A novel feature selection method in IDS system using Multi-objective Genetic Algorithm and Support Vector Machine for Smart grid Communication network

G. Manisha[a] , R. Vijayanand[b]

[a] *Department of Computer Science and Engineering, J.B.Institute of Engineering and Technology, Hyderabad, India.*

[b] *Department of Computer Science and Engineering, J.B.Institute of Engineering and Technology, Hyderabad, India.*

Abstract - ***Intrusion detection is a major security concern in neighborhood area network of smart grid. Intrusion detection system uses classifiers for detection purpose and it suffers from the irrelevant and noisy features of network traffic. Feature selection enhances the attack detection by selecting the most informative features of network traffic as input. In this paper, the task of identifying the relevant features is represented as multi-objective optimization problem with maximization of accuracy and minimization of number of informative features as objectives. This optimization problem is solved by applying Multi-objective genetic algorithm. In this work, the Support vector machine used with the optimization algorithm for the selection of diverse pareto optimal solutions. The proposed system was evaluated by the standard UNSW_NB15 dataset. The simulation results show that the proposed system finds multiple efficient solutions with high accuracy and less number of informative features and is suitable for the application of intrusion detection.***

Keywords : Multi-objective genetic algorithm, Smart meter data security, Intrusion detection system, Support vector machine

I.INTRODUCTION

Development of the smart grid has incited a wave of research efforts in attempts to secure the grid in an effective manner. Smart grid is an advanced electrical network that incorporates the thriving technologies of computation and communication networks. At the home area network level, any effort to secure the grid should have low computational complexity to be viable. Smart grid is named as Advanced metering infrastructure (AMI) and is orderly arranged as wide area network (WAN), neighborhood area network (NAN) and home area network (HAN). The communication infrastructure of AMI is shown in Fig. 1. The end user devices with communication network at customer locality increases the probability of attacks on NAN of smart grid. The ramification of any cyber attack on the grid would be devastating. Therefore, intrusion detection systems (IDS) have gained much attraction for smart grid security

Security mechanisms implemented in other networks are not suitable for smart grid because of the difference in topology, protocol and network architecture of smart grid communication [1]. Attacks like Man-in-the-middle and denial of service causes' disturbance to power distribution may cause severe damage to grid. Hence, the earlier identification of such attacks before it affects the grid is essential. Intrusion detection system (IDS) acts as a first layer of defense that detects intrusions by analyzing each packet of traffic data. Machine learning algorithms are used as classifiers for detection of attacks in IDS [4-6]. In recent years, Support vector machine(SVM) [16] gets much importance that efficiently classifies the data at rapid speed in various applications. It is a multilayer perceptron network with atleast single hidden layer that has fast learning speed and good efficiency.

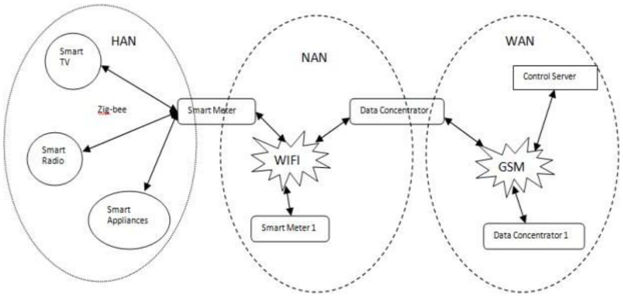

Fig. 1. Advanced Metering Infrastructure of Smart grid

The traffic data of NAN may have noisy and irrelevant features that will directly affect the performance of classifiers like SVM. Dimensionality reduction techniques improve the attack detection rate by selecting the informative features of network traffic as input to classifier. Feature selection methods, one of the dimensionality reduction techniques, selects the informative features from the original variables. Feature selection approaches are of two types [9]; filter and wrapper method. In filter method, the informative features are selected by finding the relation between input data and output whereas in wrapper method, the classification algorithms are also used in the selection of optimal features. The wrapper method can be further improved by using Evolutionary computation techniques like genetic algorithms, particle swarm optimization, whale optimization algorithm, etc.

Genetic algorithm (GA) is a generalized search technique that can be applied to select the informative features from the original dataset. In most of the existing GA- based feature selection techniques, the accuracy of classifier is used as fitness function for identifying the relevant features but in actual case multiple objectives need to be considered while selecting the relevant features. In this work, minimizing number of informative features and maximizing accuracy of classifier are used as objectives for improving the performance of IDS.

