Journal of Engineering and Applied Sciences 14 (15): 5044-5050, 2019
ISSN: 1816-949X

Automatic Segmentation of Breast Mammograms Using Hybrid Density Slicing and k-mean Adaptive Methods

Semaa Ibrahim M. Ali and Nassir H. Salman
Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq

Abstract: Medical imaging is a fundamental theme of contemporary healthcare and its engineers take mammograms, ultrasounds, X-rays and computed tomography images to analyze patient's hurts and illnesses. Segmenting the mammogram into diverse mammographic densities is strategic for risk evaluation and measurable appraisal of density variations to extract the cancer regions. Accordingly in this study, the application of density slicing and k-mean adaptive techniques has been conducted to explore the boundary of changed breast tissue areas in mammograms. The objective of the segmentation process is to perceive if density slicing and k-mean adaptive procedure have the feasibility split diverse densities for the diverse breast outlines. The density slicing is used to make available hard limitation while the thresholds are designated in accordance with user-defined and radiology. k-mean adaptive has been used to cluster the region where the initial seed was based on the mean of array multiply by 0.05. Density slicing has processed on images of numerous imaging modalities without mammograms consideration. As a result, this study is for all intents and purposes concentrated on using hybrid method of density slicing and k-mean adaptive process to achieve segmentation to augment the discernibility of diverse breast densities in mammography images. The suggested approach for the segmentation of mammograms on the source of their region into diverse densities based classifications has been investigated on mini-MIAS database. The concluding consequences show instinctive segmentation of ROI with edge map and dissimilar properties extraction for the investigation process.

Key words: Mammograms, breast cancer, image segmentation, image processing, adaptive, k-means and density slicing

INTRODUCTION

Newly, Computer-Aided Design (CAD) is a division of the predictable clinical work for detecting breast cancer on mammograms at various hospitals and screening positions in numerous nations. This indicates that CAD starts to be fundamentally in the discovery and variance analysis of countless categories of irregularities in medicinal images acquired in numerous investigations by using dissimilar imaging modalities. Actually, CAD has been one of the foremost study themes in medicinal imaging and analytical radiology. Despite primary efforts at programmed investigation of medicinal images were produced in the 1960's, thoughtful and methodical examination on CAD was initiated in the 1980's with a major transformation in the notion for application of the computer output, from automatic computer diagnostics to computer-aided analysis (Kooi *et al*., 2017).

Medical images play an imperative function in facilitating analysis and treatment and can be used to teach health care students and explaining these images will help them learn. Developments in digital imaging technology have greatly increased the number of digital images captured in recent years (Miranda *et al*., 2016). Interpreting and analyzing medical images is an important and exciting part of computer vision and pattern recognition. The development of computer-assisted diagnostic systems for cancer such as breast cancer has become very important for hospital physicians and has become a top priority for many clinical researchers and centers (Chakraborty *et al*., 2012, 2018).

A number of methods and systems for computer aided identification was applied for the masses detection (Henriksen *et al*., 2016). By Karssemeijer and Te Brake (1996) an old multiscale-based methodology was used for the detecting stellate patterns interrelated with masses and architectural deformation. This technique exhibits a good performance with a sensitivity of 90% using the MIAS database.

By Mudigonda *et al*. (2001) a density slicing method has been adopted for detecting of distrustful districts of masses and investigated the adapted flow-like textural

Corresponding Author: Semaa Ibrahim M. Ali, Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq

information in mammograms to categorize the distrustful districts as usual body districts or masses. This method shows 81% sensitivity with the mini-MIAS dataset of 56 images comprising 30 benign, 13 malignant and 13 normal samples. A development technique using an iris filter with adaptive thresholding has been employed as reported by Varela *et al*. (2007). An 88% sensitivity was realized with a lesion-based method for evaluation using a dataset of 66 cases with 130 images with masses and 134 usual images. Nevertheless, using iris filter per pixel creates the process time-consuming. A multilevel thresholding technique for the segmenting suspicious districts for designing programmed masses detection was presented by Dominguez and Nandi (2008). This technique accomplished 80% sensitivity with 57 mammographic masses images from the mini-MIAS database.

