



# Detection of Region of Interest and Stages using Mammogram images

Veena M<sup>1</sup>, M C Padma<sup>2</sup>, Dinesh M S<sup>3</sup>

PES College of Engineering Mandya, India<sup>1-3</sup>

**Abstract:** One of the risks of getting a medical diagnosis is getting breast cancer, and the number of people with it is growing all over the world. This causes the female to die day by day. The main goal of the work is to get the ROI and stages of the Breast Cancer Mammogram images. This work consists of three modules, i.e., preprocessing, segmentation and feature extraction and classification. Image is preprocessed by removing noise element and converting RGB to gray scale image. By applying morphological operation to obtain ROI for the tumor. This tumor is classified to find the stages of the tumor by multi SVM classifier.

**Keywords:** - Breast Cancer, Mammogram images, ROI, multi SVM classifier.

## 1. INTRODUCTION

One of the dangerous diseases that are spreading quickly is breast cancer. Women over 40 are more likely than younger women under 40 to get breast cancer. But a group of US researchers has found that 11% of new cases are in women who are younger than 35. On a scale from 0 to 4, there are four stages of breast cancer. Whereas stage 0 means that the cancer is not spreading and stays in the same place. Stage 4, on the other hand, is seen as an invasive.

**Tumor [T]:** The stages of the cancer are found after the ROI is found. In centimetres, the size of the tumour will be measured (cm).

T0 (T zero): Cancer is not found.

Stage1: 20mm or less in size.

Stage2: 20mm to 50mm.

Stage3: >50mm.

Stage4: It can be any size and spread all over the body.

## 2. LITERATURE SURVEY

ROI and breast cancer stages can be found in paper [1]. With this method, the stages of cancer are found by using the sobel edge detection mask and then Baye's classifier.

In paper [2], a Feature extraction of LBP texture feature, intensity value, texture, shape, and Eigen faces method is tested on an image that has already been processed. People use the support vector machine, the decision tree, and the KNN to put things into groups.

In paper [3]. Data is cleaned up and features are taken away after training. During this step, Analysis of Variance (ANOVA) is done in one direction. The last step is to use the WBCD (Wisconsin Breast Cancer (Diagnostic) Dataset) with SVM, KNN, and Decision tree.

In paper [4], they took a picture of a woman's breast with an MRI. Because breast skin is made up of big clumps of tissue, it can be hard for cancer cells to hide. Use the "tumor-free" method to get rid of this mess. The Confocal Microwave Imaging (CMI) algorithm can be used to find out where the tumour is in the breast.

In paper [5], a picture of the breast is used to make a mammogram. After the image has been preprocessed and a region of interest (ROI) of 128128 pixels has been extracted, the Dual-CT coefficient is calculated using the ROI image. KNN is a way to organise things.

In paper [6], the image is preprocessed to get rid of noise, label the image, and get rid of the background. The area of the tumour is used to choose the ROI. This tumour is found using texture analysis based on the Gray Level Co-occurrence Matrix (GLCM). Uses SVM classifier.

The Mammographic Image Analysis Society (MIAS) dataset is used in paper [7] to explain how this method works. Find the image's ROI after it has been preprocessed. The parts of GLCM are taken away.

In paper [8], mammogram images are preprocessed by taking out the region of interest (ROI) and making it the same size as the rest of the image. Find out if the image is normal or not by using the DCT transform. KNN is used to group things together.



From what has been written, it is clear that each method used to find breast cancer has its own problems. Still, there are more ways to improve how accurate breast cancer tests are. In this situation, the mammogram is suggested as a way to find ROI and different stages of breast cancer, and fig. 1 shows how the system is put together.

### 3. PROPOSED METHOD

We divide our work into preprocessing, segmentation, feature extraction and finding different stages.

