IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 25, 2019, accepted July 8, 2019, date of publication July 11, 2019, date of current version July 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2928048

TSE-IDS: A Two-Stage Classifier Ensemble for Intelligent Anomaly-Based Intrusion Detection System

BAYU ADHI TAMA[1], MARCO COMUZZI[1], AND KYUNG-HYUNE RHEE[2]
[1]School of Management Engineering, Ulsan National Institute of Science and Technology, Ulsan 44919, South Korea
[2]Department of IT Convergence and Applications Engineering, Pukyong National University, Busan 48513, South Korea

Corresponding author: Kyung-Hyune Rhee (khrhee@pknu.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korean Government (MSIT) under Grant NRF-2018R1D1A1B07048944.

ABSTRACT Intrusion detection systems (IDSs) play a pivotal role in computer security by discovering and repealing malicious activities in computer networks. Anomaly-based IDS, in particular, rely on classification models trained using historical data to discover such malicious activities. In this paper, an improved IDS based on hybrid feature selection and two-level classifier ensembles are proposed. A hybrid feature selection technique comprising three methods, i.e., particle swarm optimization, ant colony algorithm, and genetic algorithm, is utilized to reduce the feature size of the training datasets (NSL-KDD and UNSW-NB15 are considered in this paper). Features are selected based on the classification performance of a reduced error pruning tree (REPT) classifier. Then, a two-level classifier ensemble based on two meta learners, i.e., rotation forest and bagging, is proposed. On the NSL-KDD dataset, the proposed classifier shows 85.8% accuracy, 86.8% sensitivity, and 88.0% detection rate, which remarkably outperform other classification techniques recently proposed in the literature. The results regarding the UNSW-NB15 dataset also improve the ones achieved by several state-of-the-art techniques. Finally, to verify the results, a two-step statistical significance test is conducted. This is not usually considered by the IDS research thus far and, therefore, adds value to the experimental results achieved by the proposed classifier.

INDEX TERMS Two-stage meta classifier, network anomaly detection, hybrid feature selection, intrusion detection system, statistical significance test.

I. INTRODUCTION

Intrusion detection systems (IDSs) have been extensively recognized as a prominent technique for discovering and denying malevolent activities in a network [1]. As the number of malicious attacks is ceaselessly increasing, IDSs are much obliged to cope with the pruning of such attacks before they cause widespread destruction. Moreover, the present-day escalation of Internet of Things (IoT) devices and services has remarkably transformed our daily life. A large number of applications based on advanced IoT technology is successfully built and implemented, such as smart city, smart health care, smart home and vehicular networks [2]. These systems represent a further opportunity for attackers. According to [3], security is a primary barrier to the implementation of IoT network and services. This because IoT uses diverse standards and protocols, forming heterogeneous networks.

As the widespread development of IoT devices amplifies, insecure information processing is likely to put IoT networks at risk. The risk of compromising information disclosure in public spaces is particularly high with the broad development of IoT applications. Security architecture in IoT can be divided into three layers, i.e. perception layer, transportation layer, and application layer [4], [5]. Transportation layer includes network access security, which has an obligation to detect and prevent attacks. An IDS is a security mechanism which could be deployed in the transportation layer. It copes with security threats, e.g. DoS/DDoS attack, wireless LAN attack, or middle attack, which might harm the transportation security of IoT.

There are two types of IDSs, i.e., signature- and anomaly-based detection IDS. Signature-based detection deals with sniffing known attacks instantly with a lower false

The associate editor coordinating the review of this manuscript and approving it for publication was Ah Hwee Tan.

positive rate. Given its nature of dealing with known attack patterns, these techniques are less powerful when discovering *new* types of attack [6]. Anomaly-based detection, unlike signature-based detection, is able to discover novel attacks, by scanning and verifying the network patterns that are significantly different from the normal network operating patterns. As such, it constantly faces higher false positive rate. Moreover, in many cases, attackers may employ anomaly profiles disguised as normal profiles to train classification algorithms. As a result, an IDS would misapprehend anomalous patterns as normal ones. In the last decade, anomaly-based detection has gained much interest in IDS research because of the quick uprising of novel attack patterns [7], [8]. Considering the ability of network anomaly detection to discover new attack patterns, even a small detection improvement, e.g., a slightly reduced false alarm rate or a higher detection accuracy, would be extremely meaningful to avoid enterprises incurring in huge profit loss due to system performance failure and service unavailability resulting from successful attacks.

An efficient anomaly-based detection can be built using machine learning techniques. It involves solving a binary classification problem, by training a classifier to learn whether normal or anomaly usage patterns exist in the network [9]. A classification model is built using some intrusion datasets, i.e., NSL-KDD [10] or UNSW-NB15 [11], which are publicly available for benchmarking classifiers in IDS research. Various machine learning algorithms, including ensemble learning [12] and fuzzy classifier with evolutionary algorithm [13], have been proposed to improve the performance of anomaly-based IDS. More recently, deep learning [14] has been also considered, due to its prowess at uncovering complex structures of high dimensional data.

Existing solutions for anomaly-based IDS have harnessed different types of classifiers, either as individual classifiers or ensemble (meta) classifiers. When a single classification algorithm is unable to provide acceptable results, multiple classifier systems (MCSs) or classifier ensemble could be taken into account to offer a significant enhancement over individual classifier. MCSs train multiple classifiers to find a solution for the same problem [15]. In contrast to classical approaches, which build classifier model using one learner from the training set, MCSs built a set of classifiers and blend them to predict the final output.

In the past two decades, the combination of multiple classifiers has contributed to advance research in machine learning and pattern classification. Meta classifiers have been proposed in diverse real-life application domains, such as remote sensing, information security, fraud detection, health care, and recommender systems [16]. In such applications, MCSs show a plausible performance improvement over single classifiers. However, there remains underlying problems with meta classifier design, such as the classifier multeity and the choice of the appropriate techniques for combining the output of classifiers into a single one [16].

Most IDS research has focused on the utilization of long-established classification approaches either using individual classifiers, such as naive Bayes [10], [17], decision tree [10], [18]–[20], support vector machines [10], [21], [22], and naive Bayes tree [10]; or meta classifiers, i.e., bagging [23]–[25], boosting [25], [26], voting [27]–[29], random forest [10], [30], and other ensemble approaches [31]–[34]. In this paper, we propose a two-stage meta classifier for anomaly-based IDS, which utilizes two different ensembles, i.e., rotation forest [35] and bagging [36]. We demonstrate that the use of two-stage of meta classifier, combined with hybrid feature selection, can considerably improve the accuracy of anomaly-based IDS. The rationale behind choosing to design an anomaly-based IDS using a two-stage classifier ensemble is that such similar architecture model has shown remarkable accuracy in other domains, such as [37], [38]. However, in the cyber-security field, this type of design has not yet been considered.

Our contributions to the cyber-security domain are the following: (i) we propose an anomaly-based IDS based on a two-stage meta classifier, rather than an ensemble learner. The two-stage ensemble is composed by a meta classifier in the first stage whose base classifier is another meta classifier; (ii) we adopt a hybrid feature selection method to obtain a precise and accurate feature representation for the IDS problem, taking into account the fact that not all features are regarded as significant or even relevant in detecting intrusion; (iii) we conduct an extensive experimental evaluation of the proposed method to show that it produces a significant improvement of the detection rate on two different intrusion datasets when compared to several state of the art techniques; finally, (iv) we present a two-fold statistical test to demonstrate that the performance improvement shown by the proposed algorithm in respect of state of the art techniques are significant.

The remainder of the paper is organized as follows. Section II explores the existing solutions in anomaly-based IDS. A brief overview of anomaly-based IDS framework is given in Section III. This is followed by Section IV, discussing the experimental results. Finally, conclusions are drawn in Section V.

