Indian Journal of Science and Technology, Vol 9(20), DOI: 10.17485/ijst/2016/v9i20/93766, May 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Developing a Hybrid Intelligent Classifier by using Evolutionary Learning (Genetic Algorithm and Decision Tree)

Seyed Ahad Zolfagharifar[1*] and Faramarz Karamizadeh[2]

[1]Islamic Azad University Science and Research, Kohgiluyeh and Boyer Ahmad Branch, Iran; seyedahad.zolfagharifar12@gmail.com
[2]Department of Electrical and Computer Engineering, Shiraz University, Shiraz, Iran; Faramarz.karamizadeh12@gmail.com

Abstract

Objective: The objective of this paper is to give a hybrid classifier by combining the genetic algorithm and decision tree based on evolutionary learning. **Methods:** The proposed algorithm on the 8 data samples was tested. In order to implement the proposed algorithm, MATLAB software was used. In all the obtained results, standardized data sets are used, making assembly by using genetic algorithm which is very suitable. **Results:** The learning technique of sub-spaces is proposed. In this study, we tried to compare a series of different methods and updated of integrated distribution. It showed that, in cases that the number of information or the number of properties are high, the proposed hybrid classification approach that implements genetic algorithm can be used as the best approach. **Conclusion:** In this study, we tried a usual approach for clustering in error prone environments. A main excess in the precision on the tested information or on the validation is clear. It should be noted that this increasing is in comparison with the assembly classifiers which has stable accuracy.

Keywords: Combination of Genetic and Decision Tree, Consensus of Classifiers

1. Introduction

The classifier is a process to put the data within pre-defined categories[1,2]. The use of learning classifiers is an effective approach on machine learning[3]. In this version of knowing, to modify learning precision, the funding of several classifier (including MLP, SVM) combine with each other, which named first stage or production of classifiers phase. Neural networks are best method for initial classifiers[4]. The findings of theory[5,6] and experimental[7–9] indicated that when the learning collective was prior to learning of the best initial classifier that basic classifiers have acceptable performance in the error. Then, the second stage or the integration of classifiers was performed. At this stage, we summarize the classifiers utilized in the later stages as input to other classifiers, until based on the output of the initial stage are decided. Several states for the production and the integration the first stage (production of classifiers) and the second stage (the integration of classifiers) exist[10,11]. In this paper we intend in comparison of the various states of production and the integration of classifiers, then evaluated them and observe their output by the simulation deal, to achieve the optimal solution in combining them in the best way. Recognition and detection systems nowadays have many applications in various fields such as Optical Character Recognition Handwritten (HOCR). In this study, we intend the effects of classifier on the basic classifiers to increase the efficiency of the combined classifiers review.

2. Combination of Classifiers

Multiple combinations of classifiers can be an issue of public pattern recognition wherein inputs are the results of separate classifiers and output is combined decision of them. The distributer with various properties or

* Author for correspondence

techniques competes with each other and eliminates their weaknesses[12]. If multiple various classifiers vote with each other as an assembly the mistake of them are decreased significantly[7,8].

2.1 Hypothesis: Why Combination of Classifiers Works?

Following is the theorem to understand better why the combination of classifiers' performance subtilizes.

2.1.1 Condorcet Theorem

If each voter for saying true with probability of p and the likelihood that the most share of contributors opinion as true to be m next, if p>0.5 , Have to vitally m>p. Briefly, m converge into one for p>0.5 if the amount of contributors to reached infinite[10].

In overall, it proved as:

- For p>0.5, If L reached to maximum value then m = 1.
- For p<0.5, If L reached to maximum value then m = 0.
- For p = 0.5, If L reached to maximum value then m = 0.5.

In it L is the amount of contributors and m is precision of the majority vote[1].

2.2 Definition of Classifiers Assembly

A group of base classifiers which are utilized together to find the answer of the problem of sample identification and decides to raise the performance of the system, via each other when combined.

