ISSN: 0974-2115
www.jchps.com Journal of Chemical and Pharmaceutical Sciences

Hybrid methods to isolate breast tumors in mammography images

Rabab Saadoon Abdoon*, Ola Saad Khudair
Department of Physics, College of Science, University of Babylon, Hilla, Iraq
***Corresponding author: E-Mail: sr614@ymail.com**

ABSTRACT

Four mammograms, for women suffer from breast tumor, were adopted to test the performance of five proposed segmentation methods to isolate and extract tumor regions. One of the utilized images was for cancerous tumor case, while the other images were for benign. Clustering of hard scheme; soft scheme clustering; GLCM; and two adaptive hybrid techniques were implemented to extract breast tumors. K-means algorithm was implemented as hard scheme clustering with different number of clusters. In addition, soft clustering, FCM algorithm was applied with different number of segments too. First adaptive hybrid technique depends on k-means and GLCM algorithms, while the second one bases on k-means and FCM. Besides, many morphological operations were applied to get rid of any extra pixel does not belong to the tumor regions. Surface area and relative surface area of the extracted tumor regions were calculated depending on the whole area of the breast. It was found that, five is the adequate number of clusters to segment mammography images correctly. The proposed methods adequately succeeded to isolate and extract breast tumors. The extracted tumor regions were in good agreement with the radiologist delineation of the tumor regions in the adopted mammograms.

KEY WORDS: Clustering, GLCM, K-Means, FCM, Mammograms, Breast Tumors.

1. INTRODUCTION

Mammograms are low-dose x-ray images of the breast, they are commonly utilized test for early detection and characterizing breast cancer and other abnormalities in breast. Part of the emitted x-ray is absorbed by the breast while the other parts go through the breast and then hit a screen. This screen emits light and blackening a film that is located in front of it. Some of the x-rays exit the breast in different angles (scattering effect) causing image blurring. Scattered x-rays can be reduced by positioning a grid in front of the screen to absorb the scattered x-rays, but this situation requires higher dose x-rays (Maria, 2000). Digital mammograms are breast images that are stored in computer in an electronic form to make them computer readable images. Digital mammography overcomes the limitation of film mammography because it has the potential advantages: wider dynamic range and lower noise, improved image contrast and enhanced image quality (Jinshan, 2009).

Many researchers proposed different methods to segment mammograms in order to separate its contents and to isolate abnormal regions within breast. The applied methods ranged from simple image processing such as thresholding and filtering to more complicated approaches that depending on a statistical analysis of local appearance of mammogram image, utilizing wavelet transform, or implementing clustering etc., for more details see (Jinshan, 2006; Wei, 2008; Saheb, 2009; Indra, 2011; Al Mutaz, 2011; Dalmiya, 2012; Ramani, 2013; Singh; 2015).

In this study, six segmentation techniques were proposed to isolate and extract tumor regions in mammograms. These techniques are two schemes of clustering; texture feature based GLCM; enhancing based method and two adaptive hybrid techniques.

Hard Scheme K-Means Clustering: K-means is a simple unsupervised learning algorithm that solve clustering problem of hard scheme. The procedure of k-means is simple and easy. It implemented to classify a given data set by adopting a certain number of clusters fixed in a prior, and then partitioning the image's objects into many clusters in such a way that (objects) pixels within each cluster remain as close as possible to each other but far away as possible as from objects in the other clusters (Ghosh, 2013). The main idea of it, is to define k centers points, one for each cluster. The better choice of the centers location is to place them far away from each other as possible as. The next step is to associate each point in data set to the nearest center. When no point is pending, the process re-calculate new k center points of the clusters resulting from the previous step. After the new k centroids are found, a new binding has to be done between the same data set points and the nearest new center. This circulating process will continue to change k centers location step by step until no more changes are done and the centers do not move any more. K-means algorithm aims to minimize the objective function that represents the function of squared error which given by (Geethu, 2011):

$$\mathrm{J(V)} = \sum_{i=1}^{c}\sum_{j=1}^{c_i}(\|x_i - v_i\|)^2 \qquad \ldots\ldots\ldots\ldots (1).$$

Where $||x_i - v_j||$ represents the Euclidean distance between x_i and v_j; c_i is the number of data points in i^{th} cluster and c is the number of cluster centers.

