International Journal of Advanced Information Science and Technology (IJAIST) ISSN: 2319:2682
Vol.3, No.11, November 2014 DOI:10.15693/ijaist/2014.v3i11.29-35

GSO-SVM for Feature Selection & Kernel Parameter Optimization

K. Joshil Raj
Department of Computer Science,
Pondicherry University,
Puducherry, India

S Siva Sathya
Department of Computer Science,
Pondicherry University,
Puducherry, India

Kalyan Nandi
Department of Computer Science,
Pondicherry University
Puducherry, India

***Abstract* – Taking cue from the natural phenomenon of producer-scrounger process in animals where a producer searches for the food and the scrounger looks for opportunities to join, a new population based optimization technique called Group Search Optimization (GSO) has come to fore in the recent past. Among the classification algorithms Support Vector Machine (SVM) as a novel method has found applications in many areas. Its accuracy is highly dependent upon the kernel parameters and the relevant feature subset selectionThis paper proposes the use of a hybrid GSO-SVM for feature selection which can select relevant feature subsets from the classification dataset and also optimize the kernel parameters of the SVM classifier so as to achieve maximum classification accuracy. Elimination of the insignificant and useless inputs leads to a simplification of the classification problem, thereby producing faster and more accurate systems. The aim is to achieve maximum detection accuracy and to minimize computational complexity. The GSO-SVM is thus useful for parameter determination and feature selection in the SVM. The quality and effectiveness of the proposed methodology has been evaluated on standard machine learning datasets.**

Keywords – Evolutionary algorithm, Feature Selection, Group Search Optimization, Kernel Parameters, Machine Learning, Support Vector Machine

I. INTRODUCTION

Support Vector Machines (SVM) are a class of classification algorithms which have been used in a wide number of applications like cancer morphologies [1], image classification [2], gene expression data classification, text categorization, object recognition [3][4] and pattern recognition based on their characteristic features like computational efficiency, high accuracy rate, excellent generalization performance, flexibility, and effectiveness in high dimensional data. The use of kernel parameters in SVM makes it capable of dealing efficiently with nonlinear features in high dimensional features space. The suitable selection of parameter values of kernel functions such as penalty coefficient C and other kernel parameters like gamma (γ) for the radial basis function(RBF) shapes the generalization ability of SVM.

Selection of optimal feature subset from the classification data is another factor that affects the classification accuracy other than the optimal kernel parameters selection. The feature selection refers to the selection of a subset of relevant features for model construction. The idea behind the features selection technique is to weed out the data that is having irrelevant or redundant features. Redundant features provide no extra information over the currently selected features while irrelevant feature provide no information at all. Hence, evidently the lack of ability to purge irrelevant features willaffect the system performance. It must be noted that the use of feature selection has no impact on the original representation of the features rather it just selects a subset of them. The introduction of feature selection however introduces a level of complexity in the modeling task.

The process of independently selecting features and learning SVM parameters as has been the case so far, might result in loss of information related to the classification process. This paper discusses a methodology of determining optimal parameter settings for the SVM classification procedure while performing the additional task of feature selection on the dataset for selecting the most optimal subset of significant features from a complete dataset. The feature subset selection & the kernel parameter optimization of the SVM are accomplished by the use of a population based optimization algorithm, Group Search Optimizer (GSO) [5]. Each member in GSO represents a subset of features & kernel parameter values of SVM. In each iteration, GSO members evolve to generate better fitness value (classification accuracy) for specific feature subsets & specific values of the parameters. The GSO member with the highest classification accuracy gives the optimal feature subset as well as the optimal kernel parameters values. Several public datasets are employed to calculate the classification accuracy rate in order to evaluate the developed methodology.

