

A FCM BASED APPROACH FOR AUTOMATED SEGMENTATION OF BREAST MASSES IN MAMMOGRAMS

S.Vijayalakshmi,Prabha S.Nair and S.Nithyalakshmi

School of Computing Science and Engineering, Galgotias University, Greater Noida,
Uttar Pradesh

ABSTRACT

In this paper, we present an automated method to segment the breast masses in Mammograms that can be acquired in the routine clinical setting. Breast cancer is the most common deadly disease among the women. But early detection can save their life. Mass detection is an active research area for the past few years but till now no accurate method has been devised. In this paper an automated method to segment the breast masses has been proposed based on the Fuzzy C-Means Clustering method. The results show that this method can better assist as a tool in detection of breast cancer.

KEYWORDS

Breast cancer, Mammograms, Fuzzy C-Means, clustering and masses. *Mammography, Segmentation, Morphology, Fuzzy c- means clustering, Tumors.*

1.INTRODUCTION

Segmentation of medical images is the task of partitioning the data into contiguous regions representing individual anatomical objects. It is a prerequisite for further investigations in many computer-assisted medical applications, e.g. individual therapy planning and evaluation, diagnosis, simulation and image guided surgery.

In the last two decades, breast cancer has been the second leading cause of cancer deaths among women in some countries, following lung cancer [1]. An estimated 226,870 new cases of invasive breast cancer are expected to occur among women in the US during 2012; about 2,190 new cases are expected in men. Out of the 226,870 cases, 39,920 breast cancer deaths are expected [2]. Survival from breast cancer is directly related to its stage at diagnosis. The earlier the detection, the higher the chances of successful treatments [3]. Breast image analysis can be performed using X-rays, magnetic resonance or ultrasound. So far the most effective and economical breast imaging modality has been the X-ray mammography due to its simplicity, portability and cost effectiveness [4]. It is based on the difference in absorption of X-rays between the various tissue components of the breast such as fat, tumor tissue, and calcifications. Mammography has high sensitivity and specificity, even small tumors and micro calcifications can be detected on mammograms. Projection of the breast can be made from different angles [5].

The interpretation and analysis of medical images represent an important and exciting part of computer vision and pattern recognition. Developing a computer aided diagnosis system for cancer diseases, such as breast cancer, to assist physicians in hospitals is becoming of high importance and priority for many researchers and clinical centers. An extensive work had been made in the field of digital mammography, especially in the last decade. These methods varied in the diagnosis of masses in mammograms. Here are some of the most common approaches used in recent researches on this domain.

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.

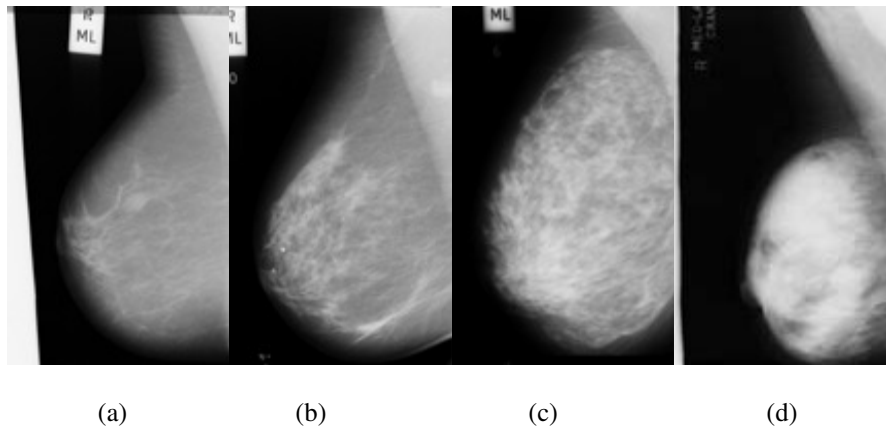


Figure1.Example mammograms, where the breast density increases from (a) BIRADS I to (d) BIRADS IV.

Bovis and Singh [6] extracted features from the global breast and used a combination of classifiers for training and testing the system. Petroudi et al. [7] used textures to obtain a visual dictionary for breast classification. Arnau Oliver & Xavier Llado [8] classified mammograms using texture information, X.Y.Wang & J.M. Garibaldi [9] implemented a technique for

comparing fuzzy and non-fuzzy clustering in diagnosing the cancers, Yao Yao [10] detected the breast cancer mass in mammograms by segmenting and using magnetic resonance imaging, Arianna Mencattini & Roberto Lojaco [11] detected breast cancer from x-ray mammography by means of wavelet and morphological operators, M.Bhattacharya & A.Das [12] segmented the digital mammograms using fuzzy logic based segmentation and multi resolution technique, W.Yang [9] enhanced contrast for digital mammography using wavelets, M.J.Bottema & G.N Lee [13] helped in diagnosing the breast cancer by implementing an algorithm which can automatically 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 presents the materials and methods utilized in the proposed work. In Section 3 the results and discussions are given . Section 4 covers Conclusions.

