Article

PSO-Driven Feature Selection and Hybrid Ensemble for Network Anomaly Detection

Maya Hilda Lestari Louk [1] and Bayu Adhi Tama [2,*]

1   Department of Informatics Engineering, University of Surabaya, Surabaya 60293, Indonesia
2   Department of Information Systems, University of Maryland, Baltimore County (UMBC),
    Baltimore, MD 21250, USA
*   Correspondence: bayu@umbc.edu

Abstract: As a system capable of monitoring and evaluating illegitimate network access, an intrusion detection system (IDS) profoundly impacts information security research. Since machine learning techniques constitute the backbone of IDS, it has been challenging to develop an accurate detection mechanism. This study aims to enhance the detection performance of IDS by using a particle swarm optimization (PSO)-driven feature selection approach and hybrid ensemble. Specifically, the final feature subsets derived from different IDS datasets, i.e., NSL-KDD, UNSW-NB15, and CICIDS-2017, are trained using a hybrid ensemble, comprising two well-known ensemble learners, i.e., gradient boosting machine (GBM) and bootstrap aggregation (bagging). Instead of training GBM with individual ensemble learning, we train GBM on a subsample of each intrusion dataset and combine the final class prediction using majority voting. Our proposed scheme led to pivotal refinements over existing baselines, such as TSE-IDS, voting ensembles, weighted majority voting, and other individual ensemble-based IDS such as LightGBM.

1. Introduction

An intrusion detection system, often known as an IDS, has the potential to make significant contributions to the field of information security research due to its capability to monitor and identify unauthorized access targeted at computing and network resources [1,2]. In conjunction with other mitigation techniques, such as access control and user authentication, an IDS is often utilized as a secondary line of defense in computer networks. In the past few decades, machine learning techniques have been applied to the network audit log to construct models for identifying attacks [3]. In this scenario, intrusion detection can be viewed as a data analytics process in which machine learning techniques are used to automatically uncover and model characteristics of a user's suspicious or normal behavior. Ensemble learning is a popular machine learning approach in which multiple distinct classifiers are weighted and combined to produce a classifier that outperforms each of them individually [4].

Tama and Lim [5] looked at how recent ensemble learning techniques have been exploited in IDS through a systematic mapping study. They argued that ensemble learning has made a significant difference over standalone classifiers, though this is sometimes the case, depending upon the voting schemes and base classifiers used to build the ensemble. This makes it challenging to design an accurate detection mechanism based on ensemble learning. Moreover, an IDS has to cope with an enormous amount of data that may contain unimportant features, resulting in poor performance. Consequently, selecting relevant features is considered a crucial criterion for IDS [6,7]. Feature selection minimizes redundant information, improves detection algorithm accuracy, and enhances generalization.

This article focuses on evaluating anomaly-based IDS by leveraging the combination of a feature selection technique and hybrid ensemble learning. More precisely, we adopt a particle swarm optimization (PSO) method as a search algorithm to traverse the whole feature space and assess potential feature subsets. Next, a hybrid ensemble learning approach, comprising two ensemble paradigms—gradient boosting machine (GBM) [8] and bootstrap aggregation (bagging) [9]—is utilized to improve the detection accuracy. Our proposed detector, combined with a feature selection technique, can substantially affect the performance accuracy of network anomaly detection with a comparable result over existing baselines. To put it in a nutshell, this article presents advancements to the existing IDS techniques.

(a) A simple yet accurate network anomaly detection using hybrid bagging and GBM ensemble is proposed. GBM is not trained independently as a classifier; rather, we use it as the base learning model for bagging in order to increase its detection performance.

(b) A PSO-guided feature selection is applied to choose the most optimal subset of features for the input of the hybrid ensemble model. The full feature set may not give substantial prediction accuracy; thus, we use an optimum feature subset derived from the PSO-based feature selection approach.

(c) Based on our experiment validation, our proposed model is superior compared to existing anomaly-based IDS methods presented in the current literature.

We break down the remaining parts of this article as follows. In Section 2, a brief survey of prior detection techniques is provided, followed by the description of the datasets and hybrid ensemble in Section 3. The experimental result is discussed in Section 4; lastly, some closing notes are given in Section 5.

2. Related Work

Ensemble learning approaches are not a novel IDS methodology. In IDS, combining multiple weak classifiers to generate a robust classifier has been discussed for a very significant period of time [5,10–15]. In this section, existing anomaly-based IDS methods employing feature selection and ensemble learning are explored briefly. It is worth mentioning that in order to give the most up-to-date literature on anomaly detectors, we have included publications published between 2020 and the present. Table 1 presents a summarization of each existing work published as an article, listed in chronological order.

Table 1. Summarization of prior anomaly-based intrusion detection techniques that employ feature selection and ensemble learning. The articles are chronologically ordered between 2020 and the present.

Author(s)	Ensemble Approach(es)	Base Learner(s)	Feature Selector	Validation Method(s)	Dataset(s)
[16]	Stacking	NN, NB, DL, SVM	IG	Hold-out	Private
[17]	AB, stacking	LR, RF	PCA	CV and hold-out	NSL-KDD, UNSW-NB15
[18]	RF, XGBoost, HGB, LightGBM	-	RF+PCA	CV	CICIDS-2018
[19]	XGBoost	-	GA	CV	CIRA-CIC-DoHBrw-2020, Bot-IoT, UNSW-NB15
[20]	RF	-	Gain ratio, Chi-squared, Pearson correlation	Hold-out	UNSW-NB15
[21]	Stacking	RF, LR	K-means	Hold-out	NSL-KDD, CIDDS-2017, Testbed