II. RELATED WORK

The issue of designing anomaly-based IDSs has been extensively researched in the literature. In this paper, we limit the review to approaches that have considered the NSL-KDD and the UNSW-NB15 datasets, i.e., the same recent datasets that we consider in this work, and that do not consider only cross-validation or hold out. The latter techniques are, in fact, not reliable enough in the context of IDSs, since training and testing are carried out using portions of the same dataset. This might lead to biased result, e.g. in some cases performance accuracy might achieve 99.9%. In this paper we use different testing sets, i.e. KDDTest+, KDDTest-21, and UNSW-NB15$_{test}$ for validation process. Therefore, we only consider works in the literature that take a similar approach. These selection criteria lead to excluding a number of approaches

([6], [19], [30], [39]–[47]), most of which using the outdated KDD Cup 99 dataset.

We firstly discuss the existing solutions considering the NSL-KDD dataset [10], i.e., an updated version of the KDD Cup 99 dataset. The work in [10] has benchmarked several individual classifiers in terms of their performance behavior on the two test datasets, i.e., KDDTest+ and KDDTest-21. The naive Bayes (NB) tree has been the best performing algorithm. A fuzzy-based classification algorithm for IDS is described in [13]. A full feature training set, e.g. KDDTrain+, and a separated test set, e.g. KDDTest+, are involved in the experiment. The fuzzy classifier improves the detection performance with respect to two performance metrics, i.e. accuracy and detection rate.

Rather than using a full feature set, Mohammadi *et al.* [18] propose a feature selection technique called Reduced Class-Dependent Feature Transformation (RCDFT). To evaluate the chosen feature set, several classification algorithms are used, i.e., decision tree (DT), multilayer perceptron (MLP), and distance-based classifier. DT performs better than MLP and distance-based classifier on the KDDTest+ dataset. In addition, the paper also evaluates other feature selection techniques, such as linear discriminant analysis (LDA), principal component analysis (PCA), and modified class-dependent feature transformation (MCDFT). Even if the false alarm rate has been lowered significantly, some classifiers still suffer from an unfavorable performance result in terms of accuracy and detection rate.

A two-layer dimension reduction and two-tier classification (TDTC) model for IDS is presented in [48]. A dimensional reduction module is used to decrease the high dimensional dataset to a lower one, with a smaller number of features. In addition, a two-tier classification module consisting of NB and certainty factor version of k-NN is applied for detecting suspicious behavior. As the proposed model is only applied on NSL-KDD, the results on different datasets are questionable.

A two-tier classifier along with LDA feature selection for IDS are proposed by [49]. The proposed classifier consists of two individual algorithms, i.e., NB and certainty factor voting version of k-NN. Its detection performance is then compared with other individual classification algorithms. According to the experiment, the model produces significant improvements of 83.4% and 4.83%, in terms of detection rate and false positive rate, respectively. In [34], authors have proposed a new tree-based ensemble technique, namely GAR-Forest. The GAR-Forest is used in combination with symmetrical uncertainty feature selection technique, showing a promising performance in terms of detection accuracy at 85.06%, using 32 features set. Nevertheless, the model bears a high false alarm rate of 12.2%.

A combination of hybrid feature selection and tree-based classifier ensemble for anomaly-based IDS is introduced in [6]. The proposed detection approach achieves 99.77% accuracy using a small size of feature set in the NSL-KDD dataset. This outperforms other similar techniques. The 10-fold cross validation (10*fcv*) is utilized as a validation method. Gradient boosting machine (GBM) is employed to detect anomaly activities in the network [9]. GBM is applied on the three different datasets, e.g. NSL-KDD, UNSW-NB15, and GPRS. It exhibits a significant performance improvement over other tree-based classifiers when it is validated both using train-test (hold out) and 10*fcv*.

More recently, a new classifier considering ramp loss function to the original one-class support vector machine for anomaly detection is developed in [50]. By using 10*fcv*, the proposed classifier obtains the best accuracy on both datasets, i.e., NSL-KDD and UNSW-NB15, when compared to other similar classification techniques, i.e., one-class SVM, ROCSVM. Previously, a genetic algorithm-logistic regression (GA-LR) wrapper approach for feature selection in network intrusion detection is initiated in [20]. A decision tree is used to evaluate the reduced feature set, obtaining 81.42% and 6.39% for accuracy and false alarm rate, respectively. A traditional ensemble approach, i.e., bagging (J48), and random forest for anomaly-based IDS are discussed in [25] and [51], respectively. Similar to [48], since the proposed classifiers are applied only on a single dataset (UNSW-NB15) the generalizability of the proposed methods is still debatable.

An anomaly detection technique based on deep learning model for Internet Industrial Control Systems (IICSs) is developed in [52]. The proposed detection model comprises a consecutive training process performed using a deep auto-encoder and deep feed-forward neural network architecture. The model is evaluated using NSL-KDD and UNSW-NB15 datasets. However, as the validation is conducted by simply dividing the dataset into training and testing set, the performance achieved is very high, which may be due to overfitting. Finally, in [21], a new feature selection technique for anomaly-based IDS called Modified Binary Grey Wolf Optimization (MBGWO) is designed. By considering a reduced feature set, the performance accuracy of SVM tested on KDDTest+ is improved when compared to other similar techniques, such as grey wolf optimizer (GWO), binary GWO, and MGWO. Nevertheless, the detection accuracy is still not able to compete against the previous works.

III. FRAMEWORK DESIGN

In this section, we first give an overview of the proposed framework at a conceptual level. The, we discuss in detail the feature selection and classifier modeling that we have adopted.

A. CONCEPTUAL FRAMEWORK OF IDS

A conceptual model of the proposed framework is given in Figure 1. The framework comprises three tiers, i.e. feature selection, classifier modeling, and validation. The first tier refers to the process of carefully choosing a feature set as the most appropriate for the anomaly detection task at hand. This is done using a hybrid technique relying on three evolutionary search techniques, i.e. particle swarm optimization (PSO), ant colony optimization (ACO), and genetic

