Improving the Anomaly Detection by Combining PSO Search Methods and J48 Algorithm

Kurniabudi
Department of Computer Engineering
Universitas Dinamika Bangsa
& Faculty of Engineering
Universitas Sriwijaya
Indonesia
kbudiz@yahoo.com

Abdul Harris
Department of Computer Sciences
line 3: *Faculty of Computer Sciences*
Universitas Dinamika Bangsa
Indonesia
abdulharris@stikom-db.ac.id

Albertus Edward Mintaria
Department of Computer Engineering
Faculty of Computer Sciences
Universitas Dinamika Bangsa
Indonesia
albertusedwardmintaria@rocketmail.com

Darmawijoyo
Faculty of Mathematics and Natural Sciences
Universitas Sriwijaya
Indonesia
darmawijoyo@yahoo.com

Deris Stiawan
Faculty of Computer Sciences
Universitas Sriwijaya
Indonesia
deris.stiawan@gmail.com

Mohd Yazid bin Idris
Department of Computing
Universiti Teknologi Malaysia
Indonesia
yazid@utm.my

Rahmat Budiarto
College of Computer Science & IT
Albaha University
Saudi Arabia
rahmat@bu.edu.sa

***Abstract*—The feature selection techniques are used to find the most important and relevant features in a dataset. Therefore, in this study feature selection technique was used to improve the performance of Anomaly Detection. Many feature selection techniques have been developed and implemented on the NSL-KDD dataset. However, with the rapid growth of traffic on a network where more applications, devices, and protocols participate, the traffic data is complex and heterogeneous contribute to security issues. This makes the NSL-KDD dataset no longer reliable for it. The detection model must also be able to recognize the type of novel attack on complex network datasets. So, a robust analysis technique for a more complex and larger dataset is required, to overcome the increase of security issues in a big data network. This study proposes particle swarm optimization (PSO) Search methods as a feature selection method. As contribute to feature analysis knowledge, In the experiment a combination of particle swarm optimization (PSO) Search methods with other search methods are examined. To overcome the limitation NSL-KDD dataset, in the experiments the CICIDS2017 dataset used. To validate the selected features from the proposed technique J48 classification algorithm used in this study. The detection performance of the combination PSO Search method with J48 examined and compare with other feature selection and previous study. The proposed technique successfully finds the important features of the dataset, which improve detection performance with 99.89% accuracy. Compared with the previous study the proposed technique has better accuracy, TPR, and FPR.**

Keywords—Feature Selection, Anomaly Detection, CICIDS2017, Correlation-Based, PSO Search

I. INTRODUCTION

More applications, protocols and devices connected on data networks , will produce a huge and heterogeneous data which in turn contributes towards the increase of dimensionality of data. The increase of dimensionality of data causing more challenges in data analysis, specifically in a case of intrusion detection system (IDS) or anomaly detection system. As mentioned in [1] Analyzing network traffic and selecting relevant features is one of the challenges of handling large volumes of network data. The same problem has been stated in [2]. The increase of dimensionality of data will challenge the feature selection and feature extraction methods.

Feature selection is considered as a method for dimensional reduction [3]. Many methods of feature selection have been developed and implemented. In [4], the information gain applied as a feature selection technique, from 41 features only 16 features have a good effect on classification algorithm. Research in [5] applies information gain on the NSL-KDD dataset, by selecting features with IG over 0.40 from 41 features is reduced to 8 features. Another researcher in [6] also implement an information gain technique on 41 features of the NSL-KDD dataset and resulted in 8 features. The experiment results on identification/classification of five attack classes/types show that accuracy of class Normal=99.7%, DoS=99.9%, Probe=96.5%, U2R 99.4% and URL=98.0%. Previous studies have shown that information gain techniques provide relatively good performance in selecting the best and relevant features that contribute to higher accuracy of anomaly detection. However, with the information gain and the gain ratio technique, user knowledge is required to determine the minimum merit value (IG value). This minimum merit will affect the number of selected features. On other hand, the downside of the NSL-KDD dataset is, it only has 41 features. Thus, a larger dataset is required to test IDS that need more features to increase its accuracy. This study overcomes the disadvantages of information gain and gain ratio techniques by using correlation-based feature selection techniques.

With the fast-growing of traffic flow in the modern network. more huge data will be produced. Therefore, the NSL-KDD dataset will no longer able to represent real-world traffic models. To overcome this problem, [7] creates a dataset namely CICIDS2017. The CICIDS 2017 is designed to represent Real-world network traffic data[8]. This circumstance motivates the researchers in this paper to utilize the dataset. In the anomaly or attack detection systems do not