Table 1. *Cont.*

Author(s)	Ensemble Approach(es)	Base Learner(s)	Feature Selector	Validation Method(s)	Dataset(s)
[22]	Majority voting	SVM, NB, LR, DT	Filter and univariate ensemble	CV	Honeypot, NSL-KDD, Kyoto
[23]	LightGBM	-	-	CV	NSL-KDD, UNSW-NB15, CICIDS-2017
[24]	RF	-	-	Hold-out	CIDDS-001, UNSW-NB15
[25]	Weighted voting	C4.5, MLP, IBL	IFA	CV	NSL-KDD, UNSW-NB15
[26]	RF	-	-	CV	NSL-KDD, UNSW-NB15, CICIDS-2017
[27]	XGBoost, RF	-	-	Hold-out	NSL-KDD, CIDDS-001, CICIDS-2017
[28]	Weighted majority voting	SVM, LR, NB, DT	Gain-ratio, Chi-squared, Information gain	Hold-out	Honeypot, NSL-KDD, Kyoto
[29]	Stacking	DT, RF, XGBoost	*SelectKbest*	CV	NSL-KDD, UNSW-NB15
[30]	LightGBM	-	DNN	Hold-out	KDD-99, NSL-KDD, UNSW-NB15

Stacking [31] has been commonly mentioned as one of the ensemble procedures. It is a general method in which a classification algorithm is trained to integrate heterogeneous algorithms. Individual algorithms are referred to as first-level algorithms, while the combiner is referred to as a second-level algorithm or meta-classifier. Jafarian et al. [16], Kaur [17], Jain and Kaur [21], Rashid et al. [29], Wang et al. [30] demonstrate that stacking generates a promising intrusion detection capability; however, most of the proposed stacking procedures do not consider LR as a second-level algorithm, as suggested by [32]. Alternatively, combiner strategies, such as majority voting [22] and weighted majority voting [25,28] may be utilized as anomaly detectors. The most prevalent mode of voting is majority rule. In this context, each algorithm casts a vote for one class label, with the class label receiving more than fifty percent of the votes serving as the final output class label; if none of the class labels acquires more than fifty percent of the votes, a rejection choice will be given, and the blended algorithm will not make a prediction. On the other hand, if individual algorithms have inequitable performance, it seems reasonable to assign the more robust algorithms more significant influence during voting; this is achieved by weighted majority voting.

Furthermore, it is possible to construct homogeneous ensembles in which an ensemble procedure is built upon a single (e.g., the same type) algorithm. Kaur [17] compares three different adaptive boosting (AB) [33] families of algorithms for anomaly-based IDS, while the rest of proposed approaches utilize tree-based ensemble learning, such as RF [18,20,24,26,27], LightGBM [18,23,30], and XGBoost [18,19,27].

In the intrusion detection field, feature selection techniques have also been exploited [34,35]. Specifically, bio-inspired algorithms have gained popularity and evolved into an alternate method for finding the optimal feature subset from the feature space [19,25,36]. Other filter-based approaches such as IG, gain ratio, chi-squared, and Pearson correlation have been intensively utilized to remove unnecessary features [16,20,22,28,29]. The filter technique assesses feature subsets according to given criteria regardless of any grouping. Information gain, for example, utilizes a weighted feature scoring system to obtain the highest entropy value. In addition, previous research indicates that feature selectors using the wrapper technique are taken into account. A wrapper-based feature selector evaluates a specific machine learning algorithm to search optimal feature subset [17,18,21,30]. Examining the above-mentioned methods for anomaly detectors, our study fills a gap by examining hybrid ensemble and PSO-based feature selection, both of which are underexplored in the existing literature.

3. Materials and Methods

This seeks assess the performance of network anomaly detection using PSO-based feature selection and hybrid ensemble. Figure 1 denotes the phases of our detection framework.
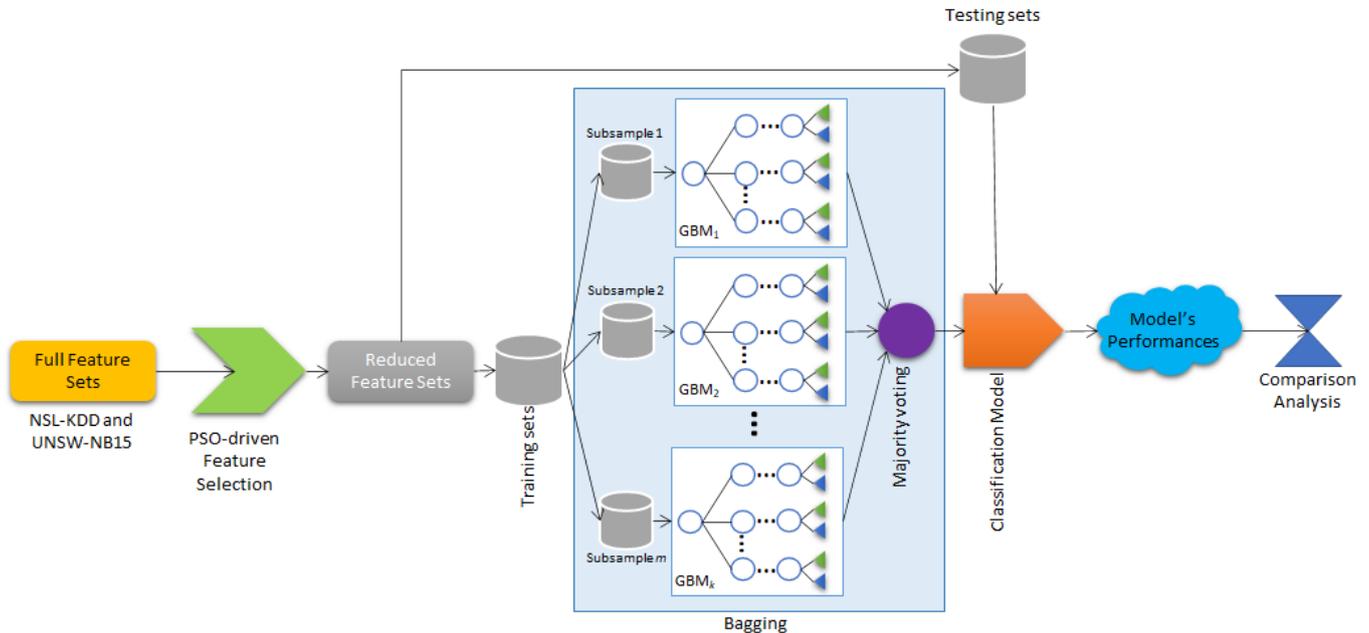

Figure 1. Proposed framework for intrusion detection based on PSO-driven feature selection and hybrid ensemble.

