Intrusion Detection Using Artificial Bee Colony With Reduced Features

P.Amudha*[1], S.Karthik[2], S.Sivakumari [3]

[1,3] Department of CSE, Faculty of Engineering, Avinashilingam Institute for Home Science and Higher Education for Women, Coimbatore – 641 108, Tamilnadu, India.
[1]amudharul@gmail.com
[2] Department of CSE, SNS College of Technology, Coimbatore 641 035, Tamilnadu, India.

Abstract

Intrusion detection is considered to be a classification problem which can detect normal traffic and other types of intrusions. The significant part is to select an efficient classification approach to construct an accurate intrusion detection system. But the performances of these methods depend on the large volume of training data. The dimensions of data can be reduced by selecting vital features using feature selection method to achieve maximum accuracy for the predicted target class. The purpose of this work is to select the features that can characterize the pattern of the network traffic and propose a hybrid algorithm to combine Artificial Bee Colony (ABC) with Support Vector Machine (SVM). Based on best features chosen, the performance of proposed algorithm is compared with other classifiers. The experimental results show that the proposed method is efficient in terms of detection rate, accuracy and false alarm rate when compared to existing methods and found to be promising for intrusion detection system.

Keywords: intrusion detection, feature selection, classification, artificial bee colony, support vector machine

Introduction

Intrusion Detection System (IDS) is a security support mechanism which has become an essential component of security infrastructure to detect attacks, identify and track the intruders. The most crucial step in building intrusion detection system is feature selection because the quantity of data is enormous that includes thousands of traffic records with number of various features. In recent years, data mining techniques on network traffic data provides a prospective solution that helps to build up better

intrusion detection systems. Due to large amount of features in intrusion detection, classifying network traffic to distinguish normal and abnormal behavior is difficult. The feature selection method reduces high dimensional dataset by removing redundant and irrelevant features and increases classification accuracy. As intrusion detection problem is considered as a classification technique, to increase the accuracy rate of classifiers, classification algorithms can be hybridized. In the recent past, computational intelligence technique such as swarm intelligence is proposed for intrusion detection [1]. Swarm Intelligence techniques aim to solve complex problems by multiple simple agents without any centralized control.

In this paper, the most important features are selected from KDDCUP'99 dataset and are evaluated using the hybridization of Artificial Bee Colony (ABC) and Support Vector Machine (SVM) algorithms. As it is seen from the results, the performance of the proposed method is highly effective compared to other results attained and seems promising for intrusion detection problem. The rest of this paper is organized as follows: Section 2 presents related work; Section 3 explains the working principle of ABC and proposed hybrid approach. Section 4 details experimental results and discussions and finally conclusion is given in section 5.

Related Work

Joshi and Varsha S. Pimprale [2] discussed that with the rapid development in communication technology, the security of computer network is one of the challenging issues and so as an Intrusion Detection system (IDS). Using data mining techniques, it is easy to find useful and interesting pattern from large volume of data. Hence the performance of Intrusion Detection system can be increased and also network security can be enhanced. The feature selection can increase the accuracy rate of intrusion detection systems. Lee and Stolfo [3] discussed that the redundant features make it difficult to detect intrusion patterns in real-life intrusion detection dataset. Generally, feature selection methods searches through the features subset and finds the best subset which still contains enough information for intrusion detection. To attain real-time intrusion detection, several methods of performing feature selection have been carried out by the research community. The main advantage of feature selection is to improve the detector's performance as it reduces its dimensionality. Ron Kohavi and George H. John[4] described the use of feature subset selection which gives the subset of features in a dataset to the supervised learning algorithm. Kira et al.[5] proposed a feature selection algorithm based on the Euclidean distance between data points. Other machine learning methods, such as Bayesian networks, decision trees, and neural networks have also been used for feature selection (Forman, 2003)[6].