Tremendous amount of work has been carried out in SVM in order to optimize its classification accuracy. Srinivas et.al in

2003 [6] combined SVM and neural networks for feature selection and classification. They calculated the classification accuracy by removing the features individually. The reduced feature sets were ranked after testing them on SVM and neural networks and hence, the ones yielding the best detection rate in the experiments were deemed to be important feature subsets. The drawback of such methods was that they required scanning of the entire search space to find the most optimal solution, implying being computationally expensive. Numerical optimization methods like gradient descent were employed as a remedy to the feature selection but they had the problem of looping around in local optima and were dependent on the initial seeds. The optimization algorithms like the evolutionary algorithms such as genetic algorithms [7], ant colony optimization [8], simulated annealing algorithm[9, 10] and particle swarm optimization [11] were used for feature selection and also optimization of SVM parameters for these algorithms are known to have better global search abilities. Jack and Nandi [12] in 2002 utilized GA to choose the ideal set of features from a dataset. The chosen subset of peculiarities is then sustained into the SVM for classification. Samanta, Al-Balushi, and Al-Araimi [13] in 2003 proposed a GA methodology to change the RBF width parameter of SVM with feature selection. Their methodology just considered adjusting the RBF width parameter for the SVM. Zhang, Jack, and Nandi [14] in 2005 developed a GA-based methodology to find an advantageous subset of features for utilization of SVM for shortcoming identification in machine condition monitoring, however, their method did not address the issue of optimizing parameter values setting for SVM.Huang and Wang [7] in 2006 exhibited an alternate GA-based strategy for feature selection and parameters determination of SVM for utilization in credit scoring

The rest of the paper is organized as follows: Section II describes the GSO algorithm. The proposed feature selection & kernel parameter optimization of the SVM classifier with GSO is described in Section III and the experimental analysis is given in Section IV. Finally Section V concludes the paper.

II. GSO ALGORITHM

Group search optimizer (GSO) [5] is a population based optimization algorithm, inspired by animal searching (foraging) behavior. GSO Optimization employs the producer–scrounger (PS) model and the animal scanning mechanism. The population of the GSO is called a group, where each individual is called a member. A group consists of three types of members: producers, scroungers and rangers. Producers perform producing strategy in the way of animal scanning mechanism; scroungers perform scrounging strategy by joining resources uncovered by others; and rangers search for the randomly distributed resources by random walks. In each generation, the best fit member is treated as the producer, and a number of members except the producer in the group are selected as the scroungers, while the remaining members are regarded as the rangers.

In an n-dimensional search space, the ith member at the kth searching bout (iteration) has a position, $X_i^k \in \mathbb{R}^n$ and a head angle $\Phi_i^k = (\Phi_{i1}^k, \Phi_{i2}^k, \ldots \Phi_{i(n-1)}^k) \in \mathbb{R}^{n-1}$ .The search direction associated with the ith member, which is represented as a unit vector $D_i^k(\Phi_i^k) = (d_{i1}^k, d_{i2}^k, \ldots d_{in}^k) \in \mathbb{R}^n$ is presented below in equation(1).

$$\begin{cases} d_{i_1}^k = \prod_{q=1}^{n-1} \cos(\Phi_{i_q}^k) \\ d_{i_j}^k = sin(\Phi_{i_{(j-1)}}^k) . \prod_{q=j}^{n-1} \cos(\Phi_{i_q}^k) \; (j = 2, \ldots, n-1 \\ d_{i_n}^k = sin\left(\Phi_{i_{(n-1)}}^k\right) \end{cases} \quad \ldots\ldots (1)$$

The scanning field of vision is an n-dimensional space, which is characterized by maximum pursuit angle $\Theta_{max} \in \mathbb{R}^1$ and maximum pursuit distance $l_{max} \in \mathbb{R}^1$. In GSO, the producer X^k_pscans at zero degree and laterally by randomly sampling three points in the search space to find the best resource at each iteration. If the producer cannot find a better search position after 'a' iterations, it will turn its head back to zero degree.