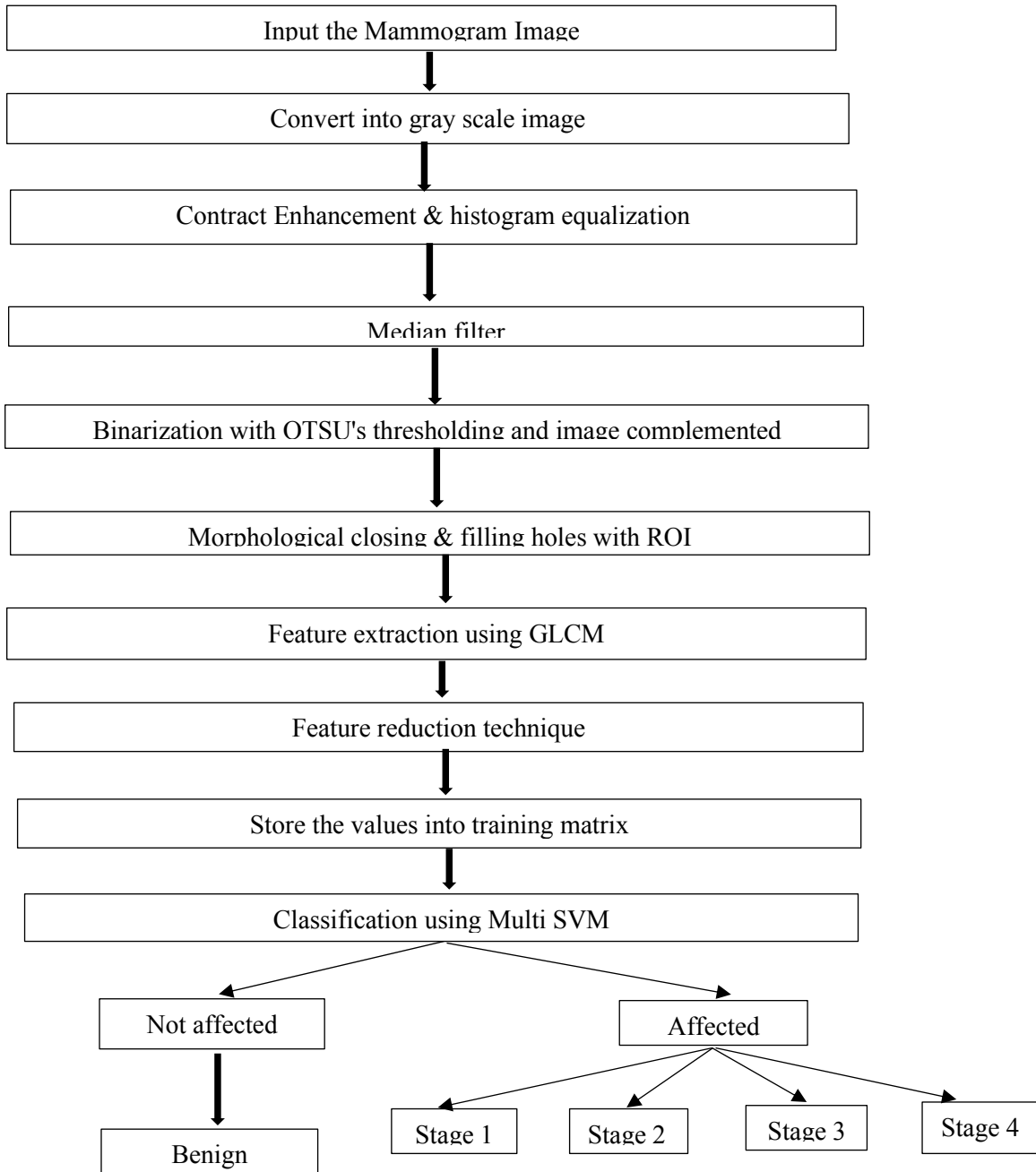


Figure 1: OVER ALL SYSTEM ARCHITECTURE

#### 3.1 Image Preprocessing

We use the median filter to get rid of noise in the image, then scale it down and turn it into a grayscale image, as shown in figure 2.

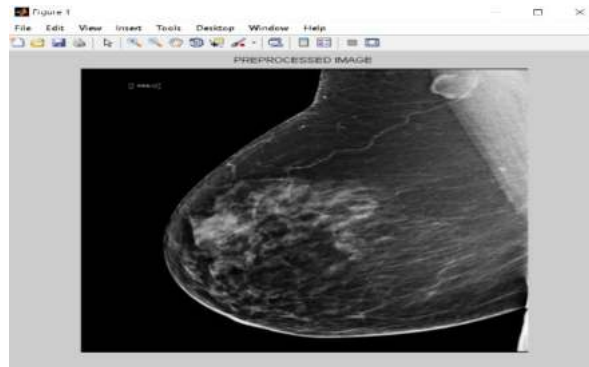


Figure 2: PREPROCESSED IMAGE

### 3.2 Image segmentation

The OTSU method is used to divide things up. OTSU Thresholding works well for faults that are long and narrow. We use morphological operation, as shown in Fig. 3, to get rid of background images.

