Improving Intrusion Detection System by Developing Feature Selection Model Based on Firefly Algorithm and Support Vector Machine

Wathiq Laftah Al-Yaseen

Abstract—**The nowadays growing of threads and intrusions on networks make the need for developing efficient and effective intrusion detection systems a necessity. Powerful solutions of intrusion detection systems should be capable of dealing with central network issues such as huge data, high-speed traffic, and wide variety in threat types. This paper proposes a wrapper feature selection method that is based on firefly algorithm and support vector machine. The firefly optimization algorithm has been effectively employed in diverse combinatorial problems. The proposed method improves the performance of intrusion detection by removing the irrelevant features and reduces the time of classification by reducing the dimension of data. The SVM model was employed to evaluate each of the feature subsets produced from firefly technique. The main merit of the proposed method is its ability in modifying the firefly algorithm to become suitable for selection of features. To validate the proposed approach, the popular NSL-KDD dataset was used in addition to the common measures of intrusion detection systems such as overall accuracy, detection rate, and false alarm rate. The proposed method achieved an overall accuracy of 78.89% compared with 75.81% for all the 41 features. The analysis results approved the effectiveness of the proposed feature selection method in enhancing network intrusion detection system.**

Index Terms—**intrusion detection system, support vector machine, firefly algorithm, wrapper feature selection method**

I. Introduction

THE recent increasing in speed of network data traffic and the growing number of attacks on computer networks have become challenges for security network researchers and practitioners. Moreover, the nowadays developments in network-based computer services require a parallel reliance on suitable security systems that is able to protect networks and computers against cyber-threats [1]. In spite of the recent advances, security issues are still on the rise. Intrusion detection system (IDS) has become an essential component of security infrastructure as they provide better defensive wall against internal and external attacks compared to other traditional security systems.

Intrusion detection systems monitor the events and activities that occur in the network to recognize the malicious ones [2]. In general, IDS can be classified as misuse- and anomaly-based detection models. Misuse-based detection models can only detect the known attacks based on their signatures that are stored in the database, whereas anomaly-based detection models can detect known and unknown attacks but with high false-positive rates [3][4][5].

Various approaches have been proposed to improve the intrusion detection systems. Many of them focus on anomaly based IDS that is designed based on machine learning algorithms like Support Vector Machine (SVM), Bayesian Tree, Naïve Bayes, and C4.5. However, IDS still faces many challenges that should be considered such as how to assure high intrusion detection rate and low false alarm rate in real-time. Furthermore, numerous features with difficulty to distinguish the association between them make the task of classification difficult. Feature selection is the critical step that should be attained before the classification process [6]. It includes identifying a subset of relevant features to be used in the classification process. The central advantages of feature selection process are improving prediction performance, reducing computation time, getting better understanding of the data, and overcoming the dimensionality issue. Moreover, the processing requirements of a classifier such as memory and disk space can be reduced [7]. Therefore, this paper proposes a wrapper feature selection model that is based on firefly algorithm and SVM model to enhance attack classification step in the intrusion detection process.

The bio-inspired optimization algorithms are popular to solve the combinatorial and complicated problems. Many of these algorithms have been well adopted in intrusion detection systems, for instance Particle Swarm Optimization [8][9][10][11][12] and Ant Colony Optimization [13][14][15]. Firefly method is one of the recognized and proficient bio-inspired optimization methods [16]. It has been successfully applied in the feature selection concept [17][18] but never employed in intrusion classification. On the other hand, SVM has many benefits that make it an appropriate solution for intrusion detection systems such as high generalization performances and the ability of training with noisy datasets. Moreover, SVM does not suffer from local minima and can assure fast execution time. Nevertheless, an obstacle of SVM is that its performance largely depends on the right selection of parameters.

The rest of this paper is organized as follows. Section 2 summarizes the related works of feature selection methods based on IDS. Section 3 gives an overview of the firefly method and support vector machine. Section 4 explains the proposed wrapper feature selection method FA-SVM and defines the datasets with performance measures. Section 5 discusses the experimental results. Finally, Section 6 concludes the paper and states the future work.

II. Related Work

Reviewing the most recent and relevant literature, several studies have considered the feasibility of improving

Manuscript received January 22, 2019; revised May 8, 2019.

Wathiq Laftah Al-Yaseen is with the Department of Computer Systems Techniques, Kerbala Technical Institute, Al-Furat Al-Awsat Technical University, 56001, Kerbala, Iraq, e-mail: inkr.wth@atu.edu.iq.

the intrusion detection systems performance by proposing enhancements for the feature selection step.

