P.GORGEL et al. / IU-JEEE Vol. 13(1), (2013), 1605-1611

CLASSIFICATION OF BREAST MASSES USING ANFIS-BASED FUZZY ALGORITHMS: A COMPARATIVE STUDY

Pelin GÖRGEL[1], Ahmet SERTBAS[1], Ainura TURUSBEKOVA[2]

[1]Department of Computer Engineering, Istanbul University (IU), Istanbul, Turkey
[2] Istanbul University (IU), Istanbul, Turkey

Abstract : This study aims to produce a diagnosis system for breast masses related to breast cancer. The dataset consisting of 60 digital mammograms is acquired from Istanbul University Faculty of Medicine Hospital. 78 masses in the mammograms are extracted manually for this study by the experts. It is a fuzzy based comperative study of malignant-benign classification for breast masses which has the accuracy of 74.36% with k-means and 93.75% with ANFIS based fuzzy c-means and subtractive clustering.

Keywords- Breast mass classification, fuzzy c-means, fuzzy subtractive algorithm, ANFIS.

1. INTRODUCTION

Breast cancer is one of the major causes of death among women and one of the most life-threatening diseases especially in developed countries nowadays. In many countries the death rate is exceedingly growing because of the late detection. Early detection of breast cancer provides patients the chance to recover. The most widely used method in the diagnosis of breast cancer is digital mammography. There can be malignant and benign masses in breast. It is very important to detect them accurately in the suspected breasts. Expert radiologists are able to detect breast cancer by looking at the images taken by device. But sometimes, when the breast masses are small or the breast tissue is thick the cancer can be misdiagnosed, or may not be seen. To help the experts in diagnosis some computer aided studies are implemented so far. In-Sung J.et al., [1] developed a system for the classification of mammographic masses as malignant or benign by adaptive k-means and ANFIS LVQ method. They achieved a classification accuracy of 86.6 %, and raised it by ANFIS LVQ method to %87.6. In their study they used backpropagation unsupervised learning method in ANFIS. Görgel P. [2] developed a system to diagnose the breast cancer. In this study Spherical Wavelet Transform (SWT) was used to obtain the features of the masses Support Vector Machines (SVM) for the diagnosis. According to the mass-tissue classification she achieved 96% accuracy rate and the number of the false positives per image was 0.05. The highest sensitivity was 88% and specificity was 98%.

This paper is organized as follows. In Section 2 clustering methods such as k-means, fuzzy c-means algorithm, and subtractive algorithm are explained. In Section 3 the data set consisting of mammogram masses is explained and the experimental results are presented and discussed. Finally Section 4 draws the conclusion and gives some final remarks.

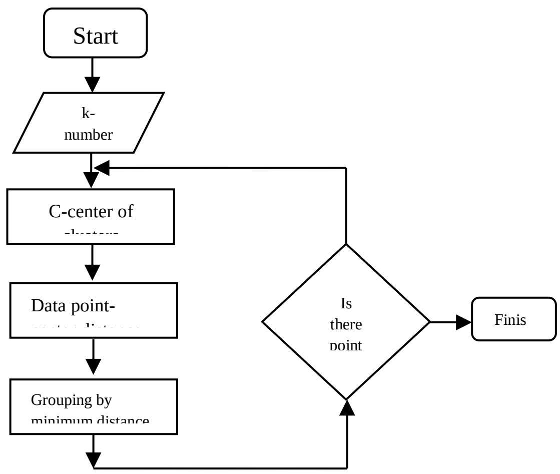
2. MATERIALS AND METHODS

2.1. K-MEANS CLUSTERING METHOD

The k-means algorithm partitions data into k sets. The solution is then a set of k centers, each of which is located at the center of the data for which it is the closest center. For membership function, each data point belongs to its nearest center, forming a Voronoi partition of the data. The objective function that the k-means algorithm optimizes is

$$E(C) = \sum_{j=1}^{k} \sum_{x_i \in C_j} \| x_i - c_j \|^2 \qquad (1)$$

In the above equation x_i is data points, c_j is the center of cluster, k is the number of clusters. Fig. 1 demonstrates k-means algorithm's running scheme.



