BIJIT - BVICAM's International Journal of Information Technology
Bharati Vidyapeeth's Institute of Computer Applications and Management (BVICAM), New Delhi (INDIA)

On the Importance of Ensembles of Classifiers

A. K. Saxena

Submitted in November 2012, Accepted in May 2013

Abstract - In this paper, a recent yet powerful technique for classification of datasets is presented. The paper contributes to highlight the importance of an ensemble approach over individual classifiers to achieve better classification accuracy of a classifier. In this paper, given dataset is divided into a number of parts to constitute an ensemble. The ensemble combines these classifiers. An unknown data pattern is tested on the ensemble. Using bagging, majority of voting technique, the performance of ensemble is determined on different sections of datasets. In the paper, six bench mark datasets are used for investigation. Each dataset is trained with 80%, 60% and 50% of the data patterns for classification. The number of classifiers in an ensemble for each data set is changed to 5,7 and 9. As a typical case, k-nearest neighbor (k-NN) classifiers are used with the values of k varying to 1,3 and 5. The classification accuracies of individual classifiers and those of ensembles are computed at each case. After extensive experiments of proposed scheme, by taking random shuffling and selection of data patterns for training and testing, it is observed that in every case, the classification accuracy obtained by ensemble is higher than that obtained by individual classifier.

Index Terms - Classification, Ensemble of classifies, bagging, k-nn classifier.

1. INTRODUCTION

There have been a significant number of research activities in the area of data analysis. The size of database keeps on increasing with useful or redundant data. The task of analysis of the data becomes complex due to presence of these redundant, mostly unwanted pieces of data, commonly called features in a formatted dataset. The role of a classifier is to divide a dataset on the basis of labels or classes of its patterns. In addition to classifying data patterns into different classes, it is also expected from a classifier to predict the label (or more often termed as class) of an unknown pattern, called test pattern. Classification has become a vital component of the study of pattern recognition [1]. Due to the huge amount of data piled up every moment on disks, web spaces and other storage devices, techniques like data mining [2,3], have become quite relevant. Classification is an important step of data mining. Classification is one of the core challenging tasks [4] in mining [5], pattern recognition [1], bioinformatics [6]. The goal of classification [7,8] is to assign a new entity into a class from a pre-specified set of classes.

Dept of CSIT, Guru Ghasidas Vishwavidyalaya, Bilaspur, Chattisgarh, India, E-Mail: amitsaxena65@rediffmail.com

A classifier needs to be trained before it can be set ready for predicting the class of unknown patterns. The learning of classifier can be made in two manners viz. supervised and unsupervised. In case of supervised learning, the class of every pattern is known in advance at the time of training. In unsupervised learning, class of the training pattern is not given. Commonly, the classifications are based on classification models (classifiers) that are induced from an exemplary set of pre-classified patterns. Alternatively, the classification utilizes knowledge that is supplied by an expert in the application domain. In a typical supervised learning setting, a set of instances also referred to as a training set is given. The labels of the instances in the training set are known and the goal is to construct a model in order to label new instances. An algorithm which constructs the model is called *inducer* and an instance of an inducer for a specific training set is called a *classifier*. There are several well established classifiers such as Fisher's Linear discriminant analysis (LDA) [25], naive Bayes classifier[26], support vector machines, SVM [27], k-Nearest neighbor [28], Neural Networks [29.], fuzzy [30, 40.]. In many examples, idea behind the construction of an ensemble is to combine the classifiers after a weak or non perfect training of individual classifiers. The ensemble so obtained outperforms every individual classier. In fact, human being tends to seek several opinions before making any important decision. Before buying very costly items or taking critical medical decisions, it is a common practice to weight the individual opinions, and combine them to reach to a final decision [9]. Recently, Mikel Galaretal[10] reported that class distribution, i.e., the proportion of instances belonging to each class in a data-set, plays a key role in classification. Sometimes imbalanced data-sets problem occurs when one class, usually the one that refers to the concept of interest (positive or minority class), is under-represented in the data-set; in other words, the number of negative (majority) instances outnumbers the amount of positive class instances [11]. The primary benefit of using ensemble systems is the reduction of variance and increase in confidence of the decision. Due to many random variations in a given classifier model (different training data, different initialization, etc.), the decision obtained by any given classifier may vary substantially from one training trial to another—even if the model structure is kept constant. Then, combining the outputs of several such classifiers by, for example, averaging the output decisions, can reduce the risk of an unfortunate selection of a poorly performing classifier. Another use of ensemble systems includes splitting large datasets into smaller and logical partitions, each used to train a separate classifier. This can be more efficient than using a single model to describe the entire data. The opposite problem, having too little data, can also be handled using ensemble systems, and this is where bootstrap-based ideas start surfacing: generate multiple classifiers, each trained on a different subset of the data, obtained through bootstrap re-sampling. While the history of ensemble systems can be traced back to some earlier studies such as [12,13],it is Schapire's 1990 paper[14] that is widely recognized as the seminal work on ensemble systems. Few more references for data fusions and combining classifiers are available in [15-21]. In this paper, study of ensemble of classifiers is presented using investigation on different datasets. It is important to submit here that there is a quite little scope of comparison of proposed scheme with others available in literature. The reason is that in each ensemble of classifiers, the constituent classifiers are well established classifiers, viz. neural networks, fuzzy, knn etc. The performances of these individual classifiers have already been widely reported in literature in several applications. For a simple implementation of the proposed scheme, k-NN classifier has been used in this paper as the constituent classifier of the ensemble. Presumably, the k-nearest neighbor algorithm[28] is considered one of the simplest machine learning algorithms. It is further to add that the objective here is not to discuss the strength of k-nn but to investigate the performance of the ensemble, irrespective of its consitiuent classifiers. However a good survey on k-nn classifier can be found at [31].

