Volume 4, No. 10, Sep-Oct 2013

ISSN No. 0976-5697

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Improve Breast Cancer Detection in Mammography Images Using Neuro-Fuzzy (ANFIS)

Aioub Zeinvand Lorestani*
Department of Computer
Science and Research Branch, Islamic Azad University
Khouzestan, Iran
zl_ayuub@yahoo.com

Dr Amir Massoud Bidgoli
Department of Computer Engineering
Islamic Azad University, Tehran North Branch
Tehran, Iran
drbidgoli@gmail.com

Dr Mashallah Abbasi Dezfoli
Department of Computer
Science and Research Branch, Islamic Azad University
Khouzestan, Iran
m.abbasi@khouzestan.srbiau.ac.ir

*Abstract***:** Breast cancer is the second largest cause of cancer deaths among women. Breast Cancer is one of the major tumor related of death in women. The most effective way to reduce breast cancer deaths is detect it earlier. This paper outlines an approach for recognizing breast cancer diagnosis using neuro-fuzzy inference technique namely ANFIS (Adaptative Neuro-Fuzzy Inference System) in order for better detection and prognosis of breast cancer in mammography images. Results that the best performances are obtained by our model compared to others cited in literature (an accuracy of 98 %).

Keywords: Breast cancer; Neuro-Fuzzy inference; ANFIS; Mammography; Accuracy.

I. INTRODUCTION

Breast cancer is the most common cancer among the women world population, affecting each year an average of 1.4 million people [2]. Breast cancer comprises 1 in 5 of all new cases of cancer, Figure 1.1. It is also the most common form of cancer death, representing 1 in 8 of all deaths from cancer, according to the International Agency of Research on Cancer [3]. The mammogram is the most efficient system to detect clinically occult illness, being the only image-based method recommended for breast cancer screening [4]. Mammography can greatly reduce the breast cancer mortality in a well-organized screening program over the population, being the breast cancer detection technique that most reduces mortality [5].

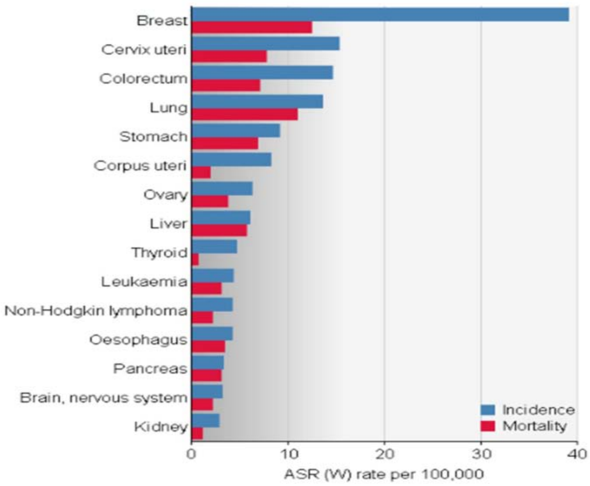

Figure 1. Cancer incidence among women world population [3].

In recent years, many studies have been done on mammography images to detect cancerous mass by computer programming and image processing methods without the interference of an individual detector to avoid the individual's fatigue, inaccuracy and vision problems.

Correct mammographic detection of asymptomatic lesions is essential to discover early breast cancer phases, increasing the treatment options and survival rate [6]. Mammography image may be of low contrast due to inadequate brightness and low dynamic range of imaging sensors. The increase of contrast may result in the increase of the dynamic range of the image. Often, before displaying the image it is necessary to make some changes. Image enhancement includes techniques such as contrast manipulation, reduction of noise and edges sharpening. The usual task of mammogram enhancement is to increase the contrast between regions of interest and background and to sharpen the edges or borders of the ROI [14].

II. IMAGE PROCESSING AND ANALYSIS ON MAMMOGRAPHIC IMAGES

Usually for edge detection of images common algorithms of displaying edges are used, but these algorithms are not suitable for medical noisy images. In this method, mathematical morphology operator is used for preprocessing and noise removing. The morphological base operator's correspond to erosion and dilation, which are inverse operators of each other. These operators decrease or increase the size of objects in binary images, respectively, being controlled by a structuring element [7].

