

How to Cite:

Ganvir, N. N., & Yadav, D. (2022). Extract roi from mammogram images using morphological and k-means clustering. *International Journal of Health Sciences*, 6(S5), 3040–3049.
<https://doi.org/10.53730/ijhs.v6nS5.9314>

Extract roi from mammogram images using morphological and k-means clustering

Neha N. Ganvir

Associate Professor, Department of E & TC, BVCOEL, Pune
Email: ganvirneha@gmail.com

D.M. Yadav

Dean Director, G.H.Raisoni COE & M, Pune
Email: dineshyadav800@gmail.com

Abstract---Cancer is leading cause of death in worldwide. Cancer cells damages all cells surrounded by it and hence covers the complete area of the body. Amongst women breast cancer is most common deadly disease as compared to men. In any type of cancer early-stage detection plays vital role as it can save the patient's life. If cancer is diagnosed early by using breast self-examination (BSE) and clinical breast examination (CBE) at 40-49 years of age, the survival rate of breast cancer reaches 100 per cent. New strategies named CAD (computer-assisted diagnosis) programs for early detection utilizing multiple mammogram datasets, such as mini-mias, DDSM, etc., CAD (Computer Aided Diagnosis) systems are mostly used for the second opinion for radiologists. Many researchers are already developing different CAD systems. Different databases are available for the researchers to study and detect the cancerous tumors. In this paper Mini-Mias database is used. Mammogram images are low contrast and may contain noise. Filtering or pre-processing is required to remove any noises present in the image, here Weiner filter is used. Two methodologies are compared here. First using morphological operated segmentation and second is by using k-means clustering algorithm. Both methodologies show segmented output of the image, but K-means clustering gives more accurate region of interest as compared to morphological. In this paper, pre-processing and segmentation is carried out. ROC output from K-Means clustering algorithm can be used to train any network to detect any cancerous tumor present in the images.

Keywords---BSE, CBE, CAD, CT, DDSM (Digital Data Screening Mammography), GLCM, Mammograms, MRI, PET.

1. Introduction

Breast cancer is leading cause for women mortality. Every year new cases of cancer spreads the news in all over the world. It has been estimated 9.6 million death over the worldwide. A total of 1,806,590 new cancer cases and 606,520 deaths are expected in the US in 2020, which is about 4,950 new cases and more than 1,600 deaths each day. The numbers are reported in "Cancer Statistics, 2020. Post cancer survival for women with breast cancer was reported 60% for Indian women, as compared to 80% in the U.S. If breast cancer detected in early stage like stage 0 and 1 then it can be cure. Breast self-examination (BSE) is the first step to examine. If any lump node observed and it has been advised to immediate contact doctor. Women with age of 40 or more than this are advised to do this BSE and CBE (clinical breast examination). The main objective of medical image study is to useful information about the physiological processes by using external and internal sources of energy and to achieve this X-rays, magnetic resonance, nuclear medicine, or ultrasound are used. The most effective and economical breast imaging modality has been X-ray mammography due to its simplicity, portability and cost effectiveness low contrast image, an important source of radiological information for breast imaging is the presence and distribution of micro calcifications in the breast, this anatomical information can be obtained with high resolution technology using X-rays, As yet there is no comprehensive imaging modality for all radiological applications and needs, although the ability to computerize and analyze medical images provides a powerful means to assist physicians. Many tests are available now like Blood tests, such as a complete blood count, mammogram of the other breast to look for signs of cancer, Breast MRI, Bone scan, Computerized tomography (CT) scan, Positron emission tomography (PET) scan. In US, once the women cross age of 40 it has been instructed to do CBE. 100% recovery is possible if the patient is identified in early stage. In this, paper pre-processing and segmentations methods discussed.

