IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-12, Dec 2018 E-ISSN: 2347-2693

Breast Cancer Segmentation Using Global Thresholding and Region Merging

Nidhi Singh[1*], S. Veenadhari[2]

[1] Department of Computer Science & Engineering, LNCT Bhopal, India
[2] Department of Computer Science & Engineering, AISECT University, India

*Corresponding Author: nnidhi_sng14@yahoo.com

Available online at: www.ijcseonline.org

Accepted: 06/Dec/2018, Published: 31/Dec/2018

***Abstract*—** Recently, more attention is being given to detect breast cancer in women. But, Due to the lack the diagnostic to suggest whether breast cancer is presented in a person is still a research issue. The proposed work gives a hybrid methodology based on global thresholding and region merging for segmentation of breast cancer in Mammogram Images. In the proposed algorithm we use wiener filtering to remove Gaussian noise then apply image normalization based on histogram shrink to enhance the quality of image. Next, Global thresholding using Otsu's method is used in order to segment the masses resulting Region of Interest(ROI) and then Region merging is used to extract segmented masses from image. Accuracy rate of the proposed method is 82% and Error rate is only 18%.

***Keywords*—** Breast Cancer, Gaussian Noise, Mammogram Mass, Otsu's Method, Region Merging

I. INTRODUCTION

Breast cancer is the most awful disease which is a found in women breast. it happens when breast cell start to grows non stopping. Breast cancer can be categories as ductul carcinoma and lobules carcinoma. When the cancer found in the milk ducts, then they are called ductal carcinoma and when the cancer found in the lobules then they are called lobular carcinoma [1]. Genetic factor or environmental factor may cause the breast cancer. The reason behind cancer due to genetic factors may be family background, individual health record, age, certain genome changes, dense breast tissue, gender etc. The reason behind cancer due to environmental factors may be poor diet, use of alcohol, less physical activity, etc. Breast tumor (cancer) are of type Benign and malignant. The benign tumors are usually non assertive and non cancerous. This type of tumor will not grow to other body parts. But, the malignant tumors are examined as cancerous and aggressive as well as grow all over the body cells i.e., metastatic in nature [2]. The classification whether a tumor is benign, malignant, and normal can be done through different techniques like SVM, neural networks; fuzzy c means clustering, etc. Different imaging techniques are possible for early detection of breast cancer. The detailed detection of the images is done using feature extraction and texture extraction techniques. This itself opens another research area as a wide variety of techniques have been used for segmentation, feature extraction, enhancement done mainly by wavelet techniques, clustering, using GLCM matrix etc which are described clearly in the related works [3].

From last decade, many researchers had proposed work on breast cancer detection and segmentation in mammogram image. Some of these based on threshold [4], gradients [5], polynomial [6], or active contours [7]. In order to detect the masses, K-means and co-occurrence matrix is described by Leonardo de Oliveira Martins et al. [8] in 2009 and SVM classifier is used. For classifying, images are separated into two groups based on masses and non masses that are shape and texture descriptor respectively Authors in [9] classify and detect breast cancer based on a Particle Swarm Optimized Wavelet Neural Network. Subashini et. al. [10] proposed breast cancer detection through ultrasound images. They removed noise by using DWT, segmentation is happened with active contour model and classification is done by using the back propagation neural network. Authors in [11] defined the computer aided diagnosis system that can detect normal and abnormal pattern in the breast. They used seven different classifiers to compare the classification results. DWT was used to select region of interest (ROI).and 20 texture feature were extracted from ROI.

Proposed work divided into four phase: in the first phase, we have applied preprocessing. In the preprocessing phase we used wiener filter to remove unwanted Gaussian noise. Then apply Normalization by histogram shrinking method to enhance the quality of input image. In the next phase we identify the global threshold value by using otsu's method and then based on the threshold value segment the image into two regions. In the next phase we use region merging technique to extract the ROI and based on ROI we segment the mammogram mass.

The remaining section is ordered as follows: Section II shows the quality parameter used in proposed technique. Section III describes working of the proposed method; Section IV consists of results and analysis; conclusions and future work are described in section V.

II. THE QUALITY PARAMETERS

Confusion matrix is used to evaluated the performance of the proposed algorithm is evaluated.

Basically, a confusion matrix provides information regarding the predicted and actual class carried out by the proposed method. The confusion matrix for benign and malignant class is given in table 1 and performance measures are given in table 2, respectively.

