Volume 6, Issue 8, August 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Comparison of Clustering Techniques for the Detection of Microcalcifications in Mammograms

[1]Ramandeep Kaur, [2]Navneet Kaur Mavi, [3]Rashwinder Singh
[1] Department of Computer Engineering, Punjabi University, Patiala, Punjab, India
[2] Assistant Professor, Department of Computer Engineering, Punjabi University, Patiala, Punjab, India
[3] Assistant Professor, Department of Electronics and Communication Engineering, CGC-CEO, Landran, Punjab, India

Abstract- Mammography is a well-known method for early detection and diagnosis of breast tumors. Microcalcifications are tiny calcium deposits which may exist in many parts in the breast tissue. So, it is a tough task to find out clusters of microcalcifaction in digital mammogram. The proposed work first uses a mechanism to enhance the low contrast mammogram which is taken from a database and then uses Fuzzy C-means clustering algorithm and K-means clustering algorithm to find out the cluster in a mammogram and to detect microcalcifications. To test the validity of proposed work results are tested on number of mammogrampic images and are compared on the basis of accuracy measures. Mean square Error (MSE), Peak Signal-to Noise ratio (PSNR) and Bit Error Rate (BER) and Time parameters are calculated to check the performance of algorithms.

Keywords- Digital mammograms, BBHE, K-Means Clustering, Fuzzy C-Means Clustering

I. INTRODUCTION

Breast cancer has caused many deaths among women of all ages in today's world. During the past few years, death rates due to breast cancer are increasing in America as compared to any other disease. Mammography is a well-known method for early detection and diagnosis of breast tumors [1]. The abnormalities in breast tissue can be of two types, first is a mass and second are known as micro-calcifications or clusters of microcalcification. Calcifications are generally calcium depositswithin the breast tissue and look like small white dots on a mammogram and may be called as specks of calcium [2]. The calcifications may appear at different locations in mammogram and may differ in sizes. In contrast, a mass is clear than a lesion and having an oval shape. Therefore it is easy to segment than microcalcifications. Moreover,microcalcifications detection is considered as adifficult task because it may appear at different locations or appear in a shape of cluster [20].Computer-aided diagnostic (CAD) systems are used by the radiologists to read and understand the mammogram easily. CAD systems are of great use for the radiologists and can increase the accuracy and efficiency with high detection rate [5] [8]. A digital mammogram consists of variouscomponents such as background tissue, high intensity pectoral region, labels, high density tumor region etcetera [1]. It is essential that before segmentation only the breast region is taken under consideration and all these artifacts should be removed. The figure shows a Medio-Lateral Oblique view of a mammogram.

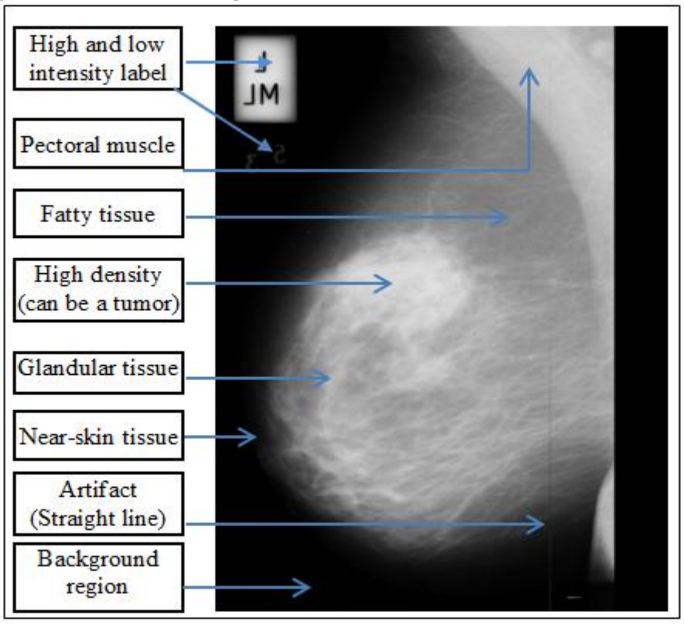

Fig. 1 The different components of mammogram.