2. Materials and Methods

It is a good idea to know how your breasts normally look and feel so you can notice any changes. Medical organizations have different recommendations for breast self-exams, though. A mammogram can show breast lumps up to 2 years before they can be felt. Different tests help determine if a lump may be cancer. Ones that are not cancerous tend to have different physical features than ones that are. Imaging tests such as mammograms and ultrasounds can often see the difference [5]. A digital mammogram hence very useful to detect cancer at early stage it uses special X-ray images to spot abnormal growths or changes in breast tissue. Many pre-processing techniques have been discussed and techniques included filtering is mentioned in [1], [2] with statics. Image pre-processing may have dramatic positive effects on the quality of feature extraction and the results of image analysis. Image pre-processing is analogous to the mathematical normalization of a data set, which is a common step in many feature descriptor methods [4]. Image used for processing is from mini-mias database. Mini-Mias has many types of images includes normal, benign, and malignant images with micro-calcified images. Microcalcifications are tiny bits of calcium deposited in breast typically size of 0.3mm to 1mm in scale [2]. Images

filtered using wiener filter. Wiener filter gives better results as compared to other filters and it has smooth response. If noise is present in the system, then it is additive white Gaussian noise (AWGN) then it is removed by using wiener filter. In a morphological operation, the value of each pixel in the output image is based on a comparison of the corresponding pixel in the input image with its neighbors. Separating the segmentation process into two parts creates its main interest: finding the markers is the intelligent part, since it requires an interpretation of the image, but the difficulty is counter well-adjusted by the fact that the resulting segmentation is largely independent of the exact shape or position of the markers. Moreover, the tedious part that is getting the contours is entirely automatic. The advantages of the method are good localization, invariance to lighting conditions, absence of parameters, high sensitivity as strongly or weakly compared objects are equally segmented [7].

The algorithm is composed of two passes preceded by a pre-processing step for simplifying small scale details of the image that might cause over-segmentation. In the first pass feature images at various scales are extracted and kept in respective level of morphological towers. In the second pass, potential features contributing to the formation of segments at various scales are detected. Finally, the algorithm traces the contours of all such contributing features at various scales [8]. When images are processed for pre-processing it may happen that noises or distortion may be added. The main goal of morphological operation is to remove this imperfection that mainly affects the shape and texture of images. Morphological operations can be very useful in image segmentation as the process directly deals with shape extraction in mammogram image. Morphology in context of image processing means description of shape and structure of the object in an image. Morphological operations work based on set theory and relied more on relative ordering of the pixel instead on their numerical significance [6]. Here, algorithm for morphological operation on mammogram image steps are as follows.

- Read the image.
- Detect the entire cell of an image.
- Binary Gradient Mask shows lines of high contrast in the image so dilate an image.
- Some holes are present in dilated image, fill those holes.
- Remove connected objects on border as the cell of interest has been successfully segmented, but it is not the only object that has been found.
- To make the segmented object look natural, smooth the object by eroding the image.
- Display the mask over the original image.
- Outlined the segmented Image.

Step wise results are discussed in results and discussion section.

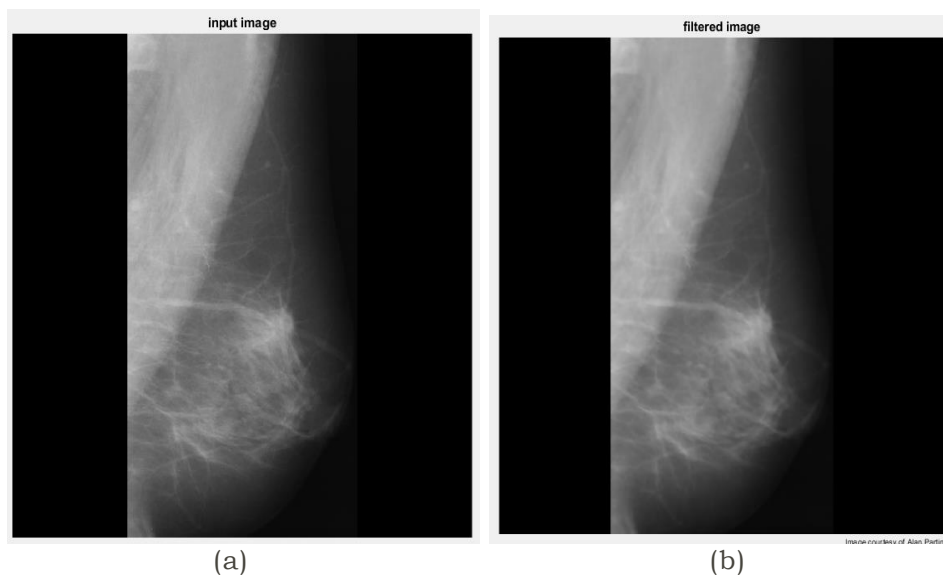
Second algorithm is segmentation by using K-means clustering algorithm. As segmentation stage is one of the most Significant steps since it affects the accuracy of the Feature Extraction & Classification. Rough k-means approach is used for segmentation of tumors from breast parenchyma. Pixel objects which belong to the tumor area are classified under smaller approximation, whereas objects which possibly belong to the same are considered as upper approximation. The difference of upper and lower approximation will result with

