

# Increasing Classifier Ensemble Efficiency using KSBC Algorithm

Elham Masoumi Nogorabi

Department of Computer Science  
Qazvin Branch, Islamic Azad University Qazvin, Iran

Hedieh Sajedi

Department of Computer Science  
School of Mathematics, Statistics and Computer  
Science, College of Science, University of Tehran,  
Tehran, Iran

## ABSTRACT

The most challenging issue in constructing a classification system based on ensemble is how to construct an appropriate ensemble of basic classifiers. In this paper, a new approach of constructing ensemble namely, KSBC, K-means Based Classifier Selection is introduced. This approach utilizes Bagging algorithm as the producer of basic classifiers. Type of all basic classifiers, decision tree or multi-layer nervous networks are considered and remained unchanged during construction of ensemble. After constructing a large number of basic classifiers, KSBC partitions them with the help of k-means clustering. Afterward, by choosing one classifier from each partition, final ensemble is constructed. Weight voting method is the assembling function of the ensemble. In addition the approaches for selection of a classifier from each partition were analyzed. The effectiveness of sampling rate on the efficiency of combining classification based on clustering of classifiers was also evaluated. Finally, several tests were carried out on a large number of standard datasets in machine learning database. Experimental results illustrate the effectiveness of the proposed method compared to other approaches.

## General Terms

Algorithms, Classification, Distributed Data Mining

## Keywords

Ensemble method, bagging, clustering

## 1. INTRODUCTION

An ensemble has a vital need for diversity, that is if an ensemble of classifiers is to be good as a plural classification, they should be diverse as enough as to cover each others' errors. Therefore, while constructing an ensemble, we need a mechanism to guarantee the diversity of the ensemble. Sometimes this mechanism is in a way that a subset of construct basic classifiers are selected and removed to maintain the diversity of the ensemble. In ensemble-based learning it is possible that the data is fractioned to smaller parts and simple and quick models are created for each part and final ensemble model is performed in a way without significant time overhead [20]. To learn any problem, many classifiers have been introduced so far [3]. Each of the proposed classifiers has a lot of strengths and weaknesses that makes them suitable for specific problem. But there is no strong solution for what is best classifier. Fortunately ensemble-based learning is a powerful approach to provide an optimal classification system for each problem. The most challenging issue in constructing an ensemble based

classification system how constructing a proper ensemble of base classifiers [20]. In this paper, we tend to analyze the effect of bagging and boosting classification methods by clustering classifiers method and improve the accuracy. In previous approaches from every cluster only one classifier was selected; one of the other aims of this study is to determine whether it is possible don't select any classifier from a cluster and have a better ensemble. To justify such an action, it has to be mentioned that in clustering of a typical dataset, some of clusters are created for a null data (include a null data practically) or Include some noise data. So the effect of removing these clusters and also the effect of training data sampling rate on the efficiency of classifiers clustering approach were analyzed. It has to be mentioned that in order to reach the diversity needed among primary classifiers, every one of them were trained on random sampling of training data which is sampling  $b\%$  of total training samples. In the second section of the article previously used approaches and their weaknesses is analyzed. Proposed approach is discussed in the third section and experimental results of applying it to standard datasets is discussed in the fourth section. The fifth part includes conclusion and future works.

## 2. RELATED WORKS

Ensemble learning is a powerful approach to provide any problem with an almost optimal classification system. Generally it is true that combining different classifiers leads to increase in classification efficiency [20]. This shows importance of diversity, in success of an ensemble. Diversity among basic classifiers signifies their independence. This means that failure in classification of basic classifiers will not take place simultaneously on one template. Kuncheva [11] has shown that, theoretically and practically, an ensemble of some basic classifiers always increases classifications accuracy provided that basic classifiers are independent. It has also been shown that ensemble philosophy can be also applied to Bayesian networks [20]. Kuncheva [17] has proposed some methods to measure an ensemble's diversity and then has suggested a model to build a diverse ensemble based on a search method. As described before, application of clustering classifiers has been discussed and also their efficiency has been studied [18], [21]. In previous works, classifiers in the ensemble have performed their learning process randomly and each one on the entire training set. Combining a number of classifiers leads to formation of classifiers with higher efficiency [2], [3]. It has been proved that combining

classifiers in the ensemble are more efficient than individual ones in average [4]. Since classifiers with different features or different methodology can complement each other and cover each other's disadvantages; combining methods often lead to improve classification efficiency [14]. Kuncheva in [11] demonstrates by kondorset theorem that usually a combination of classifiers performs better than a single one, i.e. if different classifiers are used in combination, their total error may notably decrease.

