NSL-KDD datasets, and their finding illustrates that the proposed ensemble RBF-SVM is superior to single-method approaches in terms of accuracy as it achieved 98.46%.

An ensemble-based IDS model was designed using integrated feature selection approach and an ensemble of ML classifiers comprising Bayesian Network, J48, and Naive Bayes [15]. In this model, features are reduced from 41 to 12, and majority vote is used for combing the findings. The true positive rate (TP) of the proposed model is 98.0% with a false-positive rate (FP) of 0.021%.

A hybrid classification approach was proposed to detect and forecast DDoS threats. Using the KDD Cup 99 dataset as attack data, related features were chosen based on information gain. The experimental result revealed that each step of the threat case is well divided, and they can identify DDoS threat precursors as well as the threat itself [116].

A model for Adaptive Ensemble Learning was proposed by [38] by changing the learning data ratio and constructing a MultiTree algorithm which deploys multiple decision trees. To increase detection efficiency, a number of base classifiers are chosen, including Random Forest, decision tree, deep neural network (DNN), KNN, and an adaptive voting algorithm were developed. For the validation part, the NSL-KDD dataset was used, and the MultiTree algorithm accuracy was 84.2%, while the final adaptive voting ensemble accuracy was 85.2%.

A model called SCDNN combines spectral clustering (SC) and DNN algorithms was proposed by [117]. In this model, k subsets were created from the dataset based on the similarity of the sample utilizing cluster centers as in SC. Then, the distance between data points in the training set and the test set was calculated on the basis of features similarity and was applied into the DNN algorithm to detect intrusion. NSL-KDD dataset was used for evaluation benchmark and the overall accuracy was 92.1%.

A framework with feature selection and ensemble method [118] integrates correlation-based feature selection with Bat algorithm (CFS-BA), and an ensemble of Random Forest (RF), C4.5 and Forest by Penalizing Attributes (Forest PA) is developed for the detection model. The evaluation experiments used the CIC-IDS2017, AWID, and NSL-KDD datasets. The results show that this framework has better accuracy than other research work.

A hierarchical ensemble classifier and knowledge-based method was proposed by [119]. In order to determine the specific attack class, it used a weighted voting fusion technique for specific classes to obtain a more accurate classification. The KDD Cup 99 dataset was used to prove the model. This IDS model has more complexity during the learning phase and it consumes more time in contrast to other work.

A Hybrid IDS of One Class Support Vector Machine (OC-SVM) and C5 decision tree classifier [42] was proposed to detect unknown and known intrusion. To the model was evaluated using the ADFA and NSL-KDD datasets. Their finding demonstrated that the hybrid schema has better performance than other models.

An IDS ensemble model of convolutional neural network, Random Forest, and gated recurrent unit (GRU) was proposed by [120]. NSL-KDD dataset was utilized to prove the performance of the model. The detection accuracy was 76.61% with reduced learning time and resource usage than other schema.

An IDS model with combination of ensemble (Random Forest, J48, and Reptree) and CFS algorithm suggested by [121]. Experimented on the KDD Cup 99 and NSLKDD datasets, their finding illustrates that the proposed ensemble has 99.90% for the KDD99 dataset, and 98.60% detection rate for NSLKDD. However, this model could not handle imbalance data.

A stacked ensemble classifier with a combination of gradient boosting machine, XGBoost, and Random Forest [39] was proposed and experimented on CICIDS-2017, CSIC-2010v2, UNSW-NB15, and NSL-KDD. The result shows that the proposed ensemble model has good impact on detection of attack in a Web application.

Table IV introduces a comparative analysis of different ensemble algorithms used in the literature to handle anomaly-based IDS. The table presents a comprehensive review of several ensemble classifications, showing their methods, strength, weakness and the dataset utilized for evaluation. RQ3 and RQ6 are addressed in table.

According to the reviewed articles presented in the table, different combination of classifiers and algorithms were utilized for ensemble detection. An ensemble with diversity of classifier types had significant improvements in detection accuracy and reduces the false alarm for anomaly-based IDS.

Based on the review, NSL-KDD dataset was mostly used to show the efficacy and advantages of the proposed ensemble models. Furthermore, some articles utilized different datasets to highlight their generality of their solutions, like AWID, ISCX 2012, UNSW-NB15, CIC-IDS2017 and CSIC-2010v2.

Based on the analysis of the studied articles in the review, Fig. 4 illustrates that NSL-KDD dataset was mostly utilized to highlight the effectiveness of their anomaly-based IDS models. The KDD Cup 99 dataset came in second as to be used to evaluate their solutions.