3. The Definition of Several Classifier Systems

Overall combination of classifiers in the four levels can be performed. First at a high level means in a level that outputs of several classifiers with various methods we combine. Here we have finally several basic classifiers that their decisions by way of as maximum vote with each other are combined. We call this level integration stage. The next stage is the level of classifiers. At this stage various base classifiers used for making composed several classifier systems. The next stage is the level of properties whereas it is proposed that there are a set of properties for each classifier; and a subset of properties are selected. The last stage is the level of information (see Figure 1).

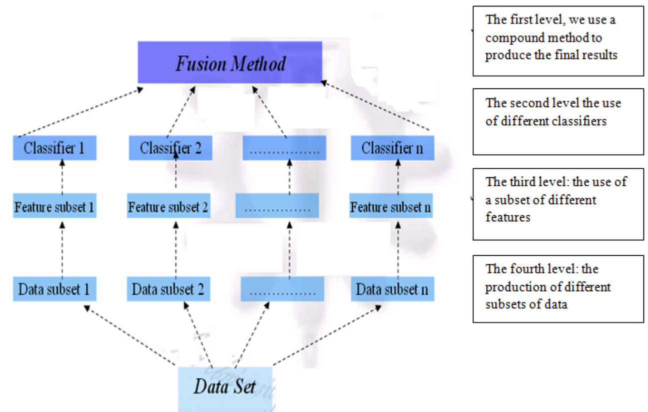

Figure 1. The different levels of production of classifiers assembly.

3.1 The Combined Output of Classifiers

The combined output of classifiers is regulated according to their type in assembly which include:

- The initial version: The outcome of each classifier of D_i for a given instance model is as correct and incorrect such for a Z data group classifier of D_i produces output vector of y_i. y_{ij} is equal to one if classifier of D_i pattern of z_j is correctly classified; And the otherwise will be zero.
- The second version of information: Each classifier of D_i a label of δ_i for an example model of X produces so that for one pattern, one vector $\delta = [\delta_1, \delta_2, \delta_n]$ occurs where L is the amount of classifiers. In this paper there is no data about amount of ensure of labels and even label are not a substitute for a pattern label.
- The third version of information: Here each classifier of D_i gives a prioritization of the most accurate label to the non-accurate label.
- The fourth version of information: Here each classifier of D_i a C-dimensional vector $[d_{i,1},d_{i,2},\ldots,d_{i,n}]$ generates, The amount of $d_{i,j}$ amount of correctness of this hypothesis that the "sample of X is a member of the ω_j group" displays. In most cases, $d_{i,j}$ is in range of 0-1. Now, deal to review the several ways of combining different classifier output.

3.1.1 The Majority Vote

We suppose outputs of classifiers as binary vectors of C-dimensional $[d_{i,1},d_{i,2},\ldots,d_{i,n}]$ that $d_{i,j}= 1$ if classifier of Di sample of X be placed in $\omega_j$group and otherwise $d_{i,j}= 0$i s. Vote of plurality for ω_k group from the following formula is calculated.

$$\sum_{i=1}^{L} d_{i,k} = \max_{j=1}^{c} \sum_{i=1}^{L} d_{i,j} \tag{1}$$

One of the disadvantages of the maximum vote is that the not guaranteed to enhance performance; by three classifiers in 60% accuracy in least of state may be majority vote reaches 40% efficiency. To best performance that achieved with a combination approach, pattern of success and failure mode called the worst performance possible. Although the model of success in instance, the high state is 91% but exists no confirmation of success pattern. Look at the Figure 2:

No.	111 *a*	101 *b*	011 *c*	001 *d*	110 *e*	100 *f*	010 *g*	000 *h*	P_{maj}	$P_{maj} - p$
Pattern of success										
1	**0**	**3**	**3**	**0**	**3**	**0**	**0**	**1**	**0.9**	**0.3**
2	2	2	2	0	2	0	0	2	0.8	0.2
3	1	2	2	1	3	0	0	1	0.8	0.2
4	0	2	3	1	3	1	0	0	0.8	0.2
5	0	2	2	2	4	0	0	0	0.8	0.2
6	4	1	1	0	1	0	0	3	0.7	0.1
7	3	1	1	1	2	0	0	2	0.7	0.1
8	2	1	2	1	2	1	0	1	0.7	0.1
9	2	1	1	2	3	0	0	1	0.7	0.1
10	1	2	2	1	2	1	1	0	0.7	0.1
11	1	1	2	2	3	1	0	0	0.7	0.1
12	1	1	1	3	4	0	0	0	0.7	0.1
Identical classifiers										
13	6	0	0	0	0	0	0	4	0.6	0.0
14	5	0	0	1	1	0	0	3	0.6	0.0
15	4	0	1	1	1	1	0	2	0.6	0.0
16	4	0	0	2	2	0	0	2	0.6	0.0
17	3	1	1	1	1	1	1	1	0.6	0.0
18	3	0	1	2	2	1	0	1	0.6	0.0
19	3	0	0	3	3	0	0	1	0.6	0.0
20	2	1	1	2	2	1	1	0	0.6	0.0
21	2	0	2	2	2	2	0	0	0.6	0.0
22	2	0	1	3	3	1	0	0	0.6	0.0
23	2	0	0	4	4	0	0	0	0.6	0.0
24	5	0	0	1	0	1	1	2	0.5	−0.1
25	4	0	0	2	1	1	1	1	0.5	−0.1
26	3	0	1	2	1	2	1	0	0.5	−0.1
27	3	0	0	3	2	1	1	0	0.5	−0.1
Pattern of failure										
28	**4**	**0**	**0**	**2**	**0**	**2**	**2**	**0**	**0.4**	**−0.2**

Figure 2. Summary of output states of 3 classifiers for ten patterns.

In Table (1) state of a (111) is saying that the amount of models that classifiers of D_1, D_2 and D_3 to do well classification action for them and the state of b (101) is saying that the amount of models that classifiers of D_1 and D_3 have done correct classification action for them and D_2 classifier has classified them incorrectly.

3.1.2 Weighted Maximum Vote

If classifiers in a hybrid assembly don't contain the equal accuracy, it is logic that more accurate classifier has had a more effects on the final decision. But, for di, j we gave the following definition:

$$d_{i,j} = \begin{cases} 1, & if \quad D_i \quad labels \quad X \quad in \quad \omega_j \\ 0, & otherwise \end{cases} \tag{2}$$

Then function of resolution for ω_j class like this obtain:

$$g_i(X) = \sum_{i=1}^{L} b_i d_{i,j} \tag{3}$$

That b_i is parameter for classifier of D_i. In instance, we propose that 3 classifier of D_i and D_2 and D_3 with precision of 0.6, 0.6 and 0.7 have. By integrating the maximum vote approach $p_{maj} = 0.696$, however by combining the weighted maximum vote by coefficient $b_1 = b_2 = 0$ and $b_3 = 1$ (that actually with this work we remove classifiers vote of D_i and D_2) $p_{maj} = p_3 = 0.7$ will be. One way to weighted to classifiers to reach upper accuracy maximum vote is as:

$$b_i \propto \log \frac{p_i}{1 - p_i} \tag{4}$$

That pi is vote precision of classifier of D_i; prodeced that the L classifier of D_i, are independent from each other[1].

4. Proposed Method

Overall, the combined classifiers at four levels or methods are conducted; combination or combination, distribution, properties and information levels.

In this research integration levels are implemented. The integration is a stage that outcomes of several classifiers integrated with various approaches. In this level each classifier weighted within each category and amount of participation of each category will be determined in final classification. This means that each classifier what percentage have has participation in consensus. In the developed approach initially by application of decision tree procedure of data sets divide to various classifiers and each classifier will be classified to subsystem categories, afterwards classifiers will be taught and the output from training and will be saved. The findings of the before stage that represent each category in the classifier of learning shows send to performance function that is defined, by using genetic algorithms the performance will be optimized. Then amount of operation and the diagnosis share of every category in the classifier obtain, and each classifier on the basis of diagnosis percentage of the sub-categories participates in consensus.