2.MATERIALS AND METHODS

2.1.Materials Used

Test images were taken from the Mammographic Image Analysis Society (MIAS) database. MIAS is an organization of UK research groups interested in the understanding of mammograms. The database is arranged in pairs of films, where each pair represents the left and right mammograms of a single patient. All the images have the size of 1024 × 1024 pixels with a spatial resolution of 200µm and a gray scale level of 8 bits. The database also provide information about the characteristics of background tissue, class of abnormality and severity of abnormality. The (x, y) coordinates of the center of lesions are also provided by the database, along with the approximate radius (in pixels) of a circle enclosing the abnormality [14]. With all these information provided, the MIAS database is considered as one of the most common sources of test mammograms that can be used in scientific researches and CAD systems

2.2.Methods Used

2.2.1.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 [15-16]; 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 radioopaque as abnormalities. Furthermore, there is low dose X-Ray radiation.

2.2.2.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 [17] 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.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 . Some of the premier operations that are instrumental for diverse image processing problems include erosion, dilation, opening and closing which are used in our proposed method and its flowchart is shown in figure 2.

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.

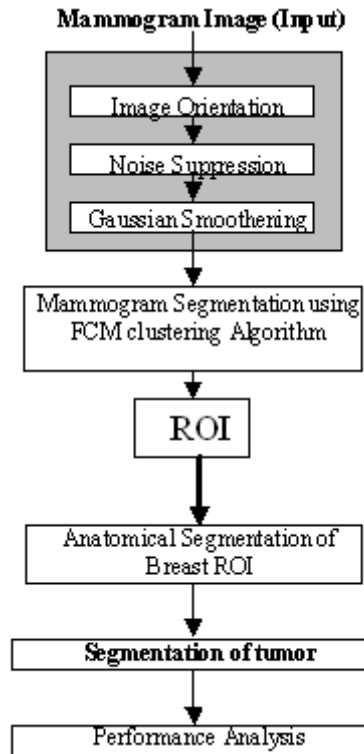


Figure 2. Flowchart of the proposed method

2.3.1.Mammogram Preprocessing

The mammogram image is transformed, so that the chest wall location, i.e., the side of the image containing the pectoral muscle, is on the upper left corner of the image. In order to determine the chest wall location, the decreasing pixel intensity of the breast tissue near the skin-air interface is used. This tissue is estimated by employing the Minimum cross-entropy threshold selection technique [18] twice in the original image. After determining the chest wall location we determine the top of the image. We extract the vertical centroid of the image and then we assume that the asymmetric region closest to the right side of the vertical centroid is the tip of the breast. The image is flipped horizontally if needed to place the asymmetric region below the vertical centroid, resulting in an image that is right way up. There are different types of noises, which appear in miniMIAS images. High intensity noise is characterized by high values of optical densities, such as labels or scanning artifacts. Tape artifacts are markings left by tapes, or other shadows presenting themselves as horizontal running strips. Such noise must be replaced by black pixels. The Gaussian smoothing operator is a 2-D convolution operator that is used to 'blur' images and remove detail and noise. It uses a kernel that represents the shape of a Gaussian ('bell-shaped') hump. In 2-D, an isotropic (i.e. circularly symmetric) Gaussian has the form:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{1}$$

The idea of Gaussian smoothing is to use this 2-D distribution as a 'point-spread' function, and this is achieved by convolution. Since the image is stored as a collection of discrete pixels we need to produce a discrete approximation to the Gaussian function before we can perform the convolution. In theory, the Gaussian distribution is non-zero everywhere, which would require an infinitely large convolution kernel, but in practice it is effectively zero more than about three standard deviations from the mean, and so we can truncate the kernel at this point. Once we obtain a suitable kernel, then the Gaussian smoothing can be performed using standard convolution methods.

The **fuzzy C-means** (FCM) algorithm follows the same principles as the **K-means** algorithm in that it compares the RGB value of every pixel with the value of the cluster center and is shown in figure 3. The main difference is that instead of making a hard decision about which cluster the pixel should belong to, it assigns a value between 0 and 1 describing "how much this pixel belongs to that cluster" for each cluster. **Fuzzy** rule states that the sum of the membership value of a pixel to all clusters must be 1. The higher the membership value, the more likely that pixel is to belong to that cluster. The FCM clustering is obtained by minimizing an objective function shown in equation (2):

$$J_m = \sum_{i=1}^N \sum_{j=1}^C u_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m < \infty \tag{2}$$

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 i th of d -dimensional measured data, c_j is the d -dimension center of the cluster, and $\|\cdot\|$ is any norm expressing the similarity between any measured data and the center.

