AUTOMATIC DETECTION OF BREAST CANCER MASS IN MAMMOGRAMS USING MORPHOLOGICAL OPERATORS AND FUZZY C –MEANS CLUSTERING

[1]S.SAHEB BASHA, [2]DR.K.SATYA PRASAD
[1]Madina Engineering College, Kadapa, (A.P) - India,
[2]Jawaharlal Nehru Technological University, Kakinada, (A.P)- India
Email: sahebshaik@gmail.com, kodati_prasad@yahoo.co.in

ABSTRACT

Breast cancer is one of the major causes of death among women. Small clusters of micro calcifications appearing as collection of white spots on mammograms show an early warning of breast cancer. Primary prevention seems impossible since the causes of this disease is still remain unknown. An improvement of early diagnostic techniques is critical for women's quality of life. Mammography is the main test used for screening and early diagnosis. Early detection performed on X-ray mammography is the key to improve breast cancer prognosis. In order to increase radiologists diagnostic performance, several computer-aided diagnosis (CAD) schemes have been developed to improve the detection of primary signatures of this disease: masses and micro calcifications. Masses are space-occupying lesions, described by their shapes, margins, and denseness properties. A benign neoplasm is smoothly marginated, whereas a malignancy is characterized by an indistinct border that becomes more speculated with time. Because of the slight differences in X-ray attenuation between masses and benign glandular tissue, they appear with low contrast and often very blurred. Micro calcifications are tiny deposits of calcium that appear as small bright spots in the mammogram. This paper presents a research on mammography images using Morphological operators and Fuzzy c – means clustering for cancer tumor mass segmentation. The first step of the cancer signs detection should be a segmentation procedure able to distinguish masses and micro calcifications from background tissue using Morphological operators and finally fuzzy c- means clustering (FCM) algorithm has been implemented for intensity – based segmentation. The proposed technique shows better results

KEY WORDS: *Mammography, Segmentation, Morphology, Fuzzy c- means clustering, Tumors.*

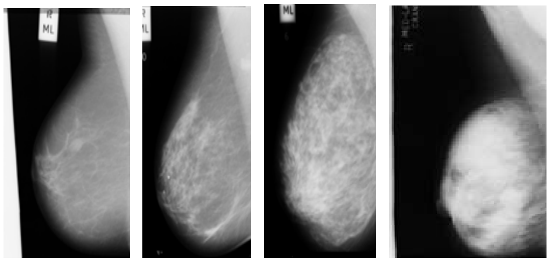
1. INTRODUCTION

Breast cancer has been one of the major causes of death among women since the last decades and it has become an emergency for the healthcare systems of industrialized countries. This disease became a commonest cancer among women. If the cancer can be detected early, the options of treatment and the chances of total recovery will increase. Intra-operative diagnosis of the disease has steadily become more important with respect to the recent introduction of sentinel lymph node biopsy.

A sentinel lymph node is classed as any node that has a direct lymphatic connection to the cancer, and would therefore be the most likely location of cancer spreading from the breast. In surgical studies it has also been indicated that the chance of finding the disease further down the chain of lymph nodes that drain the breast, is significantly smaller if the cancer cannot be found in sentinel lymph node. The American College of Radiology Breast Imaging Reporting and Data System (BIRADS) is becoming a standard on the assessment of mammographic images and uses four categories for density evaluation[9]:

- BIRADS I: the breast is almost entirely fatty,
- BIRADS II: there is some fibro-glandular tissue,
- BIRADS III: the breast is heterogeneously dense,
- BIRADS IV: the breast is extremely dense.

Figure 1 shows example mammograms of each class (the mammograms are extracted from the MIAS database). Note how the internal density of the breasts increases from BIRADS I (left) to BIRADS IV (right). It should be noted that besides density these BIRADS classes also included patterns that can be described as various textures. As such, it seems appropriate to include both aspects in an automatic classification approach.



(a) (b) (c) (d)

Figure 1. Example mammograms, where the breast density increases from (a) BIRADS I to (d) BIRADS IV.

