Classifier Ensemble Design using Artificial Bee Colony based Feature Selection

Shunmugapriya Palanisamy[1] and Kanmani S[2]

Research Scholar, Department of Computer Science and Engineering,
Pondicherry Engineering College, Puducherry, India. 605014

Professor, Department of Information Technology,
Pondicherry Engineering College, Puducherry, India. 605014

Abstract

Artificial Bee Colony (ABC) is a popular meta-heuristic search algorithm used in solving numerous combinatorial optimization problems. Feature Selection (FS) helps to speed up the process of classification by extracting the relevant and useful information from the dataset. FS is seen as an optimization problem because selecting the appropriate feature subset is very important. Classifier Ensemble is the best solution for the pitfall of accuracy lag in a single classifier. This paper proposes a novel hybrid algorithm ABCE – the combination of ABC algorithm and a classifier ensemble (CE). A classifier ensemble consisting of Support Vector Machine (SVM), Decision Tree and Naïve Bayes, performs the task of classification and ABCE is used as a feature selector to select the most informative features as well as to increase the overall classification accuracy of the classifier ensemble. Ten UCI (University of California, Irvine) benchmark datasets have been used for the evaluation of the proposed algorithm. Three ensembles ABC-CE, ABC-Bagging and ABC-Boosting have been constructed from the finally selected feature subsets. From the experimental results, it can be seen that these ensembles have shown up to 12% increase in the classification accuracy compared to the constituent classifiers and the standard ensembles Bagging, Boosting, ACO-Bagging and ACO-Boosting.

Keywords: *Feature Selection, Classification, Classifier Ensemble, Ant Colony Optimization, Bee Colony Optimization, Artificial Bee Colony, Meta-heuristic search.*

1. Introduction

Ensemble Learning has been a great topic of research during the last decade and vast amount of works have been carried out in the domain of Classifier Ensemble (CE) [2] and [3]. Classifier Ensemble is the combination of two or more classification algorithms and it is formed as a best solution to overcome the limitation of accuracy lag of a single classifier. Bagging, Boosting, Stacking, Majority Voting, Behavioral Knowledge Space and Wernecke's are some popular ensemble techniques. When the constituent classifiers in CE are of same type, it is called homogeneous CE otherwise heterogeneous CE [2] and [3].

In CE, each constituent classifier is trained over the entire feature space. Sometimes the feature space is noisy consisting of irrelevant and redundant data [1]. In such cases, the classifier consumes more time to get trained and also its misclassification rates are higher. Feature Selection (FS) is a possible solution to this problem. Feature Selection extracts the necessary and relevant data from the feature space without affecting the originality of its representation. With FS, the performance of the classifier is improved, thereby improving the efficiency of the ensemble [10].

Feature Selection has been widely used in the construction of ensembles [2]. While employing FS for ensemble construction, results would be better when FS is optimized. Swarm and evolutionary algorithms are used for optimizing feature selection resulting in optimal feature subset. In literature, Genetic Algorithm, Ant Colony Optimization, Bee Colony Optimization and Particle Swarm Optimization are used in numerous applications for optimizing FS [8], [9], [17] and [18]. ABC is a stochastic, swarm intelligent algorithm proposed by Karaboga Et. al. for constrained optimization problems [4]. Since its proposal, ABC has been proved to be successful in solving optimization problems in numerous application domains. Also ABC is proved to give promising and enhanced results in the areas where Genetic Algorithm and ant Colony Optimization have given already [4], [5], [6] and [7].

In order to enhance the classification accuracy, different algorithms for pattern classification [1], different techniques for feature selection and a number of classifier ensemble methodologies [2] and [3] have been proposed and implemented so far. The main limitation of these methods is that, none of them could give a consistent performance over all the datasets [8]. The proposed method is also attempted as an effort towards efficient feature selection optimization and ensemble construction.

In this study, classifier ensembles are constructed using optimal feature subset obtained from the combination of classifier ensemble and Artificial Bee Colony Algorithm (ABC). ABC is used to select the features and generate the feature subsets and these feature subsets are evaluated for efficiency by an ensemble made up of classifiers Decision Tree (DT), Naïve Bayes (NB) and Support Vector Machine (SVM). Each time ABC generates different feature subsets, the CE uses the average of mean accuracy of the ensemble and consensus as the fitness measure to select the feature subset.