Scroungers follow the producer adopting a random walk towards it as expressed in equation (2).

$$X_i^{k+1} = X_i^k + r \circ (X_p^k - X_i^k) \quad \text{------ (2)}$$

where $r \in \mathbb{R}^n$ is a uniform random sequence in the range (0,1). Operator "o" is the Hadamard product or the Schur product, which calculates the entry wise product of the two vectors. If a better position than the current producer is found by any of the scroungers then in the next searching bout it will switch to be a producer. This switching mechanism helps the group members to escape from local minima in the previous search bouts.Dispersed animals may adopt ranging behaviour to explore and colonize new habitats. In each generation, rangers move to the new point based on a random head angle and a random distance using the following equation (3):

$$\begin{cases} l_i = a . r_1 . l_{max} \\ X_i^{k+1} = X_i^k + l_i . d_i^k . (\Phi^{k+1}) \\ \Phi^{k+1} = \Phi^k + r_2 \alpha_{max} \end{cases} \quad \ldots\ldots (3)$$

Here, $r_1 \in \mathbb{R}^1$ is a normally distributed number with mean 0 and standard deviation 1 and $r_2 \in \mathbb{R}^{n-1}$ is a uniformly distributed random sequence in the range (0, 1).

III. PROPOSED IMPLEMENTATION: FEATURE SELECTION & PARAMETER OPTIMIZATION

Two main tasks being performed for improving the classification accuracy of SVM algorithm are feature selection and RBF kernel parameter optimization. These are being achieved by use of a novel nature inspired algorithm called GSO.

A. GSO based Feature Selection & Kernel Parameter Optimization

Each member in GSO represents a subset of features and the parameter values of RBF kernel of SVM algorithm. Thus each GSO member will have values in '2 + n' dimensions where C &γ are the two RBF kernel parameter settings in addition to the 'n' that represents the number of features. In each generation GSO members will evolve with producing, scrounging and ranging mechanism. The fitness function of each member is the SVM classification accuracy found with the corresponding feature subsets being used from the data & using the parameter values associated with that GSO member. After termination of GSO procedure, the producer (the best fit member) will indicate the optimal values of parameter settings C & γ and the feature subsets used in achieving the maximum classification accuracy.
The individual GSO member representation is as shown below:

C	γ	A_1	A_2	…..	…..	A_n

Where C &γ are the RBF kernel parameters settings & variable (A_1, A_2……A_n) are the features of the dataset. If the value of the variable (A_1, A_2…….A_n) is less than or equal to 0.5, then its corresponding feature is not chosen. Conversely, if the value of a variable is greater than 0.5, then its corresponding feature is chosen for making the feature subset. The searching range of parameter C is between 0.01 and 35,000 & for γ is between 0.0001 and 32.

B. Fitness Function

Classification accuracy and the number of selected features are the two criteria used to design a fitness function. Thus, for the GSO member with high classification accuracy and a small number of features produce a high fitness value. The member with high fitness value has high probability to affect the other members' positions in the next iteration. The fitness of i_{th} member is given by equation (4):

$$fitness_i = W_A \times acc_i + W_F \times \left[1 - \frac{\left(\sum_{j=1}^{nF} f_i\right)}{n_F}\right] \quad \text{'…….(4)}$$

Where w_A is the weight for the SVM classification accuracy. w_A is adjusted to 95%,
acc_i is the SVM classification accuracy.
w_F is the weight for the number of selected features. w_F is set to 5%.
f_i is the value of feature mask - '1' represents that feature i is selected and '0' represents that feature 'i' is not selected.
n_F is the total number of features.