TABLE IV. SUMMARY OF THE REVIEWED INTRUSION DETECTION SYSTEM

Author	Method	Strength	Weakness	Dataset
2013 [27]	Unsupervised Artificial Neural Network	Hierarchical Anomaly-based Intrusion Detection System	Overall detection accuracy of 75.49%. No class-based detection.	KDD Cup 99 NSL-KDD
2014 [110]	Support Vector Machine (SVM) And Radial Basis Function (RBF)	They develop a bagging classifier	Overall accuracy was not reasonable. No class-based detection.	NSL-KDD
2015 [47]	Hybrid Artificial Bee Colony Algorithm and Multi-Layer Perceptron	The proposed model has reasonable detection time and Error rate (0.124%)	Overall accuracy was 87.54 % No class-based detection.	NSL-KDD
2015 [15]	Ensemble of Bayesian Network, J48, and Naive Bayes	The overall accuracy and error rate was reasonable	No class-based detection.	NSL-KDD
2016 [115]	Hybrid cluster center and nearest neighbor (CANN) and synthetic minority oversampling technique (SMOTE)	Reasonable detecting R2L and U2R attacks (50% and 94%,)	Not detect the rest of attacks. Not mentioned about time.	NSL-KDD
2016 [26]	Hybrid of Fuzzy C-Means clustering algorithm and Artificial Neural Network	The model was good at finding normal and probe attacks	The results for other attack types did not yield good results even for accuracy and false alarm rate (DOS (99.96–5.33), U2R (96.78–3.22) and R2L (93.73–6.27)	KDD Cup 99
2016 [46]	Hybrid of C4.5 decision tree and the MLP classifier	Use feature selection by utilizing the Discernibility Function and MLP to provide feature subset	Detection rate R2L and U2R attacks were not satisfactory.	KDD Cup 99 ISCX dataset
2016 [28]	ANN-based IDS with back propagation and feed forward algorithms	The overall accuracy was reasonable (98.86%)	No class-based detection.	NSL-KDD
2019 [38]	Ensemble of Decision Tree, Random Forest, KNN, DNN and MultiTree	Detection based on attack class	Class based accuracy was not satisfactory. Performance metric was low.	NSL-KDD
2020 [118]	Feature Selection CFS-BA and an Ensemble of Random Forest (RF), C4.5 and Forest by Penalizing Attributes (Forest PA)	They train and test their framework with different dataset	Detection rate R2L and U2R attacks were not satisfactory.	NSL-KDD AWID CIC-IDS2017
2020 [119]	Hierarchical ensemble classifier and knowledge base method	The class-based accuracy and error rate was reasonable	The IDS model has more complexity during learning phase and it consume more time in contrast to other work.	KDD Cup 99
2020 [42]	Hybrid of One Class Support Vector Machine (OC-SVM) and C5 decision tree classifier	The overall accuracy and error rate was reasonable	No class-based detection.	ADFA NSL-KDD
2020 [120]	Ensemble of Convolutional Neural Network, Random Forest, and Gated Recurrent Unit (GRU)	The model has improvement on reduction of learning time and resource usage	Overall accuracy was not reasonable. No class-based detection.	NSL-KDD
2020 [121]	Combination of Ensemble (Random Forest, J48, and Reptree) and CFS algorithm	The overall accuracy and error rate was reasonable	The model could not handle imbalance data issue.	KDD Cup 99 NSL-KDD
2020 [39]	Ensemble of Gradient Boosting Machine, XGBoost, and Random Forest	The overall accuracy and error rate was reasonable	No class-based detection.	CICIDS-2017 CSIC-2010v2 UNSW-NB15 NSL-KDD

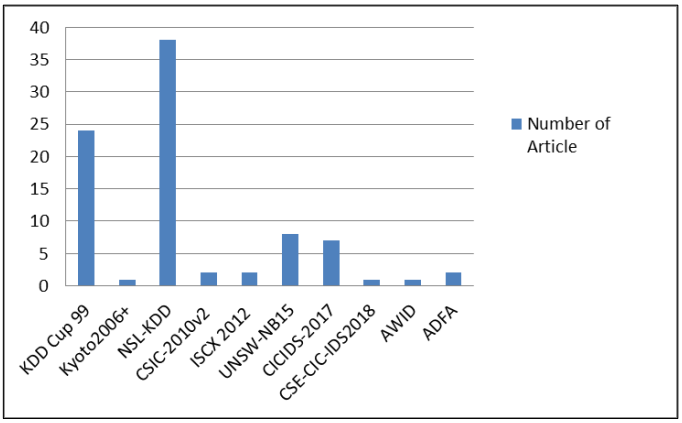

Fig. 4. Number of Datasets in the Reviewed Articles.

1) Limitations of the ensemble classification: After analyzing the data collected from the literature related to ensemble, some limitations and shortcomings of the works are identified and in order to reach maximum diversity with various boundaries of decision, the identified limitation should be considered:

a) Multiple datasets have to be utilized to prove the generality of the ensemble model.

b) In order to handle imbalance data issues in anomaly-based IDS, different types of classifiers have to be deploy in ensemble machine. Therefore, selection of various classifiers and the fusion of their outcomes empower the final result.

VI. Discussion

Upon studying and reviewing the different IDS models, we found challenges that motivate research in utilizing machine learning for feature selection and ensemble techniques in IDS. In this paper, we discuss future trends in anomaly-based IDS, in particular feature selection and ensemble techniques. Some of the critical topics in the existing research with view of future trends are described below:

1) Anomaly-based IDS datasets have a crucial impact on the proposed approaches in terms of performance assessment. To be current, it is necessary to utilized updated datasets to illustrate that the proposed solution works well with new attack types. Although KDD Cup 99 is an old dataset used by most of researchers as benchmark comparisons, the attack packets and even the features are dated 20 years ago. In addition, researchers can deploy their model on different anomaly-based IDS datasets to prove the generality of their model to detect different attacks.

