RESEARCH ARTICLE OPEN ACCESS

FCM for Malignant Detection in Mammogram

E. Vidhya Jyothi*, R. Gayathri**
*(Department of Electronics and Communication Engineering, Sri Venkateshwara College of Engineering, Chennai)
** (Department of Electronics and Communication Engineering, Sri Venkateshwara College of Engineering, Chennai)

ABSTRACT
X-ray Mammography is the current widespread method for screening breast cancer and one of the reliable techniques for diagnosis. Radiologists need information from multiple views of Mammogram in order to analyze and detect abnormalities. Computer Aided Detection (CAD) tool are used as second reader to aid the radiologist in early and efficient diagnosis. In this Paper, CAD is designed using the Proposed Algorithm is used for the detection of Cancerous tissue site in the mammogram. Hence, a modified fuzzy c-means (FCM) algorithm is presented for Mammogram image segmentation. Principally during the clinical analysis of mammogram images in CAD Process, segmentation is considered as one of the preliminary and requisite stage of image process. During segmentation, Interpretation of the noise effect has been reduced by incorporating the median filter into the proposed standard FCM Algorithm. KNN classifier does the cluster algorithm using for segmentation and classification. The efficiency of the proposed algorithm is confirmed by extensive segmentation experiments using simulation in CAD.
Keywords- Mammogram, Mammography, Computer Aided Diagnosis, Segmentation, Classification.

I. INTRODUCTION

As per statistics was conducted, the results came out with the breast cancer is the second major cause of cancer death predominantly for women in all over the world. Breast cancer occurs when normal cells become abnormal. These abnormalities are due to the out of control growth of cancer cells, which is called as malignant (cancerous) tumour. The cancer cells have potential to spread and affect other parts of the body through blood vessels and the lymphatic system though it originated in the breast. Normally breast cancers usually occur in women over the age of 50 years. Through early detection Morality rate could be decreased. A report estimated that one in eight women in the U.S. and 1 in 13 in Australia affected by breast cancer during their lifetime. As the Report stated from Tata Memorial hospital in the year (1974-78) 1 out of 1000 women in India are suffering from Breast Cancer. Prior to the physician or a woman can feel, most effective method of detecting cancer at an early stage is Mammography. Being not notify the symptoms of breast cancer, screening mammography is act as a preventive measure for women. The advanced mammography technique has employed the special x-ray machines were designed to expose the breast with less radiation and used for breast imaging. This modern technique was only existed after the late 1960s.

In most cases of women prior to menopause, the majority of new breast lumps found are being (non-cancerous). However, examine after menopause one in two new breast lumps found will be malignant.

II. LITERATURE SURVEY

Several research groups have focused on mammogram analysis and understanding for early diagnoses of breast cancer. At the outset, the statistical features of mass present in mammogram have been extracted for the diagnosis purpose. Different of statistical texture features such as uniformity, smoothness and third moments, which utilize the gray value or histogram of masses for classification. In the paper [1], the author has been used texture features derived from Spatial Gray Level Dependency (SGLD) matrices for classification of masses as normal or abnormal. Forty numbers of mass abnormalities have been considered, out of which 21 are circumscribed masses, 19 are speculated masses and 104 normal

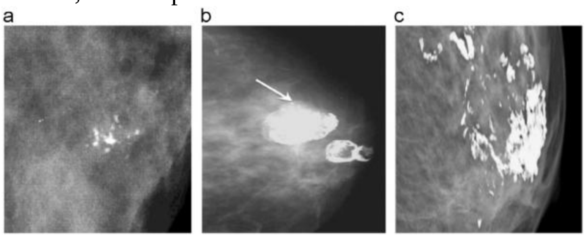


Figure (2.4) Irregular calcification (b) popcorn calcification and (c) dystrophic calcification