Support Vector Machines (SVM) proposed by Vapnik [7][8] has become a popular research method in intrusion detection due to its good generalization performance. Fatima Ardjani and Kaddour Sadouni [9] used 10-fold cross validation and optimized the performance of SVM using Particle Swarm Optimization (PSO). It utilizes the advantage of minimum structural risk with global optimizing features. The

result shows better accuracy with high execution time. Yang Lia et al. [10] proposed an IDS using Random Mutation Hill Climbing (RMHC) as search strategy and evaluated by modified linear SVM with wrapper based feature selection algorithm. Shelly Xiaonan Wu and Wolfgang Banzhaf [1] discussed that the research community is currently interested on intrusion detection system based on Computational Intelligence (CI). The characteristic features of CI systems are found to be appropriate in constructing an effective intrusion detection model. Karaboga and Basturk [11] proposed an optimization algorithm called Artificial Bee Colony (ABC) which is based on the behaviour of honeybees. Changseok Bae et al. [12] proposed Artificial Bee Colony Algorithm for Anomaly-Based Network Intrusion Detection System (A-NIDS-ABC) to optimize the solution. Mustafa Serter Uzer et al. [13] presented a hybridized method that combined ABC algorithm for selecting feature subset and Support Vector Machine for classification purpose of Medical datasets.

Artificial Bee Colony

The Artificial Bee Colony algorithm (ABC) was proposed by Karaboga and Basturk which is an optimization algorithm based on the foraging behaviour of honey bees [11]. There are three groups in ABC, namely: scouts, onlooker bees and employed bees. Scout is the bee which performs random search and the bee which visits food source is called employed bee. The bee which waits on the dance area is called onlooker bee and onlooker bee with scout is called unemployed bee. The employed and unemployed bees search for the rich food sources around the hive where, the employed bees store the information of food source and share the information with onlooker bees. The amount of food sources is equal to the amount of employed bees and also equal to the amount of onlooker bees. The location of a good food source specifies the location of a promising solution to the optimization problem [11].

The main steps of the ABC algorithm are as follows:
Algorithm 1: Artificial Bee Colony Algorithm
Input: Initial solution
Output: Optimal solution
1. Initialize Population X_h randomly (h = 1, 2, ..., SN)
2. Evaluation Population
3. REPEAT
4. FOR employed Bees
   - Produce new solution
   - Evaluate the fitness
   - Apply Greedy selection process
END FOR
5. Calculate probability of the solution
6. FOR Onlooker Bees
   - Produce new solution
   - Evaluate the fitness
   - Apply Greedy selection process

END FOR
7.   IF employed bee becomes Scout
THEN replace with new random food source solution
8.   Memorize the best solution found so far
9.   UNTIL All requirements are met
Probability of selecting a nectar source is given in equation (1)

$$P_i = \frac{F(\theta_i)}{\sum_{k=1}^{s} F(\theta_k)}$$ (1)

where,
P_i = Probability of selecting the ith employed bee
S = Number of employed bee
θ_i = Position of ith employed bee
$F(\theta_k)$ = Fitness value
Calculation of the new position is done using equation (2),

$$X_{ij}(t+1) = X_{ij}(t) + u\left(X_{ij}(t) - X_{kj}(t)\right)$$ (2)

where,
X_i = The position of onlooker bee
t = The iteration number
k = Randomly chosen solution
j = Dimension of the solution
u = Series of random variables in the range [1,-1]

Movement of Scout bees is calculated by equation (3),

$$X_{ij} = X_j^{min} + r(X_j^{max} - X_j^{min})$$ (3)

where, r is random number.

Support Vector Machine

Recently, Support Vector Machine (SVM) has been used widely in the field of machine learning. Support vector machine (SVM) is a set of related supervised learning methods introduced by Vapnik [7][8] which finds an optimal separating hyper plane and is used for classification and regression. The unique feature of SVM is the support vectors which represent the decision boundary using a subset of the training data. The main aim of support vector classification is to device an efficient way of learning good separating hyper planes in a high dimensional feature space [14]. SVMs can be trained to look for hyper planes in both linear separable data and non-linearly separable data [15]. This technique works very well with high dimensional data.