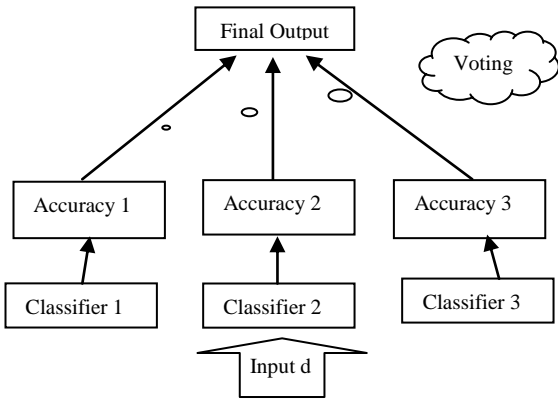


Figure 1. An ensemble with three classifiers

According to figure 1 input d is applied to three classifiers and voting is done regarding the similar outputs and the one with the most accuracy is selected as ensemble's output. Each of above classifiers is called a basic classifier. Assume that accuracy of each basic classifier is 80%. So accuracy of above combining classifier is equal to application of voting which is calculated later on. The input is considered correct only if at least two classifiers have correct selection.

Therefore total accuracy (the probability of selecting correct output) is equal to sum of probabilities that only two classifiers selected correct output plus the probability of all three classifiers correct selection which is calculated by equation (1):

$$\text{Accuracy} = \binom{3}{2} \times 0.8 \times 0.8 \times 0.8 \times 0.8^3 \quad (1)$$

It can be proved that with more basic classifiers, system output and decision will be more precise [21]. This description is true when classifiers accuracy (the probability of correct selection by each classifier) is independent and linear in relation to each other. On the other hand if their accuracy is dependent, for example in the worst case where classifiers are all the same, output of all classifiers is the same; so voting is a useless action and accuracy of combining classifier is equal to each basic classifier. So, firstly, basic classifiers should have high accuracy (better than random) and secondly they should have diversity (all should be independent and as not similar as possible). Moving forward to combining classifiers is due to their distribution feature and execution ability on different strings which improve classification efficiency. Classification and clustering as two main tasks in data mining can use methods of ensemble for two reasons; increasing efficiency and their feature of distribution. Distributional data mining can be a committee of learners, distributed decision tree or gathering diverse data of training a classifier on every one of them. As mentioned in [1] and [13], descriptive framework of distributed classification problem can be studied in three dimensions. These dimensions include data dimension, learner dimension (basic classifiers) and learner connection level dimension.

