Detection of Micro-calcification Clusters from Digital Mammogram using Medical Image Processing

Jenif D Souza W S

Assistant Professor, Department of Computer Science and Engineering, St. Joseph's College of Engineering, Chennai - 119

Abstract – **Breast cancer is the most common and leading cancers among women. Digital Mammography technique is currently used for early detection of breast cancer in women. The signs of Breast cancer are Calcification, Masses in breast, Architectural distortion and Bilateral Asymmetry. Primary sign of breast cancer is the Micro-calcification present in the breast. Earlier the detection of calcification would be better, the detection of breast cancer in early stage. But, identification of Micro-calcifications (MCs) is very difficult and time consuming for radiologists and also classification may leads to misinterpretation. Because of this, noncancerous tumor is misinterpreted as cancer which result in false positive test results and while in some cases, cancer may not be detected which result in false negative test results. In order to reduce the rate of such misinterpretation and to effectively detect Micro-calcification, a CAD based system for detection is proposed. This proposed approach is more effective and reliable for early detection of the breast cancers. The proposed method also evaluates large number of cases from a digital database for screening mammography (DDSM). This system consists of four stages such as preprocessing, segmentaton of ROI, feature extraction and finally classification. Because of this proposed system the rate of misinterpretation has reduced.**

Keywords: **Breast cancer, Computer-aided detection (CAD), Digital Mammogram, Micro-calcification.**

I. INTRODUCTION

Breast Cancer refers to the multiplication of cells in a particular location of the body. It has risk up to 12% and a risk of death of up to 5%. The signs of breast cancer are Micro-calcifications, masses present in the breast, Architectural distortions and bilateral asymmetry [1]. Important signs of breast cancer are clusters of Micro-calcifications present in the breast. Currently, mammography is the most effective tool for early detection of breast cancer. Here, Digital Mammography is used. It is a specialized form of mammography that uses digital film and computers instead of using x-ray film to examine breast tissue for breast cancer [2]. Fig 1 shows the output of digital mammogram. In mammography, due to the human factor in the screening process high degree of error is obtained. Because of this Biopsies are performed resulting in unwanted expenditure and anxiety to the patient. The error rate detected is between 10%–30% in screening studies [3].

Double reading of screening mammograms provides better results. But, manpower is a major drawback in this method [4]. In order to over this problem, many computer-aided detection and diagnosis (CAD) systems have been developed in recently [5]. For performance evaluation, a great number of mammograms image that are obtained from Digital Database of Screening Mammography (DDSM) are used. By this proposed system we can detect Micro-calcification of breast and we can also reduce the rate of false positive test results.
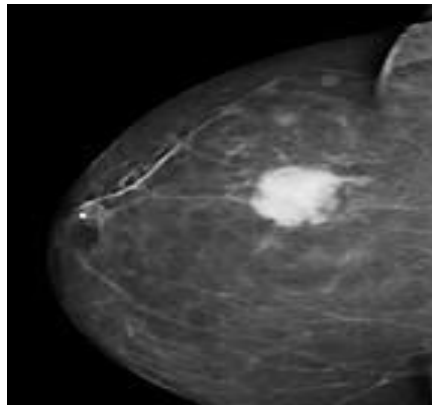

Fig 1. Shows the output image of digital mammogram

The rest of this paper is organized as follows. Section II reviews some related work based on CAD system. Section III describes the proposed approach and its architecture. Section IV presents the experimental results of the proposed work. Finally, Section V draws a conclusion of the system.

II.   RELATED WORKS

Many methods for CAD of Micro-calcifications in mammograms have been proposed. Although many types of techniques for CAD system for detection of breast cancer have been developed, some of which have achieved a high sensitivity and specificity for specific abnormalities, the automatic and accurate classification of Micro-calcification abnormalities as malignant or benign remains a challenge, most of the existing approaches have their own specific disadvantages. K-nearest neighbor and fuzzy mean are used, it has accuracy about 70% to 82% but it a very complex data space [6]. In another approach Multi-Wavelet and hard threshold is used it produce best PSNR values. But the affected area may also removed by noise removal [7].