The two-class problem for the dataset D is given as $(x_1, y_1),(x_2, y_2)\ldots(x_d, y_d)$, where x_i is the set of training tuples with associated class labels, y_i. The values of each y_i can be either +1 or -1. Let us consider an example as given in Figure 1.
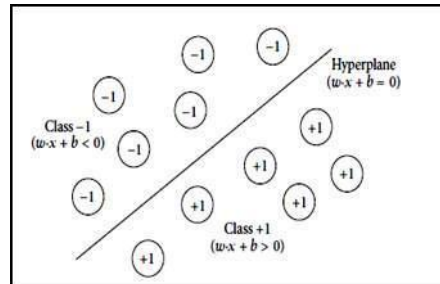
Figure 1: Linearly Separable Data

In the Figure 1, the data are linearly separable because the data points are separated into two distinct classes where y=+1 and y=-1. The hyper planes with maximal margin between them will be considered as the best separating hyper planes and it can classify the future data tuples in an accurate manner than the smaller margin. The separating hyper plane can be written as in equation (4),

$$w \cdot x + b = 0 \tag{4}$$

where, w is a weight vector and b is a bias(scalar). The maximal margin is denoted mathematically as in equation (5),

$$M = \frac{2}{\| w \|} \tag{5}$$

where, $\|w\|$ is the Euclidean norm of w.

Linear SVM is a trained SVM which can be used to classify linearly separable data. Langrangian formula can be used to rewrite the maximal margin hyper plane as the decision boundary for the classification of test or new tuples as given in equation (6),

$$d\left(x^T\right) = \sum_{i-1}^{l} y_i \alpha_i x^T + b_o \tag{6}$$

where, y_i is the class label of support vector x_i
x^T is a test tuple,
α_i is a Langrangian multiplier,
b_o is a numeric parameter,
l is the number of support vectors

For linearly separable data, the support vectors are the subset of actual training tuples. After the training process, the classifier becomes ready for prediction of the class membership on new patterns. The class of a pattern x_k is determined using equation (7),

$$class(x_k) = \begin{cases} +1 & if\ w.x_k + b\ > 0 \\ -1 & if\ w.x_k + b\ < 0 \end{cases} \tag{7}$$

Thus, the sign of $w \cdot x + b$ determines the classification of new patterns. As the classifier is described by the number of support vectors, SVMs are less prone to over-fitting. Also upper bound on the expected error rate of the SVM classifier can be computed by the number of support vectors and good generalization can be achieved regardless of the dimensionality of the data [16]. The main task of SVM classifier is to choose its right kernel which may result in different performances using different kernel functions. Multiple occurrences of kernel matrices can be derived for each kernel function by changing the kernel parameters [17]. A kernel function is a function k(x,y) with characteristic,

$$k(x, y) = \langle \phi(x), \phi(y) \rangle \tag{8}$$

The simplest of all kernel function is the dot/linear kernel $k(x,y) = x.y$. The decision function takes the form,

$$f(x) = w.x + b \tag{9}$$

The RBF Kernel is a popular kernel function used in support vector machine classification.

Proposed Hybrid ABC-SVM

Despite having better performance compared to some other swarm intelligence techniques, the basic ABC still need to be improved on intrusion detection in some aspects. One possible improvement is to increase the speed and improve the accuracy of ABC by feature selection and hybrid methods. In this work, the modified form of ABC is combined with support vector machine.

The procedure describing proposed hybrid ABC-SVM approach is as follows.
1. Initializing ABC with initial population X_h randomly, $h = 1, 2, ..., SN$.
2. Evaluating the fitness of each h = 1,2, .., SN.
3. For each employed bee and onlooker bee produce the new solution S_i, calculate the fitness value for S_i and apply greedy selection process.
4. Calculate the probability P_h for X_h.
5. Update and store the best solution.
6. After converging, the global best solution in the swarm is input to SVM classifier for training.

The two machine learning methods, Artificial Bee Colony and Support Vector machine (ABC-SVM) algorithm are combined in which the parameters of SVM are optimized using ABC. ABC begins with initial solutions that are available and iteratively explores the optimal particle. Each solution is a d-dimensional vector and represents a food source (candidate solution). In bee colony, in order to evaluate the

fitness value of the solution, an employee bee is assigned for each solution according to equation (10).

$$fit_i = \begin{cases} \frac{1}{f_i+1}, & f_i \geq 0 \\ 1 + |f_i|, & f_i < 0 \end{cases} \tag{10}$$

where, f_i represents the objective value of i-th solution.

The optimization algorithm can be confined easily in a local optimum while moving towards the global optimum; it is difficult to find an optimal solution to this function. Hence in this paper, the classical benchmark function Rastrigin [18] is implemented using Artificial Bee Colony Algorithm and is modified. Thus in equation (9), f_i is considered as Rastrigin function whose value is 0 at its global minimum (0,0,…,0). In finding the global minimum, this function is chosen, because it is considered to be one of the best test functions. Initialization range of this function is [−15, 15]. To produce many local minima, this function is combined with cosine modulation and hence the function is called multimodal.