The objective of this paper is to support the creation of an ensemble with one or more of these classifiers as constituent members and to show that under an ensemble, the classifier accuracy produced by such an ensemble using majority of voting criterion, is always higher than that obtained by using individual classifier. This is supported by investigation on six benchmark datasets.

The paper is organized as follows: Section II presents proposed ensemble scheme. Section III outlines summary of datasets used in the experiments. The details of experiments and results are discussed in Section IV. Section V addresses the strength and weakness of proposed technique by comparing it with few of the others reported. Conclusions and future research prospects are reflected in Section VI followed by references.

2. PROPOSED ENSEMBLE ALGORITHM

In this paper, simple bagging without replacement of samples, with majority of voting [11,22,23] is used for the investigation of proposed scheme. Steps of the algorithm used in the paper are given below.

Algorithm

Input: *D*, the given dataset consisting of *N* patterns. *F*, number of features in each pattern, each pattern being labeled with a class *c* and *C* is the total number of classes in *D*. *S*, is number of classifiers in the ensemble.

1. Partition the entire dataset D into two parts, training dataset, S_{tr} and testing dataset, S_{te} . Each part has same number of features. Each pattern in these two parts is labeled with one class out of *C* classes, thus
   $S_{tr} \bigcup S_{te} = D$
2. Make equal partitions of S_{tr} such that all parts except the last, will have S_{tr}/S patterns. The last part will have ($S_{tr}/S+ \ S_{tr}\%S$), where % is modulus operation on integers. The ensemble will thus have *S* number of classifiers, one for each part.
3. Invoke k-nearest neighbor classifier [32] with k =1.
4. Determine the classification accuracy, C_a of each part of the training data using k-nn, against the same test data set S_{te}. Find out the average C_a of all *S* classifiers. Determine the maximum C_a obtained in the *S* classifiers.
5. Shuffle dataset *D;* create new S_{tr} and S_{te}..
6. Iterate steps 2 to 5, *I* times. Find the C_a and maximum $C_a$in these *I* iterations. .
7. Take every pattern of S_{te} and pass it through all *S* classifiers using bagging [11,22] and majority of voting techniques to determine it's class. Repeat the process for *I* times, Calculate average and maximum C_a of the ensemble (EC_a) in these *I* trials.
8. Change the value of *k* (1,3,5)
9. Change the value of *S* (5,7,9).
10. Change the size of training data (80%,60% and 50%) and accordingly test data.

Order of algorithm: There has been a variety of work in analysis of k nearest neighbors[34,35]. In the simplest form as used here [1,33], for k-nn, the order of search is $O(kdt_rt_e)$ where *F* is number of features (dimensions) in each pattern, *k* is number of nearest neighbors, Euclidean distance is used as a metric of nearest hood between test point t_e and training pattern t_r,, *P* is the preprocessing due to shuffling and partitioning of training (and testing) datasets, talking majority decision in bag of *S* classifiers For complete algorithm proposed here, order of algorithm may be given as follows

$$O(kFt_rt_e + P)$$

The algorithm is iterated for *k* as 1,3,5; *S* as 5,7,9; and size of training dataset as 80%,60% and 50%.

Fig. 1 shows the proposed scheme. In this figure, as a typical example, five classifiers are placed in an ensemble. The parts of training data $S_1 \ldots S_5$ are used for creating five classifiers $C_1 \ldots C_5$, one classifier for one part respectively. The CA of ensemble is shown by C_e.

3. SUMMARY OF DATABASES

Table 1summarizes data sets used for the experiments. The data sets are well established and have been used in several investigations. The details of each data set can be viewed in UCI Machine Learning Repository [24]. There has been no preference to choose any particular data set for investigation in this paper.

Data Set	Total Patterns	Features	Classes	Patterns in Class1	Patterns in Class2	Patterns in Class3
Iris	150	4	3	50	50	50
Wine	178	13	3	59	71	48
Liver	345	6	2	145	200	-
Thyroid	215	5	3	150	35	30
WBC	683	9	2	444	239	-
Sonar	208	60	2	97	111	-

Table 1:Description of the Data Set Used.