In another approach vector quantization is used for clustering and texture analysis. It produces 90% accurate but the major disadvantage is absence of under and over segmentation [8]. In another method unsharp masking and cripening is used, it is better capable of enhancing and highlighting the abnormalities in details but it is computationally cheaper and produce better results . But the execution time is lower when compared to counterpart which was FCM with Euclidian distance [9]. In another approach wavelet based segmentation used to detect suspicious tumor region exactly. But the major disadvantage is this method is unsuccessful in case of benign

[10]. Wavelet analysis and genetic algorithm is used in this method, it overcomes the limitation of analyzing CC and MLO views and produce 95% sensitivity [11].

Morphological operation and wavelet transform is used in this method, it has 92. % of true MC clusters per image and has 0.08% false MC cluster per image [12]. In addition to this Wavelet and GLCM MIAS are used which is accurate when compared to other method but 14 features ofthe image are extracted. It has 94.2% accuracy, 98.8% specificity and 97.4% sensitivity [13]. An iterative modified watershed algorithm, GLCM and SVM is used in which Speckle Noise Removal and EM algorithm is used for enhancing the image. This algorithm produces 98% accuracy, 100% specificity and 97.5% sensitivity [14].

III.   PROPOSED SYSTEM

This proposed system for early detection of Micro-calcification clusters includes stages such as preprocessing, segmentation, feature extraction, classification and finally localization of affected region of the given digital mammogram image is detected. This is implemented by using Mat lab codes. Work is done on a computer with digital mammogram and data is stored for analysis. Figure 2 shows the proposed architecture.

Image Acquisition is from first screening test always included Medio Lateral Oblique (MLO) and Cranio Caudal (CC) views. This MLO and CC views of both left and right breast is done [15]. Computer aided detection (CAD) is available for Full Field Digital Mammogram (FFDM). The proposed method also evaluates large number of cases from a digital database for screening mammography (DDSM).
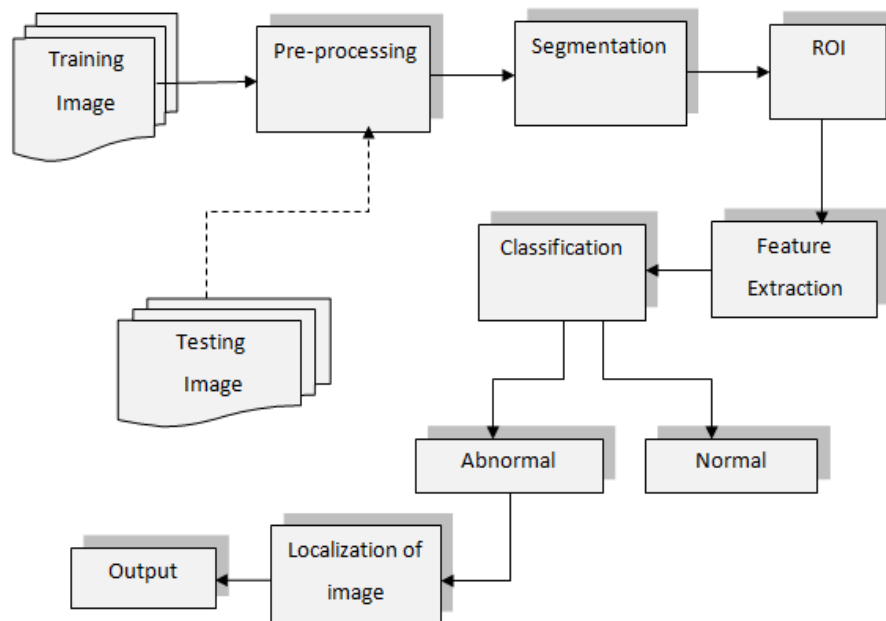

Fig. 2: Architecture of Proposed system

A. PREPROCESSING

Preprocessing of mammograms is done to improve the contrast of mammogram images. The objective of preprocessing is to improve the image quality, removing the unwanted noises and unwanted background of the mammogram images. This includes three stages such as image enhancement, noise removal and Label removal. Image enhancement is done by using Histogram equalization method in this method contrast of the image is adjusted. Then noise filtering is done by using median filter, it is used to remove the unwanted noise and high frequency components without disturbing the edges. After noise removal label is done by using bounding box algorithm which removes unwanted label from the image. Preprocessing includes the following steps.