Soft Scheme FCM Clustering: Soft clustering, FCM algorithm works by adopting membership grade to each dataset point corresponding to each cluster center point depending on the distance between the cluster center and the data point. As the data is more near to the cluster center, its belonging towards this cluster will be high. On the other hand, data point that lies far away from the center point of a cluster will have a low degree of membership to that cluster. The summation of membership of each data point must equal to one (Ghosh, 2013; Nikhil, 2005). The membership and cluster centers are updated after each iteration according to this formula (Ghosh, 2013):

$$\mu_{ij} = 1/\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{ik}}\right)^{\left(\frac{2}{m-1}\right)} \qquad \text{............ (2).}$$

The center vector can be calculated by:

$$v_i = \frac{\sum_{i=1}^{n}(\mu_{ij})^{m}x_i}{\sum_{i=1}^{n}(\mu_{ij})^{m}}\ , \quad where\ 1 \le j \le c \qquad \text{............ (3).}$$

Where 'n' is the number of data points.'v_j' represents the j^{th} cluster center. 'm' is the fuzziness index m € [1, ∞]. 'c' represents the number of cluster center.'μ_{ij}' represents the membership of i^{th} data to j^{th}, 'd_{ij}' represents the Euclidean distance between i^{th} and j^{th} cluster center data.

The objective function of fcm algorithm that have to be o minimized (Nikhil, 2005):

$$\mathrm{J(U,V)} = \sum_{i=1}^{n}\sum_{j=1}^{c}(\mu_{ij})^{m}\,\|x_i - v_i\|^2 \qquad \text{............ (4).}$$

Where, '$||x_i - v_j||$' is the Euclidean distance between i^{th} data and j^{th} cluster center(d_{ij}).

Gray Level Co-Occurrence Matrix: Texture is one of the most commonly used features interpret images, specifically medical images. Texture is a measure of the variation of the intensity of a surface quantifying properties like smoothness, coarseness, and regularity. It is used as a region descriptor in image analysis. Specifically, a textured region contains connected set of pixels that satisfy a given gray-level property that occurs repeatedly in an image region. Several methods have been applied towards the analysis and characterization of texture within medical images. One of these methods, two-dimensional dependence matrices are extensively used, they are able to capture the spatial dependence of gray-levels which contributes to the perception of texture (Chang, 1999). There is a significant variation in intensity levels between nearby pixels and within the limit of resolution (Haralick, 1973), for more details, see (Haralick, 1973; Abdoon, 2015). In this work, gray level co-occurrence matrix, glcm, that based on the texture features of processed image, was implemented to segment mammograms.

Hybrid Method Depending On K-Means and GLCM: The first hybrid method depends on k-means and GLCM. In this hybrid method, k-means segmented image was utilized as input image to GLCM algorithm to emphasize the segmentation process and to improve the performance of each of k-means and GLCM algorithms.

Hybrid Method Depending On K-Means and FCM: The second hybrid method based on k-means and FCM. In this technique, FCM algorithm was fed by k-means clustered image instead of raw mammogram image to test the performance of FCM after two pass segmentation process and to reduce the required long elapsed time of FCM algorithm.

2. EXPERIMENTAL DATASET

The adopted dataset of this work was acquired from Hilla Teaching Hospital, Province of Babylon, Iraq. These adopted images are four breast mammograms for women suffer from breast tumor. One of the utilized images is for cancerous tumor case, while the other images were for benign tumor cases. The experimental images are named from image1 to image4, each of them has 3062x2394 pixels of size. Fig.1, shows these input images.

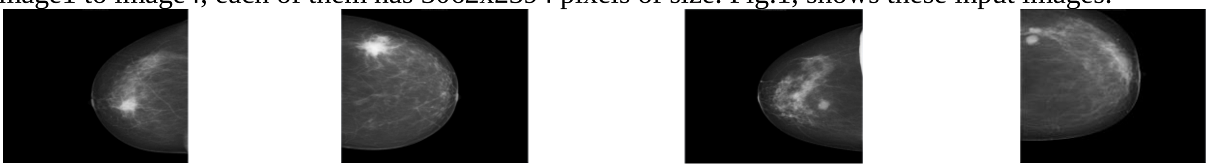

Figure.1. Input mammograms images, the second image contains cancerous tumor

Methodologies: In this work, six techniques are proposed to segment four mammography images to isolate and extract tumor region. The implemented procedure is illustrated in the block diagram of figure.2.

