**Cancer Detection and Mammogram Segmentation**

Muhammad Rizwan Khadim

Computer Science Department, University of Management and Technology, Lahore

e-mail: rizwankhadim517@gmail.com

1. **ABSTRACT**

Mammography is a specialized imaging technique in the medical field for scanning breasts. Among women breast cancer is the major reason for the increased mortality rate. Manually approach takes long hours to examine a single mammogram for detecting a cancerous region. Cancer detection through mammograms needs an automated system that is specialized in identifying cancer at its earlier stages. In this paper, we proposed an automated system that segments the mammogram masses and extracts the texture features. We applied double-thresholding for segmentation purposes and the resultant image is then morphological transformed hence ROI is cropped out. After all preprocessing steps, several models are trained to classify the images into Benign, Malignant, or normal. Finally, the performance of the classifier is measured by different matrices including precision, accuracy, and F1 score.

1. **INTRODUCTION**

“*Breast cancer is a group of diseases in which cells in breast tissue change and divide uncontrolled, typically resulting in a lump or mass (Breast Cancer Facts* -National Breast Cancer Foundation n.d.)”. After lung cancer, breast cancer is the most commonly diagnosed cancer and the leading cause of death in women. Both men and women can have abnormalities in breast tissues, but stats show there are about 100 times more cases reported in women than men [[1].](#one) In 2018, about 2.1 million new cases

were reported with a death toll of up to 627,000 [[2]](#two). Among Asian countries, Pakistan has the highest incidence rate of breast cancer: “*one in every nine women has a lifetime risk of being diagnosed with breast cancer*” [[3]](#three).

“Mammography is specialized medical imaging that uses a low-dose x-ray system to obtain a grayscale picture of the breast region. A mammography exam, called a mammogram, is one of the most reliable and effective methods for detecting breast cancer at its early stages” [[4]](#four). The advancement in medical imaging technology led to the detection of breast cancer in its earlier stages and makes diagnosis easy. Radiologists examine mammograms for abnormality detection and classify them into benign and malignant, but sometimes they fail to differentiate between false positive and false negative. Computer-aided diagnosis (CAD) offers radiologists to achieve higher diagnostic accuracy and to locate breast cancer regions.

To detect anomalies from digitized mammograms automatically, it has to be passed through some preprocessing steps. The quality of a mammogram is enhanced by removing unwanted region which includes noise, image labels, and pectoral muscle.

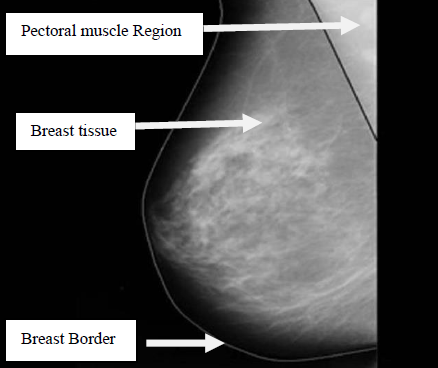


Fig. 1 Anatomy of women’s breast

Then images are sharpened which results in highlighting the breast tissues. So, image preprocessing is an indispensable part of CAD. After completing the preprocessing step, texture features of normal and abnormal breast tissues are extracted which are then later used to train the model. The model then classifies mammograms into abnormal and normal breasts.

1. **LITERATURE REVIEW**

R.S. Thakare et al. [[5]](#five) worked on breast cancer prediction and segmented the ROI with the Likelihood binarization algorithm. With adjusted thresholding values they separated similar pixels into color groups. They extracted a total of 14 Gabor features such as Mean, Standard Deviation, and GLCM from preprocessed images. Then used the SVM classifier. Samir M. Badawi et al. [[6]](#six) performed double thresholding and created a binary mask for segmentation and applied morphological transformation for removing noisy pixels. Muster et al. [[7]](#seven) removed pectoral muscle by estimating cubic polynomial and applying the canny edge detection technique. “In this technique, 10 random points are selected from the visible boundary of pectoral muscle which is then used for polynomial fitting of muscle boundary. Afterward, the cubic polynomial is used to estimate the remaining invisible pectoral muscle boundary”. Abdul Qayyum and A. Basit [[8]](#eight) created a binary mask and removed pectoral muscle by using straight-line estimation and canny edge detection techniques. Then extracted features using the statistical approach GLCM.