Aslahi et al. [19] proposed a hybrid model of GA and SVM for intrusion detection systems. This method has the capability of decreasing the features from 41 to 10. The selected features were categories into three priorities by using GA where the highest importance placed in the first priority and the lowest important in the third priority. The distribution of features was done as four features placed in the first priority, four in the second, and two in the third priority. They used the KDD'99 dataset in their experiments. The findings stated that the hybrid model could attain a positive detection of 0.973 whereas the false alarm rate was 0.017.

Rani et al. [20] introduced a hybrid detection system. They were used C5.0 decision tree as a misuse model in their approach. This model can detect the recognized attacks with low false alarm rate. Furthermore, they also applied One-Class SVM as an anomaly detection model that trained on normal traffic only chosen from the original dataset. The NSL-KDD dataset was employed in the experiments. The proposed method enhanced the detection rate and reduced the false alarm rate.

A feature selection model based on Multilayer Perception (MLP) for intrusion detection system was proposed by Ahmad et al. [21]. They combined Principal Component Analysis (PCA) and Genetic Algorithm (GA). They applied PCA to plan the features space to principal feature space and then selected the features corresponding to the highest eigenvalues. The features that were selected by PCA may lack the adequate detection for the classifier, so they adopted GA to explore the principal feature space in order to find a subset with optimal sensitivity. The feature subsets from PCA and GA will feed to train MLP classifier. The proposed method used the KDDCup'99 dataset in the evaluation; the features were reduced from 41 to 12 features only. The optimal features upgraded the detection accuracy up to 99%.

Alomari et al. [22] proposed a wrapper feature selection approach that is based on the Bees Algorithm (BA) as an exploration approach for generating a subset of features. They used SVM as a classifier to validate the subset features. Four subsets datasets were utilized with 4000 samples were generated randomly from KDDCup'99 dataset to evaluate the proposed approach. The results showed that the detection accuracy could reach up to 99% with reducing the feature group to eight features, and with 0.004 false alarm rate.

Ghanem et al. [23] introduced the Artificial Bee Colony (ABC) approach for feature selection of IDS. Their method involves two main stages: in the first stage, the subsets of features were generated of the Pareto front non-dominated solutions, while in the second stage a hybrid of a Feed Forward Neural Network (FFNN) and ABC and particle swarm optimization (PSO) were used to evaluate the feature subsets that collected from the first stage. Thus, the proposed method employed a new feature selection model named multi-objective ABC to reduce the number of network traffic features and then it used new classification approach named hybrid ABC-PSO with optimized FFNN to categorize the production data from the first stage. Moreover, a new fitness function to reduce the quantity of features was proposed to assure low false alarm rate.

Finally, Aljawarneh et al. [24] suggested a hybrid approach for intrusion detection system. In their approach, there were two main stages: at the first stage, a concept of feature selection is applied where the dataset is filtered by using the vote model based on Information Gain to select the best features that enhance the accuracy in the next stage. In the second stage, a hybrid algorithm that is composed from the following classifiers (J48, Meta Pagging, Random Tree, REPTree, AdaBoostM1, Decision Stump and Naïve Bayes) was employed to classify the samples of the testing dataset into the right classes. The results obtained based on NSL-KDD dataset pointed out that the new suggested approach improved the accuracy with a low false-positive rate and high false negative rate.

III. Background Overview

A. Firefly Algorithm

Firefly algorithm (FA) was developed by Yang [25] as a biologically stochastic global optimization approach. FA is a population-based metaheuristic where every firefly from the population is considered as a possible solution in the search space. Firefly algorithm simulates the behavior of fireflies mating and using of flash lighting to exchange information with each other [26]. In addition, they use flash lighting to attract the potential prey and provide warning mechanism. Yang [26] formulated the FA with three principles that describe the behavior of fireflies: (i) all fireflies are unisex, so that all the fireflies will be attracted to each other; (ii) attractiveness is relative to the brightness, so that any two fireflies, the less bright one will be attracted to the brighter one. However, the attractiveness decreases whenever the distance between the two fireflies increases. (iii) The firefly brightness is associated with the fitness function, if there is no firefly brighter than a current one, it will attract randomly.