The flow chart for the GSO based feature selection & parameter optimization is as shown in Figure 1:

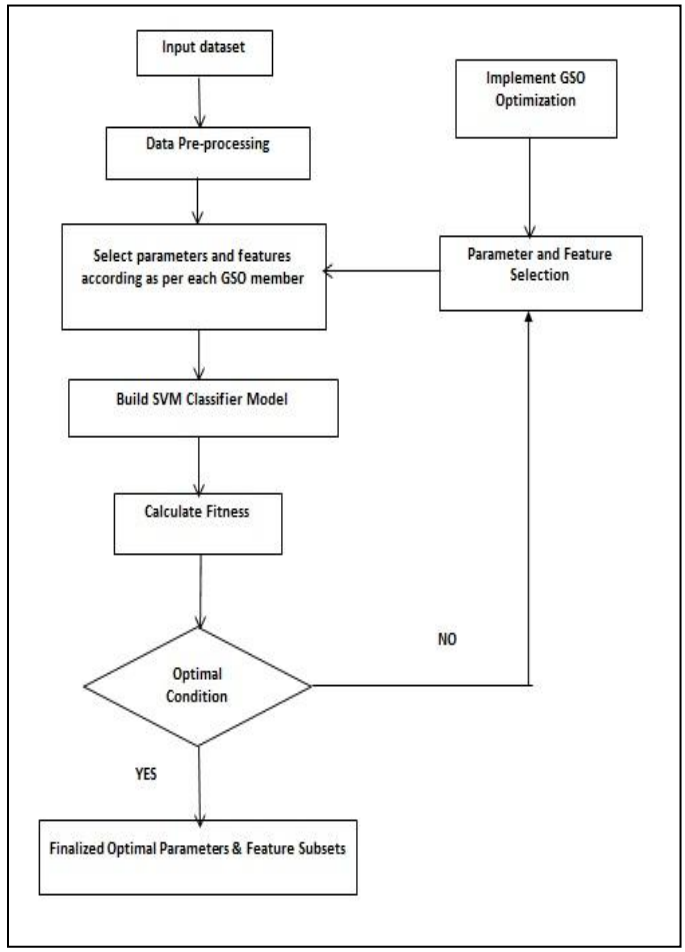

Figure 1: Flow chart flow chart for GSO based feature selection & parameter optimization

The pseudo code for the GSO based feature selection & parameter optimization is given below:

BEGIN
Input:Training Dataset
Input:Test Dataset

INITIALIZE positions X_i and head angles $\phi_i$of GSO members
CALCULATE the fitness values of initial members using SVM;

IV. EXPERIMENTAL ANALYSIS

The proposed methodology is implemented in Java using the open source java packages of WEKA 3.7.9 machine learning tool developed by Waikato University[16].Performance is evaluated on the UCI benchmark data sets.The datasets are available at the University of California at Irvine (UCI) Machine Learning repository database.

A. Experimental Settings

The parameter settings of GSO are summarized as follows. The initial population of GSO is generated uniformly at random in the search space. The initial head angle Φ^0 of each individual is set to be ($\pi/4$ . . . $\pi/4$). The constant 'a' is chosen as round ($\sqrt{m+1}$) where m is the dimension of the search space i.e. number of distinct classes in training dataset. The maximum pursuit angle $\Theta_{max}$is π/ a^2. The maximum turning angle α_{max} is set to be $\Theta_{max}/2$. The maximum pursuit distance l_{max} is calculated using equation (5) given below:

$$l_{max} = \| U - L \| = \sum_{i=1}^{m} (U_i - L_i)^2 \qquad \text{---- (5)}$$

In the proposed method U_i=1 and L_i=0 for i= {1, 2…... m}. In each generation, one best fit member is treated as the producer, 20% of the total GSO population are chosen randomly as rangers and remaining members will perform scrounge operation. The population size is taken as 48, the number of iteration as 80 in the proposed method.