use all features, instead, only a few relevant features are required by them. Thus, the challenge is to choose the best and important features in a complex dataset such as the CICIDS2017 dataset that can optimize the detection accuracy as well as the processing time. To select an important and relevant feature, experimentations are required. This paper evaluates Correlation-based Feature Selection (CFS) techniques with different search methods and its capability to search relevant features on a huge dataset. CFS is the most simple way of finding important features[9]. CFS eliminates all the features that are redundant and do not have a correlation between them that significantly improves execution speed[10]. There are 78 features in the CICIDS2017 dataset. Not all of these features are important and relevant for classifying traffic data. The main objective of this research is to analyze and design an ideal feature selection technique that can produce an important and relevant feature, which can improve the anomaly detection performance. To achieve the research objective, a novel feature selection techniques are proposed. In this study, PSO search method proposes as a feature selection technique and validates with the J48 classification algorithm. In contrast to Information Gain as previously described, the PSO search method is able to analyze and generate relevant features without user intervention. In the experiment, a combination of PSO Search methods with other search methods in correlation-based feature selection was also tested and compared. This to give a strong knowledge of the performance of correlation-based feature selection techniques in feature analysis. For validation purposes, the classification algorithm J48 is used for the detection type of traffic on the dataset. This traffic then identifies as benign or attack traffic.

This article is structured in five sections. Section I gives an introduction and research background. Section II describes a conceptual theory on feature selection method/technique and related works. Section III presents the details of the research methodology including the experimental setup and the proposed techniques. Section IV presents experimental results on the proposed feature selection techniques as well as its performance analysis and Section V concludes the work.

II. Conceptual Theory And Related Work

A. Feature Selection Method

As mentioned in [11] and supported by researchers in [12], one method for dimensional reduction of the dataset is feature selection. Feature selection can be applied to a dataset to select relevant and important features. According to [13] the feature selection process consists of four steps as follows subset generation, subset evaluation, stopping criterion, and result validation. The subset generation is a heuristic step with a search procedure for selecting a new feature subset as a candidate for evaluation. For each new subset generated will be evaluated by evaluation criteria. Stopping criterion will be fulfilled if one of the search processes in the generation of subset creation is complete, or the merit of the selected subset of features is acceptable. The final step is result validation, used as empirical evidence of the selected feature set.

The well-known feature selection methods are: filtered, wrapper, and embedded methods[14]. The filtered method uses ranking as criteria for determining relevant features. A suitable ranking criterion is used to score features or variables. A threshold value is then set to eliminate features having a score below the threshold value. Some examples of filtered methods include Correlation-based Feature Selection (CFS), Markov Blanket Filter, Fast Correlation Based Feature Selection (FCFS), etc. The wrapper method uses a predictor as a black box and the predictor's performance is determined as an objective function to evaluate the variable subset. The wrapper method includes Sequential Forward Selection, Sequential Backward Selection, Genetic Algorithm (GA), Simulated Annealing, Randomized Hill Climbing, etc. The embedded method incorporates feature selection as a part of the training process as the main approach to reduce computation overhead. Examples of the embedded method include Decision Tree, Random Forest, Naive Bayes, Support Vector Machine (SVM), methods based on regularization techniques, etc.

B. Related Work

Some researchers have carried out studies on feature selection techniques with different datasets and techniques or methods. For example, work in [15] propose a filtered-based feature selection algorithm namely Flexible Mutual Information Feature Selection (FMIFS) and combine with Least Square Support Vector Machine based IDS (LSSVM-IDS). This method implemented on the KDD CUP'99, NSL-KDD, and Kyoto 2006+ dataset. The proposed method has better accuracy and computational cost than the state-of-the-art methods. While research in [16] using a Genetic Algorithm as feature selection techniques. This method combines with the Support Vector Machine (SVM). The proposed method implemented on KDD CUP 99 and UNSW-NB15 dataset achieve high accuracy and low FPR value when implemented. Research in [17] using the combination of Maximal Information Coeffecient (MIC) and Principle Component Analysis (PCA) to create detection algorithm. The proposed algorithm test on DARPA 1999 dataset. This algorithm namely MSPCA is effectively select features and result in high detection accuracy. Another research in [18], proposed a fusion detection algorithm. A chi-square feature selection applied on NSL-KDD to select the best feature. Next, a multiclass SVM implemented to classify a network attack. The proposed method achieve high detection accuracy and FPR lower.

Furthermore, research in [19], proposed the combination of discretized differential evolution (DDE) with C4.5 classification algorithm implemented to analyze NSL-KDD dataset. The test with the reduced feature set result in the detection rate is improved with training and testing time lower. However, the most of previous studies used the NSL-KDD dataset which only has 41 features. in other words, the NSL-KDD data set can no longer be relied on to represent today's very large and complex data traffic. As mentioned in [7] the dataset that available since 1998 (including NSL-KDD) is out of date and unreliable for investigating robustness and accuracy of anomaly detection systems that need a dataset with more numbers of features. This study proposes an ideal feature selection technique that is capable of producing relevant features for anomaly detection systems on large traffic with complex data. Therefore we need a more reliable dataset. Therefore, the CICIDS2017 dataset use for the experiments. The detail explanation on the dataset is given in Sub-section III-B.

Referring to [20] research on determining a subset of features, a search method is applied to obtain the most relevant features. Search methods can use forward selection, backward elimination, or a combination. Based on related studies the authors conducted a combination of search methods and classification algorithms to get the most relevant features that can improve the detection system.

III. Research Methodology

In this section, the experimental setup, dataset, correlation-based feature selection, classification method, and the proposed technique is described.