objects in the rough boundaries [9]. The K-means algorithm is an iterative technique that is used to partition an image into K clusters. In statistics and machine learning, k-means clustering is a method of cluster analysis which aims to partition and observations into k clusters in which each observation belongs to the cluster with the nearest mean. Algorithm states that pick K cluster centers, either randomly or based on some heuristic then assign each pixel in the image to the cluster that minimizes the distance between the pixel and the cluster center. Next step is to re-compute the cluster centers by averaging all the pixels in the cluster. Repeat last two steps until convergence is attained (e.g., no pixels change clusters) [10]. To segment the breast for grouping those pixels with similar tissue appearance according to certain number of clusters. The number of clusters were determined according to the number of bins in the image histogram [11], here histogram is used.

In this paper, Image is read from database. Then image masking is done which is pre-processing for k-means clustering algorithm. Once masking is over then image is ready for partition. As discussed, clusters analysis is done. Selecting K cluster centers which is calculating distance between the pixels in the clusters. The output of this discussed in results and discussion section.

3. Result and Discussion

In this Paper, two algorithms are discussed, and results are shown. First is by applying morphological operations and second is by using k-means clustering algorithm. Here Mini-Mias mammogram images have used. Normal and malignant image has been selected. Mammograms are low contrast images and has many artifacts like tags. Images should be preprocessed properly [1], [2].