The movement of firefly *i* to another brighter (more attractive) firefly *j* based on Cartesian distance can be represented by (1).

$$x_i = x_i + \beta_0 \times e^{-\gamma r_{ij}^2} \times (x_j - x_i) + \alpha \times (rand - 0.5) \quad (1)$$

Where the first part of (1) represents the movement of attraction between two fireflies, the second part represents the attraction. β_0 is the initial attractiveness which is always set to 1, and γ is the absorption coefficient which controls the speed of convergence between fireflies. The third part of (1) is randomization, where α is a constant randomization parameter defined between [0, 1], it represents the noise of the environment that be used to provide more diversity of solutions, *rand* is a random number generated from a uniform distribution [0, 1] and adjusted to range between [– 0.5, 0.5] by expression (*rand* – 0.5). Finally, *r* represents the distance between any two fireflies (i, j) which is be defined in (2).

$$r_{ij} = \|x_i - x_j\| \quad (2)$$

Where x_i represents the position of firefly i. The pseudocode of FA can be summarized as shown in Figure 1.

B. Support Vector Machine

Support vector machine (SVM) has been a powerful technique for regression analysis and classification as a result

```
Algorithm Firefly Algorithm
Input: Population size ($n$), Maximum of iteration ($maxIter$), Absorption coefficient ($\gamma$), Randomization parameter ($\alpha$), Attractiveness value ($\beta_0 = 1$)
Output: Optimal firefly position with its fitness
Generate an initial population of $n$ fireflies $X_i (i = 1, 2, \ldots, n)$ using uniform distribution.
Evaluate all the fireflies by using a fitness function
Light intensity $I_i$ at $X_i$ is determined by fitness function
$Iteration = 0$
while ($Iteration < maxIter$) do
    $Iteration = Iteration + 1$
    for i = 1 to n do
        for j = 1 to i do
            if ($I_j > I_i$) then
                Move firefly $i$ towards firefly $j$ by using equation (1)
            end if
            Evaluate the new solution by updating the light intensity
        end for
    end for
    Rank the fireflies based fitness and find the current best
end while
```

Fig. 1. Pseudocode of Firefly Algorithm (FA)

of its robust scientific basis that can convey several salient properties that alternative approaches could hardly handle.

The data in SVM is divided into several classes (two as minimum) by a hyperplane, and it simultaneously maximizes the geometric margin and minimizes the empirical classification error. Accordingly, it is also referred to as maximum margin classifiers. The Support vector machine classifier is appraised as a machine learning mechanism that relies on statistical learning principles. This classifier is capable of developing a method to split data into dissimilar categories. This is achieved depending upon the N-dimensional hyperplane that can be quantified based on a known training dataset.

The samples of the training dataset are labeled as (x_i, y_i), i = 1, 2, ..., N, where N represents the number of data samples, y_i is a class of sample, and x_i is the training dataset. The main problem of the SVM is the determination of a maximum margin separating hyperplane from the closest points at a higher dimensional space, where the SVM computes the sum of distances between the points of the hyperplane to the closest points of the dimensional space [27]. The boundary function of the biggest margin can be determined from (3) [28].

$$Minimize\, W(\alpha) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j k(x_i, x_j) - \sum_{i=1}^{N} \alpha_i \tag{3}$$

Subject to

$$\forall i : 0 \leq \alpha_i \leq C, \quad and \sum_{i=1}^{N} \alpha_i y_i = 0$$

Where α is a vector of N variables. C is the soft margin parameter, $C > 0$.

The $k(x_i, x_j)$ represents the kernel function of the support vector machine. There is a set of kernel functions that can be used with SVM to split the samples of data into different categories. These kernel functions are listed as follows [27]; the SVM reports the best results when classifying the RBF kernel function [29].

- Linear kernel: $k(x_i, x_j) = x_i^T . x_j$.
- Polynomial kernel: $k(x_i, x_j) = (\gamma x_i^T . x_j + r)^d, \gamma > 0$.
- Radial basis function (RBF) kernel: : $k(x_i, x_j) = exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$.
- Sigmoid kernel: $k(x_i, x_j) = tanh(\gamma x_i^T . x_j + r)$

Where γ, r and d are kernel parameters.

Initially, the SVM model is an application of the Vapnik's Structural Risk Minimization (SRM) concept. Vapnik's SRM is capable of dealing with overfitting the training dataset issue adequately; that is, it has low generalization errors. A model is considered as with high generalization error or overfitted if its effectiveness becomes questionable at samples outside the training set [11].