2) Finding the appropriate feature selection schema plays an important role in anomaly-based IDS. Proper selection of feature subset helps expert machine in the learning phase to detect attacks in the testing phase. Optimization-based feature selection aims to acquire an optimal subset of features among all features in different domains. The role of new optimization-based feature selection methods in the success of anomaly-based IDS must be considered.

3) Ensemble-based modern anomaly-based IDS techniques allow multiple combinations of models or algorithms to identify new unseen cases. In the implementation, after a variety of classification models are typically constructed utilizing some portion of datasets, the various classifiers results are merged to form the final conclusion. Various schemes may be suggested for the generation of classifiers and for the combination of the ensembles.

The future trends mentioned above and open issues discussed in anomaly-based intrusion detection system should be considered by researchers in the field of anomaly-based IDS.

VII. Conclusions

Intrusion detection system is a prominent security mechanism designed to prevent intrusion, illicit entry, modification or demolition by intruders. For efficient intrusion detection process vital components like feature selection and detection mechanism have to be considered when designing the model. The article reviews the studies on feature selection and ensemble approaches utilized for anomaly-based intrusion detection systems. We discussed the main challenges in IDS, namely the dimensionality reduction in anomaly-based IDS that reduces irrelevant attributes from dataset; and how to build an appropriate feature subset selection, in order to better detect intrusion by increasing the performance metrics. Consequently, the study categorizes and discusses feature selection methods and presents their performance in detection accuracy. Another important challenge in anomaly-based IDS lies in utilizing suitable machine learning algorithms in the detection process. To illustrate their effectiveness in improving the IDS performance, this paper reviews and categorizes various machine learning schema and discussed their utilization in IDS, giving emphasize on ensemble methods as an emerging trend in anomaly-based IDS. Based on our study on anomaly-based IDS and the assessment and comparison of feature selection and detection module, we can summarize two points about how to boost the performance of anomaly-based IDS as follows:

1) Optimization-based feature selection with excellent combination and well tune up parameters will select the proper feature subset for IDSs. Through this study, it is clear that optimization-based have significant performance to design the optimal feature set. Furthermore, if their parameters are adjusted well, feature selection could be significantly enhanced.

2) Ensemble detection with different types of classification can empower the detection phase and reduce the false alarm rate. If the diversity occurred, a fusion of the outcome has better chance to detect properly.

Finally, we present some open issues and offered research trends, including the datasets used, the role of optimization-based algorithm-ms and ensemble methods, in the area of anomaly-based IDS. We expect that this review paper will furnish scientists with innovative ideas and serve as a springboard for them to undertake better studies. We acknowledge that this article has some limitations due to the scope of the review:

1) This review focused on the feature selection and ensemble detection for anomaly-based IDS.

2) This review does not focus on performance parameter which is utilized at IDS.

3) This article does not study IDS datasets in-depth, like their features, attack types, etc.

Having listed the limitations of the paper, a deep analysis on the following issues can be considered as future work:

1) Other detection methods for anomaly-based IDS, apart from the feature selection and ensemble detection methods that are discussed here, could be studied too, in order to acquire a more holistic understanding of the research area.

2) Extra studies could be performed on performance parameters which are utilized in IDS, and how we can obtain the optimal set of parameters for better detection performance.

3) An in-depth study on IDS datasets could be carried out, such as their features, attack types, etc. to understand the pattern in their attributes that may affect the detection performance.

References

[1] Patel, Q. Qassim, and C. Wills, “A survey of intrusion detection and prevention systems,” Inf. Manag. Comput. Secur., vol. 18, no. 4, pp. 277–290, 2010.

[2] J. R. Vacca, Computer and Information Security Handbook, vol. 82, no. 90001. 2013.

[3] H. I. Alsaadi, R. M. Almuttairi, O. Bayat, and O. N. Ucani, "Computational intelligence algorithms to handle dimensionality reduction for enhancing intrusion detection system," J. Inf. Sci. Eng., vol. 36, no. 2, pp. 293–308, 2020.

[4] G. T. Reddy et al., "Analysis of Dimensionality Reduction Techniques on Big Data," IEEE Access, vol. 8, pp. 54776–54788, 2020.

[5] O. Almomani, "A feature selection model for network intrusion detection system based on pso, gwo, ffa and ga algorithms," Symmetry (Basel)., vol. 12, no. 6, pp. 1–20, 2020.

[6] M. Samadi Bonab, A. Ghaffari, F. Soleimanian Gharehchopogh, and P. Alemi, "A wrapper-based feature selection for improving performance of intrusion detection systems," Int. J. Commun. Syst., vol. 33, no. 12, pp. 1–25, 2020.

[7] A. Aldweesh, A. Derhab, and A. Z. Emam, "Deep learning approaches for anomaly-based intrusion detection systems: A survey, taxonomy, and open issues," Knowledge-Based Syst., vol. 189, p. 105124, 2020.