Since the objectives are conflict in nature, the feature selection problem in IDS is modeled as multi-objective optimization problem (MOP). The result of MOP is a set of pareto optimal solutions and the plot of pareto optimal solution vectors is called pareto front. The pareto front is found by decomposing the problem into a number of single objective optimization problems and the result of all problems will be aggregated to find the final Pareto front [9]. Alternatively, Evolutionary techniques can be applied to solve the multi-objective problems by considering the problem as a whole instead of decomposition.

Multi-objective genetic algorithm (MOGA) [9] finds the Pareto front in a single run. It is similar to conventional genetic algorithm and its difference lies in the evaluation of fitness of individuals. In MOGA, niching technique is used to find the non dominated solutions that help to find the fitness of individuals. MOGA is used to solve MOP in various domains like voltage stability enhancement in power system [5], real time database system [6], intrusion detection system, etc. In [7], MOGA based IDS is designed for a classification problem and is implemented in three stages with multilayer Perceptron (MLP) as classifier. The objectives used to solve the MOP in this work are indirectly related to attack detection rate and is not suitable for IDS in smart grid.

In this paper, MOGA is applied for the multi-objective feature selection problem with conflicting objectives. The distance based method given in [12] is used to select the pareto value from the non dominated solutions of combined objectives. Support vector machineis used as classifier in this work.

II. PROPOSED SYSTEM

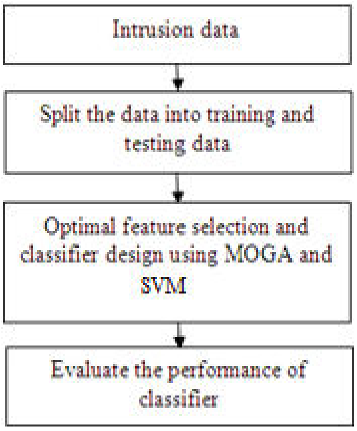
Intrusion data
Split the data into training and testing data
Optimal feature selection and classifier design using MOGA and SVM
Evaluate the performance of classifier

Fig. 2. Block diagram representation of proposed IDS

Fig. 2 shows the block diagram representation of proposed IDS. In this work, the Support vector machines used as classifier for the detection of intrusions from the traffic data. The classifier is trained by training data and is evaluated by testing data. The performance of classifier in the detection of attacks is affected from the noisy and irrelevant features present in the data. Hence, the finding of informative features from traffic data is essential to enhance the performance of classifier. In this work, the process of optimal feature selection is represented as an optimization problem with number of informative features and accuracy of classifier as objectives and MOGA is applied to solve this multi-objective optimization problem. It helps to increase the efficiency of IDS by selecting the optimal subset of features. MOGA selects the set of informative features from training data and is evaluated by SVM classifier. The fitness level of each subset of features in the GA population is evaluated on the basis of each objective and is used to calculate the scaled fitness value. The ranking operation is used to select the pareto optimal solutions from the non dominated solutions and the diversity in the selected solutions is achieved by using niching technique. The pareto front is found for the problem that has a set of solutions and the user can select the preferred solution based on the requirements.

III. PROBLEM STATEMENT

The performance of intrusion detection system depends on the informative features used to train the classifier. Selection of informative features from the dataset is necessary before giving it to machine learning algorithms. In this work, the feature selection is formulated as an optimization problem with maximizing accuracy of classifier and minimizing number of features as objectives. This is mathematically stated as,

F1 = Maximize (accuracy of the classifier) (1)

F2= Minimize (Number of informative features as input to classifier) (2)

In this work, MOGA, an extended form of GA is applied to solve this optimization problem. The selection of pareto optimal solution using MOGA based feature selection and SVM classifier for IDS is given in the subsequent sections.

IV. MULTI OBJECTIVE GENETIC ALGORITHM

The optimization problems with multiple objectives are called as MOP. Mathematically, the MOP is stated as [9],

$$F(x) = \min \{f_1(x), f_2(x), \ldots\ldots f_m(x)\}, x \epsilon S \qquad (3)$$

Where, $S - \{x \epsilon R^m : h_j(x) >= 0, g_k(x) >= 0\}$

R^m – Objective space, m – number of objective function

$g_k(x)$ – k^{th} Equality constraints, $h_j(x)$ – j^{th} Inequality constraints

The set of solutions obtained in a MOP are called pareto solutions and the set is represented as pareto front.