Table1

Actual Class	**Predicted class**	
	Positive	***Negative***
Positive	TP (true positive)	FN (false negative)
Negative	FP (false positive)	TN (true negative)

Table2

Measure	**Definition**
TPR or recall	TP/(TP + FN)
FPR	FP/(FP + TN)
precision	TP/(TP + FP)
ACC	(TP + TN/K)

The TPR calculates correctly classified malignant ROIs out of the total number of malignant ROIs. The FPR parameter calculates incorrectly classified benign ROIs out of the total number of benign ROIs. The ACC is an accuracy of observed and predicted classification.

III. PROPOSED METHODOLOGY

Breast cancer detection by proposed method is partition into following segments:

a. Local Adaptive Filter
b. Image Normalization
c. Global Thresholding
d. Region Based Segmentation

Details of all the segments are described in following sub sections.

A. Local Adaptive Wiener Filtering

Let the model for the image corrupted by white Gaussian noise is defined as:

$$y(i,j) = x(i,j) + n\ (i,j)$$

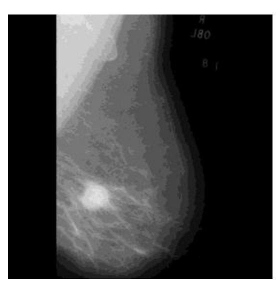
Where $x(i,j)$ is the noise free image pixel at (i,j) location and $n\ (i,j)$ is additive Gaussian noise whereas $y(i,j)$ is the image containing the noisy pixel at (i,j) location. The main purpose is to get rid of noise $n\ (i,j)$, and to gain a linear estimate $\hat{x}(i,j)$ of $x(i,j)$ which attenuates the mean squared error (MSE).

The Wiener filter [12] has a following scalar form:

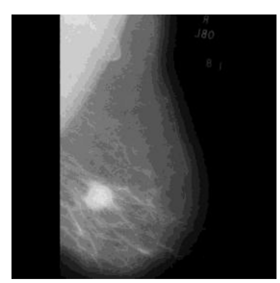
$$\hat{x}(i,j) = \frac{\sigma_x^2(i,j)}{\sigma_x^2(i,j) + \sigma_n^2(i,j)}[y(i,j) - \mu_x(i,j)] + \mu_x(i,j)$$

Where $\sigma_x^2(i,j)$ and $\mu_x(i,j)$ are the signal variances and means, respectively. Let assume that mean of the noise is to be zero. Then only we need to estimate $\mu_x(i,j)$, $\sigma_x^2(i,j)$ and $\sigma_n^2(i,j)$. We will assume that the noise mean and variance are known then we focus on the local estimation of $\mu_x(i,j)$ and $\sigma_x^2(i,j)$. The local mean and local variance are calculated over a uniform moving average window of size (2k + 1) × (2k + 1).

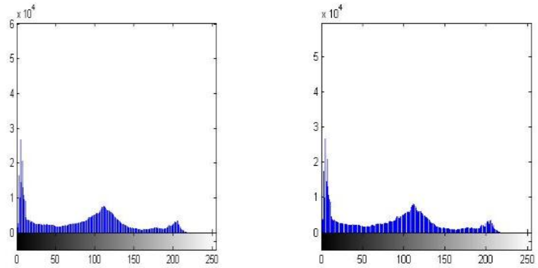
In this proposed method Wiener Filtering with 3x3 neighborhoods is applied on mammogram images. Fig. 1 (a-b) shows images before and after wiener filtering and Fig. 1(c-d) shows histograms of those images.



(a)



(b)



(c) (d)

Figure 1. Comparison Between Test Image and Histograms Before and After Filtering (a) Original Mammography Image, (b) Image After Wiener Filtering, (c) Histogram of Original Image Before Filtering, (d) Histogram of Image After Median Filtering

B. Image Normalization Using Histogram shrink

Unlike the histogram stretching, histogram shrinking is the process of compressing the gray levels of the image to normalize the image contrast. Histogram shrinking mapping function can be expressed by the following equation:

$$Shrink(I(x,y)) = \left[\frac{S_{\max} - S_{\min}}{I_{\max} - I_{\min}}\right].[I(x,y) - I_{\max}] + S_{\min}$$

$S_{\max}$ And $S_{\min}$ correspond to the maximum and minimum desired in the compressed histogram. After applying the Histogram shrink on the wiener filtered image (see Fig 1. (c)) we get the resulted image as shown in Fig. 2.