[8] W. Fang, X. Tan, and D. Wilbur, "Application of intrusion detection technology in network safety based on machine learning," Saf. Sci., vol. 124, no. December 2019, p. 104604, 2020.

[9] N. Moustafa, J. Hu, and J. Slay, "A holistic review of Network Anomaly Detection Systems: A comprehensive survey," J. Netw. Comput. Appl., vol. 128, pp. 33–55, 2019.

[10] S. Jose, D. Malathi, B. Reddy, and D. Jayaseeli, "A Survey on Anomaly Based Host Intrusion Detection System," J. Phys. Conf. Ser., vol. 1000, no. 1, 2018.

[11] H. J. Liao, C. H. Richard Lin, Y. C. Lin, and K. Y. Tung, "Intrusion detection system: A comprehensive review," J. Netw. Comput. Appl., vol. 36, no. 1, pp. 16–24, 2013.

[12] C. Liu, J. Yang, R. Chen, Y. Zhang, and J. Zeng, "Research on immunity-based intrusion detection technology for the Internet of Things," in Proceedings - 2011 7th International Conference on Natural Computation, ICNC 2011, 2011, vol. 1, pp. 212–216.

[13] W. Li, S. Tug, W. Meng, and Y. Wang, "Designing collaborative blockchained signature-based intrusion detection in IoT environments," Futur. Gener. Comput. Syst., vol. 96, pp. 481–489, 2019.

[14] F. Erlacher and F. Dressler, "FIXIDS: A high-speed signature-based flow intrusion detection system," in NOMS 2018 - 2018 IEEE/IFIP Network Operations and Management Symposium, 2018, pp. 1–8.

[15] N. F. Haq, A. R. Onik, and F. M. Shah, "An ensemble framework of anomaly detection using hybridized feature selection approach (HFSA)," IntelliSys 2015 - Proc. 2015 SAI Intell. Syst. Conf., pp. 989–995, 2015.

[16] M. Monshizadeh, V. Khatri, B. G. Atli, R. Kantola, and Z. Yan, "Performance Evaluation of a Combined Anomaly Detection Platform," IEEE Access, vol. 7, pp. 100964–100978, 2019.

[17] A. I. Hajamydeen and N. I. Udzir, "A detailed description on unsupervised heterogeneous anomaly based intrusion detection framework," Scalable Comput., vol. 20, no. 1, pp. 113–160, 2019.

[18] J. M. Estevez-Tapiador, P. Garcia-Teodoro, and J. E. Diaz-Verdejo, "Stochastic protocol modeling for anomaly based network intrusion detection," in Proceedings - 1st IEEE International Workshop on Information Assurance, IWIA 2003, 2008, vol. 02798, pp. 3–12.

[19] R. Sekar et al., "Specification-based anomaly detection: A new approach for detecting network intrusions," Proc. ACM Conf. Comput. Commun. Secur., pp. 265–274, 2002.

[20] PMG, "Maximizing the value of network intrusion detection," in A corporate white paper from the product management group of intrusion.com, 2001.

[21] W. Yassin, N. I. Udzir, A. Abdullah, M. T. Abdullah, H. Zulzalil, and Z. Muda, "Signature-Based Anomaly intrusion detection using Integrated data mining classifiers," Proc. - 2014 Int. Symp. Biometrics Secur. Technol. ISBAST 2014, pp. 232–237, 2015.

[22] V. Hajisalem and S. Babaie, "A hybrid intrusion detection system based on ABC-AFS algorithm for misuse and anomaly detection," Comput. Networks, vol. 136, pp. 37–50, 2018.

[23] M. A. Hatef, V. Shaker, M. R. Jabbarpour, J. Jung, and H. Zarrabi, "HIDCC: A hybrid intrusion detection approach in cloud computing," Concurr. Comput. , vol. 30, no. 3, 2018.

[24] I. Lorenzo-Fonseca, F. Maciá-Pérez, F. J. Mora-Gimeno, R. Lau-Fernández, J. A. Gil-Martínez-Abarca, and D. Marcos-Jorquera, "Intrusion detection method using neural networks based on the reduction of characteristics," in International Work-Conference on Artificial Neural Networks, 2009, pp. 1296–1303.

[25] S. Peddabachigari, A. Abraham, C. Grosan, and J. Thomas, "Modeling intrusion detection system using hybrid intelligent systems," J. Netw. Comput. Appl., vol. 30, no. 1, pp. 114–132, 2007.

[26] N. Pandeeswari and G. Kumar, "Anomaly Detection System in Cloud Environment Using Fuzzy Clustering Based ANN," Mob. Networks Appl., vol. 21, no. 3, pp. 494–505, 2016.

[27] L. M. Ibrahim, D. T. Basheer, and M. S. Mahmod, "A comparison study for intrusion database (KDD99, NSL-KDD) based on self organization map (SOM) artificial neural network," J. Eng. Sci. Technol., vol. 8, no. 1, pp. 107–119, 2013.

[28] B. Subba, S. Biswas, and S. Karmakar, "A Neural Network Based System for Intrusion Detection and Attack Classification," 2016 22nd Natl. Conf. Commun. NCC 2016, pp. 1–6, 2016.