Received on: 04.09.2012
Accepted on: 26.03.2013

Fig. 1: The flow chart of k-means clustering algorithm

2.2. FUZZY C-MEANS CLUSTERING METHOD

Fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters. This method was developed by Dunn [3] in 1973 and improved by Bezdek [4] in 1981 and frequently used in pattern recognition. It is based on minimization of the following objective function:

$$J_m = \sum_{j=1}^{c}\sum_{i=1}^{n}\left[u_{ij}\right)]^{m}\left\|x_i - c_j\right\|^2, \quad 1\leq m < \infty \qquad (2)$$

where m is any real number greater than 1 U_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-dimensional measured data, c_j is the d-dimension center of the cluster, and ||*|| is any norm expressing the similarity between any measured data and the center. Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership U_{ij} and the cluster centers c_j by: Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{m-1}}} \qquad (3)$$

$$c_j = \frac{\sum_i [\mu_j(x_i)]^m x_i}{\sum_i [\mu_j(x_i)]^m} \qquad (4)$$

This iteration will stop when $max_{ij}\left\{\left|\mu_j(x_i)^{(k+1)} - \mu_j(x_i)^{(k)}\right|\right\} < \varepsilon$, where ε is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m. The algorithm is composed of the following steps [4]:

1. *Initialize $U=[u_{ij}]$ matrix, $U^{(0)}$*
2. *At k-step: calculate the centers vectors $C^{(k)}=[c_j]$ with $U^{(k)}$*

$$c_j = \frac{\sum_i [\mu_j(x_i)]^m x_i}{\sum_i [\mu_j(x_i)]^m}$$

3. *Update $U^{(k)}$, $U^{(k+1)}$*

$$\mu_j() = \frac{1}{\sum_{k=1}^{C}\left(\frac{\left\|x_i - c_j\right\|}{\left\|x_i - c_k\right\|}\right)^{\frac{2}{m-1}}}$$

4. *If $\|U^{(k+1)} - U^{(k)}\| < \varepsilon$ then STOP; otherwise return to step 2.*

Fuzzy c-means algorithm has a flexible structure. The biggest advantage of the fuzzy c-means algorithm is the ability to find clusters of overlapping.

2.3. SUBTRACTIVE CLUSTERING METHOD

Subtractive clustering method was introduced by Hongxing, L in 2001 [5]. For this method, data points have to be rescaled to [0,1] in each dimension. Each data point $z_j = (x_j,\ y_j)$ is assigned a potential P_j, according to its location to all other data points:

$$P_i^* = \sum_{j=1}^{n} e^{-a\left\|x^i - x^j\right\|^2} \qquad (5)$$

where

$$\alpha = \frac{\gamma}{r_a}$$

P_i^* is the potential-value i-th data as a cluster centre.
α is the weight between i-data to j-data
x is the data point
γ is the variables (commonly set 4)
r_a is a positive constant called cluster radius.

The potential of a data point to be a cluster center is higher when more data points are closer. The data point with the highest potential, denoted by $\mathrm{P_i^*}$ is considered as the first cluster center $\mathrm{c_1 = (d_1, e_1)}$. The potential is then recalculated for all other points excluding the influence of the first cluster center according to:

$$\mathrm{P_i^* = P_i^* - P_k^*}\ \varsigma \qquad (6)$$

Where

$$\varsigma = \mathrm{e}^{-\beta\left\|\mathrm{x^i - c^k}\right\|^2} \qquad (7)$$

$$\beta = \frac{4}{\mathrm{r_b^2}} \qquad (8)$$

$$\mathrm{r_b = r_a * \eta}$$

$\mathrm{P_i^*}$-i is the new potential-value i-data.
$\mathrm{P_k^*}$- is the potential-value data as cluster centre
c is the cluster center of data
β is the weight of i-data to cluster centre
$\mathrm{r_i}$ is the distance between cluster centre
η is the quash factor

