Journal of mathematics and computer Science 12 (2014) 73 - 84

Contents list available at JMCS
Journal of Mathematics and Computer Science
Journal Homepage: www.tjmcs.com

Optimal Membership Function for Creating Fuzzy Classifiers Ensemble

M. Hassanzadeh, G. Ardeshir
Faculty of electrical and computer engineering, Babol University of Technology
m.hasanzadeh@nit.ac.ir, g.ardeshir@nit.ac.ir

Article history:
Received May 2014
Accepted July 2014
Available online August 2014

Abstract

Recent researches have shown that ensembles with more diversity classifiers have more accuracy. Six methods for measuring diversity have been introduced in this paper. These methods for measuring diversity are disagreement measure, double-fault measure, Kohavi-Wolpert variance, measurement of inter-rater agreement, measure of difficulty and generalized diversity. Six methods of measuring diversity are applied to ensemble of fuzzy classifiers produced by bagging using ANFIS as the base classifier. For an ensemble of fuzzy classifiers, relationship between membership functions and diversity has been studied. Experimental results show that using triangular membership function lead to more diverse classifiers and ensemble with more accuracy.

Keywords: Accuracy, Diversity measurement, Ensemble of Classifiers, Fuzzy Classifiers

1. Introduction

In recent year, ensemble of classifiers has been known as a method for improving the accuracy of classification. An ensemble (committee) of classifiers is a set of classifiers whose individual decisions are combined in some way (typically by voting) to classify new examples. In literature, the ensemble of classifiers is referred by different names: committees of learners, mixtures of experts, classifier ensembles, multiple classifier systems, consensus theory, etc. [1]. Hansen and Salamon in 1990 showed that an ensemble of classifiers could be more efficient than a single one if each classifier of the ensemble is different from the others in terms of the classification error [3]. This means that one of the main problems in combining classifiers is “creating diverse classifiers”. In the modeling of classifier combination, many researchers believe that the success of classifier ensembles not only depends on a set of appropriate classifiers, but also on the diversity being inherent in the classifiers. A good diversity measure would have the ability to find the extent of diversity among classifiers and estimate the improvement or deterioration in accuracy of individual classifiers when they have been combined, but there is no strict definition for measuring diversity. So, we use different definitions of diversity in our study proposed by researcher for fuzzy classifiers. In [1] ten methods and in [4] six methods have been proposed for measuring diversity and some of these measures are correlated with each other's. In this paper, a relationship between diversity and accuracy of an ensemble of fuzzy classifiers will be shown.

The rest of the paper organized as follows: section 2 describes bagging algorithm for creating ensemble of classifiers. In section 3, different methods for measuring diversity have been introduced. Anfis, which has been used as base classifier in this work, will be explained in section 4. Results of experiments will be described in section 5

2. Methods of classifiers ensemble

Using different training sets for each classifier is a method to create an ensemble. In this method, the data available for training is divided into different subsets and each set of data is used as a building block for each classifier. Two important methods of creating these subsets are Bagging and Boosting [5-8]. Since in this work we have used Bagging, it is explained in this section.

- Bagging

In this method N elements are selected randomly by replacement from an N-element set known as training set. The selected elements form a new set. It should be noted that all the elements are selected with the same probability and depending to the “replacement”, there is a possibility of one element to be selected multiple times.

The N sets of data are used to build the same number of the classifiers constructing the ensemble. After the training process, each new input data is applied to the trained classifiers and is assigned to a one of the classes by using the majority vote scheme. Figure 1 illustrates the Bagging algorithm.

In 1996, Breiman showed that the Bagging algorithm is more efficient in unstable classifiers in which a small change in the training data results in a considerable variation in classifier [5]. Neural network and decision tree are two examples of unstable classifiers.

BAGGING

Training phase

1. Initialize the parameters
   - $\mathcal{D} = \emptyset$, the ensemble.
   - L, the number of classifiers to train.
2. For $k = 1, \ldots, L$
   - Take a bootstrap sample S_k from $\mathbf{Z}$.
   - Build a classifier D_k using S_k as the training set.
   - Add the classifier to the current ensemble, $\mathcal{D} = \mathcal{D} \cup D_k$.
3. Return $\mathcal{D}$.

Classification phase

4. Run $D_1, \ldots, D_L$ on the input $\mathbf{x}$.
5. The class with the maximum number of votes is chosen as the label for $\mathbf{x}$.