[29] M. Barati, A. Abdullah, N. I. Udzir, M. Behzadi, R. Mahmod, and N. Mustapha, "Intrusion detection system in secure shell traffic in cloud environment," J. Comput. Sci., vol. 10, no. 10, pp. 2029–2036, 2014.

[30] P. Jokar and V. C. M. Leung, "Intrusion Detection and Prevention for ZigBee-Based Home Area Networks in Smart Grids," IEEE Trans. Smart Grid, vol. 9, no. 3, pp. 1800–1811, 2011.

[31] H. C. Lin, M. K. Sun, H. W. Huang, C. Y. H. Tseng, and H. T. Lin, "A specification-based intrusion detection model for wireless ad hoc networks," Proc. - 3rd Int. Conf. Innov. Bio-Inspired Comput. Appl. IBICA 2012, pp. 252–257, 2012.

[32] F. Zhang and D. Wang, "An effective feature selection approach for network intrusion detection," Proc. - 2013 IEEE 8th Int. Conf. Networking, Archit. Storage, NAS 2013, pp. 307–311, 2013.

[33] W. Meng, W. Li, and L. F. Kwok, "EFM: Enhancing the performance of signature-based network intrusion detection systems using enhanced filter mechanism," Comput. Secur., vol. 43, pp. 189–204, 2014.

[34] M. H. Aghdam and P. Kabiri, "Feature selection for intrusion detection system using ant colony optimization," Int. J. Netw. Secur., vol. 18, no. 3, pp. 420–432, 2016.

[35] H. Bostani and M. Sheikhan, "Hybrid of anomaly-based and specification-based IDS for Internet of Things using unsupervised OPF based on MapReduce approach," Comput. Commun., vol. 98, pp. 52–71, 2017.

[36] A. Althubaity, H. Ji, T. Gong, M. Nixon, R. Ammar, and S. Han, "ARM: A hybrid specification-based intrusion detection system for rank attacks in 6TiSCH networks," IEEE Int. Conf. Emerg. Technol. Fact. Autom. ETFA, pp. 1–8, 2017.

[37] Y. Wang, W. Meng, W. Li, J. Li, W. X. Liu, and Y. Xiang, "A fog-based privacy-preserving approach for distributed signature-based intrusion detection," J. Parallel Distrib. Comput., vol. 122, pp. 26–35, 2018.

[38] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, "An Adaptive Ensemble Machine Learning Model for Intrusion Detection," IEEE Access, vol. 7, pp. 82512–82521, 2019.

[39] B. A. Tama, L. Nkenyereye, S. M. R. Islam, and K. S. Kwak, "An enhanced anomaly detection in web traffic using a stack of classifier ensemble," IEEE Access, vol. 8, pp. 24120–24134, 2020.

[40] S. Dwivedi, M. Vardhan, and S. Tripathi, "An effect of chaos grasshopper optimization algorithm for protection of network infrastructure," Comput. Networks, vol. 176, no. March, 2020.

[41] A. R. Gupta and J. Agrawal, "The multi-demeanor fusion based robust intrusion detection system for anomaly and misuse detection in computer networks," J. Ambient Intell. Humaniz. Comput., vol. 12, no. 1, pp. 303–319, 2020.

[42] A. Khraisat, I. Gondal, P. Vamplew, J. Kamruzzaman, and A. Alazab, "Hybrid intrusion detection system based on the stacking ensemble of C5 decision tree classifier and one class support vector machine," Electron., vol. 9, no. 1, 2020.

[43] A. Nazir and R. A. Khan, "A novel combinatorial optimization based feature selection method for network intrusion detection," Comput. Secur., vol. 102, p. 102164, 2020.

[44] "MLP structure," https://scikit-learn.org/stable/modules/neural_networks_supervised.html.

[45] Margaret H. Dunham, "Data mining – introductory and advanced topics," Pearson Educ., pp. 106–114, 2003.

[46] A. Akyol, M. Hacibeyoglu, and B. Karlik, "Design of multilevel hybrid classifier with variant feature sets for intrusion detection system," IEICE Trans. Inf. Syst., vol. E99D, no. 7, pp. 1810–1821, 2016.

[47] M. S. Mahmod, Z. A. H. Alnaish, and I. A. A. Al-hadi, "Hybrid Intrusion Detection System Using Artificial Bee Colony Algorithm and Multi-Layer Perceptron," vol. 13, no. 2, pp. 1–7, 2015.

[48] Liao and V. R. Vemuri, "Use of k-nearest neighbor classifier for intrusion detection," Comput. Secur., vol. 21, no. 5, pp. 439–448, 2002.

[49] R. S. Naoum and Z. N. Al-Sultani, "Learning Vector Quantization (LVQ) and k-Nearest Neighbor for Intrusion Classification," World Comput. Sci. Inf. Technol. J., vol. 2, no. 3, pp. 105–109, 2012.

[50] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," Data Min. Knowl. Discov., vol. 2, no. 2, pp. 121–167, 1998.

[51] H. Eid, "Computational Intelligence in Intrusion Detection System," 2013.

