INTERNATIONAL JOURNAL OF ELECTRONICS; MECHANICAL and MECHATRONICS ENGINEERING Vol.2 Num.4 pp.(327-333)

FEATURE EXTRACTION BASED WAVELET TRANSFORM IN BREAST CANCER DIAGNOSIS USING FUZZY AND NON-FUZZY CLASSIFICATION

Pelin GORGEL[1], Ahmet SERTBAS[1], Osman N. UCAN[2]

[1]Department of Computer Engineering, Istanbul University (IU), Istanbul, Turkey
[2] Engineering and Architecture Faculty, Aydin University, Istanbul, Turkey

E-mail: paras@istanbul.edu.tr, asertbas@istanbul.edu.tr, uosman@aydin.edu.tr

Abstract- *This study helps to provide a second eye to the expert radiologists for the classification of manually extracted breast masses taken from 60 digital mammıgrams. These mammograms have been acquired from Istanbul University Faculty of Medicine Hospital and have 78 masses. The diagnosis is implemented with pre-processing by using feature extraction based Fast Wavelet Transform (FWT). Afterwards Adaptive Neuro-Fuzzy Inference System (ANFIS) based fuzzy subtractive clustering and Support Vector Machines (SVM) methods are used for the classification. It is a comparative study which uses these methods respectively. According to the results of the study, ANFIS based subtractive clustering produces ??% while SVM produces ??% accuracy in malignant-benign classification. The results demonstrate that the developed system could help the radiologists for a true diagnosis and decrease the number of the missing cancerous regions or unnecessary biopsies.*

Keywords: *Breast cancer, Fuzzy subtractive clustering, ANFIS, Wavelet transform, Support vector machines*

1. INTRODUCTION

Breast cancer is among the most life-threatening cancer especially in developed countries nowadays. 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 imaging method in the diagnosis of breast cancer is digital mammography. It is very crucial to detect both the malignant and benign masses accurately in the mammograms. In some situtations due to small masses or thick breast tissue the expert radiologists may miss the suspicious regions and so the diagnosis can fail. To deal with this problem 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%. A CAD system was developed by Delogu et al. [3] for the classification of mammographic masses as malignant or benign. They used twelve features based on shape, intensity and size of the segmented masses. In the study by Rangayyan et al. [4] combined speculation index, three shape factors, fractional concavity and compactness and achieved classification accuracy of 81.5%. Cascio et al. [5] used geometrical features about shape parameters for each region of interest to classify the masses. They used supervised neural network which achieved a sensitivity value of 82%. Tralic et al. [6] calculated three shape factors, namely Fourier descriptors, compactness and moments. Classification was performed using both single layer and multilayer perceptron neural networks and the highest accuracy was 91.5%.

This paper is organized as follows: In Section II, Fast Wavelet Transform which is used before the feature extraction is explained. Furthermore subtractive clustering method, ANFIS architecture and SVM are stated. In Section III the data set of mammogram masses is mentioned and the obtained experimental results are presented and discussed. Finally Section IV draws the conclusion and gives some final remarks.

2. MATERIALS AND METHODS

2.1. Pre-Processing for Classification

2.1.1. Fast Wavelet Transform (FWT)

Wavelets are counted as a powerful signal processing foundation of Mallat [7] in 1987. The Fast Wavelet Transform is a computationally efficient form of the discrete wavelet transform (DWT) [8]. It is a multi-resolution analysis method that provides frequency decomposition of the images or signals using scaling ($\varphi_{j,k}(x)$) and wavelet ($\psi_{j,k}(x)$) functions.

