

# Breast Cancer Diagnosis using Digital Image Segmentation Techniques

Monica\*, Singh Sanjay Kumar, Agrawal Prateek, Madaan Vishu

School of Computer Science Engineering, Lovely Professional University, Jalandhar-Delhi, G.T. Road, National Highway 1, Phagwara - 144411, Punjab, India; monicasabharwal90@gmail.com, sanjayksingh.012@gmail.com, prateek061186@gmail.com, vishumadaan123@gmail.com

## Abstract

**Background/Objectives:** The objective of this research is to work out on the computer-aided detection of Breast cancer. This research work mainly target to build up a structure of methods by using image processing and classification approach for the recognition of abnormalities in Mammograms. **Methods/Statistical Analysis:** It is observed that the breast images are analyzed after decomposition. However, the image become smaller and crucial information may be lost by virtue of image decomposition and when region of interest is applied only on the specific segment of image as above said some information get lost. **Findings:** Further, the high pass decomposed image using wavelet transform over enhance the intensity variation that may be falsely detected and cancer characteristics leading to erroneous analysis. These limitations could be overcome by denoising the given input image using the wavelet transform and analysis made on inverse transformed image. The texture features should also be considered while analyzing an image for cancer detection. A back propagation neural network is trained using the mammogram images in different categories and tested using the sample as well as unknown images 13 feature neurons are used for N/W training and testing as well. **Application/Improvements:** The N/W is trained for normal images as well as abnormal cases. The classification accuracy has been observed to the tune of 89%.

**Keywords:** BPN, Image Segmentation, NN Classifier, Statistical Features, Texture Features

## 1. Introduction

Breast cancer is common among all type of cancers occurring in females. If not detected at initial stage, it leads to very fatal results. The analysis purely depends upon the image captured by using the acquisition devices. Further, the image analysis plays the crucial role in correctly predicting the different cancer stages. Mammography is one such popular technique to analyze the breast cancer in different stages. There is greater impact on population due to this cancer.

In year 2009, estimated latest cancer cases and deaths in Females were 192,370 and 40,170 respectively. In 2010, it increased to 207,090 and 39,840 respectively. Similarly, next year (in 2011) too new cancer cases increased to 230,480, but in 2012 it decreased to 226,870 and from last two years 2013 and 2014 it again start increasing i.e., 232,340 and 232,670 respectively<sup>1</sup>.

Mammogram image enhancement and detection is done through segmentation of the tumor area in an image of mammogram<sup>2-4</sup>. Further, various algorithms and techniques were proposed for segmentation of image<sup>5-7</sup>. A new method has been proposed to extract features named Square Centroid Lines Grey Level Distribution Method (SCLGM)<sup>8-10</sup>. Various methods for mammograms classification such as SVM, KNN, QDA and LDA can be obtained from previous works<sup>11-14</sup>.

## 2. Methodology

### 2.1 Image Dataset

An online MIAS database<sup>15</sup> is used for training as well as testing. These contains three main kind of classes: (a) Background tissue (F, G, D), (b) Abnormality (7

\*Author for correspondence

Categories) (c) Severity of abnormality (B, M, N) as shown in Figure 1.

## 2.2 Image Acquisition

It is the first stage of any vision system because without an image no further processing is possible. It is basically the action of retrieving an image from some hardware based source so that it can be passed through various processes. Samples of acquired images are shown in Figure 2.

## 2.3 Image Denoising Using Haar Wavelet Transform

At this step of image processing, noise is removed from the image using Haar Wavelet Transform. In this image matrix will be decomposed in four matrices of equal and size of each part will be half of it. The four matrices contains four type of data related to image i.e., one matrix contains approximate image, other contains noise and other two matrices contains edges and hence noise can be removed easily. Figure 3 shows the effect on image after applying Haar Wavelet Transform.