4.1 Example

Suppose there are three separate devices of fruit or classifiers (A, B, C), it was intended to use of devices separate the apples and pears from each other, means ever classifier has two state and a total create 6 state. We should dedicate each class from classifiers a value by chance in fact, a value in the range of 0-1, enter that shows the amount of diagnosis the apple from the pear by each tool as a first value (Table 1).

Table 1. Applying random weight to chromosomes

Variety of modes	A classifier		B classifier		C classifier	
	Class 1 (pear)	Class 2 (apple)	Class 3 (pear)	Class 2 (apple)	Class 1 (pear)	Class 2 (apple)
Random value (random chromosome)	0.9	0.4	0	0.1	0.4	0.2

After classification determine classes and selection of basic mode that shows the each of classifier accuracy, to devices (classifier) we give several fruit as sample (data collection) and store the answers that actually is idea of each classifier to a class (Table 2). Here, 50 percent of inputs are apple and another half is pear (Figure 3).

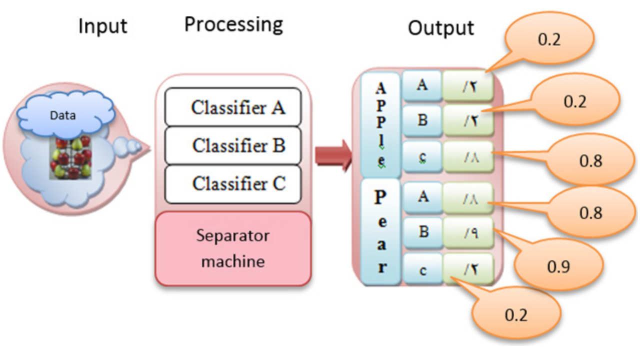

Figure 3. Method of algorithm's work.

Table 2. Calculate the new weight of chromosomes

Variety of modes	A classifier		B classifier		C classifier	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
Vote of each classifier	0.2	0.8	0.2	0.9	0.8	0.2

After specifying the opinion of ever classifier about input data now must achieve consensus of vote for each class.

$$W_{Apple} = \frac{\sum_{i=1}^{k}\left(R_i^{apple} \times D_i^{apple}\right)}{\sum_{i=1}^{k}\left(\llbracket (R) \rrbracket_i^{apple}\right)} \times P(Apple) \qquad (5)$$

In the equation of **Riapple** initial values classifiers to the second state, **Diapple** values are gained in the first implementation of device for the second state, K the amount of classifiers and P percentage of state information to the total information that in this sample, half of data apple and the other is considered pear.

In the above instance, based the given values and the consensus of votes each classifier of input information, pear or Class 1 has been specified, in continue because the percentage of classifier precision is low we change initial values. This work is done several times until precision of procedure is close to the requested level. In fact, with learned information set teach the algorithm and in finally choose among all outputs of the input data related to the best precision. And at the end use the real data regulates to test the procedure.

4.2 The Proposed Algorithm

In the developed algorithm initial decision tree algorithm of information set train by using several classifier that each classifier is composed of several level and the obtained results send to the genetic algorithm to find function of performance means to find the best possible answer in each category and then calculate the percentage of participation of each classifier in each category achieve until existence categories in each classifier participate in the consensus. First you note to how to work of each algorithm.

4.2.1 Decision Tree Procedure

This is a decision tree procedure, about a particular purpose and several features analysis to predict the situation and provide targeted sales. Decision trees are used to forecast parameters hands, classification trees are named, since the examples are placed in categories. Decision trees utilize to predict the continuous parameters are named regression trees[2]. Majority decision-tree learning procedures according to a greedy search for up-down act on the basis of trees. ID3 decision tree from the one statistical value named Information Gain were used. Until we determine how much a feature can separate examples of training based on categorization.

4.2.1.1 Entropy

The sincerity degree (irregularities) specifies a group of samples. If the group of S includes examples from a concept of target entropy of S relative to this Boolean categorization is as:

$$E(s) = -(p\log_2 p + \overline{p}\log_2 \overline{p}) \quad (6)$$

4.2.1.2 Data Gain

Data Gain is properties of entropy reduction achieved through isolated examples of this property. The (S, A) for a property such as A to group of instances of S is as:

Informationgain = Entro(s) -Σ sv|s|v ∈ values (A) * Entropy(s)

(7)

Where (A) is the group of all value of properties and VS A subsystem of S that for it A has the value of V. In the above definition the first expression is value of information entropy and the second expression is the expected value of entropy after the aweing of information[2].