Again, the data point with the highest potential $\mathrm{P_k^*}$ is considered to be the next cluster center $\mathrm{c_k}$, if

$$\frac{\mathrm{d_{min}}}{\mathrm{r_a}} + \frac{\mathrm{P_k^*}}{\mathrm{P_1^*}} \geq 1$$

with d is the minimal distance between $\mathrm{c_1}$ and all previously found cluster centers, the data point is still accepted as the next cluster center $\mathrm{c_1}$. Further iterations can then be performed to obtain new cluster centers c_2. If a possible cluster center does not fulfill the above described conditions, it is rejected as a cluster center and its potential is set to 0. The data point with the next highest potential P_k^* is selected as the new possible cluster center and re-tested. The clustering ends if the following condition is fulfilled:

$$P_k^* < \varepsilon P_i^*$$

where : ε is the reject ratio.

Indicative parameters values for r_a, η, ε and ε* have been suggested by [4]. Each cluster center is considered as a fuzzy rule that describes the system behavior of the distance to the defined cluster centers:

$$\mu_j^{ik} = e^{-\alpha\left\|x_j^i - c_j^k\right\|^2} \quad (9)$$

Eq. 9 is a common form of subtractive clustering, hence it needs to create an algorithm to process data clustering. Thus, this paper proposed an algorithm to cluster the data to train a traffic control system:

Step 1: Calculate the input data to be clusters.

X_{ij}, i=1, 2, …, n; j=1, 2, …, m.

with: n is the number of data and m is the type of data.

Step 2: Set the variables values; where η *is* quash factor, ε* is accept ratio and ε is reject ratio.

Step 3: Set the normal data values with the following model:

$X_{ij}^{norm} = \frac{X_{ij} - X_{j-min}}{X_{j-min} - X_{j-min}}$, i=1, 2, …, n; j=1, 2,…, m.

Step 4: Set the potential of each data point by the formula:

a = 1, revise to a = n

if m = 1, set

$$P_i' = \sum_{i=1}^{n} e^{-4\left\|\frac{x_i - x_k}{ra}\right\|^2}, \; i = 1, 2, \dots, n; k = 1, 2, \dots, n; i \neq k$$

if m > 1, set

$$P_i' = \sum_{i=1}^{n} e^{-4\left\|\sum_{j=1}^{m} \frac{x_i^j - x_k^j}{ra}\right\|^2}, i = 1, 2, ., n; j = 1, 2, ., n; i \neq j$$

Step 5: Set the highest potential value of data:

M= max$[P_i' \mid i = 1, 2, \dots\dots n]$

h=i, so that $D_i = M$

For $V_j = X_{hj}; j = 1, 2, \dots\dots m$, *C=0* (number of clusters) *Cnd=1, z=m*

Do Cnd ≠ 0 and Z ≠ 0

Step 6: Put the real data:

$Cnt_{ij} = Cnt_{ij}(X_{j-max} - X_{j-\min}) + X_{j-min}$

Step 8: Set the cluster sigma:

$$\sigma_j = \frac{r_j^*(X_{j-max} - X_{j-min})}{\sqrt{8}}$$

2.4. ADAPTIVE NEURO-FUZZY INFERENCE SYSTEMS (ANFIS) ARCHITECTURE

ANFIS defined by J.-S. Roger Jang in 1992 is a class of adaptive networks that are functionally equivalent to fuzzy inference systems. ANFIS represent Sugeno Tsukamoto fuzzy models [6] and uses a hybrid learning algorithm. The algorithm creates a fuzzy decision tree to classify the data into one of 2n linear regression models to minimize the sum of squared errors. The fuzzy inference system (FIS) has two inputs such as x and y and one output z. A first-order Sugeno fuzzy model has some rules as follows (Fig. 2):