IV. Intrusion Detection System based on FA-SVM Models

This section describes the proposed model of wrapper feature selection FA with SVM to improve the detection accuracy of the intrusion detection system. The NSL-KDD dataset is employed to evaluate the performance of the proposed feature selection method. This dataset has symbolic features such as protocol, service, and flag. Therefore, the proposed method has three main stages: at first, the preprocessing of data is achieved, where the symbolic features are converted to numeric ones like $protocol \in$ [0,2], $service \in$ [0,69] and $flag \in$ [0,10] then the data is normalized to [0, 1] [30]. In the next stage, the FA is applied to build a swarm of subsets of features that will be evaluated by using SVM at the final stage. The second and third stages of the proposed method are repeated many times to reach the best subset of features depending on the accuracy of SVM. Figure 2 shows the stages of the proposed method.

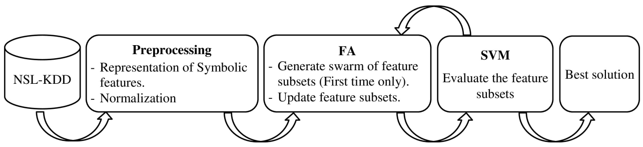
NSL-KDD
Preprocessing
- Representation of Symbolic features.
- Normalization
FA
- Generate swarm of feature subsets (First time only).
- Update feature subsets.
SVM
Evaluate the feature subsets
Best solution

Fig. 2. The stages of the proposed model

A. Dataset

The improved version of KDDCup'99 dataset [31] that called NSL-KDD dataset was chosen to evaluate the proposed model. The attacks of dataset fall into the following categories: *Denial of Service* (DoS), *Probe*, *User to Root* (U2R) and *Remote to Local* (R2L). Furthermore, NSL-KDD has two different datasets: one for training (KDDTrain+) and one for testing (KDDTest+). The test dataset includes attack types that cannot be found in the training dataset; therefore, it is an essential task for the classifier to detect the unknown

TABLE I
THE CHARACTERISTICS OF THE NSL-KDD DATASET

Category	KDDTrain+	KDDTest+
Normal	67343	9711
DoS	45927	7458
Probe	11656	2421
U2R	52	2754
R2L	995	200
Total	125973	22544

attacks. The characteristics of the NSL-KDD datasets are shown in Table I.

To evaluate the proposed method, the procedure included randomly generating training dataset with 1000 samples from KDDTrain+ and test dataset with 1000 samples from KDDTest+ dataset. Each sample has 41 features and it is labeled as normal or one of the categories (DoS, Probe, U2R, and R2L). Moreover, these features can be divided into three groups: basic (9 features), content (13 features) and traffic (19 features). Finally, the results of the full NSL-KDD dataset are calculated based on the best features that were selected from the previous phase.

B. Environment and Evaluation Measures

The proposed model then was compared with SVM classifier which was trained on all the features of the dataset (41 features). In addition, several experiments with different numbers of features (5, 10, 15, 20, 25, 30 features) were applied and compared with the 41 features. The whole experimental work has been performed on a Windows-10 PC with Intel Core i5 CPU, 12 GB RAM and @2.60 GHz. The required operations were programed using MATLAB, and multiclass classification C-SVC with RBF kernel of LIBSVM (version 3.23) was applied. The maximum number of iterations was equal to 1000 and the parameters that control the convergence of the FA are (α = 0.5, β = 1, γ = 0.1). However, the parameters of SVM are taken as (c = 1024 and γ = 0.3).

Moreover, the measures that were employed to evaluate the performance of FA-SVM are: *accuracy* (Acc), *detection rate* (DR), *false alarm rate* (FAR), *precision*, *F-score*. The details of these measures are shown as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN}$$

$$DR = TPR = Recall = \frac{TP}{TP + FN}$$

$$FAR = FPR = \frac{FP}{TN + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

$$F - score = \frac{2 \times Recall \times Precision}{Recall + Precision}$$

Where

TP: actual attack is evaluated as an attack.
FP: actual normal is evaluated as an attack.
TN: actual normal is evaluated as a normal.
FN:actual attack is evaluated as a normal.

V. EXPERIMENTAL RESULTS

The experiments of FA-SVM have been conducted with different number of features and have been implemented in different sizes of the population. Table II compares the detection accuracy of FA-SVM with a different number of features and different size of the population.