Various methods using wavelet have been proposed for feature extraction in mammograms. The efficacy of any diagnosis system based on wavelet coefficients is getting better by using multiresolution analysis [2]. In their mammogram analysis study, In order to detect the speculated mass, the diagnostics system is designed with set of statistical features is used with a binary tree classifier. The success rate was 84.2% achieved. The system based on wavelet analysis and the classifier is used to distinguish between mass and micro-calcification is ANFIS (the Adaptive Neuro-fuzzy inference system), the maximum classification achieved was 85.4%. The multiresolution analysis of digital mammogram has been studied using wavelet transform [3]; they adopted Euclidean distance used to classify the different types of clusters between micro-calcification clusters, speculated mass, circumscribed mass, ill-defined mass and normal mammogram. The utmost classification rate obtained was 87.06%. In order to separating micro calcification clusters, circumscribed mass, spiculated mass and normal classes of image, to classify the mammogram images by transforming the images into wavelet bases and then using a set of coefficients from the first level of decomposition as the feature vector [4]. The maximum classification rate achieved was 94.85%.

Abundant segmentation algorithms have been proposed for segmenting the mass region. Each has its own advantage in some viewpoint. Pectoral segmentation and articaft removal are the important pre-processing works. The morphological operators and the pectoral muscles are segment using seeded region growing method [5]. The mammogram image is enhanced by the successful methods such as the Contrast limited adaptive histogram equalization (CLAHE) and multiscale contrast enhancement algorithm.

Detection of Suspicious Lesions by Adaptive Thresholding [6] and developed suspicious lesions in mammograms are detected by, which utilizes the combination of adaptive global thresholding segmentation and adaptive local thresholding segmentation on a multiresolution representation of the original mammogram. The algorithm has been verified the algorithm with 170 mammograms in the MIAS (MiniMammographic database). The experimental results show that the detection method has a sensitivity of 91.3% at 0.71 false positives per image.

Predicting Breast Screening based Attendance by means of Machine Learning Techniques [7]. The comparative study had done and found that the efficiency and validates its accuracy of the proposed algorithm with different platforms. The Proposed algorithm was almost 82% accuracy and 89% positive predictive value and sensitivity were recorded. Predicting breast-screening attendance using machine learning (earlier mammogram) is a new field.

III. PROPOSED METHOD

3.1 Standard FCM algorithm

The FCM clustering algorithm is proposed in this paper, which is an enhancement of the hard k-means algorithm. It assigns a class or cluster membership to a data point, depending on the similarity of the data point to a particular class relative to all other classes. The standard FCM objective function of segmenting an image into *c* clusters is given in Equation. 5.1

$$G_m(a,b) = \sum_{i=1}^{u}\sum_{j=1}^{n} a_{ij}^{m} d^2(a,b)_e \qquad (5.1)$$

Subject to

$$\sum_{i=1}^{e} a_{ij} = 1 \qquad (5.2)$$

where $X = (x1, x2,. . .xj,. . .xn)$ is a data matrix with the size of $\upsilon \times n$, υ represents the dimension of each xj “feature” vector, and n represents the number of feature vectors (pixel number in the image). The feature vector X in MR images is the pixel intensity, so $\upsilon = 1$ μij is the membership of the jth data in the ith cluster ci, m presents the index of fuzziness (in our study, we choose $m= 2$), and b_i is the fuzzy cluster centroid of the ith cluster. Using the Euclidean norm, the distance metric d measure the similarity between a feature vector xj and a cluster centroid is given in the equation 5.3

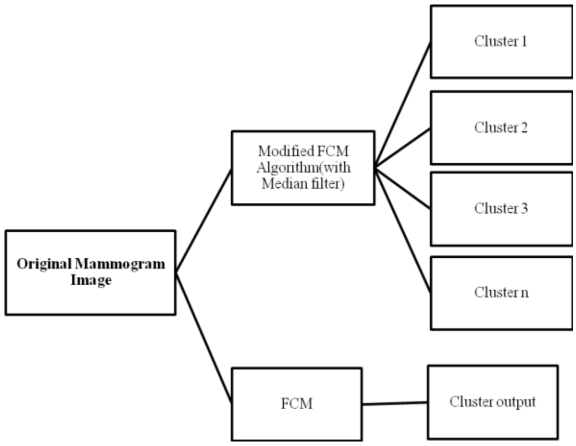
$$d^2(x_j, b_i) = II(x_j - b_i)II^2 \qquad (5.3)$$