4. EXPERIMENTS AND RESULTS

Proposed ensemble algorithm was run on an i5 machine using MATLAB software. The purpose of the investigation was to focus the strength of proposed ensemble scheme over individual classifiers. The results obtained for the six classical databases are shown in Tables 2(A-F) for Iris, Wine, Bupa Liver, Thyroid, WBC(Wisconsin Breast Cancer) and Sonar datasets respectively. In each of these tables, first column: T, (training data size) indicates the part (in percent) of the database which will be used for training only whereas the remaining part (100 – T) will be used for testing. Three sizes for training have been used in the paper viz. 80%, 60% and 50%, to reflect attitude of the proposed algorithm towards different parts of the data. The next column represents values of 'k', i.e. the k-th nearest neighbor from the testing data pattern. The measure of the distance is taken as Euclidean distance. Three values of 'k' (1,3 and 5), have been used for all these datasets. To apply bagging, each training dataset is divided into *S* number of sub sets. In the paper, *S* is set for three values: 5,7 and 9. In other words, number of classifiers in ensemble will be 5, 7 and 9 for each of the datasets. Thus for each dataset, a training part of the dataset (80/60/50 %), has *S* different subsets. For a typical training dataset with five folds or subsets

$$S_1 \bigcup S_2 \bigcup S_3 \bigcup S_4 \bigcup S_5 = S_{tr}$$

and

$$S_{tr} \bigcup S_{te} = D$$

where S_{tr} and S_{te} stand for training and testing Dataset respectively, and D is the entire dataset. As a typical case, first experiment is conducted with S=5, k=1 and training data size =80% of the total dataset. The testing data (20%) will remain as the unseen part of dataset. In this case, each of these five classifiers, $S_1 \ldots S_5$,is applied to its respective training data part, e.g. first classifier accuracy *CA* will be obtained using 1-NN between $S_1$and test dataset, second *CA* between S_2same test data set and so on. The mean (average)of these five C_a is computed. The C_a of ensemble of classifier is computed as follows. Take first test pattern from test database and find its class using first nearest neighbor (1-NN) with S_1 then find its class with S_2,S_3,S_4,S_5. The majority (mode) of values of classes so obtained in five tests will be the class accepted for the ensemble. Repeat the exercise for all the patterns in test dataset and calculate its percentage C_a. This will compute $E\ C_a$ of the ensemble. Time in execution of the whole process is also recorded. The whole exercise is repeated for five times by shuffling randomly the dataset. Compute the mean classification accuracy *Mean* C_a,, from these five iterations. Also calculate the maximum value of C_a , *Max* C_a in these five iterations. Similarly compute mean and maximum ensemble accuracy *Mean* EC_a and $MaxEC_a$ in the five iterations. Compute the mean time spent on one iteration. The mean values are shown in Tables 2(A-F). The maximum values for classification accuracies in five iterations are shown within the brackets in the same tables. This is shown by the first row of the first main sub column of Table with *S*=5. Similar exercise is repeated for *S*=7 and 9. This completes row 1 of the table. The values of k are varied to 3 and 5. Then training data size is changed to 60% and 50% and exactly same procedure is adopted. Due to space limitations, the values in tables are rounded up to two decimal places. The tables 2(A-F) are enclosed in Annexure-1.

On observing these Tables 2(A), it is noted that for iris data set, for S=5, k=1, *meanCa* =94.7 is highest when individual classifiers are considered. In this case *meanE* C_a is 96.7. For S=7, k=1, *mean* C_a =90.5 is highest for individuals, whereas *mean* C_a =96.6. For S=9, *mean* C_a =93.3, *mean* C_a =100%, Thus it is noted that *meanECA* is in each case is higher than *meanCA*. In most cases, mean C_a is same as maximum value of C_a. Typically, for 50% training data, S=7, k=1, *mean* C_a =85.7 whereas *max* C_a is 91.6. Similarly *meanE* C_a is 88.5 and *maxE* C_a =92.6. There are few more cases where mean values of CA are smaller than maximum values of C_a .Similar trend is noted in all tables 2(A-F).

As another case, Table 2(E) can be quoted which presents results on breast cancer (wbc) data. This dataset has 683 patterns divided into 444 and 239 patterns for class 1 and class2 respectively. Dataset has 9 features(attributes). With nine (S=9) 1-NN classifiers, *mean* C_a 96.6 whereas *max* C_a =97.9. The mean $E\ C_a$ =98.5 with maximum *ECA* as 99.3%, a better performance. Sonar dataset has 208 patterns divided into two classes having 97 and 111 patterns respectively. It has 60 features in each pattern. By observing Table 2(C), it is noted that *mean* C_a =61.3, with 60% training data and 1-NN, using nine classifiers (S=9), whereas mean$E\ C_a$ under similar conditions is 71.1.

It is therefore observed from study of all these tables that the values of mean C_a are always less than mean $E\ C_a$. The maximum values of C_a in few cases are greater than the mean values of C_a in five iterations. The reason for running experiments for five times is just to ensure that the performance of the classifiers can be checked under all possible patterns combinations in training and test datasets. It is again apprehended that each ensemble can contain any set of similar or combination of classifiers such as neural networks, fuzzy, Bayesian, kNN etc. The contribution of the paper is more towards showing the importance of the ensemble with majority of voting than to highlight the strength of constituent classifiers which are undoubtedly proven in literature. That is why the classification accuracies of constituent classifiers are compared with that of the ensemble and not with other constituent classifiers of the ensemble. As a typical example kNN is used in all cases.