Figure 1. Bagging algorithm [1]

3. Diversity of the ensemble

In an ensemble, the combination of the outputs of several classifiers is only useful if they disagree on some inputs [3]. We refer to the measure of disagreement as the diversity of the ensemble.

Let a labeled training set be $Tr = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots (\mathbf{x}_N, y_N)\}$, where y_i is the class label of $\mathbf{x}_i$. The base classifiers $H = \{h_1, h_2, \ldots, h_L\}$ of an ensemble are trained on the training set, and the output of a base classifier h_i on sample $\mathbf{x}_i$ is $h_i(\mathbf{x}_i)$. Since there is no strict definition for diversity, we introduce six diversity measures here.

3.1. The disagreement measure

This measure is defined based on the intuition that two diverse classifiers perform differently on the same training data. Let $\mathbf{O}$,oracle output matrix, be an $N \times L$ matrix with 1 or -1 elements. The element O_{ij} of this matrix is 1 if training sample $\mathbf{x}_i$ is classified correctly by base classifier h_i and is -1 otherwise. The disagreement measure between the two base classifiers, h_j and h_k, is measured by:

$$dis_{j,k} = \frac{n(1,-1) + n(-1,1)}{n(1,1) + n(1,-1) + n(-1,1) + n(-1,-1)} \quad (1)$$

where $n(a,b)$ is the number of training samples on which the oracle output of h_j and h_k is a and b respectively. Diversity within the whole set of base classifiers is then calculated by averaging over all pairs of base classifiers:

$$dis = \frac{2}{L(L-1)} \sum_{j=1}^{L} \sum_{k=j+1}^{L} dis_{j,k} \quad (2)$$

and we can rewrite the equation (2) as:

$$dis = \frac{2}{NL(L-1)} \sum_{j=1}^{L} \sum_{k=j+1}^{L} \left(n_{j,k}(1,-1) + n_{j,k}(-1,1) \right) \quad (3)$$

The diversity increases with the value of the disagreement measure.

3.2. The double-fault measure

This measure was introduced by Giacinto and Roli [12]. The double-fault measure of two base classifiers is calculated by:

$$DF_{j,k} = \frac{n(-1,-1)}{n(1,1) + n(-1,1) + n(1,-1) + n(-1,-1)} \quad (4)$$

Same as the disagreement measure, the diversity within the whole set of base classifiers is calculated as follows:

$$DF = \frac{2}{NL(L-1)} \sum_{j=1}^{L} \sum_{k=j+1}^{L} n_{j,k}(-1,-1) \quad (5)$$

The diversity decreases when the value of the double-fault measure increases.

3.3 Kohavi-Wolpert variance

The Kohavi-Wolpert variance was proposed by Kohavi and Wolpert [11] in their decomposition formula for the classification error of a classifier. This measure originated from the bias-variance decomposition of the error of a classifier.

The original expression of the variability of the predicted class label y for a sample x is:

$$Variance_{\mathbf{x}} = \frac{1}{2}\left(1 - \sum_{i=1}^{C} p(y = \omega_i | \mathbf{x})^2\right) \quad (6)$$

According to oracle outputs $C = 2$ and $p(y=1|\mathbf{x}) + p(y=-1|\mathbf{x}) = 1$ and this equation can be written as:

$$\begin{aligned} Variance_{\mathbf{x}} &= \frac{1}{2}\left(1 - p(y=1|\mathbf{x})^2 - p(y=-1|\mathbf{x})^2\right) \\ &= p(y=1|\mathbf{x})\, p(y=-1|\mathbf{x}) \\ &= p(O=1|\mathbf{x})\, p(O=-1|\mathbf{x}) \end{aligned} \quad (7)$$

Kuncheva and Whitaker [1] presented a modified version of Equation (7) to measure the diversity of an ensemble as:

$$KW = \frac{1}{NL^2} \sum_{i=1}^{N} l_i (L - l_i) \quad (8)$$

where l_i is the number of base classifiers that classify the training sample $\mathbf{x}_i$ incorrectly. The diversity increases with values increasing of the KW variance.

3.4 Measurement of inter-rater agreement

This measure is developed as a measure of inter-rater (inter-classifier) reliability by Fleiss in 1981[9], called k. It can be used to measure the level of agreement within a set of classifiers, hence it is also based on the assumption that a set of classifiers should disagree with one another, for being diverse. The diversity decreases when the value of k increases. The k is calculated by:

$$k = 1 - \frac{\sum_{i=1}^{N} (L - l_i) l_i}{NL(L-1)P(1-P)} \quad (9)$$

where P is the average classification accuracy of the base classifiers on the training data and can be calculated easily from oracle matrix.