The objective function is minimized when data points close to the centroid of their clusters are assigned high membership values, and low membership values are assigned to data points far from the centroid. Letting the first derivatives of G*m* with respect to μ and v equal to zero yields the two necessary conditions for minimizing G*m* as follows as in the Eqn. (5.4)

$$a_{ij} = \left(\sum_{k=1}^{c}\sum_{i} (d(x_j, b_i) / d(x_j, b_k))^{2(m-1)}\right)^{-1} \qquad (5.4)$$

To get the optimum segmentation solution, the FCM algorithm proceeds iteration process with the two necessary conditions until to satisfy the desired condition. Because of FCM clustering, a class is assigned to each data point will be associated with a membership value. By assigning the data point to the class with the highest membership value, a segmentation of the data could be obtained. The pattern of neighbors is called the "window", which slides, pixel by pixel over the entire image.

IV. PROPOSED BLOCK DIAGRAM



4.1 Mammogram Dataset

The methodology presented in this work was applied on the complete mini MIAS database. It is available online freely for scientific purposes and consists of 161 pairs of mediolateral oblique view mammograms. The films were digitized and the corresponding images were annotated according to their breast density by expert radiologists, using three distinct classes: Fatty (F) (106 images), Fatty-Glandular (G) (104 images) and Dense-Glandular (D) (112 images). Any abnormalities were also detected and described, including calcifications, well-defined, speculated or ill-defined masses, architectural distortion or asymmetry. Each pair of images in the database is annotated as Symmetric (146 pairs) or Asymmetric (15 pairs).The severity of each abnormality is also provided, i.e., benignancy or malignancy.

4.2 Pre-processing

Digitization noise, artifacts presents in raw mammogram will greatly affect the diagnosis result in a CAD system, Pre-processing need to be performed to remove them. A 2D Median filter is used to remove the digital noise.

4.3 Median Filtering

Median filtering is a nonlinear method used to remove noise from images. It is widely used as it is very effective at removing 'salt and pepper' noise while preserving edges. The median filter works by moving through the mammogram image pixel by pixel, replacing value of each 3X 3 kernel with the median value of neighboring pixels. First sorting all the pixel values from the window in numerical order, and then replacing the pixel by calculating the median (middle) pixel value.

4.4 K NEAREST NEIGHBOR

In this paper, we focus on efficient processing of K-Nearest Neighbor (KNN) for classification. K-Nearest Neighbor (KNN) algorithm is one of the most popular learning algorithms in data mining. The KNN algorithm is to find a set of k objects in the training data that are close to the test pattern, and base the assignment of a label on the majority of a particular class in this neighborhood. Given a training set M and a test pattern y, KNN computes the similarity (distance) between y and the nearest k neighbors.

The basic KNN algorithm is shown below.

Algorithm 2 Basic KNN Algorithm

Input:
M: Training set
n: Number of patterns in M
y: Test pattern
Output:
`: label of y
1: for i M 1 to n do
2: Calculate m.y; yi /, the distance between y and yi ;
3: end for
4: choose the set of k nearest training patterns for y;
5: M arg max_2L f ._ M class.cxi //;
f ._/ is a function that returns the value 1 if its
is true and 0 otherwise.
6: return;