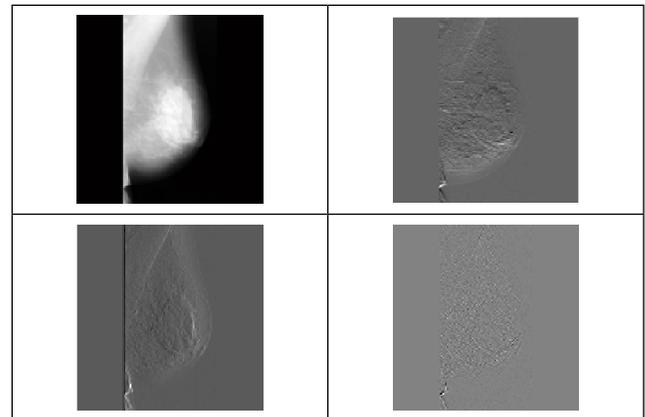


Figure 3. Haar Wavelet Transform Image.

## 2.4 Image Thresholding (Binarization) – OTSU Algorithm

Thresholding an image is a common preprocessing step. In this, the grayscale image is transformed into binary image. This conversion is mainly performed by making comparison of each pixel intensity with a threshold value. Here, this threshold value will be set by Otsu algorithm and then each pixel value will be replaced with a value that means “white” or “black” depending on the outcome of the comparison. It is the easiest method to separate foreground object from background using threshold. OTSU algorithm says maximize between class variance. A good threshold should separate pixels into tight clusters. Histogram is used for separating B and F, where histogram is a spread of intensity values across various pixels of image. In OTSU, threshold is decided by gray thresh automatically. Graythresh assumes histogram is bimodal histogram and then finds intensity value where it can be divided into two parts and then further calculations are performed.

## 2.5 Splitting Image into RGB Components

The de-noised image is now split into its R, G and B component image. Following color threshold are applied over R, G and B component images to segment the breast cancer spots in all color component images. Finally, the segmented color segments in R, G and B component images are concatenated back to get the final breast cancer spots. Using trail methods, following color threshold are applied:

- Threshold (R) = 30;
- Threshold (G) = 100;
- Threshold (B) = 20;

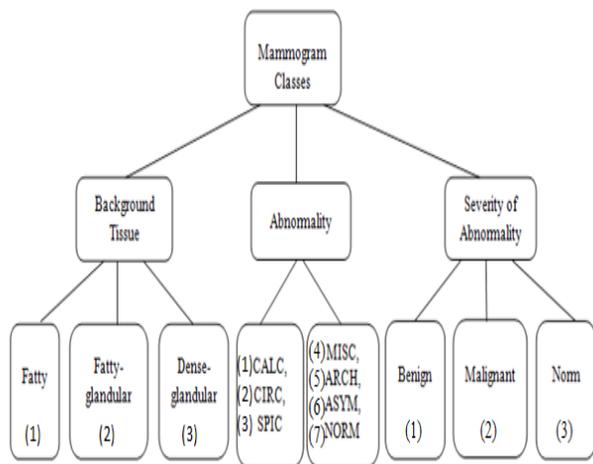


Figure 1. Classes in Mammograms.

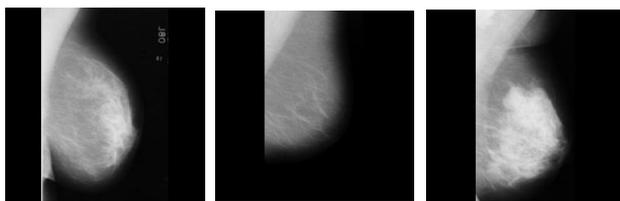


Figure 2. Acquired Images.

## 2.6 RGB to Gray Scale Conversion

The current image is in RGB format but to perform Binarization we need to convert it into Gray scale. Weighted mean is given by the following Eq. 1:

$$\text{New Gray scale image} = (0.3 \cdot R) + (0.59 \cdot G) + (0.11 \cdot B) \tag{1}$$

According to this equation, Red has contributed 30%, Green has contributed 59% which is greater in all three colours and Blue has contributed 11%. The RGB image is also called color image. It uses three 2D arrays of same size. One for R, other for G and third for B color. Each element has 8 bit value. So there are 24 bits in total in RGB image.

## 2.7 Segmentation

Segmentation is the process in which representation of an image is changed to something more useful and meaningful and to analyze the image become much easier. In this work, part is segmented from the original image which is containing cancer. Further classification is performed on that part after applying ROI.