A PSO-driven feature selection technique is applied to identify the optimum feature subsets. Next, each dataset with an optimal feature subset is split into a training set and a testing set, where the training set is used to construct a classification model (e.g., a bagging–GBM model), and the testing set is used to validate the model's performance. Finally, different combinations of ensemble methods are statistically assessed and contrasted, along with a comparison study with prior works. In the following section, we break down the datasets used in our study, as well as the concept of our anomaly-based IDS.

3.1. Data Sets

In this study, we focus on using three distinct datasets, namely, NSL-KDD [37], UNSW-NB15 [38], and CICIDS-2017 [39]. Both datasets are extensively used for appraising IDS models and have been considered as standard benchmark datasets. The NSL-KDD dataset is an enhanced variant of its earlier versions, KDD Cup 99, which was the subject of widespread debate due to data redundancy, performance bias for machine learning algorithms, and unrealistic representation of attacks. We use an original training set of NSL-KDD (e.g., KDDTrain) that contains seven categorical input features and 34 numerical input features. There are a total of 25,192 samples, which are assigned as follows: 13,449 normal samples and 11,743 attack samples.

Furthermore, two independent testing sets (e.g., KDDTest-21 and KDDTest+) are used to appraise our proposed anomaly detector. KDDTest-21 and KDDTest+ consist of 11,850 samples and 22,544 samples, respectively. On the other hand, the UNSW-NB15 dataset also contains two primary sets, i.e., UNSW-NB15-Train and UNSW-NB15-Test, which are used for training and evaluating the model, respectively. The UNSW-NB15-Train includes six categorical input features and 38 numerical input features. There are a total of 82,332 samples, 45,332 of which are attack samples and 37,000 of which are normal samples. The UNSW-NB15-Test possesses a total of 175,341 samples, including 119,341 attack samples and 56,000 normal samples. The original version of the CICIDS-2017 dataset consists of 78 numerical input features and 170,366 samples, of which 168,186 are benign and 2180

are malicious. Given that the CICIDS-2017 does not provide predetermined training and testing sets, we employ holdout with a ratio of 80/20 for training and testing, respectively. Therefore, the CICIDS-2017 training set includes 136,293 instances that are proportionally sampled from the original dataset. The characteristics of the training datasets are outlined in Table 2.

Table 2. Description of training data sets.

Dataset	#Total Samples	#Samples Labelled Normal	#Samples Labelled Anomaly	#Categorical Features	#Numerical Features
NSL-KDD	25,192	13,449	11,743	7	34
UNSW-NB15	82,332	37,000	45,332	6	38
CICIDS-2017	136,292	134,548	1744	-	78

3.2. Methods

3.2.1. PSO-Based Feature Selection

A feature selection approach is a strategy for determining a granular, concise, and plausible subset of a particular set of features. In this work, we pick a correlation-based feature selection (CFS) method [40] that measures the significance of features using entropy and information gain. At the same time, a particle swarm optimization (PSO) algorithm [41] is taken into account as a search technique. A particle swarm optimization (PSO)-based feature selection approach models a feature set as a collection of particles that make up a swarm. A number of particles are scattered across a hyperspace and each of those particles is given a position ξ_n and velocity v_n, which are entirely random. Let $\mathbf{w}$ represents the inertia weight constant, and δ_1 and δ_2 represent the cognitive and social learning constants, respectively. Next, let σ_1 and σ_2 denote the random numbers, $\mathbf{l}_n$ denote the personal best location of particle n, and $\mathbf{g}$ denote the global location across the particles. The following are thus the basic rules for updating the position and velocity of each particle:

$$\xi_n(t+1) = \xi_n(t) + v_n(t+1) \tag{1}$$

$$v_n(t+1) = \mathbf{w}v_n(t) + \delta_1\sigma_1(\mathbf{l}_n - \xi_n(t)) + \delta_2\sigma_2(\mathbf{g} - \xi_n(t)) \tag{2}$$

3.2.2. Hybrid Ensemble Based on Bagging-GBM

The proposed hybrid ensemble is constructed based on a fusion of two individual ensemble learners, i.e., bagging [9] and gradient boosting machine (GBM) [8]. In lieu of training a bagging ensemble with a weak classifier, we employ another ensemble, e.g., GBM, as the base classifier of bagging. A bagging strategy is devised using $\mathcal{K}$ GBMs built from bootstrap replicates β of the training set. A training set containing π instances will be used to generate subsamples by sampling with replacement. Some peculiar instances appear several times in the subsamples, but others do not. Each individual GBM can then be trained on each subsample. Final class prediction is determined by the majority voting rule (e.g., each voter may only choose a single class label, and the class label prediction that gathers more than fifty percent of the most votes is chosen). We present a more formal way description of bagging–GBM in Algorithm 1.

Algorithm 1: A procedure to construct bagging–GBM for anomaly-based IDS.

Building classification model:

Require: Training set $D = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_\alpha, y_\alpha)\}$; base classifier (e.g., GBM); number of GBMs $\mathcal{K}$; size of subsample γ.

