



# Breast Cancer Detection and Recognition using Image Processing Technique

Diksha Nimgade<sup>1</sup>, Preetee Karmore<sup>2</sup>, Ms. Apeksha Meshram<sup>3</sup>

Student, Department of Computer Science and Engineering, DBACER, Nagpur, India<sup>1</sup>

HOD, Department of Computer Science and Engineering, DBACER, Nagpur, India<sup>2</sup>

Hiring Coordinator, Human Resource, Prevoyance Technologies, Nagpur, India<sup>3</sup>

**Abstract:** Breast cancer is the most common malignancy in women. It is often characterized by lack of early symptoms, which results in late detection of the disease. Breast cancer can occur in both men and women. It is projected that there will be rise in new cancer cases and cancer deaths in the United States. The early diagnosis of breast cancer can increase the survival rate, as it can promote timely clinical treatment to the affected patients. The mammography is the most common method for diagnosis through which we can have knowledge of abnormalities from the input image. This paper covers image processing based analysis of mammographic image. Firstly the image is converted from 3D to 2D array. Then the image is binarized twice, first to identify breast tissue and second to identify cancer affected region. Considering region of interest that indicates affected pixels and normal pixels the percentage of disease affected areas is calculated.

**Keywords:** Breast Cancer, Mammogram, Segmentation, Thresholding, Binarization, ROI, Python 3.

## I. INTRODUCTION

Cancer is a disease in which cells in the body grow out of control. Breast cancer is the common cancer in women. Breast cancer can begin in different part of the breast. A breast is made up of three main parts lobules, duct and connective tissue. The lobules are the glands that produce milk to the nipple. The connective tissues consisting of fibrous and fatty tissues surround and hold everything together. Most breast cancer begins in the ducts or lobules. Breast cancer can spread to other part of body, it is said to have metastasized.

The most common kind of breast cancer are:

- Invasive Ductal Carcinoma

The cancer cells grow outside the ducts into other part of the breast tissue. Invasive cancer cells can also spread or metastasize, to other parts of body.

- Invasive Lobular Carcinoma

Cancer cells spread from the lobules to the breast tissues that are close by. These invasive cancer cells can also spread to other parts of the body.

There are several other less common kind of breast cancer, such as Paget's disease, external icon medullary, mucinous, and inflammatory breast cancer external icon. Ductal Carcinoma In Situ(DCIS) is a breast disease that may lead to breast cancer. The cancer cells spread to other tissue in the breast.

Breast cancer now became most common and contributes to 14% of cancers in Indian women. It is reported that with every four minutes, an Indian women is diagnosed with breast cancer. Breast cancer is on the rise, both in rural and urban India. A 2018 report of Breast Cancer statistics 1,62,468 new registered cases and 87,090 reported deaths. Cancer survival becomes more difficult in higher stages of its growth, and more than 50% of Indian women suffer from stage 3 and 4 of breast cancer. With latest study report, India's highest cancer rate is listed in the state of Kerala. Other states with high cancer rates in India include Mizoram, Haryana, Delhi and Karnataka. Mizoram accounted for the highest cancer death rates in the country, followed by Kerala and Haryana. The numbers of cancer patients are staggering and constantly rising. As per the report release by Indian Council of Medical Research, India's cancer cases could increase by 12% by 2025. The cancer incidence (new cases) in men is estimated to be 7.63 lacs and in women 8.06 lacs by 2025. Breast cancer in women is estimated to contribute 200000 cases (14.8%) and cervix cancer cases about 75000(5.4%), of the total cancer burden. The survival rates of breast cancer in India are low because of the late detection.

To increase the survival rate, the early detection of Breast Cancer is most important. So that the treatable disease and chances of survival are higher if it's detected well before it becomes critical.