II. RELATED WORK

Many researchers have provided different algorithms for the detectionofmicrocalcifications detection. Abdelai Elmoufidi et al. [1] presented a hybrid approach of K-means algorithm and Seeded based region growingto segment the breast into different regions. Moreover, Siti Salmah Yasiran et al. [2] indicated in their study that the edge detection techniques using Sobel, Prewitt's and Laplace of Gaussian operator on Active contour model produces better results as compared to previous techniques. Sze Man Kwok et al. [3] proposed a new method for pectoral muscle identification and ROI extraction which is based on straight line estimation and iterative cliff detection. The combination of fuzzy C-means segmentation with Top-hat transforms and thresholding technique was used by Akshay S.Bhardwaj and Mehmet Celenk [5] to segment the foreground from background area. A new approach used by P. Shanmugavadivu to detect microcalcifications clusters that uses intensity directed region growing method along with threshold values [6]. The study carried out by L. Vivona et al. [4] suggests that the Fuzzy Logic Clustering algorithm with a set of described features is best for microcalcification detection process and also to find the exact number of clusters. Akshay S. Bhardwaj and Mehmet Celenk [6] proposed a new method to detect MCs in a digital mammogram. The approach starts with the segmentation of the digital mammogram to isolate the breast region, using fuzzy C- Means clustering algorithm. After that, the segmented image is then further segmented using top-hat transform to identify the region of interest. Moreover, P. Shanmugavadivu et al. [7] used intensity directed region growing approaches using a threshold value to segment the microcalcifications in a mammogram. A new method that is a combination of Ostu thresholding and Morphological operators is used by M. A. Duarte et al. Anuj Kumar Singh [9] et al. introduced simple techniques such as averaging and thresholding methods. A simple and easy approach is used for the detection and segmentation of tumor region. In this process, firstly averaging filter is used for the smoothening of mammographic image. After that, Max-Mean and Least-Variance techniques are used to detect cancer parts in mammogram. M. A. Durate et al. [10] presented an automatic method for segmentation using structuring elements of different sizes to find out microcalcifications. They used a simple approach called Otsu thresholding with morphological filters for segmentation. Different methods of contrast enhancement based on histogram equalization for the preprocessing of grayscale image have been discussed by Gourav Garg et al. [11] which is the first step in mammography. We used the BBHE contrast enhancement method in our methodology.

III. PREPROCESSING

Preprocessing steps are considered as a crucial step in order to limit the search for abnormal tissues without the influence from background of the mammogram. Segmentation of breast consists of extraction of border of breast tissue, pectoral muscle detection or any label identification etcetera. Preprocessing is important on images obtained from mammographic devices. To limit the area of detection for microcalcifications, the breast region must be initially identified from the image. Mammographic preprocessing is also used to minimize the effects of noise, grandular tissues and background which may cause many FPs in the segmentation process.The flow chart represents our methodology used for comparison:

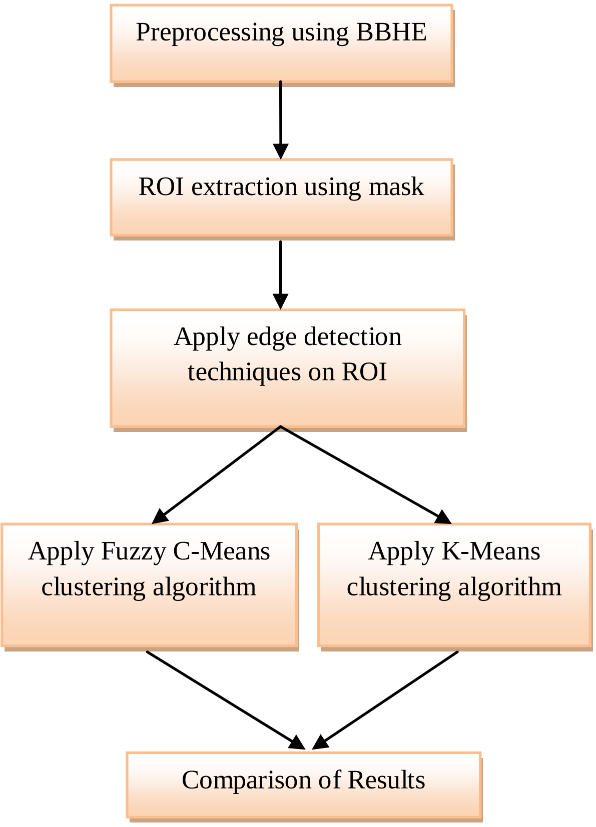

Fig. 2 Steps in Proposed methodology

Histogram equalization methods are used for preprocessing purpose. The technique used in this methodology is BBHE.

