EVALUATION OF DIFFERENT METHODS USED FOR REGION OF INTEREST EXTRACTION FROM MAMMOGRAM IMAGES

PROJECT DONE BY

LEENASRI R

BIOMEDICAL ENGINEERING

SSN COLLEGE OF ENGINEERING

1.Project Title : EVALUATION OF DIFFERENT METHODS USED FOR REGION OF EXTRACTION FROM MAMMOGRAM IMAGES.

2. Broad Subject :

Objective of this work is to evaluate the different methods that are used for region of interest extraction from mammogram images and to show how to increase the efficiency of the method that is more efficient in this evaluation.

3. Project Summary :.

Digital mammography detects various types of breast cancers. Many methods of analysing digital mammograms have been employed and yielded various success. Here we evaluate methods such as watershed transform(from region based method), K-means clustering(from clustering methods), edge detection method-sobel operator(from contour based approach) and dilation and erosion(from morphological based approach). These are simple methods that are used in Region of interest extraction. Then the overlap ratio, dice similarity index and SSIM was calculated.

4. Keywords : Watershed transform, K-means, sobel operator, morphological operations

5. Objectives :

- To collect and preprocess the mammogram images
- To segment the Region of interest by watershed transform, K-means clustering, edge detection and morphological methods.
- To superimpose the segmented ROI with the given mask to check whether the extracted ROI overlaps with the ground truth
- To find out the parameters like overlap ratio, Dice index and SSIM.

6. Introduction :

- The Digital mammography generally used to detect various types of breast cancer. In recent years, advanced imaging techniques, such as magnetic resonance imaging(MRI), ultrasound, and digital imaging have been used in breast cancer screening. But it hasn't been clear whether any of these techniques offered a better way of spotting breast cancer than standard mammography. While new techniques can help improve image quality or make diagnosis more precise, they haven't replaced the traditional mammography. The digital mammography allows radiologists to capture and manipulate the images so that the abnormalities can be seen more easily The procedure for a digital mammography is basically performed the same way as a standard mammogram.

- A region of interest is the portion of an image that you want to filter or perform other operation on. For example, in medical imaging, the boundaries of a tumor may be defined on an image or in a volume, for the purpose of measuring its size. There are many approaches that have been employed in image processing like region based approach, edge detection based segmentation, clustering based methods and morphological operations.

- REGION BASED APPROACH looks for similarities between two adjacent pixels. That is, pixels that possess similar attributes are grouped into unique regions. Here we have used watershed-based segmentation. It has its origin in mathematical morphology(theory and technique for analysing and processing of geometrical structures based on set theory, lattice theory, topology and random functions). In this method, image is regarded as topographic landscape with ridges and valleys. The watershed transform decomposes image completely and thus assigns each pixel either to a region or a watershed. The result of watershed algorithm is global segmentation, and it requires low computation time, fast simple and intuitive method. One of the major drawback of watershed segmentation is the over segmentation(overcutting that occurs during segmentation).

- CLUSTERING BASED APPROACH is to partition an image into number of disjoint groups or clusters. For example if the image holds apple and orange, most of the pixel points in the apple should be red or green, which is different form the pixel values of orange. Here we have used K-means algorithm. It is a method of vector quantization,

originally from signal processing, that aims to partition n observations into K clusters in which each observation belongs to the cluster with the nearest mean, serving as the prototype of the cluster. They are relatively simple to implement, easily adapts to new examples, scales to large data sets, generalizes to clusters of different shapes and sizes such as elliptical clusters. The disadvantages of k-means clustering is that it always depends on initial values, it has trouble in clustering data of varying size and density.

- EDGE DETECTION BASED SEGMENTATION is the most familiar approach for detecting significant discontinuities in intensity values. Edges are local changes in the image intensity Edges typically occur in the boundary between two regions. Here we used sobel operator. Sobel operator sometimes called the Sobel-Feldman operator or sobel filter, is used in image processing, particularly within the edge detection algorithms where it creates an image emphasizing edges. The main advantage of sobel operator is its simplicity which is because of the approximate gradient calculation. The major disadvantage of this operator is the signal to noise ratio(LOW).

- MORPHOLOGICAL OPERATIONS is the broad set of image processing operations based on shapes. In a morphological operation, the value of each pixel in the output image is based on the comparison of the corresponding pixel in the input image with its neighbours. The morphological operators are simple and time efficient algorithms. They rely only on the relative ordering of pixel values, not on their numerical values and therefore they are suitable for processing the binary image. The fundamental morphological operators are dilation and erosion. Dilation adds pixels to the boundaries of object while erosion removes pixels on object boundaries.

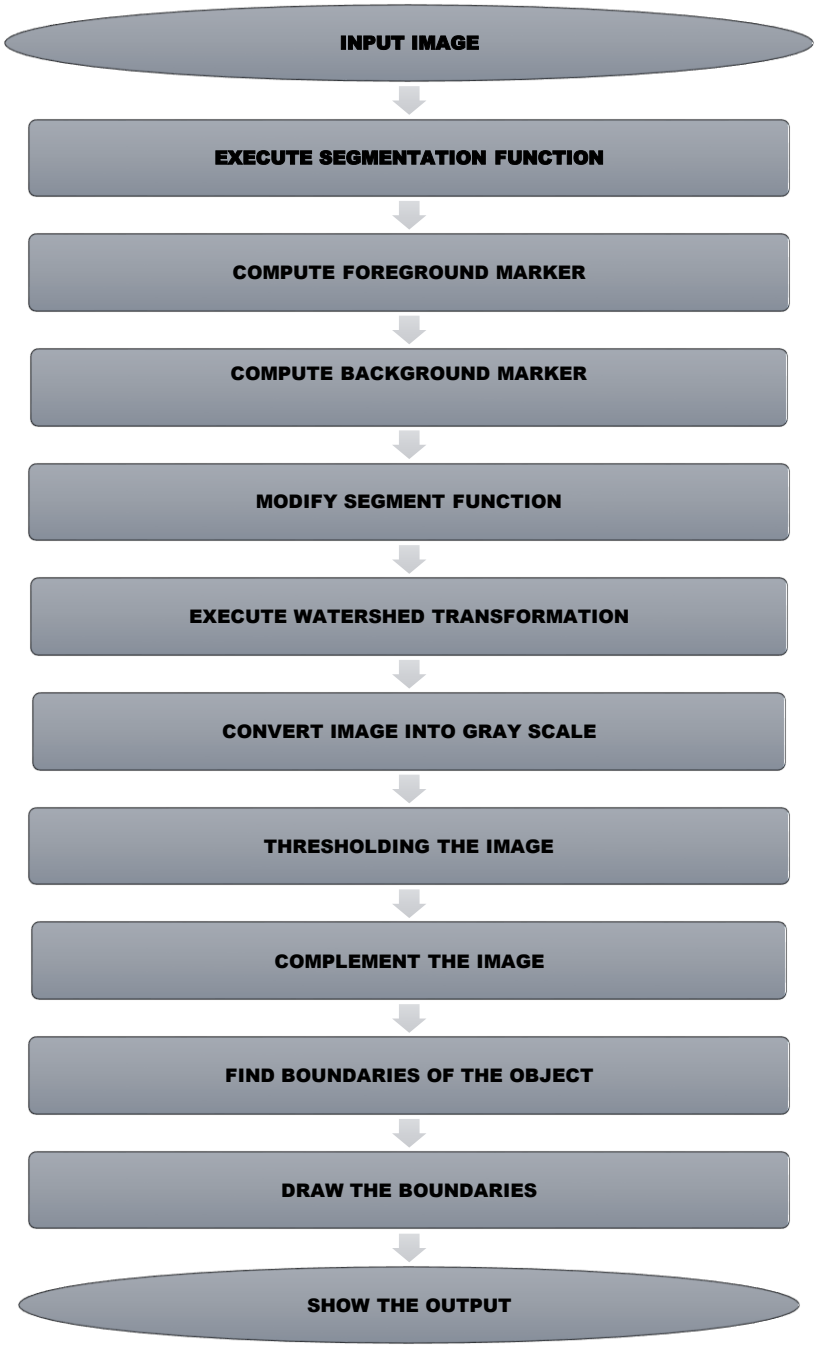
- The common advantage in all the above method is they are simple and time efficient. If these methods are used the diagnosis time is reduced. So in this paper we see the most effective method among these and a method to improve its efficiency further more.

7. Definition of the Problems :

Breast cancer is one of the leading cancers among women. In India breast cancer accounts for 14% of cancers in Indian women. It is reported that for every 4 minutes an Indian women is diagnosed with breast cancer. A 2018 report of Breast Cancer statistics recorded 1,62,468 new registered cases and 87,090 reported deaths. So early diagnosis of breast cancer is very important. One of the major part in this process in REGION OF INTEREST segmentation. It should be a time efficient, simple and effective method so that the diagnosis is done effectively and quickly. So here we evaluate the methods used for region of interest extraction and we have discussed a method to improve its efficiency.