Typically, masses have an intensive density focal district bounded by districts with increasingly lessening values. Accordingly, systems using multiple concentric layers for detecting mass as reported by Eltonsy *et al*. (2007) and Gao *et al*. (2010). A multilevel thresholding method has been used in these methods, starting from extreme to small intensity levels with a constant threshold value decrement to detect focal districts. Various concentric layers were used for detecting mass districts. While these systems have been operative, numerous parameters require serious optimization. Also, these systems may not identify the correct mass locations, if mass regions blocked by dense tissues. By Chakraborty *et al*. (2012, 2018) region growing technique organized by multilevel thresholding was given in effect performance in detecting doubtful mass regions in mammograms. By using (MIAS and local) datasets of good size, the projected technique was given sensitivity better than 90%. A fuzzy semi-supervised variety based on a modified grow cut set of rules has been reported by Cordeiro *et al*. (2016) to segment and categorize regions of interest of mammographic imaginings with a high quality of accuracy. Simple breast tissue segmenting attitude has been given in Thakran *et al*. (2018) using multi-parametric MRI images for a breast tumor. The suggested methodology hoffers automatic segmenting of fatty, FG and tumor tissues with fast and accurate results. Breast cancer is a group of abnormal cells is discovered in the breast tissues. Recently, there are frequent modalities established for the detecting breast cancer. In tissue testing Ahmad *et al*. (2012), the tissue removal is occupied from the breast tissues. The test has a highly precise result but the process to remove the tissue from the breast is very hurting and wretched. Consequently, the majority of the patients are not interested in this investigation. A mammogram (Islam *et al*., 2017) is extensively employed a method for the detecting breast cancer that has the 2D projection images of the breast. The proposed method in this study starts with the density slicing with a selected threshold by user define and radiology, then used k-mean to cluster the region which formula uses three homogeneity standards. If the pixel is nearer to the k-mean as compared to its adjacent pixels, then it is involved in the region. Otherwise, it is considered as a boundary pixel. This technique was presented for color image segmentation. The results showed a very good automatic image segmentation and edge map extracting.

MATERIALS AND METHODS

Used dataset: The dataset used in this study is an international data collection (MIAS), a British research organization interested in understanding mammograms. The original 322 images (161 pairs) are included with a resolution of 50 μ in PGM format and the real data associated with them. Films filmed in the National Breast Screening Program (NBSP) in the UK. The image is numbered to a 50 μ pixel and represented in 8 bit words per pixel. Minimized to 200 μm pixels and filled all 1024×1024 images (including 202 normal and 120 abnormal images) (Fig. 1 and 2).

Convert color dataset images into grayscale images: A grayscale digital image stands for an image that each pixel value has distinctive intensity information. This category of images, black-and-white is comprised completely of shades of gray, fluctuating from black at the vulnerable intensity to the strongest white level. Usually, each pixel value lies between 0 and 255. Sample images of mammogram images are shown in Fig. 1 and 2. In a mammogram image each pixel has one color (grayscale), that is carries only the intensity information. Pixels are ranged from the least intense (black) to the most intense (white). Pixel amounts are typically ranged from 0-255. A grayscale color is one in which the red, green and blue constituents have identical intensity in RGB space (Gonzalez and Woods, 2002). Consequently, one band of them can be used as in Fig. 3.

Removing noise and selecting region of interest: We used to the images dual methods: an image enhancement and a cropping operation one. The 1st one was applied a median filter to remove noise. The 2D median filtering has been used in a 3-by-3 region joining to take out the noise. Subsequently, it was employed to cut the black fragments of the image in addition to the in effect objects like written labels. For the images of the dataset, practically 50% of

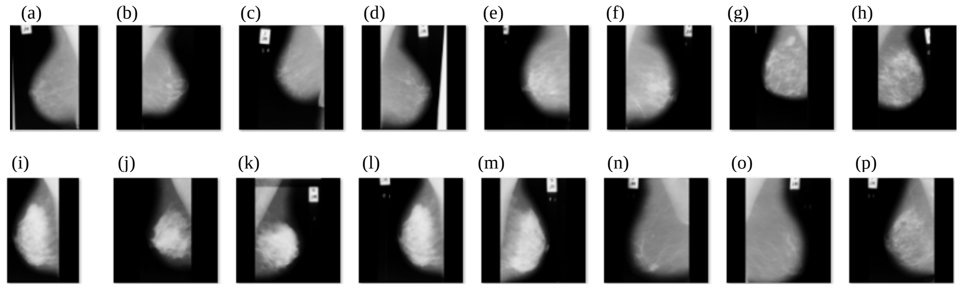

Fig. 1: Some of the mammogram images of the breast taken from dataset MIAS): a) mdb; b) mdb001; c) mbd002; d) mbd003; e) mbd004; f) mbd005; g) mbd006; h) mbd007; i) mbd008; j) mbd009; k) mbd0010; l) mbd011; m) mbd012; n) mbd013; o) mbd014; p) mbd015 and) mbd016)