Evolutionary algorithms are widely used to solve multi-objective optimization problems. Genetic algorithm with necessary modifications called as multi-objective GA has been used to solve multi-objective optimization problems [12]. GA is a meta heuristic search algorithm that mimic the nature of genetics. In GA, the possible solution to the given problem is represented as individual. The set of individuals with different characteristics are called as population and each individual is evaluated using fitness functions. Then the good individuals from the population are selected as best individual for next generation and the new individuals are generated by applying mutation and crossover operators.

These steps are repeated till the optimal results are obtained. GA with necessary modification can be applied for solving MOP. It efficiently finds pareto optimal solutions by maintaining diversity in the population using niching techniques and pareto based ranking. The flowchart in Fig. 3 shows the working of MOGA.

First the objective function value of each individual are evaluated using SVM classifier. Then the dominance of each solution is found from the objective functions of individuals.

The algorithm for finding the dominance of solutions is given below,

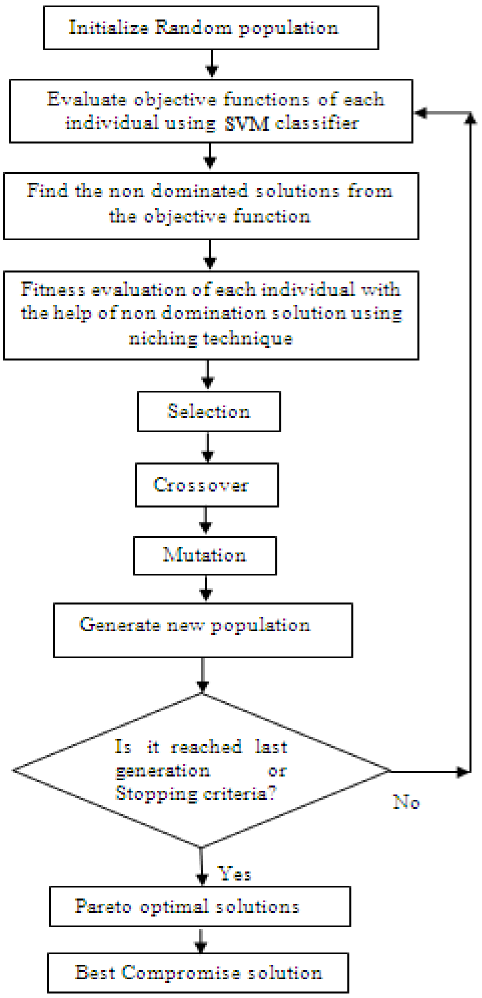
Initialize Random population
Evaluate objective functions of each individual using SVM classifier
Find the non dominated solutions from the objective function
Fitness evaluation of each individual with the help of non domination solution using niching technique
Selection
Crossover
Mutation
Generate new population
Is it reached last generation or Stopping criteria?
No
Yes
Pareto optimal solutions
Best Compromise solution

Fig. 3. Flowchart of the proposed MOGA

Input: Array of N objective functions as {(F ^{1}, F ^{1}), … (F ^{n}, F ^{n})}, i - current solution , j – all the remaining solutions

Output: S-Set of non dominant solutions

Step 1: Select the first solution as current solution i and assume S=0

Step 2: Select all the other solutions as j ie., j is not equal to i

Step 3: Check the conditions (F ^{j} < F ^{i}) and (F ^{j} < F ^{i}) for all j

Step 4: If both the conditions are true, S = S + 1

Step 5: N_i = Total number of solutions - S

Step 5: Select the next solution as current solution i

Step 6: Repeat the steps for all the solutions in the array of size N.

After finding the non dominant solutions, the ranking for the solution is introduced and is represented as,

$$R_i = 1 + N_i \qquad (4)$$

Where, R_i – Rank of the solution 'i', N_i – Number of solutions dominate i.

The non dominated solution gets the first rank and all other solutions are ranked based on its dominated position. After ranking, the solutions are arranged from lower to higher order and the fitness is assigned to each pareto solution with the help of linear mapping function. Thereafter, the solutions of each rank are averaged and are named as assigned fitness. The mapping function and averaging process makes sure the better ranked solution as top ranked output. After the fitness is calculated, the shared value of two different solutions 'i' and 'j' is need to be calculated for finding the niche count value n_i.

$$Sh(d_{ij}) = \begin{cases} 1-(d_{ij}/\sigma), & \text{if } d_{ij} < \sigma\,; \\ 0, & \text{otherwise.} \end{cases} \qquad (5)$$

$$d_{i,j} = \sqrt{\sum_{k=1}^{M} \left(\frac{f_k^i - f_k^j}{f_k^{max} - f_k^{min}} \right)^2} \qquad (6)$$