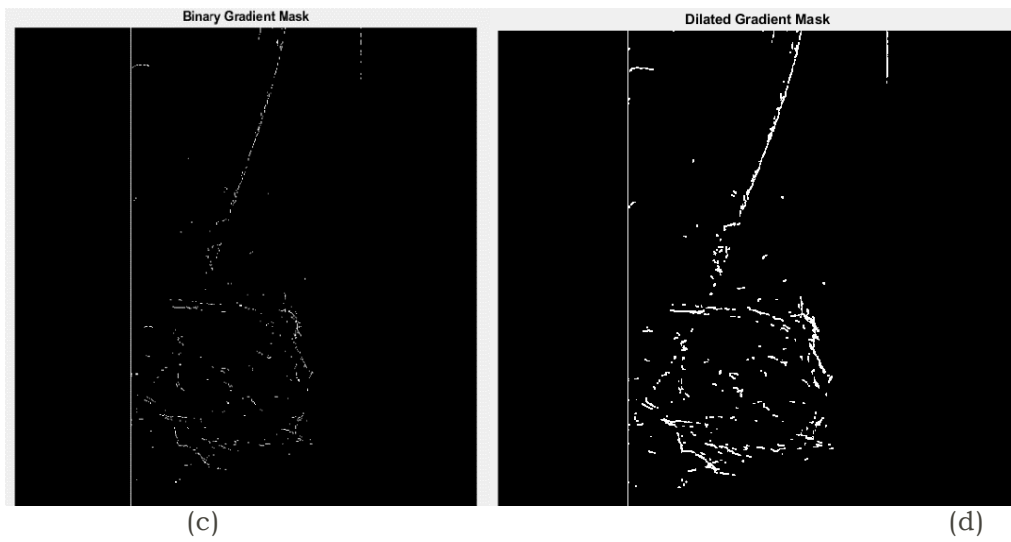
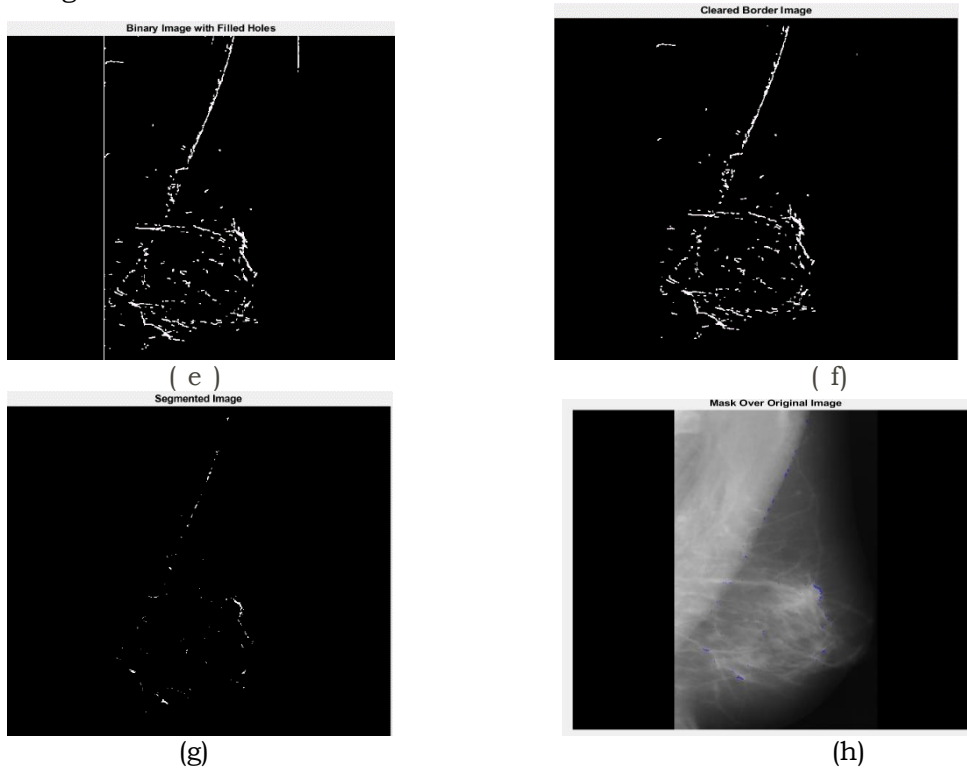
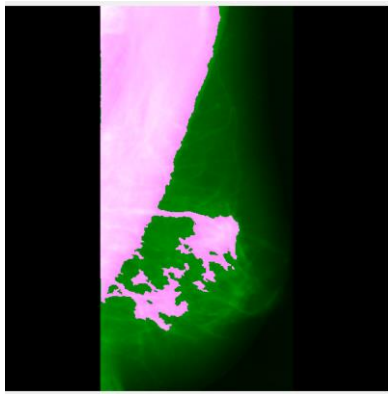
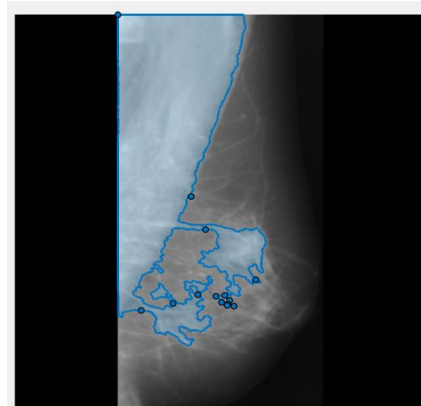


Figure 1: Mammogram Image (a) Input Mini-Mias Mammogram Image (b) Wiener filter Image (c) Binary Gradient Mask (d) Dilated gradient Mask Image.