This paper is organized in six sections. Section 2 describes about feature selection and its types. In Section 3, a brief description of Artificial Bee Colony is presented. Section 4 outlines the proposed method ABCE: the ABC based feature selection and the ensemble construction are explained in this section. The experiments and results are discussed in section 5 and the paper is concluded in section 6.

2. FEATURE SELECTION (FS)

Feature selection is viewed as an important preprocessing step for different tasks of data mining especially pattern classification [10], [11] and [12]. When the dimensionality of the feature space is very high, FS is used to extract the informative features from the feature space and the uninformative ones will be removed. Otherwise the uninformative features tend to increase the complexity of computation by introducing noisy and redundant data into the process. With FS, the features are ranked based on their importance making the feature set more suitable for classification without affecting the original feature representation and accuracy of prediction [10]. It has been proved in the literature that "classifications done with feature subsets obtained by FS have higher prediction accuracy than classifications carried out without FS" [19].

A number of algorithms have been proposed to implement FS. FS algorithms related to pattern classification fall into two categories: Filter Approach and Wrapper Approach. When the process of FS is independent of any learning algorithm, it is called filter approach. It depends on the general characteristics of the training data and uses the measures such as distance, information, dependency and consistency to evaluate the feature subsets selected [10] and [20]. On the other hand when a classifier is involved, it is called wrapper approach. The feature subset that results from a wrapper approach depends on the classification method used and two different classifiers can lead to two different feature subsets. Compared to filter approach the feature subsets obtained through wrapper are always effective subsets but it is a time consuming process [20] and [24]. Independent of filter and wrapper approaches, evolutionary algorithms are also used for searching the best subset of features through the entire feature space [8], [9], [17] and [18].

3. The Artificial Bee Colony Algorithm (ABC)

ABC is a swarm intelligent and meta-heuristic search algorithm proposed by Karaboga [4] and since then it has been used widely in many fields for solving optimization problems [5], [6], [7], [17] and [18]. ABC is inspired by the foraging behavior of honey bee swarms. The ABC algorithm employs three types of bees in the colony: employed bees, onlooker bees and scout bees. Initially, the food source positions are generated (N) and the population of employers is equal to the number of food sources. Each food source represents a solution of the optimization problem. Each employed bee is assigned a food source and they exploit the food sources and pass the information of nectar content to the onlookers. The number of onlookers is equal to the number of employed bees. Based on the information gained, the onlookers exploit the food sources and its neighborhood until the food sources become exhausted. The employed bee of exhausted food sources becomes a scout. Scouts then start searching for new food source positions. The nectar information represents the quality of the solution available from the food source. Increased amount of nectar increases the probability of selection of a particular food source by the onlookers [5]. The ABC algorithm is given in Fig.1 [4].

1. Initialize the food source positions
2. Evaluate the food sources
3. Produce new food sources(solutions) for the employed bees
4. Apply greedy selection
5. Calculate the fitness and probability values
6. Produce new food sources for onlookers
7. Apply greedy selection
8. Determine the food source to be abandoned and allocate its employed bee as a
   Scout for searching the new food sources
9. Memorize the best food source found
10. Repeat steps 3-9 for a pre-determined number of iterations

Fig.1 Steps of the ABC algorithm

4. The Artificial Bee Classifier Ensemble (ABCE)

In the proposed study, the ensemble is constructed as a combination of ABC algorithm and a CE consisting of the classifiers Decision tree, Naïve Bayes and Support Vector Machine [1] and [16]. In ABCE, ABC algorithm is used as a feature selector and feature subset generator; the ensemble classifier is used as the evaluator to evaluate the feature subsets generated. The classifier ensemble helps ABC in picking up the best feature subset by evaluating each configuration suggested by ABC. The ABC algorithm helps in efficient CE construction suggesting the best feature subset for the ensemble to work with. Hence both ABC and CE try to enhance the performance of each other in the proposed method. The steps of ABCE are given in Fig.2.