Thresholding Methods come in two different kinds.

**Global Thresholding:** Grey values are used to set the global threshold values.

**Local Thresholding:** It divides original image into several sub regions

OTSU is a type of global threshold that depends only on the grey values of the image. Before the image is processed, it is broken up into black, grey, and white parts. The cluster is split into three parts based on the pixels.

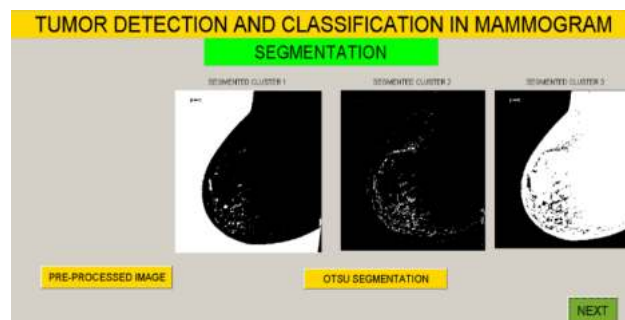


Figure 3: OTSU SEGMENTATION IMAGE

#### Morphological image processing:

With this operation, you can change how thick the border is around the image. Figure 4 shows the Dilation and Closing operation that we use for our work, and Figure 5 shows the extracted ROI.

**Dilation:** By adding pixels around an object's edges, dilation makes it bigger. This process depends on the building block. Dilation of an Image 'A' by a structured element 'B' is written by  $A \oplus B$ .

$$g(x, y) = \begin{cases} 1 & \text{if 'B' hits 'A'} \\ 0 & \text{Otherwise} \end{cases}$$

**Closing:** Most of the time, this operation gets rid of small holes and fills in spaces in the shape. Figure 4 shows it as  $A \bullet B$ .

$$A \bullet B = (A \oplus B) \ominus B$$

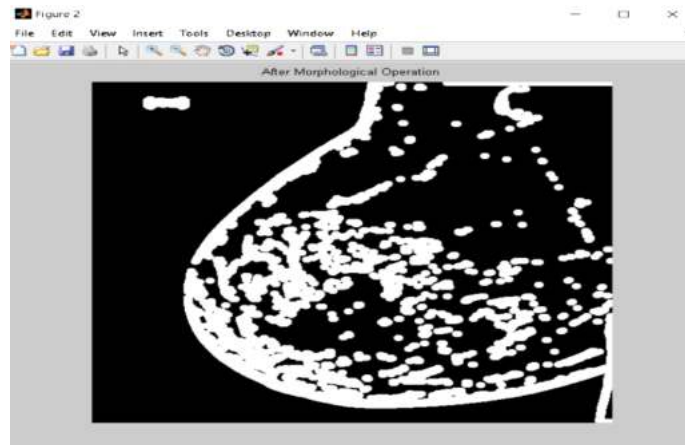


Figure 4: MORPHOLOGICAL OPERATION OF THE BREAST.

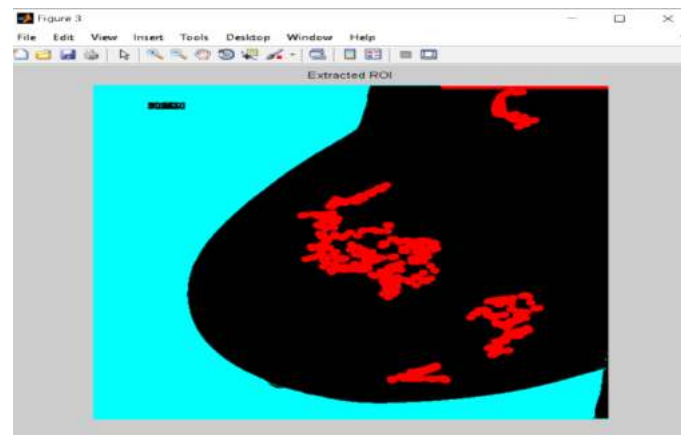


Figure 5: EXTRACTION OF ROI

### 3.3 Feature extraction and classification

We pull out features with the help of GLCM. After finding the ROI, the type of cancer can be determined. Table 1 explains how the formula for the GLCM works.