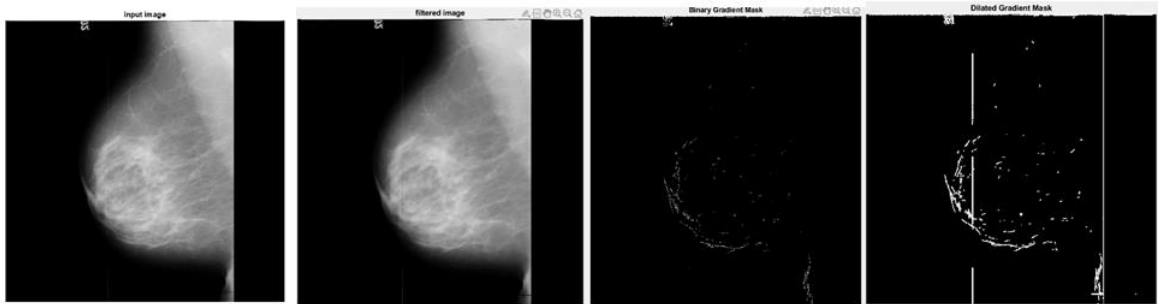


(i)



(j)

Figure 2: Mammogram Image (e) Binary Image filled with holes, (f) Cleared Border Image, (g) Segmented Image, (h) Pre-processing segmentation Mask. (i) Masked Image, (j) Locating Region of Interest Image using k-means clustering



(a)

(b)

(c)

(d)

Figure 3: Mammogram Benign Image (a) Input Mini-Mias Mammogram Image (b) Wiener filter Image (c) Binary Gradient Mask (d) Dilated gradient Mask Image