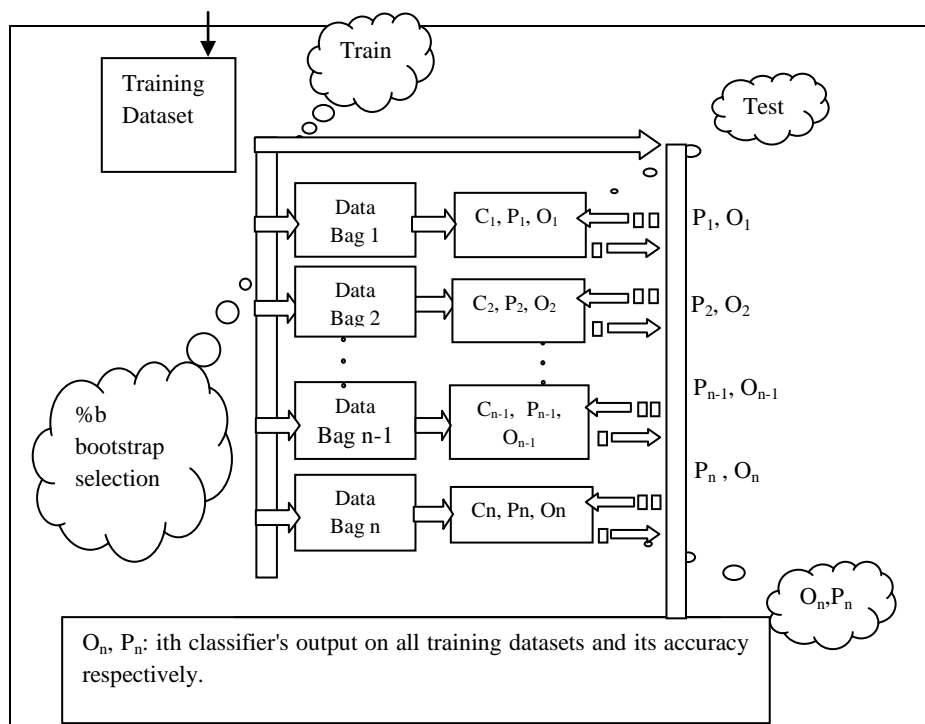


Figure 2. Training phase of bagging method

Many methods have been proposed to create diversity. For example different subsets of train samples are created and independent classifiers are created upon them. Bagging and boosting methods are among most famous methods [4], [6]. In these methods, rather than using all the samples, their subsets are used to train basic classifiers. The term bagging is first used by [3] abbreviating for bootstrap aggregating. The idea of bagging is simple and interesting: the ensemble is made of classifiers built on bootstrap copies of the train set. Using different train set, the needed diversity for ensemble is obtained. In order to promote model diversity, bagging trains each model in the ensemble using a randomly-drawn subset of the training set [4]. Figure2 illustrate the training phase of bagging method in generally. Boosting involves incrementally building an ensemble by training each new classifier to emphasis the training instance that previous classifiers misclassified [14], [16]. Figure3 illustrate the training phase of boosting method in generally.

In another level, features are used to produce diversity among classifiers, in a way each classifier uses a subset of features [10], [14], [5]. In a different look, different kinds of classifiers are used in combination to make the diverse [11]. In all above methods if we create many classifiers there is a possibility that some of them always give the same output. So, after creating some classifiers with a given method, a selecting mechanism should be applied to select appropriate classifiers and our proposed approach is a proper solution.

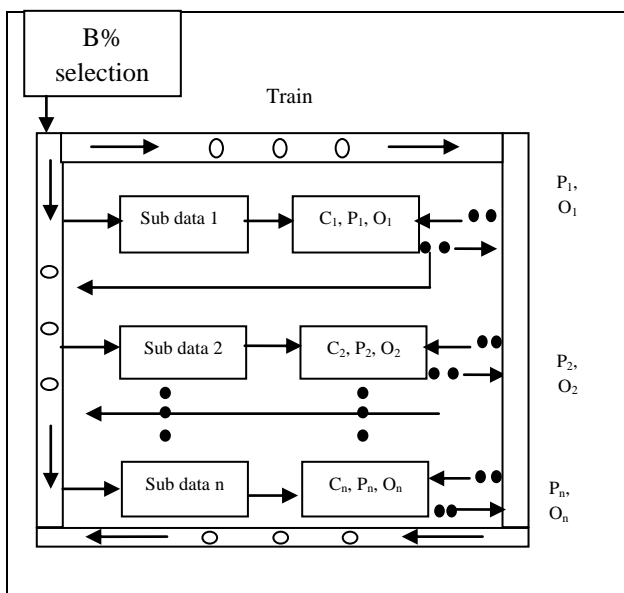


Figure 3. Training phase of Boosting method

### 3. PROPOSED APPROACH KSBC

The main idea of combining classification based on k-means clustering (KSBC) is to use the most diverse subsets of basic classifiers obtained by Bagging and Boosting methods. In fact, firstly some basic classifiers are produced based on two famous mechanisms of creating ensembles (Bagging and Boosting). Then produced classifiers are partitioned by an algorithm according to their outputs on the entire set of training data. After that one classifier is selected randomly

from each cluster/partition. Since each cluster is created based on classifier outputs, it is so likely that by selecting one classifier from each cluster and considering them as an ensemble, the formed ensemble is a diverse one and is better than ensembles created by Bagging and Boosting approaches; i.e. from each cluster the classifier with the least distance from the centre of cluster is selected. The pseudo code of KSBC algorithm training part is shown in figure 4. In this figure, bagging approach is used to create basic classifiers.

---

Input:  
 TS: Training Set  
 L: Labels of Training Set  
 n: Ensemble Size  
 b: Ration of Sub sampling from TS

Output:  
 E: Ensemble of Classifiers  
 P: Accuracies of Classifiers  
 O: Outputs of Classifiers

```

m=length (TS)
For i = 1: n
    Bag = sub sample (TS, b)
    Train (Classifier, Bag)
    For j = 1: m
        O (i, j) = Test (Classifier, TS (j))
    End
    P (i) = Accuracy of Classifier (i)
    E (i) = Classifier
End
    
```