A. Proposed Feature Selection Method

In this paper, a new feature selection technique is proposed. The objective of the propose technique is to select the most important features from the CICIDS2017 dataset. The selected feature is then used to classify normal and attack class of traffic. To meet the objective, this paper proposes a combination of the feature selection approach. For a clear view, Figure 1 illustrates the workflow of the proposed technique.

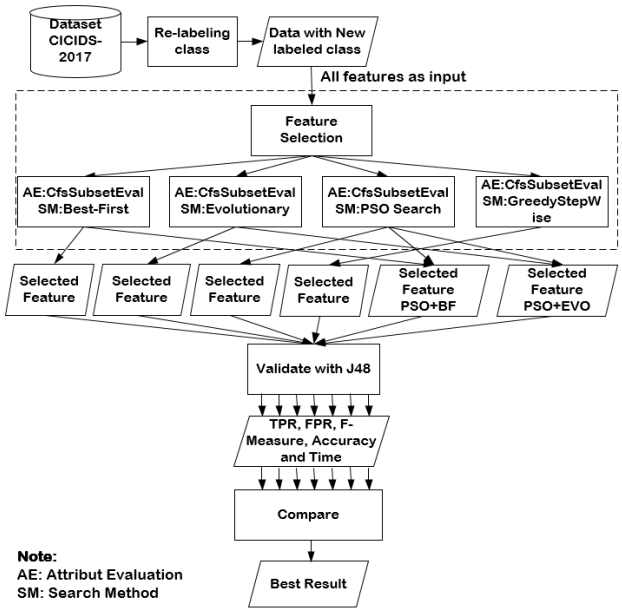

Fig. 1. The Propose Method used in the Experiment

In the initial experiment, feature analysis was carried out with a single search method such as Best-First (BF), Evolutionary, Greedy-stepwise (GS), and PSO-search and validated with the J48 classification algorithm, to test the ability to detect normal traffic and attacks with selected features. Furthermore, testing the combination of the PSO-Search search method and other search methods is carried out and the results are compared. So that it is obtained the most ideal method for detecting traffic anomalies. In this study, the attack traffic is considered as anomalous traffic. Beside propose a combination PSO Search method of correlation-based feature selection with J48 classification algorithm, a couple of combinations of the selected features from the PSO Search method combine with Best-First Search and PSO Search combine with Evolutionary Search examined and validated through a training process using the J48 classification algorithm.

B. CICIDS2017

The CICIDS2017 is proposed by researchers in [21], to overcome the limitation of publicly available IDS dataset that meet real-world network traffic criteria [7]. The CICIDS2017 is the valid dataset[22], it is the largest and most commonly used dataset[23]. The 20% of MachineLearningcsv from the CICIDS2017 dataset used in this experiment are described in Table 1. This dataset consists of 78 features and normal traffic and attack traffic. CICIDS2017 dataset is a huge volume of data and high class imbalanced. To handle class imbalanced, as suggest in [24] re-labeled class is done as shown in Table 1.

TABLE I. Class distribution on 20% of CICIDS2017 dataset

New Labels	Old Labels	Number of Instances	Fraction to Majority Class	Fraction to Total Instance
Normal	Benign	454,306	1	80.245
Bot	Bot	367	0.00081	0.065
Brute Force	FTP- Patator, SSH-Patator	2,717	0.00598	0.480
Dos/DdoS	DDoS, DoS, GoldenEye, DoS Hulk, DoS Slow, httptest, DoS slowloris, Heartbleed	76,445	0.16827	13.503
Infiltration	Infiltration	6	0.00001	0.001
Portscan	PortScan	31,882	0.07018	5.631
Web Attack	Web Attack–Brute Force, Web Attack–Sql Injection, Web Attack–XSS	426	0.00094	0.075
Total Instances		**566,149**		

C. Units

As mentioned before, this research uses Correlation-Based Feature Selection (CFS). This method is proposed by researchers in [25]. Correlation-based Feature Subset Selection is applied to calculate the merit of a subset of features with k number of features in (1).

$$Ms = R_{FC} = \frac{k_{rfc}}{\sqrt{k+k(k-1)_{rff}}} \quad (1)$$

where, RFC = correlation between the class and the features;
rfc = average value of feature-class correlation;
rff = average value of feature-class correlation;
k = number of feature in feature set start with 0 (empty)

Correlation coefficient value is the level of eligibility of the selected subset which is considered as one of the performance metrics to identify the best subset selection technique. The level of correlation must be higher than the correlation between features and class attributes and must be the lowest among its features [26]. As the number of features increases, the level of correlation between classes and features increases, because newly added features are lack of correlation to the selected features and may have good dominance over higher correlations with class [27].

The correlation-based feature selection has been widely used in many research and using vary search method, for

example : research in [28] implement Best first, Greedy Stepwise, Exhaustive search, Genetic search, Random search and Scatter search VI to select relevant feature from NSL-KDD dataset. Research in [29] using the Best-First Search method for finding an important feature on KDD CUP 99 dataset. Another research in [30], using a correlation-based feature selection technique to reduce 41 features of NSL-KDD. These techniques compare with chi-square feature selection. Correlation more effectively reduces the feature of NSL-KDD from chi-square.