3.1 Brightness Preserving Bi-Histogram Equalization (BBHE):

Using this method, the histogram of an image is divided into two equal parts which are based on the average intensity of all pixels. Moreover, this method is used to preserve the brightness of image.

This technique is represented as:

$X=X_L \text{ U } X_U$

Where $X_L=\{X(i, j)/X(I, j) \leq X_m, ¥ X(I, j) € X\}$

And $X_U=\{X(i, j)/X(I, j) > X_m ¥ X(i, j) € X\}$

X_m is the mean of image X and X_m is decomposed into sub images X_L and X_U.

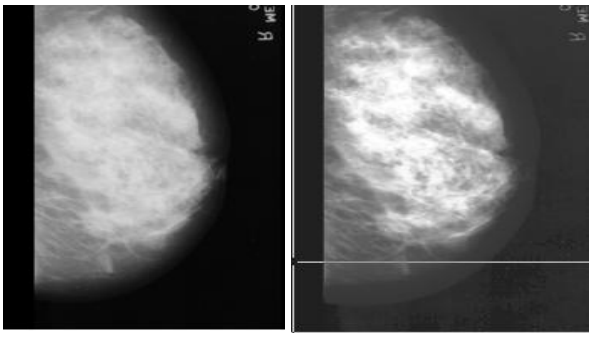

Fig.3 (a) Original image mdb236 (b) Enhanced image

After preprocessing only the breast region is separated using a mask and this region is called as Region of Interest (ROI). Further detection is carried out on this extracted ROI.

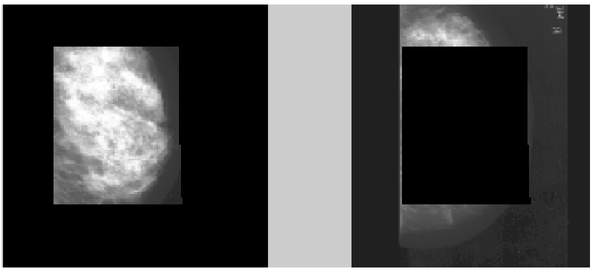

Fig. 4 (a) Extracted ROI (b) Non-ROI.

3.2 Edge Detection:

In the third step, various edges are detected from the region of interest which is carried out using three operators: Sobel operator, Prewitt's operator and Laplace of Gaussian and these edge detected images are used for clustering mechanisms. The 3×3 masks used by these three operators are as shown in fig. 4(a). Moreover after edge detection, morphological opening and closing are used to fill the holes and extra edge to the background as shown in fig. 5(a).

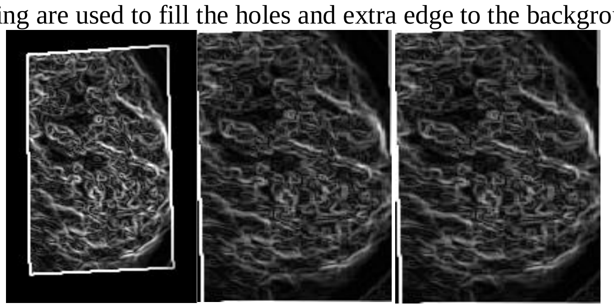

Fig.5 (a) Sobel Operator (b) Prewitt's Operator (c) LoG Operator

IV. SEGMENTATION

4.1 Fuzzy C-Means Clustering:

Fuzzy Clustering analysis involves assigning data points to clusters (also called classes) such that pixels in the same class are similar in each way, whilst pixels that belong to different classes are as dissimilar as possible. The classes or clusters are identified by means similarity measures such as connectivity, distance and intensity value.Clustering is done by assigning membership to each data point called as pixel corresponding to each cluster center and is done on the basis of distance between the selected pixel and cluster center. In this technique, more the pixel is near to the center of a particular cluster more is its membership towards that cluster. After a single iteration occurs, membership and cluster centers are updated every time which follows the formula as under:

$\mu_{ij}=1/\sum_{k=1}^{c}\left(D_{ij}/D_{ik}\right)^{\frac{2}{m}-1}$ (1)

$v_j=\left(\sum_{i=1}^{n}\left(u_{ij}\right)^m x_i\right)/\left(\sum_{i=1}^{n}\left(u_{ij}\right)^m\right), ¥\ 1,2,3 \ldots\ldots c$ (2)

Where,

'n' is the number of data points.