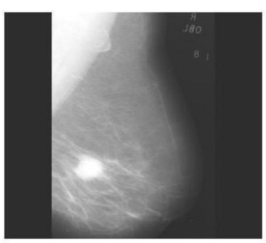


Figure 2. Image After Normalization using histogram shrinking

C. Global Thresholding Using Otsu's Method

Otsu's method is an optimal global thresholding method proposed by Otsu [13]. In this method optimal threshold is selected automatically and stably based on the global property of the histogram.

Let L is total number of gray level presented in given image. Then pixels have range of gray level between 0 to L-1. Let the number of pixels at level i is denoted by n_i . Suppose that a threshold k is chosen such that C1 is the set of pixels with levels [0,1,2,…..,k] and C2 is the set of pixels with levels [k+1,…..,L-1]. The Otsu method select the optimum threshold value k that maximize the between class variance, defined as

$$\sigma_B^2(k) = P_1(k)[m_1(k) - m_G]^2 + P_2(k)[m_2(k) - m_G]^2$$

The Probability $P_1(k)$ that a pixel is assigned to class C1 is given by:

$$P_1(k) = \sum_{i=0}^{k} p_i$$

Similarly, the probability of pixels assigned to class C2 is defined as:

$$P_2(k) = \sum_{i=k+1}^{L-1} p_i = 1 - P_1(k)$$

The Global mean m_G is defined as:

$$m_G = \sum_{i=0}^{L-1} ip_i$$

The mean intensity for class C1 is given by

$$m(k) = \sum_{i=0}^{k} ip_i$$

The between class variance is given by

$$\sigma_B^2(k) = \frac{[m_G P_1(k) - m\ (k)]^2}{P_1(k)[1 - P_1(k)]^2}$$

The optimal threshold value k is selected from largest value of $\sigma_B^2(k)$. The ratio of the between class variance to the total image intensity variance,

$$\eta(k) = \frac{\sigma_B^2(k)}{\sigma_G^2}$$

Is a measure of reparability of image intensities into two classes (foreground and background) which can be shown to be in range 0≤η (k*) ≤ 1 Where k* is the optimal threshold.

After Applying Global Thresholding Using Otsu Method on the image getting from previous step is shown if Fig. 3.

Figure 3. Shows Image After applying Global Thresholding Using Otsu Method.

D. Region-based segmentation

Region-based segmentation is a technique for partitioning the image into different regions depends on homogeneity of the regions. Each pixel into the region has the some similar characteristics. Region based segmentation work as follow:

In the first step we have to select a set of seed points. Selection of the seed point depends on the factor such as pixels in a certain gray-level range etc.

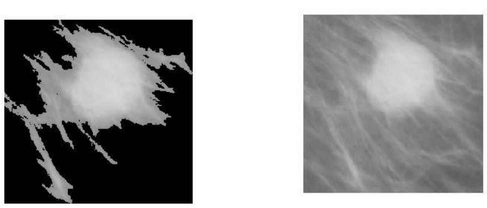
In the next step regions are then grown from these seed pixels to adjacent pixels depending on a region membership criterion. The criterion could be, for example, pixel intensity, gray level texture, or color.

Repeat step 2 for each of the newly added pixels; stop if no more pixels can be added.

Fig 4 show the Extracted ROI from Original Image

Figure 4. Extracted ROI from output of global Thresholding by region growing method

Fig. 5(a) shows the ROI extraction from original image and 5(b) shows the final Output mass Image After cleaning.



(a) (b)

Figure 5. (a) ROI extraction from original image (b).Final Output mass Image After cleaning

Step by step process of proposed method is described by Flow chart as shown in fig. 6.

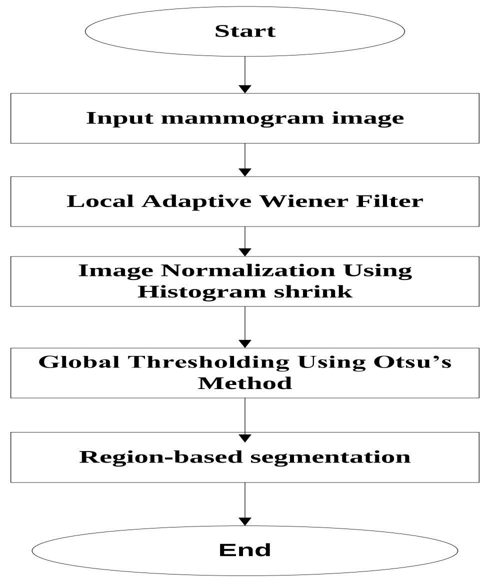


Figure 6. Shows Flow Chart of Proposed method

IV. RESULTS AND DISCUSSION

The proposed algorithm has been tested on MIAS dataset. We have randomly selected 50 sample images from MIAS dataset [14]. MIAS consists of 322 mammogram images of size 1024 × 1024 pixels. Simulations have been carried out on 3.2 GHz processor with 4 GB RAM.