B. Performance Evaluation

To measure the performance of the proposed Feature selection & parameter optimization of SVM classifier using GSO, the experiments were carried out on the following standard datasets: Australian,Breast Cancer, Corral, Ionosphere, Iris, Soybean, Sonar, Vehicle, Vote, Vowel from the UCI machine learning repository.The descriptions of the UCI datasets used in the experiments are given in Table 1.
The k-fold method proposed by Salzberg [15]wasemployed in this study. In this experiment, the value of kis set to 10. Thus, the dataset was split into 10 parts, witheach part of the data sharing the same proportion of eachclass of data.Nine data parts were applied in the training process, while the remaining one was utilized in the testing.The standard benchmark dataset is shown in Table 1. The selected feature subsets & the optimized parameter values for each of the standard benchmark dataset without feature selection and with feature selection are shown in Table 2 & Table 3 respectively. The standard benchmark dataset is shown in Table 1. These two tables also give the comparison with the classification accuracy achieved using the complete dataset & gives the number of features selected in thereduced feature subsets of each dataset. The optimized values of the RBF kernel parameters are also given in the table. The searching range of parameter C of SVM was between 0.01 and35,000, while the searching range of parameter γ of SVMwas between 0.0001 and 32.Table 4 gives the name of list of selected features for each of the dataset used in the experiments. The classification accuracy comparison between the complete dataset & the reduced dataset is depicted in the figure 2. The results obtained by the developed GSO_SVM approach for the following dataset: Australian,Breast Cancer, Ionosphere, Iris,Sonar, Vehicle, Vowel with and without feature selection were compared with the gridsearch , as shown in Table 5.

Table 1: Description of the UCI standard datasets used for experimental results

Dataset	No. of classes	No. of instances	No. of Features
Australian	2	690	14
Breast Cancer	2	699	9
Corral	2	128	6
Ionosphere	2	351	34
Iris	3	150	4
Soybean	19	683	35

Sonar	2	208	60
Vehicle	4	846	18
Vote	2	435	16
Vowel	11	990	13

Table 2: Classification accuracy(%) results without Feature Selection

Dataset	**Result for without Feature Selection**			
	No of features	***Value of Optimized Parameter***		***Accuracy (%)***
		γ	***C***	
Australian	14	1.0E-4	4704	86.52
Breast Cancer	9	1.0E-4	5537	96.84
Corral	6	1.96	9150	100
Ionosphere	34	0.4305	15103	95.44
Iris	4	1.0E-4	1419	97.33
Soybean	35	1.075	2998	93.85
Sonar	60	0.4108	3162	88.94
Vehicle	18	0.1836	2848	81.44
Vote	16	0.2124	6161	94.94
Vowel	13	1.057	22084	99.69

Table 3: Classification accuracy(%) results with Feature Selection

Dataset	**Result for with Feature Selection**			
	Selectedfeatures	***Value of Optimized Parameter***		***Accuracy (%)***
		γ	***C***	
Australian	5	10.0	8711	87.39
Breast Cancer	7	1.07	6650	97.13
Corral	4	2.44	9266	100
Ionosphere	17	0.6454	14138	96.58
Iris	3	1.0E-4	2013	97.33
Soybean	34	0.0799	12126	95.75
Sonar	32	3.003	14877	91.82
Vehicle	12	1.357	10334	83.38
Vote	6	3.35	7111	96.55
Vowel	9	1.959	24861	99.79