Sreedevi Sa and Elizabeth Sherly [[9]](#nine) proposed a novel approach to detecting pectoral muscle. They used Gray level thresholding, Canny edge detection, and 8 connected labeling technique.

1. **METHODOLOGY**

Our proposed methodology is shown in the block diagram below Fig. 2

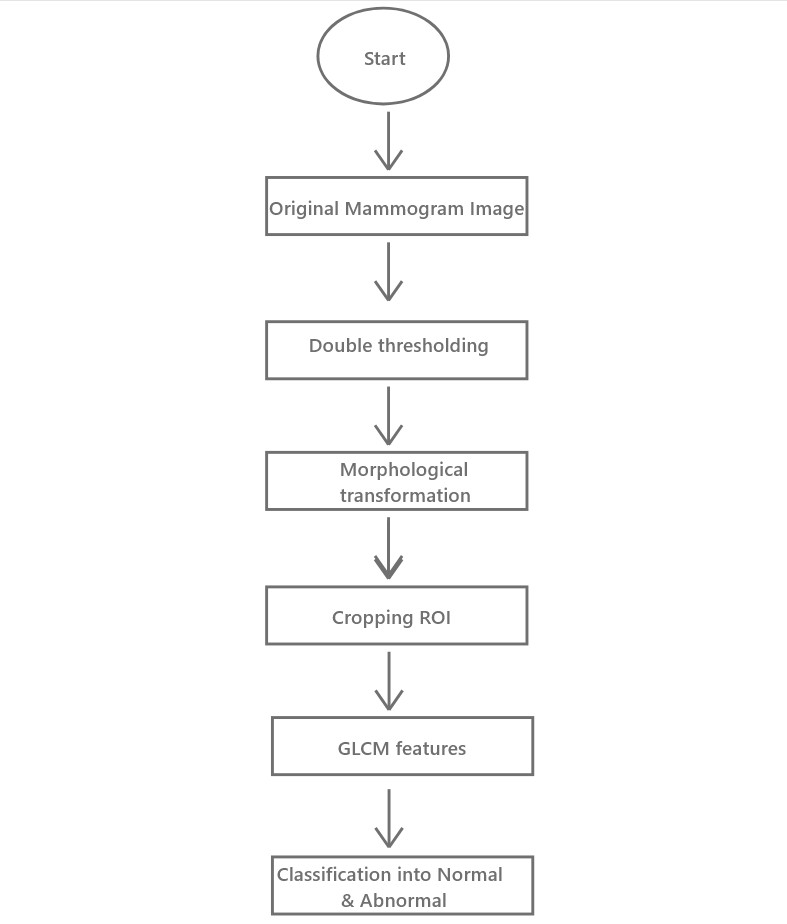


Fig. 2 Proposed Methodology

First of all, double thresholding is performed using the Otsu’s technique which is used in generating binary masks. This mask contains the breast region and other artifacts including labels and pectoral muscle. In order to remove labels from the mask morphological technique is applied onto the mask. A kernel of 7x7 is used to erode the mask and then dilated by using the kernel of the same size. This results in label-free images.

To crop the masked portion from the image bitwise is performed. The resultant images contain the breast portion attached to the pectoral muscle as

shown in fig. 3.