---

Figure 4. The pseudo code of KSBC algorithm training part

N subsets of training set are produced as sampling without displacement.  $i^{th}$  data subset which is sampled as  $b\%$  of training set, is called  $i^{th}$  data bag and is shown as  $DB_i$ . This is obvious that the number of samples in  $DB_i$  is  $m*b/100$ . Then a basic classifier is trained on each  $DB_i$  and classifier trained on  $DB_i$  is called  $C_i$ . Then  $C_i$  is tested on the entire training set TS and its accuracy is calculated. The classifier's  $i^{th}$  output on  $j^{th}$  data from TS is a vector shown by  $O_{ij}$ .  $O_{ij}^k$  is about the degree of  $i^{th}$  classifier's certainty that  $j^{th}$  data belongs to  $k^{th}$  class.  $i^{th}$  classifier's output on all training datasets is shown by  $O_i$  and its accuracy is shown by  $P_i$ . The only difference between approaches of producing basic classifiers in proposed method and Bagging method lies in sampling rates. In bagging method, the value of parameter  $b$  is always 100 but in our proposed method, to produce basic classifiers the parameter value is set to a number between 30 and 100.

When boosting method is used to create basic classifiers in the ensemble, a subset including  $b\%$  of training set is created at the beginning. Then the first basic classifier is trained upon this subset. After that this first basic classifier is tested upon the entire training set which leads to creation of  $O_1$  and  $P_1$ . A subset including  $b\%$  of training set is selected using  $O_1$  and  $O_2$  and  $P_2$  are produced. This mechanism continues until  $i^{th}$   $b\%$  subset of training set is created according to  $O_1, O_2 \dots O_{i-1}$ . The only difference between approaches of producing basic classifiers in proposed method and Boosting method lies in the amount of parameter  $b$ . In

Boosting method  $b$  is always 100 but in proposed method this parameter is a number between 30 and 100.

The pseudo code of KSBC framework is shown in figure 5. This framework is explained generally in figure 7 too. In the framework of KSBC a dataset DC is created at first that each data is a classifier. The  $i^{\text{th}}$  data of DC is the  $i^{\text{th}}$  classifier. Assume that our data has  $f$  features.  $P^{\text{th}}$  feature of  $i^{\text{th}}$  data is shown as  $X_{ip}$ , that is obtained from equation (2):

$$X_{ip} = O_{ij}^k \quad (2)$$

Input:

E: Ensemble of Classifiers  
P: Accuracies of Classifiers  
O: Outputs of Classifiers  
th: Threshold of Selecting a Classifier

Output:

SC: Ensemble of Classifiers  
SP: Accuracies of Classifiers  
SOC: Size of Cluster

[n m] = size (O)

C = K-means (O, r)

Cluster of Classifiers (1: r) = { }

For  $i = 1: n$

CC (i) = CC (i)  $\cup$  {C (i)}

Acc\_C(i) = Acc\_C(i)  $\cup$  {P(i)}

End

$j=0$

For  $i = 1: r$

SOC = | CC (i) |

If (SOC > th)

$j=j+1$

Temp = Random Select (1: SOC)

SC (j) = CC (temp)

SP (j) = Acc\_C (temp)

End

End

Figure 5. The pseudo code of KSBC algorithm

Where  $j$  and  $k$  are calculated by equations (3) and (4), in which  $c$  is the number of classes.

$$j = \lceil p / c \rceil \quad (3)$$

$$k = p - j * c \quad (4)$$

Features of DC dataset are viewpoints of different basic classifiers on main dataset's data. So we have a new dataset with  $n$  data (each of which is a classifier) and  $f$  features ( $F=m*c$ ). Parameter  $n$  which is a predefined parameter by user is the number of classifiers produced by Bagging or Boosting approaches. After creating dataset DC, we divide it to many partitions by a clustering algorithm. The number of partitions is shown by  $r$ . Classifiers in a partition have similar output. This means they have low diversity. So it is better to select one classifier from each cluster. In order to avoid null classifications, clusters with fewer members than a threshold number are omitted. Assume that  $E$  is an ensemble of  $n$

classifiers  $\{C_1, C_2 \dots C_n\}$  and there is  $c$  classes in the problem. then assume that by applying the ensemble on data  $o_j$  a matrix  $D_j$  like equation (5) is resulted:

$$D^j = \begin{bmatrix} O_{1j}^1 & O_{2j}^1 & \cdot & O_{nj}^1 \\ \cdot & \cdot & \cdot & \cdot \\ O_{1j}^{c-1} & O_{2j}^{c-1} & \cdot & O_{nj}^{c-1} \\ O_{1j}^c & O_{2j}^c & \cdot & O_{nj}^c \end{bmatrix} \quad (5)$$