5. DISCUSSIONS ON THE COMPARATIVE STUDY OF PROPOSED TECHNIQUE

The proposed technique has been used for different applications e.g. in [36] , researchers used ensemble classifier for fMRI data analysis. There are various strong merits of the proposed scheme including high possibility of getting better classification accuracy from an ensemble than an individual classifier; the individual classifiers of the ensemble need not to be perfectly trained, mostly these are weak learners, thereby reducing the time and efforts of training them; the fact is also confirmed when different sizes of the training dataset is taken (80%,60% and 50%) still a good accuracy is achieved; there is a scope for feature selection and dimensionality reduction of the dataset, under different combinations of features, the ensemble can be called to predict a reasonable good accuracy.

Although it is difficult to find a common platform to compare the performance of proposed technique with some other used in different context, yet few results are being discussed here for the purpose

For iris data, the accuracy obtained in [37] is 94.7 for CBA scheme 96.6 for Neural Network system, where as with the proposed technique it is 100% for 9 classifiers in the ensembles with 80% training data for validation with k as 1.

For thyroid data [7], the accuracy is 95% with time as 0.913 seconds. In proposed scheme, the accuracy is 95.4 with S=4, k=1, time = 0.50 seconds.

For wine dataset, accuracy in [7] is although 89% but time taken is 1.34 seconds whereas in proposed scheme accuracy is 81.7 but time is quite less, 0.53 seconds ($k=1,S=9$, training data T = 60%).

For WBC data, in [38], the classification accuracy is 90% with time taken is 48 seconds whereas in the proposed technique, the accuracy is 98% ($k=1,S=9$, training data T = 80%) with time = 1.6 seconds.

For sonar data, the accuracy obtained in [39], is 81% whereas the accuracy obtained by proposed technique is approximately 79% ($k=1,S=5$, training data T = 80%).

It is again reminded that the proposed technique focuses on the use and importance of an ensemble of classifiers and not of an individual classifier.

One possible inability of the proposed technique is that it does not address or attempt to modify the original structure of any individual constituent classifiers. If a classifier originally does not fit suitable for a particular dataset or on a specific nature of data, the ensemble by no means will be able to improve its performance. Moreover for a large set of data such as micro array gene data, the performance of the proposed technique is subject to test.

6. CONCLUSION

In this paper, a recent yet important scheme of classification has been presented. A classifier can produce good classification accuracy for one dataset, but performs poorer when presented with different dataset or even different section of the same dataset. If however, multiple classifiers are trained for small sections of the databases, and are combined in the form of an ensemble, then such an ensemble can produce better classification accuracy. To justify it, six bench mark datasets, iris, BUPA liver, thyroid, sonar, breast cancer and wine have been used for empirical study. The size of the training part of each dataset is taken as. 80%,60% and 50%. The number of classifiers in the ensemble is taken as 5,7 and 9. The k-nearest neighbor has been used as classifier with the values of k as 1,3 and 5 under each case. Experiments were conducted to evaluate the classification accuracies of all six datasets. In order to provide diversity in training and testing datasets, the experiments were iterated for five times with shuffling of dataset. The mean and the maximum classification accuracies of individual classifiers on each sub sets of training datasets were computed. The same were computed for ensemble of the classifiers using majority of voting. The results produced in these two cases, show that the classification accuracy of each individual classifier in general is lower than that of the classification accuracy obtained by their ensemble. Thus it is concluded from these investigations that an ensemble is a good approach to determine the class of an unseen data pattern. The scheme can be applied to many other datasets. Moreover, other classifiers like neural network, fuzzy etc. can be included in the ensemble. This study can also be extended to explore the possibility of feature selection or dimensionality reduction.

REFERENCES

[1]. Duda, R.O., Hart, P.E., Stork, D.G., *Pattern Classifi*cation. John Wiley and Sons (Asia) Pte.Ltd., second .ed. 2006.

[2]. Kamber, M., Han, J.,Pei,*Data mining: Concepts and techniques*,2nd ed. CA: Morgan Kaufmann Publisher. San Francisco, 2011.

[3]. Jean-Marc Adamo, Data Mining for Association Rules and Sequential Patterns: Sequential and Parallel Algorithms, Springer, 2001.

[4]. Misra, B.B., Dehuri, S., Dash, P.K., Panda, G., "Reduced Polynomial Neural Swarm Net for Classification Task in Data Mining", IEEE Congress on Evolutionary Computation,2008b

[5]. Kosala, R., Blockeel, H., "Mining Research: A Survey". ACM SIGKDD Explorations. 2 (1), 2000, pp 1–15,.

[6]. Baldi, P., Brunak, S., *Bioinformatics: The Machine Learning Approach*. MIT Press, Cambridge, MA.,1998.