Table 4: List of Features selected for each dataset

Benchmark Dataset	**Total no of Features**	**No of Selected Features**	**Name of Selected Features**
Australian	14	5	A1, A7, A8, A9, A13
Breast Cancer	9	3	Clump_Thickness, Cell_Shape_Uniformity, Single_Epi_Cell_Size, Bare_Nuclei, Bland_Chromatin, Normal_Nucleoli, Mitoses
Corral	6	4	A0, A1, B0, B1
Ionosphere	34	17	a03, a06, a08, a11, a13, a16, a18, a19, a20, a21, a23, a25, a26, a30, a31, a32, a33
Iris	4	3	Sepallength, petallength, petalwidth
Soybean	35	34	Date, plant-stand, precip, temp, hail, crop-hist, area-damaged, severity, seed-tmt, germination, plant-growth, leaves, leafspots-halo, leafspots-marg, leafspot-size, leaf-shread, leaf-malf, leaf-mild, stem, lodging, stem-cankers, canker-lesion, fruiting-bodies, external-decay, mycelium, int-discolor, sclerotia, fruit-pods, fruit-spots, seed, seed-discolor, seed-size, shriveling, roots
Sonar	60	32	attribute_1, attribute_2, attribute_4, attribute_5, attribute_6, attribute_7, attribute_9, attribute_10, attribute_12, attribute_14, attribute_16, attribute_17, attribute_19, attribute_20, attribute_22, attribute_32, attribute_36,attribute_37, attribute_38, attribute_39, attribute_40,attribute_41, attribute_42,attribute_43, attribute_48,attribute_51, attribute_54,attribute_55, attribute_56,attribute_57, attribute_59,attribute_60
Vehicle	18	12	COMPACTNESS, CIRCULARITY, DISTANCE CIRCULARITY, PR.AXIS ASPECT RATIO, SCATTER RATIO, PR.AXIS RECTANGULARITY, MAX.LENGTH RECTANGULARITY, SCALED VARIANCE_MAJOR, SKEWNESS ABOUT_MINOR, KURTOSIS ABOUT_MAJOR,KURTOSIS ABOUT_MINOR, HOLLOWS RATIO
Vote	16	6	adoption-of-the-budget-resolution, physician-fee-freeze, mx-missile, immigration, education-spending, export-administration-act-south-africa.
Vowel	13		Train or Test, Speaker Number, Sex, Feature 0, Feature 1, Feature 3, Feature

Benchmark Dataset	Total no of Features	No of Selected Features	Name of Selected Features
			6, Feature 8, Feature 9

Table 5: Experimental results of the developed GSO- SVM approach with and without feature selection and grid search (%)

Dataset	GSO-SVM without Feature Selection	GSO-SVM with Feature Selection	Grid search
Australian	86.52	87.39	84.54
Breast Cancer	96.84	97.13	96.64
Ionosphere	95.44	96.58	93.08
Iris	97.33	97.33	96.00
Sonar	88.94	91.82	87.90
Vehicle	81.44	83.38	84.28
Vowel	99.69	99.79	98.91

IV. CONCLUSION AND FUTURE SCOPE

Appropriate feature subset selection and model parameter setting have a heavy impact on the classification accuracy. In this paper, a GSO based parameter optimization & feature subset selection for a SVM classifier is proposed. This new methodology is based on Group Search Optimizer, a population based optimization algorithm developed from the group search behavior of animals in search of food. The optimum values of RBF kernel parameters settings C &γ are selected in addition to reducing the dataset for classification by selecting the optimum feature subsets. Feature subset selection influences the selection of appropriate kernel parameters and vice versa [56], thus obtaining the optimal feature subset and SVM kernel parameters simultaneously. This methodology was tested on the UCI benchmark datasets and shows that classification accuracy has significantly increased, thereby proving this technique to be an effective tool for improving the overall classification performance. Results of this study have RBF kernel function into consideration, however similar procedure can be employed for optimizing other kernel parameters.

REFERENCES

[1] S. Mukherjee, P. Tamayo, D. Slonim, A. Verri, T. Golub, J. Mesirov, and T. Poggio. "Support vector machine classification of microarray data'', AI Memo 1677, Massachusetts Institute of Technology, 1999.

[2] Guyon, I., and Elisseeff, A., 2003, "An introduction to variable and feature selection'',*Journal of Machine Learning Research*, Vol. 3, pp. 1157-1182.

[3] Keerthi, S. S., & Lin, C.-J. (2003),"Asymptotic behaviors of support vector machines with gaussian kernel'',*Neural Computation*, 15, 1667–1689

[4] T. Evgeniou, M. Pontil, C. Papageorgiou, and T. Poggio,- "Image representations for object detection using kernel classifiers''.*In Asian Conference on Computer Vision*, 2000.