V. SIMULATION RESULTS

I. Original Mammogram Image

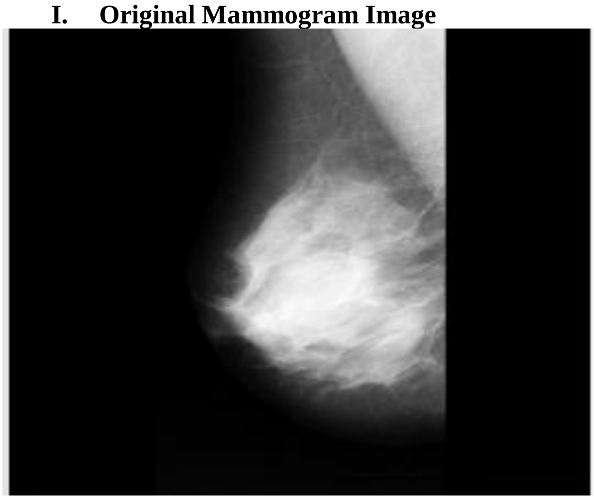


Figure (2.5) Original mammogram image=

Modified FCM algorithm Output

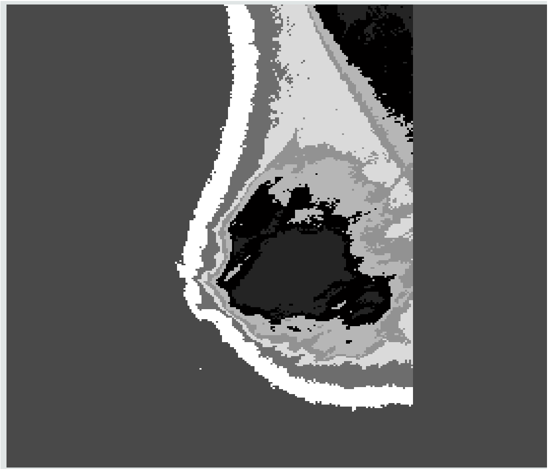


Figure (2.6) Multi Cluster segmentation Image

The above image in figure (2.6) represents the mulitcluster image of mammogram after simulation of original mammogram image using fussy cluster mean algorithm in MATLAB. The inference is the darkest region found out from the segmented image represents the suspicious dense tissue.

