Breast Cancer Detection Using Image Processing Techniques

Siddhartha Gupta
School Of Electrical Engineering (SELECT)
VIT Vellore
Vellore, India
Siddgupta07395@gmail.com

Neha Sinha
School Of Electrical Engineering (SELECT)
VIT Vellore
Vellore, India
nehasinhaneha1@gmail.com

Sudha R
Staff member at Division of production system
University, West, Sweden
Trollhättan
ishuma@gmail.com

Challa Babu*
School Of Electrical Engineering (SELECT)
VIT Vellore
Vellore, India
babu2342@gmail.com

Abstract— **Cancer is the uncontrolled multiplication of group of cells in a particular location of the body and is the second largest disease leading to the death of women in the world. The disease can be cured if it is detected in early stages. A lot of research has been done to find out the tumor correctly but a 100% accurate method has not been found. Research on breast cancer detection using digital image processing is not new but many new approaches in this field is being considered to accurately predict the tumor region. The present approach is to detect the tumor region visually as well as to figure out in which region the tumor is mostly concentrated. This work majorly focuses on finding out the best algorithm/s to detect the tumor present in the breast. In the proposed work, a variety of algorithms has been applied but the best one suited for cancer detection is the combination of K Means, Closing, Dilation and Canny Edge Detection algorithm.**

Keywords— Image Processing, Breast Cancer, K-Means, Dilation, Closing, Edge Detection

I. INTRODUCTION

Cancer cells are termed as abnormal tumors. Breast cancer is any form of abnormal tumor that develops from breast cells. Breast cancer is the second biggest cause of deaths in women mostly aged between 30-55 years. In order to differentiate and identify the tumor region, first of all mammograms are taken and Region of Interests (ROI) is taken into consideration. Image enhancement is performed on the mammograms and then various algorithms are applied. In the existing literature different algorithms for clustering the image data has been presented.

Breast Tumor detection done in the past consisted of performing Image Processing and extracting the region of interest followed by processing that region using various algorithms and finally applying neural nets to find out the accuracy of the detection [1-3] but the main drawback of these methods was that the algorithms and methodology was pretty linear in nature. The focus was on getting the work done with good accuracy but there was no mention of better alternatives or the removal of noise by using any other filters or algorithms [4]. Also, the Closing and Dilation algorithm has been very rarely used in any application. These algorithms are very useful if they are used along with some good algorithms as they remove noise to a good extent in images which helps in reducing running time in pre-processing. A lot of research on finding out the right algorithm for reducing running time during pre-processing has been done [5]. Table1 shows some algorithms/techniques used in the past and their efficiency for better understanding. The approach of this works with three to five variation of algorithms and gives to the ground truth images detailed comparison with results as to how a tumor can be detected and what is the best way to achieve it [6]. The method explained in this paper consists of a tree shape structure and it tells the best route to follow for good accuracy and the least convergence time. The detailed result with respect to accuracy for each path has been mentioned [7]. A bunch of algorithms for the cancer detection application has been explained in this paper, the best one mentioned consists of K-Means Algorithm, Closing Algorithm, Dilation Algorithm, and Canny Edge Detection Algorithm.

Section 2 explains the past research that has been done and contains specific remarks explaining the accuracy of a particular technique or algorithm. Section 3 contains the entire Methodology of the paper and it consists of the various algorithms used along with their outcomes. In Section 4, results along with the final accuracy achieved is mentioned and the paper has been concluded in section 5 with References in the next section.

II. LITERATURE REVIEW

The best suited algorithm for cancer detection by using image processing techniques are listed in Table 1.