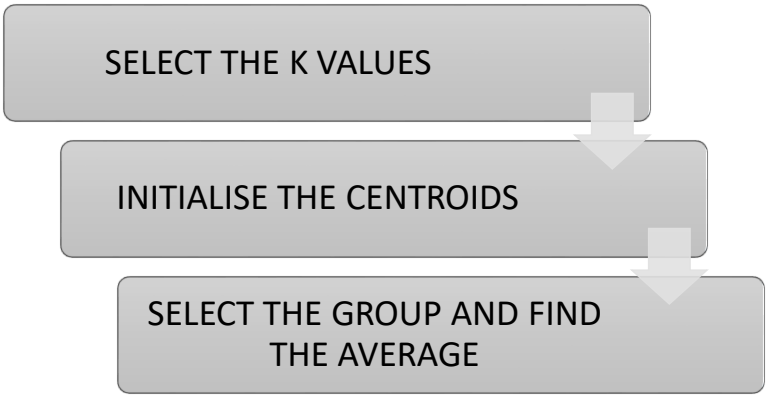
8. DETAILED EXPLANATION OF METHODS USED:

8.1. WATERSHED TRANSFORM:

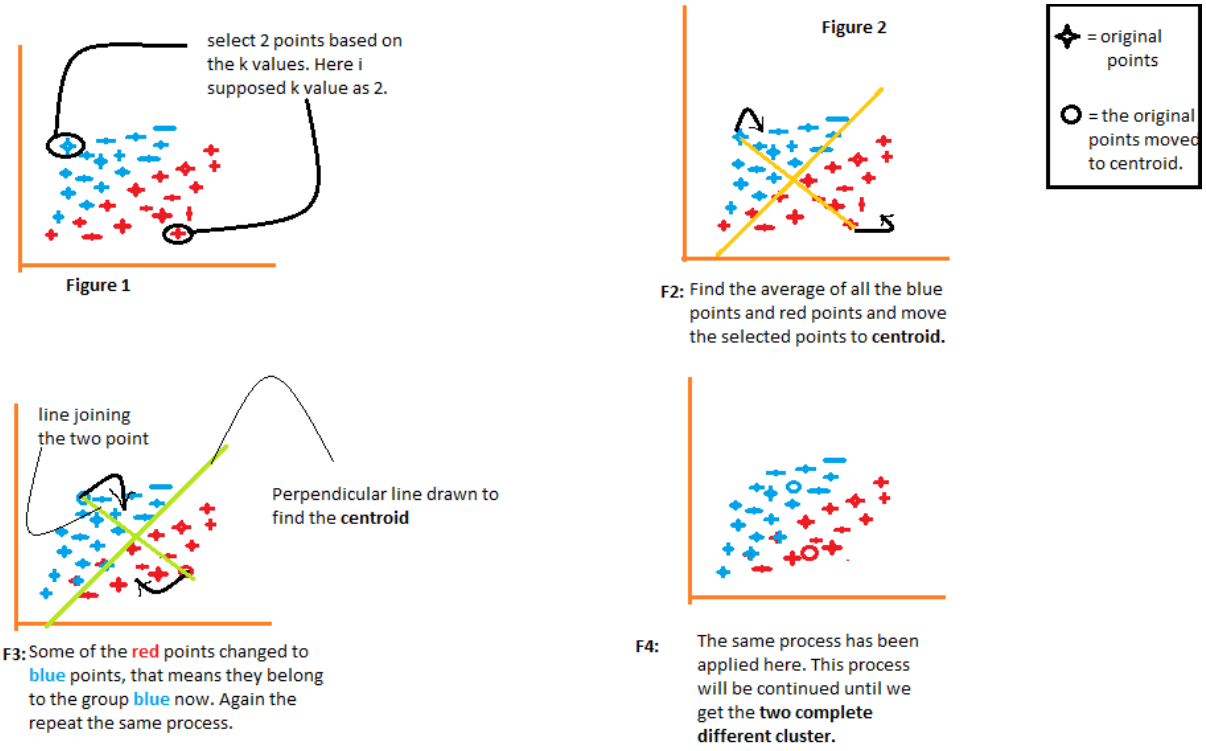


8.2. K-MEANS CLUSTERING:

k-means clustering tries to group similar kinds of items in form of clusters. It finds the similarity between the items and groups them into the clusters. K-means clustering algorithm works in three steps. Let’s see what are these three steps.



Let us illustrate this with a example:



- Figure 1 shows the representation of data of two different items. the first item has shown in blue color and the second item has shown in red color. Here I am choosing the value of K randomly as 2. There are different methods by which we can choose the right k values.
- In figure 2, Join the two selected points. Now to find out centroid, we will draw a perpendicular line to that line. The points will move to their centroid. If you will notice there, then you will see that some of the red points are now moved to the blue points. Now, these points belong to the group of blue color items.

- The same process will continue in figure 3. we will join the two points and draw a perpendicular line to that and find out the centroid. Now the two points will move to its centroid and again some of the red points get converted to blue points.
- The same process is happening in figure 4. This process will be continued until and unless we get two completely different clusters of these groups.

8.3. SOBEL OPERATOR:

This operator uses two 3x3 kernels which are convolved with original image to calculate approximations of the derivatives- one for horizontal changes and one for vertical. If we define A as a source image, and G_x and G_y are two images which at each point contains the horizontal and vertical derivative approximations respectively, the computation.

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

$$\mathbf{G} = \sqrt{\mathbf{G}_x^{\,2} + \mathbf{G}_y^{\,2}}$$

Magnitude

8.4. MORPHOLOGICAL OPERATORS:

1. **Erosion :** Erosion is one of the basic operators in the area of morphology, the other being dilation. It is typically applied to binary images, but there are versions that work

on grayscale images. The basic effect of the operator on a binary image is to erode away the boundaries of regions of foreground pixels (i.e. white pixels, typically). Thus areas of foreground pixels shrink in size, and holes within those areas become larger. In erosion operator takes two pieces of data as inputs. The first is the image which is to be eroded. The second is a (usually small) set of coordinate points known as a structuring element (also known as a kernel). It is this structuring element that determines the precise effect of the erosion on the input image. In mathematical term erosion is defined as- For image A and structuring element B in the Z^2, erosion is defined as

$$A \ominus B = \{z \mid (B)_z \cap A^C \neq \phi\}$$

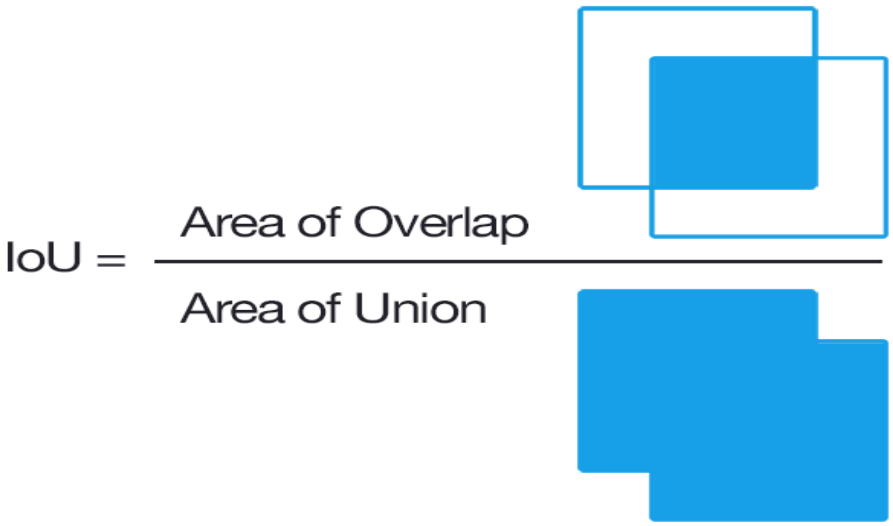
This equation indicates that erosion of A by B is the set of all points Z such that B, translated (shifted by z), is a subset of A i.e., B is entirely constant with A. B is entirely contained within A. Erosion reduces the number of pixels from the object boundary. The number of pixels removed depends on the size of structuring element.

2. **Dilation:** The basic effect of the operator on a binary image is to gradually enlarge the boundaries of regions of foreground pixels (i.e. white pixels, typically). Thus areas of foreground pixels grow in size while holes within those regions become smaller. The dilation operation usually uses a structuring element for probing and expanding the shapes contained in the input image. Dilation is defined as where A is the image and B is the structuring element. Hence dilation of A with B is a set of all displacement, Z , such that (B) and A overlap by at least one element. Dilation adds pixel to the boundaries of objects in an image. The number of pixels added depends on the shape size of the structuring element. Here is the formula for dilation

$$A \oplus B = \left\{ z \mid (\hat{B})_z \cap A \neq \phi \right\}$$

9. PARAMETERS USED TO DETERMINE THE EFFICIENCY:

- **OVERLAP RATIO:** To find the ratio of overlap between the extracted ROI and ground truth



- **DICE INDEX:** The Sørensen–Dice coefficient is a statistic used to gauge the similarity of two samples**.**

  FORMULA: $DSC(A,B)= 2(A \cap B)/(A+B)$

- **SSIM:** The Structural SIMilarity (SSIM) index is a method for measuring the similarity between two images.