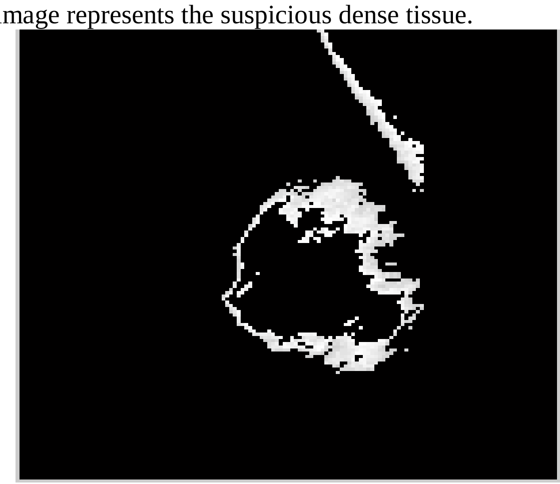


Figure (2.7) Single cluster segmentation image

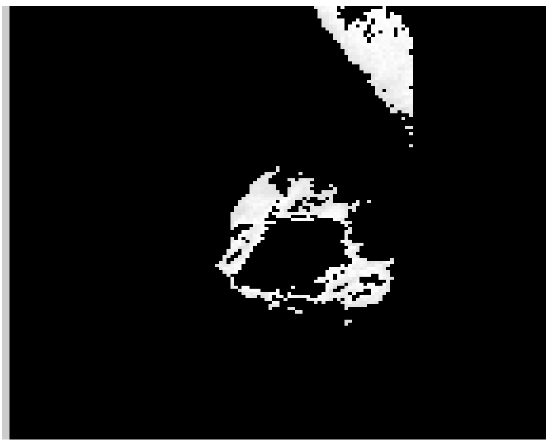


Figure (2.8) Single cluster segmentation image
Original Mammogram Image

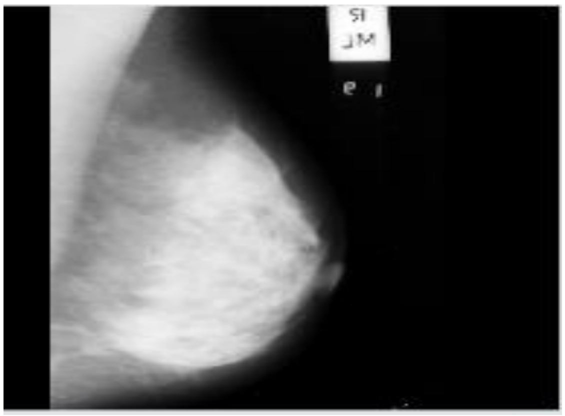


Figure (3.1) Original mammogram image

Modified FCM algorithm Output

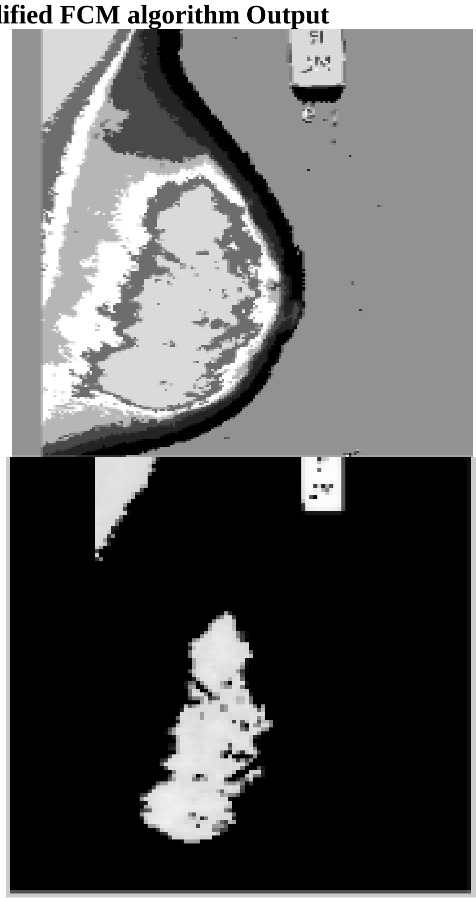


Figure (3.2) Multi and Single Cluster Segmentation Image

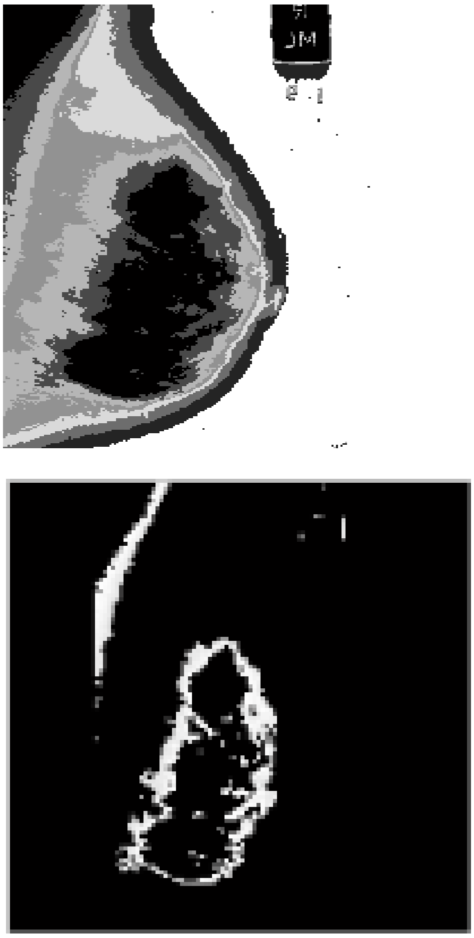


Figure (3.3) Multi and Single Cluster Image of Mammogram

The above image in figure (3.3) represents the mulitcluster image and single cluster image of mammogram respectively after simulation of original mammogram image using Fussy Cluster Mean algorithm (FCM) in MATLAB. The inference is the darkest region found out from the segmented image represents the suspicious dense tissue.