[5] S. He, Q. H. Wu, and J. R. Saunders, "Group Search Optimizer: An Optimization Algorithm Inspired by Animal Searching Behavior," *IEEE Transactions On Evolutionary Computation*, vol. 13,no. 5, pp. 973-990, October 2009.

[6] Mukkamala Srinivas, Sung Andrew H., "Identifying Important Features for Intrusion Detection Using Support Vector Machines and Neural Networks", *Proceedings of the 2003 Symposium on Applications and the Internet* (SAINT'03).

[7] C.-L. Huang, C.-J. Wang," A GA-based attribute selection and parameter optimization for support vector machine'', *Expert Syst.* Appl. 31 (2) (2006) 231–240.

[8] X. Zhang, X. Chen, Z. He, "An ACO-based algorithm for parameter optimization of support vector machines'', *Expert Syst.* Appl. 37 (9) (2010) 6618–6628.

[9] P.-F. Pai, W.-C. Hong, "Support vector machines with simulated annealing algorithms in electricity load forecasting'', *Energy Conversion Manage*. 46 (2005) 2669–2688

[10] P.-F. Pai, W.-C. Hong, "Software reliability forecasting by support vector machines with simulated annealing algorithms'', *J. Syst. Softw.* 79 (2006) 747–755.

[11] X.C. Guo, J.H. Yang, C.G. Wu, C.Y. Wang, Y.C. Liang, "A novel LS-SVMs hyper- parameter selection based on particle swarm optimization'', *Neurocomputing* 71 (2008) 3211–3215.

[12] Jack, L. B., & Nandi, A. K. (2002)., "Fault detection using support vector machines and artificial neural networks, augmented by genetic algorithms",. *Mechanical Systems and Signal Processing*, 16, 373–390

[13] B. Samanta, K.R. Al-Balushi, S.A. Al-Araimi, "Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection'', *Eng. Appl. Artif. Intell.* 16 (2003) 657–665.

[14] L. Zhang, L.B. Jack, A.K. Nandi, "Fault detection using genetic programming'', *Mech. Syst. Signal Process*. 19 (2005) 271–289.

[15] Salzberg, S. L. (1997), "On comparing classifiers: pitfalls to avoid and a recommended approach'', *Data Mining and Knowledge Discovery*, 1, 317–327.

[16] WEKA software, Machine Learning, The University of Waikato,Hamilton,NewZealand.url:http://www.cs.waikato.ac.nz/ml/weka

Authors Profile

K JOSHIL RAJ received the **B.Tech.**degree in Electrical Engineering from the College of Military Engineering, Pune, India, in 2004 &**M.Tech.** degree in Computer Science & Engineering from Defence Institute of Advance Technology, Pune, India, in 2011. He is currently pursuing **Ph.D.**in Computer Science &

Engineering from Pondicherry University, Pondicherry, India. His research interests include Computational Intelligence, Bio Inspired Algorithms, Cyber Security, Neural Networks, fuzzy logic, Remote sensing & GIS.

Dr. S. Sivasathyareceived the **B. Tech.** degree in Computer Science & Engineering from Pondicherry University, Pondicherry, India, in 1993. She received her **M.Tech.** and**Ph.D.** degree in Computer Science from the same university in 1999 and 2009, respectively. She is currently an Associate Professor in the Department of Computer Science, School Engineering and Technology in Pondicherry University, Pondicherry, India. Her research interests include Genetic Algorithms, Bioinformatics, Intrusion Detection & Response, Computer Based Music Analysis.

Kalyan Nandireceived the **B.Sc.**degree in Computer Applications(Major) from the Ramakrishna Mission Vidyamandira College, Belur, Calcutta University, West Bengal, India, in 2011&**M.Sc**. degree in Computer Science from Pondicherry University, Pondicherry, India. He is currently doing **M.Tech.** in Computer Science and engineering fromPondicherry University, Pondicherry, India. His research interests includeBio Inspired Algorithm, Machine learning, Distributed System&Cloud Computing.