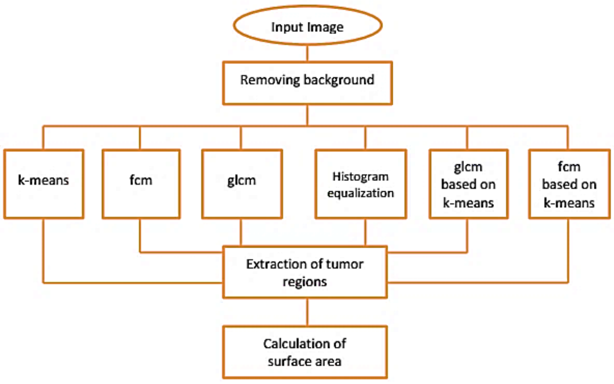

Figure.2. Block diagram of the implemented procedure

3. EXPERIMENTS AND RESULTS

The experiments and results are presented as follows:

Background Removing: The first step of this work is removing the background to minimize the number of mathematical processes and to reduce implementing time. This process was achieved automatically by using image moments depending technique that proposed firstly by (Abdoon, 2015). Figure 3 presents the results of this step.

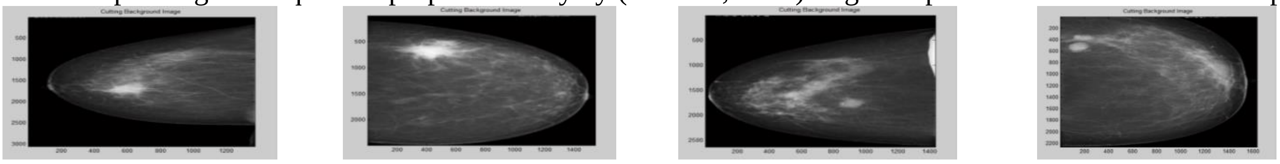

Figure.3. Input mammograms after removing background

Hard Scheme K-Means Clustering: Different number of clusters were adopted to cluster the experimental images by implementing k-means algorithm. It was found that five clusters are the suitable number of clusters to segment breast mammograms and the resultant segmented images of this method for the four images are presented in figure.4, first line, while the extracted tumor regions are shown in second line.

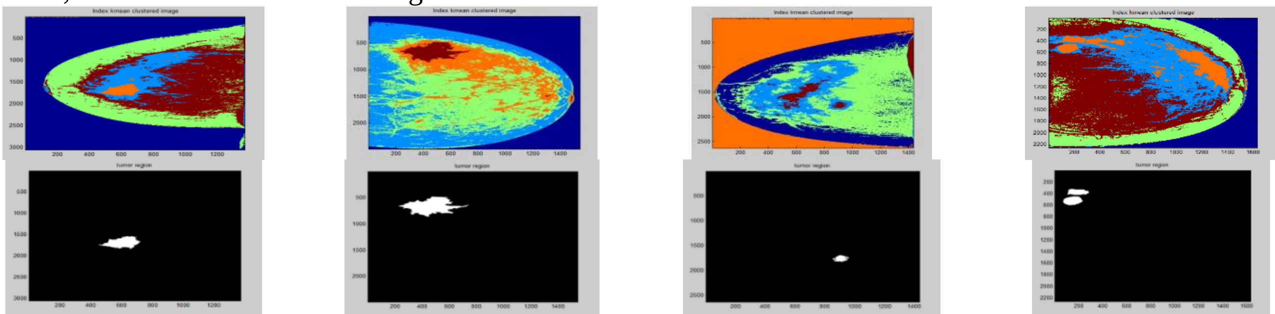

Figure.4. Results of implementing k-means algorithm on image1 to image4 from left to right respectively. Tumor regions are extracted by applying morphological operations with disk shaped element and choosing the tumor regions

Soft Scheme FCM Clustering: FCM algorithm was applied with different number of clusters to segment the experimental mammograms. It found that five clusters are the proper number of clusters to cluster breast mammogram images into its contents and the resultant segmented images and the extracted tumor regions of this algorithm are showed in figure.5, first and second row respectively.