$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k) \quad (1)$$

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) \quad (2)$$

In the above equations j and k determines the scaling and wavelet functions' width and the position respectively while the value $2^{j/2}$ controls the amplitude. (3) and (4) illustrate the approximation and detail coefficients respectively in the two-dimensional wavelet transform. $f(x,y)$ is used for the image and m and n are for the image size. The index i is H for horizontal, V for vertical and D for diagonal details.

$$W_\varphi(j,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\varphi_{j,m,n}(x,y) \quad (3)$$

$$W_\psi^i(j,m,n) = \frac{1}{\sqrt{MN}}\sum_{x=0}^{M-1}\sum_{y=0}^{N-1} f(x,y)\psi^i_{j,m,n}(x,y) \quad (4)$$

The FWT is implemented via digital filters and downsamplers as formulated in (5) and (6). After a FWT, four sub-images one of which is the approximation image and the others are horizontal, vertical and diagonal detail images are obtained. Low pass ($h_\varphi(n)$) and high pass ($h_\psi(n)$) filters are used for the approximation and detail coefficients respectively. After the filtering step, downsampling is implemented for the scale changing as seen in Fig. 1. The fast wavelet transform is continued until the sub-images reach the optimum contrast.

$$W_\psi(j,k) = h_\psi(n) * W_\varphi(j+1,n) \quad (5)$$

$$W_\varphi(j,k) = h_\varphi(n) * W_\varphi(j+1,n) \quad (6)$$

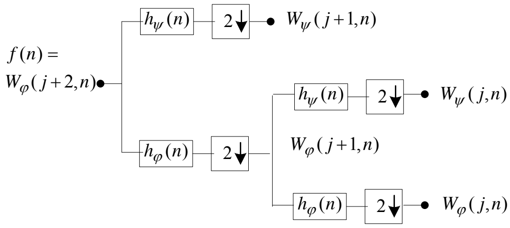

Figure 1. A two scale Fast Wavelet Transform

2.1.2. Feature Extraction

In this study we extract some features about the mass size, geometrical shape and boundary after applying the FWT. The preferred features related with size are as follows (Table I): In a region *Area* is the actual scalar number of pixels, *Centroid* is the center and *BoundingBox* is the smallest rectangle containing the region. *Filled Area* is the number of on pixels in filled image and *Equiv Diameter* ($\sqrt{4 * Area/\pi}$) is the diameter of a circle with the same area as the region. The features related with geometrical shape are as follows: *Euler Number* is the number of objects in the region minus the number of holes in those objects and *Extrema* is the extremal points in the region. *Convex Hull* is the smallest convex polygon that can contain the region. *Solidity* is the proportion of the pixels in the convex hull that are also in the region. Finally the features related with the boundary are as follows: *Major Axis Length* is the length (in pixels) of the major axis of the ellipse that has the same second-moments as the region while *Minor Axis Length* is the length (in pixels) of the minor axis of the ellipse that has the same second-moments as the region.

Table I. Extracted features

Area	Bounding Box
Centroid	Filled Area
Extrema	Convex Hull
Eccentricity	Major Axis Length
Orientation	Minor Axis Length
Symmetry	Mean Center-Border Distance
Solidity	Equiv Diameter
Extent	Euler Number

Eccentricity is the eccentricity of the ellipse that has the same second-moments as the region and it is the ratio of the distance between the foci of the ellipse and its major axis length. *Orientation* means the angle (in degrees) between the x-axis and the major axis of the ellipse that has the same second-moments as the region. *Extent* represents the proportion of the pixels in the bounding box that are also in the region. We also develop two further features which one is boundary based *Mean Center-Border Distance* representing the similarity between a circle and the mass and the other is shape based *Symmetry*. These calculated numeric features listed below provide feature matrices to the fuzzy inference system.