  FORMULA: $r^*(x, y) = \sigma_{xy}\sigma_x\sigma_y$

  when $\sigma_x\sigma_y \neq 0$, 1 when both standard deviations are zero, and 0 when only one is zero.

10.DIGITAL IMAGE PROCESSING USING MATLAB MATLAB (MATrix LABoratory) :

It is software, which provides an interactive environment for numerical computations and graphics analysis. This software is especially designed and used for mathematical analysis using matrix computations. Also this software has a large variety different tool boxes having graphic solving capabilities and it can be extended using programming, which are written in its own programming language. MATLAB has a large collection of toolboxes in a variety of

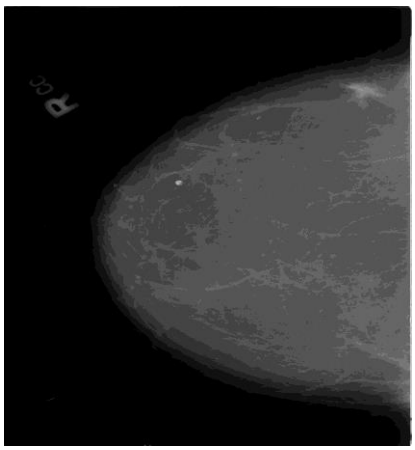
domains. Some examples of MATLAB toolboxes are control system, signal processing, neural network, image processing, and system identification. The toolboxes consist of functions that can be used to perform computations in a specific domain.

11. IMPORTANCE OF THE PROJECT IN THE CONTEXT OF CURRENT STATUS

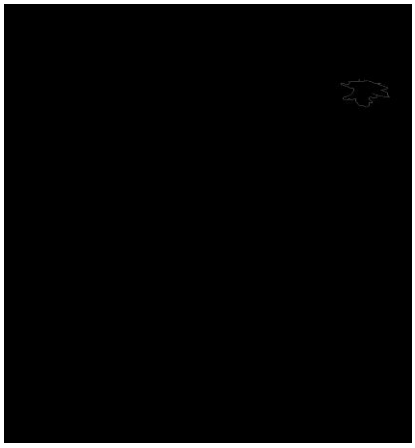
The breast cancer among women is increasing very rapidly. Digital image processing techniques help in manipulation of the digital images through the use of computers. The processing of images is faster and more cost-effective in this method. One needs less time for processing, as well as less film and other photographing equipment. It is more ecological to process images. No processing or fixing chemicals are needed to take and process digital images. It is important to diagnose the breast cancer in less time and in a very efficient manner. This evaluation is important in current situation because the methods discussed here are very simple and time saving method. The efficient method is found out and a way to increase its efficiency further more is also discussed. So this will be very important in future diagnosis of breast cancer.

12. DATABASE DETILS: *(MINI-DDSM)*

This is the light weight version of the popular DDSM(Digital Database for Screening Mammography). The data set comes along with the age/density attributes, patient folders, original identification filename, and lesion binary mask.

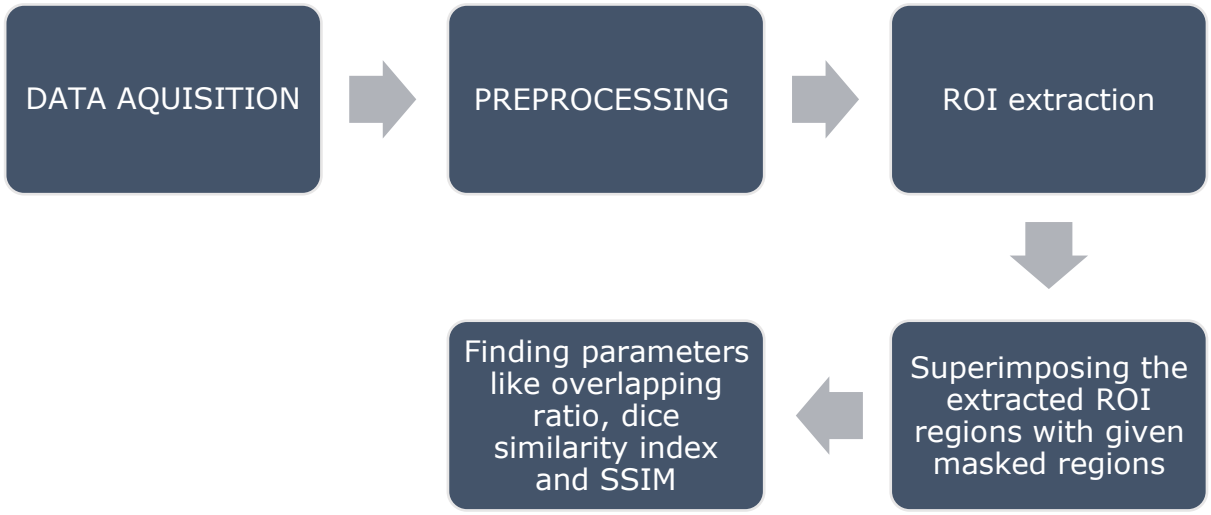


MAMMOGRAM IMAGE



MASK(GROUND TRUTH)

13. METHODOLOGY:



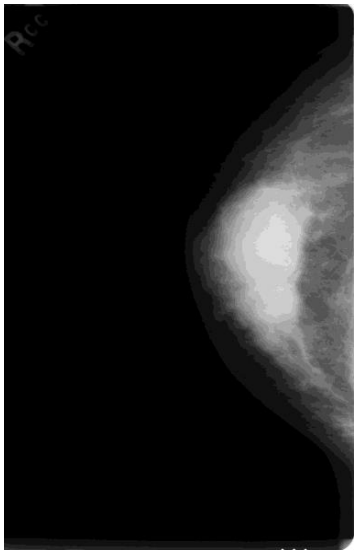
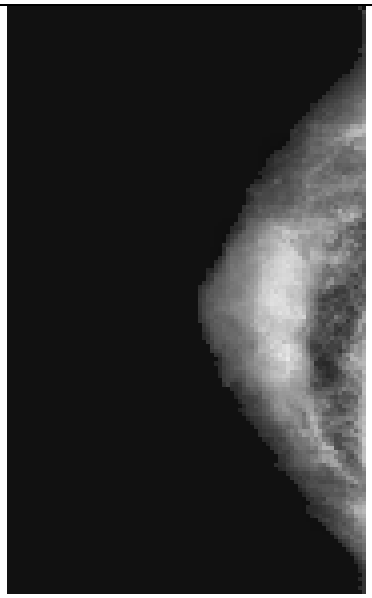
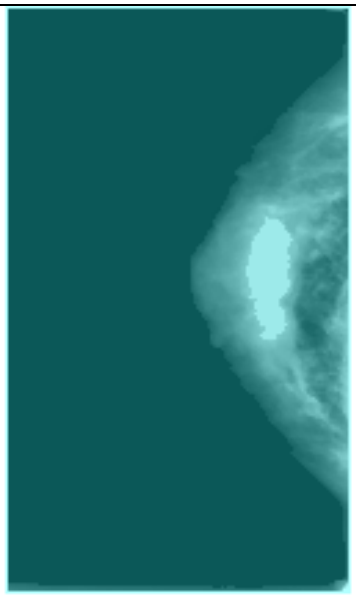
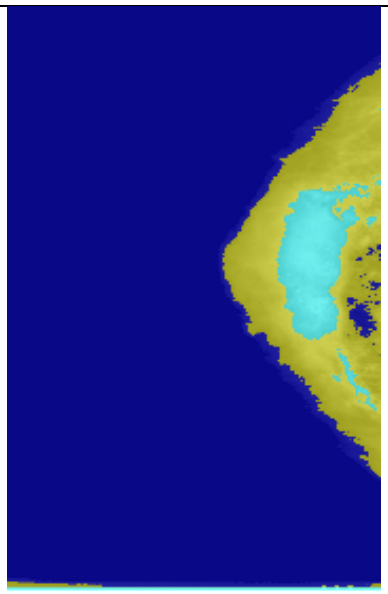
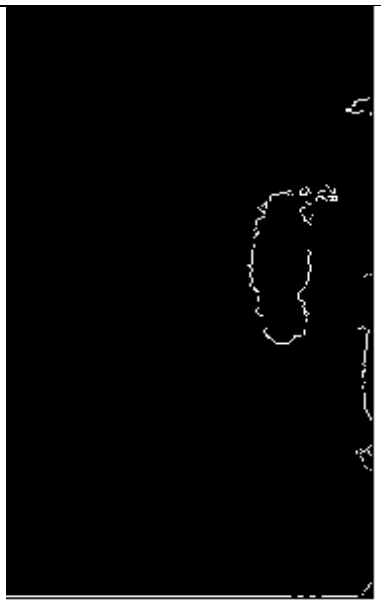
13.1. PREPROCESSING:

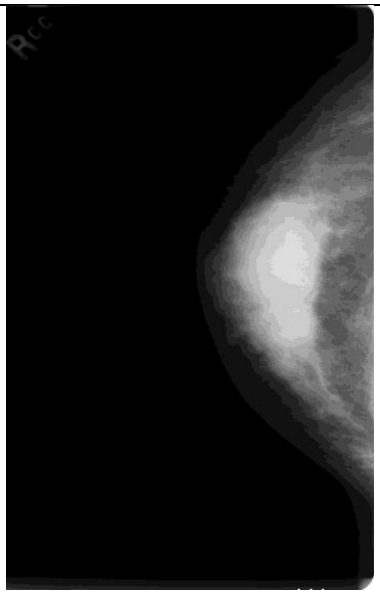
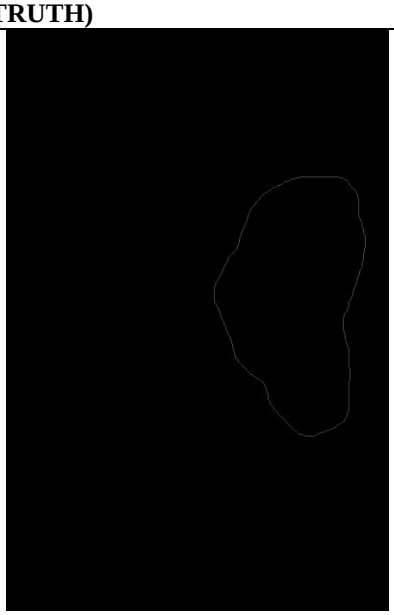
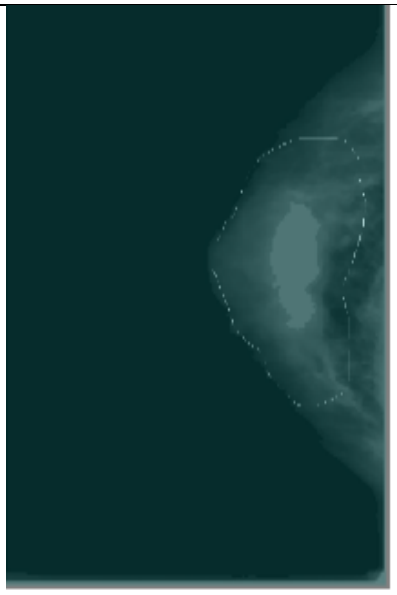
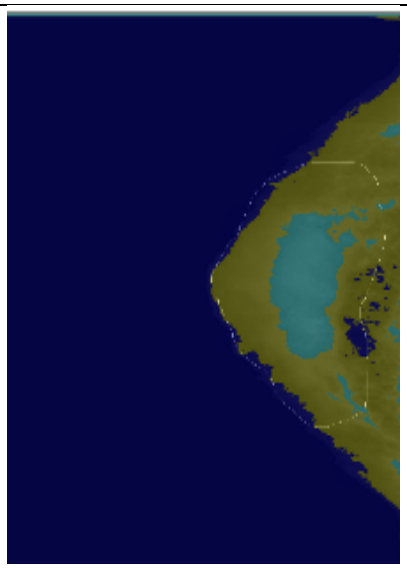
- Noise removal using Median filter.
- Contrast enhancement using CLAHE(Contrast Limited Adaptive Histogram Equalization).

13.2. REGION OF INTEREST:

- The Region of Interest is extracted by the methods mentioned above and the parameters like overlapping ratio, dice similarity index and SSIM are calculated

14. RESULTS:

ORIGINAL IMAGE	PREPROCESSING	WATERSHED ROI
C_0005 		
KMEANS ROI	**ROI BY DETECTING EDGES**	**ROI BY DILATON AND EROSION**
		

ORIGINAL IMAGE	MASKED IMAGE (GROUND TRUTH)	SUPERIMPOSED IMAGE WITH WATERSHED ROI
		
SUPERIMPOSED IMAGE WITH KMEANS ROI	**SUPERIMPOSED IMAGE WITH ROI BY DETECTING EDGES**	**SUPERIMPOSED IMAGE WITH ROI BY DILATON AND EROSION**
		

2D PLOT

MASK PLOT	WATERSHED PLOT	K-MEAN PLOT
[555, 230, 395, 478]	[682, 356, 172, 224]	[659, 320, 220, 302]
SOBEL OPE. PLOT	**MORPHOLOGICAL OPE. PLOT**	
[652, 290, 216, 298]	[603, 320, 317, 287]	

OVERLAP RATIO

WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOGICAL OPERATION
0.2041	0.3519	0.3409	0.4819

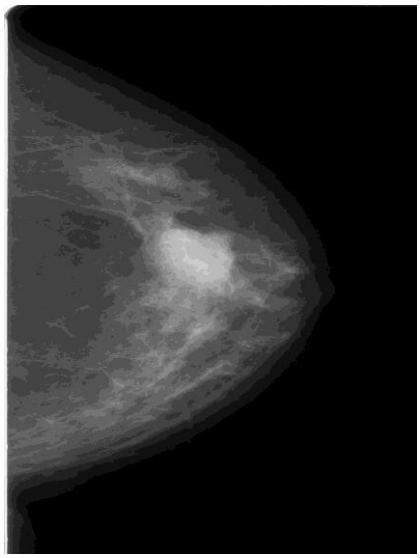
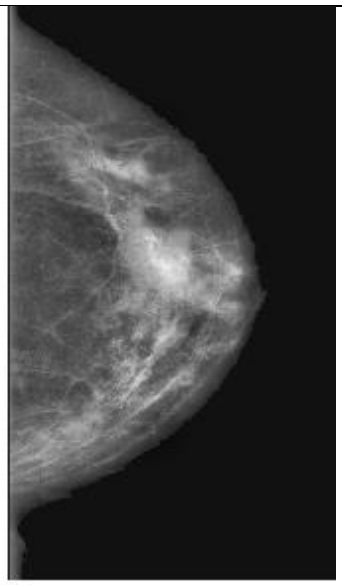
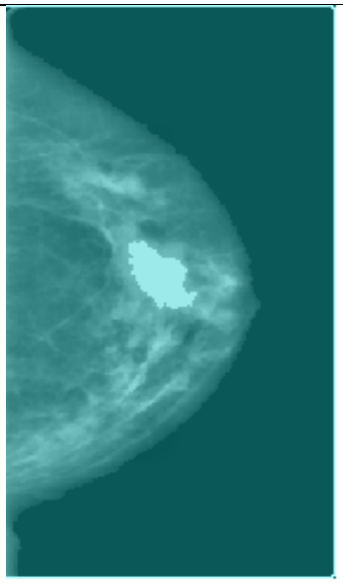
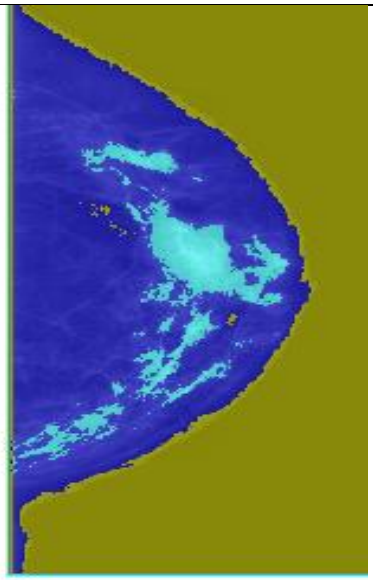
DICE INDEX

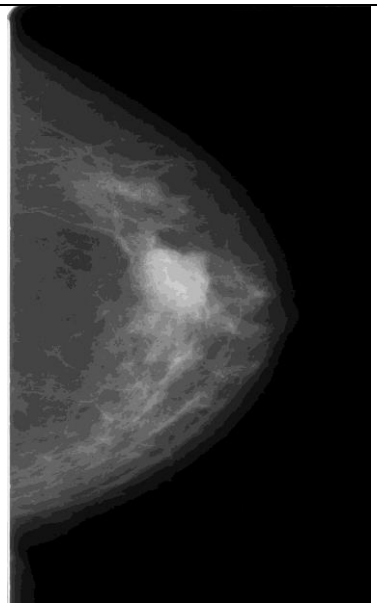
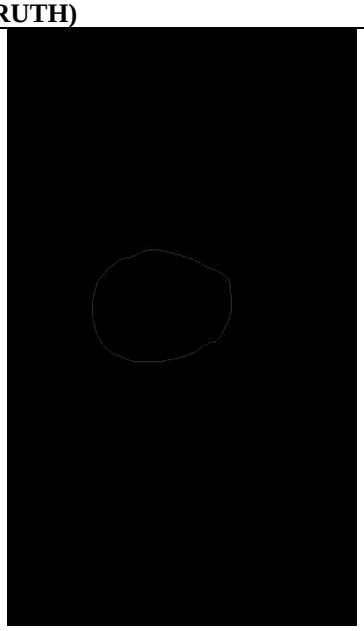
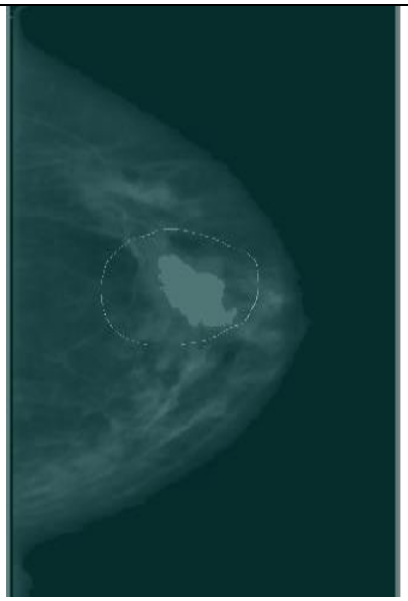
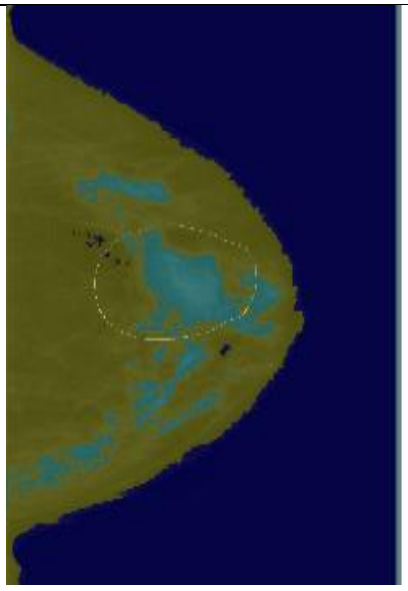
WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOG -ICAL OPERATOR
0.1731	0.1734	0.4426	0.5429