First, Top-hat operator and then Bottom-hat operator have been implemented on the original image. Then, all Top-hat operator image pixels are subtracted from Bottom-hat operator image pixels. Figure 2-1 shows the result of the operation on mammography image.

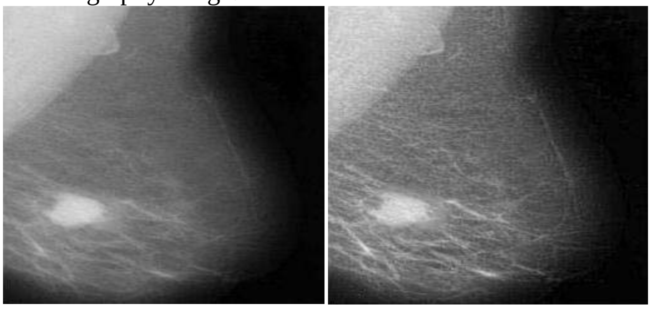

Figure 2. On the left: the original image, on the right: the improved image [11].

In the following, in order to remove bright calcium particles (benign areas of the image) by a disc-shaped structural element, with a diameter of 2 pixels, the image is eroded. The result of this conversion has been shown in the following figure.

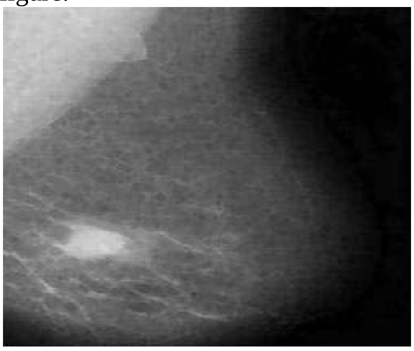

Figure 3. Converting erosion to an image [11].

The pixels that are close to tumor threshold limit can be extracted as candidate area to the tumor and from them on only processing operation should be implemented on the areas. Threshold limit for image database has been considered as 190 and all pixels that are more than 190 considered as candidate pixel.

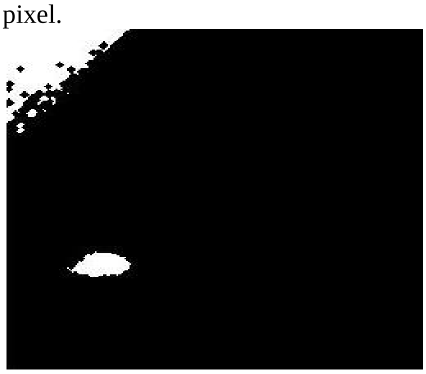

Figure 4. Extracting candidate areas from the mammography image [11].

In medical and technical drawings, a commonly used approach is to explain certain features of the drawing with text labels that are arranged on its boundary. After segmenting candidate areas, the light pixels that are connected together get the same label. This labeling may cause the features of each area to be classified separately and
The related patterns of each area will be analyzed separately.

III. EXTRACTED FEATURES FOR CLASSIFYING CANDIDATE AREAS

First, confirm that you in this section features of the candidate areas are extracted from mammography image. After labeling candidate areas, features of each area are extracted to send to the classifier and the training stage is implemented. The desired features that answered appropriately in training section are as the following:

- The total intensity of the pixels around the candidate area.
- The average intensity of pixels of candidate area.
- The intensity standard deviation of the pixels of candidate area.
- Maximum intensity of the pixels of candidate area.
- Minimum intensity of the pixels of candidate area.
- The entropy intensity of the pixels of candidate area.
- Total pixels of candidate area.
- Total gradient in X direction.
- Total gradient in Y direction.

IV. CLASSIFYING CANDIDATE AREAS

After extracting features, the features of each candidate area are put into the desired class so that training and testing could be done. Moreover, it can be detected whether it is of cancerous or non-cancerous class.
In order for classification, usually the following categories are applied:
1. Neural networks
2. Fuzzy logic
3. Neuro-fuzzy system

In this study, neuro-fuzzy method was applied [15]. The applied neuro-fuzzy system of this study is takagi-sugeno. The system possesses parameters, rules and functions of compatible fuzzy member which is optimized during training by means of back propagation method. Training the system is supervisory. Even though all the rules and parameters are optimized automatically, but if necessary, parameters can be set up manually in every stage. This is one of the best features of neuro-fuzzy systems.
In this study, for implementing the system neuro-fuzzy of toolbox matlab 7 and the system by the control parameters are optimized.