'v_j' is used to represent the j^{th} cluster center.

'm' is the fuzziness index and m € [1, ∞].

'c' is the number of cluster centers.

'μ_{ij}' represents the membership of i^{th} data point to j^{th} cluster center.

'D_{ij}' iscalled as the 'Euclidean distance' between i^{th} data value and j^{th} cluster center.

Primary objective of fuzzy c-means algorithm is to reduce the following:

$J(U,V)=\sum_{i=1}^{n}\sum_{i=1}^{n}\left(u_{ij}\right)^m \left\|x_i - v_j\right\|^2$ (3)

Where,

'$||x_i - v_j||$' is the Euclidean distance

Fuzzy C-Means Clustering is performed under various steps:

Let X = {$x_1, x_2, x_3, x_4 \ldots, x_n$} is the set of data values or pixels and

V = {$v_1, v_2, v_3 \ldots, v_c$} is the set of centers of clusters.

1) First of all, randomly select 'c' cluster center.

2) Then, calculate the fuzzy membership 'μ_{ij}' using:

$\mu_{ij}=1/\sum_{k=1}^{c}\left(D_{ij}/D\right)^{\frac{2}{m}-1}$ (4)

3) After the calculation of membership function, compute the fuzzy centers 'v_j' using:

$V_j=\left(\sum_{i=1}^{n}\left(u_{ij}\right)^m x_i\right)/\left(\sum_{i=1}^{n}\left(u_{ij}\right)^m\right), ¥\ 1,2,3 \ldots\ldots c$ (5)

4) Repeat steps 2) and 3) until the value of 'J' is reached at minimum value
or $||U^{(k+1)} - U^{(k)}|| < \beta$.

Where,

'k' is iteration count.

'β' is called as 'Termination criterion 'between[0,1].

'U'= $(\mu_{ij})_{n*c}$' is the fuzzy membership matrix.

'J' is the objective function.

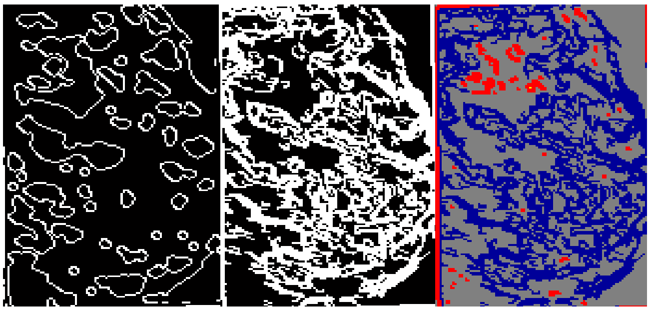

Fig. 6 (a) Sobel Edge detected and Morphologically refined image (b)-(c) Fuzzy segmented image.

4.2 K-means Clustering Algorithm

K-Means clustering is the process of grouping the given data values into different clusters. K-means clustering is a popular method used in image segmentation process. The K-means algorithm is used to segment a mammogram into specified clusters with their cluster centers. The basic algorithm is as follows:

1) First of all, select K cluster centers, which are based on some predefined parameters.
2) When a pixel value in the image is having minimum distance between the selected pixel value and the center of cluster, assign that pixel to the cluster.
3) Calculate the average of all pixels in the cluster and again compute the center of new cluster.
4) Repeat the previous steps (2) and (3)
5) Stop when all the clusters are obtained.

Distance is the squared or absolute difference between a selected pixel and a cluster center that is close to cluster centre. Moreover, the difference may be based on intensity of pixel, texture, or some other parameters. K can be selected on the random basis or manually.

Input is K number of clusters

E: dataset {$e_1, e_2, e_3 \ldots\ldots e_n$} of n values

Output of k-Means is set of K clusters as specified in input of K-mean Algorithm.

$$e(x, y) = \sum |x_i - y_i|^2$$

1) First of allcalculate the mean of data values within each cluster.

2) Update the mean value by repeating the whole process until no change.