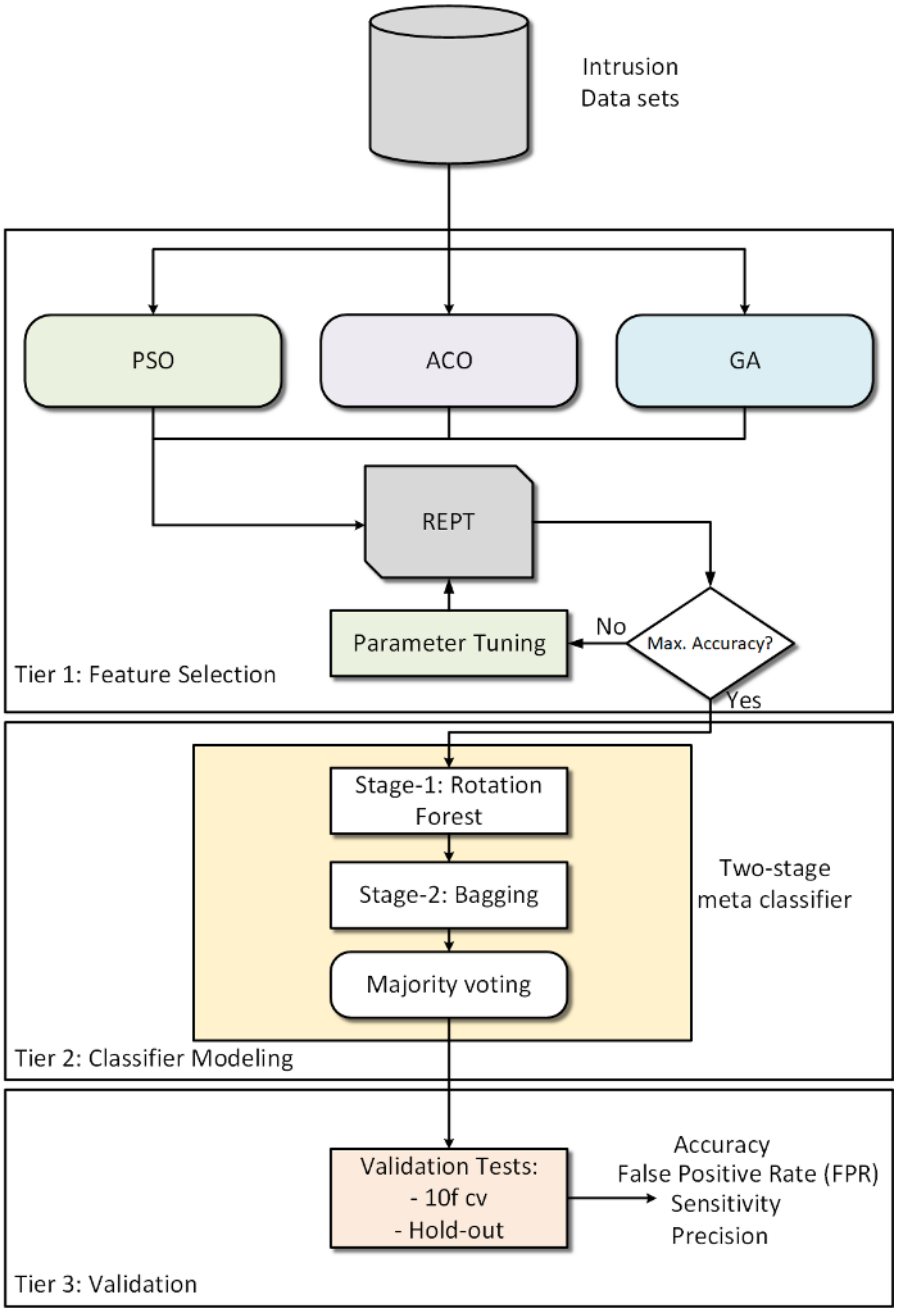

FIGURE 1. A conceptual framework for an anomaly-based IDS.

algorithms (GA). The feature selection method is described in depth in Section III-B.

In the second tier, a two-stage meta classifier for classification is designed. This tier is responsible for building classification model through the combination of two meta classifiers, i.e. rotation forest (RF) and bagging (BG). Since BG requires weak classifier, a conjunctive rule (CR) [53] classifier is chosen as base classifier. Following this, other meta combinations and a single classifier can be taken into consideration, e.g. bagging of CR (BG-CR), rotation forest of CR (RF-CR), and CR. These classifiers are further used as the basis of our classification analysis using statistical significance tests provided in Section IV-C. The two-stage meta classifier is described in depth in Section III-C.

Lastly, in the third layer, the proposed two-stage meta classifier is evaluated. This validation is performed using 10-fold cross validation (10*fcv*) [54]. We also consider a validation test using simple hold-out (train-test) approach applied on each provided test set for overall comparison with existing techniques. In addition, four performance measures that are frequently used in IDS research are taken into account. These are accuracy, false positive rate (FPR), sensitivity (also known

as recall), and precision. The experimental results are presented in Section IV.

B. TIER 1: HYBRID FEATURE SELECTION

A feature selection technique can be seen as a procedure for selecting a precise, compact and accurate subset of features from a given feature set. In this work, we choose a correlation-based feature selection, which estimates the importance of features using entropy and information gain [55]. In particular, irrelevant, noisy, and redundant features have to be excluded from the dataset in this tier.

We consider an evolutionary approach to feature selection, using three distinct evolutionary search techniques: PSO [56], GA [57], and ACO [58]:

- *Particle swarm optimization*. In this technique, a feature set is represented by particles in a swarm. Several particles are placed in an hyperspace in which each particle possesses random location x_i and velocity v_i. Let ω be the inertia weight constant, with c_1 and c_2 be the cognitive and social learning constant, respectively. Let also n_1 and n_2 be random numbers, p_i be the personal best location of particle i, and g be the global location among the particles. Then, the fundamental rules for updating the position and speed of each particle are:

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{1}$$

$$v_i(t+1) = \omega v_i(t) + c_1 n_1 (p_i - x_i(t)) + c_2 n_2 (g - x_i(t)) \tag{2}$$

- *Genetic algorithm*. In this technique, a set of features is represented by a chromosome. The existence of particular feature in a feature set is determined using binary value, either 1 (present) or 0 (missing). In addition, the Goldberg method is frequently taken into consideration to obtain the best feature set, while a k-fold cv is used by subset evaluator to examine the input features. In the experiment, it is necessary to specify the parameters such as initial population, mutation, crossover probability, and k.
- *Ant colony optimization*. In this techniques, adopting a graph representation, features are denoted by nodes and the selection of the best possible next feature is denoted by edges. The final feature subset is obtained through an ant search in the graph. The search stopping criterion is set to check a minimum number of visited nodes [59]. In addition, in order to evaluate which features are more informative among the currently chosen features, a probabilistic transition rule is utilized. Let k be the number of ants, J_i^k the set of ant k's unvisited features, η_{ij} the heuristic merit of picking feature j when presently at feature i, $\tau_{ij}(t)$ the amount of virtual pheromone on edge (i, j), then the likelihood of an ant at feature i to be willing to travel to feature j at time t is:

$$p_{ij}^k(t) = \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ij}]^\beta}{\sum_{l \in J_i^k} [\tau_{il}(t)]^\alpha \cdot [\eta_{il}]^\beta} \tag{3}$$

Several experiments have been carried out by tuning the size of particle, the number of ants, and the population size of PSO, ACO, and GA, respectively. A feature set is then selected by considering the maximum classification accuracy of a REPT classifier [60]. REPT is chosen due to its simplicity and speed in generating decision trees. It reduces the size of decision trees by pruning segments of the tree that contribute only marginally to sample classification. The classification accuracy of REPT is evaluated using subsampling (Monte-Carlo cross validation) technique. Subsampling is very similar to classical bootstrap [61]. It draws a training set D_{train} from D, whilst the remaining part of the dataset D_{test} is used for testing. The process is then repeated in a given number of iterations k. In the experiment, we choose $k = 40$ and sampling ratio 80/20.

C. TIER 2: A TWO-STAGE META CLASSIFIER

A meta classifier trains multiple individual classifiers, either in a parallel or serial manner. In order to construct a two-stage meta classifier, we employ two original ensembles, that is, rotation forest and bagging, which work as follows:

- *Rotation forest*. The goal of this meta classifier is to produce accurate and diverse classification algorithms. To create some feature subset projections, rotation forest uses principle component analysis (PCA). A number of independent feature subsets are trained using the same classification algorithm. Then, a full feature set for each classifier is collected, arranging the ensemble [35]. Let F and L be the feature set and number of subsets, respectively. The rotation forest splits randomly F into L subsets. PCA is then applied independently on each subset, and the new extracted features are collected by pooling all principle components. A dataset D is transformed into a new feature space, from which a classifier C_i is able to create a model. Independent split of the feature set yields the diversity of the extracted features.
- *Bagging*. In this meta classifier, several base individual classifiers are trained independently in parallel [36]. Let D be the original training set, which has n sample size. A number of M bootstrap samples $D_1, D_2, \ldots, D_M$ are randomly created from D. Next, an individual classifier C_i is trained on each bootstrap sample D_i. Finally, majority voting is taken to predict the final output of the new test instances. The final prediction C^* on a test instance, bagging feeds to its individual classifiers $C_1, C_2, \ldots, C_M$, collects all of the outputs, the votes of the label, and decides the winner label.