1. $\kappa \leftarrow 1$
2. **repeat**
3. $D_\kappa \leftarrow$ replacement-based subsample of γ instances from D.
4.   Construct classifier h_κ using GBM on D_κ.
5. $\kappa \leftarrow \kappa + 1$
6. **until** $\kappa > \mathcal{K}$

Evaluating classification model:

Require: An object deserving of a classification $\mathbf{x}$.

Output: Final class label prediction τ

1. $Counter_1, \ldots, Counter_y \leftarrow 0$
2. **for** $i = 1$ to $\mathcal{K}$ **do**
3. $vote_i \leftarrow h_i(\mathbf{x})$
4. $Counter_{vote_i} \leftarrow Counter_{vote_i} + 1$
5. **end for**
6. $\tau \leftarrow$ the most prevalent class label chosen by constituents.
7. Return τ

3.2.3. Evaluation Criteria

3.3. Metrics

The objective of a performance evaluation is to ensure that the proposed model works correctly with the IDS datasets. In addition, such an assessment seeks specific criteria so that the effectiveness of the proposed model can be better justified. As an anomaly-based IDS is a binary classification problem, we utilize various performance indicators that are relevant to the task, such as accuracy (Acc), precision, recall, balanced accuracy (BAcc), AUC, F1, and MCC. It is important to note that various metrics have been applied in prior research, except for BAcc and MCC, which have not been widely utilized. Balanced accuracy shows benefits over general accuracy as a metric [42], while MCC is a reliable measure that describes the classification algorithm in a single value, assuming that anomalous and normal samples are of equal merit [43]. More precisely, BAcc is specified as the arithmetic mean of the true positive rate (TPR) and true negative rate (TNR) as follows.

$$\text{BAcc} = \frac{1}{2} \times (TPR + TNR) \tag{3}$$

MCC assesses the strength of the relationship between the actual classes a and predicted labels p:

$$\text{MCC} = \frac{\text{Cov}(a, p)}{\sigma_a \times \sigma_p} \tag{4}$$

where $\text{Cov}(a, p)$ is the covariance between the actual classes a and predicted labels p, while σ_a and σ_p are the standard deviations of the actual classes a and predicted labels p, respectively.

3.4. Validation Procedure

As stated in Section 3.1, except for the CICIDS-2017 dataset, each intrusion dataset was built with a predefined split between training and testing sets. As a result, we utilized such a training/testing split (e.g., hold-out) as a validation strategy in the experiment. The hold-out procedure was repeated five times for each classification algorithm to verify that the performance results were not achieved by chance. The final performance value was calculated by averaging the five performance values.

4. Results and Discussion

The experimental assessment of the proposed framework is presented and discussed in this section. The final subsets of the NSL-KDD and UNSW-NB15 derived by PSO-based feature selection are taken from our earlier solutions reported in [6,7]. Here, 38 optimal features from the NSL-KDD and 20 optimal features from the UNSW-NB15 were employed, respectively. In contrast, the proposed feature selection identifies 17 optimal features from the original CICIDS-2017 dataset.

Furthermore, we appraised the potency of the proposed model under several ensemble strategies corresponding to different ensemble sizes. The size of the ensemble was determined by the number of base classifiers (e.g., GBM in our example) used to train the ensemble (e.g., bagging in our case). For instance, GBM-2 indicates that two GBMs were included when training the bagging ensemble, and so on. The experiment was conducted on a Linux operating system, 32 GB, and Intel Core i5 using the *R* program. Figure 2 shows the performance average with five times of hold-out for each ensemble strategy. The plot also depicts the performance of the base classifier as a standalone classifier. Taking AUC, F1, and MCC metrics as examples, the proposed model surpasses the individual classifier in all datasets considered by a substantial margin.
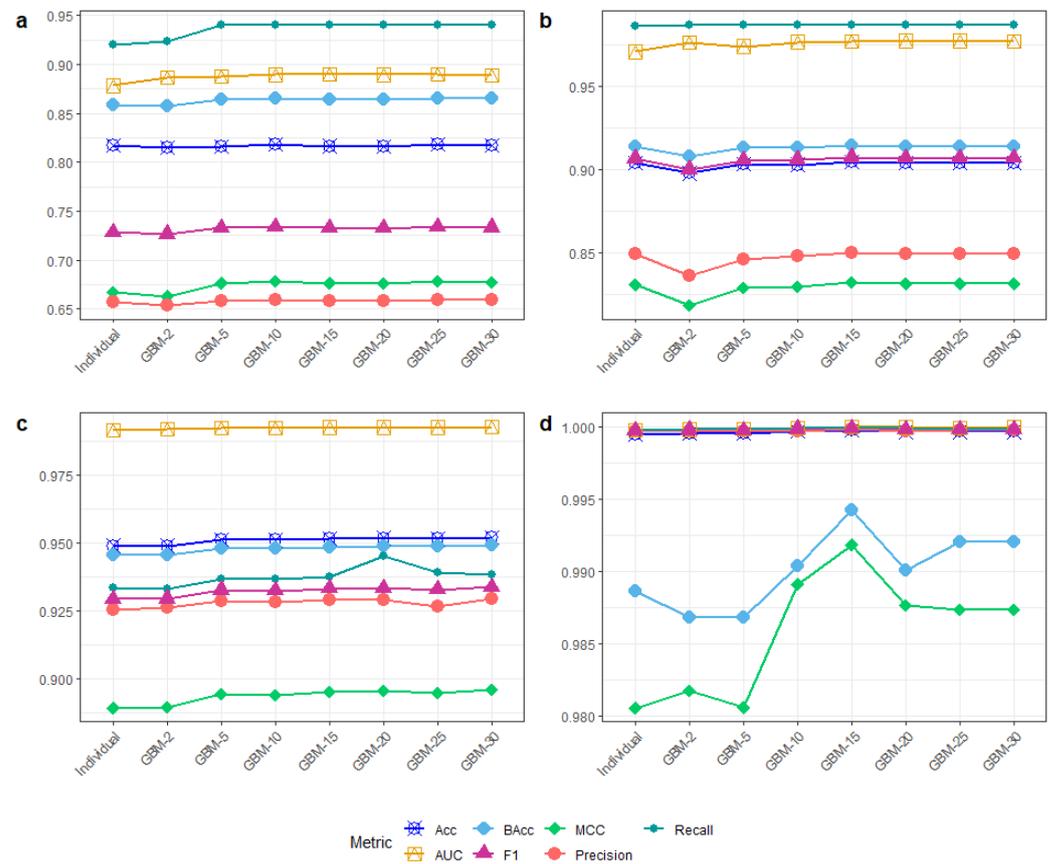