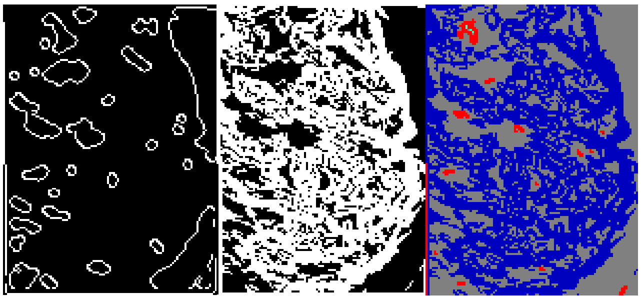

Fig. 7 (a) Edge Detected and Refined image (b)-(c) K-Means Segmentation

V. EXPERIMENTAL RESULTS

The K-Means and Fuzzy C-Means clustering methods are tested using 23 images taken from mini-MIAS database which contains images of benign and malignant masses, calcifications and normal tissues. These mammographic images are collected by mammographic image analysis society. Bit Error Rate (BER), Peak Signal-to-Noise Ratio (PSNR), time and Mean Square Error (MSE) parameters are calculated on Different images and comparison is performed using these two clustering techniques. Moreover, True Positive, True Negative, False Positiveand False Negative rates are compared to measure the accuracy.

Table 1 shows the values of MSE, PSNR, BER parameters on 23 mammograms using K-Means Clustering algorithm:

Image	Mdb209	Mdb211	Mdb213	Mdb216	Mdb218	Mdb219	Mdb223	Mdb226	Mdb227	Mdb231	Mdb233	Mdb236
MSE	6.62e+03	8.47e+03	6.18e+03	1.20e+04	8.57e+03	1.04e+04	4.94e+03	4.98e+03	7.02e+03	6.97e+03	5.90e+03	1.01e+04
PSNR	9.91	8.84	10.21	7.31	8.8	7.92	11.19	11.15	9.66	9.69	10.42	8.08
BER	0.1	0.11	0.09	0.13	0.11	0.12	0.08	0.08	0.1	0.1	0.09	0.12

Image	Mdb238	Mdb239	Mdb240	Mdb241	Mdb245	Mdb248	Mdb249	Mdb252	Mdb253	Mdb256	Mdb290
MSE	5.62e+03	1.36e+04	1.38e+04	7.48e+03	6.46e+03	7.34e+03	8.46e+03	6.23e+03	1.19e+04	9.63e+03	4.58e+03
PSNR	10.63	6.79	6.71	9.38	10.02	9.47	8.85	10.18	7.35	8.29	11.54
BER	0.09	0.14	0.14	0.1	0.09	0.1	0.11	0.09	0.13	0.12	0.08

Similarly, table 2 shows the values of MSE, PSNR and BER using Fuzzy C-means clustering technique:

Image	Mdb209	Mdb211	Mdb213	Mdb216	Mdb218	Mdb219	Mdb223	Mdb226	Mdb227	Mdb231	Mdb233	Mdb236
MSE	6.64e+03	8.49e+03	6.19e+03	1.20e+04	8.58e+03	1.05e+04	4.94e+003	4.99e+03	7.03e+03	6.98e+03	5.91e+03	1.10e+04
PSNR	9.9	8.84	10.21	7.31	8.79	7.91	11.19	11.14	9.66	9.68	10.41	8.07
BER	0.1	0.11	0.09	0.13	0.11	0.12	0.08	0.08	0.1	0.1	0.09	0.12

Image	Mdb238	Mdb239	Mdb240	Mdb241	Mdb245	Mdb248	Mdb249	Mdb252	Mdb253	Mdb256	Mdb290
MSE	5.62e+03	1.36e+03	1.38e+03	7.49e+03	6.46e+03	7.33e+03	8.41e+03	6.23e+03	1.19e+04	9.63e+03	4.55e+03
PSNR	10.62	6.78	6.7	9.38	10.02	9.46	8.85	1.18	7.35	8.2	11.54
BER	0.09	0.14	0.14	0.1	0.09	0.1	0.11	0.09	0.13	0.12	0.08

Table 3 shows the comparison of FCM and K-Means clustering techniques. As lower the mean square error and higher the value of PSNR the results achieved are better.

Table 3 Comparison of K-Means and Fuzzy C-Means techniques in terms of TP, TN, FP, FN.

Algorithm	TP	FN	TN	FP
Fuzzy C-Means	18	2	2	1
K-Means	15	3	3	2

Both the clustering algorithms using with edge detection technique shows very effective results when individually applied on database images. From the results it can be seen Fuzzy C-means algorithm matches greater numbers of clusters as compared to K-means algorithm.