[7]. Saxena, Patre,Dubey, "An Evolutionary Feature Selection Technique Using Polynomial Neural Network", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 4, No 1, July 2011 ISSN (Online): pp. 1694-0814 www.IJCSI.org.

[8]. Mitchel, T.M.,*Machine Learning*. McGraw Hill,1997.

[9]. Polikar R, “Ensemble based systems in decision making”, IEEE Circuits Syst Mag 6(3):pp. 21–45, 2006.

[10]. Mikel Galaretal. , “A Review on Ensembles for the Class Imbalance Problem: Bagging-, Boosting-, and Hybrid-Based Approaches”, IEEE Transaction on Systems, Man and Cybernetics, Part-C,Applications and reviews, pp. 1-22, 2011.

[11]. Polikar,“Bootstrap methods”, in IEEE Signal Processing Magazine,July 2007.

[12]. B.V. Dasarathy and B.V. Sheela, “Composite classifier system design: Concepts and methodology,” Proc. IEEE, vol. 67, no. 5, pp. 708–713, 1979.

[13]. L.K. Hansen and P. Salamon, “Neural network ensembles,” IEEE Trans. Pattern Anal. Machine Intell., vol. 12, no. 10, pp. 993–1001, 1990.

[14]. R.E. Schapire, “The strength of weak learnability,” Machine Learning, vol. 5,no. 2, pp. 197–227, June 1990.

[15]. Okun, Oleg,Supervised and Unsupervised Ensemble Methods and their Applications, Springer, 2008,

[16]. Oleg Okun, *Feature Selection and Ensemble Methods for Bioinformatics: Algorithmic Classification and Implementations*. , IGI Global, Hershey, PA, 2011.

[17]. Verma, B.,”Cluster-Oriented Ensemble Classifier: Impact of Multicluster Characterization on Ensemble Classifier Learning”, IEEE Transactions on Knowledge and Data Engineering, vol 24(4), pp 605-618, 2011.

[18]. E. Bauer and R. Kohavi, “An empirical comparison of voting classification algorithms: Bagging, boosting, and variants,” Machine Learning, vol. 36, no. 1-2, pp. 105–139,1999. [Online].Available:http://citeseer.ist.psu.edu/bauer99empirical.html

[19]. D. Ruta and B. Gabrys, “An overview of classifier fusion methods,” Computing and Information Systems, vol. 7, pp. 1–10, 2000.

[20]. A. Al-Ani and M. Deriche, “A new technique for combining multiple classifiers using the Dempster-Shafer theory of evidence,” Journal of Artificial Intelligence Research, vol. 17, pp. 333–361, 2002.

[21]. L. K. Hansen and P. Salomon, “Neural network ensembles,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12(10), pp. 993–1001, October 1990.

[22]. Lior Rokach, “Ensemble-based classifiers”, Artificial Intelligence Rev pp. 33:1–39, Springer, 2010.

[23]. Saxena,Mondol,Mir, “Improving the Classification ccuracy with Ensemble of Classifiers”, proc. Of National Conference NCMIRA 12, India,21-23 Dec 2012, pp. 52-56.

[24]. http://www.ics.uci.edu/$\sim$mlearn/MLRepository.html

[25]. Fisher, R. A. (1936). "The Use of Multiple Measurements in Taxonomic Problems". Annals of Eugenics 7 (2): 179–188.

[26]. Domingos, Pedro & Michael Pazzani (1997) "On the optimality of the simple Bayesian classifier under zero-one loss". Machine Learning, 29:103–137

[27]. Cortes, Corinna; and Vapnik, Vladimir N.; "Support-Vector Networks", Machine Learning, 20, issue 3, 273-297, 1995

[28]. Cover TM, Hart PE, "Nearest neighbor pattern classification". IEEE Transactions on Information Theory 13 (1): 21–27, 1967.

[29]. Haykin, S. , Neural Networks: A Comprehensive Foundation, Prentice Hall, 1999.

[30]. Angelov P., X. Zhou, “Evolving Fuzzy-Rule-based Classifiers from Data Streams”, IEEE Transactions on Fuzzy Systems, ISSN 1063-6706, special issue on Evolving Fuzzy Systems, December 2008, vol. 16, No6, pp.1462-1475.

[31]. Liangxiao Jiang, Wuhan Zhihua Cai; Dianhong Wang; Siwei Jiang, “Survey of Improving K-Nearest-Neighbor for Classification”, in Proc. of IEEE Fourth International Conference on Fuzzy Systems and Knowledge Discovery, FSKD 2007. Vol.1, pp. 679-683.

[32]. www.mathworks.in/help/stats/knnsearch.html

[33]. Machine Learning, Tom Mitchell, McGraw Hill, 1997

[34]. Gayathri, K., Marimuthu, A.; “Text document pre-processing with the KNN for classification using the SVM” ,in Proc. Of 7th International IEEE Conference on Intelligent Systems and Control (ISCO), 2013, pp. 453-457.