In this work, a new procedure for creating a two-stage meta classifier for an anomaly-based IDS is proposed. The proposed approach, unlike typical meta classifiers, which are frequently built from simple weak classifiers, considers two meta classifiers. Roughly speaking, it is a two-stage classification algorithm, where a meta classifier acts as a base model of another meta classifier. As illustrated in Figure 2, BG is chosen as a base model of RF, where another weak classifier, namely conjunctive rule (CR) [53] is chosen as

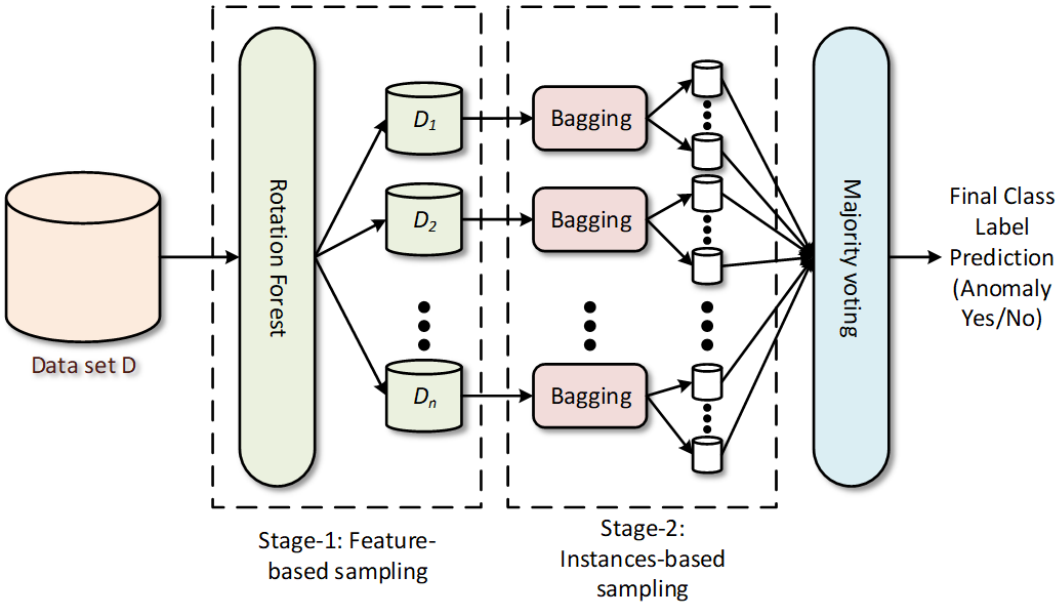

FIGURE 2. Proposed procedure of constructing a two-stage meta classifier.

a base model of BG. CR is prevalently recognized as an inductive learner, where the objective of rule induction is to generate a set of rules from the data [62].

In practice, several combinations of meta learners could be considered, since there exist a large number of meta learners in the literature. However, the combination that we propose is expected to maximize diversity, since RF and BG have different induction strategies, by taking into account the features (vertical induction) and samples (horizontal induction) of a training set, respectively. In the first stage, RF creates a feature set of D into L feature subsets. Subsequently, each feature subset is divided into M subsamples in the second stage classifier. Majority voting is utilized as an operator to aggregate the final class label prediction.

Let T be the total number of classifiers. Then, given the two meta learners, it holds that $T = L \times M$ classifiers. Let us assume that T classifiers $\{h_1, \ldots, h_T\}$ are specified and our goal is to concatenate h_i to predict the class label from a set of l possible class label $\{c_1, \ldots, c_l\}$. Suppose also that, for a given sample x, the final prediction of h_i is prescribed as a l-dimensional vector $(h_i^1(x), \ldots, h_i^l(x))^T$, where $h_i^j(x)$ is the output of h_i for the class label c_j. Hence, $h_i^j(x) \in \{0, 1\}$ holds value 1 if h_i estimates c_j as the class label, and 0 otherwise. Majority voting grants each classifier to vote one class label and the final class prediction $H(x)$ is chosen in accordance with the one that receives more than half of the votes, that is:

$$H(x) = \begin{cases} c_j & \text{if } \sum_{i=1}^{T} h_i^j(x) > \frac{1}{2} \sum_{k=1}^{l} \sum_{i=1}^{T} h_i^k(x) \\ rejection & \text{otherwise} \end{cases} \quad (4)$$

IV. EXPERIMENT RESULT AND DISCUSSION

Datasets and experimental settings are discussed in Section IV-A, while Section IV-B and Section IV-C present the result of experiments for feature selection and intrusion detection (including statistical tests), respectively.

A. INTRUSION DETECTION DATASETS

We consider the following publicly available intrusion detection datasets that are widely adopted in previous works:

- *NSL-KDD* [10]. It is an improved version of the KDD Cup 99 dataset, which does not have redundant samples, thus preventing classifiers to have a biased result. It comprises 42 features and a class label attribute. We consider 20% of dataset, so-called KDDTrain+, in the model training. KDDTrain+ consists of 25,192 samples, with 13,499 anomalous and 11,743 normal samples. In addition, we take into account two separated test sets, i.e., KDDTest+ (22,544 samples) and KDDTest-21 (11,850 samples), which are provided specifically for performance benchmark analysis.
- *UNSW-NB15* [11]. This dataset, unlike NSL-KDD, is an original version of an intrusion detection dataset that has appeared more recently. The full training set (UNSW-NB15$_{train}$) is composed by 42 features, with 37,000 samples in the normal class and 45,332 samples in the anomaly class. A specialized testing set (UNSW-NB15$_{test}$) is also used in the experiment. UNSW-NB15$_{test}$ has 175,341 samples.

All experiments discussed in the remainder of this paper were run on a Linux machine with 32G RAM memory and Intel Xeon Processor. The classifiers are implemented using the *RWeka* library [63].

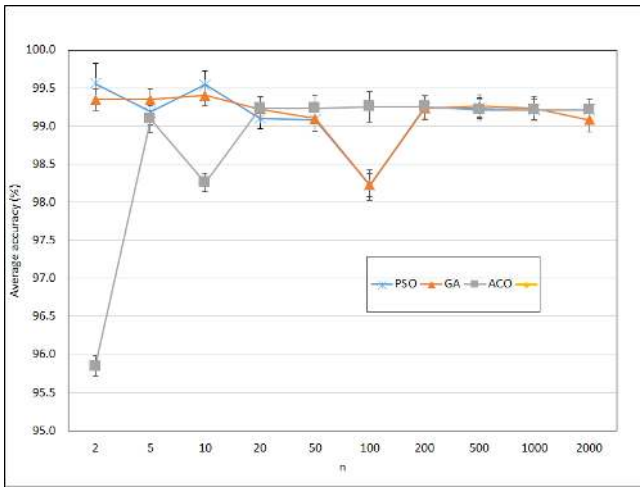

FIGURE 3. Classification accuracy of REPT for each search technique on NSL-KDD.

B. RESULTS OF FEATURE SELECTION

Experiments to determine the best configuration for feature selection are carried out by changing the value of the parameter n, representing the number of particles, population size, and ants in PSO, GA, and ACO, respectively, all other conditions being the same. Parameters setting for PSO are set as follows: c_1, c_2, number of generations, mutation type and mutation probability is set to 2, 2, 30, bit-flip, and 0.01, respectively. In GA, the initial population, maximum number of generations, mutation, crossover probability, k, and random seed number are set to 30, 30, 0.01, 0.9, 10 and 1, respectively. In ACO, the local pheromone update (α) and the relative importance of pheromone versus heuristic (β) are set to 0.8 and 1, respectively.