3.5 The measure of "difficulty"

This measure was introduced by Hansen and Salamon [3]. Defining a discrete random variable V, $V_i = (L - l_i)/L$ for a sample $\mathbf{x}_i$ which is randomly drawn from the training set, the measure of difficulty was defined as the variance of V over the whole training set as:

$$diff = \text{var}(V_i) \quad (10)$$

The diversity increases with decreasing values of the measure of difficulty.

3.6. Generalized diversity

This measure was proposed by Partidge and Krzanowski [10]. The heuristic behind this measure is similar to that of the Double-Fault measure. Given two classifiers, Partidge and Krzanowski argued that maximum diversity is achieved when failure of one classifier is accompanied by correct classification of data by the other classifier and minimum diversity occurs when two classifiers fail together. Therefore, for a sample $\mathbf{x}_i$ that is randomly drawn from the training set, the generalized diversity is defined as:

$$GD = 1 - \frac{\sum_{j=1}^{N} \frac{j(j-1)}{L(L-1)} T_j}{\sum_{j=1}^{N} \frac{j}{L} T_j} \quad (11)$$

where T_j is the probability that $l_i = j$.

Six methods for measuring diversity are listed in TABLE 1. For more details on measuring diversity see [1] and [4].

TABLE 1. Different methods for measuring diversity.

Name of measure	formula
The disagreement measure	$dis = \frac{2}{L(L-1)} \sum_{j=1}^{L} \sum_{k=j+1}^{L} dis_{j,k}$
double-fault measure	$DF = \frac{2}{L(L-1)} \sum_{j=1}^{L} \sum_{k=j+1}^{L} n_{j,k}(-1,-1)$
Kohavi-Wolpert variance	$KW = \frac{1}{NL^2} \sum_{i=1}^{N} l_i (L - l_i)$
Measurement of inter-rater agreement	$k = 1 - \frac{\sum_{i=1}^{N} (L - l_i) l_l}{NL(L-1)P(1-P)}$
Generalized diversity	$GD = 1 - \frac{\sum_{j=1}^{N} \frac{j(j-1)}{L(L-1)} T_j}{\sum_{j=1}^{N} \frac{j}{L} T_j}$
The measure of "difficulty"	$V_i = \frac{L - l_i}{L}$, $diff = \text{var}(V_i)$

4. ANFIS as the base classifier

ANFIS (Adaptive-Network-based Fuzzy Inference System), which has been used as the base classifiers, is a fuzzy inference system which is implemented under the framework of adaptive networks [13]. An adaptive network can be considered as a superset of feed-forward neural networks with supervised learning. ANFIS is a type of neurofuzzy network which has the fuzzy rules embedded within the neural network. Figure

2 shows the structure of an adaptive network. Node functions are represented by squares if they have parameters to be set, which make them adaptive, and by circles if they do not have parameters [13]. The links have no associated weights and they only represent direction flow.

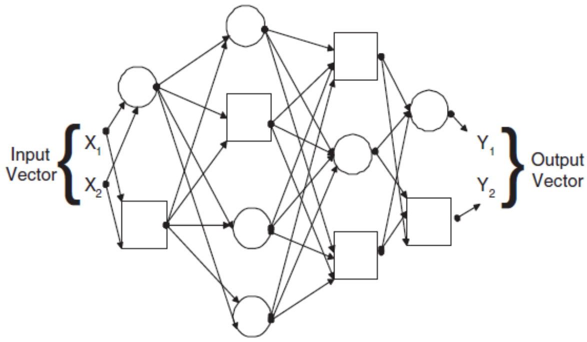


Figure 2. Structure of ANFIS [14]

5. Experiment setup and results

First, we evaluate some experiments to choose best membership function in fuzzy classifiers ensemble to increase diversity. Diversity of fuzzy classifiers combined via bagging, has been measured for several datasets. Anfis is used as the base classifier and six diversity measures were calculated for seven different membership functions. Figure 3 and TABLE 3 show the results for iris dataset and the results for glass dataset are shown in Figure 4 and TABLE 4. Figure 5 and TABLE 5 show the results for an artificial dataset (1000 tow dimensional samples with 4 classes). The results for another artificial dataset (1000 three dimensional samples with 7 classes) are shown in Figure 6 and TABLE 6. Eighty percent of samples were used for training data and 20% for test. The bold numbers in tables 3 to 5 show the best diversity in each column.