The result of the segmented masses for three mammogram images is shown in fig. 7 to fig. 9. In this proposed method we observed that preprocessing step greatly affect the segmentation process.

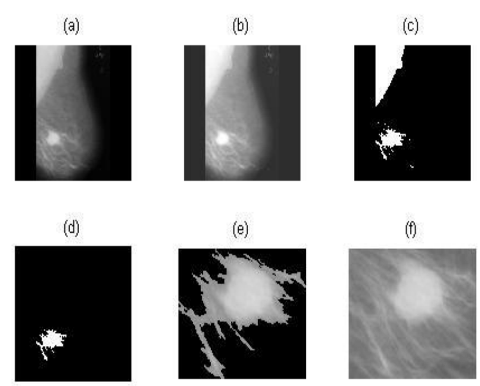


Figure 7. Result of Segmented masses. (a) Original image mdb.jpg, (b) Mammography image after Wiener Filter Normalization step, (c) segmented image by Otsu method (d) ROI extraction by region growing (e) ROI extraction from original image (f) Final Output mass Image After cleaning

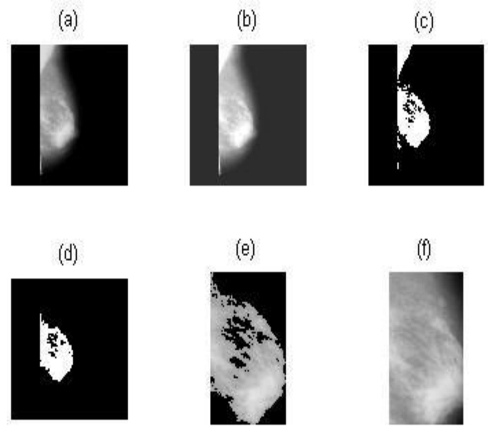


Figure 8. Result of Segmented masses (a) Original image mdb072.jpg, (b) Mammography image after Wiener Filter Normalization step, (c) segmented image by Otsu method (d) ROI extraction by region growing (e) ROI extraction from original image (f) Final Output mass Image After cleaning

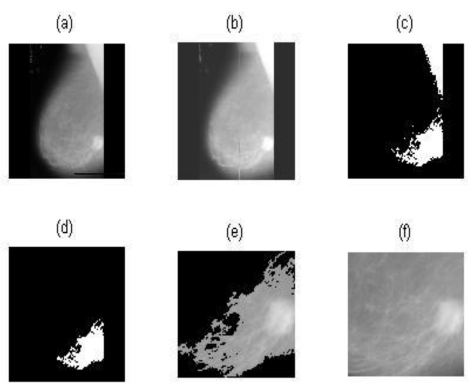


Figure 9. Result of Segmented masses (a) Original image mdb271.jpg, (b) Mammography image after Wiener Filter Normalization step, (c) segmented image by Otsu method (d) ROI extraction by region growing (e) ROI extraction from original image (f) Final Output mass Image After cleaning

50 MIAS images are used in the experiments. 35 of these are cancerous image and 15 normal images. With the proposed algorithm, out of these 35 cancerous image, 29 patients was segmented correctly and 6 the cancerous region of the tumor was found to be erroneous. When the proposed algorithm is applied on the remaining 15 patients on normal images, 3 images, and cancer tissue was determined, and in the case of 12, it was segmented by the algorithm that these images do not contain cancer. In the data table shown in Table 3; accuracy rate 82% and Error rate is 18%.

Table 3: confusion matrix

Actual Class	**Predicted class**	
	Cancerous image	***Normal Image***
Cancerous image	29	06
Normal Image	03	12

V. CONCLUSION AND FUTURE SCOPE

Proposed work starts by using wiener filter to remove unwanted Gaussian noise. Then apply Normalization by histogram shrinking method to enhance the quality of input image. In the next step identify the global threshold value by using Otsu's method and then based on the threshold value segment the image into two regions. Finally, using region merging technique extract the ROI and based on ROI segment the mammogram mass. The proposed techniques are implemented and tested under MATLAB environment on 50 mammography images (normal and cancer) to obtain ROI. Accuracy rate of the proposed method is 82% and Error rate is only 18%.