The candidate solution (new solution) is chosen by the employee bee from the neighbor food sources and which then applies greedy selection strategy by calculating the Rastrigin function as in equation (11).

$$Min \ f(x) = 10n + \sum_{i=1}^{n} \left[x_i^2 - 10\cos(2\pi x_i) \right] \tag{11}$$

In order to evaluate its performance using 10-fold cross validation method, for each candidate solution, SVM classifier is built. The algorithm uses the best fit food source determined by the onlooker bee is selected for the next location of n-candidate solution (food source). Thus, on the average, succeeding population of each candidate solution will be better than its predecessor. This procedure continues until the performance of SVM converges. Sequential Minimal optimization (SMO) is used in the training stage of SVM. SMO algorithm is a popular optimization method used to train the support vector machine. The RBF kernel function is widely used in SVM classification and the RBF kernel which is represented as feature vectors in some input space, is defined as,

$$K(x,y) = \exp\left(-\frac{\| x - y \|^2}{2\sigma^2} \right) \tag{12}$$

where, - $\|x - y\|^2$ may be recognized as the squared Euclidean distance between the two feature vectors and $2\sigma^2$ is a free parameter. SVM training algorithm automatically produces, the number of support vectors, the weights and the threshold of RBF and yield excellent results [19]. Setting the two major RBF parameters in SVM, C and γ, must be appropriate. Parameter C signifies the cost of the penalty. Parameter γ has a

much greater influence on classification outcomes than *C* because if γ is very large, then it will result in over-fitting, while a small value leads to under-fitting.

Experimental Results and Discussions

Dataset Description and Experimental Setup

In this paper, intrusion detection dataset KDDCup'99 [20] dataset which is derived from UCI Machine Learning Repository [21] is used for experiment. It is a collection of LAN simulated TCP data of U.S. Air Force which contains 41 features which specifies the status of a connection as either normal, or attack type. Actually KDDCUP consists of three datasets namely, Whole KDD, 10% KDD and Corrected KDD and the detail is shown in Table 1.

Table 1: Number of instances in KDDCUP'99 datasets

Dataset	Number of Instances					
	Normal	DoS	Probe	U2R	R2L	Total
Whole KDD	972,780	3,883,370	41,102	52	1,126	4,898,430
10% KDD	97,277	391,458	4,107	52	1,126	494,020
Corrected KDD	60,593	229,853	4,166	70	16,347	311,029

For experimental analysis 10%KDDCup'99 dataset is used. The dataset contains large number of redundant instances and is shown in Table 2. The duplicate instances are removed during pre-processing and datasets are generated.

Table 2. Statistics of duplicate instances in 10% KDDCUP'99 dataset

	Original instances	Distinct instances	Reduction rate	Selected instances
Normal	97,277	87832	9.71%	8783
DoS	391,458	54572	86.06%	7935
Probe	4,107	2131	48.11%	2131
U2R	52	52	0%	52
R2L	1,126	999	11.28%	999
Total	494,020	145,586	70.53%	19,900

The number of features in the dataset is reduced by forming a new feature subset considering a single feature. Initially, accuracy and detection rate are estimated for the first feature in the dataset using a classifier. This procedure is continued for all the features in the dataset. Then the features are arranged in sorted manner based on their efficiency and the features are selected if their accuracy and detection rate are greater than threshold value calculated using all the features.

The performance metrics like accuracy, detection rate and false alarm rate are recorded for the classification algorithms. The confusion matrix includes four

classification performance indices: *True Positive* (*TP*), *True Negative* (*TN*), *False Positive* (*FP*), and *False Negative* (*FN*) are given in Table 3. It is usually used to evaluate the performance in the two-class classification problem.

Table 3. Confusion Matrix

Actual Class	Predicted Class	
	Positive	Negative
Positive	True Positive	False Negative
Negative	False Positive	True Negative

- *True Positive* (*TP*): The number of of attacks that are correctly identified.
- *True Negative* (*TN*): The number of normal records that are correctly classified.
- *False Positive* (*FP*): The number of normal records incorrectly classified
- *False Negative* (*FN*): The number of attacks incorrectly classified.