The result of image segmentation is a set of segments that collectively cover the entire image, or a set of contours extracted from the image. Complete segmentation process to get segmented part is shown in Figure 4. A few segmented images are shown in figure 5.

## 2.8 Region of Interest (ROI) Extraction

The region of interest is extracted by mapping the coordinates of pixels in segmented part in above step to that of the original input image. This will extract the infected part from the original image and then feature extraction

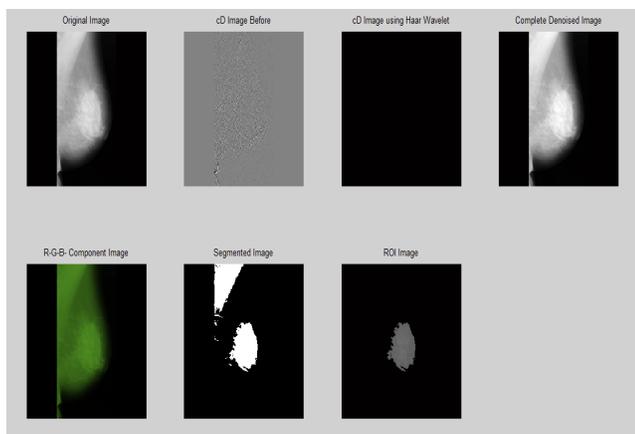


Figure 4. Complete Steps during Segmentation.

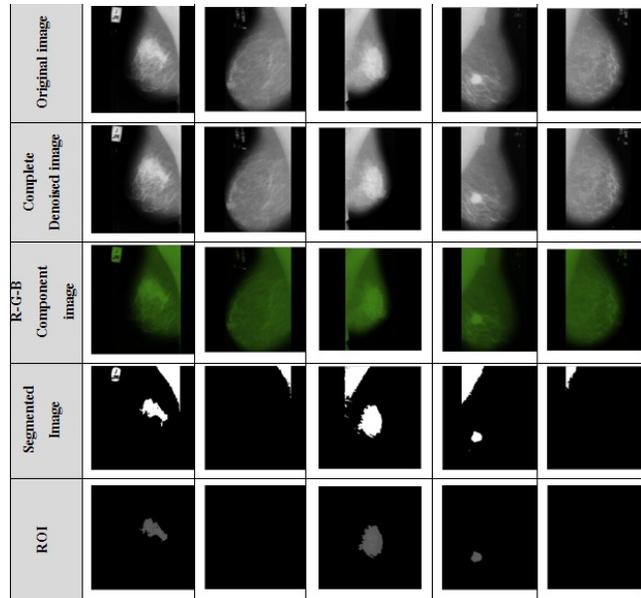


Figure 5. Sample of Segmented Images.

algorithm is applied only on the region of interest. The ROI is extracted from the original image using the pixel mapping between segmented images.

## 2.9 Features Extraction

- Texture features like Contrast, Energy, Homogeneity, and Correlation are computed.
- Statistical features like Entropy and Mean intensity are also calculated.
- Color moments like Standard Deviation (SDR, SDG, and SDB) and Mean (MR, MG, and

MB) of Red, Green & Blue Component, and SD Gray are manipulated.

## 2.10 Classification

A neural network is a system that reduces the operation of the human brain. In this network, processors in large number works parallelly, every processor apply its own small knowledge and make access data<sup>16</sup>. Neural network works in three layers.

**Layer 1:** It is the input layer. We have used here 13 nodes (features) in input layer.

**Layer 2:** It is the hidden layer. We have used here 2 layers.

**Layer 3:** It the last layer and called as Output layer.

We are using here back propagation neural network, as it is assumed to be work better than the feed forward neural network.

### 3. Results and Discussion

The presented work is for detection of breast cancer from images (mammograms). The images are acquired from online source. The images obtained are in pgm format and we have computed properties of GLCM matrix. The GLCM matrix properties are invariant to rotation as the matrix is contrast based variation in the input image. The algorithm gives vital information for early detection of breast cancer and features based analysis can be very helpful in pre and post treatment plan. For the testing and training purposes, 100 images are used. The algorithm is applied to all the images and the results are summarised in the below tables. A threshold value of all the properties may be set so as to predict the different stages of breast cancers. Samples of features extracted from images are shown below in Table 1.