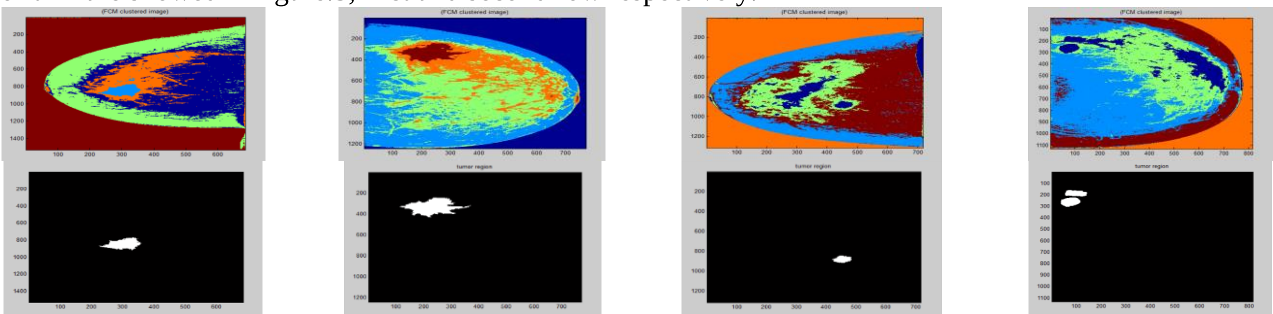

Figure.5. Results of implementing fcm algorithm on image1 to image4 from left to right respectively

Gray Level Co-Occurrence Matrix, GLCM: In this step of the work, GLCM algorithm was implemented to segment the experimental mammograms to isolate the breast tumor regions. The results of this method are illustrated in figure.6.

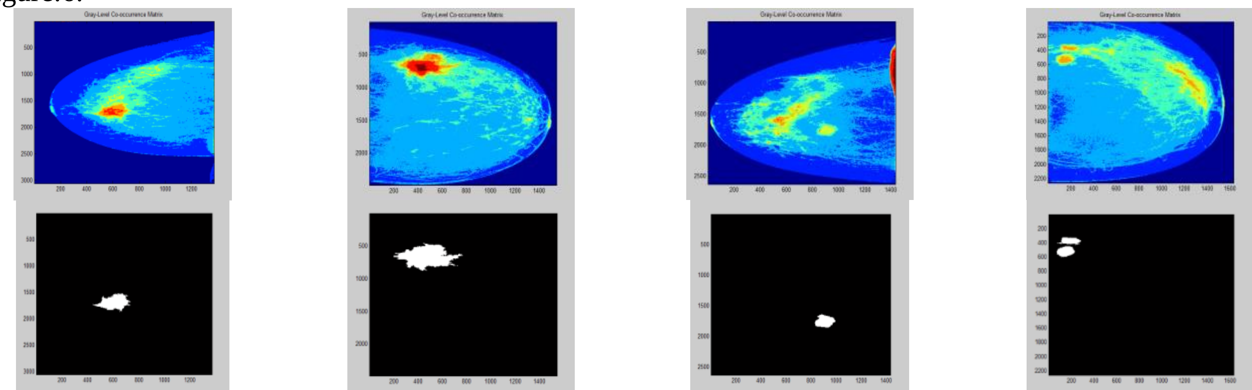

Figure.6. Results of applying glcm algorithm on image1 to image4 from left to right respectively

Hybrid Method Based of K-Means and GLCM: K-means segmented images were utilized as input images of GLCM algorithm to improve the performance of each of them. The results of this hybrid method are presented in figure.7.

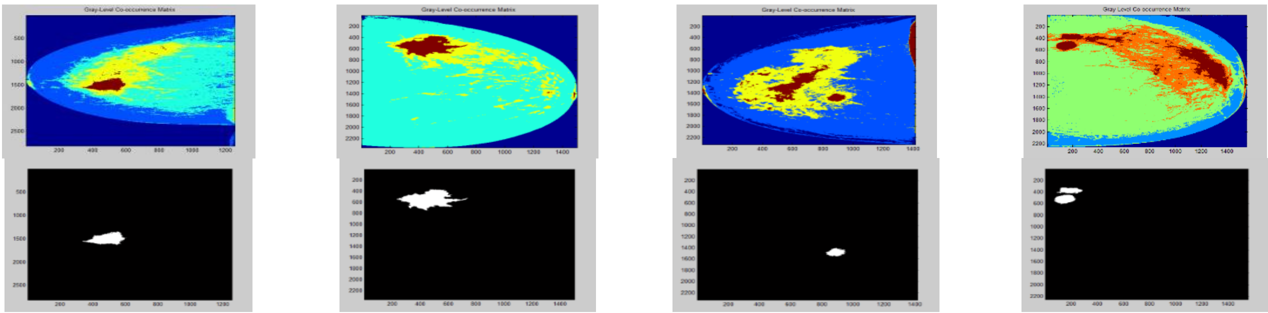

Figure.7. Results of applying first hybrid method on image1 to image4 from left to right respectively

Hybrid Method Based of K-Means and FCM: K-means segmented images were fed as input images of FCM algorithm to reduce the long required elapsed time of FCM and to test the performance of each of k-means and FCM. The results of this hybrid technique are showed in figure.8. The manually delineation of the tumor regions by the radiologist is illustrated in figure.9, in blue.