SSIM

WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOGI -CAL OPERATION
0.3559	0.5521	0.5997	0.5423

GRAPH:

ORIGINAL IMAGE	PREPROCESSING	WATERSHED ROI
D_4181 		
KMEANS ROI	**ROI BY DETECTING EDGES**	**ROI BY DILATON AND EROSION**
		

ORIGINAL IMAGE	MASKED IMAGE (GROUND TRUTH)	SUPERIMPOSED IMAGE WITH WATERSHED ROI
		
SUPERIMPOSED IMAGE WITH KMEANS ROI	**SUPERIMPOSED IMAGE WITH ROI BY DETECTING EDGES**	**SUPERIMPOSED IMAGE WITH ROI BY DILATON AND EROSION**
		

2D PLOT

MASK PLOT	WATERSHED PLOT	K-MEAN PLOT
[247, 363, 409, 206]	[379, 402, 221, 142]	[188, 343, 522, 389]
SOBEL OPE. PLOT	**MORPHOLOGICAL OPE. PLOT**	
[276, 348, 439, 226]	[306, 384, 338, 152]	

OVERLAP RATIO

WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOGICAL OPERATION
0.3725	0.4149	0.7442	0.6098

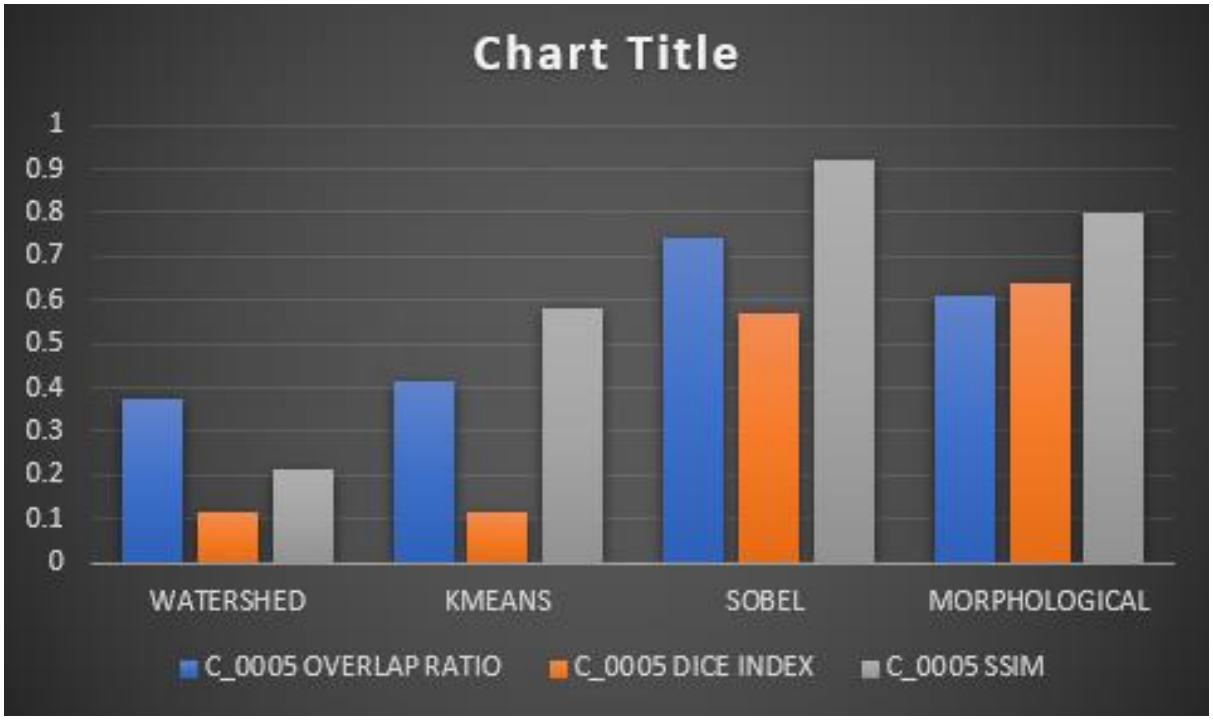
DICE INDEX

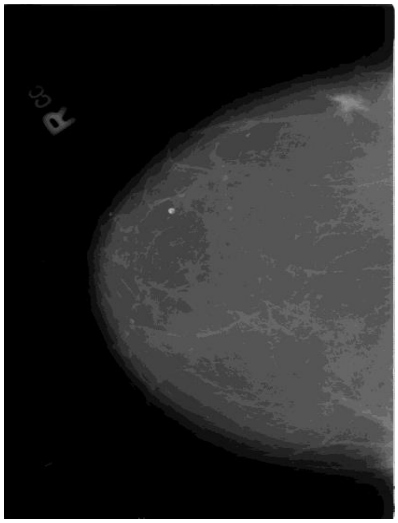
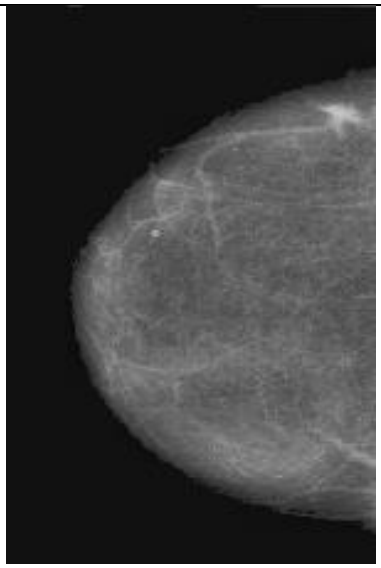
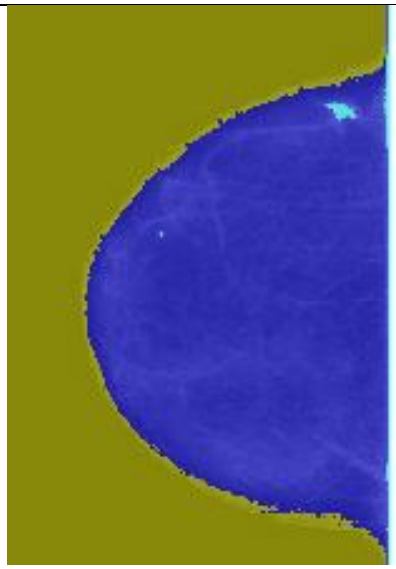
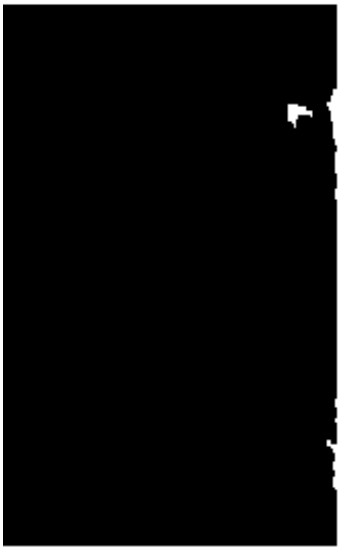
WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOG-ICAL OPERATOR
0.1141	0.1151	0.5693	0.6384