**Table 1.** Features Extracted from the Images

Image No.	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
1	0.384	0.984	0.497	0.988	1.036	36.535	31.157	2.042	2.049	1.058	2.008	2.066	1.005
2	0.609	0.975	0.470	0.977	1.106	49.332	32.437	2.056	2.071	1.049	2.062	2.021	1.065
3	1.233	0.949	0.451	0.962	1.137	52.007	34.250	7.038	3.010	3.038	7.026	3.024	3.062
4	0.979	0.959	0.464	0.969	1.112	59.498	35.295	7.052	3.041	3.022	7.086	3.086	3.028
5	0.186	0.992	0.509	0.995	0.998	64.481	21.765	2.062	1.078	1.095	2.092	1.038	1.016
6	0.337	0.986	0.500	0.990	1.023	70.091	21.141	7.080	1.011	3.016	7.036	1.085	3.058
7	0.503	0.979	0.481	0.985	1.057	49.860	26.071	7.059	2.093	3.058	7.001	2.081	3.061
8	0.706	0.971	0.476	0.978	1.081	61.969	26.237	7.048	2.027	3.026	7.048	2.023	3.005
9	0.728	0.968	0.491	0.975	1.067	53.564	27.402	7.017	1.026	3.020	7.061	1.082	3.081
10	0.370	0.984	0.501	0.989	1.026	43.175	28.811	2.080	1.071	1.086	2.078	1.020	1.099

Table 2 describes the parameters used during training the samples and testing them. First we have trained a sample of 100 images with BPN. 13 features act as inputs and

**Table 2.** Neural Network Parameters

Sr. No.	Parameters	Values
1	No. of Input Neurons= No. of Features	13
2	No. of Training Samples	100
3	Target MSE	0.001
4	No. of hidden layers	2
5	Categories in BGT case	3 (1, 2, 3)
6	Categories in ABNORM case	7 (1, 2, 3, 4, 5, 6, 7)
7	Categories in SEV case	3 (1, 2, 3)

**Table 3.** Sample of Classification Results

Image No.	BPN O/P	Expected O/P	BPN	Exp.	Result
1	(1.963, 1.809, 0.850)	(2.000, 2.000, 1.000)	(2,2,1)	(2,2,1)	0
2	(1.623, 1.974, 0.801)	(2.000, 2.000, 1.000)	(2,2,1)	(2,2,1)	0
3	(2.795, 6.337, 2.732)	(3.000, 7.000, 3.000)	(3,7,3)	(3,7,3)	0
4	(2.384, 6.197, 2.777)	(3.000, 7.000, 3.000)	(0,7,3)	(3,7,3)	3
5	(0.818, 1.826, 0.922)	(1.000, 2.000, 1.000)	(1,2,1)	(1,2,1)	0
6	(0.884, 5.690, 2.932)	(1.000, 7.000, 3.000)	(1,7,3)	(1,7,3)	0
7	(1.612, 6.142, 2.372)	(2.000, 7.000, 3.000)	(2,7,0)	(2,7,3)	3
8	(1.892, 5.608, 2.725)	(2.000, 7.000, 3.000)	(2,7,3)	(2,7,3)	0
9	(0.998, 5.954, 2.972)	(1.000, 7.000, 3.000)	(1,7,3)	(1,7,3)	0
10	(0.986, 1.623, 0.982)	(1.000, 2.000, 1.000)	(1,2,1)	(1,2,1)	0

trained for 100 images. It is not possible to train all at once because they have subcategories. So we have trained them all through a function. Three back propagation neural networks are trained for BGT, ABNORM and SEV cases.

Further testing was performed using test images within the training samples. We have total set of 100 images out of which we have tested 100 images. Detailed Results obtained after applying NN classifier i.e., images are selected from the training samples and then tested them. We get number of correctly identified images as 89 and wrongly identified images as 11. That means accuracy becomes 89%, which is actually a good result. A sample of some results is shown in Table 3.