The results of REPT on NSL-KDD dataset for each search technique is presented in Figure 3. It is obvious that PSO with $n = 2$ indicates the best classification result with an accuracy of 99.557±0.134%. This case generates a set of 37 features (see Figure 4), namely: protocol_type, service, flag, src_bytes, dst_bytes, land, wrong_fragment, urgent, hot, num_failed_logins, logged_in, num_compromised, root_shell, su_attempted, num_file_creations, num_shells, num_outbound_cmds, is_host_login, is_guest_login, count, srv_count, serror_rate, srv_serror_rate, rerror_rate, srv_rerror_rate, same_srv_rate, srv_diff_host_rate, dst_host_count, dst_host_same_srv_rate, dst_host_diff_srv_rate, dst_host_same_src_port_rate, dst_host_srv_diff_host_rate, dst_host_serror_rate, dst_host_srv_serror_rate, dst_host_rerror_rate, and dst_host_srv_rerror_rate.

The results for the UNSW-NB15 dataset are visualized in Figure 5 and 6. PSO with $n = 5$ achieves the best classification accuracy of 97.055±0.125%, resulting in the following feature set (19 features): service, state, sbytes, dbytes, sttl, dttl, sinpkt, swin, dtcpb, tcprtt, ackdat, dmean, response_body_len, ct_state_ttl, ct_srt_dport_ltm, ct_dst_sport_ltm, ct_dst_src_ltm, ct_srv_dst, and is_sm_ips_ports.

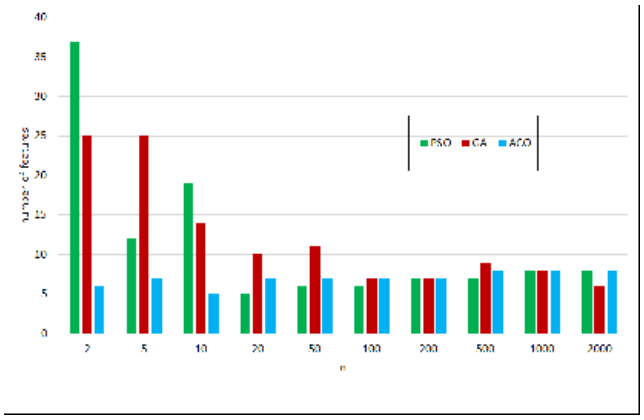

FIGURE 4. Number of selected features on NSL-KDD.

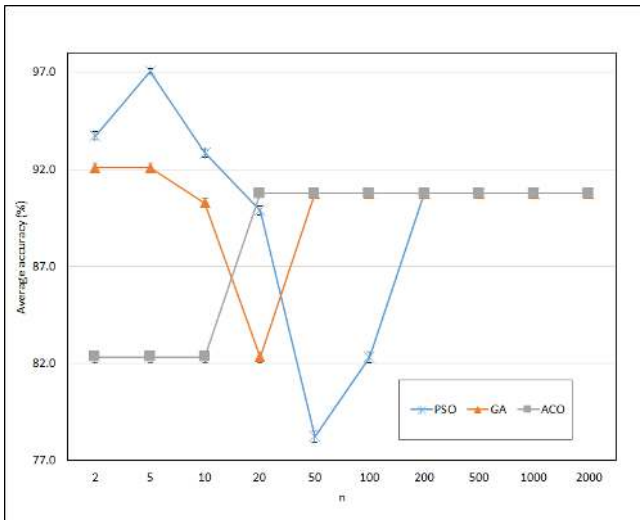

FIGURE 5. Classification accuracy of REPT for each search technique on UNSW-NB15.

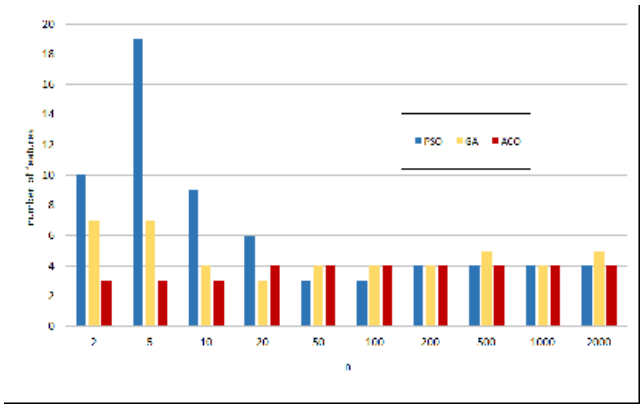

FIGURE 6. Number of selected features on UNSW-NB15.

The two selected feature sets discussed above are used in the next section for evaluating the performance of the two-stage classification model in the second tier of our framework.

C. INTRUSION DETECTION CLASSIFICATION ANALYSIS

This section evaluates the performance of the proposed two-stage classifier against other classifiers, namely bagging of

TABLE 1. Results of average accuracy (%) with standard deviations, Friedman average rank, and Iman-Davenport test in terms of 10-fold cross validation.

Method	NSL-KDD	UNSW-NB15	Friedman rank	Iman-Davenport p-value
Proposed	**96.388±1.093**	**81.533±2.959**	**1.0**	0.03407
BG-CR	94.026±0.639	76.632±0.385	3.3	
RF-CR	96.060±0.980	80.428±1.278	2.0	
CR	93.592±0.716	76.632±0.385	3.8	

∗ Best result is indicated in bold

TABLE 2. Results of the pairwise comparison using Friedman post-hoc test, where the proposed approach is chosen as a control classifier.

Benchmark	Friedman post-hoc p-value
BG-CR vs. Proposed	0.081
RF-CR vs. Proposed	0.439
CR vs. Proposed	0.033

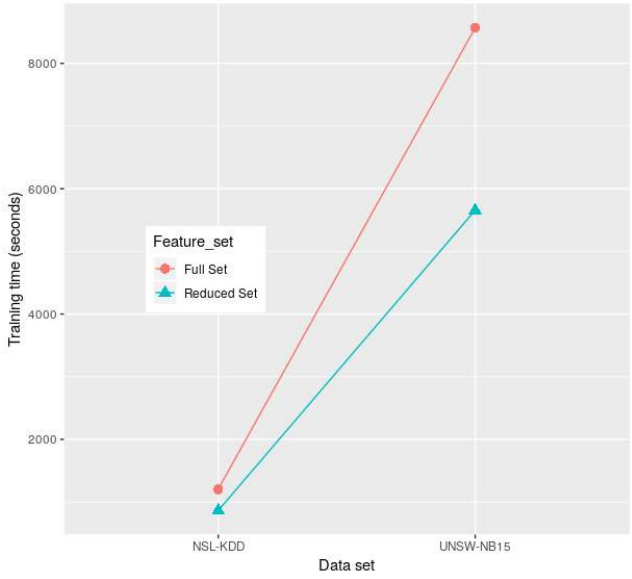

FIGURE 7. Training time taken by the proposed model (reduced set) and original full feature set.

CR (BG-CR), rotation forest of CR (RF-CR), and CR. For each classifier, we present the average accuracy using 10-fold cross-validation, i.e. 10*fcv*. We also report the results of performance comparisons using statistical tests.