978-1-5386-8190-9/19/$31.00 © 2019 IEEE

TABLE I. DIFFERENT TYPES OF ALGORITHMS AND THEIR ADVANTAGES AND DISADVANTAGES

Author & Year	**Title**	**Parameters Considered**	**Remarks**
Mohammad Naved Qureshi 2018 [8]	An Improved Method for Image Segmentation Using K-Means Clustering with Neutrosophic Logic	Novel Approach	K means gave better results with images containing noise or not containing noise
Spandana Paramkusham 2013 [9]	Automatic Detection of Breast Lesion Contour and Analysis using Fractals through Spectral Methods	Image Processing accuracy was 82%. Combined with ML accuracy was 98%.	Many techniques were used in this paper but it still is pretty linear with no emphasis on why they are using just this algorithm and nothing else.
Tobias Chrisiian Cahoon 2000 [10]	Breast Cancer Detection Using Image Processing Techniques	Reduce the error upto 10%	The results had lots of noise in the final stages as well as it was not very reliable.
Liying Yuan 2015 [11]	Adaptive Image Edge Detection Algorithm Based on Canny Operator	Results were Good	The algorithm can extract full skull image edge whose detail is richer, positioning is more accurate, and it is not affected by noise easily.
Huili Zhao 2010 [12]	Improvement of Canny Algorithm Based on Pavement Edge Detection	Accuracy increased by 10%	The algorithm was better for images having pavement detection application.
Peizhen Wang 2013 [13]	Research on Edge Detection Algorithm for Plate Image	Accuracy was 90% Convergence time was slower than normal	Plate Images edges were clearly detected after using this algorithm.
Syed Mohammad Abid Hasan 2016 [14]	Depth edge detection by image-based smoothing and morphological operations	Computation time was 0.01 sec	The noise was greatly removed but the algorithm is not good for real time application.
G.T. Shrivakshan 2012 [15]	A Comparison of various Edge Detection Techniques used in Image Processing	None	Canny Edge detection gave the best results but under the noisy conditions Laplacian Edge Detection algorithm proved beneficial
M. Fathy 1995 [16]	An image detection technique based on morphological edge detection and background differencing for real-time traffic analysis	100% accuracy with 5% in some special cases	Edge Detection was very successful even in the past and was perfectly integrated with this project.
Zhang Jin-Yu 2009 [17]	Edge Detection of Images Based on Improved Sobel Operator and Genetic Algorithms	Running time was 0.92 sec	Accuracy was not very high with this algorithm although anti-noise and runtime capability was increased.
Larry S. Davis 1975 [18]	A Survey of Edge Detection Techniques	A review Paper	This paper tells the advantages of the edge detection techniques and shows that it is quite beneficial when it comes to medium level noise images.
Pragati Kapoor 2010 [19]	Image Processing for Early Diagnosis of Breast Cancer Using Infrared Images	New Approach	Edge detection segmented the thermal images quite nicely and was successful in giving them a strong foundation to work their respective algorithms
Nameirakpam Dhanachandra 2015 [20]	Image Segmentation using K-means Clustering Algorithm and Subtractive Clustering Algorithm	The Peak to Signal Noise Ratio was 34.80 and The Root mean square error was 0.0077	The Proposed method has better results when compared to using just the K means algorithm
R.Meena Prakash 2017 [21]	Segmentation of thermal infrared breast images using K-Means, FCM and EM algorithms for breast cancer detection	FCM segmentation gives good accuracy	The drawback of K-Means segmentation is that it results in empty clusters in certain cases
Mrs. Sandhya G 2013 [22]	Automated Detection of Cancer Tissues in Mammograms Using Advanced K-Means Clustering with Homomorphic Filtering	Advanced K means accuracy was 90%	The Algorithm is a success but it's linear in nature due to a lack of resources. The accuracy is also good.
Paweł Filipczuk 2012 [23]	Breast Fibro adenoma Automatic Detection Using K-Means Based Hybrid Segmentation Method	Final Accuracy of 85.78% was achieved	K means algorithm is used to segment the images and then train using KNN, Naïve Bayes Algorithm
Tse-Wei Chen 2010 [24]	Edge Adaptive Image Segmentation Based on K Means	Running Time was 218 ms which is quite efficient.	The Paper is to compare the results between using Edge segmentation based on K means and the K means algorithm.
H.P. Ng 2006 [25]	Medical Image Segmentation Using K-Means Clustering And Improved Watershed Algorithm	Accuracy was 93%-94% and fewer partitions after using K means algorithm	By using K means before watershed algorithm the drawbacks of sensitivity to noise and over partitioning
Ala Balti 2011 [26]	Segmentation and Enhancement of Fingerprint Images Using K-means	The sensitivity of the algorithm was 75.9764	The accuracy after using the K means algorithm was Highly improved

978-1-5386-8189-3/19/$31.00 © 2019 IEEE

The proposed work consists of a tree approach like process as shown in Fig. 1.