4.1 ABC Feature Selector

The ABC algorithm is used to optimize the process of feature selection and increases the predictive accuracy of the classifier ensemble. First, the classifier ensemble (made up of DT, SVM and NB) is used to evaluate the discriminating ability of each feature F_i in the dataset. Then, the ensemble accuracy (x_i) of each feature F_i is calculated by employing 10-fold cross validation [2] and [3] for each of the classifier.

Each employed bee is assigned a binary bit string made of '0's and '1's. The length of the binary bit string is equal to the number of features in the dataset and is used to represent the feature selection by each employed bee. A '1' means the feature is selected and a '0' means the feature is not selected. The population of the employed bees and onlooker bees are equal to the feature size (m) of the dataset as features are considered as food sources here.

1.Cycle =1
2. Initialize ABC parameters
3. Evaluate the fitness of each individual feature
4. Repeat
5. Construct solutions by the employed bees
    For *i* form 1 to *m*
    - Assign feature subset configurations (binary bit string) to each employed bee
    - Produce new feature subsets v_i
    - Pass the produced feature subset to the Classifier Ensemble
    - Evaluate the fitness (fit_i) of the feature subset based on the ensemble's mean accuracy and consensus
    - Calculate the probability p_i of feature subset solution

6. Construct solutions by the onlookers
    For *i* form 1 to *m*
    For *j* form 1 to *m*
    - Select a feature based on the probability p_i
    - Compute v_i using x_i and x_j
    - Apply greedy selection between v_i and x_i

7. Determine the scout bee and the abandoned solution
8. Calculate the best feature subset of the cycle
9. Memorize the best optimal feature subset
10. Cycle = Cycle + 1
11. Until pre-determined number of cycles is reached
12. Employ the same searching procedure of bees to generate the optimal feature subset configurations
13. Construct the ensembles ABCE, ABC-Bagging and ABC-Boosting using the best optimal feature subset

Fig. 2 Steps of ABCE

Each employed bee is allocated a feature and it evaluates the fitness of the feature by using the mean accuracy of the ensemble and the consensus [8]. The fitness for each feature (feature subset) pointed by the employed bee is calculated using the equations (1), (2) and (3).

$$fitness_1(s) = \frac{\sum_{j=1}^{m} accuracy_j(s)}{m} \tag{1}$$

$$fitness_2(s) = consensus(s) \tag{2}$$

$$fit_i = \frac{fitness_1(s) + fitness_2(s)}{2} \tag{3}$$

$accuracy_j(s)$ is the predictive accuracy of the j^{th} classifier in the ensemble and $consensus(s)$ specify the classification accuracy using consensus upon the s^{th} feature subset [8]. The first part of the fitness (mean accuracy) checks whether the feature subset has superior power on accurate classification with the whole classifier ensemble and targets to optimize it. So, the mean accuracy helps in increasing the generalization ability of the feature subset. The second part of the fitness (consensus) checks for the optimality of the feature subset in producing high consensus classification [8].

The onlooker bee gains information from the employed bee and calculates the probability of selecting a feature using equation (4). Then the onlooker computes the new solution v_i using the ensemble accuracies of the feature the employed bee is pointing to and the feature the onlooker bee has selected. If the new solution v_i is greater than x_i , the employed bee will be pointing to feature subset consisting of the feature it was previously pointing and the newly selected feature. If v_i is not greater than x_i then, the employed bees feature will be retained and the newly selected feature is neglected. The new solution v_i is computed by using equation (5).

$$p_i = \frac{fit_i}{\sum_{i=1}^{m} fit_i} \tag{4}$$

$$v_i = x_i + \varphi_i (x_i - x_j) \tag{5}$$

where, x_i is the ensemble accuracy of the feature allocated to the employed bee and x_j is the ensemble accuracy of the feature the onlooker has selected. φ_i is an uniformly distributed real random number in the range [0,1]. This way, each time the employed bee is assigned a new feature subset, the onlooker exploits and tries to produce new feature subset configuration.