In order to compare multiple classifiers, a common procedure is as follows [64]. First, omnibus tests, e.g. Friedman rank [65] and Iman-Davenport [66] are applied to determine the ranking of classifiers and to identify if at least one of the classifiers has performance difference among the competitors, respectively. More specifically, the goal of Iman-Davenport is to test whether all the classification algorithms perform equally, or, on the contrary, some of them hold a significant difference. Second, if such significant difference is found, then a pair-wise test, e.g., Friedman post-hoc with the corresponding p-value correction is used for multiple comparisons. Regarding post-hoc tests, the alternative of multiple comparisons normally rely on three possible options, i.e., pair-wise comparisons, comparison with control, and all pair-wise comparisons. In this paper, comparison with control is chosen, in which the proposed classifier is considered as a control classifier. In order to be marked as significant, the tests must be lower than a specified threshold p-value (0.05 in our case). Table 1 shows the average accuracy (along with the relative standard deviations) and Friedman average rank, as well as the result of the Iman-Davenport test obtained by different classifiers. Note that a lower ranking corresponds to a better classification performance.

The proposed classifier emerges as the clear best performer. The proposed classifier is, in fact, associated with the lowest (e.g. best) mean rank. The p-value $= 0.03407 < 0.05$ means that the null hypothesis, which indicates statistical equivalence among all algorithms, can be rejected. In addition, as the null hypothesis is rejected, it is necessary to carry out a post-hoc procedure using Friedman post-hoc test [67] to identify the pairs that indicate statistical differences between the four classification algorithms. As we have mentioned previously, the proposed approach is chosen as a control classifier as it holds smaller mean rank. The Friedman post-hoc results are shown in Table 2. It clearly shows how proposed classifier outperforms other classifiers. Also, from Table 2, it can be argued that the performance difference between the proposed classifier and the remaining classifiers, i.e. BG-CR, RF-CR, and CR is significant (p-value $= 0.081$), not too significant (p-value $= 0.439$), and highly significant (p-value $= 0.033$), respectively.

To extend this benchmark, we have compared the proposed two-stage classifier with the performance achieved by previous studies that use the datasets KDDTrain+ for training and KDDTest+ and KDDTest-21 for testing. We also include the result obtained by [10], where the NSL-KDD dataset has been firstly proposed. These results are shown in Table 3 and Table 4. Based on the experimental validation on KDDTest+, the highest detection accuracy is achieved by the proposed approach, which outperforms the most recent anomaly-based IDS techniques, i.e. SVM [21], bagging (J48) [25], and two-tier classifier [49]. Besides having superior detection accuracy, the proposed approach also outperforms significantly other approaches in terms of sensitivity and precision. Even though our proposed classifier does not perform best in terms of FPR metric, it is still comparable as being able to outperform GAR-Forest [34]. Moreover, according to a validation test applied on KDDTest-21, the proposed approach clearly outperforms classifiers available in the current literature, regardless of the evaluation metrics (see Table 4).

Table 5 shows the results regarding the UNSW-NB15$_{test}$ dataset. Here, other existing approaches, i.e., [17], [20], [51], which have used the same datasets are considered. The proposed classifier yields performance accuracy of 91.27%, which is substantially higher than other classifiers. However, in terms of FPR, there exists a slight difference of 2.51%

TABLE 3. Performance benchmark with some of the existing approaches on KDDTest+.

Method	Year	Feature selection	Accuracy (%)	FPR (%)	Sensitivity (%)	Precision (%)	Statistical test
Proposed (Two-stage ensemble)	2019	Hybrid	**85.797**	11.7	**86.8**	**88.0**	Yes
SVM [21]	2019	MBGWO	81.58	-	-	-	No
Bagging (J48) [25]	2018	Gain ratio	84.25	2.79	-	-	No
Two-tier classifier [49]	2017	LDA	83.240	4.8	-	-	No
TDTC [48]	2016	Two-layer	84.82	5.56	-	-	No
GAR-Forest [34]	2016	No	82.399	14.3	82.4	85.8	No
GAR-Forest [34]	2016	InfoGain	83.641	13.3	86.6	84.7	No
GAR-Forest [34]	2016	CFS	82.976	14.9	83.0	84.7	No
GAR-Forest [34]	2016	SU	85.056	12.2	85.1	87.5	No
SVM [22]	2014	Yes	82.37	15	-	-	No
Decision tree [18]	2012	RCDFT	80.141	**2.5**	-	67.0	No
Fuzzy classifier [13]	2011	No	82.740	3.9	86.7	-	No
NBTree [10]	2009	No	82.020	-	-	-	No
Random forest [10]	2009	No	80.670	-	-	-	No
SVM [10]	2009	No	69.520	-	-	-	No
Decision tree [10]	2009	No	81.050	-	-	-	No
Naive bayes [10]	2009	No	76.560	-	-	-	No

∗ Best result is indicated in bold

TABLE 4. Performance benchmark with some of the existing approaches on KDDTest-21.

Method	Year	Feature selection	Accuracy (%)	FPR (%)	Sensitivity (%)	Precision (%)	Statistical test
Proposed (Two-stage ensemble)	2019	Hybrid	**72.52**	**18.00**	**72.50**	**85.00**	Yes
Decision tree [18]	2012	RCDFT	58.80	27.67	-	-	No
NB tree [10]	2009	No	66.16	-	-	-	No

∗ Best result is indicated in bold

TABLE 5. Performance benchmark with some of the existing approaches on UNSW-NB15$_{test}$.

Method	Year	Feature selection	Accuracy (%)	FPR (%)	Sensitivity (%)	Precision (%)	Statistical test
Proposed (Two-stage ensemble)	2019	Hybrid	**91.27**	8.90	**91.30**	**91.60**	Yes
Two-stage classifier [51]	2018	Information gain	85.78	15.64	-	-	No
Decision tree [20]	2017	GA-LR	81.42	**6.39**	-	-	No
Decision tree [17]	2016	No	85.56	15.78	-	-	No
Logistic regression [17]	2016	No	83.15	18.48	-	-	No
Naive Bayes [17]	2016	No	82.07	18.56	-	-	No
Neural network [17]	2016	No	81.34	21.13	-	-	No
Expectation-maximization [17]	2016	No	78.47	23.79	-	-	No

∗ Best result is indicated in bold

between our approach and a most recent technique, (DT-GALR). Surprisingly, our proposed approach has performed better by 6.74% than two-stage classifier [51], in terms of accuracy and FPR.

The comparison analysis shown in Table 3 - 5 show that the proposed approach is very competitive as an effective approach for the anomaly-based intrusion detection task. In addition to the performance analysis, the statistical significance tests prove that the better performance of the proposed classifier is statistically significant when compared to state of the art techniques. Note that statistical tests are usually not provided by the other approaches in the literature that we have considered in this paper.

Finally, Figure 7 shows the execution time of the proposed classifier. The training time is calculated based on the computation time required for classification modeling. It is worth mentioning that the proposed model whose considerable reduced the training time when the optimal number features, obtained as the output of tier 1, is considered. For practical implementation, the time performance is acceptable, since the classification has to be trained only once and can then be used as an off-line anomaly detection tool in the network.

V. CONCLUSION

In this paper, a novel method for anomaly-based intrusion detection system based on the combination of hybrid feature selection and two-stage meta classifier has been proposed and discussed. Two intrusion datasets (NSL-KDD and UNSW-NB15) have been employed to evaluate the performance of the proposed approach. Based on the statistical significance tests, it could be concluded that the proposed approach outperforms other state of the art individual classifier and meta-classifiers, such as conjunctive rule (CR), bagging of CR (BG-CR), rotation forest of CR (RF-CR). The proposed method yields a superior result in terms of accuracy, specificity, and precision metric when validated against pre-specified testing sets, i.e. KDDTest+, KDDTest-21, and UNSW-NB15$_{test}$. For future work, it is required to validate the proposed approach in solving a multi-class

classification problem, which represents incoming network traffic as normal or some attack groups.