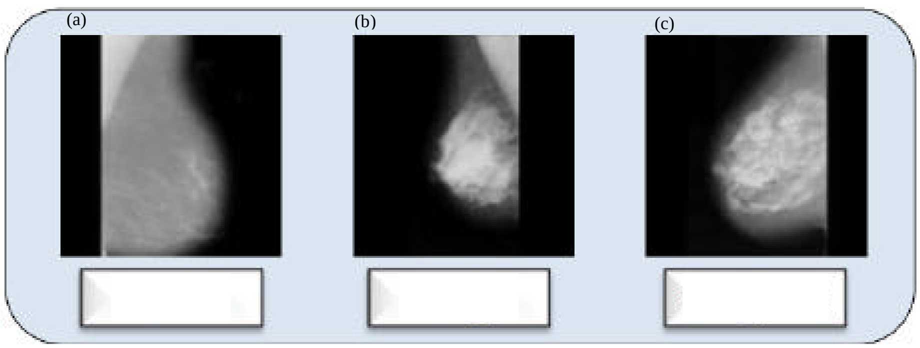

Fig. 2: Samples of Benign and Malignant mammogram images: a) Normal; b) Benign and c) Malignant

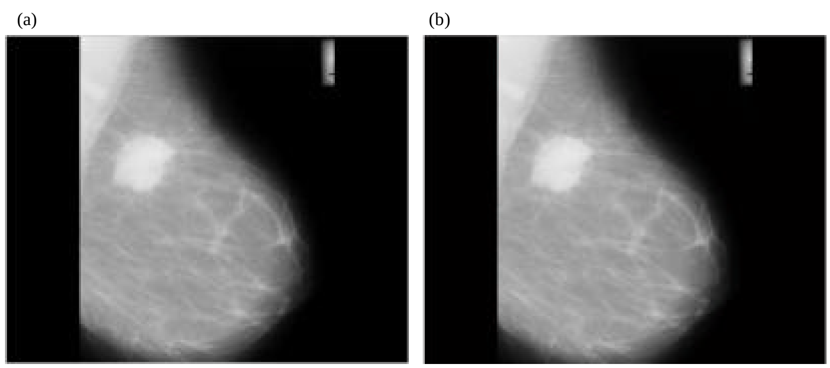

Fig. 3: a) Original image and b) Gray level mammogram image

the entire images involved a black background with noteworthy noise. Cropping can be used to remove the undesirable portions of the image, typically in the marginal areas of interest. Primarily, a threshold magnitude has been used to convert gray mammogram image to a binary mammogram. This threshold has computed from the smallest intensity magnitudes amid the preliminary dual weightiest peaks of mammogram image histogram. Those peaks characterize the background and the breast. Then, the associated constituent is employed to extract the prevalent constituent of the breast image that increases the grayscale image contrast by converting the values by Contrast-Limited Adaptive Histogram Equalization (CLAHE) method. Subsequently, the grayscale image is converted to a binary image. The resultant image BW swaps entirely pixels in recorded image with luminance bigger than white level with 0.08 value and substitutes all other pixels with the value of 0 for black level. The connected components found in BW. By computing the number of connected objects to evaluate the area of them, the maximum area is within the breast region. So, the lowest and highest axies and lowest and highest y-axes are detected (Gonzalez and Woods, 2002). Subsequently, the region of interest from the original image is cropped as illustrated in Fig. 4.