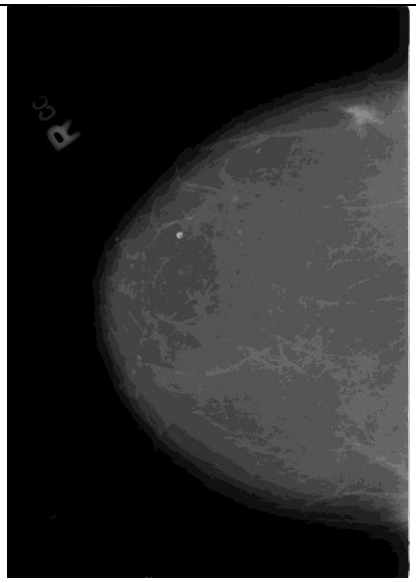
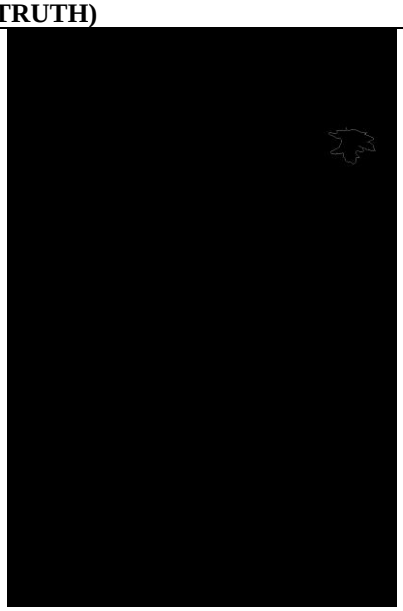
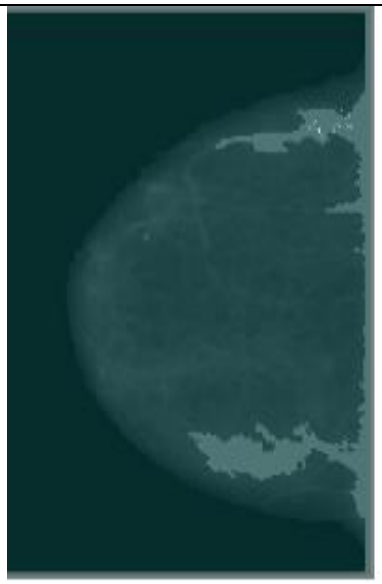
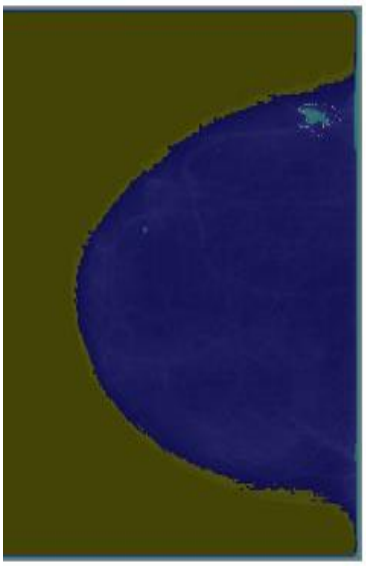
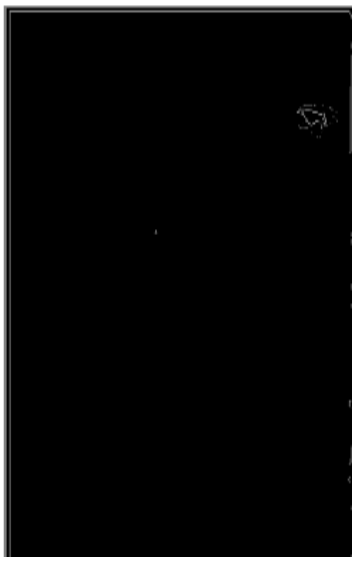
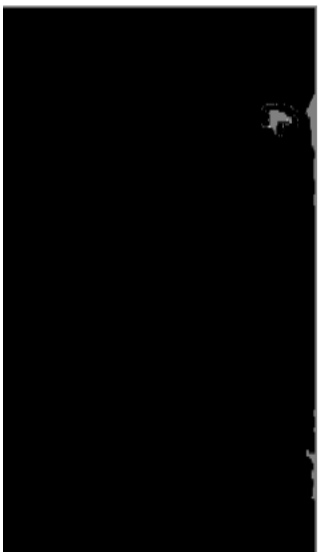
SSIM

WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOGI-CAL OPERATION
0.2134	0.5838	0.9719	0.8017

GRAPH:



ORIGINAL IMAGE	PREPROCESSING	WATERSHED ROI
B_3022 		
KMEANS ROI	**ROI BY DETECTING EDGES**	**ROI BY DILATON AND EROSION**
		

ORIGINAL IMAGE	MASKED IMAGE (GROUND TRUTH)	SUPERIMPOSED IMAGE WITH WATERSHED ROI
		
SUPERIMPOSED IMAGE WITH KMEANS ROI	**SUPERIMPOSED IMAGE WITH ROI BY DETECTING EDGES**	**SUPERIMPOSED IMAGE WITH ROI BY DILATON AND EROSION**
		

2D PLOT

MASK PLOT	WATERSHED PLOT	K-MEAN PLOT
[825, 164, 123, 77]	[594, 166, 381, 102]	[808, 164, 133, 80]
SOBEL OPE. PLOT	**MORPHOLOGICAL OPE. PLOT**	
[815, 164, 115, 80]	[839, 167, 112, 77]	

OVERLAP RATIO

WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOGICAL OPERATION
0.2359	0.7990	0.7637	0.8043

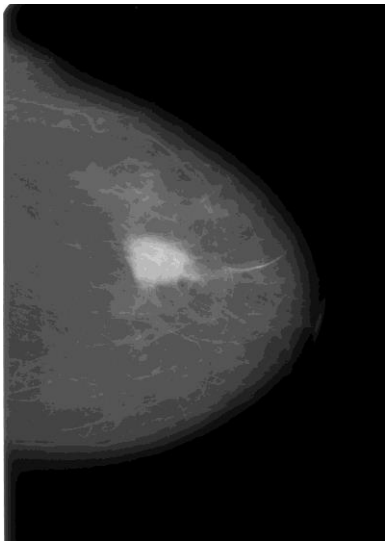
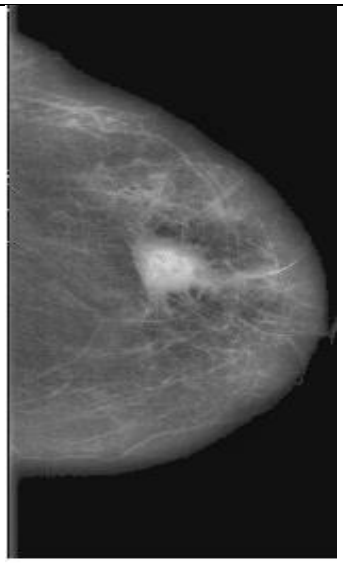
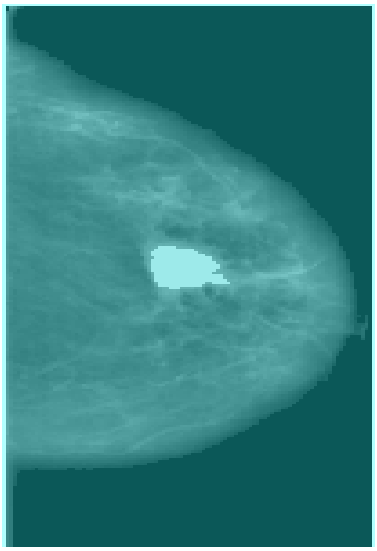
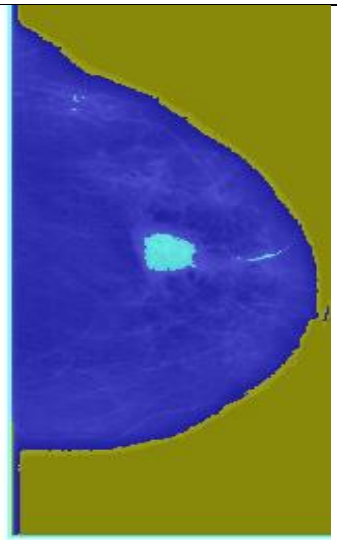
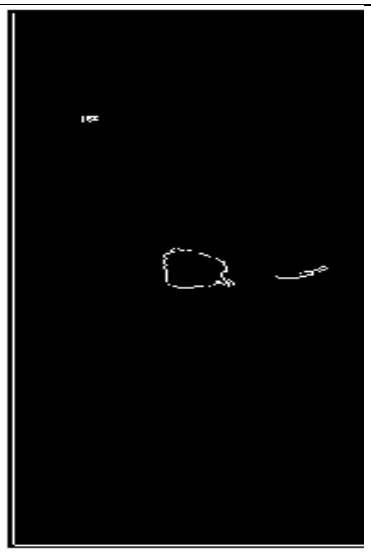
DICE INDEX