Figure 2. Performance average of all classification algorithms on KDDTest-21 (**a**), KDDTest+ (**b**), UNSW-NB15-Test (**c**), and CICIDS-2017 (**d**).

We next analyzed the performance difference of all algorithms using statistical significance tests. Here, we adopted two statistical omnibus tests, namely the Friedman test and the Nemenyi posthoc test [44]. Performance differences across classification algorithms were calculated by Friedman rank, as illustrated in Table 3. Each algorithm was given a rank for each dataset based on the MCC score, and the average rank of each algorithm was then determined. Table 3 demonstrates that bagging with 30 GBMs (e.g., GBM-30) was the

top-performing algorithm, followed by GBM-15. Interestingly, GBM-2 was the weakest performer, failing to outperform a standalone GBM model.

Table 3. Friedman rank matrix of all classifiers relative to each dataset with respect to MCC metric. Bold indicates the best rank, while the second best is underlined. The Friedman test indicates that performance differences across algorithms are significant (p-value < 0.05).

Dataset	GBM-10	GBM-15	GBM-2	GBM-20	GBM-25	GBM-30	GBM-5	Individual
CICIDS-2017	2	1	6	3	4	5	7	8
KDDTest-21	2	4	8	6	1	3	5	7
KDDTest+	6	1	8	3	4	2	7	5
UNSW-NB15-Test	6	3	7	2	4	1	5	8
Average rank	4.00	<u>2.25</u>	7.25	3.50	3.25	**2.75**	6.00	7.00
p-value				0.01197				

The Nemenyi test employs the Friedman rank; if such average differences are more than or equal to a critical difference (CD), then the performances of such algorithms are substantially different. Figure 3 illustrates that there are no significant performance differences across the benchmarked algorithms, as no average rank exceeds the critical difference (CD) of the Nemenyi test. As shown by a horizontal line, all algorithms are linked. As a final comparison, our best-proposed model (e.g., GBM-30) is compared against existing solutions for anomaly-based IDS. We contrast the efficacy of our proposed scheme to those with a comparative validation approach (e.g., hold-out using predetermined training/test sets).
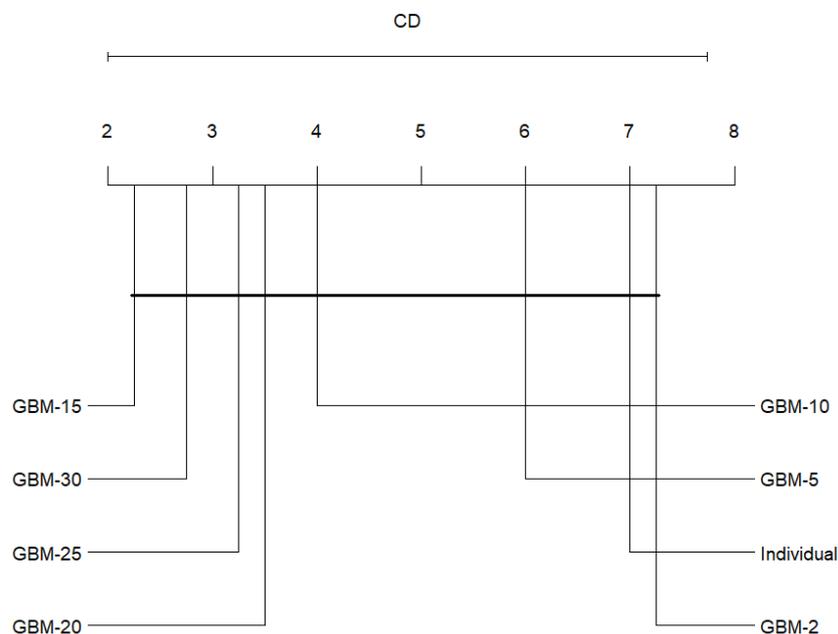

Figure 3. Critical difference plot based on Nemenyi test with respect to MCC metric. Critical difference (CD) is at 5.74, which exceeds the average rank, while all classifiers are tied altogether.

Table 4 compares the performance of our proposed model (e.g., GBM-30) against that of a variety of existing studies published in the latest scientific literature. The proposed model achieves the highest FPR, recall, AUC, and F1 metrics on KDDTest+. Nonetheless, compared to [45], there are minor variations in accuracy and precision measures. Except for the precision metric, our proposed model is the best performer on the KDDTest-21 across all performance criteria. Similarly, on UNSW-NB15-Test and CICIDS-2017, our proposed model outperforms all other models in all performance measures except the FPR metric.

In general, our proposed model is shown to be a feasible solution for anomaly-based IDS, at least for the public datasets addressed in this study. Specifically, with respect to the lowering of FPR and increasing recall, AUC, and F1 scores, our suggested model has shown a significant improvement over the existing studies. In addition, we show the computational time required for individual GBM as well as GBM-15 on the reduced and full feature sets for each dataset in Figure 4. Our feature selection technique significantly lessens the training and testing complexity by roughly one-third compared to the complete feature set, particularly when large datasets such as CICIDS-2017 and UNSW-NB15 are employed.