Now ensemble decides whether data  $o_j$  belongs to class  $q$  according to equation (6):

$$q = \arg \max_{i=1}^c \left| \sum_{k=1}^n w_k * D_{i k}^j \right| \quad (6)$$

in which  $w_j$  is the weight of  $j^{\text{th}}$  classifier effect which in optimal situation results from equation (7) [7]:

$$w_j = \log \frac{p_j}{1 - p_j} \quad (7)$$

$$p_k = \frac{\sum_{j=1}^m \sum_{i=1}^c |L_i^j - O_{kj}^i|}{c * m} \quad (8)$$

in which  $p_j$  is the weight of of  $j^{\text{th}}$  classifier effect on training set (TS). Notice that in equal terms we use random function in equation (6). Assume vector  $L^j$  for data  $o_j$ ,  $L^j_O$  is one if  $o_j$  belongs to class  $q$  and is zero if not. Now the accuracy of classification  $C_u$  on TS is calculated from equation (8). Figure 6 shows a block diagram of this method. In the proposed method for constructing ensemble, after a number of base classifiers were constructed based on bagging and boosting method, KSBC partitions them with the help of k-means clustering. Then randomly classifiers are selected from each cluster. If at least two classifiers had each cluster, are store all classifiers and the act of voting is done. Finally the accuracy of this method is obtained. But if that there were less than two classifiers for each cluster, the cluster will be removed.

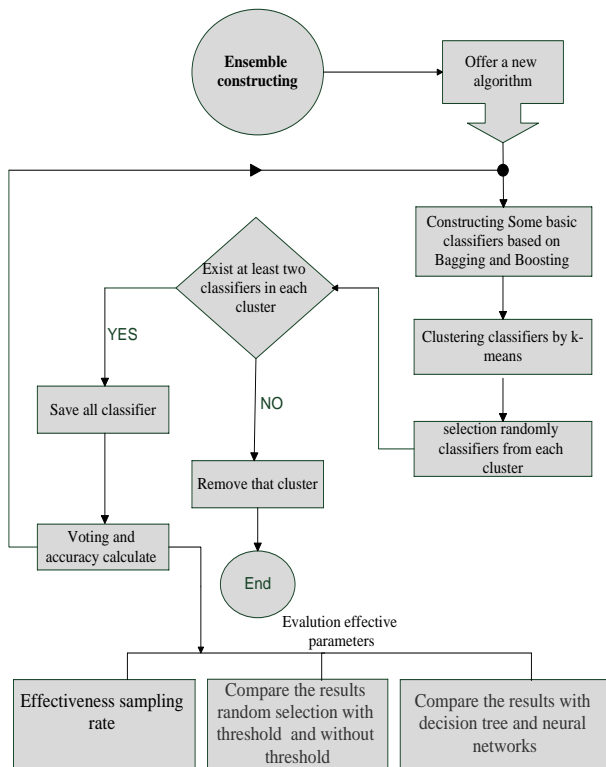


Figure 6. Block diagram the accuracy calculation

#### 4. EXPERIMENTAL RESULTS

In this section experimental results of KSBC on some standard datasets are reported. All accuracy is calculated by 4-fold-cross-validation method. A short explanation of datasets used in this paper is given in table 1.

Table 1. A short explanation of datasets used

Dataset	Number of classes	Number of features	Number of instances
Iris	3	4	150
Wine	3	13	178
Yeast	10	8	1484
SA-Heart	2	9	462
Breast Cancer	2	9	683
Bupa	2	6	345
Glass	6	9	214
Galaxy	7	4	323
Half-Ring	2	2	400
Ionosphere	2	34	351
Balance Scale	3	4	625

All datasets are extracted from Machine Learning website [2] except for halfring dataset which is handmade and is considered as a difficult dataset for machine learning algorithms. All parameters of classifiers are constant during creation of an ensemble. Criterion of decision tree is Gini criterion and parameter of pruning is set to 2.

Gianito method does not use any particular method for produce basic classifiers. It is a method for producing a number of classifiers without sampling rate. All classifiers generated in this way on the whole set of training data, the learning process have followed. Figure 7 represents the effect of sampling rate on KSBC algorithm performance with basic classifiers of decision tree. These graphs report the average accuracy calculated on all datasets.