First, basic structure of the system should be identified. The basic structure includes number of inputs, outputs, fuzzy conditional rules and form of membership functions. Number of inputs is considered as equal as the dimensions of applied feature vector.

Number of outputs of the system equals 1 in every state. This means that only one neuron is considered in outer layer and the system is taught in a way that the output figure of the neuron equals to the number of desired class. Block structure of the system has been illustrated in figure 4-1. As discussed earlier, only the numbers of the inputs of the system are same as the number of features.

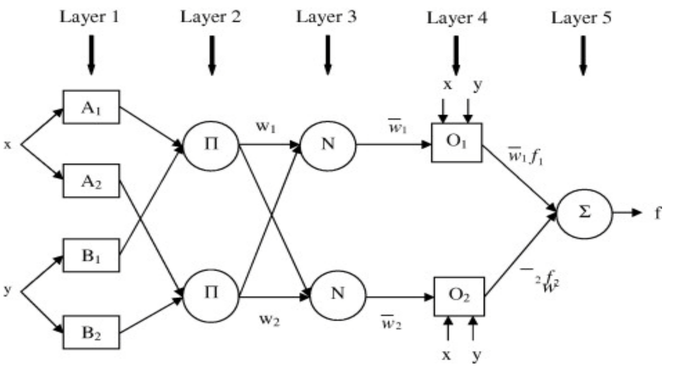

Figure 5. The proposed neuro-fuzzy system block structure [15].

The Takagi Sugeno inference engine used rules are as follows:

- ***Rule 1***: if (in1 is in1cluster1) and (in2 is in2cluster1) and (in3 is in3cluster1) and (in4 is in4cluster1) and (in5 is in5cluster1) and (in6 is in6cluster1) and (in7 is in7cluster1) and (in8 is in8cluster1) and (in9 is in9cluster1) then (out1 is out1cluster1).
- ***Rule 2***: if (in1 is in1cluster2) and (in2 is in2cluster2) and (in3 is in3cluster2) and (in4 is in4cluster2) and (in5 is in5cluster2) and (in6 is in6cluster2) and (in7 is in7cluster2) and (in8 is in8cluster2) and (in9 is in9cluster2) then (out1 is out1cluster2).

As discussed earlier, training this system is supervised learning. Therefore, for training the system a goal vector is established which corresponds to the number of training classes. Since one neuron has been considered, the target vector is defined as follows:

Target Vector T :{ 1, 2, 3… n+1} n: number of classes.

As it was mentioned, the number of classes is considered as one class more than the actual number of classes so that unknown classes can be attributed to them. However, in the images tested in the present study, this class is not much of application, because in these images every class is predetermined and the system is trained to all of the classes. Now we can start training the system. The program is designed in a way that to start training two input values are requested from the user, one is the number of training samples and the other is the number of periods or iterations to reach the goal. To compare the achievements, we use 70 for training Suggested System and Network MLP and the rest 30 are used for the test. More accurate structure of the system has been illustrated in figures 5 and 6.