Bovis and Singh [5] extracted features from the global breast and used a combination of classifiers for training and testing the system. Petroudi et al. [6] used texturess to obtain a visual dictionary for breast classification. Arnau Oliver & Xavier Llado [9] classified mammograms using texture information, X.Y.Wang & J.M. Garibaldi [12] implemented a technique for comparing fuzzy and non-fuzzy clustering in diagnosing the cancers, Yao Yao [13] detected the breast cancer mass in mammograms by segmenting and using magnetic resonance imaging, Arianna Mencattini & Roberto Lojacono [7] detected breast cancer from x-ray mammography by means of wavelet and morphological operators, M.Bhattacharya & A.Das [8] segmented the digital mammograms using fuzzy logic based segmentation and mutiresolution technique, A.Laine & W.Yang [9] enhanced contrast for digital mammography using wavelets, M.J.Bottema & G.N Lee [4] helped in diagnosing the breast cancer by implementing an algorithm which can automatatically extract the image features. The techniques so far, utilized either one or two features to detect or to extract breast cancer mass in mammograms..

In this paper we have presented a novel approach to automatically detect the breast cancer mass in mammograms using morphological operators and fuzzy c – means clustering algorithm. The proposed approach utilizes the morphological operations for segmenting the mammograms to distinguish the masses and micro calcifications and fuzzy c – means clustering for intensity based segmentation to extract the features of cancer masses from the mammograms.

This paper is organized as follows: Section 2 introduces the concepts and techniques utilized in the proposed work. Section 3 discusses the segmentation of mammograms using morphological operations & the identification of masses in mammograms using fuzzy c –means clustering algorithm. Section 4 shows Experimental results & Section 5 covers Conclusions.

2.1 X-RAY MAMMOGRAPHY

X-Ray Mammography is commonly used in clinical practice for diagnostic and screening purposes. Screening mammography has been recommended as the most effective method for early detection of breast cancer. Mammography provides high sensitivity on fatty breast and excellent demonstration of micro calcifications [3]; it is highly indicative of an early malignancy. Due to its low cost, it is suitable for mass screening program. Mammography has its limitations. It is less reliable on dense breast of young women or women underwent a surgical intervention in the breast because glandular and scar tissues are as radiopaque as abnormalities. Furthermore, there is low dose X-Ray radiation.

2.2 MRI OF THE BREAST

Magnetic Resonance Imaging is the most attractive alternative to Mammography. MRI is sensitive for detecting some cancers which could be missed by mammography. In addition, MRI can help radiologists and other specialists determine how to treat breast cancer patients by identifying the stage of the disease. It is highly effective to image breast after breast surgery or radiation therapy. To be effective, contrast-enhanced breast MRI is carried out by injecting in the patient’s body of a paramagnetic contrast agent. This method is based on the hypothesis that, after the injection of the agent, abnormalities enhance more than normal tissues due to their increased vascularity, vascular permeability and interstitial spaces [1] MRI forms 3D uncompressed image. It can perform with all women including who are not suitable for mammography, such as young women with dense breast and women with silicone-filled breast implants. Since it uses magnetic fields, MRI has no harmful effects on human bodies. However, MRI takes rather long time to perform and has high cost which is more than ten times greater than mammography. Its low resolution limits its application to very small lesions or micro calcifications.

2.3 IMAGE PROCESSING TOOLS FOR TUMOR DETECTION IN MAMMOGRAMS

Since screening mammography is currently the main test for early detection of breast cancer, a huge number of mammograms need to be examined by a limited number of radiologists, resulting misdiagnoses due to human errors by visual fatigue. In order to improve the diagnostic accuracy and efficiency, computer-aided diagnosis has been introduced in the screening process. Currently, there are several image processing methods proposed for the detection of tumors in mammograms. Various technologies such as fractal analysis [1], discrete wavelet transform and Markov random field have been used. In [2], a multiple circular path convolution neural network architecture has been designed for the analysis of tumor and tumor-like structures. In [3], Petrick *et al*. reported a two-stage adaptive density-weighted contrast enhancement (DWCE) algorithm for tumor detection in mammograms. These studies focus on two types of breast cancer: micro calcifications and masses. The performance of various methods reported in the literature in most cases has been measured on different data sets. The choice of database used by these researchers can influence the performance of their algorithms significantly [4]. Due to time limit, this paper aims to implement some conventional imaging processing techniques to detect mass in mammograms from Digital Database for Screening Mammography (DDSM) of University of South Florida.

3.1 MATHEMATICAL MORPHOLOGY

One of the most rewarding areas of Image processing is Mathematical Morphology. Set theory forms the substratum of Mathematical Morphology. The objects in an image are analogous to the sets in Mathematical Morphology. The geometric relations amidst the points of such sets serve as the crux for the morphological operations [10]. Some of the premier operations that are instrumental for diverse image processing problems include erosion, dilation, opening and closing.