Where, Sh - Sharing function values of i and j

D_{ij} - the distance between the solution i and j

σ_s - Sharing parameter having the greatest distance in the same niche

f_k - function of kth objective α - 1

The niche count value for each solution is found to maintain diversity among the solutions of each rank. This calculation of n_i is called niching operation and is implemented by summing all the sharing function values.

$$nc_i = \sum_{j=1}^{\mu(r_i)} Sh(d_{ij}) \quad (7)$$

Where, nc - the number of ranking solutions

Finally, the shared fitness value is calculated from the fitness of a solution by corresponding niche count. Each shared fitness value has different levels of fitness and the fitness value of solutions in less crowded region has the optimal solution. This procedure is repeated till processing all ranks. In this work, tournament selection is used to select the individuals with high fitness value and two point crossover is used to generate new individuals. The non-uniform mutation operator is applied to make random changes in the population. The procedure is repeated till maximum number of generations or the stopping criterion is reached and the pareto front is obtained as output.

From the pareto front, user can select the best solution based on the requirement. Due to the user inability of selecting the exact best solution, fuzzy set theory is used to select the best solution. It assigns a linear membership function for each objective and is represented as,

$$m_i = \begin{cases} 1, & F_j \geq MF_j \\ \frac{MF_j - F_j}{MF_j - MNF_j} & MF_i < F_i < MNF_i \\ 0, & F_j \leq MNF_j \end{cases}$$

Where, MF_j and MNF_j are the maximum and minimum value of the j^{th} objective function. The equation maps the solution between 0 and 1. The membership function for each solution is calculated by,

$$m^k = \frac{\sum_1^N m_i^k}{\sum_{k=1}^M \sum_{i=1}^N m_i^k}$$

Where N is the number of objectives and M is the one number of non dominated solutions.

Finally, the solution having maximum membership function is selected as best compromise solution.

V. CLASSIFICATION USING SUPPORT VECTOR MACHINE

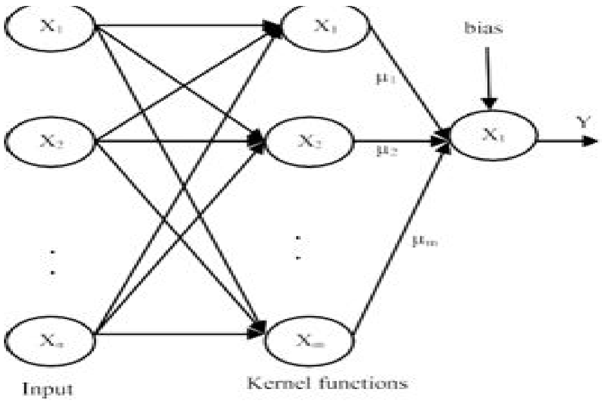

Fig. 4. support Vector Machine Architecture

SVM classifier is a supervised learning algorithm based classifier that is designed with the basic idea of separating class with a straight line given by the Equation

$$f(x) = W^T X + b = 0$$

where W is weight vector and b is bias. In complex datasets, classification is performed by curve instead of line. Thus for achieving the classification in com- plex dataset, the input data is initially mapped into feature space and is linearly separated by a separable hyperplane.

The mapping operation is carried out by 'kernel' function and choosing different kernel functions will produce different convergence rate. Linear, Quadratic, Polynomial, Multilayer Perceptron (MLP) and Gaussian are some of the widely used kernel func- tions in SVM classifier. The linear kernel function has low accuracy in the classification of complex datasets where some data cannot be separated linearly. So, the other kernel functions are used in most cases. The architecture of SVM, where X is the input data having 'n' number of features and is given to 'm' kernel functions. In this work, MLP kernel function [14] is used to map the input data into kernel space for finding the maximum margin hyper- plane. The MLP kernel function of SVM has number of hidden layers that focus on the minimization of error functions with the help of support vector co- efficient 'k' and bias value '0'. The kernel function maps the input data into the default scale value of [+1, –1]. The training of hidden layers for the input data X_i is represented as,

$$c_i = \frac{\sum_{m=1}^{n_i} x_m^i}{n_i}$$

The minimization of error tunes the hidden layer margin which is reflected in maximal margin of hyper plane. The training is repeated till each output data are correctly matched. In general, the separable line with global margin (2/ w) is selected by most of the classifiers for classifying data with maximum accuracy [15].