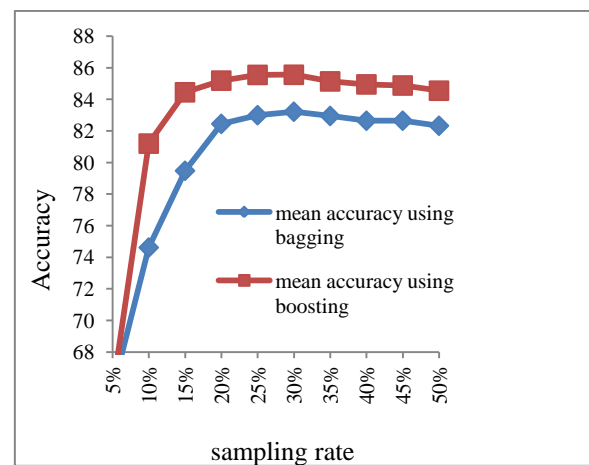


Figure 7. Average efficiency of KSBC algorithm on all datasets with  $r=11$ (number of cluster) and  $n=151$ (number of classifiers), appropriate cluster threshold of 2 and basic classifiers of decision tree

It can be understood from figure 7 that sampling rates more than a threshold value not only doesn't increase classification accuracy but also decreases it. One choice is 30-35%. Table 2 shows results of different values with approaches of Bagging, Boosting and Gianito, parameter  $n=151$ , sampling rate of 30%, appropriate cluster threshold of 2 and basic classifiers of nervous networks. This table has Bagging and Boosting approaches for comparison. As you can see proposed approach has a significant advantage over other approaches. The bold value in each dataset shows the highest accuracy.

**Table2. Results of different values with approaches of Bagging, Boosting and Gianito, parameter n=151, sampling rate of 30%, appropriate cluster threshold of 2 and basic classifiers of nervous networks**

Dataset	Ada boost	Arc-X4	Bagging	Classifier Selection By Bagging	Classifier Selection By Arc-X4	Cluster and Selection
Iris	94.67	96.62	96.62	<b>97.97</b>	97.33	93.33
Wine	94.38	96.59	96.06	<b>97.19</b>	95.51	95.23
Yeast	59.50	60.85	<b>61.19</b>	<b>61.19</b>	60.85	60.56
SA-Heart	69.48	<b>73.04</b>	72.39	71.52	71.09	70.18
Breast Cancer	96.49	<b>97.06</b>	96.91	96.91	96.47	96.19
Bupa	68.41	70.06	71.22	<b>72.09</b>	68.02	71.98
Glass	66.36	66.04	66.98	<b>67.45</b>	66.04	67.05
Galaxy	85.14	<b>87.00</b>	85.62	85.62	84.52	<b>87.00</b>
Half-Ring	<b>97.25</b>	<b>97.25</b>	95.75	<b>97.25</b>	<b>97.25</b>	<b>97.25</b>
Ionosphere	89.17	90.03	88.51	<b>90.31</b>	87.64	88.51
Balance Scale	93.12	93.27	91.99	91.35	92.95	<b>95.75</b>
Average	85.52	86.45	85.50	<b>86.57</b>	85.70	86.10

Table 3 repeats the results of table 2 with decision tree. What was concluded from table 2 can be seen here too. As you can see proposed approach with Bagging algorithm has great advantage over other ones. The degree of this advantage

is higher here than table 2. This is for rather instability of basic classifiers of decision tree compared to nervous networks.

**Table 3. Results of different values with approaches of Bagging, Boosting and Gianito, parameter n=151, sampling rate of 30%, appropriate cluster threshold of 2 and basic classifiers of decision tree. (The bold value in each dataset shows the highest accuracy)**

Dataset	AdaBoost	Arc-X4	Bagging	Classifier Selection By Bagging	Classifier Selection By Arc-X4	Cluster and Selection
Iris	94.67	<b>96.62</b>	95.27	<b>96.62</b>	95.95	94.59
Wine	96.63	96.07	<b>97.19</b>	<b>97.19</b>	95.51	92.61
Yeast	<b>54.78</b>	53.17	53.98	53.98	52.09	54.51
SA-Heart	67.32	70.00	71.30	<b>72.61</b>	69.70	68.04
Breast Cancer	96.19	95.74	96.32	<b>96.47</b>	95.05	93.68
Bupa	66.96	70.64	72.09	<b>72.97</b>	66.28	64.53
Glass	70.09	65.04	<b>70.28</b>	<b>70.28</b>	62.26	60.85
Galaxy	71.83	70.59	<b>73.07</b>	72.45	70.28	70.94
Half-Ring	<b>97.25</b>	<b>97.25</b>	95.75	<b>97.25</b>	95.75	95.75
Ionosphere	91.17	90.31	<b>92.31</b>	91.45	89.74	87.64
Balance Scale	91.52	94.44	93.60	<b>94.72</b>	94.24	94.44
Average	84.54	84.45	85.16	<b>86.38</b>	83.38	82.88

Since each of the classifiers that have been partitioned in a similar outputs, and this means that they have little variation. So it's best to be selected from each cluster one classifier. So selection random a classifier of each cluster with threshold, for avoid being of the noisy classification, Clusters with fewer members than a threshold value are removed. As expected in the case selection a classifier of each cluster