Equations 12 to 14 describe *CA*, *DR* and *FAR* respectively.

$$Classification\ accuracy = (TP + TN)/(TP + TN + FP + FN) \qquad (12)$$

$$Detection\ Rate(DR) = TP/(TP + FN) \qquad (13)$$

$$False\ Alarm\ Rate(FAR) = FP/(TN + FP) \qquad (14)$$

The *k*-fold cross-validation method is used for improving the classifier reliability and is used for the test result to be more valuable [22]. In this paper, 10-fold cross validation is used in which the original sample is divided into random 10 subsamples, one of which is used for testing and the remaining sub-samples are used for training. This process is repeated for 10- times and average of each fold gives the accuracy of the algorithm [23].

Results and Discussions

The proposed feature selection method is applied on the datasets and then classification algorithms are applied on each datasets with reduced features. The reduced number of features selected for each dataset is given in Figure 2 and Figure 3 presents the process of feature selection.
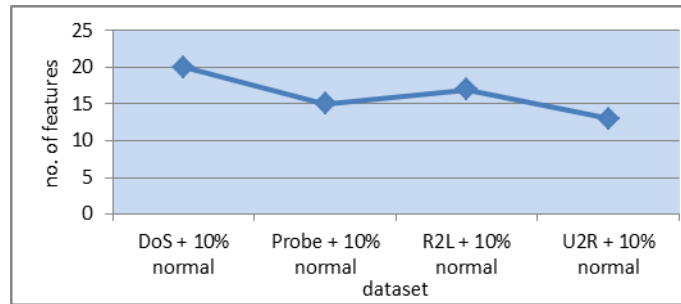
**Figure 2:** Number of features selected
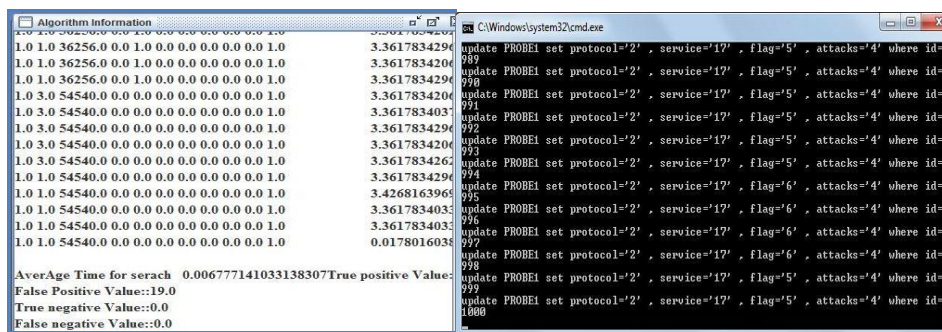
Figure 3: Feature Selection Process

Classification accuracies are obtained by applying 10-fold cross-validation. The results reveal that the accuracy of hybrid ABC-SVM is 98.22% for DoS+ 10% normal dataset, 97.31% for Probe+ 10% normal dataset, 98.33% for R2L+ 10% normal dataset and 98.57% for U2R+ 10% normal dataset. Based on the values obtained, proposed hybrid approach obtains highest accuracy percentage when compared to SVM and ABC classification based algorithms. Figure 4 specifies a comparison chart on the accuracy rate obtained by the proposed method. The results also indicate that the ABC provides better accuracy rate for Probe+ 10% normal and U2R+ 10% normal datasets.
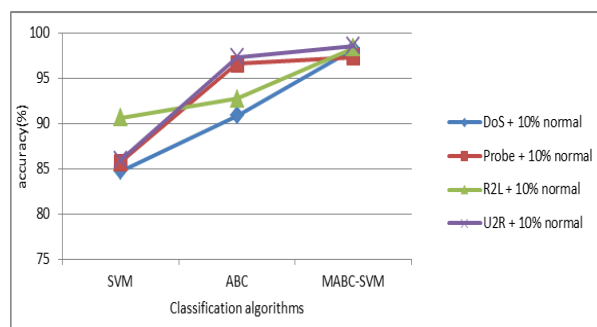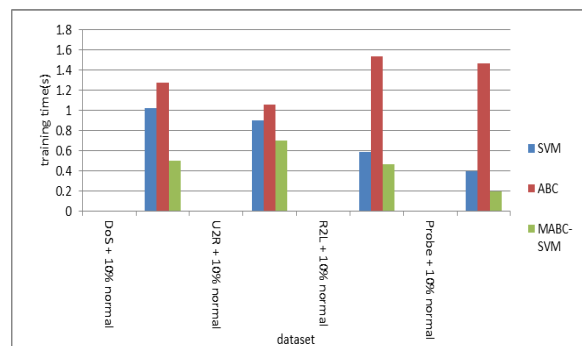