D. Classification Algorithm

Classification is one of the machine learning functions. The classification method has been widely used in intrusion detection research. Many methods have been developed and applied to solve the security problem. In this research to validate the feature selection method, the J48 classification algorithm is used. J48 or knowing as the C4.5 algorithm is one of the decision tree algorithm, the pseudocode of the algorithm is shown in figure 2[31].

```
1: Create a root node N;
2: IF (T belongs to same category C)
       {leaf node = N;
         Mark N as class C;
         Return N;
       }
3: For i=1 to n
       {Calculate Information_gain  (Ai);}
4: ta= testing attribute;
5: N.ta = attribute having highest information_gain;
6: if (N.ta == continuous )
         { find threshold;}
7: For (Each T in splitting of T)
8:       if (T is empty)
           {child of N is a leaf node;}
           else
           {child of N= dtree T)}
10: calculate classification error rate of node N;
11: return N;
```

Fig. 2. C4.5 (J48) Pseudocode

E. Measurement Metrics

For performance evaluation of anomaly detection in this research, the confusion matrix as listed in table 2 is used.

TABLE II. CONFUSION MATRIX

Actual Class	Predicted Class	
	Positive	***Negative***
Positive	TP	FN
Negative	FP	TN

Entries in Table 2 are defined as follows,

- TP (True Positive) is attack class that is properly defined as an attack;
- FP (False Positive) is a normal class defined as an attack;
- TN (True Negative) is a normal class which is defined as normal;
- FN (False Negative) is a class of attack defined as normal.

By the above definitions, the performance of classification can be measured in terms TPR (True Positive Rate), FPR (False Positive Rate), Accuracy, and F-Measure using the formulas in (2) - (5).

$$TPR = \frac{TP}{(TP+FN)} \quad (2)$$

$$FPR = \frac{FP}{(FP+TN)} \quad (3)$$

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (4)$$

$$F-Measure = \frac{2*Precision*Recall}{Precision+Recall} \quad (5)$$

F. Experimental Setup

For experiment purpose, a notebook powered by Intel Core i7 processor, 2.70 GHz and 8 GB RAM with Windows 10 as Operating System is used. Software Weka3.9 with heap size 3072 MB is utilized as an analysis tool. Weka is a machine learning tool[32] and for feature selection technique, Weka Library function used in this experiment. The experiments were carried out in three stages: 1) Feature selection stage is performed for correlation-based feature selection with different search methods include CFs-BestFirst, CFs-Evolutionary, CFs-GreedyStepWise, and CFs-PSO. This stage produces selected features; 2) Next is the training stage where selected features will be trained by classification algorithm J48, and 3) Detection performance evaluation stage. In this stage, true positive rate (TPR), false-positive rate (FPR), Accuracy, and Processing Time are measured to compare the detection performance of the techniques implemented during the experiments. The 10-fold cross-validation for feature selection and classification was used in the evaluation. The reason for using Cross-validation 10-fold is that it reduces computation time while maintaining an accuracy classification algorithm[33]. In the 10-fold cross-validation test, The input dataset is randomly divided into 10 fold by the same size. From the 10 fold, 9 fold are used as training data and 1 fold section for testing data. Furthermore, this process is repeated 10 times until each fold is tested. This evaluation method has been used in research [5], [32] and [34].

IV. RESULT AND ANALYSIS

In this section, Firstly, the experiment results from the feature selection using various methods are presented in Sub-section IV-A. Sub-section IV-B presents the results of correlation-based feature selection and followed by the results of the learning process of the J48 classification algorithm. The performance analysis and discussion about the findings from the experiments are presented in Sub-section IV-C. Lastly, Sub-section IV-D discusses the comparison results.

A. Experimental Data

The CICIDS2017 dataset used in this experiment has 78 features consisting of 77 network data traffic information features and 1 feature as a label. Table 3 shows the 77 features analyzed using the proposed feature selection technique.

B. Selected Features from Proposed Method

The proposed technique is a kind of simple way of feature selection by combining selected features from Best First, PSO, and Evolutionary searches. Feature selection using GS and BS produces 6 features with exactly the same feature types. While evolutionary produces 8 features. PSO Search produces 15 features, The proposed method, PSO Search produces 15 features, PSO + BF 16 features and PSO

+ Evo produces 20 features. The list of features produced by each feature selection technique is presented in table 4.