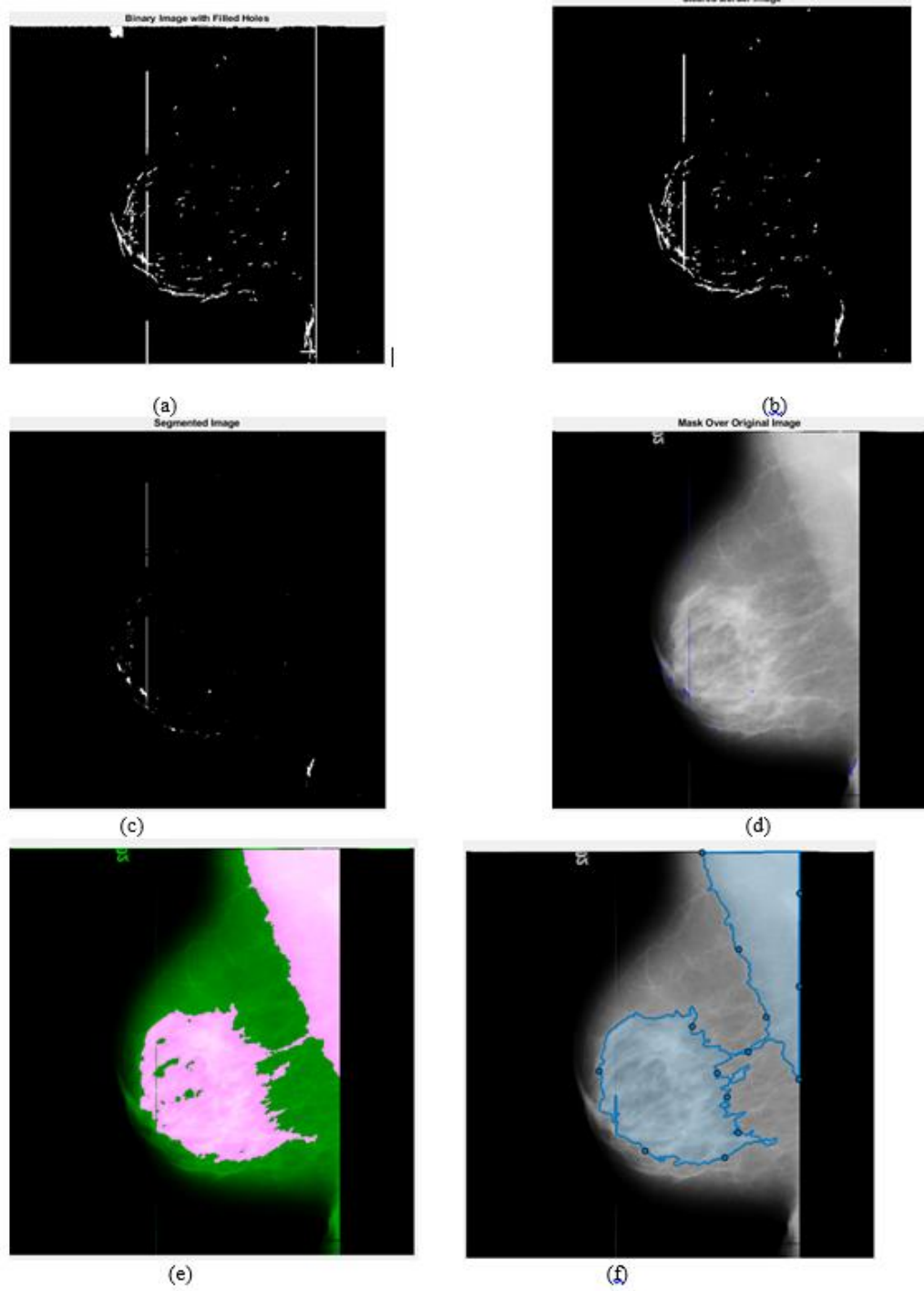


Figure 4: Mammogram Image (a) Binary Image filled with holes, (b) Cleared Border Image, (c) Segmented Image, (d) Pre-processing

segmentation Mask, (e) Masked Image, (f) Locating Region of Interest Image using k-means clustering.

Here, MATLAB platform is used for analyzing images. Figure 1 and 2 shows normal image from database, and figure 3 and 4 shows image outputs of malignant image from Mini-Mias database. Wiener filter has advantages of smooth filtering and maintaining aspect ratios of the image. Two types of segmented images have shown here first by performing morphological operations and second by using k-means clustering. Figure 1 and 2, fig 3 and 4 (a), (b), (c), (d) and (e), (f) and (g) shows morphological operated images. Figure 2, (h),(i) and (j) shows k-means clustering segmented images and figure 4 (d), (e) and (f) shows the same. Images are taken from Mini-Mias Database mathematical Morphology, dealing with geometrical features is a well-suited technique for segmentation purpose. Figure (a) shows mammogram image of normal and malignant image. This image is required to be processed; fig (b) shows Weiner filtered image. Selection of wiener filter is based on as per [1] as this image is again gone through morphological operations. The beauty of morphological operation is discussed in material and methods, detect a cell using edge detection and basic morphology. An object can be easily detected in an image if the object has sufficient contrast from the background. Figure (c) is used to detect the entire cell. A binary mask containing the segmented cell, calculate the gradient image and apply a threshold. Use edge and the Sobel operator to calculate the threshold value. Tune the threshold value and use edge again to obtain a binary mask that contains the segmented cell. Figure (d) is dilated image. The binary gradient mask shows lines of high contrast in the image. These lines do not quite delineate the outline of the object of interest. Compared to the original image, there are gaps in the lines surrounding the object in the gradient mask. Figure (e) shows binary image filled with holes, as the dilated gradient mask shows the outline of the cell quite nicely, but there are still holes in the interior of the cell. Figure (f) shows cleared border image as the cell of interest has been successfully segmented, but it is not the only object that has been found. Any objects that are connected to the border of the image can be removed. Finally, to make the segmented object look natural, smooth the object by eroding the image twice with a diamond structuring element, figure (g) shows segmented image after using morphological operation. Image can be visualized by masking the image as shown in figure (h). Second methodology is to segment the image using k-means clustering. It is discussed briefly in material and methods. Before processing to k-means clustering, the image should be pre-processed by using segmentation mask as shown in figure (i). After masking, segmentation is carried out with locating region of interest as shown in figure (j) in figure 2. It has been observed that as compared to morphological segmentation k-means clustering gives more properly segmented output and clearly identify region of interest. This region of interest can process further for extract useful features to detect images cancerous or not.