## 4. Conclusion

The proposed work is targeted to be implemented on mammogram images as obtained from online image data base. The speed of the algorithm primarily depends upon the image size and therefore is expected to vary from image to image. Haar wavelet is used for speedy image decomposition for image denoising. The denoised images are exposed to feature extraction algorithm followed by back propagation neural network system for classification. 13 feature neurons are used for N/W training and testing as well. The N/W is trained for normal images as well as abnormal cases. The classification accuracy has been observed to the tune of 89% and can be further improved with more training samples. Rigorous testing is performed taking test samples from within the training set and samples from out of the training set.

## 5. Acknowledgment

I would like to convey my gratefulness to MIAS data base for providing me sample images of mammograms. This paper has the assent of all co-authors and authorities of my institute, where this study has been carried out and there exists no conflict of interest anywhere.

## 6. References

1. American Cancer Society. Cancer Facts, Figures 2009-2016. Date Accessed: 17/08/2015; Available from: <http://www.cancer.org>.
2. Singh AK, Gupta B. A Novel Approach for Breast Cancer Detection and Segmentation in a Mammogram. *Procedia Computer Science*. 2015; 54:676–82.
3. Fiallos CB, Perez MG, Conci A, Andaluz VH. Institute of Electrical and Electronics Engineers: De Ambato U. T. Automatic Detection of Injuries in Mammograms Using Image Analysis Techniques. 2015; p. 245–48.
4. Gowri DS, Amudha T. A Review on Mammogram Image Enhancement Techniques for Breast Cancer Detection. *International Conference on Intelligent Computing Applications*. 2014; p. 47–51.
5. Sharma N, Mishra M, Shrivastava M. Colour Image Segmentation Techniques and Issues: an Approach. *International Journal of Scientific and Technology Research*. 2012 May; 1(4):9–12.
6. Guzman-Cabrera R, Guzman-Sepulveda JR, Torres-Cisneros M, May-Arrijo DA, Ruiz-Pinales J, Ibarra-Manzano OG, Avina-Cervantes G, Parada AG. Digital Image Processing Technique for Breast Cancer Detection. *International Journal of Thermophysics*, 2013 Sep; 34(8):1519–31.
7. Deepa S, Subbiah Bharathi V. Efficient ROI Segmentation of Digital Mammogram Images using Otsu's N thresholding method. *International Journal of Engineering Research and Technology (IJERT)*. 2013 Jan; 2(1):1-6.
8. Elfarrar BK, Abuhaiba ISI. New feature extraction method for mammogram computer aided diagnosis. *International Journal of Signal Processing, Image Processing and Pattern Recognition*. 2013 Feb; 6(1):13–36.
9. Arpana M, Kiran P. Feature Extraction Values for Digital Mammograms. *International Journal of Soft Computing and Engineering (IJSCE)*. 2014 May; 4(2):183–87.
10. Gaikar V, Mhaske R. Clustering of Breast Cancer Tumor using third order GLCM feature. *International Conference on Green Computing and Internet of Things (ICGCLOT)*. 2015 Oct; p. 318–22.
11. Elmannan ME. Implementation of Practical Computer Aided Diagnosis System for Classification of Masses in Digital Mammograms. *International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering*. 2015 Sep; p. 336-41.
12. Sasikala S, Ezhilarasi M, Sudharsan P, Sivakumari CLY. Performance Analysis of Various Segmentation Techniques in Breast Mammogram Images. *International Conference on Intelligent Computing Applications*. 2014 Mar; p. 228–32.
13. Armen Sahakyan HS. Segmentation of the Breast Region in Digital Mammograms and Detection of Masses. *International Journal of Advanced Computer Science and Applications (IJACSA)*. 2012; 3(2):102–5.
14. Snehal A Mane, Kulhalli KV. Mammogram Image Features Extraction and Classification for Breast Cancer Detection. *International Research Journal of Engineering and Technology (IRJET)*. 2015 Oct; 2(7):810–14.
15. Data Base of Mammograms. Date Accessed: 17/08/2015; Available from: <http://peipa.essex.ac.uk/info/mias.html>.
16. Shiffman Daniel 2016. The Nature of Code. Date Accessed: 05/02/2016; Available from: <http://natureofcode.com/book/chapter-10-neural-networks>.