From Table II, the ratio 83.7% indicates the best detection accuracy of FA-SVM, so that it can be compared with the result of SVM when applying on the total 41 features of the dataset. Furthermore, the performance of the proposed FA-SVM model when the number of features above 10 is better than when implement with 41 features (80%). Table III, Table IV and Table V compares the performance of FA-SVM (10, 20 and 30 features) with SVM (41 features) based on detection rate, precision and F-score respectively. The proposed method shows the high improvement in the results of SVM.

The best subset of features that were selected by FA-SVM (10, 20 and 30 features) are shown in Table VI. Moreover, the ROC of comparison between FA-SVM and SVM is shown in Figure 3.

Moreover, in order to confirm that the proposed method has significant results, 10 randomly testing datasets with

TABLE II
COMPARISON OF DETECTION ACCURACY

No. of features	Population size				
	20	40	60	80	100
5	77	78.8	76.2	76.7	76.7
10	78.5	79.6	80.1	80.1	80.5
15	79.4	80.4	80.8	80.3	81.4
20	80.8	81.5	80.9	81.2	80.8
25	81.8	81.7	81.5	81.7	81.8
30	**83.7**	82.5	82.3	82.3	82.1

TABLE III
COMPARISON OF DETECTION RATES

Category	SVM	FA-SVM		
		10 features	20 features	30 features
Normal	96.73	96.73	96.3	96.08
DoS	88.39	76.49	88.69	89.29
Probe	61.05	90.53	77.89	87.37
U2R	0	0	0	0
R2L	0.94	16.98	0.94	12.26

TABLE IV
COMPARISON OF PRECISION

Category	SVM	FA-SVM		
		10 features	20 features	30 features
Normal	74.25	76.03	74.79	77.5
DoS	94.29	98.85	94.6	96.46
Probe	67.44	62.32	79.57	78.3
U2R	0	0	0	0
R2L	100	100	100	92.86

TABLE V
COMPARISON OF F-SCORE

Category	SVM	FA-SVM		
		10 features	20 features	30 features
Normal	84	85.14	84.19	85.8
DoS	91.24	86.24	91.55	92.74
Probe	64.09	73.82	78.72	82.59
U2R	0	0	0	0
R2L	1.9	29.03	1.87	21.67

TABLE VI
THE BEST SUBSET OF FEATURES SELECTED BY FA-SVM

No. of Features	The features
10 features	3, 11, 15, 30, 8, 7, 12, 19, 2, 23
20 features	7, 13, 14, 24, 41, 18, 12, 37, 27, 11, 20, 19, 23, 36, 30, 28, 3, 39, 2, 8
30 features	34, 3, 32, 21, 22, 13, 29, 28, 26, 24, 27, 36, 33, 14, 5, 23, 30, 20, 25, 15, 6, 8, 7, 38, 10, 9, 39, 2, 41, 16

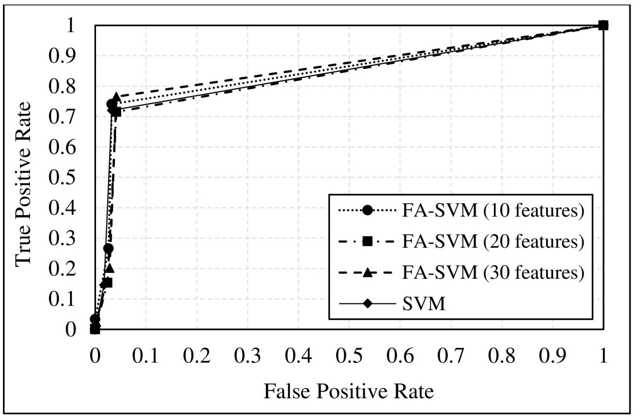

Fig. 3. ROC curve for comparing the performance of FA-SVM with SVM using a random dataset with 1000 samples

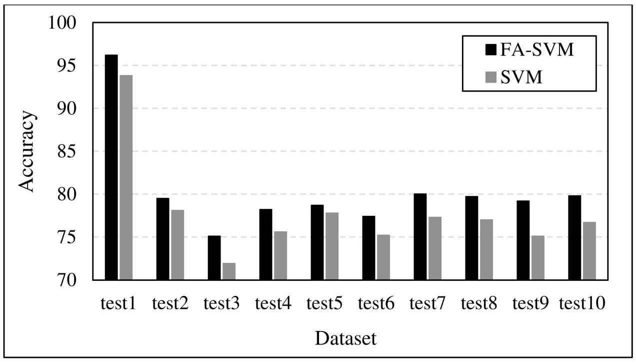

Fig. 4. Variation in the accuracy between FA-SVM and SVM based on 10 randomly testing datasets