After all possible features are exploited for forming the feature subset, the nectar content gets accumulated towards better feature subset configuration. If any employed bee has not improved, then the employed bee becomes a scout. The scout is assigned a new binary feature set based on the equation (6).

$$x_i^j = x_{\min}^j + rand[0,1](x_{\max}^j - x_{\min}^j) \tag{6}$$

Where x_{max}^j and x_{min}^j represents the lower and upper bounds of the dimension of the population. The bees keep executing the same procedure for a pre-determined number of runs to form the best feature subset.

Hence ABC is used to select and rank different features based on their importance. So, relevant features are extracted and the computation complexity due to irrelevant and noisy features is greatly reduced. Apart from this, for large datasets especially with large number of features, the performance of classifiers is affected because it has to handle more number of features. By using ABC, the number of features is scaled down based on their importance and the computational speed of the classifier is increased.

4.2 The Ensemble Classifier

The classifiers Decision Tree, Naïve Bayes and SVM are put together to form the classifier ensemble in the proposed method. In the proposed study, when the bees keep executing their searching procedure, the feature subset selected by each of the bees is input to the classifier ensemble. The three classifiers consider the candidate feature subsets one at a time, get trained with the combination of features and classify the test set. After the classifiers have finished, the ABC algorithm calculates the mean accuracy of the classifier ensemble and consensus using equations (1) and (2). The fitness is then calculated as the average of mean accuracy and consensus. The fitness *(fit_i)* is used as the evaluation criterion for selecting the best feature subset combination.

In the proposed method, the classifier ensembles ABC-CE, ABC-Bagging and ABC-Boosting are constructed using the finally selected feature subset. ABC-CE is formed by the majority vote [2] of the three classifiers, Decision Tree, Naïve Bayes and SVM. Bagging [2], [3] & [13] and Boosting are most famous CE methods which have been used in numerous Pattern Classification domains [2], [3] & [14]. ABC-Bagging is constructed by the combination of C4.5 Bagging with ABC selected feature subset. ACO-Boosting is constructed by Boosting the C4.5 decision tree along with ABC selected feature subset.

5. Experiments and Results

The datasets used, the implementation and the results of AC-ABC are discussed in this section.

5.1 Datasets

The performance of the proposed method ABCE discussed in this study has been implemented and tested using 10 different medical datasets. Heart-C, Dermatology, Hepatitis, Lung Cancer, Pima Indian Diabetes, Iris, Wisconsin Breast Cancer, Lmphography, Diabetes and Stalog-Heart are the datasets used. These datasets are taken from UCI machine learning repository

[15] and their description is given in Table I. The reasons for selecting these datasets are that they have been used in numerous classifier ensemble and feature selection proposals for experimental proof. The datasets are chosen such that the number of features is in a varied range and large number of instances, so that the effect of feature selection by ABCE is easily visible.

Table 1: Datasets Description

Dataset	Instances	Features	Classes
Heart-C	303	14	2
Dermatology	366	34	6
Hepatitis	155	19	2
Lung Cancer	32	56	2
Pima	768	8	2
Iris	150	4	3
Wisconsin	699	9	2
Lymph	148	18	4
Diabetes	768	9	2
Heart-Stalog	270	13	2

5.2 Implementation of ABCE

Classifications of the datasets are implemented using WEKA 3.6.3 Software from Waikato University [16] and feature selection using ABC has been implemented using Net Beans IDE. Decision Tree is implemented by using J48 algorithm, SVM by the LIBSVM package and Naïve Bayes by the Naïve Bayes classification algorithm from WEKA.

The artificial bees search for the best feature subset configuration with the following parameter initializations for ABC:

Population Size p	: 2 * No. of features in the data set
Dimension of the population	: $p \times N$
Lower Bound	: 1
Upper Bound	: N
Maximum Number of iteration	: Equal to the number of features
No. of runs	: 10
φ	: 0.3

With these parameter settings, the best optimal feature subset is recorded after executing a specified number of cycles. After every iteration, the employed bees pass the selected features to the classifier ensemble for evaluation. The mean accuracy of the classifiers in the ensemble and the consensus upon the feature subset are calculated using equations (1) and (2). The fitness measure for each feature subset is the average of the mean accuracy and the consensus and it is calculated using equation (3). The onlookers decide upon a feature subset with a probability which depends on the fitness. The number of features selected and the ensemble accuracy of ABCE is given in Table 2.