The numbers 1,2,…,7 on x-axis are representative of several membership functions as follows:

1: Generalized Bell membership function

2: Pi-shaped membership function

3: Triangular membership function

4: Diff. sigmoidal membership function

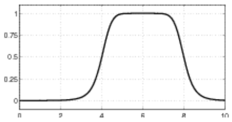
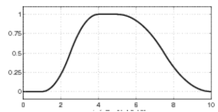
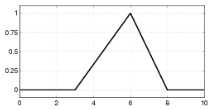
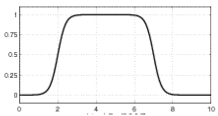
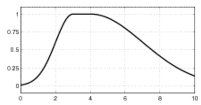
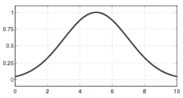
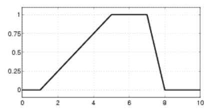
5: Gaussian2 membership function

6: Gaussian membership function

7: Trapezoidal membership function

Mathematical formula and shape of membership functions are mentioned in TABLE 2.

TABLE 2. Mathematical formula and shape of membership functions

	Membership function	Mathematical formula	Shape
1	Generalized Bell	$f(x;a,b,c)=\dfrac{1}{1+\left\|\dfrac{x-c}{a}\right\|^{2b}}$	
2	Pi-shaped		
3	Triangular	$f(x;a,b,c)=\begin{cases}0 & x\le a\\ \dfrac{x-a}{b-a} & a\le x\le b\\ \dfrac{c-x}{c-b} & b\le x\le c\\ 0 & x\ge c\end{cases}$	
4	Diff. sigmoidal	$f(x;a,c)=\dfrac{1}{1+e^{-a(x-c)}}$	
5	Gaussian2		
6	Gaussian	$f(x;\sigma,c)=e^{\frac{-(x-c)^2}{2\sigma^2}}$	
7	Trapezoidal	$f(x;a,b,c)=\begin{cases}0 & x\le a\\ \dfrac{x-a}{b-a} & a\le x\le b\\ 1 & b\le x\le c\\ \dfrac{d-x}{d-c} & c\le x\le d\\ 0 & x\ge d\end{cases}$	

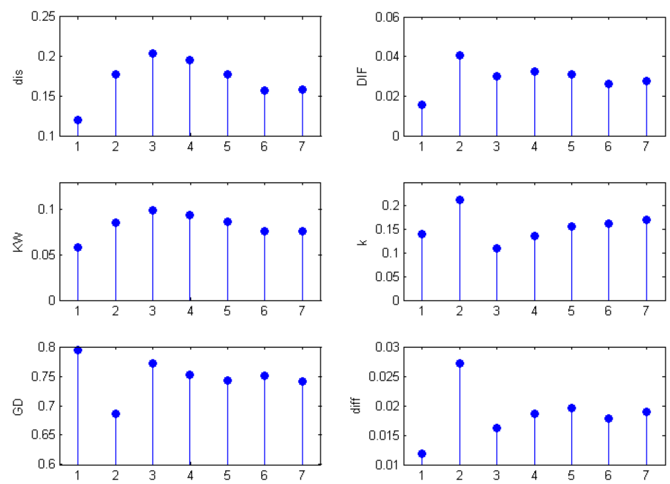


Figure 3. Diversity measurements for iris dataset

TABLE 3. Diversity measurements for iris dataset, rows indicate different membership functions and columns show different method for measuring diversity

	Dis	DF	KW	k	GD	diff
Gbell	0.1196	**0.0155**	0.0578	0.141	0.7944	**0.0119**
Pi	0.1771	0.0406	0.0856	0.2126	0.6857	0.0271
Tri	**0.2041**	0.0302	**0.0986**	**0.1108**	**0.7717**	0.0162
Dsig	0.1951	0.0322	0.0943	0.1359	0.752	0.0187
gauss2	0.1779	0.0308	0.086	0.1559	0.743	0.0196
Gauss	0.1567	0.0261	0.0757	0.1623	0.7502	0.0179
Trap	0.1578	0.0275	0.0763	0.1701	0.7416	0.019