REFERENCES

[1] Saurabh PrasadLori Mann BruceJohn E Ball," A Multi-classifier and Decision Fusion Framework for Robust Classification of Mammographic Masses" International Conference of the IEEE Engineering in Medicine and Biology , 2008.

[2] Islam M.J., Ahmadi M., Sid-Ahmed M.A. (2010) Computer-Aided Detection and Classification of Masses in Digitized Mammograms Using Artificial Neural Network. Advances in Swarm Intelligence. ICSI 2010.

[3] S. Baeg, N. Kehtarnavaz,"Texture based classification of mass abnormalities in mammograms", 13th IEEE Symposium on Computer-Based Medical Systems, 2000..

[4] R. C. Gonzalez and R. E. Woods, Digital Image Processing, 3 ed.: Pearson Prentice Hall, 2007.

[5] A.J. Mendez, P.J. Tahoces, M.J. Lado, M. Souto, J.L. Correa, and J.J. Vidal, J.J, "Automatic Detection of Breast Border and Nipple in Digital Mammograms", Computer Methods and Programs in Biomedicine, vol. 49, pp. 253–262, 1996.

[6] R. Chandrasekhar, and Y. Attikiouzel, Y. Automatic, "Breast Border Segmentation by Background Modeling and Subtraction", in: M.J. Yaffe (Ed.), Proceedings of the 5th International Workshop on Digital Mammography (IWDM), Medical Physics Publishing, Toronto, Canada, 2000, pp. 560–565.

[7] M. A. Wirth, A. Stapinski, "Segmentation of the breast region in mammograms using active contours", in Visual Communications and Image Processing, pp.1995–2006.

[8] Leonardo de Oliveira Martins, E.C. da Silva, A.C. Silva, A. C. de Paiva," Classification of Breast Masses in Mammogram Images Using Ripley's K Function and Support Vector Machine", International Workshop on Machine Learning and Data Mining in Pattern Recognition ,MLDM 2009.

[9] J. Dheeba, N. Albert Singh, S. Tamil Selvi 'Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach', Journal of Biomedical Informatics 49 (2014) 45–52, 2014 Elsevier Inc.

[10] K.Subashini, K.Jeyanthi, 'Masses detection and classification in ultrasound images', IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS),Volume 9, Issue 3 Ver. II (May -Jun. 2014), PP 48-51

[11] Milos Radovic,Marina Djokovic,Aleksandar Peulic,Nenad Filipovic,"Application of data mining algorithms for mammogram classification",13th IEEE International Conference on BioInformatics and BioEngineering,2014

[12] F. Jin, P. Fieguth, L. Winger and E. Jernigan,"Adaptive Wiener Filtering Of Noisy Images And Image Sequences " , IEEE conference 2013.

[13] Otsu N.,"A Threshold Selection Method from Graylevel Histograms", IEEE Transactions on Systems, Man&Cybernatics, Vol. 9, Issue 1, 1979, pp. 62-66.

[14] [Online] Available: ttp://www.wiau.man.ac.uk.services/MIAS / MIASweb.htm.

Authors Profile

Mr. C T Lin pursed Bachelor of Science from University of Taiwan, Taiwan in 2006 and Master of Science from Osmania University in year 2009. He is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Computational Sciences, Department of Electronic and Communication, University of Taiwan, Taiwan since 2012. He is a member of IEEE & IEEE computer society since 2013, a life member of the ISROSET since 2013, ACM since 2011. He has published more than 20 research papers in reputed international journals including Thomson Reuters (SCI & Web of Science) and conferences including IEEE and it's also available online. His main research work focuses on Cryptography Algorithms, Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining, IoT and Computational Intelligence based education. He has 5 years of teaching experience and 4 years of Research Experience.

Mr C H Lin pursed Bachelor of Science and Master of Science from University of New York, USA in year 2009. He is currently pursuing Ph.D. and currently working as Assistant Professor in Department of Telecommunication, University of New York, USA since 2012. He is a member of IEEE & IEEE computer society since 2013, a life member of the ISROSET since 2013 and ACM since 2011. He has published more than 20 research papers in reputed international journals including Thomson Reuters (SCI & Web of Science) and conferences including IEEE and it's also available online. His main research work focuses on Cryptography Algorithms, Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining, IoT and Computational Intelligence based education. He has 5 years of teaching experience and 4 years of Research Experience.