TABLE III. LIST OF ALL FEATURES TO ANALYZE

Feature ID.	Feature Names	Feature ID.	Feature Names	Feature ID.	Feature Names	Feature ID.	Feature Names
f1	Bwd Packet Length Std	f21	Flow IAT Mean	f41	Packet Length Std	f61	Bwd Avg Bulk Rate
f2	Flow Bytes/s	f22	Flow IAT Max	f42	Packet Length Variance	f62	Subflow Fwd Packets
f3	Flow Packets/s	f23	Flow IAT Min	f43	FIN Flag Count	f63	Subflow Fwd Bytes
f4	Flow IAT Std	f24	Fwd IAT Total	f44	SYN Flag Count	f64	Subflow Bwd Packets
f5	Fwd IAT Std	f25	Fwd IAT Mean	f45	RST Flag Count	f65	Subflow Bwd Bytes
f6	Bwd IAT Std	f26	Fwd IAT Max	f46	PSH Flag Count	f66	Init_Win_bytes_forward
f7	Bwd Packets/s	f27	Fwd IAT Min	f47	ACK Flag Count	f67	Init_Win_bytes_backward
f8	Destination Port	f28	Bwd IAT Total	f48	URG Flag Count	f68	act_data_pkt_fwd
f9	Flow Duration	f29	Bwd IAT Mean	f49	CWE Flag Count	f69	min_seg_size_forward
f10	Total Fwd Packets	f30	Bwd IAT Max	f50	ECE Flag Count	f70	Active Mean
f11	Total Backward Packets	f31	Bwd IAT Min	f51	Down/Up Ratio	f71	Active Std
f12	Total Length of Fwd Packets	f32	Fwd PSH Flags	f52	Average Packet Size	f72	Active Max
f13	Total Length of Bwd Packets	f33	Bwd PSH Flags	f53	Avg Fwd Segment Size	f73	Active Min
f14	Fwd Packet Length Max	f34	Fwd URG Flags	f54	Avg Bwd Segment Size	f74	Idle Mean
f15	Fwd Packet Length Min	f35	Bwd URG Flags	f55	Fwd Header Length	f75	Idle Std
f16	Fwd Packet Length Mean	f36	Bwd Header Length	f56	Fwd Avg Bytes/Bulk	f76	Idle Max
f17	Fwd Packet Length Std	f37	Fwd Packets/s	f57	Fwd Avg Packets/Bulk	f77	Idle Min
f18	Bwd Packet Length Max	f38	Min Packet Length	f58	Fwd Avg Bulk Rate		
f19	Bwd Packet Length Min	f39	Max Packet Length	f59	Bwd Avg Bytes/Bulk		
f20	Bwd Packet Length Mean	f40	Packet Length Mean	f60	Bwd Avg Packets/Bulk		

TABLE IV. SELECTED FEATURES BY VARIOUS SEARCH METHODS

Search Method	# Selected Features	Features ID. Of Selected Features
Best First	6	f1, f8, f13, f19, f67, f69
Greedy Stepwise	6	f1, f8, f13, f19, f67, f69
Evolutionary	8	f12, f13, f18, f20, f51, f67, f69, f73
PSO (Proposed-1)	15	f1, f8, f18, f19, f27, f31, f42, f50, f63, f65, f66, f67, f69, f70, f72
BestFirst+ PSO (Proposed-2)	16	f1, f8, f13, f18, f19, f27, f31, f42, f50, f63, f565, f66, f67, f69, f70, f72
PSO+Evo(Proposed-3)	20	f1, f8, f12, f13, f18, f19, f20, f27, f31, f42, f50, f51, f63, f65, f66, f67, f69, f70, f72, f73

C. Performance Analysis

The features selected from each feature selection method is then used by the J48 algorithm to identify/classify normal traffic and attacks on the dataset. The TPR, FPR, Accuracy, and processing time are recorded. The TPR of attack identification/classification for each feature selection method is listed in Table 5.

Comparing the TPR of weighted attack identification/classification, the propose method achieves the highest TPR value of 0.999 in detect the Normal, DoS/DDoS and PortScan. The proposed technique of PSO+J48, PSO+BF+J48 and PSO+Evo+J48 almost have the same achievement. Furthermore, only using selected features of BF and GS the J48 can identify/classify infiltration attacks, even in small percentages. Based on the experiment result, only the propose method can detect Bot and Web Attack.

Overall, using the selected features resulted from the PSO+Evo feature selection method gives the best attacks identification/classification results as the TPR values are higher than others.

TABLE V. COMPARING THE TPR OF METHODS

Class	Methods						
	All+J48	***BF+J48***	***Evo+J48***	***GS+J48***	***PSO+J48***	***PSO+BF+J48***	***PSO+EVO+J48***
Normal	0.999	0.977	0.975	0.977	0.999	0.999	0.999
DoS/DDoS	0.999	0.998	0.994	0.998	0.999	0.999	0.999
PortScan	0.999	0.996	0.996	0.996	0.999	0.999	0.998
Bot	0.717	0.468	0.465	0.468	0.715	0.716	0.715
Web Attack	0.983	0.120	0.104	0.120	0.933	0.933	0.923
Infiltration	0.000	0.417	0.000	0.417	0.000	0.000	0.017
Brute Force	0.996	0.996	0.982	0.996	0.996	0.996	0.995

The FPR of classification using different results of feature selection methods are listed in Table 6. The experiment results shown is Table 6 indicate the propose method has the lowest FPR value of 0.002 for detection the Normal traffic compare with others. Overall, the FPR value of attack detection using selected features from the proposed technique is lower than the FPR value of Best-First, Evolutionary, and GreedyStepWise methods. The main objective of attack detection systems development is to

identify the attack with high accuracy, high TPR, and low FPR values. Thus, the proposed techniques are promising. But still, have a problem in detection infiltration attack.