TABLE VII
COMPARISON DETECTION RATES BETWEEN PROPOSED METHOD AND DIFFERENT CLASSIFIERS BASED ENTIRE KDDTEST+

Method	Measure	Normal	DoS	Probe	U2R	R2L	OA
SVM	DR	97.23	77.78	67.58	6	7.23	75.81
	Precision	66.26	96.28	79.96	0	95.22	
	F-score	78.81	86.05	73.25	0	13.43	
Bayesian Net	DR	94.91	62.64	69.48	21	12.82	70.82
	Precision	65.35	96.97	68.85	5.64	81.15	
	F-score	77.4	76.11	69.16	8.89	22.14	
Naïve Bayes	DR	77.88	70.29	86.45	23	14.81	68.1
	Precision	71.75	89.48	53.67	2.69	75.98	
	F-score	74.69	78.73	66.23	4.82	24.79	
SMO	DR	97.41	79.31	70.71	10.5	1.71	75.09
	Precision	65.36	97.27	90.3	63.64	75.81	
	F-score	78.23	87.38	79.31	18.03	3.34	
MLP	DR	97.65	72.46	58.69	0	0	72.34
	Precision	63.13	92.58	84.28	0	0	
	F-score	76.68	81.29	69.19	0	0	
C4.5	DR	95.4	82,8	59.27	1.5	4.25	75.38
	Precision	65.88	95.47	76.41	0	0	
	F-score	77.94	88.68	66.76	0	0	
Random Forest	DR	97.39	79.11	60.51	0.5	0.18	74.65
	Precision	64.34	96.3	85.67	0	0	
	F-score	77.49	86.86	70.92	0	0	
NN	DR	97.49	78.28	64.11	5.5	1.2	74.97
	Precision	66	96.05	14.16	0	0	
	F-score	78.71	86.26	2.21	0	0	
Proposed Method	DR	97.49	81.52	77.7	7	12.93	**78.89**
	Precision	76.03	62.32	98.85	0	100	
	F-score	85.14	86.24	73.82	0	29.03	

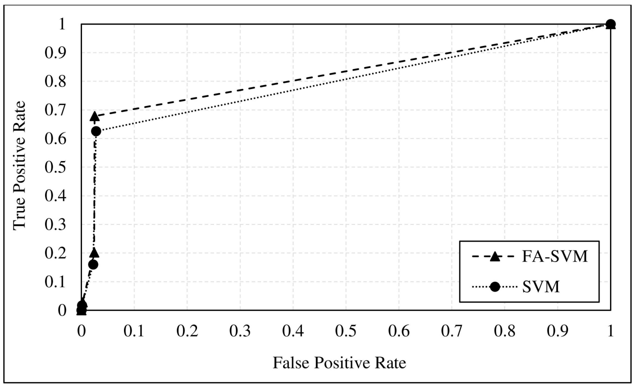

Fig. 5. ROC curve for comparing the performance of FA-SVM and SVM using entire NSL-KDD dataset

1000 samples were generated. The overall accuracies of FA-SVM with all 10 testing datasets overcome the results of SVM as shows in Figure 4. Furthermore, the t-test shows that the proposed method significantly improved the overall accuracy, where the p-value is 0.00000491287.

To compare accurately, Table VII compares the proposed model with different classifiers SVM, Bayesian Network, Naïve Bayes, SMO, MLP, C4.5, Random Forest and Neural Network when using the entire KDDTest+ dataset. The performance of the proposed model based detection rates when using 10 features from Table VI is also superiority on the all these classifiers with 41 features. Regarding accuracy, generally, this model succeeded in attaining about 78.89% good performance with an acceptable rate of false alarm (2.5%) compared to best classifier SVM with 41 features which achieved a 75.81% performance with FAR of 2.8%. Moreover, the proposed FA-SVM introduces balanced of results based all measures when comparing to other methods. Furthermore, the ROC curve of comparison FA-SVM (10 features) and SVM (41 features) is shown in Figure 5.

Moreover, a comprehensive evaluation between the proposed model and other related works implemented on the

TABLE VIII
THE EFFECTIVENESS OF THE PROPOSED MODEL WITH RESPECT TO OTHER RELATED ONES

Model	No. of Features	Overall Accuracy
CNN [32]	41	77.8
Fuzzy + NN [33]	41	78.87
ACO [15]	20	78.7
ANN [34]	29	76.3
Proposed model	**10**	**78.89**

entire KDDTest+ was also performed (see Table VIII). These works were achieved on five categories not binary classification. The proposed method proved to be powerful in comparison with the previous methods based on the number of feature and the overall accuracy criteria. We can see from Table VIII, the overall accuracy of our proposed method with only 10 features exceeded on the best method which be used Fuzzy with Neural Network and 41 features.