Step 1: HISTEQ enhances the contrast of images by transforming the values in an intensity image, or the values of an indexed image, so that the histogram of the output image approximately matches a specified histogram. This allows for areas of lower local contrast to gain a higher contrast. Figure 3 (a) shows enhanced image for the input mammogram image.

Step 2: MEDFILT2(A) performs median filtering of the matrix A using the default 3-by-3 neighborhood. MEDFILT2 pads the image with zeros on the edges, so the median values for the points within [M N]/2 of the edges may appear distorted. IMNOISE add noise to image. J = IMNOISE(I, 'salt & pepper', D) adds "salt and pepper" noise to the image I, where D is the noise density. The default for D is 0.05. Figure 3 (b) shows image with salt and pepper noise and noise removed images.

Step 3: Labels in the image such as L-CC, R-CC, L-MLO, and R-MLO are marked on the corner of the image. Those labels are removed by using bounding box algorithm. REGIONPROPS measure properties of image regions. RECTANGLE('Position', [x y w h]) adds a rectangle at the specified position.

B. SEGMENTATION

In Segmentation region of interest (ROI) is recognized. Segmentation is the process which is used to detect the various abnormal areas present in the image. It is the second main stage of the proposed system. Segmenting the breast region from surrounding regions is carried out in order to make it easier to extract the suspicious tissues. During this process the mammogram image is segmented into several other regions. Segmentation is done to detect the possible clusters that contain Micro-calcification. ROI is the sum of the areas of all interested shapes. ROI is segmented and for that segmented region edge detection is done by using canny edge detector. Image segmentation includes following steps.

Step 1: The image is subjected to segmentation by setting a threshold value. Image segmentation is the process of partitioning a digital image into multiple segments. Figure 3 (c) shows segmented ROI image. Finally the Region of Interest is extracted from the image.

Step 2: It takes a gray scale segmented image as input, and produces an image showing the positions of tracked intensity discontinuities as output. EDGE takes intensity or a input I, and returns a binary image BW of the same size as the input image. It displays 1's if the function finds edges in that I and 0's otherwise. The Canny method finds edges by checking the local maxima of the gradient of I. By using the derivative of a Gaussian filter gradient is calculated. The method

uses two thresholds, to detect strong edges and weak edges, and includes the weak edges in the output only if they are connected to strong edges. Figure 3 (d) shows the edge detected image.

C. FEATURE EXTRACTION

Feature extraction is done to remove the cluster that belongs to background or normal tissues. Features are extracted by using GLCM (Gray Level Co-occurrence Matrix) algorithm. This GLCM extracts 26 features from the image. The co-occurrence matrix is a tool for extracting second-order texture information from images given. In this study, fifteen features were derived from each GLCM. The features are angular second moment, entropy, contrast, local homogeneity, correlation, shade, prominence, variance, sum average, sum entropy, difference entropy, sum variance and difference variance. The mean and range of these were calculated, comprising a total of 30 GLCM features.

GLCM can be an m x n x p array of gray-level co-occurrence matrices. Each GLCM is normalized so that its sum is one. Properties can be a list of strings, a cell array containing strings, or a space separated string. In the proposed work feature such as homogeneity, correlation, intensity, energy, contrast and entropy are extracted for the segmented image. Figure 3 (e) shows the feature extracted image for the segmented region.

D. CLASSIFICATION

Classification is the process in which many characters are extracted and classified into the various abnormalities such as malignant or benign. Many different types of classifiers can be used in this step for classification of the abnormal areas. Classification is done by using SVM algorithm and then edge is detected to localize the affected by using prewitt edge detector. Classification includes two stages.