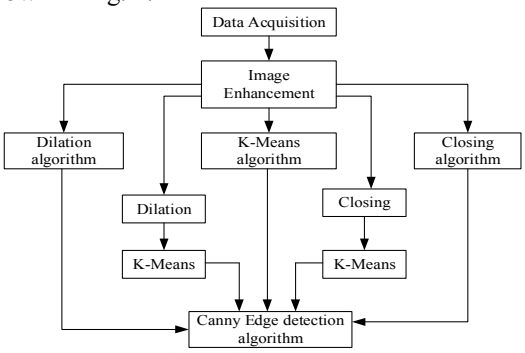

Fig. 1. The Tree Process

III. METHODOLOGY

The main components constituting methodology are i) Image Acquisition ii) Image Preprocessing iii) Image Enhancement iv) Morphological Processing v) Image Segmentation and Data Extraction. They are explained in detail below:-

A. Image Acquisition

The primary database is from The Mammographic Image Analysis Society (MIAS) database of digital mammograms (v1.21) obtained from University of Cambridge which consists of 322 images. Five other images has also been downloaded from Google Images to test the algorithm to some noisy images as well. The resolution of MIAS is about 50 micron. Some of the images has also been acquired from Google images for experimentation purposes. Most of the images have significant tumor present in them which can be seen easily while some have small cysts present in them that are hard to find. Fig. 2 consists of some images that was found online:

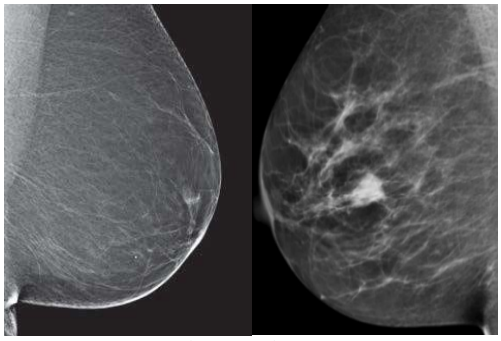

Fig. 2. A mammogram of a Normal Breast vs a mammogram of a Cancerous Breast

B. Image Preprocessing

As the pixel size of the image is too large to perform an operation or to extract some useful features so the Region of Interest (ROI) of the image was extracted to perform different operations so that the useful features can be extracted easily.

C. Image Enhancement

Image Enhancement is mainly done with the goal that the outcome is more appropriate than the original picture for a particular application. The spatial domain techniques are performed mainly because they operate directly on the pixels. The main Transformations applied were:

a) LOG Transformation

$$g(m,n) = clog_2(f(m,n)) \quad (1)$$

Where f (m, n) is the pixel value of the given image at a point.

b) Power Transformation

$$g(m,n) = [f(m,n)]^{\gamma} \quad (2)$$

Where gamma can be greater than one or smaller than one

c) Negative Transformation

$$g(m,n) = 255 - f(m,n) \quad (3)$$

This transformation is especially suited for improving white and dim detail inserted in dark locales of a picture, particularly when the black territories are prevailing in size.

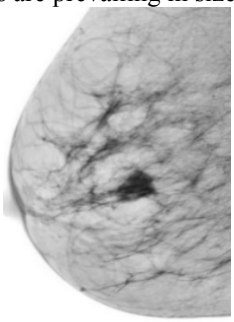

Fig. 3. Final Image after applying every transformation

D. Morphological Processing

Morphological image processing is a gathering of non-linear operations identified with the shape or morphology of features in a picture. The Morphological processes applied were closing, dilation and K Means Clustering.

(a) Closing

Closing is a very important tool in image processing. Like its dual operator opening, it can be derived from the fundamental operations of erosion and dilation. Closing is comparative somehow to Dilation on the grounds that it will in general amplify the limits of foreground (bright) regions in a picture (and reduce background color gaps in such locales) and it is less damaging to the original boundary shape. Likewise with other morphological operators, the correct task is dictated by an organizing component or a structuring element. The impact of the operator is to safeguard background regions that have a comparative shape to this structuring element, or then again that can totally contain the organizing component, while dispensing with every single other regions of background pixels.

Simply the idea of "closing" is to remove false negatives.

$$A \bullet B = (A \oplus B) \ominus B \quad (4)$$

Where A is the image used and B is the kernel that is sliding across that image whereas $\oplus$ represents dilation and $\ominus$ represents erosion.

978-1-5386-8189-3/19/$31.00 © 2019 IEEE

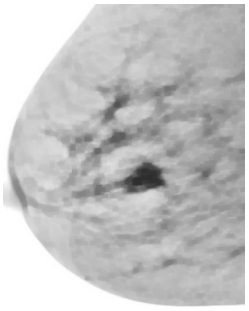

Fig. 4. Image after Applying Closing

(b) Dilation

The fundamental impact of the operator on a binary image is to bit by bit broaden the limits of areas of foreground pixels (i.e. white pixels, commonly). Along these lines, zones of foreground pixels develop in size while holes inside the locales wind up a little. Basically a slider slides around the region and checks for black regions and if the entire area of pixels is not black, it is converted to white. It eliminates the black noise.

$$A \oplus B = \cup_{b \in B} A_b \quad (5)$$

Where A_b is defined as the translation of A by b.