[35]. Efendi Nasibov, , Cagin Kandemir-Cavas; “Efficiency analysis of KNN and minimum distance-based classifiers in enzyme family prediction”, Computational Biology and Chemistry, Elsevier, vol 33(6), 2009, pp. 461, 464.

[36]. Ludmila I. Kunchevaa,, Juan J. Rodríguezb, “Classifier ensembles for f MRI data analysis: an experiment”, Elsevier, Magnetic Resonance Imaging,28(2010), 583-593.

[37]. Prachitee Shekhawat1,Sheetal S. Dhande, “Building an Iris Plant Data Classifier Using Neural Network Associative Classification”, International Journal of Advancements in Technology http://ijict.org/, vol 2(4), 2011,pp 491-506.

[38]. Saxena,Patre,Dubey, “Investigating a novel GA-based feature selection method using improved KNN classifiers”, Int. J. Information and Communication Technology, Vol. 3, No. 3, 2011,pp 274-288.

[39]. Saxena,Pal,Kothari, “Evolutionary methods for unsupervised feature selection using Sammon’s stress function” , Fuzzy Information and Engineering Volume 2, Number 3, 229-247, DOI: 0.1007/s12543-010-0047-4, Springer, http://www.springerlink.com/content/1616-8658/2/3/

[40]. P. C. Jha, Shivani Bali and P. K. Kapur, “Fuzzy Approach for Selecting Optimal COTS Based Software Products Under Consensus Recovery Block Scheme”,

BIJIT - BVICAM's International Journal of Information Technology, Vol. 3(1), 2011

Annexure-1

Table 2: Results obtained for six datasets used in ensemble of classifiers

(A) Iris data

T	k	S=5			S=7			S=9		
		Mean C_a (Max C_a)	MeanE C_a(MaxE C_a)	Mean Time	Mean C_a (Max C_a)	MeanE C_a (MaxE C_a)	Mean Time	Mean C_a (Max C_a)	MeanE C_a (MaxE C_a)	Mean Time
80	1	94.7(94.7)	96.7(96.7)	0.43	90.5(90.5)	96.6(96.6)	0.48	93.3(93.3)	100(100)	0.47
	3	88.7(88.7)	96.7(96.7)	0.46	89.0(89.0)	90.0(90.0)	0.49	92.5(92.5)	100(100)	0.48
	5	94.0(94.0)	96.7(96.7)	0.47	88.5(88.5)	93.3(93.3)	0.50	83.7(83.7)	96.6(96.6)	0.51
60	1	94.7(94.7)	98.3(98.3)	0.50	89.3(89.3)	95.0(95.0)	047	90.5(90.5)	96.7(96.7)	0.51
	3	91.4(91.4)	98.4(98.4)	0.49	85.0(85.0)	91.7(91.7)	052	79.5(79.5)	95.0(95.0)	0.52
	5	89.4(89.4)	91.7(91.7)	0.49	80.5(80.5)	95.0(95.0)	052	62.5(62.5)	93.4(93.4)	056
50	1	91.7(91.7)	98.7(98.7)	0.49	**85.7(91.6)**	**88.5(92.6)**	0.47	86.9(90.1)	94.4(94.6)	0.44
	3	86.1(86.1)	86.7(86.7)	0.46	80.2(80.6)	93.6(96.0)	0.52	77.9(82.8)	91.7(92.0)	0.53
	5	78.4(78.4)	96.0(96.0)	0.49	62.5(68.6)	81.3(86.6)	0.51	66.5(69.3)	81.9(86.6)	0.55

(B) liver data

T	k	S=5			S=7			S=9		
		Mean C_a (Max C_a)	MeanE C_a (MaxE C_a)	Mean Time	Mean C_a (Max C_a)	MeanE C_a (MaxE C_a)	Mean Time	Mean C_a (Max C_a)	MeanE C_a (MaxE C_a)	Mean Time
80	1	59.1(63.8)	66.9(73.9)	0.73	56.6(56.7)	60.9(60.9)	0.69	53.9(56.8)	60.6(71.0)	0.81
	3	61.5(63.2)	66.9(69.6)	0.73	59.3(59.6)	61.7(62.3)	0.74	60.1(60.1)	68.2(68.2)	0.68
	5	59.8(61.2)	62.9(65.2)	0.76	59.1(59.1)	71.0(71.0)	0.74	60.5(63.6)	66.4(71.0)	0.72
60	1	55.5(57.1)	59.3(64.5)	0.73	54.7(57.1)	60.6(62.3)	0.71	56.7(56.7)	59.4(59.4)	0.68
	3	57.1(58.4)	60.7(63.1)	0.69	58.7(59.1)	65.5(65.9)	0.72	59.0(59.7)	67.9(68.8)	0.74
	5	58.9(60.0)	61.6(63.0)	0.69	60.1(60.7)	66.2(66.6)	0.72	58.4(58.4)	64.5(64.5)	0.69
50	1	56.4(56.9)	61.5(65.7)	0.83	55.3(59.3)	61.3(69.2)	0.73	57.5(57.7)	65.5(69.2)	0.73
	3	60.8(60.8)	67.4(67.4)	0.72	58.1(58.6)	65.2(66.9)	0.72	57.4(57.9)	69.1(72.1)	0.78
	5	56.5(60.8)	59.8(69.2)	079	57.9(59.4)	64.5(66.9)	0.73	56.9(57.6)	64.4(66.2)	0.86