3.2 FUZZY LOGIC

Zadeh introduced the theory of fuzzy logic in the late 1960s. Formerly Lukasiewicz had created the multivalued logic and the fuzzy logic is considered a rediscovery of that approach. Since various real world scenarios could not be represented by two values the fuzzy set approach was introduced. Fuzzy sets, fuzzy membership functions, and fuzzy rules form the elemental components of the fuzzy logic decision making systems. A membership function forms an analogous part of a fuzzy set.

3.3 MORPHOLOGICAL SEGMENTATION

This section details the segmentation of mammograms for identifying the cancer in breasts. The proposed approach utilizes mathematical morphology operations for the segmentation. The morphological operations are applied on the grayscale mammography images to segment the abnormal regions[14,15]. Erosion and dilation are the two elementary operations in Mathematical Morphology. An aggregation of these two represents the rest of the operations [11]. The symbols □, Θ, o, and • , respectively denote the four fundamental binary morphological operations: dilation, erosion, opening, and closing. A function *f* (*x*, *y*) denotes the image, where (*x*, *y*) □ *R*2 or Z2 , or simply *f* , and the function *h*(*x*, *y*) , or *h* will act as the structuring element. The four operations are defined as follows:

Dilation:

$$(f \oplus h)\ (\mathrm{x,y}) = \sup_{(\mathrm{r,a})\in \mathrm{H}} \{\mathrm{x - r, y - s}) + \mathrm{h(r,s)}\}$$

Erosion:

$$(f \Theta h)\ (\mathrm{x,y}) = \inf_{(\mathrm{r,a})\in \mathrm{H}} \{f(x+r, y+s) - h(r,s)\}$$

Opening: $f \circ h = (\mathrm{f\Theta h}) \oplus \mathrm{h}$

Closing: $f \bullet h = (f \oplus h)\Theta h$

Where sup {} and inf {} denote the supremum and infirmum operation, respectively, and *H* □ *R*2 or *Z* 2 is the support of *h*(*x*, *y*) In our approach, we have used binary open morphological operation for the segmentation of mammography images. Erosion and Dilation are merged to form a powerful operator called Opening. Commonly this operator gentles the frontiers of an image, breaches narrow Isthmuses and annihilates thin Protrusions. Opening operation is obtained by doing Dilation on Eroded Image. Generally, objects that are adjacent are spaced, objects that are adjoined are detached and the holes within the objects are enlarged by Opening. The grey-scale mammographic images are segmented using binary open morphological operation. The segmented regions are further processed using fuzzy c- means algorithm.

3.4 FUZZY C-MEANS ALGORITHM

The FCM algorithm, also known as Fuzzy ISODATA, is one of the most frequently used methods in pattern recognition[12]. It is based on the minimization of the objective function (1) to achieve a good classification. J is a squared error clustering criterion, and solutions of minimization of (1) are least-squared error stationary points of J.

$$J_m = \sum_{i=1}^{N}\sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2 , \quad 1 \le m < \infty \qquad \text{--- } 1$$

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:

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

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

This iteration will stop when $\max_{ij}\left\{\left| u_{ij}^{(k+1)} - u_{ij}^{(k)} \right|\right\} < \varepsilon$, 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:

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

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m \cdot x_i}{\sum_{i=1}^{N} u_{ij}^m}$$

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

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

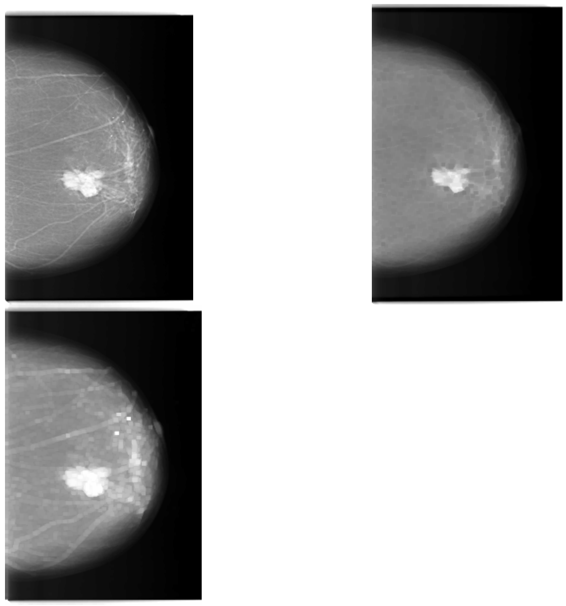
4. *If $\| U^{(k+1)} - U^{(k)} \| < \varepsilon$ then STOP; otherwise return to step 2*

Clustering is a multivariate analysis technique widely adopted in medical diagnosis studies and pattern recognition areas[16]. By examining the underlying structure of a dataset, cluster analysis aims to class data into separate groups according to their characteristics. The clustering is performed such that spectra held within a cluster are as similar as possible, and those found in opposing clusters as dissimilar as possible. Thus different cells types found within biological tissue can be separated and characterized.