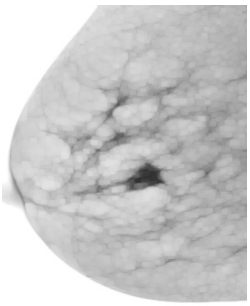

Fig. 5. Image after applying Dilation

(c) K Means Clustering

K means clustering is a method commonly used to automatically divide a data set into k groups [figure6]. It starts with taking k samples of clusters and then repeatedly refining them as follows:

1. Each instance d_i is assigned to its closest cluster
2. Each cluster center C_i is updated with the average or mean of its elements present in it.

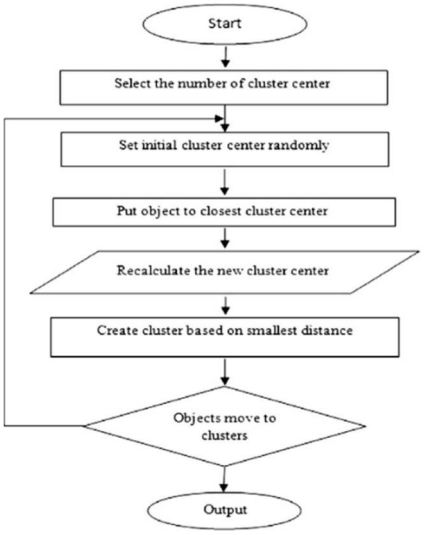

Fig. 6. Flowchart explaining the working of K means Algorithm

Through K means we gave only 3 segments in our sample which hide the unwanted feature and the tumor is clearly visible after the process is performed.

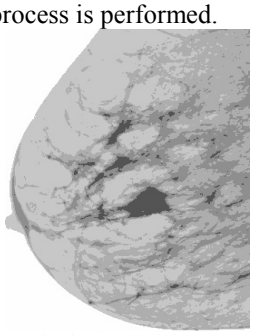

Fig. 7. Image acquired after applying K-Means Clustering

(d) Closing With K Means

Closing has been applied before k means algorithm to remove the noise and make the image smooth as much as possible before applying the edge detection algorithm.

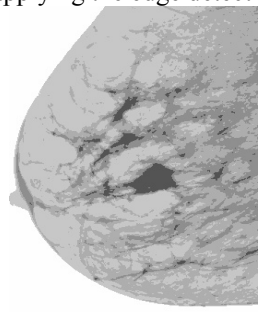

Fig. 8. Image Acquired after applying Closing with K means Algorithm

(e) Dilation with K means algorithm

The same as closing has been done with the dilation algorithm as well to check the accuracy of the tumor detection.

978-1-5386-8189-3/19/$31.00 © 2019 IEEE

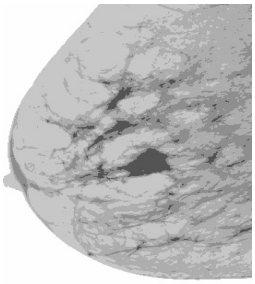

Fig. 9. Image Acquired after applying Dilation with K means Algorithm

E. Image Segmentation and Data Extraction

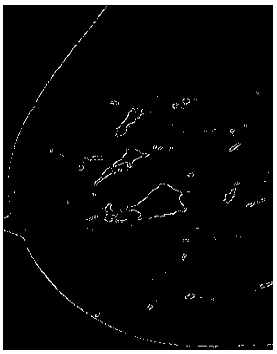

Fig. 10. Image after Applying Canny Edge Detection to a K Means Image

For this purpose Edge Detection algorithm is used. Edge detection is a technique in which the boundaries of the objects contained in an image is found. An edge is basically where the pixel value of the image changes drastically. An Edge detection Algorithm works by detecting discontinuities in brightness. Canny Edge Detection algorithm implemented in python has been used. After applying all the algorithms the tree process looks like as shown in Fig. 1.

IV. Results and Discussion

The individual results of each algorithm has been shown in the above section without applying the edge detection. But it is not possible to get accurate results using individual algorithms. So each algorithm has been implemented with edge detection to enhance the accuracy of the detection process. For better understanding of the algorithms, two processes are taken into account namely process 1 and process 2. K means, Closing and Dilation are individually applied with edge detection in process 1 whereas in process 2, the two algorithms has been combined to enhance the quality of image i.e. Closing and K Means, Dilation and K means.