(C) sonar data

T	k	S=5			S=7			S=9		
		Mean C_a (Max C_a)	MeanE C_a(MaxE C_a)	Mean Time	Mean C_a (Max C_a)	MeanE C_a (MaxE C_a)	Mean Time	Mean C_a (Max C_a)	MeanE C_a (MaxE C_a)	Mean Time
80	1	70.3(71.4)	78.1(78.6)	0.79	66.1(68.4)	69.5(73.8)	0.78	66.9(66.9)	77.6(78.6)	0.72
	3	65.5(66.6)	72.4(74.9)	0.76	57.3(64.9)	58.6(73.8)	0.80	48.5(56.3)	47.1(64.3)	0.76
	5	60.9(60.9)	66.6(66.6)	0.76	59.9(60.5)	69.0(69.0)	0.76	54.4(56.3)	58.1(61.9)	0.77
60	1	65.2(69.4)	71.9(74.7)	0.73	61.3(61.9)	66.7(67.4)	0.76	61.3(61.3)	71.1(71.1)	0.80
	3	59.0(59.3)	62.2(65.1)	0.75	55.6(55.6)	60.2(60.2)	0.79	57.0(58.4)	63.1(63.9)	0.76
	5	57.3(60.2)	61.4(61.4)	0.74	55.4(55.9)	63.4(66.2)	0.74	54.2(55.2)	59.8(62.7)	0.78
50	1	65.6(65.6)	69.2(69.2)	0.73	62.7(63.8)	73.1(75.0)	0.77	59.9(60.7)	68.1(72.1)	0.75
	3	59.1(59.4)	61.5(64.4)	0.72	57.1(58.2)	63.8(64.4)	0.77	56.8(57.2)	67.5(69.2)	0.78
	5	58.7(59.6)	61.5(63.5)	0.75	51.4(51.4)	50.9(50.9)	0.74	56.4(56.7)	62.9(72.1)	0.80

(D) thyroid data

T	k	S=5			S=7			S=9		
		Mean C_a (Max C_a)	MeanE C_a(MaxE C_a)	Mean Time	Mean C_a (Max C_a)	MeanE C_a (MaxE C_a)	MeanTime	Mean C_a (Max C_a)	MeanE C_a (MaxE C_a)	Mean Time
80	1	90.2(90.2)	95.4(95.4)	0.50	87.4(87.4)	93.0(93.0)	0.55	77.9(79.3)	86.5(88.4)	0.53
	3	86.5(86.5)	88.4(88.4)	0.53	83.0(83.0)	90.7(90.7)	0.59	79.6(81.9)	80.9(83.7)	0.62
	5	83.3(83.3)	86.1(86.1)	0.54	74.2(76.4)	72.5(76.7)	0.53	77.4(77.7)	69.8(69.8)	0.59
60	1	82.6(92.0)	84.4(96.5)	0.55	76.4(76.4)	81.4(81.4)	0.53	80.2(82.8)	79.7(80.2)	0.57
	3	79.5(79.5)	82.5(82.5)	0.51	75.6(75.9)	76.7(76.7)	0.56	71.6(72.0)	68.1(70.9)	0.64
	5	78.1(78.1)	80.2(80.2)	0.58	73.0(73.0)	72.1(72.1)	0.51	74.3(75.3)	70.0(70.9)	0.62
50	1	81.3(81.3)	81.3(81.3)	0.57	86.8(86.8)	92.5(92.5)	0.55	80.7(80.7)	85.4(85.9)	0.61
	3	81.1(81.1)	85.0(85.0)	0.55	79.0(79.8)	79.8(80.4)	0.59	72.0(74.1)	67.5(71.9)	0.62
	5	78.5(78.5)	82.2(82.2)	0.57	70.4(70.7)	66.9(67.3)	0.59	80.4(80.4)	77.6(77.6)	0.59