[52] A. Chalak, "Data Mining Techniques for Intrusion Detection and Prevention System," 2011, vol. 11, no. 8, pp. 200–203.

[53] F. Kuang, W. Xu, and S. Zhang, "A novel hybrid KPCA and SVM with GA model for intrusion detection," Appl. Soft Comput. J., vol. 18, pp. 178–184, 2014.

[54] S. J. Horng et al., "A novel intrusion detection system based on hierarchical clustering and support vector machines," Expert Syst. Appl., vol. 38, no. 1, pp. 306–313, 2011.

[55] W. Feng, Q. Zhang, G. Hu, and J. X. Huang, "Mining network data for intrusion detection through combining SVMs with ant colony networks," Futur. Gener. Comput. Syst., vol. 37, pp. 127–140, 2014.

[56] C. F. Tsai, Y. F. Hsu, C. Y. Lin, and W. Y. Lin, "Intrusion detection by machine learning: A review," Expert Syst. Appl., vol. 36, no. 10, pp. 11994–12000, 2009.

[57] S. Agrawal and J. Agrawal, "Survey on anomaly detection using data mining techniques," Procedia Comput. Sci., vol. 60, no. 1, pp. 708–713, 2015.

[58] N. Ben Amor, S. Benferhat, and Z. Elouedi, "Naive bayesian networks in intrusion detection systems," in 14th European Conference On Machine Learning 17th European Conference On Principles And Practice Of Knowledge Discovery In Databases, 2003.

[59] D. M. Singh, N. Harbi, and M. Zahidur Rahman, "Combining Naive Bayes and Decision Tree for Adaptive Intrusion Detection," Int. J. Netw. Secur. Its Appl., vol. 2, no. 2, pp. 12–25, 2010.

[60] J. E. Dickerson and J. A. Dickerson, "Fuzzy network profiling for intrusion detection," in PeachFuzz 2000. 19th International Conference of the North American Fuzzy Information Processing Society-NAFIPS (Cat. No. 00TH8500), 2000, pp. 301–306.

[61] K. Bajaj and A. Arora, "Dimension Reduction in Intrusion Detection Features Using Discriminative Machine Learning Approach.," … J. Comput. Sci. Issues (IJCSI …, vol. 10, no. 4, pp. 324–329, 2013.

[62] Z. Muda, W. Yassin, M. N. Sulaiman, and N. I. Udzir, "Intrusion detection based on K-Means clustering and Naïve Bayes classification," 2011 7th Int. Conf. Inf. Technol. Asia Emerg. Converg. Singul. Forms - Proc. CITA'11, pp. 1–6, 2011.

[63] W. Yassin, N. I. Udzir, and Z. Muda, "Anomaly-Based Intrusion Detection Through K- Means Clustering and Naives Bayes Classification," Proc. 4th Int. Conf. Comput. Informatics, ICOCI 2013, no. 049, pp. 298–303, 2013.

[64] M. Barati, A. Abdullah, R. Mahmod, N. Mustapha, and N. I. Udzir, "Features Selection for Ids in Encrypted Traffic Using Genetic Algorithm," Proc. 4th Int. Conf. Comput. Informatics, ICOCI 2013, no. 038, pp. 279–285, 2013.

[65] R. Xu and D. C. Wunsch, "Survey of clustering algorithms," 2005.

[66] E. Vasilomanolakis, S. Karuppayah, M. Muhlhauser, and M. Fischer, "Taxonomy and survey of collaborative intrusion detection," ACM Comput. Surv., vol. 47, no. 4, pp. 1–33, 2015.

[67] A. Fahad, Z. Tari, I. Khalil, A. Almalawi, and A. Y. Zomaya, "An optimal and stable feature selection approach for traffic classification based on multi-criterion fusion," Futur. Gener. Comput. Syst., vol. 36, pp. 156–169, 2014.

[68] A. Fahad, Z. Tari, I. Khalil, I. Habib, and H. Alnuweiri, "Toward an efficient and scalable feature selection approach for internet traffic classification," Comput. Networks, vol. 57, no. 9, pp. 2040–2057, 2013.

[69] Z. Liu, R. Wang, M. Tao, and X. Cai, "A class-oriented feature selection approach for multi-class imbalanced network traffic datasets based on local and global metrics fusion," Neurocomputing, vol. 168, pp. 365–381, 2015.

[70] E. De La Hoz, E. De La Hoz, A. Ortiz, J. Ortega, and A. Martínez-Álvarez, "Feature selection by multi-objective optimisation: Application to network anomaly detection by hierarchical self-organising maps," Knowledge-Based Syst., vol. 71, pp. 322–338, 2014.

[71] Y. Li, J. L. Wang, Z. H. Tian, T. B. Lu, and C. Young, "Building lightweight intrusion detection system using wrapper-based feature selection mechanisms," Comput. Secur., vol. 28, no. 6, pp. 466–475, 2009.

[72] K. Deb, A. Member, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multi-objective genetic algorithm:NSGAII," vol. 6, no. 2, pp. 182–197, 2002.

[73] H. Zhang, G. Lu, M. T. Qassrawi, Y. Zhang, and X. Yu, "Feature selection for optimizing traffic classification," Comput. Commun., vol. 35, no. 12, pp. 1457–1471, 2012.