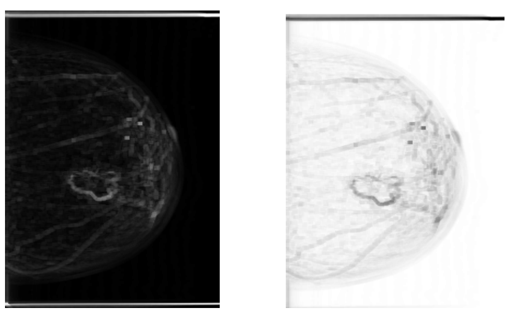
4. RESULTS & DISCUSSIONS:

This section details the results of the automatic detection of breast cancer mass in mammograms using morphological operators and fuzzy c – means clustering.

In this analysis, the first procedure is determining the seed regions. When dealing with mammograms, it is known that pixels of tumor regions tend to have maximum allowable digital value. Based on this information, morphological operators are used such as Dilation and Erosion to detect the possible clusters which contain masses[10,11]. Image features are then extracted to remove those clusters that belong to background or normal tissue as a first cut. Features used here include cluster area and eccentricity. The fuzzy c-means clustering algorithm is used as a segmentation strategy to function as better classifier & aims to class data into separate groups according to their characteristics.

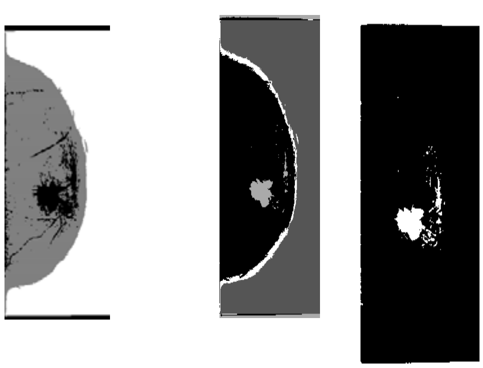


Original Image   Dilated Image   Eroded Image



Morph- Gradient   Visualization

Figure 2. Morphological results



(a)   (b)   (c)

Figure 3. Results from three separate runs with FCM

FCM is helpful in early stage of clustering in medical diagnosis. The cancerous mode is separated from the fatty breast regions using FCM. As the number of clusters increases, more and more information is obtained about the tissue which cannot be identified by pathologists.

5. CONCLUSIONS:

Beast cancer is one of the major causes of death among women. So early diagnosis through regular screening and timely treatment has been shown to prevent cancer. In this paper we have presented a novel approach to identify the presence of breast cancer mass in mammograms. The proposed work utilizes morphological operators for segmentation and fuzzy c- means clustering for clear identification of clusters.

The morphological operations and FCM is a new approach, using this we have successfully detected the breast cancer masses in mammograms. The results indicates that this system can facilitate the doctor to detect breast cancer in the early stage of diagnosis process.

REFERENCES:

[1] Lei Zheng and Andrew K. Chan, "An artificial intelligent algorithm for tumor detection in screening mammogram," *IEEE transactions on medical imaging*, vol. 20, no. 7, July, 2001.

[2] Huai Li, Shih-Chung B.Lo, Yue Wang, Lisa Kinnand, and Matthew T. Freedman, "A multiple circular path convolution neural network system for detection of mammographic masses," *IEEE transactions on medical imaging,* vol. 21, no, 2, February, 2002.

[3] H.P.Chan, N.Petrick, and B.Sahiner, "Computer-aided breast cancer diagnosis," Artificial intelligence techniques in breast cancer diagnosis and prognosis, Series in machine perception and artificial intelligence,Vol.39, World Scientific Publishing Co.Pte.Ltd,2000, pp.179-264.

[4] M.J. Bottema, G.N.Lee and S.Lu, "Automatic image feature extraction for diagnosis and prognosis of breast cancer," Artificial intelligence techniques in breast cancer diagnosis and prognosis, Series in machine perception and artificial intelligence, Vol.39, World Scientific Publishing Co.Pte.Ltd, 2000, pp. 17-54.