2.2. Classification and Diagnosis

2.2.1 Fuzzy subtractive clustering

The subtractive clustering is used to determine the number of clusters of the data being proposed, and then generates a fuzzy model [9]. The purpose of this algorithm is to estimate both the number and initial locations of cluster centers [10]. The subtractive clustering method partitioned the training data into groups called clusters. By the end of clustering, a set of fuzzy rules will be obtained. The FIS is generated with minimum number of rules. The clustering is carried out in a multidimensional space; the related fuzzy sets must be obtained. Let the cluster set be $Z_1, Z_2, .. Z_n$ for n data. The subtractive clustering algorithm steps are as follows:

- The initial potential value for each data point

(Z_i) as in (7) is computed.

$$P_i = \sum_{j-1}^{n} e^{-d(Z_i - Z_j)} \quad (7)$$

In (7) d is equal to $4/r^2$ where r is the neighborhood for each cluster. If the point falls outside this neighborhood region it has little influence to the potential value.

A point is the first center if its potential value $P(1)$ is equal to the maximum of initial potential value

($P(1)^*$) as demonstrated in (8).

$$P(1)^* = \max(P(1)(Z_i)) \quad (8)$$

A threshold (δ) is defined for the decision to continue or stop the cluster center search.

$$\delta = \mu \times P(1)^* \quad (9)$$

In (9) μ is the reject ratio and $P(1)^*$ is the potential value of the first cluster center.

The previous cluster center from further consideration is subtracted and the remaining points' potential values are adjusted using (10).

$$P_i = P_i - P(k)^* e^{-d(Z_i - Z_k^*)^2} \quad (10)$$

Where Z_k^* the point of the kth is cluster center and $P(k)^*$ is its potential value. This procedure is continued until the maximum potential value in the current iteration is equal to or less than the threshold δ .

2.2.2. ANFIS Architecture

The ANFIS is the abbreviated of adaptive neuro-fuzzy inference system [11]. This method is a fuzzy inference system (FIS) using a backpropagation tries to minimize the Root Mean Square Error (RMSE). As in the artificial neural network systems the input passes through the input layer (by input membership function) and the output could be seen in output layer. The fuzzy rules are learned by the system with through the training process of the ANFIS. Assume that the considered FIS has two inputs x and y and one output f (Fig. 2). For Sugeno fuzzy model, a common rule set with two fuzzy if–then rules is as follows:

Rule 1: If x is $A1$ and y is $B1$, then $f1 = p1x + q1y + r1$

Rule 2: If x is $A2$ and y is $B2$, then $f2 = p2x + q2y + r2$.

The necessary processes are implemented in five layers as seen in Fig. 3 and the overall output is calculated.

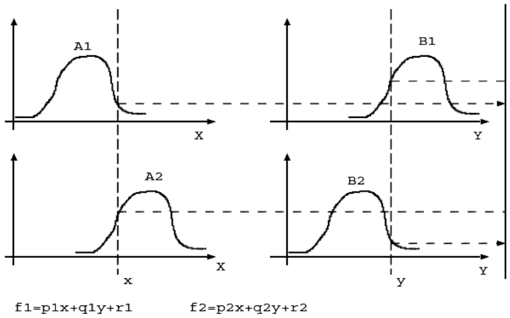

Figure 2. Sugeno fuzzy model

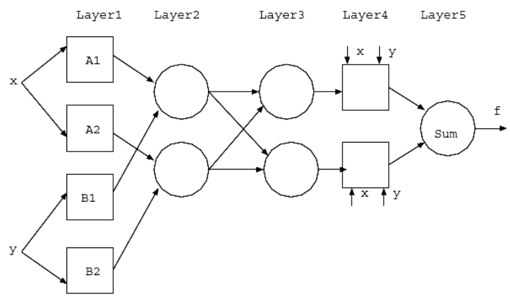

Figure 3. ANFIS architecture with two inputs and an output

2.2.3. Support Vector Machines (SVM)

Support Vector Machine (SVM), introduced by V. Vapnik in 1995 [12], is a method to estimate the data classification function [13]. The basic idea of SVM is to construct a hyperplane as the decision surface in such a way that the margin of separation between positive and negative examples is maximized [14]. A classification task usually involves separating data into training and testing sets. Each instance in the training set contains one target value and several attributes. The goal of SVM is to produce a model (based on the training data) which predicts the target values of the test data given the test data attributes only. SVM uses a kernel function in which the nonlinear mapping is implicitly embedded. In Cover's theorem, a function can be considered as a kernel provided that it satisfies Mercer's conditions [15]. The following relation should be maximized to optimize the SVM classifier boundary in a given training set of instance-label pairs $(x_i, y_i), i = 1,...,l$ where $x_i \in R^n$ and $y \in \{1,-1\}^l$:

$$L(c) = \sum_{i=1}^{l} c_i - \frac{1}{2}\sum_{i,j=1}^{l} y_i y_j c_i c_j K(x_i, x_j), \quad 0 \le c_i \le P \tag{11}$$

While

$$\sum_{i=1}^{l} y_i c_i = 0, \quad w = \sum_{i=1}^{N} c_i y_i x_i, \quad c_i \left[y_i (w^T x_i + b) - 1 + \xi_i \right] = 0 \tag{12}$$

where P is a user-specified positive parameter to control tradeoff between the SVM complexity and the number of non-separable points, l shows number of samples and $K(x_i, x_j)$ is the SVM kernel. Here a solution to $c = (c_1, c_2, ..., c_l)$ is obtained where c_i is a Lagrange coefficient. The slack variables ξ_i are used to relax the constraints of the canonical hyperplane equation. In a typical SVM the kernel function plays an important role in implicitly mapping the input vector into a high-dimensional feature space, in which better separability can be achieved.

2.3. Proposed Algorithm

Firstly both of the malignant and benign masses are passed through a two scale FWT in the classification of breast masses. Because it has been experienced that the sub-images become too blurred after the second scale decomposition. We obtain 8 different coefficient matrices using a two scale FWT. These are approximation (A), horizontal (H), vertical (V) and diagonal (D) coefficient matrices for the first and second scales labeled $A_1, H_1, V_1, D_1, A_2, H_2, V_2$ and D_2. Since the first scale detail coefficients are generally composed of poor information, only the mean of those matrices ($M(H_1, V_1, D_1)$) is used. Consequently, six coefficient matrices are used totally for each mass. The next step feature extraction is applied to those matrices separately. The pre-processed mass images are then classified for diagnosis. For the classification subtractive clustering based ANFIS and SVM methods are implemented respectively to make a comparison. The flow chart of the proposed algorithm is illustrated in Fig. 4.

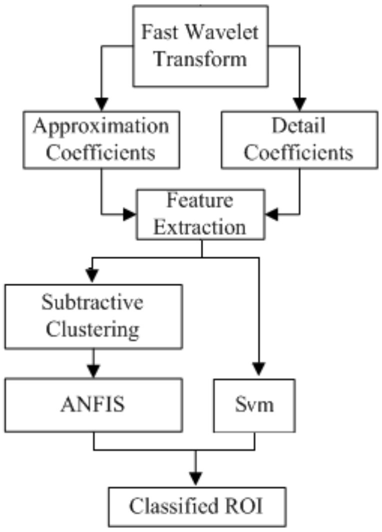

Figure 4. The flow chart of the proposed method.

In fuzzy subtractive – ANFIS method, the fuzzy rules are trained maximum 50 epochs until RMSE converges to zero. A reject ratio of 0.5 is used and the radius is specified as 0.6. In Eq.13 x, y and N demonstrate targets, outputs and data size respectively in the ANFIS.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - y_i)^2}{N}} \tag{13}$$

3.THE DATASET AND EXPERIMENTAL RESULTS

The digital mammogram dataset used for this study consists of 35 malignant and 43 benign masses (Fig. 5,6). The mammograms are acquired from the patients of Istanbul University Faculty of Medicine Hospital in Turkey.