[74] Y. Zhu, J. Liang, J. Chen, and Z. Ming, "An improved NSGA-III algorithm for feature selection used in intrusion detection," Knowledge-Based Syst., vol. 116, pp. 74–85, 2017.

[75] R. Kohavi and G. H. John, "Wrappers for feature subset selection," Artif. Intell., vol. 97, no. 1–2, pp. 273–324, 1997.

[76] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," J. Mach. Learn. Res., vol. 3, no. Mar, pp. 1157–1182, 2003.

[77] G. H. John, R. Kohavi, and K. Pfleger, "Irrelevant features and the subset selection problem," in Machine Learning Proceedings 1994, Elsevier, 1994, pp. 121–129.

[78] M. H. C. Law, M. A. T. Figueiredo, and A. K. Jain, "Simultaneous feature selection and clustering using mixture models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 26, no. 9, pp. 1154–1166, 2004.

[79] F. Amiri, M. Rezaei Yousefi, C. Lucas, A. Shakery, and N. Yazdani, "Mutual information-based feature selection for intrusion detection systems," J. Netw. Comput. Appl., vol. 34, no. 4, pp. 1184–1199, 2011.

[80] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution," Proceedings, Twent. Int. Conf. Mach. Learn., vol. 2, pp. 856–863, 2003.

[81] A. R. A. Yusof, N. I. Udzir, A. Selamat, H. Hamdan, and M. T. Abdullah, "Adaptive feature selection for denial of services (DoS) attack," 2017 IEEE Conf. Appl. Inf. Netw. Secur. AINS 2017, vol. 2018-Janua, pp. 1–4, 2018.

[82] H. Chen, R. Cheng, J. Wen, H. Li, and J. Weng, "Solving large-scale many-objective optimization problems by covariance matrix adaptation evolution strategy with scalable small subpopulations," Inf. Sci. (Ny)., 2018.

[83] Y. Xue, B. Zhao, T. Ma, and A. X. Liu, "An evolutionary classification method based on fireworks algorithm.," IJBIC, vol. 11, no. 3, pp. 149–158, 2018.

[84] K. Chen, F.-Y. Zhou, and X.-F. Yuan, "Hybrid particle swarm optimization with spiral-shaped mechanism for feature selection," Expert Syst. Appl., vol. 128, pp. 140–156, 2019.

[85] R. Vanaja and S. Mukherjee, "Novel Wrapper-Based Feature Selection for Efficient Clinical Decision Support System," in International Conference on Intelligent Information Technologies, 2018, pp. 113–129.

[86] Y. Zhang, D. Gong, and J. Cheng, "Multi-objective particle swarm optimization approach for cost-based feature selection in classification," IEEE/ACM Trans. Comput. Biol. Bioinforma., vol. 14, no. 1, pp. 64–75, 2017.

[87] B. Agarwal and N. Mittal, “Hybrid approach for detection of anomaly network traffic using data mining techniques,” Procedia Technol., vol. 6, pp. 996–1003, 2012.

[88] B. M. Aslahi-Shahri et al., “A hybrid method consisting of GA and SVM for intrusion detection system,” Neural Comput. Appl., vol. 27, no. 6, pp. 1669–1676, 2016.

[89] B. Ma and Y. Xia, “A tribe competition-based genetic algorithm for feature selection in pattern classification,” Appl. Soft Comput., vol. 58, pp. 328–338, 2017.

[90] T. Mehmod and H. B. M. Rais, “Ant colony optimization and feature selection for intrusion detection,” Lect. Notes Electr. Eng., vol. 387, pp. 305–312, 2016.

[91] F. Kuang, S. Zhang, Z. Jin, and W. Xu, “A novel SVM by combining kernel principal component analysis and improved chaotic particle swarm optimization for intrusion detection,” Soft Comput., vol. 19, no. 5, pp. 1187–1199, 2015.

[92] M. R. Gauthama Raman, N. Somu, K. Kirthivasan, R. Liscano, and V. S. Shankar Sriram, “An efficient intrusion detection system based on hypergraph - Genetic algorithm for parameter optimization and feature selection in support vector machine,” Knowledge-Based Syst., vol. 134, pp. 1–12, 2017.

[93] M. H. Ali, B. A. D. Al Mohammed, A. Ismail, and M. F. Zolkipli, “A New Intrusion Detection System Based on Fast Learning Network and Particle Swarm Optimization,” IEEE Access, vol. 6, pp. 20255–20261, 2018.

[94] S. Sarvari, N. F. Mohd Sani, Z. Mohd Hanapi, and M. T. Abdullah, “An Efficient Anomaly Intrusion Detection Method with Feature Selection and Evolutionary Neural Network,” IEEE Access, vol. 8, pp. 70651–70663, 2020.

[95] T. G. Dietterich, “Ensemble methods in machine learning,” in International workshop on multiple classifier systems, 2000, pp. 1–15.

[96] G. Folino and F. S. Pisani, “Combining ensemble of classifiers by using genetic programming for cyber security applications,” Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9028, no. February, 2015.