Table 4. Comparison of the proposed model's outcomes to that of previous network anomaly detectors. Bold indicates the best values.

Ref.	Method	Feature Selection	Acc (%)	FPR (%)	Precision (%)	Recall (%)	AUC	F1
KDDTest+								
[45]	Stacking	-	**92.17**	2.52	-	-	-	-
[46]	Autoencoder	-	84.21	-	-	87.00	-	-
[23]	LightGBM	-	89.79	9.13	-	-	-	-
[26]	MFFSEM	RF	84.33	24.82	74.61	97.15	-	0.841
[28]	Weighted majority voting	GR, IG, and χ^2	85.23	12.8	**90.3**	-	-	0.855
This study	Hybrid ensemble	PSO	90.39	**1.59**	84.94	**98.68**	**0.9767**	**0.907**
KDDTest-21								
[47]	Voting ensemble	CFS-BA	73.57	12.92	73.6	-	-	-
This study	Hybrid ensemble	PSO	**81.72**	**2.1**	65.87	**94.00**	**0.8886**	**0.7332**
UNSW-NB15-Test								
[45]	Stacking	-	92.45	11.3	-	-	-	-
[26]	MFFSEM	RF	88.85	**2.27**	-	80.44	-	-
[20]	RF	GR, χ^2, and PC	83.12	3.7	-	-	-	-
[23]	LightGBM	-	85.89	14.79	-	-	-	-
[30]	LightGBM	DNN	88.34	12.46	-	-	-	0.881
This study	Hybrid ensemble	PSO	**95.20**	4.03	**92.93**	**93.84**	**0.9925**	**0.9338**
CICIDS-2017								
[48]	Rough set theory + Bayes	FPE	97.95	-	-	96.37	-	0.9637
[21]	Stacking	*K*-Means	98.0	**0.2**	97.0	98.0	-	0.98
[49]	ICVAE-BSM	-	99.86	-	99.68	99.68	-	0.9968
This study	Hybrid ensemble	PSO	**99.98**	2.6	**99.99**	**99.99**	**1.00**	**0.9998**

Lastly, we discuss two main implications of our study as follows. First, most previous comparisons were made on particular performance metrics. Our work, however, aims to examine a more trustworthy metric (e.g., MCC) that creates more accurate estimates for the proposed model [43]. The MCC measure could be used to judge future work, especially for detecting network anomalies. Second, a strategy for detecting intrusions should ideally have a low proportion of false positives. Unfortunately, it is nearly impossible to prevent false positives in network anomaly detection. Our work, however, produces the lowest false positive rate on the NSL-KDD dataset and fair results on the UNSW-NB15 and CICIDS-2017.
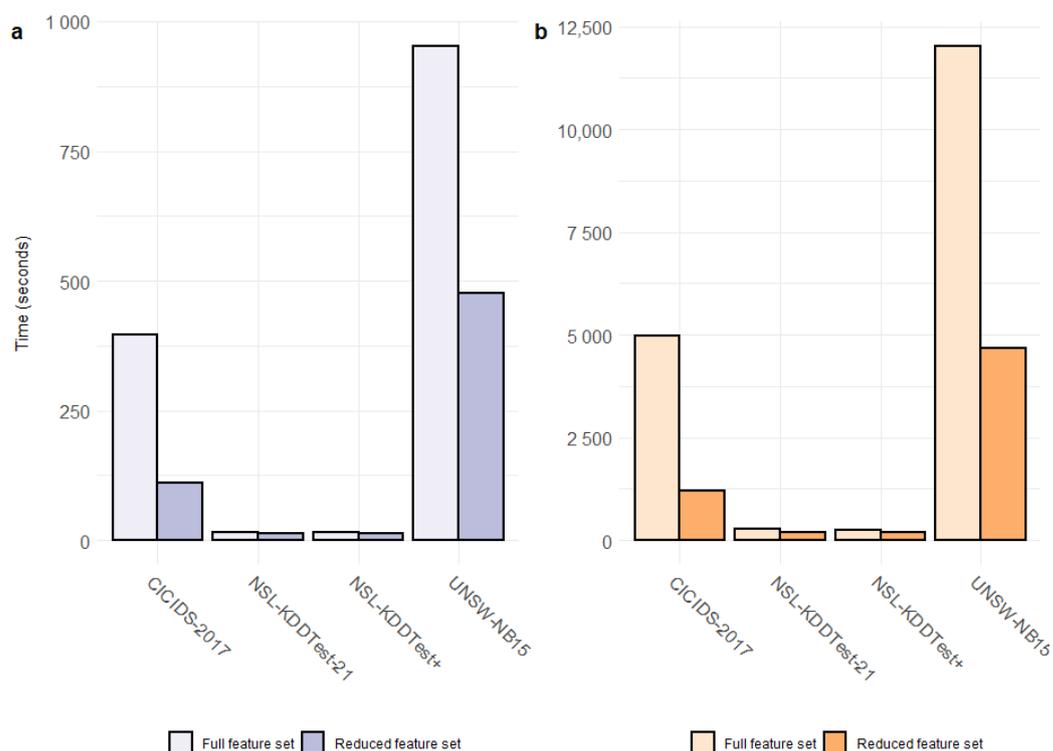
Figure 4. Training and testing complexity for individual GBM (**a**) and GBM-15 (**b**) on reduced and complete feature sets for each data set.