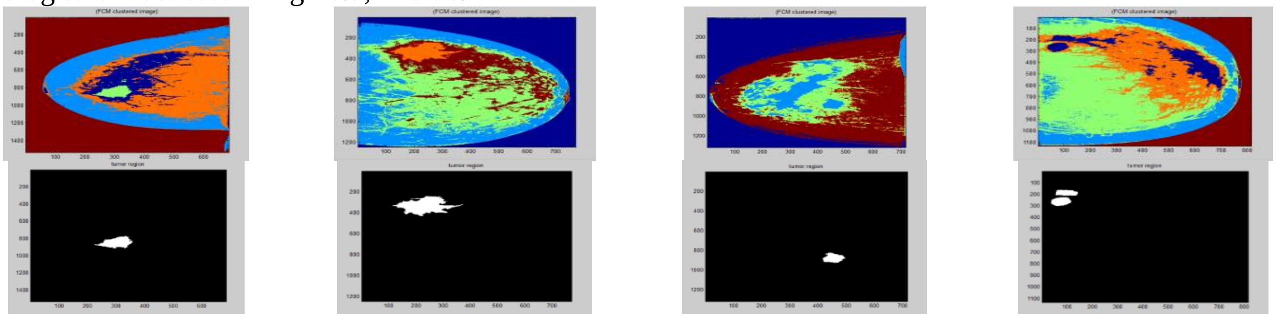

Figure.8. Results of implementing second hybrid method on image1 to image4 from left to right respectively

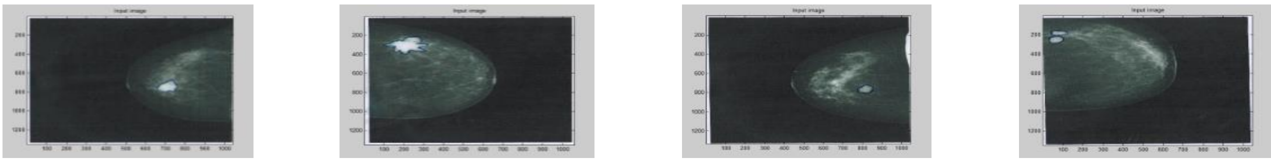

Figure.9. Doctor delineation of the tumor regions in blue color on image1 to image4 from left to right respectively

The surface area of the extracted tumor regions was calculated and presented in table.1. The relative surface area of the extracted tumor regions was calculated after determining the area of the corresponding whole breast. The whole area of the breast was found after segmenting each mammogram image into two segments only, background and foreground. The values of the calculated areas of whole breast were: 1836038, 2694011, 1840592 and 2551689 pixels for image1 to image4 respectively. Table.2, presents the calculated values of the relative surface area.

Table.1. Calculated surface area of the extracted tumor regions by implementing the proposed techniques

Image	**Surface Area (pixel)**				
	K-means	**FCM**	**GLCM**	**GLCM based on K-means**	**FCM Based on K-means**
Image 1	45352	44008	51569	42769	42744
Image 2	93819	89868	110857	93819	94588
Image 3	22195	21768	24497	22195	22028
Image 4	32185	32284	31877	32185	33712

Table.2. Relative surface area of the extracted tumor regions by implementing the proposed techniques

Image	**Relative Surface Area**				
	K-means	**FCM**	**GLCM**	**GLCM based on K-means**	**FCM Based on K-means**
Image 1	0.02470	0.02396	0.02808	0.02329	0.02328
Image 2	0.03482	0.03335	0.04114	0.03482	0.03511
Image 3	0.01205	0.01182	0.01330	0.01205	0.01196
Image 4	0.01266	0.01265	0.01249	0.01261	0.01321