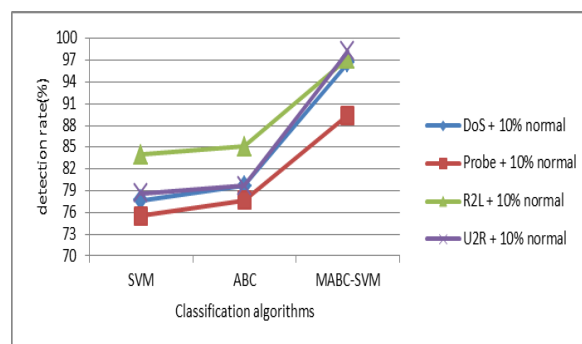**Figure 4:** Comparison on Accuracy Rate
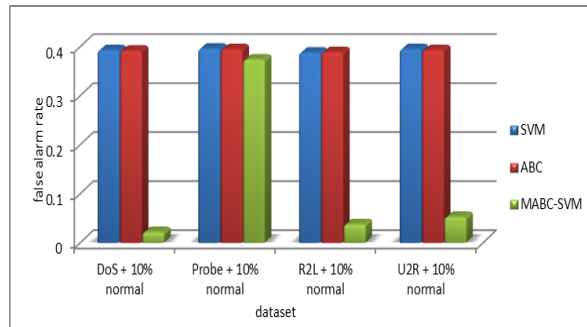
Figure 5: False Alarm Rate of Proposed Algorithm

In an effective intrusion detection system, it is preferred to have high detection rate and low false alarm rate. Figure 5 illustrates the performance of the algorithms based on false alarm rate. The proposed hybrid algorithm effectively detects attack with low false alarm rate for all the datasets ranging from 0.02 to 0.05 except Probe+10% normal dataset.

Figure 6: Time taken for proposed and existing algorithms

Figure 7: Comparison on detection rate

Figure 6 reveals that SVM takes less time for U2R, R2L and Probe datasets but provides less accuracy percentage than the other. The basic ABC algorithm consumes more time to build a model for all datasets. However the modified ABC-SVM takes

less time for Probe+10%normal dataset compared to other datasets. Comparatively the proposed algorithm takes less time and provides better accuracy and low false alarm rate for all datasets than other existing algorithms. It is indicated from the Figure 7, the detection rate is very low for SVM and ABC algorithms in DoS+10%normal, Probe+10%normal and U2R+10%normal dataset. In R2L+10%normal dataset, both SVM and ABC algorithms presents an acceptable detection rate (ranging from 84% to 97%) compared to other datasets. As U2R+10%normal dataset contains only 52 instances, classification accuracy and detection rate presented on U2R+10%normal dataset is satisfactory using hybrid algorithm. By using the hybrid algorithm, highest DR 96.67%, 97.24% and 98.24% are obtained for DoS+10%normal, R2L+10%normal and U2R+10%normal datasets respectively.

Conclusion

In the KDDCup'99 dataset used in this work contains large number of redundant instances and 41 features which affects the success of the classifier and the processing time. Hence the duplicate instances were removed during pre-processing and feature subsets were formed. In this work, feature selection method based on single feature was proposed to evaluate its effect on the performance of classification algorithms. The feature selection method effectively reduced the number of features needed to improve the performance of the classifiers. To achieve a reliable performance of the classifier, 10-fold cross validation method was used. The experiment results proved that the proposed hybrid ABC-SVM model is effective on the KDDCup'99 dataset. The hybrid approach outperformed other algorithms in terms of accuracy, detection rate and false alarm rate. In addition, the time used to detect an intrusion of the proposed model is much less than the conventional SVM and ABC algorithms.

References

[1]   Shelly Xiaonan Wu, Wolfgang Banzhaf, 2010, "The use of computational intelligence in intrusion detection systems: A review," Applied Soft Computing, pp.1-35.