Table 2: Feature Selection and Ensemble Accuracy Achieved through ABCE

Dataset	No. of Features	Features Selected by ABCE	Predictive Accuracy (ABCE)(%)
Heart-C	14	7	86.92
Dermatology	34	24	98.55
Hepatitis	19	11	81.26
Lung Cancer	56	27	89.25
Pima	8	6	80.08
Iris	4	2	96.00
Wisconsin	9	4	96.99
Lymph	18	9	96.69
Diabetes	9	5	83.12
Heart-Stalog	13	6	84.07

Three classifier ensembles ABC-CE, ABC-Bagging, ABC-Boosting are then constructed using the optimal feature subset selected by the proposed ABCE method.

The classification accuracies achieved by these three ensembles are given in Table 3. Also in Table 3, the performance of ABCE is compared with ACO based ensemble, Bagging C4.5 and Boosting C4.5. 10-fold Cross Validation has been used to evaluate the accuracy of the constructed ensembles [1], [2] and [3].

When the ABCE method is applied to the datasets and the ensembles are constructed using the features output by ABCE, the recognition rates for all the ten datasets are improved significantly and this is shown in Fig.4.

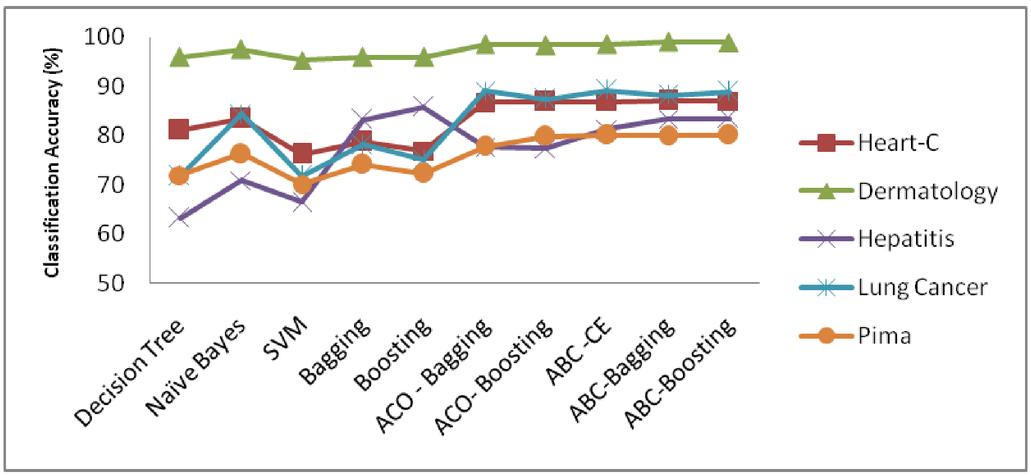
From the data represented in Table 2, Table 3 and Fig. 3, it can be inferred that:

i. Feature selection definitely increases the classification accuracy and speeds up the process of classification
ii. For all datasets except Hepatitis and Diabetes, ABCE has given the highest recognition rates
iii. For Hepatitis, Boosting has given the highest accuracy and ABCE has given better performance compared to ACO
iv. For diabetes, ACO has the leading performance and accuracy of ABCE is marginally low compared to ACO
v. For Heart-c, Iris, Pima and Wisconsin, feature subset obtained is almost of same size as in ACO
vi. For Lung Cancer, Lymph and Stalog, size of the feature set is minimized to a greater level with good prediction accuracies. This very well explains the effectiveness of the proposed method
vii. Convergence of the search space is achieved quickly