GRAPH

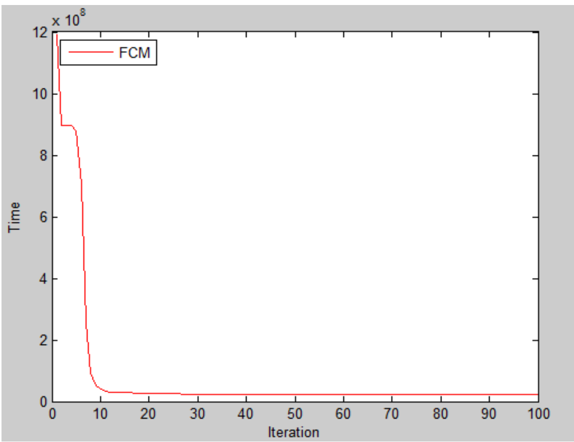


Figure (2.9) Graph shows Iteration Vs Time

Original Mammogram Image

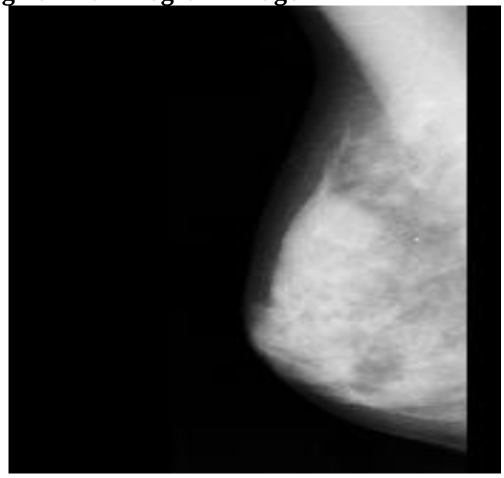


Figure (3.4) Original mammogram image

Modified FCM algorithm Output

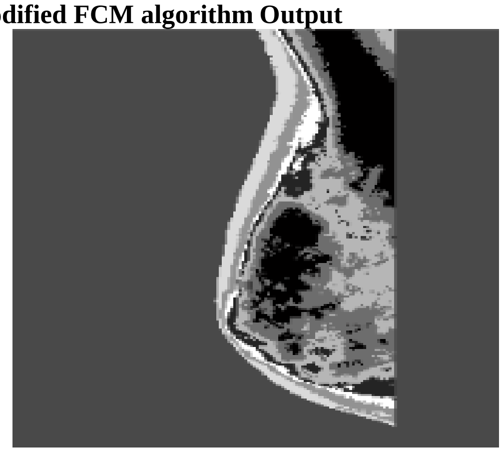


Figure (3.5) Multi Cluster Segmented Image

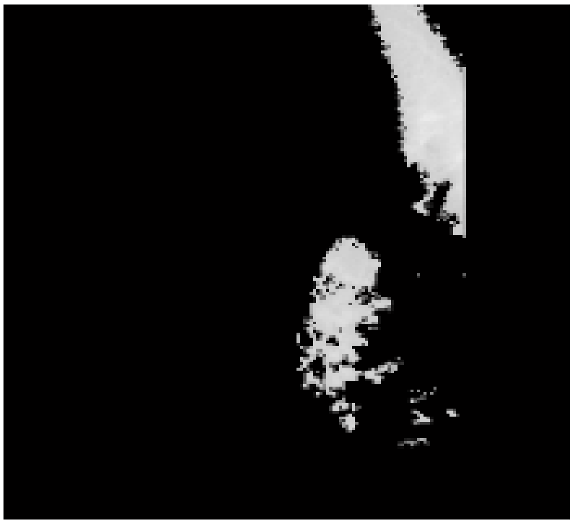


Figure (3.6) Single Cluster Segmented Image

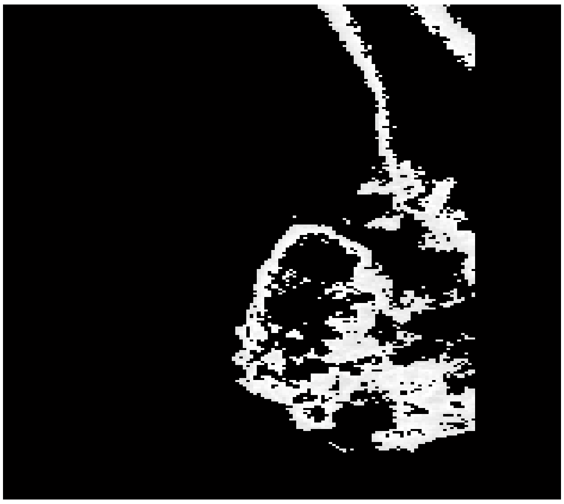


Figure (3.7) Single Cluster Segmented Image

Performance Table

Input data	**Performance result of FCM**
Md001	93.35
Mdoo2	95.42
Md001	93.35
Mdoo2	92.24
Md001	90.67
Mdoo2	95.71
Md001	93.28
Mdoo2	94.63
Md001	96.35
Mdoo2	91.04