TABLE VI. COMPARING THE FPR OF METHOD.

Class	Methods						
	All+J 48	***BF+J 48***	***Evo+ J48***	***GS+J 48***	***PSO+ J48***	***PSO +BF +J48***	***PSO+ EVO +J48***
Normal	0.002	0.007	0.011	0.007	0.002	0.002	0.002
DoS/DDoS	0.000	0.020	0.023	0.020	0.000	0.000	0.000
PortScan	0.000	0.000	0.001	0.000	0.000	0.000	0.000
Bot	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Web Attack	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Infiltration	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Brute Force	0.000	0.000	0.000	0.000	0.000	0.000	0.000

The average Accuracy of attack detection/classification using the selected features resulted from each method is show in figure 3. The experiment result show the accuracy of All+J48 method achieve 99.98%, BF+J48 98.04%, Evo+J48 97.73, GS+J48 98.04, PSO+J48 99.89%, PSO+BF+J48 99.98% and PSO+Evo+J48 99.88%. Overall, The proposed method achieves accuracy above 99,88%. Although, the accuracy of the classification of attacks using all features also has the same accuracy with the proposed method, but still has the limitation that will describe in the next.

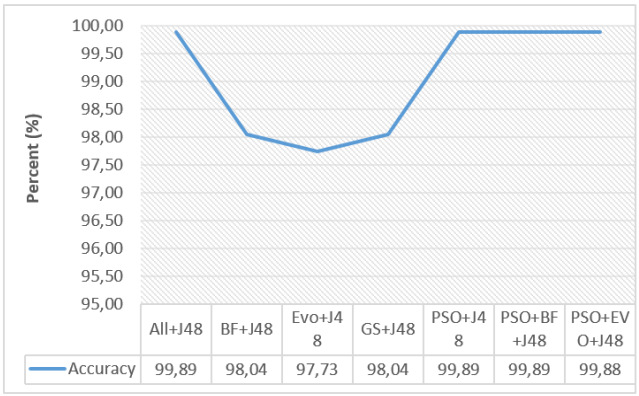

Fig. 3. Accuracy Proposed Method

To ensure the test results in this study, the F-measure value is presented in table 7. The F-Measure value is also used to measure the performance of the classification machine. The results show that the method used has a good performance in detecting normal traffic types, DoS / DDoS, PortScan, Bot, Web Attack, and Brute Force. However, the proposed method still has problems detecting the type of infiltration attack. This will be the aim of further research.

TABLE VII. COMPARING THE F-MEASURE OF METHOD.

Class	Methods						
	All+J 48	***BF+J 48***	***Evo+ J48***	***GS+J 48***	***PSO+ J48***	***PSO +BF +J48***	***PSO+ EVO +J48***
Normal	0.999	0.988	0.986	0.988	0.999	0.999	0.999
DoS/DDoS	0.998	0.938	0.929	0.938	0.999	0.999	0.999
PortScan	0.996	0.995	0.993	0.995	0.996	0.996	0.996
Bot	0.806	0.628	0.627	0.628	0.821	0.820	0.817
Web Attack	0.976	0.212	0.184	0.212	0.945	0.946	0.942
Infiltration	0.000	0.560	Nan	0.560	Nan	Nan	0.041
Brute Force	0.997	0.986	0.971	0.986	0.997	0.997	0.997

Another way to look at classification engine performance is the processing time. The observations on the processing time of the identification/classification using selected features resulted from each feature selection method are presented in Figure 4. The results show that the processing time of the J48 classification algorithm using all feature (78 features) is take a longer time. Based on the experimental results, it can be concluded that the number of features analyzed greatly influences the processing time. This also shows that the ability of the feature selection algorithm to select the best feature also affects the performance of the detection engine.

D. Comparison with previous studies

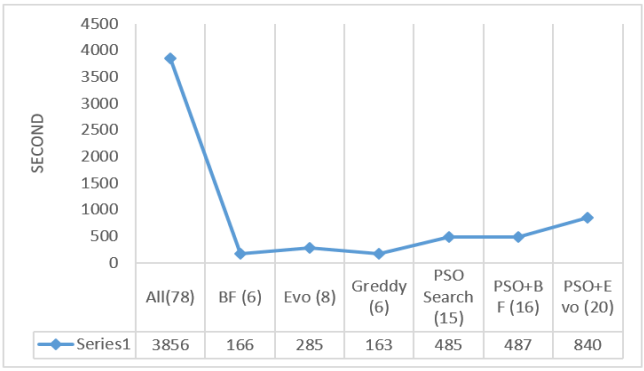

Fig. 4. Comparing the Process Time Accuracy Proposed Method

The objective of this study is to proposed a feature selection technique that effectively improved an anomaly detection system. To validate the proposed technique has been achieved the research objective, a comparison with previous studies is done. Table 8 present the experiment result and performance of previous studies, the proposed method outperforms the previous studies in terms of accuracy. As listed in Table 7, the proposed-1 (PSO+J48) and the proposed-2 (PSO+BF+J48) achieves 99.89% of accuracy while the proposed-3 (PSO+Evo+J48) achieves 99.88% of accuracy.