Table 3: Classification Accuracy of the Ensembles by 10 Fold Cross Validation

Dataset	**Bagging (C4.5)**	**Boosting (C4.5)**	**ACO - Bagging**	**ACO-Boosting**	**ABC -CE**	**ABC-Bagging**	**ABC-Boosting**
Heart-C	78.88	76.9	86.75	86.85	86.92	**87.11**	86.99
Dermatology	95.90	95.90	98.58	98.35	98.55	**99.07**	98.94
Hepatitis	83.23	**85.81**	77.65	77.45	81.26	83.40	83.44
Lung Cancer	78.12	75	89.05	87.37	**89.25**	88.19	88.95
Pima	74.09	72.4	77.72	79.82	80.08	79.94	**80.16**
Iris	95.33	93.33	93.74	93.35	96.00	**96.34**	**96.34**
Lymph	95.14	96.42	78.44	78.35	**96.99**	95.20	95.91
Wisconsin	74.09	72.4	88.94	87.65	**96.69**	96.00	93.12
Diabetes	79.05	83.11	83.91	**84.11**	83.12	83.88	83.96
Heart Stalog	80	80.37	80.99	82.12	84.07	83.72	**84.99**

(All numbers are in percentage)



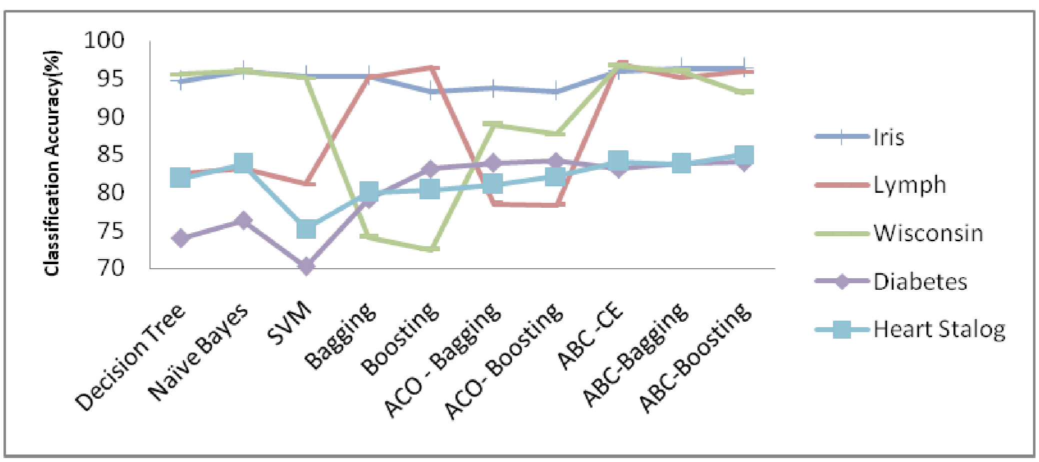


Fig.3 Graph Showing the Comparison of Predictions for the Ten UCI Datasets by the Constituent Classifiers, Traditional Ensembles and Ensembles Constructed Using ACO and ABCE

viii. In the graphical representation, the curve tends to elevate on the ABCE methods for most of the datasets, which shows the betterment of the proposed method

6. Conclusion

In this paper, a new method of classifier ensemble ABCE has been proposed and implemented. ABCE is proposed by combining the multi-objective ABC with a Classifier Ensemble (CE) and has been used to optimize the feature selection process. This method has resulted in optimal selection of feature subsets and the effectiveness of the proposed method can be seen from the results obtained. The ensembles ABC-CE, ABC-Bagging and ABC-Boosting developed using the selected feature subset, has given classification accuracies increased by 12% than the constituent classifiers and the ensembles Bagging, Boosting, ACO-Bagging and ACO-Boosting.

References

[1] R.O. Duda, P.E. Hart and D.G. Stork, Pattern Recognition, John Wiley & Sons, Inc 2nd edition, 2001.

[2] L.I.Kuncheva, Combining Pattern Classifiers, Methods and Algorithms, Wiley Interscience, 2005.

[3] Polikar R., "Ensemble based Systems in decision making" IEEE Circuits and Systems Mag., vol. 6, no. 3, pp. 21-45, 2006.

[4] Karaboga and B. Basturk, "A Powerful and Efficient Algorithm for Numerical Function Optimization: Artificial Bee Colony (ABC) Algorithm", Journal of Global Optimization, Springer Netherlands, vol.39, no.3, 2007, pp.459-471

[5] Karaboga, "An Idea Based on Honey Bee Swarm for Numerical Optimization", Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department, 2005.