REFERENCES

[1] B. A. Tama and K.-H. Rhee, "An extensive empirical evaluation of classifier ensembles for intrusion detection task," *Comput. Syst. Sci. Eng.*, vol. 32, no. 2, pp. 149–158, 2017.
[2] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, Oct. 2010.
[3] S. Sicari, A. Rizzardi, L. A. Grieco, and A. Coen-Porisini, "Security, privacy and trust in Internet of Things: The road ahead," *Comput. Netw.*, vol. 76, pp. 146–164, Jan. 2015.
[4] L. Liu and S. Lai, "ALOHA-based anti-collision algorithms used in RFID system," in *Proc. Int. Conf. Wireless Commun., Netw. Mobile Comput. (WiCOM)*, Sep. 2006, pp. 1–4.
[5] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, and D. Qiu, "Security of the Internet of Things: Perspectives and challenges," *Wireless Netw.*, vol. 20, no. 8, pp. 2481–2501, Nov. 2014.
[6] B. A. Tama and K.-H. Rhee, "HFSTE: Hybrid feature selections and tree-based classifiers ensemble for intrusion detection system," *IEICE Trans. Inf. Syst.*, vol. 100, no. 8, pp. 1729–1737, 2017.
[7] B. A. Tama and K. H. Rhee, "Performance analysis of multiple classifier system in DoS attack detection," in *Proc. Int. Workshop Inf. Secur. Appl.* Cham, Switzerland: Springer, 2015, pp. 339–347.
[8] X.-S. Gan, J.-S. Duanmu, J.-F. Wang, and W. Cong, "Anomaly intrusion detection based on PLS feature extraction and core vector machine," *Knowl.-Based Syst.*, vol. 40, pp. 1–6, Mar. 2013.
[9] B. A. Tama and K.-H. Rhee, "An in-depth experimental study of anomaly detection using gradient boosted machine," *Neural Comput. Appl.*, vol. 31, no. 4, pp. 955–965, 2019.
[10] M. Tavallaee, E. Bagheri, W. Lu, and A.-A. Ghorbani, "A detailed analysis of the KDD CUP 99 data set," in *Proc. 2nd IEEE Symp. Comput. Intell. Secur. Defence Appl.*, Jul. 2009, pp. 1–6.
[11] N. Moustafa and J. Slay, "UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set)," in *Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS)*, Nov. 2015, pp. 1–6.
[12] B. A. Tama and K.-H. Rhee, "Performance evaluation of intrusion detection system using classifier ensembles," *Int. J. Internet Protocol Technol.*, vol. 10, no. 1, pp. 22–29, 2017.
[13] P. Krömer, J. Platoš, V. Snášel, and A. Abraham, "Fuzzy classification by evolutionary algorithms," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2011, pp. 313–318.
[14] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015.
[15] Z.-H. Zhou, *Ensemble Methods: Foundations and Algorithms*. Boca Raton, FL, USA: CRC Press, 2012.
[16] M. Wo niak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Inf. Fusion*, vol. 16, pp. 3–17, Mar. 2014.
[17] N. Moustafa and J. Slay, "The evaluation of network anomaly detection systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set," *Inf. Secur. J., Global Perspective*, vol. 25, nos. 1–3, pp. 18–31, 2016.
[18] M. Mohammadi, B. Raahemi, A. Akbari, and B. Nassersharif, "New class-dependent feature transformation for intrusion detection systems," *Secur. Commun. Netw.*, vol. 5, no. 12, pp. 1296–1311, 2012.
[19] A. Khraisat, I. Gondal, and P. Vamplew, "An anomaly intrusion detection system using C5 decision tree classifier," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining*. Cham, Switzerland: Springer, 2018, pp. 149–155.
[20] C. Khammassi and S. Krichen, "A GA-LR wrapper approach for feature selection in network intrusion detection," *Comput. Secur.*, vol. 70, pp. 255–277, Sep. 2017.
[21] Q. M. Alzubi, M. Anbar, Z. N. M. Alqattan, M. A. Al-Betar, and R. Abdullah, "Intrusion detection system based on a modified binary grey wolf optimisation," *Neural Comput. Appl.*, to be published.
[22] M. S. Pervez and D. M. Farid, "Feature selection and intrusion classification in NSL-KDD cup 99 dataset employing SVMs," in *Proc. 8th Int. Conf. Softw., Knowl., Inf. Manage. Appl. (SKIMA)*, Dec. 2014, pp. 1–6.
[23] S. S. S. Sindhu, S. Geetha, and A. Kannan, "Decision tree based light weight intrusion detection using a wrapper approach," *Expert Syst. Appl.*, vol. 39, no. 1, pp. 129–141, 2012.
[24] M. Govindarajan and R. M. Chandrasekaran, "Intrusion detection using neural based hybrid classification methods," *Comput. Netw.*, vol. 55, no. 8, pp. 1662–1671, 2011.
[25] N. T. Pham, E. Foo, S. Suriadi, H. Jeffrey, and H. F. M. Lahza, "Improving performance of intrusion detection system using ensemble methods and feature selection," in *Proc. Australas. Comput. Sci. Week Multiconf.*, 2018, p. 2.
[26] W. Hu, W. Hu, and S. Maybank, "AdaBoost-based algorithm for network intrusion detection," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 38, no. 2, pp. 577–583, Apr. 2008.
[27] S. Mukkamala, A. H. Sung, and A. Abraham, "Intrusion detection using an ensemble of intelligent paradigms," *J. Netw. Comput. Appl.*, vol. 28, no. 2, pp. 167–182, 2005.
[28] R. Perdisci, G. Gu, and W. Lee, "Using an ensemble of one-class SVM classifiers to harden payload-based anomaly detection systems," in *Proc. 6th Int. Conf. Data Mining (ICDM)*, 2006, pp. 488–498.
[29] 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.
[30] R. Primartha and B. A. Tama, "Anomaly detection using random forest: A performance revisited," in *Proc. Int. Conf. Data Softw. Eng. (ICoDSE)*, Nov. 2017, pp. 1–6.
[31] J. B. D. Cabrera, C. Gutiérrez, and R. K. Mehra, "Ensemble methods for anomaly detection and distributed intrusion detection in mobile ad-hoc networks," *Inf. Fusion*, vol. 9, no. 1, pp. 96–119, 2008.
[32] G. Giacinto, R. Perdisci, M. Del Rio, and F. Roli, "Intrusion detection in computer networks by a modular ensemble of one-class classifiers," *Inf. Fusion*, vol. 9, no. 1, pp. 69–82, 2008.
[33] E. de la Hoz, A. Ortiz, J. Ortega, and E. de la Hoz, "Network anomaly classification by support vector classifiers ensemble and non-linear projection techniques," in *Proc. Int. Conf. Hybrid Artif. Intell. Syst.* Berlin, Germany: Springer, 2013, pp. 103–111.
[34] N. K. Kanakarajan and K. Muniasamy, "Improving the accuracy of intrusion detection using GAR-Forest with feature selection," in *Proc. 4th Int. Conf. Frontiers Intell. Comput., Theory Appl. (FICTA)*. New Delhi, India: Springer, 2016, pp. 539–547.
[35] J. J. Rodriguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1619–1630, Oct. 2006.
[36] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, no. 2, pp. 123–140, 1996.
[37] A. Marqués, V. García, and J. S. Sánchez, "Two-level classifier ensembles for credit risk assessment," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 10916–10922, 2012.
[38] S. Bashir, U. Qamar, F. H. Khan, and L. Naseem, "HMV: A medical decision support framework using multi-layer classifiers for disease prediction," *J. Comput. Sci.*, vol. 13, pp. 10–25, Mar. 2016.
[39] A. I. Saleh, F. M. Talaat, and L. M. Labib, "A hybrid intrusion detection system (HIDS) based on prioritized k-nearest neighbors and optimized SVM classifiers," *Artif. Intell. Rev.*, vol. 51, no. 3, pp. 403–443, 2017.
[40] Ü. Çavu o lu, "A new hybrid approach for intrusion detection using machine learning methods," *Appl. Intell.*, vol. 49, no. 7, pp. 2735–2761, 2019.
[41] K. A. Taher, B. M. Y. Jisan, and M. M. Rahman, "Network intrusion detection using supervised machine learning technique with feature selection," in *Proc. Int. Conf. Robot., Elect. Signal Process. Techn. (ICREST)*, Jan. 2019, pp. 643–646.
[42] I. Ullah and Q. H. Mahmoud, "A two-level hybrid model for anomalous activity detection in IoT networks," in *Proc. 16th IEEE Annu. Consum. Commun. Netw. Conf. (CCNC)*, Jan. 2019, pp. 1–6.
[43] S. Rezvy, M. Petridis, A. Lasebae, and T. Zebin, "Intrusion detection and classification with autoencoded deep neural network," in *Proc. Int. Conf. Secur. Inf. Technol. Commun.* Cham, Switzerland: Springer, 2018, pp. 142–156.
[44] H. Choi, M. Kim, G. Lee, and W. Kim, "Unsupervised learning approach for network intrusion detection system using autoencoders," *J. Supercomput.*, to be published.
[45] F. Gottwalt, E. Chang, and T. Dillon, "CorrCorr: A feature selection method for multivariate correlation network anomaly detection techniques," *Comput. Secur.*, vol. 83, pp. 234–245, Jun. 2019.
[46] A. Niranjan, D. Nutan, A. Nitish, P. D. Shenoy, and K. Venugopal, "ERCR TV: Ensemble of random committee and random tree for efficient anomaly classification using voting," in *Proc. 3rd Int. Conf. Converg. Technol. (I2CT)*, Apr. 2018, pp. 1–5.