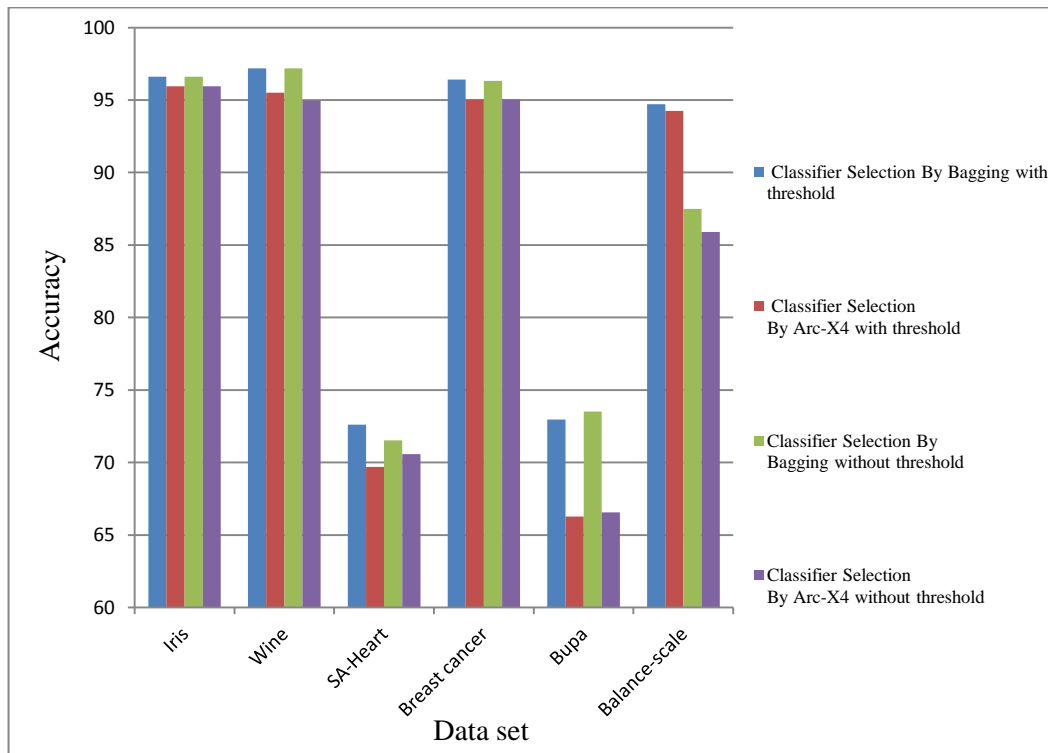
randomly without threshold with generator bagging as base classifiers, because the noisy data are too, classification accuracy reduced. Table 4 proves the claim. So accuracy in the proposed method with selection random a classifier of each clustering, with threshold and generator bagging in majority the dataset is best case.

**Table 4. Compare the results of the proposed method with selection random a classifier of each cluster, with cluster threshold 2 and selection random a classifier of each cluster without threshold with basic classifiers of decision tree**

Dataset	Arc-X4	Bagging	Classifier Selection By Bagging with threshold	Classifier Selection By Arc-X4 With threshold	Classifier Selection By Bagging without threshold	Classifier Selection By Arc-X4 without threshold
Iris	<b>96.62</b>	95.27	<b>96.62</b>	95.95	96.62	95.95
Wine	96.07	97.19	<b>97.19</b>	95.51	97.19	94.95
SA-Heart	70.00	71.30	<b>72.61</b>	69.70	71.52	70.57
Breast Cancer	95.74	96.32	<b>96.47</b>	95.05	96.32	95.05
Bupa	70.64	72.09	72.97	66.28	73.55	66.57
Balance- Scale	94.44	93.60	<b>94.72</b>	94.24	87.50	85.90

Figure 8 shows Comparative average accuracies of two methods proposed with selection random a classifier of each clustering, with threshold and selection random a classifier of

each clustering, without threshold over six dataset by considering a decision tree as each of the base classifiers.