Step 1: Classification is done by using SVM (Support Vector Machine) algorithm. It non-linearly maps the input data to some high dimensional values, where the data given are linearly separated, finally it provide better classification performance. SVMTRAIN(TRAINING, Y) trains a support vector machine (SVM) classifier on data taken from two groups.

Training is contains numeric matrix of data. Rows of training data correspond to observations and columns correspond to features. Y is a column vector that contains the known class labels for training. Each element of Y specifies the corresponding row of training belongs to. Training and Y must have the same number of rows. SVMSTRUCT contains information about the trained classifier, including the support vector, which is used by SVMCLASSIFY for classification.

Step 2: Localization of the image is done by using prewitt operator. It is an edge detection method which detects the affected region of the breast. The Prewitt method finds edges using the derivative. This method returns edges at where the gradient of I is maximum. This method specifies the sensitivity threshold for the Prewitt method. EDGE method ignores all edges that are not stronger than Threshold. If you do not specify Threshold, or if threshold is empty, Edge chooses the value automatically. Figure 3 (f) shows the location in which the breast cancer is present.

IV. EXPERIMENTAL RESULTS

The experimental result of the proposed system is show in figure 3. This include enhanced image and noise filtered image as a result of preprocessing after that segmentation is done. In segmentation ROI is segmented and then edge detection is done. Feature extraction is done next to segmentation here; features are extracted by using GLCM algorithm. Finally classification of image is done in this stage normal and abnormal breast are classified and for the abnormal affected region localization of cancer is done by using prewitt operator.

This proposed system is evaluated by measuring the accuracy, specificity and sensitivity. This is measured by knowing the values of True Positive (TP), True Negative (TN), False Positive (FP) and False Negative values (FN). There are some terms that are used TNR (1), FPR (2), Accuracy (3) sensitivity (4) and Specificity (5).
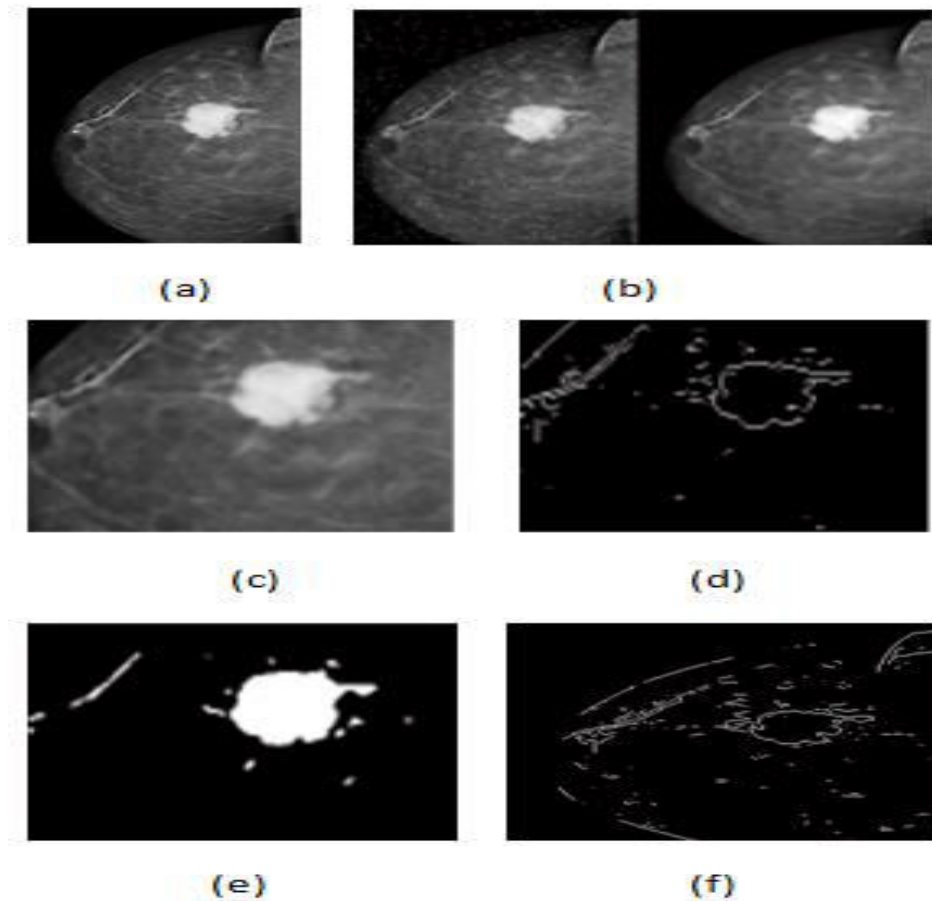