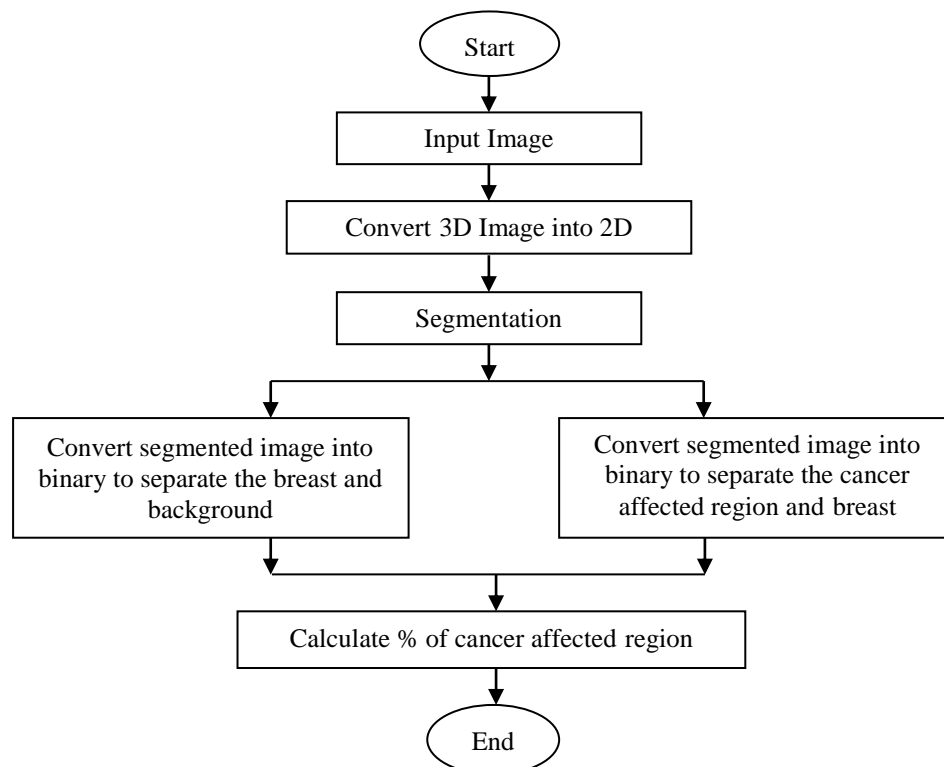
Mammography is the widely used technique for detection and recognition of breast cancer. It uses an x-ray system of a low-energy( usually 28kVp) to look inside the breast. Every small tumor and macro calcification can be detected using image processing technique to process on mammogram. Later an automated model is to be developed for detection of abnormal mass in breast.



## II. DATABASE RESOURCES

The image database used in this paper is the MIAS database. Mammogram Image Analysis Society is an organization of UK research groups that interested in the understanding of mammograms and generate database for research. The database contains 322 digitized mammograms and it is decreased to a 200-micron pixel frame and the images are 1024x1024. A community that consists of a sum of 321 mammographic images whereas 206 normal images, 51 malignant and 62 benign images. The MIAS database is collected in the Portable Grey Map setup. Each image is 8-bit grey level intensity range images among 256 various grey level intensity(0 – 255). In this analysis used 20 benign, 20 normal also 20 malignant mammograms that are thick and fatty glandular. All mammographic images are changed to word in .jpg format.

## III. METHODOLOGY



### A. Conversion of 3D(color) image into 2D(ayscale)

The mammogram image taken from the database is look like a grayscale(2D array) image but it is in color(3D array). For further processing, the image should be in grayscale. The RGB image can be read as three images(a Red scale image, a Green scale image, a Blue scale image)stacked on top of each other. RGB image is basically a  $M*N*3$  array(3D array) of pixel intensities level. The intensity level of each color Red, Blue and Green varies from 0 to 255. Similarly, A Grayscale image can be viewed as a single layered image. In python, a grayscale image is basically  $M*N$  array(2D array) whose value have been scaled to represent intensities. Our key idea is to convert RGB image pixel(3D array) which a triple value corresponding to Red, Blue and Green color component of image at a specified spatial location to a single value by calculating a weighted sum of all three color components.

- Algorithm for conversion

1. Read RGB color image into python environment.
2. Extract Red, Blue and Green color component from RGB image into 3 different 2D matrices.
3. Create a new matrix with the same number of rows and columns as RGB image, containing all zeros.
4. Convert each RGB pixel values at location(i, j) to grayscale values by forming a weighed sum of Red, Green and Blue color components and assign it to corresponding location(I, J) in new matrix.



*B. Segmentation*

Image segmentation is the process of partitioning the digital image into multiple segments(sets of pixels, also known as image objects). The aim of segmentation is to simplify and/or change the representation of an image into something which is more meaningful and clear to analyse. Image segmentation is the process of assigning a label to every pixel in an image. So that pixels with the same label share certain characteristics. The result of an image segmentation is a set of segments that covers the entire image, or set of counters extracted from the image. Each of the pixels in a region is similar with respect to some characteristics or computed properly, such as color, intensity, texture.