The key advantages of the proposed approach is the excellent enhancing of detection accuracy with using a few features compared to the other methods and also the short time of training and testing model due to the high reduction in the number of features that reached up to 76%.

VI. Conclusion and Feature Work

The present study proposed a wrapper feature selection model that combines the Firefly Algorithm (FA) with the support vector machine technique (SVM). The proposed model is a novel feature selection method (FA-SVM) that is able to reduce the number of features efficiently, and to improve the detection accuracy and false alarm rate of the SVM classifier. To evaluate the efficiency of the proposed model, the NSL-KDD benchmark was employed and compared with the SVM. The analysis revealed that FA-SVM can determine the best features of the dataset such that improving the classification of SVM as a classifier for IDS. Therefore, the future work can focus on combining FA with other classifiers and comparing it to other feature selection approaches in order to assess its quality.

References

[1] C.-J. Tu, L.-Y. Chuang, J.-Y. Chang, C.-H. Yang *et al.*, "Feature selection using pso-svm," *International Journal of Computer Science*, 2007.

[2] C. Kruegel and T. Toth, "A survey on intrusion detection systems," in *TU Vienna, Austria*. Citeseer, 2000.

[3] H.-J. Liao, C.-H. R. Lin, Y.-C. Lin, and K.-Y. Tung, "Intrusion detection system: A comprehensive review," *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 16–24, 2013.

[4] P. Kukiełka and Z. Kotulski, "New unknown attack detection with the neural network–based ids," in *The State of the Art in Intrusion Prevention and Detection*. Auerbach Publications, 2014, pp. 276–301.

[5] L.-S. Chen and J.-S. Syu, "Feature extraction based approaches for improving the performance of intrusion detection systems," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol. 1, 2015, pp. 18–20.

[6] H.-Y. Lin, "Effective feature selection for multi-class classification models," in *Proceedings of the World Congress on Engineering*, vol. 3, 2013.

[7] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[8] S. M. H. Bamakan, B. Amiri, M. Mirzabagheri, and Y. Shi, "A new intrusion detection approach using pso based multiple criteria linear programming," *Procedia Computer Science*, vol. 55, pp. 231–237, 2015.

[9] L. Xiao, Z. Shao, and G. Liu, "K-means algorithm based on particle swarm optimization algorithm for anomaly intrusion detection," in *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on*, vol. 2. IEEE, 2006, pp. 5854–5858.

[10] S. Srinoy, "Intrusion detection model based on particle swarm optimization and support vector machine," in *Computational Intelligence in Security and Defense Applications, 2007. CISDA 2007. IEEE Symposium on*. IEEE, 2007, pp. 186–192.

[11] A. A. Aburomman and M. B. I. Reaz, "A novel svm-knn-pso ensemble method for intrusion detection system," *Applied Soft Computing*, vol. 38, pp. 360–372, 2016.

[12] H. Zheng, M. Hou, and Y. Wang, "An efficient hybrid clustering-pso algorithm for anomaly intrusion detection," *Journal of Software*, vol. 6, no. 12, pp. 2350–2360, 2011.

[13] W. Feng, Q. Zhang, G. Hu, and J. X. Huang, "Mining network data for intrusion detection through combining svms with ant colony networks," *Future Generation Computer Systems*, vol. 37, pp. 127–140, 2014.

[14] H.-H. Gao, H.-H. Yang, and X.-Y. Wang, "Ant colony optimization based network intrusion feature selection and detection," in *Machine Learning and Cybernetics, 2005. Proceedings of 2005 International Conference on*, vol. 6. IEEE, 2005, pp. 3871–3875.

[15] M. H. Aghdam and P. Kabiri, "Feature selection for intrusion detection system using ant colony optimization." *IJ Network Security*, vol. 18, no. 3, pp. 420–432, 2016.

[16] M. Sayadi, R. Ramezanian, and N. Ghaffari-Nasab, "A discrete firefly meta-heuristic with local search for makespan minimization in permutation flow shop scheduling problems," *International Journal of Industrial Engineering Computations*, vol. 1, no. 1, pp. 1–10, 2010.