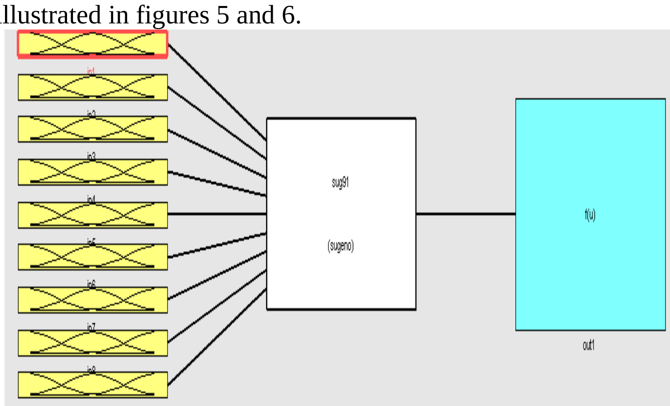

Figure 5. Structure of the proposed neuro-fuzzy system

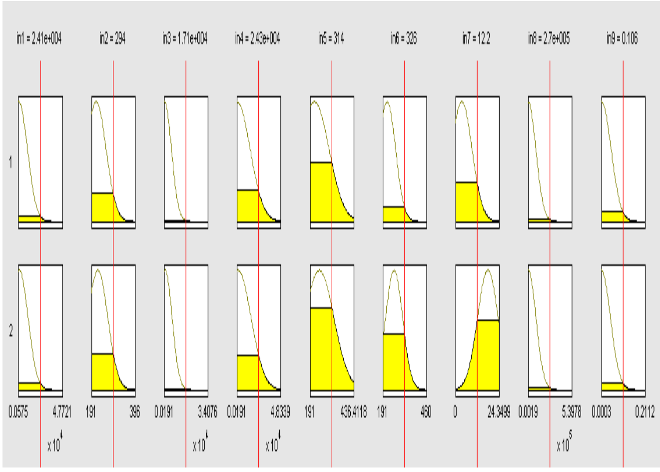

Figure 6. System membership functions after training

The number of iterations of reaching the goal indicates that the system at the training stage in each training attempts several times to reach its goal. In other words, the parameter indicates the number of optimizing cycles and coefficient adjustment in back propagation mechanism of the system. This works the same for MLP as well. The value should be chosen in a way that root mean square error of the training is minimized. By increasing the iterations, the required time for training increases exponentially. On the other hand, we cannot consider the number of iterations as being too large because it takes too much time for training.

V. Conclusion

This section consists of the results obtained from implementing the algorithms in the proposed method in order to extract tumor area from mammography images. The results of implementing algorithm have been presented based on 2 criteria: sensitivity and specificity that are defined as follows:

$sensitivity$: the ratio of the pixels that have identified as tumor to the total pixels of tumor.

$specificity$: The ratio of the number of pixels that have been detected as healthy to the total pixels of healthy.

High level of the two criteria indicates better performance of the methods. The two criteria can be computed applying the following formulas:

$$sensitivity = \frac{T_p}{T_p + F_n}$$

$$specificity = \frac{T_n}{T_n + F_n}$$

- *Tp:* the number of abnormal pixels that are detected as abnormal.
- *Tn:* the number of normal pixels that are detected as normal.
- *FP:* the number of normal pixels that are detected as abnormal.
- *FN:* the number of abnormal pixels that are detected as normal.

TableI. The results of implementing the proposed algorithm on the 5 images.

The result image	Tp	Tn	Fp	Fn	sensitivity	specificity
First	115	191153	9	12	.9055	.9799
second	586	190676	7	9	.9848	.9898
Third	136	191888	8	16	.8947	.9898
Fourth	106	191901	4	11	.9059	.9999
Fifth	780	191201	23	8	.9898	.9998

The required average time for in order to find the situation using CPU 1.8 GHz and 1G RAM was calculated in less than 80 second. During extracting the tumor area, choosing high threshold value (higher than 190) may result in increasing sensitivity and decreasing feature. The optimal value of the threshold experimentally was determined 190 for the desired image base.

Table II. Comparing the proposed algorithm with the previous methods.

	The proposed algorithm	***sensitivity***	***specificity***
Mammography images	***Matosi[9]***	0.90	0.91
	Arfan[1]	0.93	0.96
	The prposed method	0.95	0.98

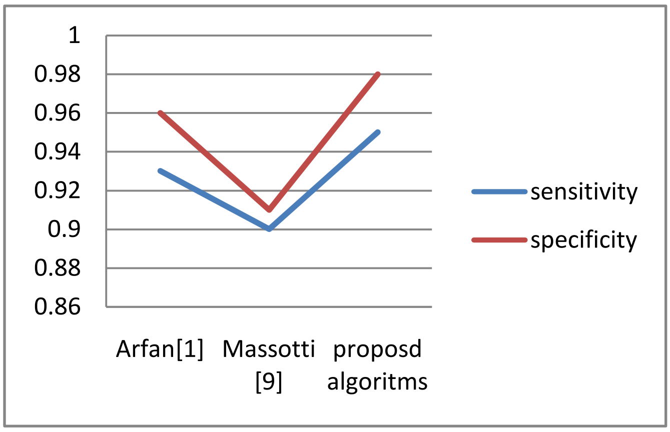

Figure 7. Comparative chart of the proposed algorithm with the previous methods.