4.2.2 Genetic Algorithm

Genetic algorithm is one of the most powerful methods of meta-heuristic to find minimum and maximum points of a target function. The algorithm's work is the way in which the parameters of search space in the form of strings called chromosomes are formed. Each chromosome represents a solution to the target problem. Chromosomes with together constitute a set that is called population and at the start of operation elements of initial population are randomly selected. The algorithm repeatedly imposed on population elements of two operator the cutting and mutation and from a population creates new population. Usually to results of a population are called generation.

Finally after finite repeated the desired results produce in the last generation. It is obvious that all the answers are not always the optimal solution. To determine the being optimal amount of each answer from a criterion are used that the objective function is called. In action the objective function to each population chromosome from one generation, a value assigns, that this amount, indicates the fitness of the answer than the other answers of same generation.

5. Evaluation Criteria

It is important to know how the classifier does as well. Performance of concept classification is fusion and complex the most important part of it is the classification accuracy[1]. In one approach at least can be named reason that we wish to know the amount of generalization of the classification; one is to see if the classification be done useful sufficiently and another compare its performance with other similar work.

For evaluating the performance of the proposed algorithm from the two criteria is used. One variance error or error of error and another statistical analysis, these two criteria are accurate criteria and in the completed tasks as the main criteria of classifier quality from them were used[1].

5.1 Evaluation of the Various Methods

To measure the various procedures and comparing approaches by offering method initially approaches and the developed technique on information groups (Table Appendix 1) implemented (on each of the ten runs) then the percentage of precision and mistake of each approach in each run obtain and in an Excel file will be entered, the average and standard deviation of the accuracy and error of each method we gain on each set. The average obtained in decreasing order until we algorithms according to precision will be saved and evaluate which approach is the most accurate in information collection.

5.2 Accuracy in Various Ways

For performing this research, we classifies twenty classification algorithm instances as Table Appendix 2 table, choose stochastically and with information set they've conducted them 10 times, and save findings. A weka software algorithm is used to run the sample. Weka, a collection of machine learning procedure and tools for information preprocessing is. This software is designed which can be quickly, existing methods as flexibility set on the new set of data. In accordance with the broad range of learning algorithms, this software includes a variety of tools for preprocessing of data. This diverse and comprehensive toolbox is available through a usual display so that the applicant can evaluate with each other the different methods and techniques that are most suitable to address the issues, recognize.

5.3 Comparative results

If average of results of the existing classifiers in previous part put on together and sort the results based on the data (Table Appendix 3 and 4) tables show a better result of work.

5.4 Displayed Diagrams

To better analyze of findings in this section by using multiple graphs show the situation of classifiers (Figure 4).

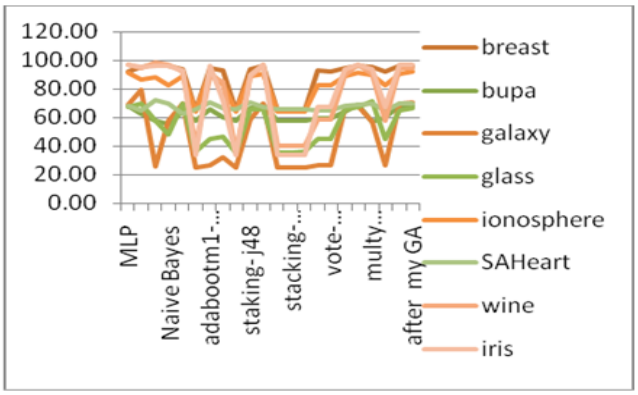

Figure 4. Average of implementation of classifiers.