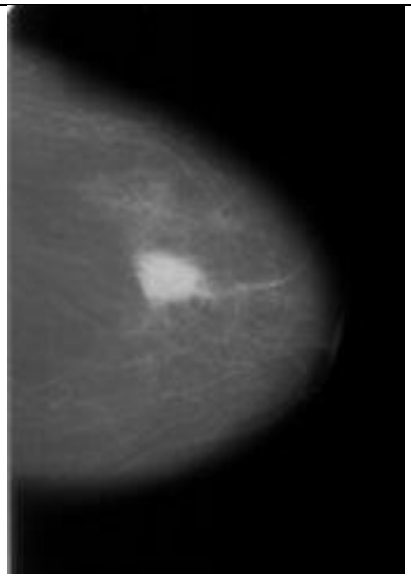
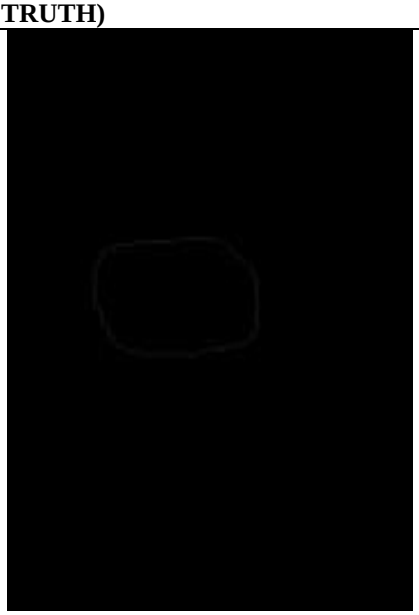
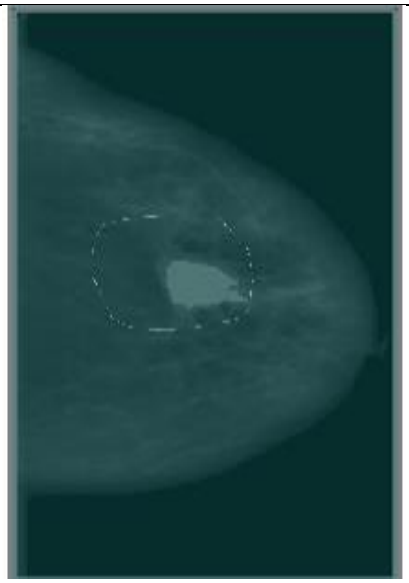
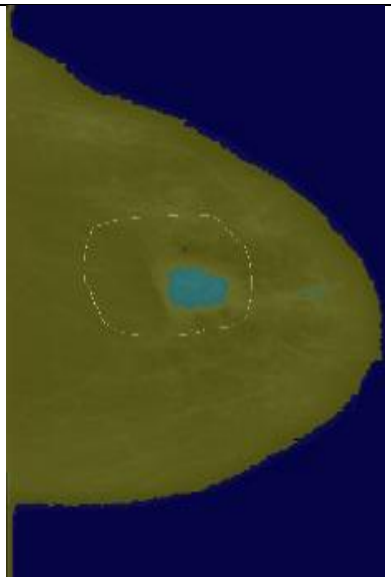
WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOG-ICAL OPERATOR
0.0655	0.0631	0.6031	0.7448

SSIM

WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOGI-CAL OPERATION
0.7404	0.9993	0.9997	0.9995

GRAPH:

ORIGINAL IMAGE	PREPROCESSING	WATERSHED ROI
D_4178 		
KMEANS ROI	**ROI BY DETECTING EDGES**	**ROI BY DILATON AND EROSION**
		

ORIGINAL IMAGE	MASKED IMAGE (GROUND TRUTH)	SUPERIMPOSED IMAGE WITH WATERSHED ROI
		
SUPERIMPOSED IMAGE WITH KMEANS ROI	**SUPERIMPOSED IMAGE WITH ROI BY DETECTING EDGES**	**SUPERIMPOSED IMAGE WITH ROI BY DILATON AND EROSION**
		

2D PLOT

MASK PLOT	WATERSHED PLOT	K-MEAN PLOT
[215, 356, 412, 209]	[403, 426, 204, 107]	[378, 436, 213, 108]
SOBEL OPE. PLOT	**MORPHOLOGICAL OPE. PLOT**	
[352, 410, 275, 140]	[352, 399, 365, 199]	

OVERLAP RATIO

WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOGICAL OPERATION
0.2535	0.2672	0.4471	0.4037

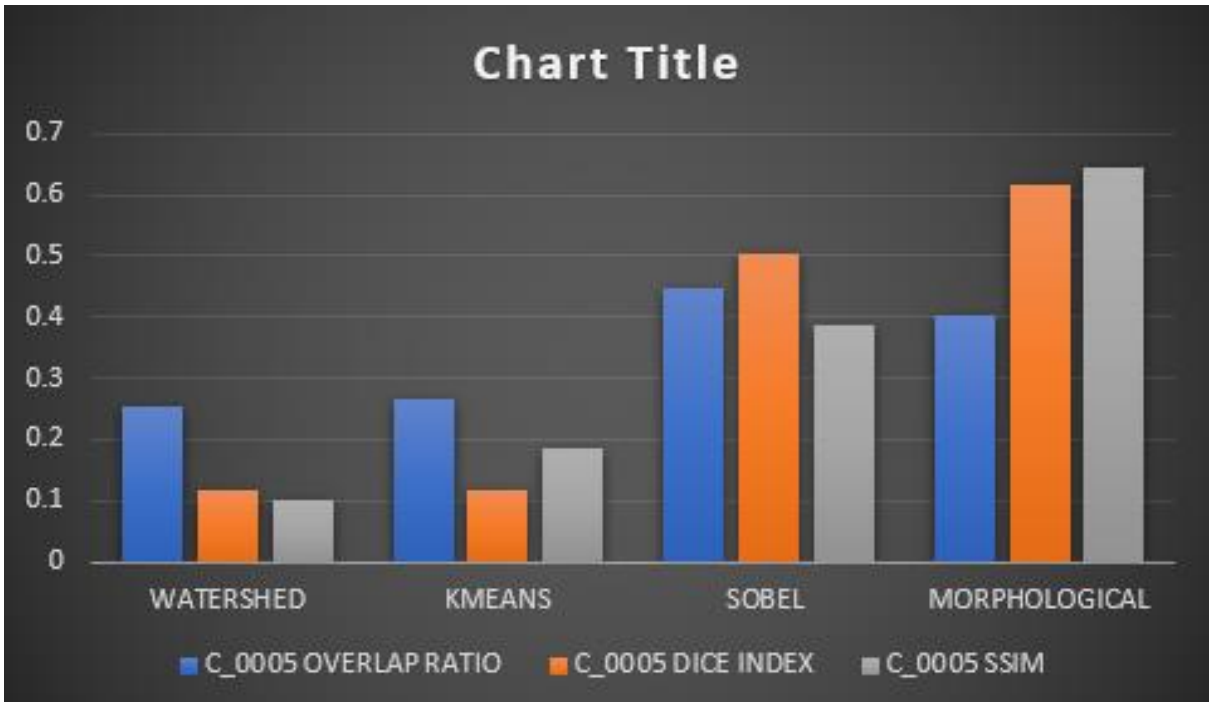
DICE INDEX

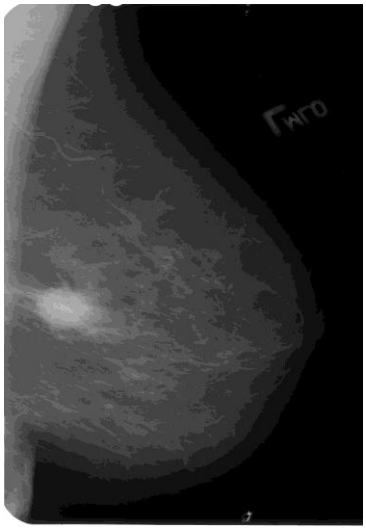
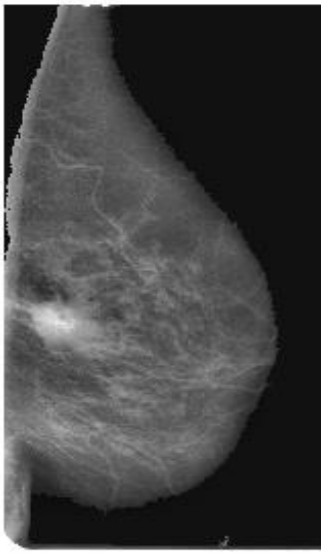
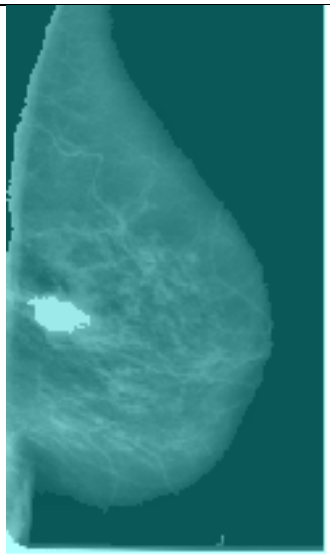
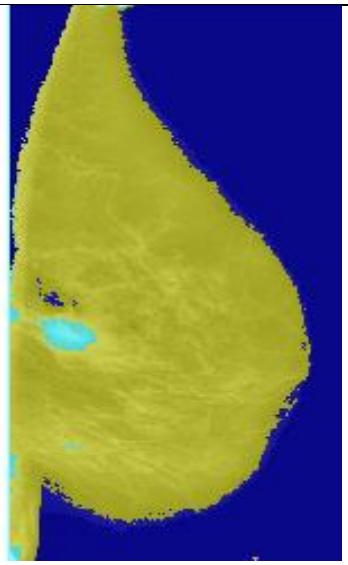
WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOG-ICAL OPERATOR
0.1173	0.1181	0.5016	0.6149