5. Conclusions

An anomaly-based intrusion detection system (IDS) was proposed to thwart any malicious attack and was recognized as a viable method for detecting novel attacks. This work investigated a novel anomaly-based intrusion detection system (IDS) strategy that combines particle swarm optimization (PSO)-guided feature selection with a hybrid ensemble approach. The reduced feature subset was utilized as input for the hybrid ensemble, which was a combination of two well-known ensemble paradigms, including bootstrap aggregation (Bagging) and gradient boosting machine (GBM). The proposed model revealed a substantial performance gain compared to existing studies using the NSL-KDD, UNSW-NB15, and CICIDS-2017 datasets. More specifically, our anomaly detector achieved the lowest FPR at 1.59% and 2.1% on KDDTest+ and KDDTest-21, respectively. With respect to the accuracy, recall, AUC, and F1 metrics, our proposed model consistently surpassed previous research across all datasets considered.

List of Acronyms

AB	Adaboost
AUC	Area Under ROC Curve
BA	Bat Algorithm
CFS	Correlation-based Feature Selection
CV	Cross Validation
DL	Deep Learning
DNN	Deep Neural Network
DT	Decision Tree
FPE	Feature Probability Estimation
GA	Genetic Algorithm
GR	Gain Ratio
HGB	Histogram-based Gradient Boosting
IBL	Instance-based Learning
IFA	Improved Firefly Algorithm
IG	Information Gain
LR	Logistic Regression
MCC	Matthew Correlation Coefficient
MLP	Multilayer Perceptron
NB	Naive Bayes
NN	Neural Network
PC	Pearson Correlation
PCA	Principle Component Analysis
RF	Random Forest
SVM	Support Vector Machine

References

1. Ghorbani, A.A.; Lu, W.; Tavallaee, M. *Network Intrusion Detection and Prevention: Concepts and Techniques*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2009; Volume 47.
2. Bhattacharyya, D.K.; Kalita, J.K. *Network Anomaly Detection: A Machine Learning Perspective*; CRC Press: Boca Raton, FL, USA, 2013.
3. Thakkar, A.; Lohiya, R. A review on machine learning and deep learning perspectives of IDS for IoT: Recent updates, security issues, and challenges. *Arch. Comput. Methods Eng.* **2021**, *28*, 3211–3243. [CrossRef]
4. Rokach, L. *Pattern Classification Using Ensemble Methods*; World Scientific: Singapore, 2010; Volume 75.
5. Tama, B.A.; Lim, S. Ensemble learning for intrusion detection systems: A systematic mapping study and cross-benchmark evaluation. *Comput. Sci. Rev.* **2021**, *39*, 100357. [CrossRef]
6. Tama, B.A.; Rhee, K.H. HFSTE: Hybrid Feature Selections and Tree-Based Classifiers Ensemble for Intrusion Detection System. *IEICE Trans. Inf. Syst.* **2017**, *100D*, 1729–1737. [CrossRef]
7. Tama, B.A.; Comuzzi, M.; Rhee, K.H. TSE-IDS: A two-stage classifier ensemble for intelligent anomaly-based intrusion detection system. *IEEE Access* **2019**, *7*, 94497–94507. [CrossRef]
8. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
9. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
10. Khraisat, A.; Gondal, I.; Vamplew, P.; Kamruzzaman, J. Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity* **2019**, *2*, 20. [CrossRef]
11. Resende, P.A.A.; Drummond, A.C. A Survey of Random Forest Based Methods for Intrusion Detection Systems. *ACM Comput. Surv.* **2018**, *51*, 1–36. [CrossRef]
12. Aburomman, A.A.; Reaz, M.B.I. A survey of intrusion detection systems based on ensemble and hybrid classifiers. *Comput. Secur.* **2017**, *65*, 135–152. [CrossRef]
13. Thakkar, A.; Lohiya, R. A survey on intrusion detection system: Feature selection, model, performance measures, application perspective, challenges, and future research directions. *Artif. Intell. Rev.* **2022**, *55*, 453–563. [CrossRef]
14. Lohiya, R.; Thakkar, A. Application domains, evaluation data sets, and research challenges of IoT: A Systematic Review. *IEEE Internet Things J.* **2020**, *8*, 8774–8798. [CrossRef]
15. Thakkar, A.; Lohiya, R. A review of the advancement in intrusion detection datasets. *Procedia Comput. Sci.* **2020**, *167*, 636–645. [CrossRef]
16. Jafarian, T.; Masdari, M.; Ghaffari, A.; Majidzadeh, K. Security anomaly detection in software-defined networking based on a prediction technique. *Int. J. Commun. Syst.* **2020**, *33*, e4524. [CrossRef]
17. Kaur, G. A comparison of two hybrid ensemble techniques for network anomaly detection in spark distributed environment. *J. Inf. Secur. Appl.* **2020**, *55*, 102601. [CrossRef]

18. Seth, S.; Chahal, K.K.; Singh, G. A novel ensemble framework for an intelligent intrusion detection system. *IEEE Access* **2021**, *9*, 138451–138467. [CrossRef]

19. Halim, Z.; Yousaf, M.N.; Waqas, M.; Sulaiman, M.; Abbas, G.; Hussain, M.; Ahmad, I.; Hanif, M. An effective genetic algorithm-based feature selection method for intrusion detection systems. *Comput. Secur.* **2021**, *110*, 102448. [CrossRef]

20. Nazir, A.; Khan, R.A. A novel combinatorial optimization based feature selection method for network intrusion detection. *Comput. Secur.* **2021**, *102*, 102164. [CrossRef]

21. Jain, M.; Kaur, G. Distributed anomaly detection using concept drift detection based hybrid ensemble techniques in streamed network data. *Clust. Comput.* **2021**, *24*, 2099–2114. [CrossRef]

22. Krishnaveni, S.; Sivamohan, S.; Sridhar, S.; Prabakaran, S. Efficient feature selection and classification through ensemble method for network intrusion detection on cloud computing. *Clust. Comput.* **2021**, *24*, 1761–1779. [CrossRef]