In the chart above we can see differences between the means. In this chart clearly there are significant differences between the information, for example, information set of galaxy that with green Colour is displayed note that, it would obvious that the outcomes of this data set from 20% to 80% that related to the classifier of KNN is and correct is diagnosed. To view the mean and standard deviation the percentage of the accuracy of classifiers to data set the Figure 5 is in continue. As you can see this chart shows the mean with red line shows square up of the red line and the square below demonstrates the criteria deviation of the findings.

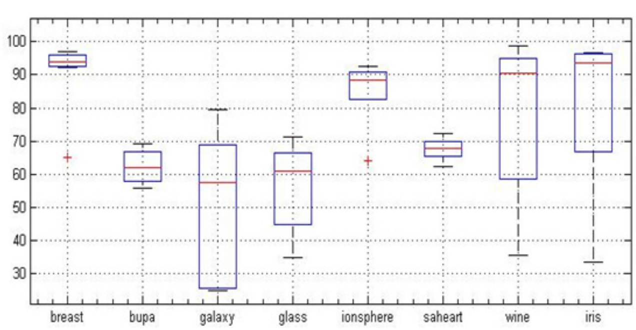

Figure 5. Mean and standard deviation for results of classifiers implementation.

6. Conclusion

According to the growing need for intelligent computing effectively of last few decades, enhancing a categorization algorithm with better efficiency is important. Previous studies showed that problems can be solved by conventional classifications algorithms however about complicated issues with the most complexity, they are inefficient.

Even resolution problems while that our understanding from the relationships among categories is low can be so complicated. In order to overcome this problem, the learning technique of sub-spaces is proposed. In this paper, we tried to compare a series of different methods and developed a hybrid intelligent classification approach. Results indicated that in the cases where the amount of data or the amount of features is high, proposed hybrid classification method based on genetic algorithm is the best method. In this study, we tried a usual approach for categorization in error prone environments.

7. References

1. Kuncheva L. Combining pattern classifiers, methods and algorithms. New York: Wiley; 2005.
2. Fukunaga K. Introduction to statistical pattern recognition. Orlando, FL: Academic Press, Inc; 1972.
3. Vapnik V, Lerner A. Pattern recognition using generalized portrait method. Automation and Remote Control. 2009; 24:774–80.
4. Japkowicz N, Stephen S. The class imbalance problem: A systematic study. Intelligent Data Analysis Journal. 2002; 6(5):429–49.
5. Akbani R, Kwek S, Jakowicz N. Applying support vector machines to imbalanced datasets. Proceedings of European Conference on Machine Learning; Pisa, Italy. 2014. p. 39–50.
6. Wu G, Chang EY. Class-boundary alignment for imbalanced dataset learning. Proceedings of the ICML'03 Workshop on Learning from Imbalanced Data Sets; Washington, DC. 2003.
7. Skalak B. Prototype and feature selection by sampling and random mutation hill climbing algorithms. Proc 11th Int Conf on Machine Learning; New Brunswick, N.J. Morgan Kaufmann, Los Altos, CA. 1994. p. 293–301.
8. Kohonen T. Improved versions of learning vector quantization. Proc International Joint Conference on Neural Networks; San Diego, CA. 1990 Jun. p. I545–50.
9. Godara S, Singh R. Evaluation of predictive machine learning techniques as expert systems in medical diagnosis. Indian Journal of Science and Technology. 2016 Mar; 9(10):1–14.

10. Marquis JA. Sur les elections par scrutiny. Histoire de l'Academie Royale des Sciences. 1984; 44:31–4.
11. Montazeri-Gh M, Mahmoodi-k M. An optimal energy management development for various configuration of plug-in and hybrid electric vehicle. Journal of Central South University. 2015; 22:1737–47.
12. Mohana S. A position balanced parallel particle swarm optimization method for resource allocation in cloud. Indian Journal of Science and Technology. 2015 Feb; 8(S3):182–8.