[47] B. A. Tama and K.-H. Rhee, "An integration of PSO-based feature selection and random forest for anomaly detection in IoT network," in *Proc. MATEC Web Conf.*, vol. 159, 2018, p. 1053.
[48] H. H. Pajouh, R. Javidan, R. Khayami, D. Ali, and K.-K. R. Choo, "A two-layer dimension reduction and two-tier classification model for anomaly-based intrusion detection in IoT backbone networks," *IEEE Trans. Emerg. Topics Comput.*, to be published.
[49] H. H. Pajouh, G. Dastghaibyfard, and S. Hashemi, "Two-tier network anomaly detection model: A machine learning approach," *J. Intell. Inf. Syst.*, vol. 48, no. 1, pp. 61–74, 2017.
[50] Y. Tian, M. Mirzabagheri, S. M. H. Bamakan, H. Wang, and Q. Qu, "Ramp loss one-class support vector machine; A robust and effective approach to anomaly detection problems," *Neurocomputing*, vol. 310, pp. 223–235, Oct. 2018.
[51] W. Zong, Y.-W. Chow, and W. Susilo, "A two-stage classifier approach for network intrusion detection," in *Proc. Int. Conf. Inf. Secur. Pract. Exper.* Cham, Switzerland: Springer, 2018, pp. 329–340.
[52] A.-H. Muna, N. Moustafa, and E. Sitnikova, "Identification of malicious activities in industrial Internet of Things based on deep learning models," *J. Inf. Secur. Appl.*, vol. 41, pp. 1–11, Aug. 2018.
[53] W. W. Cohen, "Fast effective rule induction," in *Machine Learning Proceedings*. Amsterdam, The Netherlands: Elsevier, 1995, pp. 115–123.
[54] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. Roy. Stat. Soc. B (Methodol.)*, vol. 36, no. 2, pp. 111–147, 1974.
[55] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, Univ. Waikato, Hamilton, New Zealand, 1999.
[56] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. Comput. Simulation*, vol. 5, Oct. 1997, pp. 4104–4108.
[57] D. E. Goldberg and J. H. Holland, "Genetic algorithms and machine learning," *Mach. Learn.*, vol. 3, nos. 2–3, pp. 95–99, 1988.
[58] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*, no. 1. Oxford, U.K.: Oxford Univ. Press, 1999.
[59] R. Jensen and Q. Shen, "Fuzzy-rough data reduction with ant colony optimization," *Fuzzy Sets Syst.*, vol. 149, no. 1, pp. 5–20, 2005.
[60] J. R. Quinlan, "Simplifying decision trees," *Int. J. Hum.-Comput. Stud.*, vol. 51, no. 2, pp. 497–510, 1999.
[61] B. Bischl, O. Mersmann, H. Trautmann, and C. Weihs, "Resampling methods for meta-model validation with recommendations for evolutionary computation," *Evol. Comput.*, vol. 20, no. 2, pp. 249–275, 2012.
[62] P. Clark and T. Niblett, "The CN2 induction algorithm," *Mach. Learn.*, vol. 3, no. 4, pp. 261–283, 1989.
[63] K. Hornik, C. Buchta, and A. Zeileis, "Open-source machine learning: R meets Weka," *Comput. Statist.*, vol. 24, no. 2, pp. 225–232, 2009.
[64] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.
[65] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Statist. Assoc.*, vol. 32, no. 200, pp. 675–701, Dec. 1937.
[66] R. L. Iman and J. M. Davenport, "Approximations of the critical region of the fbietkan statistic," *Commun. Statist.-Theory Methods*, vol. 9, no. 6, pp. 571–595, 1980.
[67] M. Friedman, "A comparison of alternative tests of significance for the problem of m rankings," *Ann. Math. Statist.*, vol. 11, no. 1, pp. 86–92, 1940.

BAYU ADHI TAMA received the Ph.D. degree from Pukyong National University, South Korea. He is currently a Postdoctoral Researcher with the Industrial Artificial Intelligence Laboratory, Pohang University of Science and Technology (POSTECH), South Korea. Prior to this, he was a Postdoctoral Researcher with the School of Management Engineering, Ulsan National Institute of Science and Technology (UNIST), South Korea. His research interests include machine learning and artificial intelligence techniques applied for cyber security, medical informatics, and industrial applications. He received the Full Scholarship for his doctoral study from the Korean Government and an award for his excellent academic achievement from the Ministry of Education, South Korea. He serves as a Reviewer in many top-rank conferences and journals.

MARCO COMUZZI received the Ph.D. degree from the Politecnico di Milano, Italy, in 2007. Since then, he has held academic positions at the Eindhoven University of Technology and the City, University of London. Since 2016, he has been with the School of Management Engineering, Ulsan National Institute of Science and Technology (UNIST), South Korea, where he is currently an Associate Professor and the Head of the UNIST Blockchain Research Center. He has published more than 50 papers in international academic conference and journals. His research interests include enterprise systems design, blockchain and, principally, business process management. Recently, his research is focusing on the application of machine learning techniques for predictive monitoring using process event logs.

KYUNG-HYUNE RHEE received the M.S. and Ph.D. degrees from the Korea Advanced Institute of Science and Technology (KAIST), South Korea, in 1985 and 1992, respectively. He was a Senior Researcher with the Electronic and Telecommunications Research Institute (ETRI), South Korea, from 1985 to 1993. He was also a Visiting Scholar with the University of Adelaide, the University of Tokyo, and the University of California at Irvine. He has served as the Chairman of the Division of Information and Communication Technology, Colombo Plan Staff College for Technician Education, Manila, Philippines. He is currently a Professor with the Department of IT Convergence and Application Engineering, Pukyong National University, South Korea. He also currently serves as the President of the Korea Institute of Information Security and Cryptology (KIISC). His research interests include key management and its applications, mobile communication security, and security evaluation of cryptographic algorithms.