Table II. Number of the masses in training and test sets

	Train set	Test set
Malignant	23	12
Benign	29	14

The malignant and benign masses have been pre-extracted by a radiologist manually. One-third of the data set (26 masses) was used for testing and the others were used for training. 14 masses are benign and the others are malignant among 26 test masses (Table II). Using these data sets 47 fuzzy rules are extracted in the whole test process. Before the ANFIS training process the initial error was 0.14. As seen in Table III, when the network is trained 50 epochs which produce best classification, the training error is decreased to 0.012 and the classification accuracy reaches to 92%. If the epoch number is increased or decreased, the error changes but the accuracy becomes stable at 88% each time. In Table IV the confusion matrices are given obtained after different training epochs.

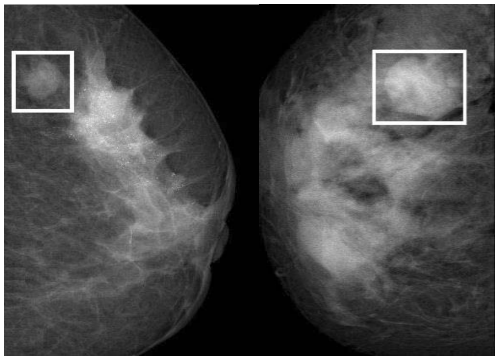

Figure 5. Samples from the dataset with framed benign masses

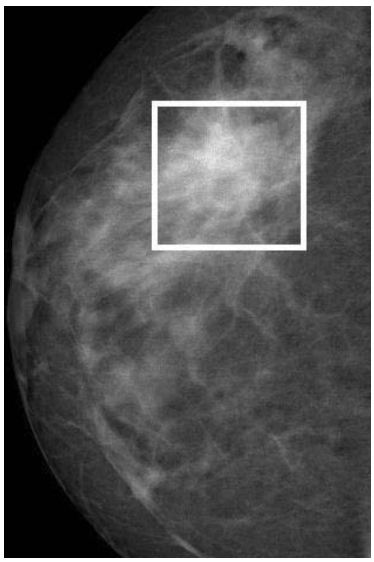

Figure 6. Samples from the dataset with framed malign mass

Table III. Experimental results of the Fuzzy Subtractive - ANFIS method. (Tr.Ep. : Training epoch, Tr. Err. : Training error, A_z. : Area under the ROC curve, Sen.(Rec.) : Sensitivity (Recall), Spe. : Specificity, Pre. : Precision, FMs. : FMeasure, Acc. %: Accuracy)

Tr. Ep.	Tr. Err.	A_z	Sen. (Rec.)	Spe.	Pre.	FMs.	Acc. %
20	0.041	0.88	75	100	100	86	88
30	0.032	0.88	83	93	91	87	88
40	0.024	0.89	100	79	80	89	88
50	**0.012**	**0.93**	**100**	**86**	**86**	**92**	**92**
60	0.020	0.89	100	79	80	89	88
70	0.028	0.89	92	86	85	88	88
80	0.035	0.88	83	93	91	87	88

Table IV. Confusion matrix of the Fuzzy Subtractive - ANFIS method (Tr.Ep.: Training epoch, TP: True Positives, TN: True Negatives, FP: False Positives, FN: False Negatives)

	Tr. Ep. 20	Tr. Ep. 30	Tr. Ep. 40	Tr. Ep. 50	Tr. Ep. 60	Tr. Ep. 70	Tr. Ep. 80
TP	9	10	12	**12**	12	11	10
TN	14	13	11	**12**	11	12	13
FP	0	1	3	**2**	3	2	1
FN	3	2	0	**0**	0	1	2

Additionally SVM method with 3-fold cross validation is applied to the pre-processed masses. The comparative results are demonstrated in Table V-VI.