[5] K. Bovis & S. Singh. "Classification of mammographic breast density using a combined classifier paradigm." In

Med. Image Underst. Anal., pp. 177–180. 2002.

[6] S. Petroudi, T. Kadir & M. Brady. "Automatic classification of mammography parenchyma patterns: A statistical approach." In *IEEE Conf. Eng. Med. Biol. Soc.*, volume 1, pp. 798–801. 2003.

[7] Arianna Mencattini & Roberto Lojacono " Breast cancer segmentation by means of wavelet analysis and morphological operators.

[8] M.Bhattacharya & A.Das, " Fuzzy logic based segmentation of Micro calcification in Breast using Digital Mammograms considering Mutiresolution.

[9] Arnau Oliver & Xavier Llado, " Classifying Mammograms using texture information".

[10] S..Saheb Baha and Dr. K.Satya Prasad, "Morphological image processing in Bio-Medical Application", Proceedings of PCEA-IFTOMM- International conference – PICA-2006, held on 11th to 14th of July at Nagapur and NCBME-2006 – National conference on Bio-Medical Engineering held on 28th to 29th of March at Mumbai.

[11] S.Saheb Baha and Dr. K.Satya Prasad, "Automatic detection of Hard Exudates in Diabetic Retinopathy using Morphological Segmentation and Fuzzy Logic" in IJCSNS International Journal of Computer Science and Network Security, VOL.8 No.12, December 2008- P- 211 – 218

[12] X.Y.Wang & J.M.Garibaldi, " A comparison of fuzzy & non- fuzzy clusterings techniques in cancer diagnosis".

[13] Yao Yao, " Segmentation of Breast cancer mass in mammograms & detection using magnetic resonance imaging".

[14] DIGTAL IMAGE PROCESSING
- Rafael C. Gonzalez
- Richard E. Woods
ADDISON-WESLEY
An imprint of Pearson Education, 1st Edition.

[15] DIGITAL IMAGE PROCESSING
-Nick Efford
ADDISON-WESLEY
An imprint of Pearson Education, 1st Edition.

[16] Fuzzy Logic with Engineering Applications, International edition 1997
-Timothy J. Ross
McGRAW-HILL International Edition.

BIOGRAPHY:

S.Saheb Basha received B.E degree in Electronics & Communication Engineering from Gulbarga University, Gulbarga, Karnataka, India in 1992 and M.Tech degree in Digital Systems & Computer Electronics from JNTU College of Engineering, Anantapur, Andhra Pradesh, India in 2002 and currently pursuing his Ph.D under the supervision of Prof. K. Satya Prasad garu at JNT University, Kakinada. He worked as Assistant Engineer in Production Department from 1992 t0 1995 in Bull Power Systems Ltd, Hyderabad, India and later as Senior Engineer in Design & Development Section from 1995 to 1998 in Bull Power Systems Ltd, Hyderabad, India. He joined in Madina Engineering College, Kadapa, Andhra Pradesh as Assistant Professor in Electronics & Communication Department in 1999 and presently he is in the capacity of Professor & Vice- Principal. He published 3 technical papers in National & International conferences & Journals. He is a life member of ISTE & IE.

Dr. K. Satya Prasad received B Tech. degree in Electronics and Communication Engineering from JNTU college of Engineering, Anantapur, Andhra Pradesh, India in 1977 and M. E. degree in Communication Systems from Guindy college of Engg., Madras University, Chennai, India in 1979 and Ph.D from Indian Institute of Technology, Madras in 1989. He started his teaching carrier as Teaching Assistant at Regional Engineering College, Warangal in 1979. He joined JNT University, Hyderabad as Lecturer in 1980 and served in different constituent college's viz., Kakinada, Hyderabad and Anantapur and at different capacities viz., Associate Professor,

Professor, Head of the Department, Vice Principal and Principal. Presently he is working as Director of Evaluation in JNT University, Kakinada, A.P. He has published more than 50 technical papers in different National & International conferences and Journals and Authored one Text book. He has guided 4 Ph.D scholars and at present 12 scholars are working with him. His areas of Research include Communications, Signal Processing, Image Processing, Speech Processing, Neural Networks & Adhoc wireless networks etc. Dr Prasad is a Fellow member of various professional bodies like IETE, IE, and ISTE.