Rule1:

If x is A1 and y is B1, then f1 = p1x + q1y + r1

Rule2:

If x is A2 and y is B2, then f2 = p2x + q2y + r2

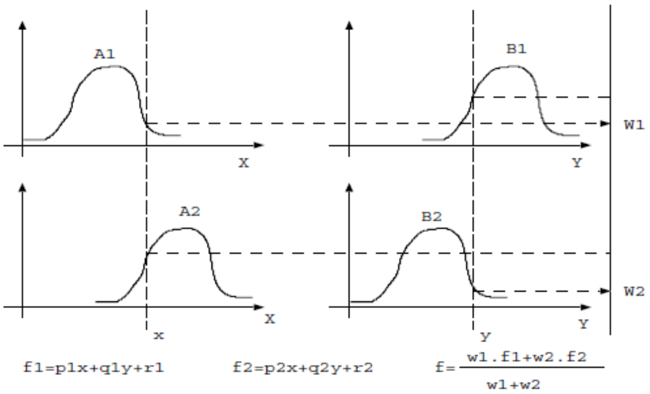


Fig. 2: Sugeno type FIS

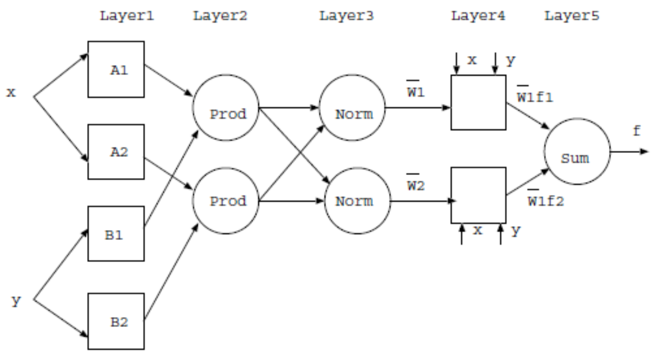


Fig. 3: ANFIS architecture

Fig. 3 demonstrates the ANFIS architecture with its layers. $O_{l,i}$ is the output of the ith node of the layer l. Every node i in this layer is an adaptive node with a node function

$O_{1,i}$= μA_i (x) for i = 1, 2, or

$O_{1,i}$ = μB_{i-2} (x) for i = 3, 4

x (or y) is the input node i and A_i (or B_{i-2}) is a label associated with this node and $O_{1,i}$ is the membership grade of a fuzzy set (A_1, A_2, B_1, B_2,). Typical membership function is written in Eq.10.

$$\mu A_i \text{ (x)} = \frac{1}{1+\left|\frac{x_i-c_i}{a_i}i\right|^{2b_i}} \text{ ; } \forall r \tag{10}$$

where a_i, b_i, c_i are the parameter set. The output is the product of all the incoming signals.

$O_{2,i} = w_i$, $= \mu A_i$ (x μB_i (y), i = 1, 2

The i-th node calculates the ratio of the i*th* rulet's firing strenght to the sum of all rulet's firing strengths.

$O_{3,i} = \overline{w_i} = \frac{w_i}{w_1+w_2}$, i = 1, 2

Outputs are called normalized firing strengths. Every node i in this layer is an adaptive node with a node function:

$O_{4,i} = \overline{w_i}\, f_i = \overline{w_i}\, (p_x + q_i y + r_i)$

Where $\overline{w_i}$ is the normalized firing strenght from layer 3 and $\{p_i, q_i, r_i\}$ is the parameter set of this node. The single node in this layer is a fixed node labeled sum, which computes the overall output as the summation of all incoming signals:

Overall output $= O_{5,i} = \sum_i \overline{w_i}\, f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$

The ANFIS can be trained by a hybrid learning algorithm presented by Jang [7]. In the forward pass the algorithm uses least-squares method to identify the consequent parameters on the layer 4. The errors are propagated to the backward and the premise parameters are updated by gradient descent.