[97] M. P. Sesmero, A. I. Ledezma, and A. Sanchis, “Generating ensembles of heterogeneous classifiers using stacked generalization,” Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 5, no. 1, pp. 21–34, 2015.

[98] A. A. Aburomman and M. B. I. Reaz, “A survey of intrusion detection systems based on ensemble and hybrid classifiers,” Comput. Secur., vol. 65, pp. 135–152, 2017.

[99] L. Breiman, “Bagging predictors,” Mach. Learn., vol. 24, no. 2, pp. 123–140, 1996.

[100] L. Breiman, “Random forests,” Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[101] L. Breiman, “Pasting small votes for classification in large databases and on-line,” Mach. Learn., vol. 36, no. 1–2, pp. 85–103, 1999.

[102] Sangeeta Bhattacharya and S. Selvakumar, “LAWRA: a layered wrapper feature selection approach for network attack detection,” Secur. Commun. NETWORKS, vol. 2, pp. 71–81, 2015.

[103] M. Ambusaidi, X. He, P. Nanda, and Z. Tan, “Building an intrusion detection system using a filter-based feature selection algorithm,” IEEE Trans. Comput., vol. PP, no. 99, p. 1, 2016.

[104] S. Mohammadi, H. Mirvaziri, M. Ghazizadeh-Ahsaee, and H. Karimipour, “Cyber intrusion detection by combined feature selection algorithm,” J. Inf. Secur. Appl., vol. 44, pp. 80–88, 2019.

[105] S. Hosseini and B. M. H. Zade, “New hybrid method for attack detection using combination of evolutionary algorithms, SVM, and ANN,” Comput. Networks, vol. 173, no. March, p. 107168, 2020.

[106] H. Alazzam, A. Sharieh, and K. E. Sabri, “A feature selection algorithm for intrusion detection system based on Pigeon Inspired Optimizer,” Expert Syst. Appl., vol. 148, 2020.

[107] A. Golrang, A. M. Golrang, S. Y. Yayilgan, and O. Elezaj, “A novel hybrid ids based on modified NSGAII-ANN and random forest,” Electron., vol. 9, no. 4, pp. 1–19, 2020.

[108] Z. Zhang, J. Wen, J. Zhang, X. Cai, and L. Xie, “A Many Objective-Based Feature Selection Model for Anomaly Detection in Cloud Environment,” IEEE Access, vol. 8, pp. 60218–60231, 2020.

[109] N. V Chawla, L. O. Hall, K. W. Bowyer, T. E. Moore, and W. P. Kegelmeyer, “Distributed pasting of small votes,” in International Workshop on Multiple Classifier Systems, 2002, pp. 52–61.

[110] M. Govindarajan, “Hybrid Intrusion Detection Using Ensemble of Classification Methods,” Int. J. Comput. Netw. Inf. Secur., vol. 6, no. 2, pp. 45–53, 2014.

[111] R. E. Schapire, “The strength of weak learnability,” Mach. Learn., vol. 5, no. 2, pp. 197–227, 1990.

[112] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” J. Comput. Syst. Sci., vol. 55, no. 1, pp. 119–139, 1997.

[113] D. H. Wolpert, “Stacked generalization,” Neural networks, vol. 5, no. 2, pp. 241–259, 1992.

[114] A. KumarShrivas and A. Kumar Dewangan, “An Ensemble Model for Classification of Attacks with Feature Selection based on KDD99 and NSL-KDD Data Set,” Int. J. Comput. Appl., vol. 99, no. 15, pp. 8–13, 2014.

[115] M. R. Parsaei, S. M. Rostami, and R. Javidan, “A Hybrid Data Mining Approach for Intrusion Detection on Imbalanced NSL-KDD Dataset,” vol. 7, no. 6, pp. 20–25, 2016.

[116] A. R. Yusof, N. I. Udzir, and A. Selamat, “An Evaluation on KNN-SVM Algorithm for Detection and Prediction of DDoS Attack,” Springer Int. Publ. Switz., vol. 9799, no. 61272374, pp. 841–852, 2016.

[117] T. Ma, F. Wang, J. Cheng, Y. Yu, and X. Chen, “A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks,” Sensors (Switzerland), vol. 16, no. 10, 2016.

[118] Y. Zhou, G. Cheng, S. Jiang, and M. Dai, “Building an efficient intrusion detection system based on feature selection and ensemble classifier,” Comput. Networks, vol. 174, no. March, 2020.

[119] M. Sarnovsky and J. Paralic, “Hierarchical Intrusion Detection Using Machine Learning and Knowledge Model,” Symmetry (Basel)., vol. 12, no. 203, pp. 1–14, 2020.

[120] A. Andalib and V. Tabataba Vakili, “An Autonomous Intrusion Detection System Using an Ensemble of Advanced Learners,” 2020 28th Iran. Conf. Electr. Eng., 2020.

[121] C. Iwendi, S. Khan, J. H. Anajemba, M. Mittal, M. Alenezi, and M. Alazab, “The use of ensemble models for multiple class and binary class classification for improving intrusion detection systems,” Sensors (Switzerland), vol. 20, no. 9, pp. 1–37, 2020.