SVM classifier is generally designed from known labeled training data and its accuracy is analyzed using testing data. The Multiclass SVM is proved to be as accurate in the classification of multiple classes than any other technique. The number of classifiers used in Multiclass SVM depends on the number of classes in the dataset. In this model each classifier is trained by the dataset with consideration of classifier's corresponding class as '1' and all other classes as '0'. The classifier trained with such kind of classification dataset produces high accuracy. The individual classifier output is noted for all dataset and the final prediction is based on the confidence value for each class. Mathematically it is represented as,

$$\text{class} = \underset{i=1...n}{\text{arg max}} \ \text{of } g_i$$

where n represent the number of classes, g_i is the individual value of ith classifier and class is the output of high confidence value classifier. Once the classifier is trained and tested, they are ready to classify the data in real environment.

SVM classifier has good capability on dealing with nonlinear data that makes it as most suitable for intru- sion detection systems [16]. SVM based IDS has excellent performance by implementing in various orders like distributed [13, 16], hierarchical [17], etc. The usage of SVM in IDS requires extensive memory because of large dimensionality in network datasets and also affects the accurate detection ratio.

VI. EXPERIMENTAL RESULTS AND EVALUATION

In this section, the performance of proposed MOGA based IDS for NAN of smart grid is evaluated using the standard intrusion dataset. The proposed system is coded in Matlab 2014a and is evaluated by the datasets in a system having INTEL I3 processor with 2GB RAM.

The performance of proposed MOGA with ELM model is further validated by standard UNSW_NB15 dataset. Although the proposed system is evaluated with experimental dataset, the standard dataset is used to validate the experimental process. In this work, an intrusion dataset UNSW_NB15 dataset generated by the UNSW repository is used for the evaluation purpose. It has ten different classes of data of which 9 are attacks and 1 normal class data. The attacks relevant to NAN of smart grid are considered in this work as 7 and the features are reduced from 49 to 45 for getting better results. In this work, 6388 data (Normal:3280, Reconnaissance:712, DoS:116, Fuzzers:919, Backdoor:17, Generic: 1337, Worms: 7) are selected from the dataset. The parameters assigned to get the optimal results are Number of generation=20, population size=20, crossover value=0.6 and mutation value=0.05. The pareto front obtained by the proposed MOGA+ELM for the population size of 20 using UNSW_NB15 dataset is shown in Fig. 4. From the pareto front, the solution with 86.07 accuracy and 16 informative features is select as best optimal solution and the feature subset of the solution are given in Table 1. It shows that MOGA with ELM classifier has select the optimal subset of features effectively and is suitable for IDS.

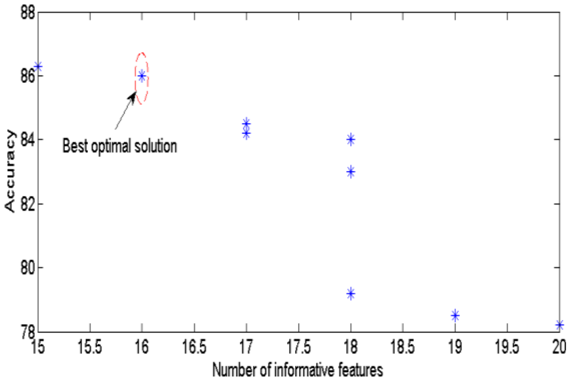

Fig. 5. Pareto front of MOGA + SVM on UNSW_NB15 dataset with population size of 20

Table- I: Details of the best optimal solution obtained from the pareto front

Number of features	Feature subset	Accuracy
16	2, 3, 4, 8, 10, 11, 13, 16, 21, 22, 23, 25, 27, 28, 29, 33	89.04

The Attack detection rate and the execution time of different classifiers with MOGA are given in Table. II and III. It shows that the proposed system has better performance than RF and LR classifiers.

Table- II: Attack detection rate of MOGA with different classifiers

Techniques	**Accuracy**
MOGA + SVM	87.57
MOGA + RF	81.34
MOGA + LR	62.76

Table- III: Execution time of MOGA with different classifiers

Techniques	**Time (Seconds)**
MOGA + SVM	38.37
MOGA + RF	1432
MOGA + LR	897.12

From the results, MOGA with SVM classifier has better performance than Random forest classifier and in the aspect of execution time SVM clearly dominates RF with large gap. Similarly, LR classifier executes faster than RF but is slower than SVM and has poor attack detection ratio. Thus, MOGA with SVM classifier selects the most optimal subset of features effectively and dominates RF and LR classifier in the intrusion detection system.