Fig 3. (a) Enhanced image (b) noise filtered image (c) Segmented image (d) Edge detected image (e) Feature extracted image (g) localization of image

True positive rate (TPR): TPR is the ratio of the number of malignant cases classified to the total number of malignant case in the tests set.

$$TNR = \frac{TP}{FN+TP} \qquad (1)$$

False positive rate (FPR): FPR is the ratio of the number of benign cases that are incorrectly classified to the total number of benign in test set.

$$FPR = \frac{FP}{TN+FP} \qquad (2)$$

$$Accuracy = \frac{TP+TN}{TN+FP+FN+TP} \qquad (3)$$

$$Sensitivity\ (TPR) = \frac{Number\ of\ true-Positive\ marks}{Number\ of\ malignants} \qquad (4)$$

$$Specificity\ (TNR) = \frac{Number\ of\ true-Neagtive\ marks}{Number\ of\ benigns} \qquad (5)$$

Free-response receiver operating characteristic (FROC) curve is for evaluating the performance of detection of Micro-calcification. An FROC plot is obtained by calculating of the true-positive detection ratio with average number of false positives per image. True positive detection ratio means how many true abnormalities that are correctly detected by computerized scheme and false positive per image refers to how many true abnormalities are misinterpreted as false.

FROC CURVE
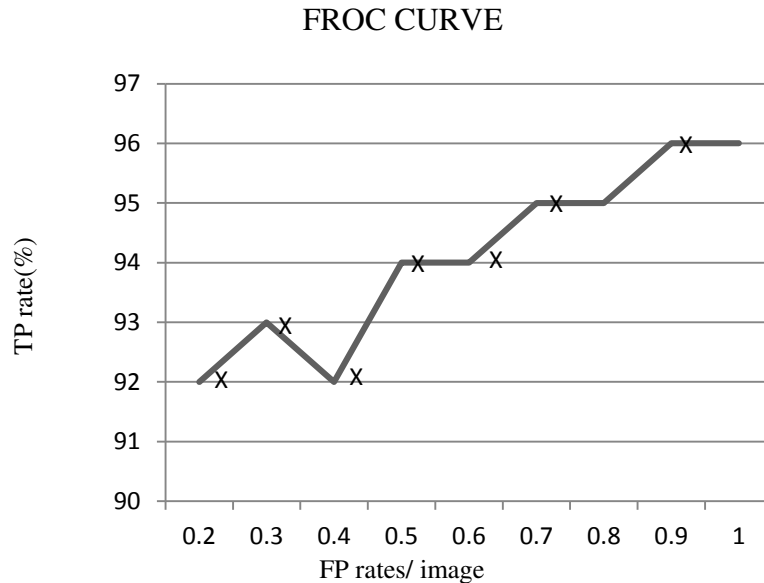
The result is calculated by using these formulas. Here, the rate of False Positive, rate of True Negative, Accuracy, sensitivity and specificity of the detection rate of mammogram images are detected for this proposed system.

V.   CONCLUSION

CAD is an important tool for early detection of breast cancer. Compared to double reading, CAD reduces the workload of radiologists and also reduces the rate of misinterpretation. CAD-based readings can provide an improved diagnostic accuracy for radiologists. From this proposed system Micro-calcification present in the breast is detected in the early stage and the rate of misinterpretation of mammogram image is reduced. From this method the rate of obtaining False Positive and False Negative test rate is reduced. Experimental results were confirmed that Micro-calcification clusters can be accurately detected by using this proposed system.