4. Conclusion

Cancer is deadly disease in worldwide. Breast cancer is most common death amongst women. It is very important to detect cancer in early stage. BSE and CBE are common technique to detect lumps or nodes in breast. In this paper, preprocessing segmentation technique with filter is discussed. Mini-Mias

database is used, normal and malignant images are used to perform different operations. Figure 1,2,3,4 shows different morphological operated images and figure (i) and (j) shows preprocessed k-means clustering used segmented images with highlighting region of interest. Mammograms are low contrast images and included marks, labels, and tags etc. Filtering is important to remove any noise present in the image. Weiner filter has smooth filtering performance and morphological operations are performed which works on the shapes of image. Outlined border is properly can observe from the results. Next methodology has been performed on mammogram images are k-means clustering segmentation. First partial stretching enhancement is applied to the image to improve the quality of the image before applying to K-means clustering algorithm. It is an unsupervised algorithm, and it is used to segment the interest area from the background. Figure shows region of interest from the mammogram image. K-means clustering gives best segmentation result as compared to morphological. This output of ROC can be processed further to extract different features to train any neural network to detect whether the image is cancerous or not.

5. References

- [1] Neha N.Ganvir, D.M.Yadav, "Filtering Method for Pre-processing Mammogram Images for Breast Cancer Detection", International Journal of Engineering and Advanced Technology. ISSN: 2249 – 8958, Volume-9 Issue-1, (pp. 4222-4229),Oct 2019.
- [2] Neha N.Ganvir, D.M.Yadav , "Clustered Micro-Calcifications Extraction From Mammogram Images Using Cellular Automata Segmentation With Anisotropic Diffusion Filtering" , International Journal of Scientific & Technology Research. ISSN 2277-8616, VOLUME 9, ISSUE 04, (pp. 1217-1221), April 2020.
- [3] Marcotegui B., Meyer F. (1994), "Morphological Segmentation of Image Sequences. In: Serra J., Soille P. (eds) *Mathematical Morphology and Its Applications to Image Processing*". Computational Imaging and Vision, vol 2. Springer, Dordrecht. https://doi.org/10.1007/978-94-011-1040-2_14.
- [4] Krig S. (2014), "Image Pre-Processing. In: *Computer Vision Metrics*". Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4302-5930-5_2.
- [5] <https://www.webmd.com/breast-cancer/breast-cancer-detection>.
- [6] Srisha, Ravi & Khan, Am. (2013)., "Morphological Operations for Image Processing : Understanding and its Applications", Conference: National Conference on VLSI, Signal processing & Communications.
- [7] Meyer F. (2005), "Morphological segmentation revisited. In: Bilodeau M., Meyer F., Schmitt M. (eds) *Space, Structure and Randomness*". Lecture Notes in Statistics, vol 183. Springer, New York, NY. https://doi.org/10.1007/0-387-29115-6_13.
- [8] S. Mukhopadhyay and B. Chanda, "Multiscale morphological segmentation of gray-scale images," in *IEEE Transactions on Image Processing*, vol. 12, no. 5, pp. 533-549, May 2003, doi: 10.1109/TIP.2003.810757.
- [9] Rinantha, K., & Suryasa, W. (2017). Comparative study for better result on query suggestion of article searching with MySQL pattern matching and Jaccard similarity. In *2017 5th International Conference on Cyber and IT Service Management (CITSM)* (pp. 1-4). IEEE.

- [10] Bora V.B., Kothari A.G., Keskar A.G. (2012), “ *Mammogram Segmentation Using Rough k-Means and Mass Lesion Classification with Artificial Neural Network*”. In: Hassanien A.E., Salem AB.M., Ramadan R., Kim T. (eds) *Advanced Machine Learning Technologies and Applications. AMLTA 2012. Communications in Computer and Information Science*, vol 322. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-35326-0_7.
- [11] Patel, Dr. Bhagwati & Sinha, Professor G. (2010), “ *An Adaptive K-means Clustering Algorithm for Breast Image Segmentation*”. *International Journal of Computer Applications*. 10. 10.5120/1467-1982.
- [12] Shahine, Osama & Kelash, Hamdy & Faragallah, Osama & Attiya, Gamal. (2012). “*Breast Mass Detection in Mammograms using Modified K-means Clustering*”. 10.13140/2.1.4115.4561.