[17] E. Emary, H. M. Zawbaa, K. K. A. Ghany, A. E. Hassanien, and B. Parv, "Firefly optimization algorithm for feature selection," in *Proceedings of the 7th Balkan Conference on Informatics Conference*. ACM, 2015, p. 26.

[18] M. Goodarzi and L. dos Santos Coelho, "Firefly as a novel swarm intelligence variable selection method in spectroscopy," *Analytica chimica acta*, vol. 852, pp. 20–27, 2014.

[19] B. Aslahi-Shahri, R. Rahmani, M. Chizari, A. Maralani, M. Eslami, M. Golkar, and A. Ebrahimi, "A hybrid method consisting of ga and svm for intrusion detection system," *Neural computing and applications*, vol. 27, no. 6, pp. 1669–1676, 2016.

[20] M. S. Rani and S. B. Xavier, "A hybrid intrusion detection system based on c5. 0 decision tree and one-class svm," *International journal of current engineering and technology*, vol. 5, no. 3, pp. 2001–2007, 2015.

[21] I. Ahmad, A. Abdullah, A. Alghamdi, K. Alnfajan, and M. Hussain, "Intrusion detection using feature subset selection based on mlp," *Scientific research and essays*, vol. 6, no. 34, pp. 6804–6810, 2011.

[22] O. Alomari and Z. A. Othman, "Bees algorithm for feature selection in network anomaly detection," *Journal of applied sciences research*, vol. 8, no. 3, pp. 1748–1756, 2012.

[23] W. A. H. Ghanem and A. Jantan, "Novel multi-objective artificial bee colony optimization for wrapper based feature selection in intrusion detection," *International journal of advance soft computing applications*, vol. 8, no. 1, 2016.

[24] S. Aljawarneh, M. Aldwairi, and M. B. Yassein, "Anomaly-based intrusion detection system through feature selection analysis and building hybrid efficient model," *Journal of Computational Science*, vol. 25, pp. 152–160, 2018.

[25] X.-S. Yang, *Nature-inspired metaheuristic algorithms*. Luniver press, 2010.

[26] X. S. Yang, "Firefly algorithms for multimodal optimization," in *International symposium on stochastic algorithms*. Springer, 2009, pp. 169–178.

[27] V. Golmah, "An efficient hybrid intrusion detection system based on c5. 0 and svm," *International Journal of Database Theory and Application*, vol. 7, no. 2, pp. 59–70, 2014.

[28] C.-W. Hsu, C.-C. Chang, C.-J. Lin *et al.*, "A practical guide to support vector classification," 2003.

[29] F. Kuang, W. Xu, and S. Zhang, "A novel hybrid kpca and svm with ga model for intrusion detection," *Applied Soft Computing*, vol. 18, pp. 178–184, 2014.

[30] M. Sabhnani and G. Serpen, "Application of machine learning algorithms to kdd intrusion detection dataset within misuse detection context." in *MLMTA*, 2003, pp. 209–215.

[31] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, "A detailed analysis of the kdd cup 99 data set," in *Computational Intelligence for Security and Defense Applications, 2009. CISDA 2009. IEEE Symposium on*. IEEE, 2009, pp. 1–6.

[32] M. Zhu, K. Ye, and C.-Z. Xu, "Network anomaly detection and identification based on deep learning methods," in *International Conference on Cloud Computing*. Springer, 2018, pp. 219–234.

[33] R. A. R. Ashfaq, X.-Z. Wang, J. Z. Huang, H. Abbas, and Y.-L. He, "Fuzziness based semi-supervised learning approach for intrusion detection system," *Information Sciences*, vol. 378, pp. 484–497, 2017.

[34] B. Ingre and A. Yadav, "Performance analysis of nsl-kdd dataset using ann," in *2015 International Conference on Signal Processing and Communication Engineering Systems*. IEEE, 2015, pp. 92–96.

Author (Wathiq Laftah Al-Yaseen) received his B.Sc. degree in computer science from the University of Basrah in 2000. He received his M.Sc. degree in Computer Science from the University of Babylon, Iraq in 2003. He received his Ph.D. degree in Computer Science in 2017 from FTSM/UKM, Malaysia. He is currently a Lecturer in the Department of Computer Systems Techniques at Kerbala Technical Institute in Al-Furat Al-Awsat Technical University, Kerbala, Iraq. His research interests include Artificial Intelligence, Network Security, Machine Learning and Bioinformatics.