VII.CONCLUSION

In this paper, MOGA with SVM classifier based IDS is proposed with the objectives of maximizing accuracy and minimizing number of features for securing NAN of smart grid. Genetic algorithm is used to optimize this multi-objective problem and the optimal subsets of informative features are obtained by using support vector machine as classifier. It has been validated with single objective approach and UNSW_NB15 dataset. The simulation results verify the proposed system is suitable for identifying intrusions in NAN of Smart grid.

REFERENCES

1. Brycent Chatfield, Intrusion Detection for Smart Grid Communication Systems Spring 2017
2. Kasongo SM, Sun Y. A deep gated recurrent unit based model for wireless intrusion detection system. Cakovec: ICT Express; 2020.
3. R.Vijayanand, D. Devaraj A Novel Feature Selection Method Using Whale Optimization Algorithm and Genetic Operators for Intrusion Detection System in Wireless Mesh Network March 2020 IEEE PP(99):1-1.
4. G. Lu, D. De, and W. Z. Song, “Smartgridlab: A laboratory-based smart grid testbed,”in2010 First IEEE International Conference on Smart Grid Communications, Oct2010, pp. 143–148.
5. Y. Zhang, L. Wang, W. Sun, R. C. G. II, and M. Alam, “Distributed intrusion detectionsystem in a multi-layer network architecture of smart grids,”IEEE Transactions onSmart Grid, vol. 2, no. 4, pp. 796–808, Dec 2011.
6. Brethier. R, Sanders, W.H and Khurana. H. Intrusion Detection for Advanced Metering Infrastructure: Requirementsand Architectural Directions. In: First International Conference on Smart Grid Communications (SmartGridComm), Gaithersburg, IEEE Xplore, 2010, 350 –355.
7. Akkaya. K, Rabieh. K, Mahmoud. Mand Tonyali. S, Customized Certificate Revocation Lists for IEEE 802.11s-based Smart Grid AMI networks,IEEE Trans on Smart Grid,6(5)(2015), 2366-2374.
8. Cardenes. A, Berthier. R, Bobba. R.B, Huh. J.H, Jetc.heva J.G, Grochocki. D and Sanders, A framework for Evaluating Intrusion Detection Architectures in Advanced Metering Infrastructures, IEEE Transon smart Grid, 5(2) (2014) 906 –915
9. Bandyopadhyay, S. and Saha, S.: 2013,Some Single- and Multiobjective Optimization Tech-niques In Unsupervised Classification Similarity Measures, Classical and Metaheuristic Approaches, and Applications, Springer-Verlag, Berlin.
10. D’Cruz,N.,Radford,A.D.andGero,J.S.: 1983,AParetooptimizationformulationforbuildingperformance and design,Engineering Optimization,7(1), 17-33.
11. D. E. Goldberg, *Genetic algorithms in search, optimization and machine learning*, Boston: Addison-Wesley, 1989.
12. Q. Zhang and H. Li, “MOEA/D: A multiobjective evolutionary algorithm based on decomposition,” in IEEE Transactions on Evolutionary computation, vol. 11, no.6, 2007, pp. 712-731.
13. D. Subhadrabandhu, S. Sarkar and F. Anjum, A framework for misuse detection in ad hoc networks-part I, *IEEE Journal on Selected Areas in Communications* **24** (2006), 274–289.
14. J.A.K. Suykens, T. Van Gestel, De Brabanter, B. De Moor and J. Vandewalle, Least Squares Support Vector Machines, World Scientific, Singapore; (2002).
15. R. Collobert and S. Bengio, Link between preceptrons, MLP and SVMs, In: *2004 International Conference on Machine Learning* (2004), 23–31.
16. F. Kuang, W. Xu and S. Zhang, A novel hybrid KPCA and SVM with GA model for intrusion detection, *Applied Soft Computing* **18** (2014), 178–184.
17. V.S. Feng and S.Y. Chang, Determination of a wireless networks parameters through parallel hierarchical support vector machines, *IEEE Transactions on Parallel and Distributed Systems* **23** (2012), 505–512.

BIOGRAPHIES

Mrs. G. Manisha is a PG scholar in J.B. Institute of Engineering and Technology. Her research interest includes machine learning and Network Security.

Dr. R. VIJAYANAND is an Assistant professor in Department of Computer Science and Engineering, J.B.Institute of Engineering and Technology, Hyderabad, India.