TABLE VIII. COMPARISON OF ACCURACY OF CLASSIFICATION TECHNIQUES BASED ON ITS SELECTED

Studies	Technique	Dataset	Accuracy
(Ahmim et al., 2019) [22]	the combination of three different classifiers namely, REP Tree, JRip algorithm and Forest PA	CICDS 2017	96.66%
(Abdulhammed et al., 2019) [35]	Dimensional reduction using auto-encoder and PCA. Random Forest	CICIDS 2017	99.60%
(Ustebay, Turgut, and Aydin, 2019) [23]	Recursive feature elimination using random forest. Deep Learning Multilayer Perceptron (DMLP)	CICIDS 2017	91.00%
The Proposed-1	Feature selection using Correlation-Based with PSO Search method and J48 Classifier	CICIDS 2017	99.89%
The proposed-2	Feature selection using Correlation-Based with BF+PSO Search method and J48 Classifier	CICIDS 2017	99.89%
The proposed-3	Feature selection using Correlation-Based with PSO+Evo Search method and J48 Classifier	CICIDS 2017	99.88%

V. CONCLUSION

In order to produce important and relevant features, it is necessary to carry out feature analysis and testing. The best feature selection technique will produce the best and relevant features that contribute to the improvement of the classification algorithm's performance. To provide the best and important features, the PSO-search method on the correlation-based feature selection technique combine with J48 the classification algorithm is proposed in this paper. The combination of the other search method to select the important features also investigated. The proposed technique has tested and the results show that the attacks detection using the J48 algorithm and the propose method outperforms the existing techniques/methods in term of Accuracy, TPR, and FPR. In other words, the selected features resulted from PSO Search and the combinations of PSO+BF also PSO+Evo improve TPR and Accuracy of the J48 classification algorithm. Surprisingly, the proposed method shows better accuracy compared to other existing techniques. Furthermore, by utilizing the selected features from the proposed feature selection techniques, the J48 algorithm can detect normal and attacks traffic. By optimizing the feature selection method/technique in turn, it will improve significantly the accuracy of the anomaly detection system. In the future, the authors of this paper consider the improvement of the anomaly detection system for imbalanced data.

ACKNOWLEDGMENT

This project is funded by Universitas Dinamika Bangsa as a part of research program Nomor:01/MOU/LPPM-STIKOM DB/VII/2019 in collaborate with Comnets Lab, Universitas Sriwijaya.

REFERENCES

[1] Sheena, K. Kumar, and G. Kumar, "Analysis of Feature Selection Techniques: A Data Mining Approach," *Int. Conf. Eng. Technol.*, vol. 4, no. Icaet, pp. 17–21, 2016.

[2] H. El Merabet and A. Hajraoui, "A survey of malware detection techniques based on machine learning," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 1, pp. 366–373, 2019.

[3] L. Wang, Y. Wang, and Q. Chang, "Feature selection methods for big data bioinformatics: A survey from the search perspective," *Methods*, vol. 111, no. August, pp. 21–31, 2016.

[4] K. Rai, M. S. Devi, and A. Guleria, "Decision Tree Based Algorithm for Intrusion Detection," *Int. J. Adv. Netw. Appl.*, vol. 07, no. 04, pp. 2828–2834, 2016.

[5] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," *J. Comput. Sci.*, vol. 25, pp. 152–160, 2018.

[6] S. K. Pandey, "Design and performance analysis of various feature selection methods for anomaly-based techniques in intrusion detection system," *Secur. Priv.*, vol. 2, no. 1, p. e56, 2019.

[7] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," *ICISSP 2018 - Proc. 4th Int. Conf. Inf. Syst. Secur. Priv.*, vol. 2018-Janua, no. Cic, pp. 108–116, 2018.

[8] A. Binbusayyis and T. Vaiyapuri, "Identifying and Benchmarking Key Features for Cyber Intrusion Detection: An Ensemble Approach," *IEEE Access*, vol. 7, pp. 106495–106513, 2019.

[9] J. Jabez, S. Gowri, S. Vigneshwari, J. A. Mayan, and S. Srinivasulu, "Anomaly Detection by Using CFS Subset and Neural Network with WEKA Tools," *Inf. Commun. Technol. Intell. Syst. Proc. ICTIS 2018*, vol. 2, pp. 675–682, 2019.

[10] A. Niranjan, D. H. Nutan, A. Nitish, P. D. Shenoy, and K. R. Venugopal, "ERCR TV: Ensemble of Random Committee and Random Tree for Efficient Anomaly Classification Using Voting," *2018 3rd Int. Conf. Converg. Technol. I2CT 2018*, pp. 1–5, 2018.

[11] S. Chormunge and S. Jena, "Efficient feature subset selection algorithm for high dimensional data," *Int. J. Electr. Comput. Eng.*, vol. 6, no. 4, pp. 1880–1888, 2016.

[12] J. Li *et al.*, "Feature selection: A data perspective," *ACM Comput. Surv.*, vol. 50, no. 6, 2017.

[13] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018.