REFERENCES

[1] Giger M.L, N Karssemeijer. Computer-aided diagnosis in medical imaging. IEEE Trans Med Imaging, 20 (12) 2001, pp. 1205–1208.

[2] K Doi, "Computer-aided diagnosis in medical imaging: achievements and challenges', Congress on Medical Physics and Biomedical Engineering, 2009, p. 96

[3] KH Ng and M. Muttarak, "Advances in mammography have improved early detection of breast cancer," J. Hong Kong College Radiol, 2003, vol. 6, no. 3, pp. 126–131.

[4] R. Bird, T. Wallace, and B. Yankaskas, "Analysis of cancers missed at screening mammography," Radiology, 1992 vol. 184, no. 3, pp. 613–617.

[5] J. Brown, S. Bryan, and R. Warren, "Mammography screening: An incremental cost effectiveness analysis of double versus single reading of mammograms," Br. Med. J, 1996. vol. 312, no. 7034, pp. 809–812.

[6] R. J. Ferrari, R. M. Rangayyan, R. A. Borges, and A. F. Fr`ere, "Segmentation of the fibro-glandular disc in mammograms using Gaussian mixturemodeling," Med.Biol.Eng.Comput., vol.42, no.3, pp.378–387, 2004.

[7] T. Chrisiian Cahoon, Melanie A. Sutton, James C. Bezdek, Breast cancer detection using image processing techniques IEEE International Conference on Fuzzy Systems, 2000, 973- 976.

[8] Kother Mohideen, Arumuga Perumal, Krishnan and Mohamed Sathik, Image De noising and Enhancement Using Multi wavelet with Hard Threshold In Digital Mammographic Images, International Arab Journal of eTechnology, 2011, 2(1).

[9] Rao VNT, Govardhan A. Efficient segmentation and classification of mammogram images with fuzzy filtering. Indian Journal of Science and Technology. 2015; 8(15):1–8.

[10] Shanmugavadivu P, Sivakumar V. Fractal-based detection of Micro-calcification clusters in digital mammograms. Proceedings of ICECIT; 2012 Dec 21-23; Anantapur, Andhra Pradesh, India; Elsevier. p. 58–63.

[11] Wang TC, Karayiannis NB. Detection of Micro-calcifications in digital mammograms using wavelets. IEEE Transactions on Medical Imaging. 1998; 17(4):498–509.

[12] Zhang X, Homma N, Goto S, Kawasumi Y, Ishibashi T, Abe M, Sugita N, Yoshizawa M. A hybrid image filtering method for computer-aided detection of Micro-calcification clusters in mammograms. Journal of Medical Engineering. 2013; 1–8.

[13] Beura S, Majhi B, Dash R. Mammogram classification using two dimensional discrete wavelet transform and graylevel co-occurrence matrix for detection of breast cancer. Neurocomputing. 2015; 154:1–14.

[14] Jenefer MB. Cyrilraj V. An efficient image processing methods for mammogram breast cancer detection. Journal of Theoretical and Applied Information Technology. 2014; 69(1):32–9.

[15] Sampat, M.P., Markey, M.K., Bovik, A.C.: Computer-Aided Detection and Diagnosis in Mammography. In: Bovik, A.C. (ed.) Handbook of Image and Video Processing. Elsevier Academic Press, Amsterdam, 2005.

AUTHOR BIOGRAPHY

Ms. Jenif D Souza W S received M.E in Computer Science and Engineering from St. Joseph's College of Engineering, Chennai, Tamil Nadu, India in 2017 and B.E from SCAD College of Engineering and Technology, Tirunelveli, Tamil Nadu, India. She is working as Assistant Professor in St. Joseph's College of Engineering, Chennai, Tamil Nadu, India. Her research interests are Image processing, Machines learning and Theory of computation.