Table.1. FEATURE EXTRACTION FORMULAE

| Sl.No | GLCM Feature       | Formula  |
|-------|--------------------|--|
| 1.    | Contrast           | $\sum_{i,j=0}^{N-1} P_{ij} (i-j)^2$  |
| 2.    | Correlation        | $\sum_{i,j=0}^{N-1} P_{ij} \left[ \frac{(i-\mu_i)(j-\mu_j)}{\sqrt{(\sigma_i^2)(\sigma_j^2)}} \right]$          |
| 3.    | Dissimilarity      | $\sum_{i,j=0}^{N-1} P_{ij}  i-j $  |
| 4.    | Energy             | $\sum_{i,j=0}^{N-1} P_{ij}^2$  |
| 5.    | Entropy            | $\sum_{i,j=0}^{N-1} P_{ij} (-\ln P_{ij})$  |
| 6.    | Homogeneity        | $\sum_{i,j=0}^{N-1} \frac{P_{ij}}{1+(i-j)^2}$  |
| 7.    | Mean               | $\mu_i = \sum_{i,j=0}^{N-1} i(P_{ij}), \quad \mu_j = \sum_{i,j=0}^{N-1} j(P_{ij})$                             |
| 8.    | Variance           | $\sigma_i^2 = \sum_{i,j=0}^{N-1} P_{ij} (i-\mu_i)^2, \quad \sigma_j^2 = \sum_{i,j=0}^{N-1} P_{ij} (j-\mu_j)^2$ |
| 9.    | Standard Deviation | $\sigma_i = \sqrt{\sigma_i^2}, \quad \sigma_j = \sqrt{\sigma_j^2}$   |



### 3.3.1 METHODOLOGY

#### Algorithm for Feature Extraction and Classification

**Purpose:** Finding ROI and different stages.

**Input:** Mammogram Image.

**Output:** Stages of Breast Cancer with ROI.

```
stats = graycoprops(glems, 'Contrast Correlation Energy Homogeneity');
```

```
Contrast = stats.Contrast;
```

```
Correlation = stats.Correlation;
```

```
Energy = stats.Energy;
```

```
Homogeneity = stats.Homogeneity;
```

```
Mean = mean2(seg_img);
```

```
Standard_Deviation = std2(seg_img);
```

```
Entropy = entropy(seg_img);
```

```
RMS = mean2(rms(seg_img));
```

```
%Skewness = skewness(img)
```

```
Variance = mean2(var(double(seg_img)));
```

```
a = sum(double(seg_img(:)));
```

```
Smoothness = 1-(1/(1+a));
```

```
Kurtosis = kurtosis(double(seg_img(:)));
```

```
Skewness = skewness(double(seg_img(:)));
```

```
% Visualize Results
```

```
if result == 1
```

```
    R1 = 'T1 Type tumor ';
```

```
    set(handles.text35, 'string', R1);
```

```
    helpdlg(R1);
```

```
    disp(R1);
```

```
elseif result == 2
```

```
    R1 = 'T2 Type tumor ';
```

```
    set(handles.text35, 'string', R1);
```

```
    helpdlg(R1);
```

```
    disp(R1);
```

```
elseif result == 3
```

```
    R1 = 'T3 Type tumor ';
```

```
    set(handles.text35, 'string', R1);
```

```
    helpdlg(R1);
```

```
    disp(R1);
```

```
elseif result == 4
```

```
    R1 = 'T4 Type tumor ';
```

```
    set(handles.text35, 'string', R1);
```

```
    helpdlg(R1);
```

```
    disp(R1);
```

```
elseif result == 5
```

```
    R1 = 'Benign cancer ';
```

```
    set(handles.text35, 'string', R1);
```

```
    helpdlg(R1);
```

```
    disp(R1);
```

```
end
```

## 4. EXPERIMENTL RESULTS AND DISCUSSION:

There are many phases in different places (ROI). The dataset we used to figure out ROI and Stages came from both Kaggle and the UPMC Breast Tomography/FFDM collection database. Figure 6 shows the different stages.