3. THE DATASET AND EXPERIMENTAL RESULTS

In this paper we perform a classification of breast masses based on fuzzy clustering algorithms. There can be malignant or benign masses in breast tissue. Generally, breast masses related to breast cancer are malignant. In this study breast masses were taken from mammographic images of the patients in Radiology Department of Istanbul University Medical Faculty Hospital. Mammographic images were scanned by digital mammogram device. In the data set there are 78 masses which are acquired from 60 different mammographic images. In this study these masses are distinguished from each other as malignant or benign. For classification, clustering algorithms such as k-means algorithm, fuzzy c-means algorithm and subtractive algorithm are used. During classification procedures fuzzy c-means and subtractive algorithms were applied on the basis of ANFIS.

Table 1 demonstrates the performance metrics of the k-means. K-means algorithm is used with different distance functions to find the most suitable distance function for this problem. As declared in Table 1, the highest sensitivity is reached as 97.14%, with Euclidean distance function.

Table 1: Breast mass classification accuracy based on k-means algorithm

Distance function	**Sensitivity (%)**	**Specificity (%)**	**Positive predictive value (%)**	**Negative predictive value (%)**	**Benign-Malign classification accuracy (%)**
City	20	**86,05**	53,85	56,92	56,41
Cityblock	57,14	72,09	45,71	72,09	60,26
Euclidean	**97,14**	11,63	47,22	83,33	50
Cosine	94,29	58,14	**64,71**	**92,59**	**74,36**
Correlation	71,43	62,79	60,98	72,97	66,66

The highest specificity 86.05%, is provided with City distance function. On the other hand the highest positive (64.71%) and negative predictive values (92.59%) are obtained with Cosine distance function. In Table 2 the confusion matrix of breast mass classification based on k-means algorithm is shown.

Table 2: Confusion matrix using k-means algorithm

Distance function	**True positive (TP)**	**False positive(FP)**	**True negative (TN)**	**False negative(FN)**	**True classified masses number**	**ROC Area**
City	7	6	**37**	**28**	44	0,589
Cityblock	16	19	31	12	47	0,530
Euclidean	**34**	**38**	5	1	39	0,544
Cosine	**33**	18	25	2	**58**	0,671
Correlation	25	16	27	10	52	0,671

The highest ROC area (0.671) is obtained both with correlation and cosine distance functions.

For fuzzy c-means algorithm with ANFIS, the data set is separated into train and check data. K-fold cross validation is used in the implementation of both fuzzy c-means and subtractive algorithm. The original sample is randomly partitioned into k subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing data, and the remaining (k − 1) subsamples are used as training data. The cross-validation process is then repeated k times (the folds), with each of the k subsamples used exactly once as the validation data. The *k* results from the folds then can be averaged (or otherwise combined) to produce a single estimation [8]. As seen in Table 3, the highest accuracy is 93.75% when *k* is 5. And when k is chosen according to the highest classification accuracy, the sensitivity is 100% and specificity is 90.91%. The more comprehensive results which change with the number of iterations are listed in Table 4, when *k* is 4 and 5 separately. The performance metrics in Table 3-4 all belong to fuzzy c-means algorithm.

Table 3: Performance metrics of fuzzy c-means algorithm changing with *k* value

Training set	**Checking set**	**Optimum Number of iterations**	**Sensitivity (%)**	**Specificity (%)**	**Benign-Malign classification accuracy (%)**
k=2 39	39	100	%23,53	%59,09	%46,15
k=3 52	26	100	%80	**%93,75**	%88,46
k=4 58	20	80	%85,71	%92,71	**%90**
k=5 62	16	80	**%100**	%90,91	**%93,75**

Table 4: Performance metrics of fuzzy c-means algorithm changing with the number of iterations