The calculation of the centroid of the k th cluster is achieved using equation (3):

$$v_k = \frac{\sum_{i=1}^n \mu_{ik}^m p_i}{\sum_{i=1}^n \mu_{ik}^m} \tag{3}$$

The fuzzy membership table is calculated using the original equation (4):

$$\mu_{ik} = \frac{1}{\sum_{l=1}^c \left(\frac{|p_i - v_k|}{|p_i - v_l|} \right)^{\frac{2}{m-1}}} \tag{4}$$

This algorithm has been extended for clustering of color images in the RGB color space. Hence, the computation given in equation (2) to compute the Euclidean distance between the values p_i and v_k is modified to incorporate RGB colors, and is shown in equation (5):

$$|p_i - v_k| = \sqrt{\sum_{i=1}^n (p_{iR} - v_{kR})^2 + (p_{iG} - v_{kG})^2 + (p_{iB} - v_{kB})^2} \quad (5)$$

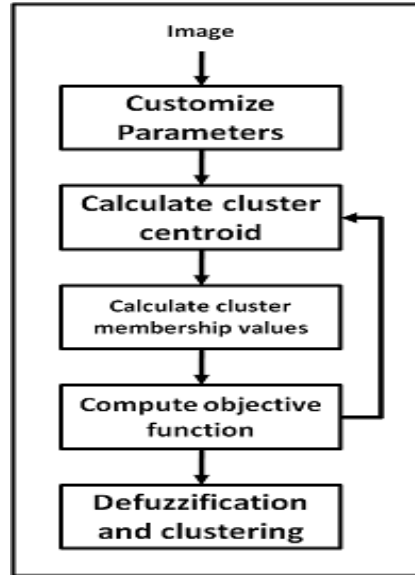


Figure 3. Flowchart for FCM

As mentioned earlier, this is an iterative process. The pseudo-code is as follows:

Step 1: Set the number of clusters, the fuzzy parameter (a constant > 1), and the stopping condition

Step 2: Initialize the fuzzy partition matrix

Step 3: Set the loop counter k = 0

Step 4: Calculate the cluster centroids, calculate the objective value J

Step 5: For each pixel, for each cluster, compute the membership values in the matrix

Step 6: If the value of J between consecutive iterations is less than the stopping condition, then stop; otherwise,

set k=k+1 and go to step 4

Step 7: Defuzzification and segmentation

3.RESULTS AND DISCUSSION

The original image as well as the segmented results are presented in table 1. The accuracy of this technique was evaluated through quantitative measures derived through the comparison of each segmented mammogram representing an abnormal mass, as “mask” with its corresponding “gold standard” and is shown in table 2. The gold standard is generated by manually segmenting the breast region representing an abnormal mass from each mammogram. The boundary of the regions is then manually traced to extract the abnormal breast region to generate a ground truth (GT) image.

Table 1: Original Mammogram and the detected mass

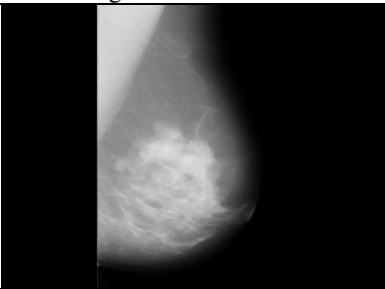
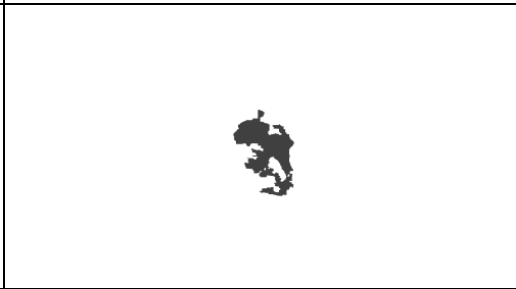
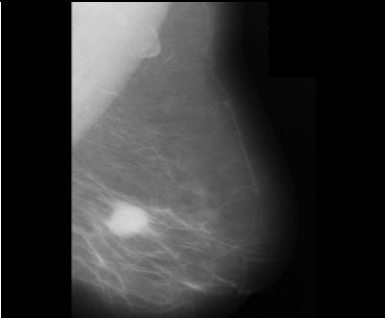

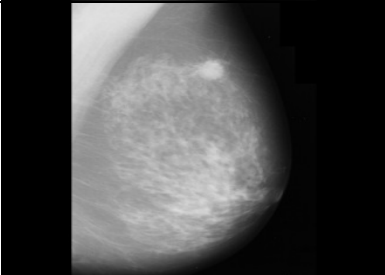
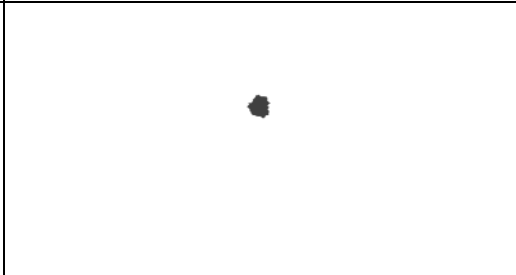
Original Mammogram	Mass detection
	