Table 4 represents the accuracy of FCM and K-Means clustering.

Algorithm	Accuracy
Fuzzy C-Means	86%
K-Means	78.26%

Table 5 Represents the time comparison of FCM and K-Means algorithms.

Clustering Algorithm	Average Time Taken (in sec.)
Fuzzy C-Means	6.638 sec.
K-Means	2.334 sec.

VI. CONCLUSION AND FUTURE SCOPE

In this research, we presented work a comparison of clustering algorithm for microcalcifications detection. The proposed method uses K-Means and Fuzzy C-Means clustering algorithms that are applied on the digital mammogram with three edge detection techniques i.e. Sobel operator, Prewitt's operator and Laplace of Gaussian operator. However the time taken by K-Means algorithm is less as compared to fuzzy C-Means algorithm but after testing on 23 database images, Fuzzy C-Means clustering shows 86% accuracy and K-means shows 78.26%. In addition, the Laplace of Gaussian operator does not show very good results when used with clustering techniques whereas the results shown by Prewitt's and Sobel are almost similar. Moreover, both algorithms are compared using MSE, BER and PSNR. So, from the results it is clear that the Fuzzy C-Means algorithm produces more accurate results as compared to K-Means algorithm. In addition, the time taken by K-Means algorithm for segmentation is less as compared to Fuzzy C-Means algorithm but the clusters formed by the Fuzzy C-Means technique are more compact.

In future, detection rate can be enhanced by using both the clustering techniques or by using other classification measures. Feature extraction techniques can also be used to gain better results.

REFERENCES

[1] Abdelali Elmoufidi, Khalid el Fahssi, said Jai-Andalossai, Abderrahin Sekkaki "Automatically Density Based Breast Segmentation for Mammograms By Using Dynamic K-Means Algorithm And seed Based region Growing" *IEEE Instrumentation And Measurement Society*, 2015.

[2] Siti Salmah Yasiran,Rozi Mahmud "Microcalcifications Segmentation Using Three Edge Detection Techniques" *IEEE International Conference on Electronic design, System and Applications*, pp. 207-211, 2012.

[3] Sze Man Kwok, Ramachandran Chandrasekhar, Member, IEEE, Yianni Attikiouzel, Fellow, IEEE, and Mary T. Rickard "Automatic Pectoral Muscle Segmentation on Mediolateral Oblique View Mammograms, "*IEEE transactions on medical imaging*, Vol. 23, pp.1129-1140, 2004.

[4] L.Vivona, D. Cascio, R. Magro, F.Fauci, G. Raso "A Fuzzy Logic C-Means Clustering Algorithm to Enhance Microcalcifications clusters in digital Mammograms" *IEEE Nuclear Science Symposium Conference Record*, 3048-3050, 2011

[5] Akshay S. Bhardwaj and Mehmet Celenk "Detection Of Microcalcification With Top Hat Transforms and The Gibbs Random fields" *IEEE*, pp. 6382-6385, 2015.

[6] P. Shanmugavadivu and S. G. Lakshmi Narayan "Segmentation of Microcalcifications in mammogram Images using Intensity-directed Region growing" ICCCI, January 2013

[7] Indra Kanta Maitra, Sanjay Nag, Sammer Kumar "Techniques For Preprocessing of digital mammogram" *Computer Methods And Programs In Biomedicine*, pp. 175-188, 2012.

[8] Fatemeh Pak, HamidrezaRrashidy Kanam, Afsaneh Alikhassi "Breast Cancer Detection And Classification In Digital Mammography Based On Non Subsampled Contourlet Transform (NSCT) And Super Resolution" *Computer Methods And Programs In Biomedicine*, June 2015.

[9] Mario Mustra, Mislav Grgic "Robust Automatic Breast and Pectoral Muscle Segmentation For Scanned Mammograms" *Signal Processing*, pp. 2187-2887, 2013.

[10] M. A. Duarte, A. V. Alvarenga,C.m. Azevedo "Automatic Microcalcifications segmentation Procedure Based on Otsu Method and Morphological filters" *IEEE*, pp. 102-106, April 2011.

[11] Gaurav Garg, Poonam Sharma "An Analysis of HistogramEqualization Method for Brightness Preserving and Contrast enhancement" *International Journal of Innovations and advancement in Computer Science, IJIACS* ,vol. 3, pp. 1-11, 2014.