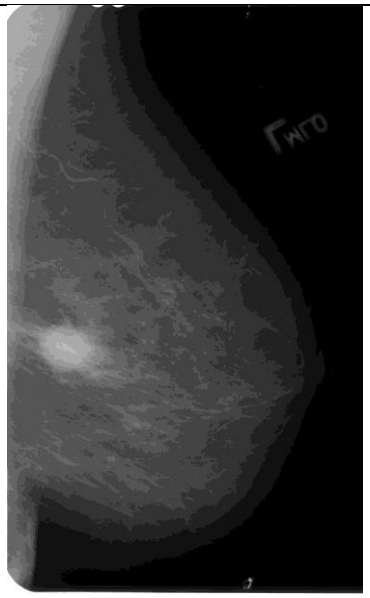
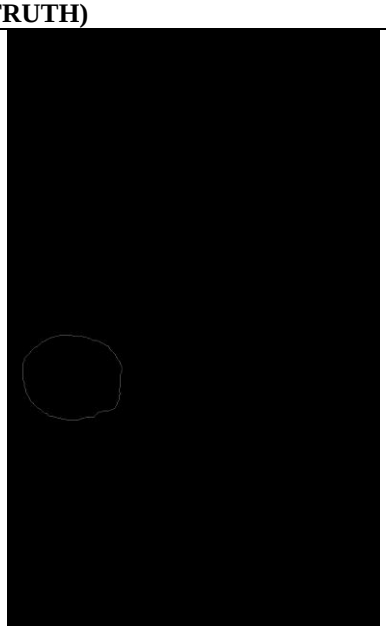
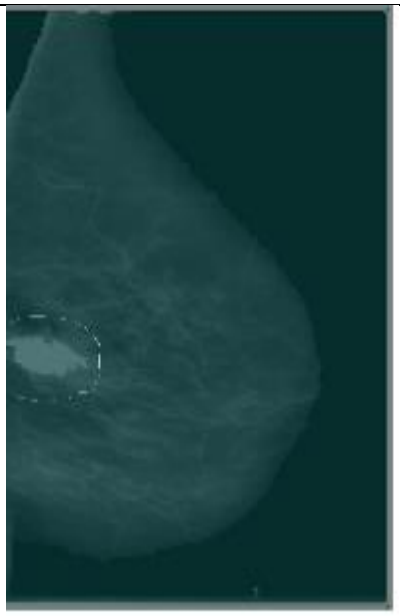
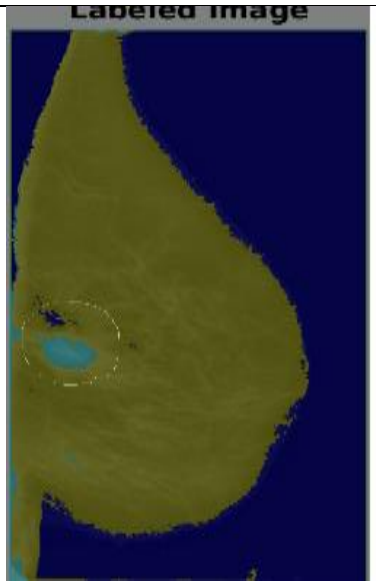
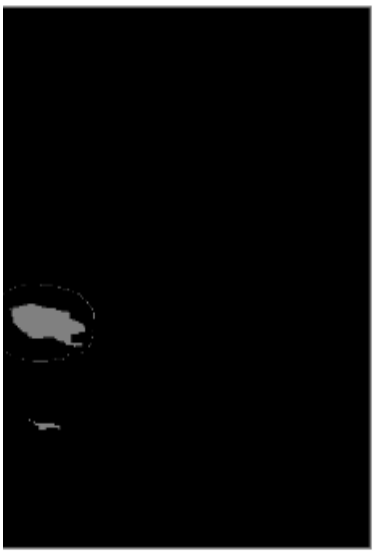
SSIM

WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOGI-CAL OPERATION
0.0996	0.1863	0.3870	0.6435

GRAPH:



ORIGINAL IMAGE	PREPROCESSING	WATERSHED ROI
B_3412 		
KMEANS ROI	**ROI BY DETECTING EDGES**	**ROI BY DILATON AND EROSION**
		

ORIGINAL IMAGE	MASKED IMAGE (GROUND TRUTH)	SUPERIMPOSED IMAGE WITH WATERSHED ROI
		
SUPERIMPOSED IMAGE WITH KMEANS ROI	**SUPERIMPOSED IMAGE WITH ROI BY DETECTING EDGES**	**SUPERIMPOSED IMAGE WITH ROI BY DILATON AND EROSION**
		

2D PLOT

MASK PLOT	WATERSHED PLOT	K-MEAN PLOT
[39, 508, 288, 147]	[87, 533, 213, 102]	[22, 518, 308, 158]
SOBEL OPE. PLOT	**MORPHOLOGICAL OPE. PLOT**	
[39, 518, 309, 179]	[43, 515, 315, 150]	

OVERLAP RATIO

WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOGICAL OPERATION
0.5132	0.7655	0.6780	0.7980

DICE INDEX

WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOG -ICAL OPERATOR
0.0781	0.0787	0.5692	0.6388

SSIM

WATERSHED	K-MEANS	SOBEL OPERATOR	MORPHOLOGI -CAL OPERATION
0.9581	0.9968	0.9950	0.9975

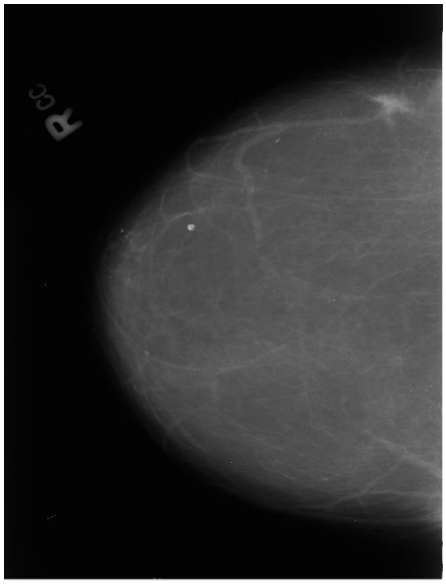
GRAPH:

15. INFERENCE:

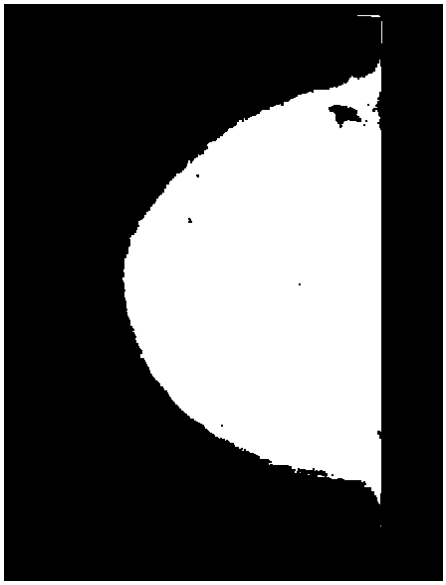
The above process was done for over 100 images and the average values are

	WATERSHED	KMEANS	SOBEL OPERATOR	MORPHOLOGICAL OPERATION
OVERLAP RATIO	0.3252	0.5245	0.5984	0.6215
DICE INDEX	0.1087	0.1095	0.5463	0.6521
SSIM	0.7211	0.6646	0.7506	0.7960

- So the above table states that the overlap ratio, dice index and SSIM are high for ROI extraction by morphological operators(ie. Dilation and erosion). The morphological concepts constitues a powerful set of tools for extracting regions of interest in an image. A significant advantage in terms of implementation is the fact that **dilation and erosion** are the primitive methods.
- The overlap ratio of the roi with morphological operators can be increased by integrating them with the thresholding techniques.



ORIGINAL IMAGE



ROI EXTRACTED IMAGE

- In this case the average overlapping ratio was found to be 0.8250. So we conclude that among those four methods of region of interest extraction morphological operators show better ratios than other methods. For more better results we can integrate the morphological operators with thresholding methods.

18. Reference :

- **A REVIEW OF MAMMOGRAPHIC REGION OF INTEREST CLASSIFICATION**

SB Yengec Tasdemir, K Tasdemir… - … Reviews: Data Mining …, 2020 - Wiley Online Library

https://onlinelibrary.wiley.com/doi/abs/10.1002/widm.1357

- **BREAST CANCER DETECTION IN DIGITAL MAMMOGRAMS**

KL Kashyap, MK Bajpai… - 2015 IEEE international …, 2015 - ieeexplore.ieee.org

https://ieeexplore.ieee.org/abstract/document/7294523