4. CONCLUSIONS

In this study, four mammograms, contain tumor regions, were adopted to test the performance of five proposed segmentation methods to isolate and extract tumor regions. Clustering k-means and FCM; GLCM; and two adaptive hybrid techniques: first one based on GLCM and k-means; while the other depends on k-means and FCM; were implemented to extract breast tumors. K-means algorithm was implemented as hard scheme clustering method with different number of clusters. In addition, soft clustering, FCM algorithm was applied with different number of segments too. From the results, it was found that, five clusters are the adequate number of clusters to segment mammography images of breast correctly utilizing the two schemes of clustering. The results proved that, the proposed methods adequately succeeded to isolate and extract breast tumors. The differences in some area values of the extracted tumor regions belong to the utilized disk radius of morphological structuring elements. The resulted extracted tumor regions were in good agreement with the delineation of the radiologist doctor of the tumor regions in the adopted mammograms.

5. ACKNOWLEDGEMENT

Authors like to express their thanks to radiologist doctor Fatin Fadil Al-Saffar for delineation of tumor regions of the adopted mammograms**.**

REFERENCES

Abdoon RS, Abood LK, Ali SM, Adaptive Techniques for Brain Tumor Detection in MRI Lap LAMBERT Academic Publishing, Germany, 2015.

Al Mutaz MA, Safaai D, Nazar Z, Masses Detection in Digital Mammogram by Gray Level Reduction using Texture Coding Method, International Journal of Computer Applications, 29, 2011, 19-23.

Cascio D, Fauci F, Magro R, Raso G, Bellotti R, De Carlo F,Tangaro S, DeNunzio G, Quarta M, Forni G, Lauria A, Fantacci ME, Retico A, Masala GL,Oliva P, Bagnasco S, CheranS C, Lopez TE, Mammogram Segmentation by Contour Searching and Mass Lesions Classification with Neural Network, IEEE Transactions On Nuclear Science 53, 2006, 2827.

Chang KI, Bowyer KW, Sivagurunath M, Evaluation of Texture Segmentation Algorithms IEEE Conf on Computer Vision and Pattern Recognition, 1, 1999, 294.

Dalmiya S, Dasgupta A, Datta SK, Application of Wavelet based K-means Algorithm in Mammogram Segmentation, International Journal of Computer Applications, 52, 2012, 15.

Geethu GL, Suruliandi A, Anatomical Structure Segmentation in MRI Brain Images Proceedings of ICETECT, IEEE Xplore 786, 2011.

Ghosh S, Dubey SK, Comparative Analysis of K-Means and Fuzzy C-Means Algorithms, (IJACSA) International Journal of Advanced Computer Science and Applications, 4, 2013, 35.

Haralick RM, Shanmugam K, Dinstein I, Textural Features for Image Classification, IEEE Trans on Systems, Man, and Cybernetics, Smc-3,1979, 610.

Indra KM, Sanjay N, Samir KB Identification of Abnormal Masses in Digital Mammography Images, International Journal of Computer Graphics, 2, 2011, 17.

Jinshan T, Rangaraj MR, Jun X, Yongyi Y, Computer-Aided Detection and Diagnosis of Breast Cancer with Mammography, Recent Advances IEEE Transactions on Information Technology in Biomedicine 13, 2009, 236.

Maria GB, Computer Aided Detection of Masses in Digital Mammograms, Ph. D thesis, Electrical and Information Engineering, Radboud University of Nijmegen, Netherlands, 2000.

Nikhil RP, Kuhu Pal, James MK, and James CB, A Possibilistic Fuzzy c-Means Clustering Algorithm, IEEE Transactions on fuzzy systems, 13, 2005, 517.

Ramani R, SuthanthiraVanitha N, Valarmathy S, The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images, IJ Image, Graphics and Signal Processing, 5, 2013, 47.

Saheb SB, Satya K Prasad, Automatic Detection of Breast Cancer Mass in Mammograms Using Morphological Operators and Fuzzy C-Means Clustering, Journal of Theoretical and Applied Information Technology, 2009.

Singh AK and Gupta B, A Novel Approach for Breast Cancer Detection and Segmentation in a Mammogram Procedia Computer Science, 54, 2015, 676.

Wei P, Li J, Zhao S, Lu D and Chen G, a method of detection micro-calcifications in mammograms using wavelets and adaptive thresholds IEEE, Bioinformatics and Biomedical Engineering, 8, 2008, 2361.