Process1: The images were taken by applying the base algorithm in Subsection 3.4 and is processed by applying Edge Detection mentioned in subsection 3.5.

The base algorithms were K means, Closing and Dilation. The images received after applying the algorithms are given in Fig. 11.

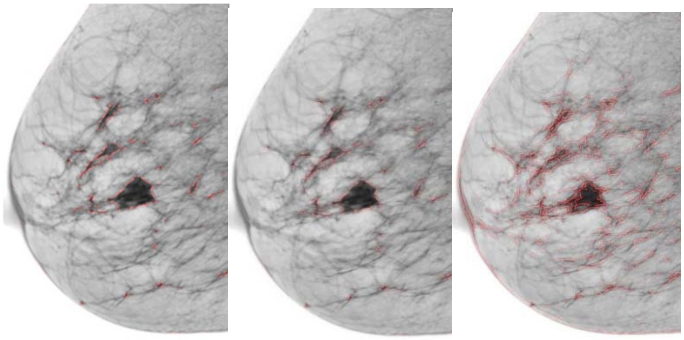

Fig. 11. From Left-Result showing K means, Closing and Dilation

Process 2:

In process 2 the images from closing with K means and Dilation with K means are taken from subsection D and is processed by applying Edge Detection mentioned in subsection E. The images received after applying the algorithms are given in Fig. 12.

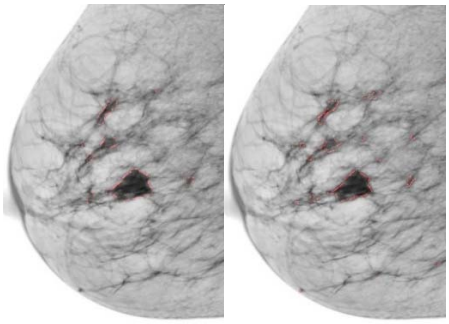

Fig. 12. From left-Result after applying (i) Closing and K Means (ii) Dilation and K Means

978-1-5386-8189-3/19/$31.00 © 2019 IEEE

V. CONCLUSION

K means algorithm performs well to detect breast tumor but it seldom fails if the image has too much noise. To make the K means more reliable, dilation and closing algorithm are considered because these when applied alone are not able to give satisfactory results. For more advantages, dilation with K means has been performed to reduce the effect of noise and hence to increase the accuracy by 10%. The tree process focuses on to suggest that the results obtained when the dilation algorithm is applied before K Means algorithm is more accurate than the individual. The overall accuracy is 60%-80% for process 1 and 85%- 90% for process 2.

REFERENCES

[1] Charan, S., Khan, M.J. and Khurshid, K., 2018, March. Breast cancer detection in mammograms using convolutional neural network. In Computing, Mathematics and Engineering Technologies (iCoMET), 2018 International Conference on (pp. 1-5). IEEE.

[2] Song, H., Men, A. and Jiang, Z., 2017. Breast tumor detection using empirical mode decomposition features. IEEE Access.

[3] Tan, Y.J., Sim, K.S. and Ting, F.F., 2017, November. Breast cancer detection using convolutional neural networks for mammogram imaging system. In Robotics, Automation and Sciences (ICORAS), 2017 International Conference on (pp. 1-5). IEEE.

[4] Sangeetha, R. and Murthy, K.S., 2017, January. A novel approach for detection of breast cancer at an early stage using digital image processing techniques. In Inventive Systems and Control (ICISC), 2017 International Conference on (pp. 1-4). IEEE.

[5] Paramkusham, S., Rao, K.M. and Rao, B.P., 2013, September. Early stage detection of breast cancer using novel image processing techniques, Matlab and Labview implementation. In Advanced Computing Technologies (ICACT), 2013 15th International Conference on (pp. 1-5). IEEE.

[6] Padmanabhan, S. and Sundararajan, R., 2012, December. Texture and statistical analysis of mammograms: A novel method to detect tumor in Breast Cells. In Machine Vision and Image Processing (MVIP), 2012 International Conference (pp. 157-160). IEEE.

[7] Selvathi, D. and AarthyPoornila, A., 2017, July. Performance analysis of various classifiers on deep learning network for breast cancer detection. In Signal Processing and Communication (ICSPC), 2017 International Conference on (pp. 359-363). IEEE.