Appendix

Table Appendix 1. Set of data used

Row	Set	Explanation
1	Breast	This data set is related to risk factors for breast cancer in women of America.
2	Bupa	This data set is analysis of some liver disorders
3	Galaxy	This data set is classified set of radar return information from ionosphere
4	Glass	This data set is related to glass, produced glasses
5	ionosphere	this of this data set for classification of information of eleven radar That the signal sent or received to the ionosphere
6	SA-Heart	of this data set related to men who have been diagnosed with heart disease
7	wine	This data set is related to recognition of beverages
8	Iris	This data set related to understand the lily flowers

Table Appendix 2. Table of sample classification algorithms

Row	Name of input classifier	Base classifier (output)
1	MLP	
2	IBK or KNN	
3	SMO or SVM	
4	Naive Bayes	
5	random forest	
6	random tree	
7	stacking	bagging
8	adabootm1	Decision stump
9	bagging	decision stump
10	multy scheme	decision stump
11	vote	decision stump
12	staking	desionstump
13	multy scheme	filtered classifier
14	staking	j48
15	stacking	MLP
16	vote	MLP
17	vote	rbf classifier
18	vote	reptree
19	stacking	tree_j48
20	vote	zero R

Table Appendix 3. The results of different methods of assembly classification on the various set (highlight values the best results in each data set)

Name of input classifier	breast	bupa	galaxy	glass
after my GA	96.05	68.02	70.92	66.09
MLP	92.34	68.32	68.76	67.06
IBK or KNN	95.67	62.96	79.60	69.44
SMO or SVM	96.93	58.09	25.33	57.57
Naive Bayes	96.31	55.68	57.86	48.22
ranodom tree	93.79	64.00	70.31	69.95
random forest	65.01	57.97	24.77	35.51
adabootm1- Decision stump	94.58	66.00	26.07	44.81
bagging- decision stump	93.22	60.75	31.67	46.45
staking- decision stump	65.01	57.97	24.77	35.00
staking- j48	93.75	66.78	58.33	65.19
staking- MLP	95.99	68.99	69.81	65.93
stacking- tree_j48	65.01	57.97	24.77	35.28
stacking- bagging	65.01	57.97	24.77	34.95
vote- zeroR	65.01	57.97	24.77	35.51
vote - rbf classifair	92.80	59.97	26.01	44.81
vote- decision stump	92.49	59.54	26.04	44.86
vote - reptree	94.41	64.55	66.69	65.37
vote- MLP	96.15	69.04	68.58	66.78
multy scheme-filtered classifier	95.14	57.62	56.87	71.17
multy scheme- decision stump	92.50	60.55	26.01	44.81
mean	88.54	62.49	48.29	54.68
SD	12.41	4.44	21.56	13.63

Table Appendix 4. The results of different methods of assembly classification on the various set (highlight values the best results in each dataset)

Name of input classifier	ionosphere	SA-Heart	wine	iris
before my ABC	92.71	68.82	92.08	96.00
after my ABC	92.71	69.57	94.34	96.67
before my GA	90.57	70.22	95.14	96.67
after my GA	92.29	69.68	95.14	96.67
MLP	91.51	68.48	97.19	96.47
IBK or KNN	86.75	64.81	94.83	95.40
SMO or SVM	88.03	72.36	98.82	96.00
Naive Bayes	82.48	70.15	97.36	95.53
ranodom tree	88.77	62.36	92.19	92.80
random forest	64.10	65.37	39.89	33.33

adabootm1- Decision stump	91.11	70.56	89.27	95.67
bagging- decision stump	83.05	66.32	85.34	72.20
staking- decision stump	64.10	65.37	35.45	33.33
staking- j48	89.37	70.58	88.76	89.53
staking- MLP	90.94	67.58	97.42	96.67
stacking- tree_j48	64.10	65.37	39.89	33.33
stacking- bagging	64.10	65.37	39.89	33.33
vote- zeroR	64.10	65.37	39.89	33.33
vote - rbf classifair	82.48	64.72	58.60	66.67
vote- decision stump	82.54	65.19	59.10	66.67
vote - reptree	89.37	68.31	90.28	93.40
vote- MLP	91.17	68.74	97.19	96.67
multy scheme-filtered classifier	90.03	70.00	90.95	93.53
multy scheme- decision stump	82.45	65.15	58.20	66.67
mean	83.29	67.52	77.80	77.77
SD	10.61	2.58	23.61	25.43