*C. Thresholding segmentation*

Segmentation in image processing plays an important role in detection the region of interest from background. It's input are images and the output are the properties/features from those images. Segmentation of intensity images(such as mammographic images)has four types: thresholding technique, boundary base method, region base method and hybrid technique that used both boundary and region criteria. The thresholding in mammogram images is based on separated the histogram into background and breast tissue. Depending on the value of threshold all pixels less than the threshold are classified as background, and the remaining pixels are breast.

The one of the best automatic thresholding selection technique is the Otsu method. Its concept based on selected the optimal threshold that maximizes the separability classes in gray level.

*D. Formulation*

To analyse and process any image we should know that an image is generated from a set of pixels denoted as  $n_i$

$$n = \sum_{i=0}^{L-1} n_i \dots\dots\dots(1)$$

Gray level histogram is normalize and regarded as a probability distribution:

$$h_i = \frac{n_i}{n} \dots\dots\dots(2)$$

The gray level of image is  $[0..L-1]$ . Where the gray level 0 is the darkest and the gray level L-1 is the lightest. The probability of occurrence of the two classes can be denoted as the following:

$$W_1(t) = \sum_{i=0}^t h \qquad W_2(t) = \sum_{i=t+1}^{L-1} h(i) \dots\dots\dots(3)$$

The mean and variance of the foreground and background are denoted respectively as the following.

$$\mu_1(t) = \sum_{i=0}^t ih(i), \qquad \sigma_1^2(t) = \sum_{i=0}^t (i - \mu_1(t))^2 h(i) \dots\dots\dots(4)$$

$$\mu_2(t) = \sum_{i=t+1}^{L-1} ih(i), \qquad \sigma_2^2(t) = \sum_{i=t+1}^{L-1} (i - \mu_2(t))^2 h(i) \dots\dots\dots(5)$$

To separate the object from the background specific thresholding algorithm is used for each image to get an optimal threshold.

*E. Otsu method*

Otsu method is based on established the optimal thresholding hold that(minimize intra-class variance) or maximize inter class variance. The following equations represents the within-class variance and the between class variance respectively.

$$\sigma_w^2(t) = \omega_1(t) \sigma_1^2(t) + \omega_2(t) \sigma_2^2(t) \dots\dots\dots(6)$$

$$\sigma_B^2(t) = \omega_1(t)(\mu_1(t) - \mu_T(t))^2 + \omega_2(t)(\mu_2(t) - \mu_T(t))^2 \dots\dots\dots(7)$$

The final form of between-class variance can also be denoted as the following.

$$\sigma_B^2(t) = \omega_1(t)\omega_2(t)(\mu_2(t) - \mu_1(t))^2 \dots\dots\dots(8)$$



Otsu method known as: it's simple, effective and easier to apply and choosing the optimal threshold is automatically and stably.

#### F. Disease percentage calculation

$$\text{Disease percentage} = (TP_a / (TP_a + TP_n)) * 100$$

Where,  $TP_a$  = Total number of cancer affected pixels

$TP_n$  = Total number of normal pixels

### IV. EXPERIMENTAL RESULT

Following are some mammographic images taken from the database. After applying above technique on images, following results has been obtained. Here, fig 1.1, fig .2.1, fig 3.1 and fig 4.1 are the original images saved in 3D array. The images are look like as a grayscale images but it is saved in 3D. For further processing it is converted in 2D in python environment.

Once the image is in 2D array, It goes to segmentation. Segmentation done twice using thresholding approach. The first to identifying breast tissue(or to separate breast from background). Using Otsu threshold method the threshold value is calculated automatically and it can separate breast from background successfully. Fig 1.2, fig 2.2, fig 3.2 and fig .4.2 are the binary images show the breast tissue.

The Second segmentation is to identifying cancer affected region(or to separate the tumor from background). To binaries the image and to decide threshold value we have tried whole range of the pixel intensity(0 to 255) and comes out to most perfect value to every image in the database. The threshold value is 190 which is applicable to every image in the database. Using this threshold value the binarization is done as shown in fig 1.3,fig 2.3, fig .3.3 and fig 4.3.

On the below binary images(image for cancer affected pixels)some pixels is not a cancer affected pixels it is included from light reflection. The top left and right cornered white pixels are only from the light reflection and the white pixels in circular portion is the cancer affected pixels. So to increase the accuracy of disease we have ignored cornered pixels. Fig 1.4, fig 3.4 and fig 4.4 show the region of interest(ROI) without light reflection.