[14] Y. Dhote, S. Agrawal, and A. J. Deen, "A Survey on Feature Selection Techniques for Internet Traffic Classification," *Proc. - 2015 Int. Conf. Comput. Intell. Commun. Networks, CICN 2015*, pp. 1375–1380, 2016.

[15] M. A. Ambusaidi, X. He, P. Nanda, and Z. Tan, "Building an intrusion detection system using a filter-based feature selection algorithm," *IEEE Trans. Comput.*, vol. 65, no. 10, pp. 2986–2998, 2016.

[16] H. Gharaee and H. Hosseinvand, "A New Feature Selection IDS based on Genetic Algorithm and SVM," pp. 1–6, 2016.

[17] Z. Chen, C. K. Yeo, B. S. L. Francis, and C. T. Lau, "Combining MIC feature selection and feature-based MSPCA for network traffic anomaly detection," *2016 3rd Int. Conf. Digit. Inf. Process. Data Mining, Wirel. Commun. DIPDMWC 2016*, pp. 176–181, 2016.

[18] I. Sumaiya Thaseen and C. Aswani Kumar, "Intrusion detection model using fusion of chi-square feature selection and multi class SVM," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 29, no. 4, pp. 462–472, 2017.

[19] E. Popoola and A. Adewumi, "Efficient feature selection technique for network intrusion detection system using discrete differential evolution and decision tree," *Int. J. Netw. Secur.*, vol. 19, no. 5, pp. 660–669, 2017.

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

[21] I. Sharafaldin, A. Gharib, A. H. Lashkari, and A. A. Ghorbani, "Towards a Reliable Intrusion Detection Benchmark Dataset," *Softw. Netw.*, vol. 2017, no. 1, pp. 177–200, 2017.

[22] A. Ahmim, L. Maglaras, M. A. Ferrag, M. Derdour, and H. Janicke, "A Novel Hierarchical Intrusion Detection System based on Decision Tree and Rules-based Models," *2019 15th Int. Conf. Distrib. Comput. Sens. Syst. (DCOSS). IEEE*, pp. 228–233, 2019.

[23] S. Ustebay, Z. Turgut, and M. A. Aydin, "Intrusion Detection System with Recursive Feature Elimination by Using Random Forest and Deep Learning Classifier," *Int. Congr. Big Data, Deep Learn. Fight. Cyber Terror. IBIGDELFT 2018 - Proc.*, pp. 71–76, 2019.

[24] R. Panigrahi and S. Borah, "A detailed analysis of CICIDS2017 dataset for designing Intrusion Detection Systems," *Int. J. Eng. Technol.*, vol. 7, no. 3.24 Special Issue 24, pp. 479–482, 2018.

[25] M. A. Hall, "Correlation-based Feature Selection for Machine Learning," *PhD Thesis*, no. April, 1999.

[26] S. Bahl and S. K. Sharma, "Performance Analysis of User to Root Attack Class Using Correlation Based Feature Selection Model," 2015.

[27] S. Bahl and D. Dahiya, "Enhanced Intrusion Detection System for Detecting Rare Class Attacks using Correlation based Dimensionality Reduction Technique," vol. 9, no. March, 2016.

[28] S. Bahl, "Features Contribution for Detecting Attacks of an Intrusion Detection System," vol. 13, no. 9, pp. 5635–5653, 2017.

[29] S. Yılmaz Gündüz and M. N. ÇETER, "Feature Selection and Comparison of Classification Algorithms for Intrusion Detection," *ANADOLU Univ. J. Sci. Technol. A - Appl. Sci. Eng.*, vol. 19, no. 1, pp. 206–218, 2018.

[30] K. A. Taher, B. M. Yasin Jisan, and M. M. Rahman, "Network Intrusion Detection using Supervised Machine Learning Technique with Feature Selection," *2019 Int. Conf. Robot. Signal Process. Tech.*, pp. 643–646, 2019.

[31] S. Aljawarneh, M. B. Yassein, and M. Aljundi, "An enhanced J48 classification algorithm for the anomaly intrusion detection systems," *Cluster Comput.*, pp. 1–17, 2017.

[32] M. Reazul, A. Rahman, and T. Samad, "A Network Intrusion Detection Framework based on Bayesian Network using Wrapper Approach," *Int. J. Comput. Appl.*, vol. 166, no. 4, pp. 13–17, 2017.

[33] M. Nikhitha and M. A. Jabbar, "K Nearest Neighbor Based Model for Intrusion Detection System," *Int. J. Recent Technol. Eng.*, vol. 8, no. 2, pp. 2258–2262, 2019.

[34] M. Nawir, A. Amir, N. Yaakob, and O. B. Lynn, "Effective and efficient network anomaly detection system using machine learning algorithm," *Bull. Electr. Eng. Informatics*, vol. 8, no. 1, pp. 46–51, 2019.

[35] R. Abdulhammed, H. Musafer, A. Alessa, M. Faezipour, and A. Abuzneid, "Features dimensionality reduction approaches for machine learning based network intrusion detection," *Electron. (Switzerland). MPDI*, vol. 8, no. 3, p. 322, 2019.