[2]   Joshi, S.A., Varsha S.Pimprale, 2013, "Network Intrusion Detection System (NIDS) based on Data Mining", International J. of Engineering Science and Innovative Technology (IJESIT), 2(1), pp.95-98.

[3]   Lee, W., Stolfo, S.J., 2000," A framework for constructing features and models for intrusion detection systems," Information System Security, 3(4), pp.227–261.

[4]   Ron Kohavi, George H. John., 1997, " Wrappers for feature subset selection," Artificial Intelligence, 97, pp.273-324.

[5]   Kira, K., and Rendell, L.A.., 1992, "The feature selection problem: traditional methods and new algorithm," Proc. 9th National Conference on Artificial Intelligence, pp: 129–134.

[6]   Forman, G., 2003,"An extensive empirical study of feature selection metrics for text classification," J. Machine Learning Research, 3, pp.1289–1306.

[7]   Vapnik, V.N., 1998, Statistical Learning Theory, John Wiley and Sons.

[8]   Vapnik,V.N., 1995, The Nature of Statistical Learning Theory, Springer-Verlag.

[9]   Fatima Ardjani, Kaddour Sadouni, 2010, "Optimization of SVM multiclass by particle swarm (PSO-SVM)," International J. Modern Education and Computer Science, 2, pp.32–38.

[10]   Yang Lia, Jun-Li Wang, Zhi-Hong Tian, Tian-Bo Lu, Chen Young, 2009, " Building lightweight intrusion detection system using wrapper based feature selection mechanisms," Computers and Security, 28, pp.466-475.

[11]   Karaboga, D., and Basturk, B., 2008," On the Performance of Artificial Bee Colony (ABC) Algorithm," Applied Soft Computing, pp. 687-697.

[12]   Changseok Bae, Wei-Chang Yeh, Mohd Afizi Mohd Shukran, Yuk Ying Chung and Tsung-Jung Hsieh, 2012, "A Novel Anomaly-Network Intrusion Detection System Using ABC Algorithms," International J. of Innovative Computing, Information and Control, 8(12), pp. 8231-8248.

[13]   Mustafa Serter Uzer, Nihat Yilmaz, Onur Inan, 2013, " Feature Selection Method Based on Artificial Bee Colony Algorithm and Support Vector Machines for Medical Datasets Classification", The Scientific World Journal, pp.1-11.

[14]   Nello Cristianini, John Shawe-Taylor, 2000, An Introduction to Support Vector Machines and other Kernel-based Learning Methods, Cambridge University Press.

[15]   Pang-Ning Tan, Michael Steinbach, Vipin Kumar, 2006, Introduction to Data Mining, Pearson Education.

[16]   Jaiwei Han, Micheline Kamber, 2006, Data Mining Concepts and Techniques, 2nd Edition, Morgann Kaufmann.

[17]   Wu Zhili, 2004, " Kernel Based Learning Methods for Pattern and Feature Analysis," Ph.D Thesis, Hong Kong Baptist University.

[18]   Rajesh A. Thakker, Shojaei Baghini, M., and. Patil, M. B., 2009, Automatic Design of Low-Power Low-Voltage Analog Circuits Using Particle Swarm Optimization with Re-Initialization," J. of Low Power Electronics, 5, pp. 1 –12.

[19] Sivakumari, S., Praveena Priyadarsini, R., Amudha, P., 2009, " Performance Evaluation of SVM Kernels using Hybrid PSO-SVM," ICGST-AIML J., 9(1), pp.19-25.

[20] http://kdd.ics.uci.edu/ Databases/kddcup99/10 percent.gz.

[21] UCI Repository of Machine Learning Databases, University of California, Irvine, Dept. of Information and Computer Science. http://www.ics.uci.edu/~mlearn/ML Repository.html

[22] Francois, D., Rossi, F., Wertz, V., and M. Verleysen, 2007," Resampling methods for parameter-free and robust feature selection with mutual information," Neurocomputing,70(7-9), pp.1276–1288.

[23] Diamantidis, N. A., Karlis, D.,and Giakoumakis, E. A., 2000," Unsupervised stratification of cross-validation for accuracy estimation," Artificial Intelligence, 116(1-2), pp. 1–16.