Fig 2.3 is without light reflection, hence fig 2.4 will be remain same.

Disease Percentage calculation:

1. For fig 1.1  
Disease percentage =  $2341/39408 * 100$   
=5.940 %
2. For fig 2.1  
Disease percentage =  $29649/40930 * 100$   
=72.438 %
3. For fig 3.1  
Disease percentage =  $2025/38374 * 100$   
=5.277 %
4. For fig 4.1  
Disease percentage =  $5933/47191 * 100$   
=12.572 %



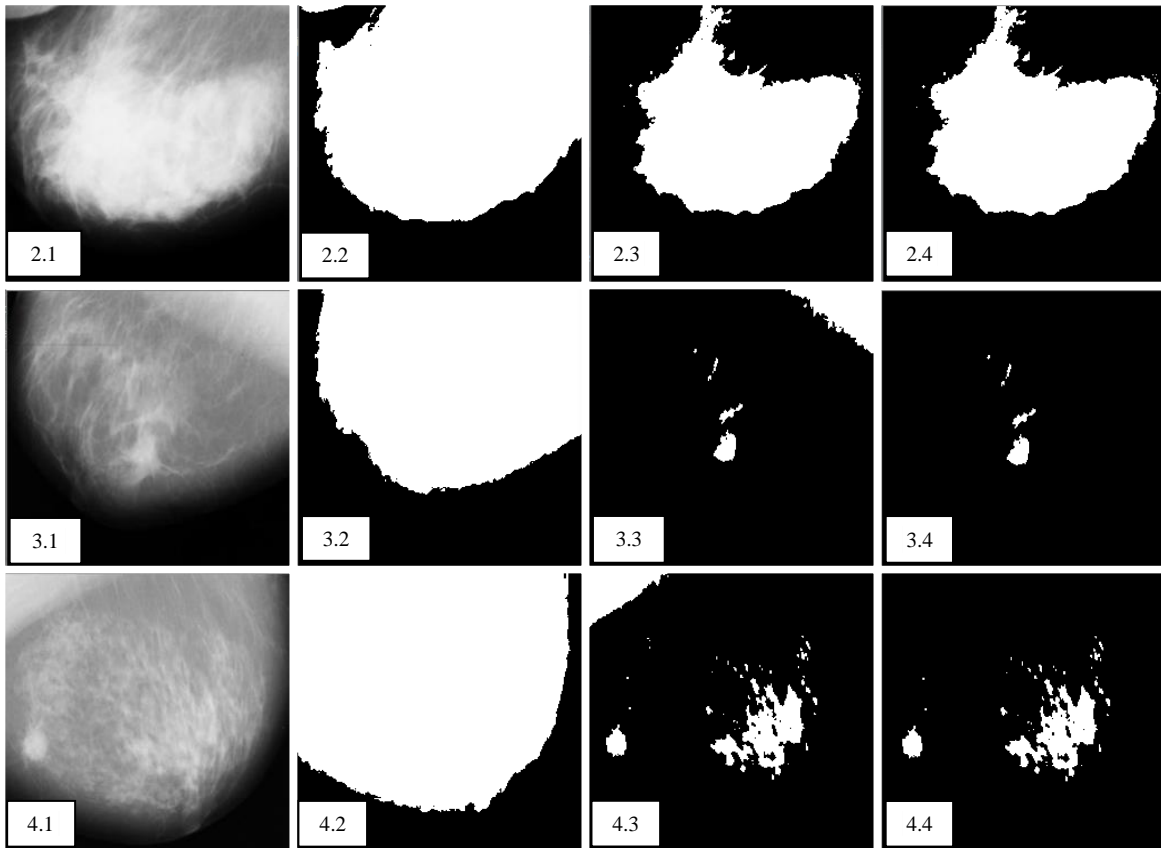


Image Resource :- <http://peipa.essex.ac.uk/info/mias.html>

Input Image taken from MIAS Database :- fig 1.1, fig 2.1, fig 3.1, fig 4.1

Binaries Image to identify Breast Tissue :- fig 1.2, fig 2.2, fig 3.2, fig 4.2

Binaries Image to identify Cancer Affected Region :- fig 1.3, fig 2.3, fig 3.3, fig 4.3

Removal of Light Reflection :- fig 1.4, fig 2.4, fig 3.4, fig 4.4

## V. CONCLUSION

In this paper, we have applied threshold method to binaries the image. The main purpose of this paper is to improve accuracy of disease percentage. As the accuracy is more it helps to increase survival rate. In some cases of mammogram images there may be light reflection on top left and right corner. That cornered white pixels are from light reflection and may not be cancer affected pixels. To find the accurate ROI(tumor) we have ignored that white pixels coming from the light reflection and calculated new ROI without reflection. Using this approach the most accurate results are obtained. These results are appreciated by the breast surgeon.