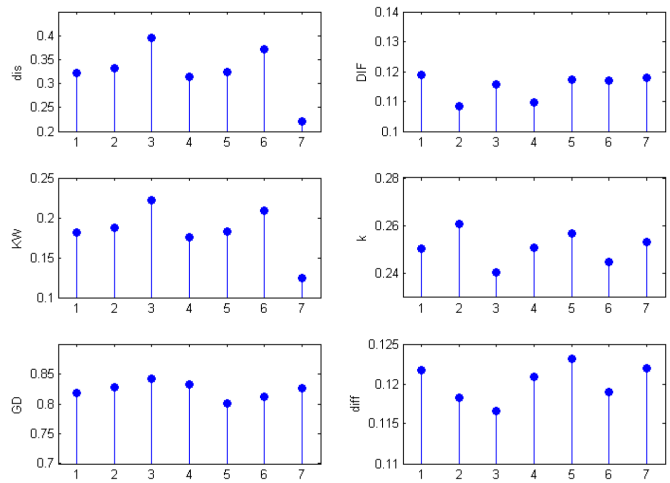


Figure 4. Diversity measurements for glass dataset

TABLE 4. Diversity measurements for glass dataset, rows indicate different membership functions and columns show different method for measuring diversity

	Dis	DF	KW	k	GD	diff
Gbell	0.3223	0.1190	0.1815	0.2503	0.8181	0.1217
Pi	0.3327	**0.1084**	0.1873	0.2609	0.827	0.1183
Tri	**0.3951**	0.1157	**0.2224**	**0.2405**	**0.8424**	**0.1166**
Dsig	0.3136	0.1098	0.1760	0.2509	0.8321	0.1209
gauss2	0.3238	0.1173	0.1823	0.2569	0.8001	0.1232
Gauss	0.3710	0.1171	0.2089	0.2449	0.8119	0.1190
Trap	0.2210	0.1179	0.1244	0.2531	0.8265	0.1220

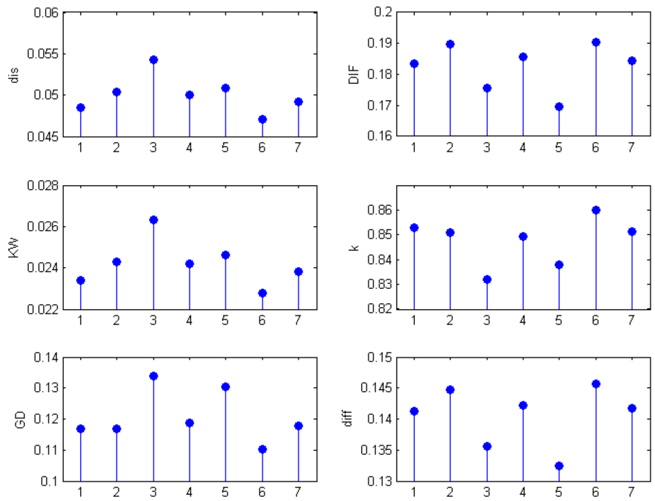


Figure 5. Plots of diversity measures for artificial dataset- 1000 tow dimensional samples with 4 classes

TABLE 5. Diversity measures for artificial dataset- 1000 tow dimensional samples with 4 classes

	Dis	DF	KW	k	GD	diff
Gbell	0.0485	0.1834	0.0234	0.8527	0.1167	0.1412
Pi	0.0504	0.1899	0.0243	0.8509	0.117	0.1447
Tri	**0.0543**	0.1756	**0.0263**	**0.832**	**0.134**	**0.1356**
Dsig	0.05	0.1856	0.0242	0.8495	0.1188	0.1422
gauss2	0.0509	**0.1695**	0.0246	0.838	0.1304	0.1325
Gauss	0.0471	0.1905	0.0228	0.86	0.1101	0.1457
Trap	0.0492	0.1844	0.0238	0.8513	0.1177	0.1417