Table V. The comparative results (Sensitivity (Recall), Spe.: Specificity, Pre. : Precision, FMs. : FMeasure, Acc. %: Accuracy)

	Sen. (Rec.)	Spe.	Pre.	FMs.	Acc. %
Fuzzy Subtractive ANFIS	100	86	86	92	92
SVM	92	85	86	89	88

Table VI. Confusion matrix of two methods (TP: True Positives, TN: True Negatives, FP: False Positives, FN: False Negatives)

	TP	TN	FP	FN
Fuzzy Subtractive ANFIS	12	11	2	1
SVM	12	12	2	0

4. CONCLUSION

Computer aided diagnosis systems used for medical decision provide medical data to be examined in shorter time and in more detail and early diagnosis. The research presented in this article aims to decrease the mortality rate related to breast cancer by reducing the number of malignant masses which radiologists may miss by means of computer aided techniques. We developed a program in MATLAB 7.6 using FWT and feature extraction for pre-processing. Fuzzy subtractive clustering based ANFIS and SVM methods were used respectively for the classification as malignant or benign. According to the results the best accuracy is performed as 92% by fuzzy subtractive based ANFIS with 50 epochs. On the other hand the highest accuracy of the SVM method is 88%. The performance difference depends to the ANFIS architecture which enables the system to learn the problem until the error decreases to a desired value. Consequently one can see that this method is efficient for solving the real world problems related with breast cancer diagnosis using FWT multi-resolution decomposition and ANFIS rule extraction and learning methods. The satisfying performances demonstrates that this study is valuable to improve early diagnosis and reduce the number of unnecessary biopsies.

5. 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] P. Delogu, M. E. Fantacci, P. Kasae, A. Retico, "Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier", *Computers in Biology and Medicine*, vol. 37, pp. 1479 – 1491, 2007.

[4] R. M. Rangayyan, N. R. Mudigonda, J. E. L. Desautels, "Boundary modelling and shape analysis methods for classification of mammographic masses", *Medical and Biological Engineering and Computing*, vol. 38, no.5, pp. 487-496, 2000.

[5] D. Cascio, F. Fauci, R. Magro, G. Raso, "Mammogram segmentation by contour searching and mass lesions classification with neural network ", *IEEE Transactions on Nuclear Science,* vol.53, no.5, pp. 2827-2833, 2006.

[6] D. Tralic, J. Bozek, S. Grgic, "Shape analysis and classification of masses in mammographic images using neural networks", *18th International Conference on Systems, Signals and Image Processing (IWSSIP),* Sarajevo, 2011.

[7] S. Mallat, "A compact multiresolution representation: The wavelet model", *IEEE Computer Society Workshop on Computer Vision*, Washington, 1987.

[8] R. C. Gonzales, *R. E. Woods, Digital Image Processing using MATLAB,* Prentice Hall, USA, 2004.

[9] M. Alata, M. Molhim, "Optimizing of fuzzy C-means clustering Algorithm using GA", *World Academy of Science, Engineering and Technology*, vol. 39, 2008.

[10] A. Elmzabi, M. Bellafkih, M. Ramdani, "An adaptive fuzzy clustering approach for the network management", *World Academy of Science, Engineering and Technology,* vol.31, pp. 425-430, 2007.

[11] P. Tahmasebi, A. Hezarkhani, "Application of Adaptive Neuro- Fuzzy Inference System for Grade Estimation; Case Study", *Australian Journal of Basic and Applied Sciences,* vol.4, no.3, pp. 408-420, 2010.

[12] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998

[13] C.J.C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Kluwer Academic Publishers, Dordrecht, 1998.

[14] G. B. Junior, A. C. Paiva et al., Classification of breast tissues using Moran's index and Geary's coefficient as texture signatures and SVM, Computers in Biology and Medicine, 39 (2009), 1063 -1072.

[15] J. Yan, B. Zhang, N. Liu et al., Effective and efficient dimensionality reduction for Large-Scale and Streaming Data Preprocessing, IEEE Transactions on Knowledge and Data Engineering, 18 (2006), 320-333.