VI. CONCLUSION

In the field of Medical image segmentation, one of the most popular methods FCM has been widely used in image segmentation. On the other hand, traditional FCM was always suffers from noisy images. Although many researchers have build up a range of algorithms based on FCM, none of them is perfect. A modified FCM based clustering algorithm is proposed here. In the proposed algorithm, both local and non-local information is incorporated, and a ROM method is exploited to control the trade off between them. The algorithm is formulated by modifying the distance measurement of the standard FCM algorithm to allow the category of a pixel to be influenced by other pixels and to suppress the noise effect during segmentation. We test our algorithm on synthetic square image, simulated and Mammogram images, with different noise levels. We can conclude that our proposed algorithm yields a robust and precise segmentation.

REFERENCES

[1] L. Vibha, G.M. Harshavardhan, K. Pranaw, P. Deepa Shenoy, K.R. Venugopal, L.M. Patnaik, “Classification of mammograms using decision trees”, in: Pro- ceedings of the 10th IEEE International Database Engineering and Applica- tions Symposium (IDEAS’06), 2006, pp. 263–266.

[2] Liu S, Babbs CF, Delp EJ. “Multiresolution detection of spiculated lesions in digital mammograms”. IEEE Transactions on Image Processing 2001;10(6):874–84.

[3] Rashed EA, Ismail IA, Zaki SI. “Multiresolution mammogram analysis in multilevel decomposition”. Pattern Recognition Letters 2007; 28:286–92.

[4] Ferreira CBR, Borges DL. “Analyses of mammogram classification using a wavelet transform decomposition” Pattern Recognition Letters 2003; 24:973–82.

[5] Jawad Nagi, Sameen Abdul Kareem, Farrukh Nagiand Syed Khaleel Ahmed.”Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms”,IEEE EMBS Conference on Biomedical Engineering & Sciences 2010;87-92.

[6] Kai Hu, Xieping Gao and Fei Li. “Detection Of Suspicious Lesions by Adaptive Thresholding Based on Multiresolution Analysis in Mammogram”. IEEE Transactions on Instrumentation and Measurement, Vol.60, PP. 462-472, 2011.

[7] Vikraman Baskaran, Aziz Guergachi, Rajeev K. Bali and Raouf N. G. Naguib, Senior Member, IEEE“Predicting Breast Screening Attendance Using Machine Learning Techniques Member”, IEEE Trans. ON Information Technology In Biomedicine, VOL. 15, NO. 2, p.251-259, MAR. 2011

[8] Maurice Samulski* and Nico Karssemeijer, “Optimizing case-Based detection performance in a multiview CAD system for Mammography”, IEEE Trans. Med Imag, Vol. 30, no. 4, pp.1001-1009, Apr. 2011.

[9] Mousa R, Munib Q, Moussa A. “Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural”, Expert Systems with Applications 2005; 28:713–23

[10] Ferreira CBR, Borges DL. “Analyses of mammogram classification using a wavelet transform decomposition”, Pattern Recognition Letters 2003; 24:973–82.

[11] Xinbo Gao, Ying Wang, Xuelong Li and Dacheng Tao. "On Combining Morphological Component Analysis and Concentric Morphology Model for Mammographic Mass Detection" IEEE Transactions on Information Technology in Biomedicine 2010; 14: 266-273.

[12] A. Jemal, L. X. Clegg, E.Ward, L. A. G. Ries, X.Wu, P.M. Jamison, P.A. Wingo, H.L. Howe, R. N. Anderson, and B. K. Edwards, "Annual report to the nation on the status of cancer, 1975–2001, with a special feature regarding survival," Cancer, vol. 101, no. 1, pp. 3–27, 2004.