(E) wbc data

T %	k	S=5			S=7			S=9		
		Mean C_a (Max C_a)	MeanE C_a(MaxE C_a)	Mean Time	Mean C_a (Max C_a)	MeanE C_a (MaxE C_a)	MeanTime	MeanE C_a (MaxE C_a)	MeanE C_a (MaxE C_a)	Mean Time
80	1	95.4(97.2)	96.5(98.5)	1.35	95.5(97.8)	96.2(98.5)	1.50	96.6(97.9)	98.5(99.3)	1.60
	3	96.9(97.4)	96.6(97.0)	1.59	96.9(97.5)	97.5(98.5)	1.71	94.4(94.9)	95.2(96.3)	1.48
	5	95.7(96.5)	96.3(97.0)	1.53	96.5(97.7)	96.6(97.8)	1.43	96.3(97.2)	96.6(97.8)	1.51
60	1	95.3(96.3)	96.5(97.4)	1.55	94.9(96.0)	96.2(96.7)	1.66	95.8(96.6)	98.3(99.2)	1.54
	3	96.5(97.6)	97.4(98.5)	1.55	95.6(96.0)	96.5(97.4)	1.95	94.9(95.6)	95.0(98.2)	1.51
	5	95.7(98.1)	95.9(98.5)	1.56	96.5(97.2)	96.3(97.0)	1.66	94.5(96.0)	94.9(95.9)	1.71
50	1	94.6(94.9)	95.3(95.3)	1.42	95.1(95.4)	95.6(96.2)	1.63	94.3(95.0)	95.8(96.7)	1.71
	3	96.1(96.9)	96.5(97.4)	1.78	95.5(96.8)	96.2(97.4)	1.62	94.4(95.5)	94.9(95.6)	1.92
	5	95.6(96.3)	95.9(96.2)	1.72	95.6(95.8)	95.8(95.9)	1.50	96.0(96.6)	96.0(96.5)	1.63

(F) Wine data

T %	k	S=5			S=7			S=9		
		Mean C_a (Max C_a)	MeanE C_a(MaxE C_a)	Mean Time	Mean C_a (Max C_a)	MeanE C_a (MaxE C_a)	MeanTime	Mean C_a (Max C_a)	MeanE C_a (MaxE C_a)	Mean Time
80	1	60.4(66.6)	58.3(69.4)	0.50	67.2(69.4)	76.6(83.3)	0.49	62.9(62.9)	75.0(75.0)	0.56
	3	64.6(65.0)	65.0(66.6)	0.54	64.5(68.5)	68.3(75.0)	0.57	70.1(70.1)	77.7(77.7)	0.61
	5	65.7(67.7)	68.8(72.2)	0.52	65.4(66.3)	60.5(61.1)	0.55	70.0(70.0)	75.0(75.0)	0.61
60	1	63.5(67.6)	67.6(73.2)	0.53	68.8(68.8)	77.5(78.8)	0.51	70.7(70.7)	81.7(81.7)	0.53
	3	67.0(67.3)	66.5(67.6)	0.55	64.5(69.2)	69.3(76.1)	0.60	61.6(61.6)	64.7(64.7)	0.57
	5	69.8(70.7)	73.8(76.0)	0.54	70.9(72.6)	74.3(76.0)	0.60	57.1(57.1)	61.9(61.9)	0.60
50	1	61.9(63.8)	64.7(67.4)	0.50	65.5(65.6)	69.6(69.6)	0.57	61.2(61.2)	67.4(67.4)	0.55
	3	69.2(70.5)	73.4(74.1)	0.55	59.7(66.3)	63.3(69.6)	0.57	62.5(62.5)	69.6(69.6)	0.62
	5	65.1(71.4)	74.6(75.2)	0.55	66.1(66.1)	70.7(70.7)	0.62	59.2(59.8)	65.1(66.2)	0.61

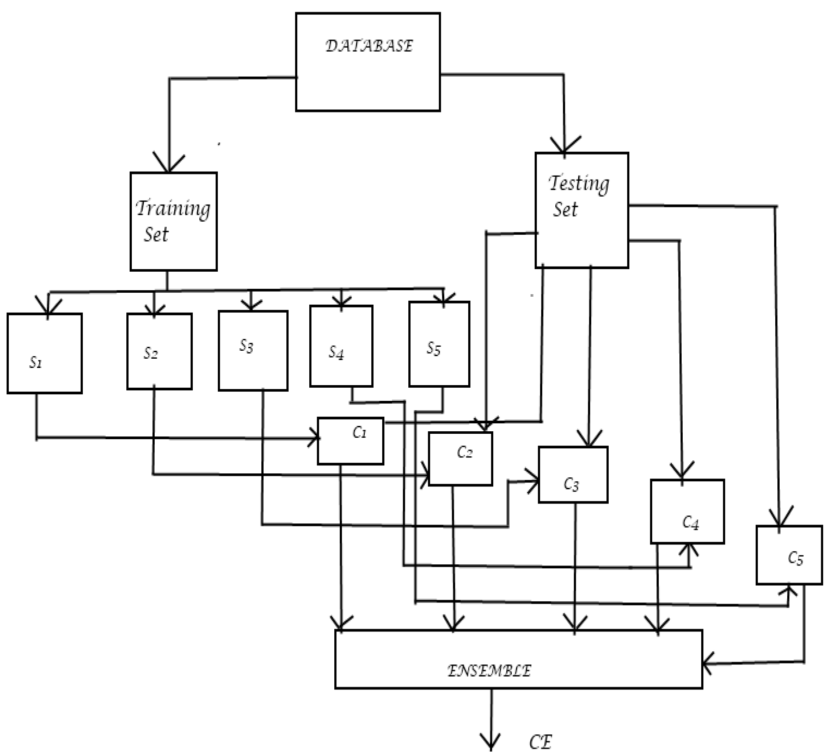

Figure1: "Representation of ensemble algorithm for number of classifiers, S=5"