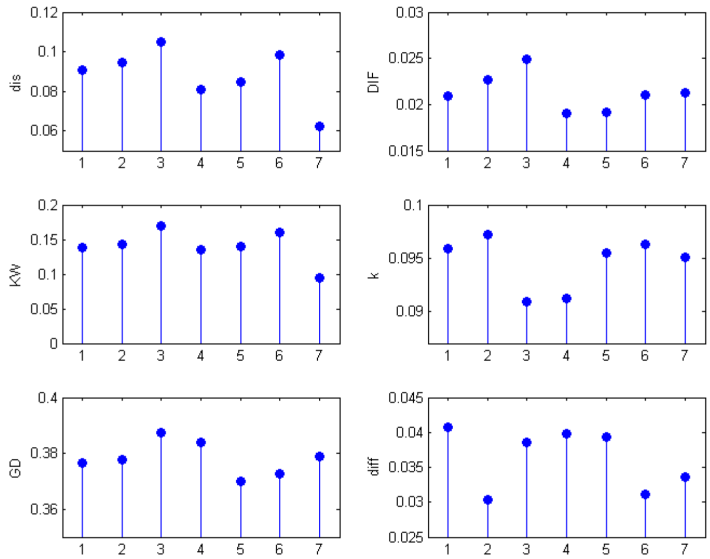


Figure 6. Diversity measurements for artificial dataset-1000 three dimensional samples with 7 classes

TABLE 6. Diversity measurements for artificial dataset-1000 three dimensional samples with 7 classes

	dis	DF	KW	k	GD	diff
gbell	0.0909	0.0209	0.1389	0.0959	0.3766	0.0408
pi	0.0943	0.0227	0.1434	0.0972	0.3778	**0.0304**
tri	**0.1047**	0.0249	**0.1703**	**0.0909**	**0.3873**	0.0385
dsig	0.0808	**0.0191**	0.1352	0.0912	0.3840	0.0398
gauss2	0.0848	0.0192	0.1396	0.0955	0.3700	0.0393
gauss	0.0983	0.0210	0.1599	0.0963	0.3726	0.0312
trap	0.0621	0.0213	0.0953	0.0951	0.3789	0.0336

Among 6 different methods of measuring diversity applied to 4 datasets, (24 cases in total shown in tables 3 to 5), the triangular membership function lead to better results in 16 cases. In other 8 cases, generalized bell membership function and gaussian2 membership function were better respectively. According to above results, it seems that the triangular membership function is the best choice and can be used in a fuzzy system.

6. Conclusion

In this paper, six measures have been studied for evaluation of diversity measurement in ensemble of classifiers. Results for different data sets show that in ensemble of fuzzy classifiers, using triangular membership function leads to more accurate results than other membership functions. Moreover, increasing diversity leads to more accurate classifying. Hence, in ensemble of fuzzy classifiers, diversity of base classifiers using in ensemble should be noticed and considered in design

7. References

[1] Kuncheva L.I., C.J. Whitaker. Measures of diversity in classifier ensembles, Machine Learning, 1 (2003), pp. 181-207

[2] Kuncheva.L. Combining Pattern Classifiers. Methods and Algorithms , John Wiley & Sons, 2004.

[3] Hansen, L. Salamon, P. Neural network ensembles, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12, (1990), pp. 993-1001.

[4] E. K. Tang · P. N. Suganthan · X. Yao, An analysis of diversity measures, Machine Learning vol. 65, (2006), pp. 247–271.

[5] L. Breiman, Bagging predictors. Machine Learning 24 (2), (1999), pp. 123_140.

[6] L. Breiman, Arcing classifiers. The Annals of Statistics, 26(3), (1998), pp. 801–849.

[7] D. Optiz and R. Macline, " Popular ensemble methods: an empirical. study," Journal of Artificial Intelligence Research, vol. 11,(1999).

[8] Lior Rokach, Taxonomy for characterizing ensemble methods in classification tasks: A review and annotated bibliography, Computational statistics and data analysis, vol. 53, (2009), pp. 4046-4072.

[9] Fleiss, J. Statistical methods for rates and proportions. John Wiley & Sons, 1981.

[10] Patridge, D., & Krzanowski, W. J. ,Software diversity: Practical statistics for its measurement and exploitation. Information & Software Technology, 39, (1997), pp. 707–717.

[11] Kohavi, R., & Wolpert, D. , Bias plus variance decomposition for zero-one loss functions, Proc. 13th Int. Conference on Machine Learning, (1996), pp. 275–283.

[12] Giacinto, G.,& Roli, F. , Design of effective neural network ensembles for image classification processes. Image Vision and Computing, 19(9/10), (2001), pp. 699–707.

[13] Jang J.S.R.,“ANFIS: Adaptive-Network-Based Fuzzy Inference System”,, IEEE Trans. Systems, Man, Cybernetics, 23(5/6),(1993), pp. 665-685.

[14] Juana Canul-Reich, Larry Shoemaker and Lawrence O. Hall, Ensembles of Fuzzy Classifiers, IEEE International Conference on Fuzzy Systems, (2007).