	
	

Table 2. Performance analysis

Performance measures	Result
Accuracy (Percentage agreement)	0.9985
Dice similarity coefficient (DSC)	0.8794
Error rate	0.0015
TNSensitivity (Percentage of Correct Estimation)	0.9419
Specificity (True Negative Fraction/Rate)	0.9301
False Positive Fraction/Rate	0.0697
Under estimation fraction (UEF)	0.0007
Over estimation fraction (OEF)	0.0004

4.CONCLUSION

In this paper we have presented a new approach to find out the presence of cancer mass in mammograms. The proposed work utilizes the combination of morphological operators for segmentation and fuzzy c- means clustering for clear identification of clusters. This is a new approach, and using this we have successfully detected the breast cancer masses in mammograms. The results indicate that this system can facilitate the physician in the identification of breast cancer in the early stage of diagnosis process itself.

REFERENCES

- [1] F. Moayedi, R. Boostani, A. R. Kazemi, S. Katebi And E. Dashti; "Subclass Fuzzy-Svm Classifier As An Efficient Method To Enhance The Mass Detection In Mammograms", Iranian Journal Of Fuzzy Systems, Vol. 7, No. 1, Pp. 15-31, 2010.
- [2] CONE HEALTH, the network for exceptional care, Regional Cancer Statistics January 2012.
- [3] VALLIAPPAN Raman, PUTRA Sumari and MANDAVA Rajeswari; "A Theoretical Methodology and Prototype Implementation for Detection Segmentation Classification of Digital Mammogram Tumor by Machine Learning and Problem Solving Approach", International Journal of Computer Science Issues (IJCSI), Vol. 7, Issue 5, September 2010.
- [4] H.B.Kekre, Tanuja K. Sarode, Saylee M. Gharge; "Tumor Detection in Mammography Images using Vector Quantization Technique", International Journal of Intelligent Information Technology Application, Vol. 2, No. 5, pp. 237-242, 2009.
- [5] Valliappan Raman, Putra Sumari, H.H.Then, and Saleh Ali K. Al-Omari; "Review on Mammogram Mass Detection by Machine Learning Techniques", International Journal of Computer and Electrical Engineering, Vol. 3, No. 6, December 2011.
- [6] K. Bovis & S. Singh. "Classification of mammographic breast density using a combined classifier paradigm." InMed. Image Underst. Anal., pp. 177–180. 2002.
- [7] 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.
- [8] Arnau Oliver & Xavier Llado, "Classifying Mammograms using texture information".
- [9] X.Y.Wang & J.M.Garibaldi, " A comparison of fuzzy & non- fuzzy clusterings techniques in cancer diagnosis".
- [10] Yao Yao, " Segmentation of Breast cancer mass in mammograms & detection using magnetic resonance imaging".
- [11] Arianna Mencattini & Roberto Lojacono " Breast cancer segmentation by means of wavelet analysis and morphological operators.
- [12] M.Bhattacharya & A.Das, " Fuzzy logic based segmentation of Micro calcification in Breast using Digital Mammograms considering Mutiresolution.
- [13] 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.
- [14] The MAMMOGRAPHIC IMAGE ANALYSIS SOCIETY (MIAS), mini mammographic image database guide.
- [15] 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.
- [16] 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

- [17] 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.
- [18] A.D.Brink and N.E.Pendock, "Minimum Cross-entropy Threshold Selection, Pattern Recognition", 29(1): 179-188, 1996.

Authors

Vijayalakshmi S was born in the year 1975. She received the B.Sc. degree in Computer Science from Bharathidasan University, Trichirapalli, India in 1995, the MCA degree from the same University in 1998 and the M.phil. degree from the same University in the year 2006. She received her doctorate in 2013. She has been working as an Assistant Professor (grade-III), Galgotias University, Greater Noida, Uttar Pradesh, India she has 17 years of teaching experience and 10 years of research experience. She has published many papers in the area of image processing especially in medical imaging.



Prabha Shreeraj Nair was born in Chatishgarch, India, in 1973. She received the B.E. Degree in Computer Technology from Nagpur University, in 1996 and Masters in Computer Science and Engineering from Kakatiya University, in 2007. She has been working as an Assistant Professor, Galgotias University, Greater Noida, Uttar Pradesh, India she has 18 years of teaching experience. She is engaged in research on Software Engineering, Augmented realit



Nithyalaskshmi pursuing B.E (CSE) Degree in Veltech Multitech SR&DR engineering college .