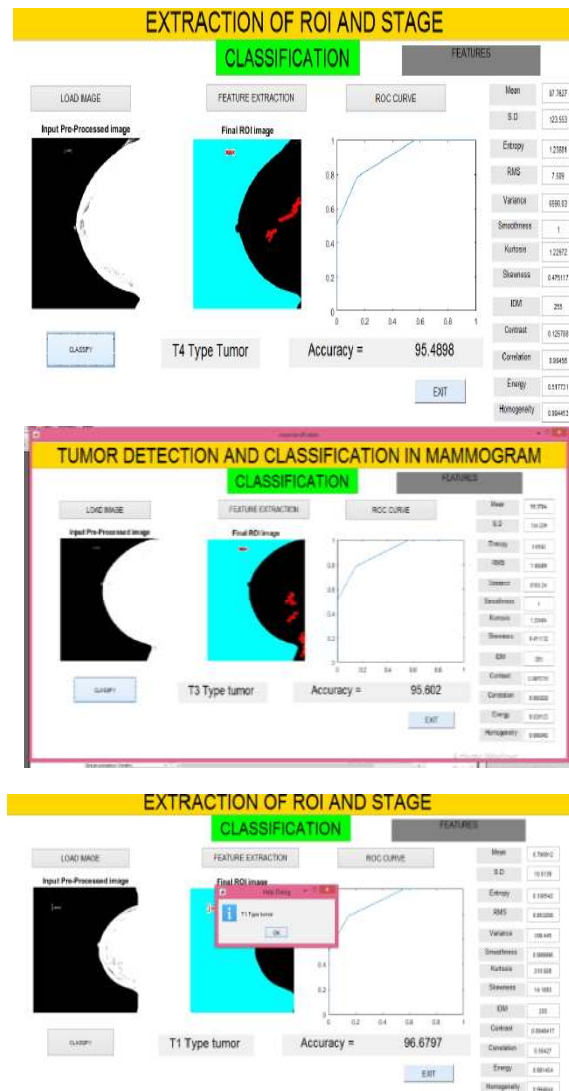


Figure 6: ROI AND DIFFERENT STAGES OF MAMMOGRAM IMAGES.

## CONCLUSION AND FUTURE ENHANCEMENT

Early detection of breast cancer is a challenge and there is a scope for improvement. Mammography is the best way to find breast cancer early, and a lot of people can be cured at the early stages. As a result, ROI and various stages of breast cancer can be detected using mammo-gram views. Four phases such as Pre-processing, segmentation, feature extraction, and classification will be used. MultiSVM and KNN classifiers are used, the accuracy obtained is 90% and 96%, respectively. Each modalities have their individual limits, therefore fusing features from two modalities i.e., Mammogram and DBT views and it is utilized to advance the precision in the detection of breast cancer.

## REFERENCES

- [1] A Novel Approach for Detection of Breast Cancer at an early stage using Digital Image processing techniques, Author : Sangeetha R, Dr. Srikanth Murthy K. (ICISC-2017) 2017 IEEE.
- [2] Study of Breast Cancer Detection Methods using Image Processing with Data Mining Techniques. Author: Dr D Devakummari, V.Punithavathi. IJPAM, ISSN:1311-8088 Volume 118 No 18 2018, 2867-2873.
- [3] A Comparative Study of Breast Cancer Diagnosis Using Supervised Machine Learning Techniques. Author: Madhuri Gupta & Bharat Gupta, IEEE, ISBN:978-1-5386-3452-3.
- [4] Early Breast Cancer Detection via UWB Based on MRI-derived Model. Author: Xia Xiao, Qinwei Li, Liang Wang. IEEE.



- [5] Early Detection and Classification of Breast Cancer. Author V.Vishrutha and M. Ravishankar, Vol,1 AISC 327.
- [6] Breast Cancer Diagnosis Using Image Processing and Machine Learning for Elastography Images. IEEE 2019.
- [7] A Research on Detection and Classification of Breast Cancer using k-means GMM and CNN Algorithms. Author: S. Shamy and J. Dheeba (IJEAT) August 2019.
- [8] Classification of Region of Interest in Mammograms using Dual Contourlet Transform and Improved KNN. Author: Min Dong, Zhe Wang, Cheghui Dong Hindawi 2017.
- [9] <https://www.cancer.net/cancer-types/breast-cancer/stages>.