[8] Qureshi, M.N. and Ahamad, M.V., 2018. An Improved Method for Image Segmentation Using K-Means Clustering with Neutrosophic Logic. Infinite Study.

[9] Paramkusham, S., Rao, K.M.M. and Rao, B.P., 2013. Automatic Detection of Breast Lesion Contour and Analysis Using fractals through spectral methods. In Advances in Computer Science, AETACS.

[10] Cahoon, T.C., Sutton, M.A. and Bezdek, J.C., 2000. Breast cancer detection using image processing techniques. In Fuzzy Systems, 2000. FUZZ IEEE 2000. The Ninth IEEE International Conference on (Vol. 2, pp. 973-976). IEEE.

[11] Yuan, L. and Xu, X., 2015, August. Adaptive image edge detection algorithm based on canny operator. In Advanced Information Technology and Sensor Application (AITS), 2015 4th International Conference on (pp. 28-31). IEEE

[12] Zhao, H., Qin, G. and Wang, X., 2010, October. Improvement of canny algorithm based on pavement edge detection. In Image and Signal Processing (CISP), 2010 3rd International Congress on (Vol. 2, pp. 964-967). IEEE.

[13] Wang, P., Meng, X., Zhang, K., Zheng, S. and Pan, R., 2013, May. Research on edge detection algorithm for plate image. In Control and Decision Conference (CCDC), 2013 25th Chinese (pp. 5123-5127). IEEE.

[14] Hasan, S.M.A. and Ko, K., 2016. Depth edge detection by image-based smoothing and morphological operations. Journal of Computational Design and Engineering, (pp. 191-197).

[15] Shrivakshan, G.T. and Chandrasekar, C., 2012. A comparison of various edge detection techniques used in image processing. International Journal of Computer Science Issues (IJCSI).

[16] Fathy, M. and Siyal, M.Y., 1995. An image detection technique based on morphological edge detection and background differencing for real-time traffic analysis. Pattern Recognition Letters, (pp.1321-1330).

[17] Jin-Yu, Z., Yan, C. and Xian-Xiang, H., 2009, April. Edge detection of images based on improved Sobel operator and genetic algorithms. In Image Analysis and Signal Processing, 2009. IASP 2009. International Conference on (pp. 31-35). IEEE.

[18] Davis, L.S., 1975. A survey of edge detection techniques. Computer graphics and image processing, (pp. 248-270).

[19] Kapoor, P. and Prasad, S.V.A.V., 2010, February. Image processing for early diagnosis of breast cancer using infrared images. In Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on (Vol. 3, pp. 564-566). IEEE.

[20] Dhanachandra, N., Manglem, K. and Chanu, Y.J., 2015. Image segmentation using K-means clustering algorithm and subtractive clustering algorithm. Procedia Computer Science, 54 (pp.764-771).

[21] Prakash, R.M., Bhuvaneshwari, K., Divya, M., Sri, K.J. and Begum, A.S., 2017, March. Segmentation of thermal infrared breast images using K-means, FCM and EM algorithms for breast cancer detection. In Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017 International Conference on (pp. 1-4). IEEE.

[22] Sandhya, G., Gowda, M.S., Swamy, L.N., Raju, G.T. and Vasumathi, D., 2013, December. Automated detection of cancer tissues in mammograms using advanced K-Means clustering with homomorphic filtering. In Circuits, Controls and Communications (CCUBE), 2013 International conference on (pp. 1-4). IEEE.

[23] Filipczuk, P., Kowal, M. and Obuchowicz, A., 2012, May. Breast fibroadenoma automatic detection using k-means based hybrid segmentation method. In Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on (pp. 1623-1626). IEEE.

[24] Chen et al., 2010, September. Edge-adaptive image segmentation based on seam processing and K-Means clustering. In Image Processing (ICIP), 2010 17th IEEE International Conference on (pp. 3049-3052). IEEE.

[25] Ng, H.P. et al., 2006, March. Medical image segmentation using k-means clustering and improved watershed algorithm. In Image Analysis and Interpretation, 2006 IEEE Southwest Symposium on (pp. 61-65). IEEE.

[26] Balti, A., Sayadi, M. and Fnaiech, F., 2011, March. Segmentation and enhancement of fingerprint images using K-means, fuzzy C-mean algorithm and statistical features. In Communications, Computing and Control Applications (CCCA), 2011 International Conference on (pp. 1-5). IEEE.

978-1-5386-8189-3/19/$31.00 © 2019 IEEE