**Figure 8. Comparative results table 4 using basic classifiers of decision tree**

## 5. CONCLUSION AND FUTURE WORKS

In this paper a new approach is proposed to improve the efficiency of classification. The proposed approach utilizes a revised version of bagging ensemble to produce ensemble basic classifiers. Type of all basic classifiers decision tree or multi-layer nervous networks are considered and remain unchanged during constructing the ensemble. After constructing a large number of basic classifiers, KSBC partitions them with the help of k-means clustering. After that basic classifiers are partitioned by k-means clustering

algorithm then a random classifier is selected from each cluster. Weight voting method is the assembling function of the ensemble methods. In this paper the effectiveness of clusters number on the efficiency of KSBC was analyzed. The effectiveness of sampling rate is also shown that sampling rates more than 30% value not only doesn't increase classification accuracy but also decreases it, and exactly constant after sampling rates 30%. So we demonstrated that optimal ensemble size is 30% of total set created by using bagging approach. Results obtained using selected classifier

randomly from each cluster, with threshold and selection randomly classifier of each cluster without appropriate threshold, shows that first method by bagging classifier has better performance in most data sets. Proposed approach with bagging algorithm has great advantage over other ones, and in this research it can be understood that by using decision tree, effectiveness of proposed approach can be shown more obviously for rather instability of basic classifiers of decision tree compared to nervous networks, other than random selection can be used such as selecting the closest classifier to cluster centre.

## 6. REFERENCES

- [1] Amigo E., Gonzalo J., Artiles J., and Verdejo F, 2011, "Combining Evaluation Metrics via the Unanimous Improvement Ratio and its Application to Clustering Tasks", *Journal of Artificial Intelligence Research*, vol. 42, pp. 689-718.
- [2] Blake C.L., Merz C.J., 1998, "UCI Repository of machine learning databases".
- [3] Breiman L., 1996, "Bagging Predictors", *Journal of Machine Learning*, pp. 123-140.
- [4] Breiman L., 2001, "Random Forests", *Machine Learning*, pp. 5-32.
- [5] Dasgupta S., 2010, "Which Clustering Do You Want? Inducing Your Ideal Clustering with Minimal Feedback", *Journal of Artificial Intelligence Research*, vol. 39, pp. 581-632.
- [6] Freund Y., Schapire R.E., 1997, "A Decision Theoretic Generalization of Online Learning and an Application to Boosting", *Journal Computer Systems*, vol. 55, no. 1, pp. 119-139.
- [7] Giacinto G., Roli F., 2001, "An approach to the automatic design of multiple classifier systems", *Pattern Recognition Letters*, pp. 25–33.
- [8] Günter S., Bunke H., 2002, "Creation of Classifier Ensembles for Handwritten Word Recognition Using Feature Selection Algorithms", *Proceedings of the Eighth International Workshop on Frontiers in Handwriting Recognition*, pp. 183-190.
- [9] Hatami N., Ebrahimpour R., 2007, "Combining Multiple Classifiers: Diversify with Boosting and Combining by Stacking", *International Journal of Computer Science and Network Security*, vol. 7, no .1.
- [10] Khashei, M., Bijari, M., 2010, "An Artificial Neural Network Model for Time series Forecasting", *Expert Systems with Applications*, pp. 479–489.
- [11] Kuncheva L.I., 2005, "Combining Pattern Classifiers, Methods and Algorithms", Wiley.
- [12] Kuncheva, L.I. and Whitaker, C., 2003, "Measures of diversity in classifier ensembles and their relationship with ensemble accuracy", pp. 181–207.
- [13] Kurland O., Krikon E., 2011, "The Opposite of Smoothing: A Language Model Approach to Ranking Query-Specific Document Clusters", *Journal of Artificial Intelligence Research*, vol. 41, PP. 367-395.
- [14] Minaei-Bidgoli B., Topchy A.,William F., 2004, "Ensembles of Partitions via Data Resampling".
- [15] Minaei-Bidgoli B., Topchy A.P., Punch W.F., 2004, "Ensembles of Partitions via Data Resampling", *ITCC*, pp.188-192.
- [16] Mokeddem D., Belbachir H., 2008, "Distributed Data Mining using Ensemble Learning method".
- [17] Parvin H., Minaei-bidgoli B., Beigi A., 2011, "A new classifier ensembles framework", *springer*, pp. 110-119.
- [18] Parvin H., Minaei-Bidgoli B., Shahpar H., 2011, "Classifier Selection by Clustering", *3rd Mexican Conference on Pattern Recognition (MCP2011)*, LNCS, Springer, Heidelberg, pp. 60–66.
- [19] Pazos, A.B.P., Gonzalez, A.A., Pazos, F.M., 2009, "Artificial NeuroGlial Networks", *Encyclopedia of Artificial Intelligence*, pp. 167–171.
- [20] Peña J.M., "Finding Consensus Bayesian Network Structures", *Journal of Artificial Intelligence Research*, vol. 42, pp. 661-687.
- [21] Yang T., 2006, "Computational Verb Decision tree", *International Journal of Computational Cognition*, vol. 4, no. 4, pp. 34-46.