23. Liu, J.; Gao, Y.; Hu, F. A fast network intrusion detection system using adaptive synthetic oversampling and LightGBM. *Comput. Secur.* **2021**, *106*, 102289. [CrossRef]

24. Al, S.; Dener, M. STL-HDL: A new hybrid network intrusion detection system for imbalanced dataset on big data environment. *Comput. Secur.* **2021**, *110*, 102435. [CrossRef]

25. Tian, Q.; Han, D.; Hsieh, M.Y.; Li, K.C.; Castiglione, A. A two-stage intrusion detection approach for software-defined IoT networks. *Soft Comput.* **2021**, *25*, 10935–10951. [CrossRef]

26. Zhang, H.; Li, J.L.; Liu, X.M.; Dong, C. Multi-dimensional feature fusion and stacking ensemble mechanism for network intrusion detection. *Future Gener. Comput. Syst.* **2021**, *122*, 130–143. [CrossRef]

27. Gupta, N.; Jindal, V.; Bedi, P. CSE-IDS: Using cost-sensitive deep learning and ensemble algorithms to handle class imbalance in network-based intrusion detection systems. *Comput. Secur.* **2022**, *112*, 102499. [CrossRef]

28. Krishnaveni, S.; Sivamohan, S.; Sridhar, S.; Prabhakaran, S. Network intrusion detection based on ensemble classification and feature selection method for cloud computing. *Concurr. Comput. Pract. Exp.* **2022**, *34*, e6838. [CrossRef]

29. Rashid, M.; Kamruzzaman, J.; Imam, T.; Wibowo, S.; Gordon, S. A tree-based stacking ensemble technique with feature selection for network intrusion detection. *Appl. Intell.* **2022**, *52*, 9768–9781. [CrossRef]

30. Wang, Z.; Liu, J.; Sun, L. EFS-DNN: An Ensemble Feature Selection-Based Deep Learning Approach to Network Intrusion Detection System. *Secur. Commun. Netw.* **2022**, *2022*, 2693948. [CrossRef]

31. Wolpert, D.H. Stacked generalization. *Neural Netw.* **1992**, *5*, 241–259. [CrossRef]

32. Ting, K.M.; Witten, I.H. Issues in stacked generalization. *J. Artif. Intell. Res.* **1999**, *10*, 271–289. [CrossRef]

33. Schapire, R.E.; Freund, Y. Boosting: Foundations and algorithms. *Kybernetes* **2013**, *42*, 164–166. [CrossRef]

34. Thakkar, A.; Lohiya, R. Fusion of statistical importance for feature selection in Deep Neural Network-based Intrusion Detection System. *Inf. Fusion* **2022**, *90*, 353–363. [CrossRef]

35. Thakkar, A.; Lohiya, R. Attack classification using feature selection techniques: A comparative study. *J. Ambient Intell. Humaniz. Comput.* **2021**, *12*, 1249–1266. [CrossRef]

36. Thakkar, A.; Lohiya, R. Role of swarm and evolutionary algorithms for intrusion detection system: A survey. *Swarm Evol. Comput.* **2020**, *53*, 100631. [CrossRef]

37. Tavallaee, M.; Bagheri, E.; Lu, W.; Ghorbani, A.A. A detailed analysis of the KDD CUP 99 data set. In Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications, Ottawa, ON, Canada, 8–10 July 2009; pp. 1–6.

38. Moustafa, N.; Slay, J. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In Proceedings of the 2015 Military Communications and Information Systems Conference (MilCIS), Canberra, Australia, 10–12 November 2015; pp. 1–6.

39. Sharafaldin, I.; Lashkari, A.H.; Ghorbani, A.A. Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSp* **2018**, *1*, 108–116.

40. Hall, M.A. Correlation-Based Feature Selection for Machine Learning. Ph.D. Thesis, The University of Waikato, Hamilton, New Zealand, 1999.

41. Kennedy, J.; Eberhart, R.C. A discrete binary version of the particle swarm algorithm. In Proceedings of the 1997 IEEE International Conference on Systems, Man, and Cybernetics, Computational Cybernetics and Simulation, Orlando, FL, USA, 12–15 October 1997; Volume 5, pp. 4104–4108.

42. Brodersen, K.H.; Ong, C.S.; Stephan, K.E.; Buhmann, J.M. The balanced accuracy and its posterior distribution. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 3121–3124.

43. Chicco, D.; Tötsch, N.; Jurman, G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData Min.* **2021**, *14*, 13. [CrossRef] [PubMed]

44. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.

45. Tama, B.A.; Nkenyereye, L.; Islam, S.R.; Kwak, K.S. An Enhanced Anomaly Detection in Web Traffic Using a Stack of Classifier Ensemble. *IEEE Access* **2020**, *8*, 24120–24134. [CrossRef]

46. Ieracitano, C.; Adeel, A.; Morabito, F.C.; Hussain, A. A novel statistical analysis and autoencoder driven intelligent intrusion detection approach. *Neurocomputing* **2020**, *387*, 51–62. [CrossRef]

47. Zhou, Y.; Cheng, G.; Jiang, S.; Dai, M. Building an efficient intrusion detection system based on feature selection and ensemble classifier. *Comput. Netw.* **2020**, *174*, 107247. [CrossRef]

48.   Prasad, M.; Tripathi, S.; Dahal, K. An efficient feature selection based Bayesian and Rough set approach for intrusion detection. *Appl. Soft Comput.* **2020**, *87*, 105980. [CrossRef]

49.   Zhang, Y.; Liu, Q. On IoT intrusion detection based on data augmentation for enhancing learning on unbalanced samples. *Future Gener. Comput. Syst.* **2022**, *133*, 213–227. [CrossRef]