[6] D. Karaboga, B. Basturk, "Artificial Bee Colony (ABC) Optimization Algorithm for Solving Constrained Optimization Problems", LNCS: Advances in Soft Computing: Foundations of Fuzzy Logic and Soft Computing, Springer-Verlag, 2007. vol. 4529/2007, pp.789-798.

[7] Fei Kang, Junjie Li, Haojin Li, Zhenyue Ma and Qing Xu, An Improved Artificial Bee Colony Algorithm, Proc. IEEE International Workshop on Intelligent Systems and Applications, 2010, pp 1 - 4.

[8] Zili Zhang and Pengyi Yang, "An Ensemble of Classifiers with Genetic Algorithm Based Feature Selection", IEEE Intelligent Informatics Bulletin, 2008, Vol 9, No.1, pp. 18-24.

[9] Nadia Abd-Alsabour and Marcus Randall, Feature Selection for Classification Using an Ant Colony System, Proc. Sixth IEEE International Conference on e–Science Workshops, 2010, pp 86- 91.

[10] Dash.M and Liu.H., "Feature Selection for Classification", Intelligent Data Analysis, Vol.39, 1997, No. 1, pp. 131 – 156.

[11] Witten, I. H., & Frank, E. (2005). Data mining: Practical machine learning tools and techniques. San Francisco: Morgan Kaufmann.

[12] Han, J., and Kamber, M.: Data mining concepts and techniques, Academic Press, 2001.

[13] L. Breiman, "Bagging predictors," Machine Learning, 1996, vol. 24, no. 2, pp.123-140.

[14] Y. Freund, R.E. Schapire, "Experiments with a new boosting algorithm", Proceeding of the Thirteenth International conference on Machine Learning, 1996, 148-156.

[15] A.Frank, A. Asuncion, UCI Machine Learning Repository, (http://archive.ics.uci.edu/ml. Irvine, CA: University of California, School of Information and Computer Science (2010))

[16] WEKA: A Java Machine Learning Package, http://www.cs.waikato.ac.nz/˜ml/weka/.

[17] N.Suguna and K.G.Thanushkodi, "A novel Rough Set Reduct Algorithm for Medical Domain based on Bee Colony Optimization", Journal of Computing, Vol. 2(6), 2010, 49 -54.

[18] N.Suguna and K.G.Thanushkodi, "An Independent Rough Set Approach Hybrid with Artificial Bee Colony Algorithm for Dimensionality Reduction", American Journal of Applied Sciences 8 (3), 2011, pp 261 – 266.

[19] Laura E A Santana, Ligia Silva Anne M P Canuto, Fernando Pintro and Karliane O Vale, "A Comparative Analysis of Genetic Algorithm and Ant Colony Optimization to Select Attributes for an Heterogeneous Ensemble of Classifiers", Proc. IEEE Congress on Evolutionary Computation, 2010, pp. 465-472.

[20] ShiXin Yu, Feature Selection and Classifier Ensembles: A Study on Hyperspectral Remote Sensing Data, Ph.D. Thesis, The University of Antwerp, 2003.

P.Shunmugapriya received her M.E. degree in 2006 from Department of Computer Science and Engineering, FEAT, Annamalai University, Chidambaram. She had been working as a Senior Lecturer for the past 7 years in the Department of Computer Science and Engineering, Sri Manakula Vinayagar Engineering College, Affiliated to Pondicherry University, Puducherry. Currently she is working towards her Ph.D degree in Optimal Design of Classifier Ensembles using Swarm Intelligent, Meta-Heuristic Search Algorithms. Her areas of interest are Artificial Intelligence, Ontology based Software Engineering, Classifier Ensembles and Swarm Intelligence.

Dr. S. Kanmani received her B.E and M.E in Computer Science and Engineering from Bharathiyar University and Ph.D in Anna University, Chennai. She had been the faculty of Department of Computer Science and Engineering, Pondicherry Engineering College from 1992 onwards. Presently she is a Professor in the Department of Information Technology, Pondicherry Engineering College, Puducherry. Her research interests are Software Engineering, Software testing, Object oriented system, and Data

Mining. She is Member of Computer Society of India, ISTE and Institute of Engineers, India. She has published about 50 papers in various international conferences and journals.