Train set	**Check set**	**Number of iterations**	**Sensitivity (%)**	**Specificity (%)**	**Benign-Malign classification accuracy (%)**
k=4 58	20	20	42,85	53,86	50
		40	85,71	69,23	75
		60	71,43	92,31	85
		80	85,71	92,71	**90**
		100	85,71	69,23	75
		250	85,71	69,23	75
		500	50	80	65
k=5 62	16	20	83,33	90	87,5
		40	83,33	90	87,5
		60	83,33	90	87,5
		80	100	90,91	**93,75**
		100	83,33	90	87,5
		250	83,33	90	87,5
		500	83,33	90	87,5

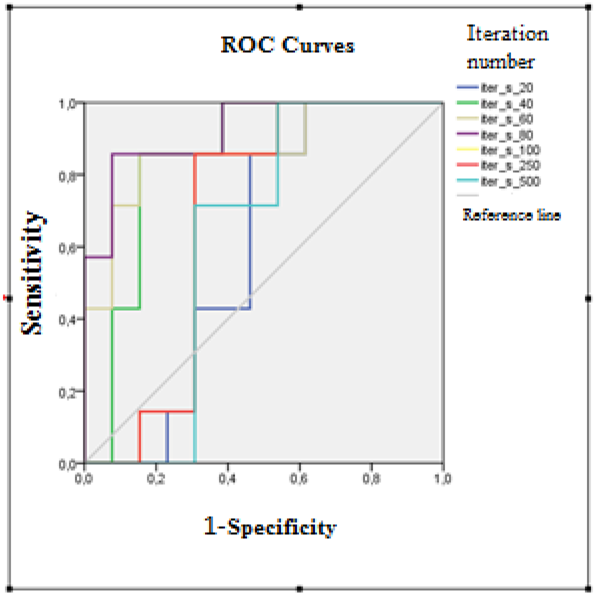


Fig. 4: ROC curves of the fuzzy c-means algorithm results

Fig. 4 demonstrates the ROC curves obtained with different iteration numbers with fuzzy c-means algorithm.

Consequently according to the results of the subtractive algorithm with ANFIS (Table 5), the highest classification accuracy is 93.75% as well when *k* is 5. And when k is chosen according to the highest classification accuracy, the sensitivity is 100% and specificity is 90%. The more comprehensive results which change according to the number of iterations are listed in Table 6, when *k* is 4 and 5 separately.

Table 5: Performance metrics of subtractive algorithm changing with *k* value

Train set	**Check set**	**Optimum Number of iterations**	**Sensitivity (%)**	**Specificity (%)**	**Benign-Malign classification accuracy (%)**
k=2 39	39	100	47,06	54,54	51,28
k=3 52	26	100	80	**93,75**	88,46
k=4 58	20	80	**100**	84,61	90
k=5 62	16	80	**100**	90	**93,75**

Table 6: Performance metrics of subtractive algorithm changing with the number of iterations

Training set	Checking set	Number of iteration	Sensitivity (%)	Specificity (%)	Benign-Malign classification accuracy (%)
k=4 58	20	20	100	76,92	85
		40	100	76,92	85
		60	100	76,92	85
		80	100	76,92	85
		100	100	84,65	**90**
		250	100	76,92	85
		500	100	76,92	85
k=5 62	16	20	83,33	90	87,45
		40	83,33	90	87,45
		60	100	90	93,75
		80	100	90	93,75
		100	100	90	93,75
		250	100	90	93,75
		500	100	90	93,75