VI. DISCUSSION

Radiological imaging plays a significant role in medical diagnosis by increasing image quality, upgrading imaging systems and computer technology. Image interpretation by a radiologist is affected by nonsystematic research of disease patterns and it can be varied in different individuals. Computer-based systems could be applied as the second option in pathological diagnosis. They improve the quality of medical images interpretation with a better review of different disease patterns and the accurate decision in diagnosis.

An important factor which analyzes the record of computer-based systems is sensitivity parameter. It is a term which expresses percentage of diagnosed injuries in each image. In the present study, an attempt has been made to decrease mammography image features for processing and detecting cancer applying neuro-fuzzy system. By subtracting the value of feature matrix, the algorithm complexity decreases and we can achieve the desired answer with less number of training. Since the number of training in this system compared with fuzzy systems or multi-layer neural networks has been less and the time of diagnoses has decreased.

VII. REFERENCES

[1] Arfan Jaffar M. , Naveed N., Zia S., Ahmed B., Tae-Sun Choi(2011) “DCT features based malignancy and abnormality type detection method for mammograms” International Journal of Innovative Computing, Information and Control, Volume 7, Number 9.

[2] Autier, P., Boniol, M., lavecchia, C., Vatten, L., Gavin, A., Héry, C., et al. (2010). Disparities in breast cancer mortality trends between 30 European countries: retrospective trend analysis of WHO mortality database. *BMJ 2010*, 341:c3630.

[3] Ferlay, J., Shin, H., Bray, F., Forman, D., Mathers, C., & Parkin, D. (2010). *GLOBOCAN 2008*. Retrieved from Cancer Incidence and Mortality Worldwide: IARC cancerbase: http://globocan.iarc.fr.

[4] Chagas, F., Rodrigues, D., Tavares, J., Reis, A., Miranda, D., & Duarte, R. (2007). An Application of Hough Transform to Identify Breast Cancer in Images. *Vipimage.*

[5] Eastman, G. W., Wald, C., & Crossin, J. (2006). Getting Started in Clinical Radiology - From image to diagnosis. Thieme.

[6] Lee, C. (2002). Screening mammography: proven benefit, continued controvery. *Radiology Clinics of North America, 40*, pp. 395-407.

[7] Gonzalez, R., & Woods, R. (2002). *Digital Image Processing* (2nd Edition ed.). New Jersey: Prentice Hall, Inc.

[8] Mamdani E. H. And Assilian S.,(1975), "An experiment in linguistic synthesis with a fuzzy logic controller", International Journal of Man-Machine Studies, Vol. 7, No.1, pp. 1-13.

[9] Massotti, M. (2006). A ranklet-based image representation for mass classification in digital mammograms. Medical Physics, 33(10), 3951–3961.

[10] Mencattini A., Salmeri M.,(2008)" Mammographic Images Enhancement and Denoising for Breast Cancer Detection Using Dyadic Wavelet Processing",IEEE transactions on instrumentation and measurement, VOL. 57, NO. 7.

[11] Mias Data Set:/Http://Peipa.Essex.Ac.Uk/Ipa/Pix/Mias.

[12] Menhaj M. B.(2009) "Fundamentals of Neural Networks, Amirkabir University of Technology, (in Farsi).

[13] Mohamed M.E. ,Ibrahima F., Brahim B. S.(2012)“A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multi resolution representation” Computers in Biology and Medicine 42 ,pp. 123–128.

[14] Cheng, H., Cai, X., Chen, X., Hu, L., & Lou, X. (2003). Computer-aided detection and classication of microcalcifications in mammograms: A survey. *Pattern Recognition, 36*, pp. 2967 – 2991.

[15] Robert F. (2000) “Introduction To Neuro-Fuzzy Systems” Advances In Intelligent And Soft Computing, Vol. 2.