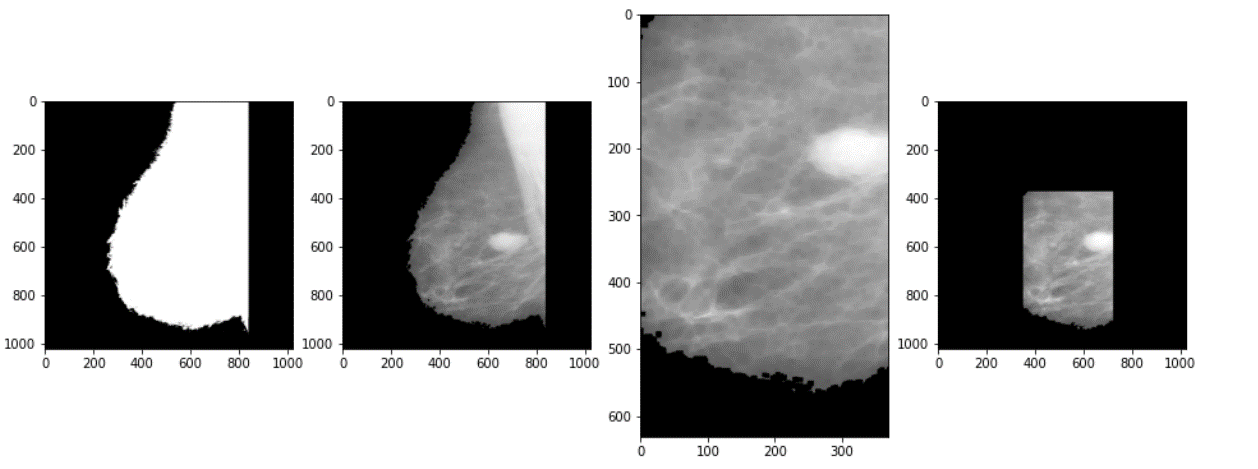


Fig. 3 Image segmentation process

To remove the pectoral muscle, we simply cropped the ROI with the most suitable coordinates that generalize the whole dataset by setting Width, Height, X, and Y to 370, 631, 350, and 376 respectively.

After the image segmentation and preprocessing we extracted the texture feature from the mammograms by using Gray Level Co-occurrence matrix (GLCM). We created a GLCM matrix at the angle of 0, 45, 90, and 135 degrees. The results are shown in fig. 4 below.

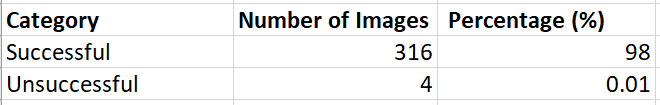


Fig. 4 Results for pectoral muscle segmentation

Six features i.e. contrast, correlation, energy, dissimilarity, ASM, and homogeneity were extracted from each GLCM matrix such that the length of

the feature vector was 1x24. These features were extracted from all the images (322) present in the dataset including both normal and abnormal tissues.

The resultant dataset is huge in dimensionality which might lead to under any classification model. Thus, extensive Principle Component Analysis (PDA) is performed for feature selection.

The dataset we used for this purpose is MINI DDSM containing 207 normal, 67 benign, and 51 malignant mammograms. All images at the resolution of 1024\*1024 pixels. All the images have a spatial resolution of 50 microns and are in MLO view.

1. **RESULTS AND DISCUSSION**

Logistic regression is used to train the model, the loss function we used for this classifier is binary cross-entropy and sigmoid as a step function. After the training the model, we yield an accuracy of 92.8% and an F1 score of 0.90. Furthermore, the best model can be achieved by reducing the dimensionality and setting the p-value to 0.4-0.5.

1. **CONCLUSION**

The proposed image segmentation has the potential to be used in CAD. Cropping the ROI by estimated co-ordinates data has shown systematically more effective than remove pectoral muscle by contouring.

Features are extracted by using GLCM features with 4 different degrees at a distance of 1. A supervised machine learning model is trained and accuracy is measured.

1. **REFERENCES**

[1] Samir M. Badawy, Alaa A. Hefnaw: Mammogram Segmentation A Qualitative Study

<https://www.researchgate.net/publication/320801723_Breast_Cancer_Detection_with_Mammogram_Segmentation_A_Qualitative_Study>

[2] Abdul Qayyum, A. Basit, “Automatic Breast Segmentation and Cancer Detection via SVM in Mammograms,” 978-1-5090-3552-6/16/$31.00 \_c 2016 IEEE

[3] Better Reporting and Awareness Campaigns Needed for Breast Cancer in Pakistani Women: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7936924/>

[4] AlirezaOsarech, BitaShadgar, “A Computer Aided Diagnosis System for Breast Cancer”, International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011

[5] R.S. Thakare, Dr. S. M. Deshmukh: “Automatic breast segmentation and cancer detection using SVM”. ISO 3297:2007 Certified pp. 2020

[6] Samir M. Badawy, Alaa A. Hefnaw: Mammogram Segmentation A Qualitative Study

<https://www.researchgate.net/publication/320801723_Breast_Cancer_Detection_with_Mammogram_Segmentation_A_Qualitative_Study>

[7] M. Mustra and M. Grgic, “Robust automatic breast and pectoral muscle segmentation from scanned mammograms,” Signal processing, vol. 93, no. 10, pp. 2817–2827, 2013.

[8] Abdul Qayyum, A. Basit, “Automatic Breast Segmentation and Cancer Detection via SVM in Mammograms,” 978-1-5090-3552-6/16/$31.00 \_c 2016 IEEE

[9] S Sreedevi, A Novel Approach for Removal of pectoral muscles in Digital Mammogram pp. 2015