## REFERENCES

- [1]. S. Punitha, A. Amuthan, K. Suresh Joseph, "Benign and malignant breast cancer segmentation using optimized region growing technique", Future Computing and Informatics Journal 3(2018) 348-35
- [2]. Said Pertuz, German F Torres, Rrulla Tamini, Joni Kamarainen, "Open framework for mammography based cancer risk assessment", IEEE-EMBS International Conference on Biomedical and Health Informatics at Chicago(may 2019)
- [3]. Hongyu Wang, Jun Feng, Qirong Bu, Feihong Liu, Min Zhang, Yu Ren, Yi Lv, "Brest mass detection in digital mammogram based on gestalt psychology", Hindawi Journal of Healthcare Engineering volume 201
- [4]. A. Jemal, R. Siegel, E Ward et al, "Cancer statistics, 2008. CA: A cancer journal for clinicians", vol.58, no.2, pp. 71-96



- [5]. Z. Wang, G. Yu, Y. Zhao, Q. Qu, "breast tumor detection in digital mammography based on extreme learning machine", *Neurocomputing*, vol.128, no. 5, pp. 175-184
- [6]. Verma Brijesh, Zhang ping, "A novel neural-genetic algorithm to find the most significant of features in digital mammogram", *Applsoft comput* 2007;7:612-25
- [7]. Dheebea J, Tamil Selvi S, "A Swarm optimized neural network system for classification of microcalcification in mammogram", *J Med Syst* 2012;36:3051-61
- [8]. Parthima Guruprasad, "Overview of different thresholding method in image processing", *TEQIP Sponsored 3<sup>rd</sup> National Conference*, jun 2020
- [9]. Ali El-Zaart, Moumena Alhadithi, "Mammogram images thresholding for breast cancer detection using different thresholding methods. *Advance in Breast Cancer Research*",02(03).10.4236/abcr.2031.23013
- [10]. Shruti S, Binu Xavier, Ravi Kumar A, Yeshwanth S, Dr. Mahalinga V Mandi, "Breast cancer classification using python programming in machine learning", *International Journal of Engineering Research & Technology(IJERT)*, ISSN:2278-0181, vol. 9 Issue 08, August-2020
- [11]. Breast cancer fact and figures 2003-2004. American Cancer Society Simon Hadush Nrea, Yaacob Girmay Gezahegn, Abiol Sinamo, Gebrekirstos Hagos, "Breast cancer detection using convolutional neural networks", *Accepted paper at AI4AH, ICLR 2020*
- [12]. Kavneet Kaur, Ms. Navneet Kaur, Mr. Rashwinder Singh, "Breast cancer detection in mammogram images using Watershed and thresholding technique", *International Journal of Engineering Research and General Science* Volume 4, Issue 3, May-june,2016
- [13]. Ammari H, "Biomedical Imaging Modalities", *An Introduction to Mathematics and Emerging Biomedical Imaging* 2008, 198p. ISBN: 978-3-540-79552-0
- [14]. M. Garcia, A. Jemal, E. Ward, M. Center, Y. Hao, R. Siegel, M. Thun, *Global, Cancer Facts & Figures* (American Cancer Society, Atlanta, GA, 2007)
- [15]. T.M. Deserno (ed.), *Biomedical Image Processing* (Springer, Berlin, 2010)
- [16]. H.D. Cheng, X. Cai, X. Chen, L. Hu, X. Lou, *Pattern Recognit.* 36, 2967 (2003)
- [17]. M.J. Bottema, G.N. Lee, S. Lu, *Automatic image feature extraction for diagnosis and prognosis of breast cancer*, in *Artificial Intelligence Techniques in Breast Cancer Diagnosis and Prognosis*, Series in Machine Perception and Artificial Intelligence, vol 39 (World Scientific Publishing Co. Pte. Ltd., Singapore, 2000), pp. 17-54
- [18]. Shalinia G, Dr. T.N.R Kumar, "Comparison of Breast Cancer Detection using Probabilistic Neural Network and Support Vector Machine", *International Research Journal of Engineering and Technology*, volume 06, issue-6 June 2019