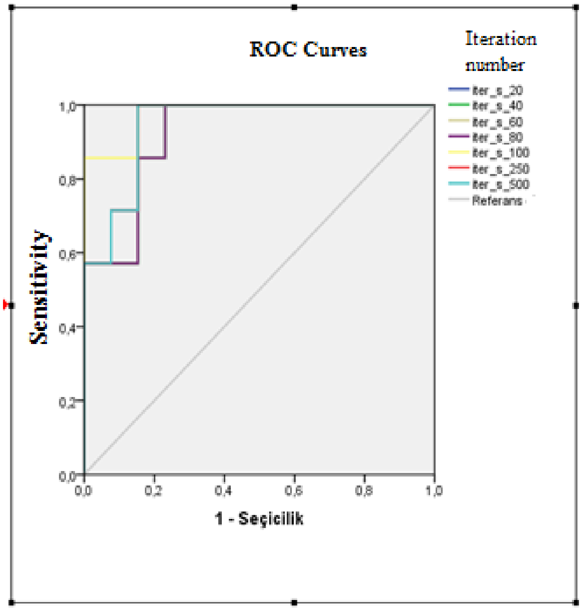


Fig. 5: ROC curves of the subtractive algorithm results

Fig. 5 demonstrates the ROC curves obtained with different iteration numbers with subtractive algorithm. The highest results achieved by using the mentioned three algorithms (K-means, fuzzy c-means and subtractive) are listed in Table 7. As seen in Table 7 the classification accuracy of 93.75% is reached at 80 iterations with Anfis fuzzy c-means and 60 iterations with Anfis subtractive algorithm. Table 8 demonstrates the elapsed time with the related algorithms.

Table 7: The optimum results with K-means, fuzzy c-means and subtractive algorithms based on ANFIS

Classification method	Inference System	Number of iterations	Sensitivity (%)	Specificity (%)	Positive pred. value (%)	Negative pred. value (%)	Benign-Malign classification accuracy (%)
k-means	-	-	94,29	58,14	64,71	92,59	74,36
Fuzzy c-means	ANFIS	80	100	90,91	83,33	100	93,75
Subtractive	ANFIS	60	100	90	85,71	100	93,75

Table 8: Classification time

Algorithm	Elapsed Time
K-means	0.489
Anfis Fuzzy c-means	2.784
Anfis Subtractive	467.528

4. CONCLUSIONS

In this paper, a database consisting of 78 breast masses is used. In the classification process, k-means, fuzzy c-means and fuzzy subtractive methods have been applied to the masses with ANFIS structure. K-means algorithm performs the poorest result among these algorithms. Fuzzy c-means and subtractive algorithm provide the same classification accuracy of 93.75%. On the other hand the elapsed time of subtractive algorithm is much more than fuzzy c-means algorithm. The highest sensitivity is 100% with both fuzzy c-means and subtractive algorithm in optimum. The highest specificity is 90.91% with fuzzy c-means algorithm. According to the satisfying results it is reasonable to use fuzzy c-means and subtractive algorithms with ANFIS structure for classification the breast masses as malignant or benign for a true diagnosis.

REFERENCES

[1] In-Sung J., Devinder T. and Wang G.N. Neural Network Based Algorithms for diagnosis and classification of breast canser tumor. Department of Industrial and Information Engineerin, Ajou University, South Korea, 2011.

[2] Gorgel P. , "Cancer Region diagnosis of 2-dimensional mammographic data using image processing techniques", İstanbul University, The Institute of Sciences, Computer Engineering Department, PhD thesis, 2011.

[3] DUNN, J.C., 1974, A Fuzyy Relative of ISODATA Process and Its Use in Detecting Compact, Well Separated Clusters, Journ., Cybern., 3, 95-104.

[4] BEZDEK, J.C., 1981, "Pattern Recognition with Fuzzy Objective Function Algorithms", Plenum Press, New York.

[5] Hongxing, L. et al. 2001. Fuzzy Neural Intelligent System, Mathematical Foundation and the Application in Engineering. CRC Press LLC.

[6] SUGENO, M., 1977, "Fuzzy measures and fuzzy integrals: a survey," (M.M. GUPTA, G. N. SARIDIS, and B.R. GAINES, editors) Fuzzy Automata and Decision Processes, pp. 89-102, North-Holland, NY.

[7] JANG, J.-S. R. and C.-T. SUN, 1997, "Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence", Prentice Hall.

[8] Geisser, Seymour (1993). Predictive Inference. New York, NY: Chapman and Hall. ISBN 0412034719.