Removing pectoral muscle: Mammograms illustrate a projection of the breast that is feasibly prepared from

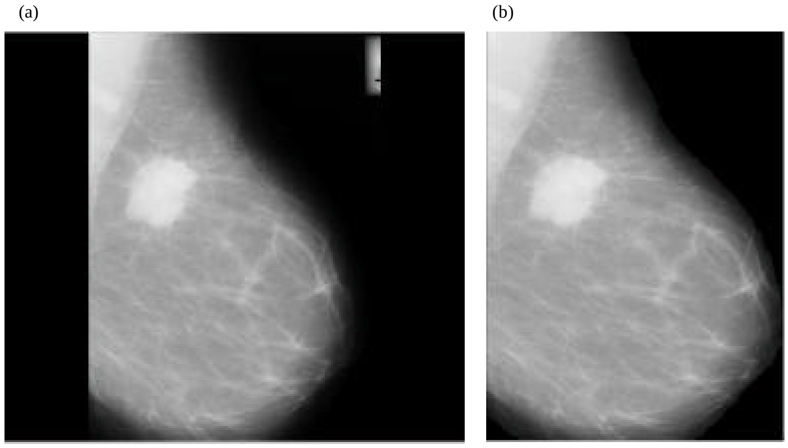

Fig. 4: a) Remove noise by the median filter and b) Select ROI

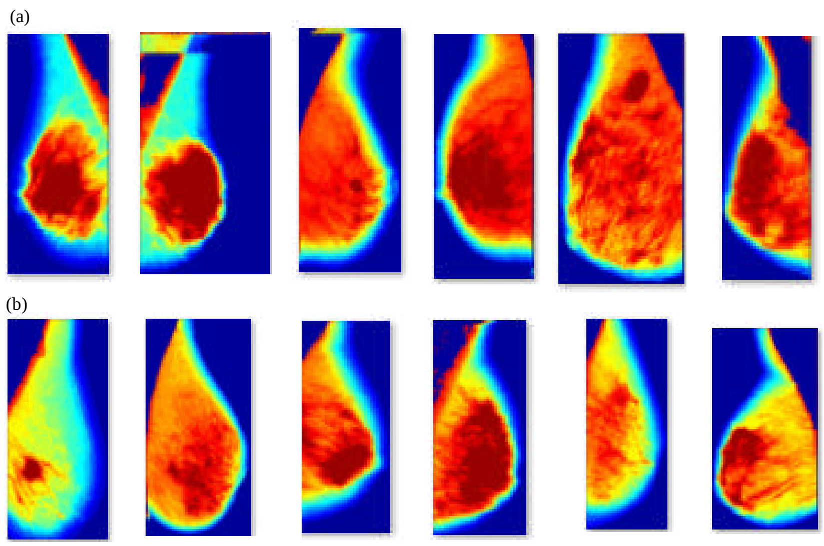

Fig. 5: Density slicing method results tested on the dataseta: a) Benign test) results samples Image 1-6 and b) Image 1-6 malignant test results

dissimilar angles. The dualistic utmost mutual projections are mediolateral oblique and craniocaudal. The benefit of the mediolateral oblique projection is that practically the entire breast is discernible including lymph nodules. The foremost drawback is the fragment of the displayed pectoral muscle in the higher image fragment that is covered over a breast portion. The craniocaudal vision is reserved from above, consequential in an image that on occasion does not display the area on the point of the chest wall. In this study, the previous one has a specific advantage but pectoral muscle recognition is a tricky task in the breast segmentation practice. It is essential to detect the pectoral muscle and describes the Region of Interest (ROI) for auxiliary investigations. Firstly, regulate image intensity values and convert it to BW. Then, subtract the cropped image and the adjusted image will be close to the morphological operation to the binary image BW. Finally, fill all the hole in the image and display the cropped image without the filled image.

Density slicing: The mammogram represents typically low contrast and hence, contrast enrichment step is required for this resolve by using filtering techniques. Density or level slicing splits pixel values into sequences of intervals or "slices" with diverse colors applicable to each slice. This method is frequently implemented on distinct band images to highpoint transformations. It is frequently employed to show changes in vegetation and in thermal imagery (Mudigonda *et al*., 2001) (Fig. 5).

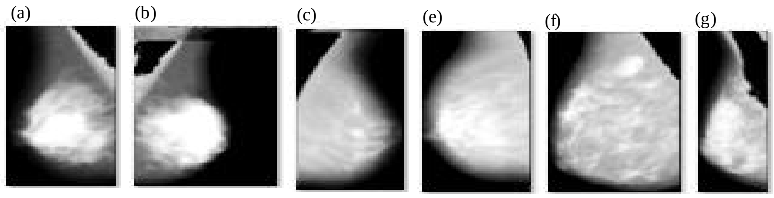

Fig. 6: The samples of gray level threshold images: a) Image 1; b) Image 2; c) Image 3; d) Image 4; e) Image 5 and f) Image 6

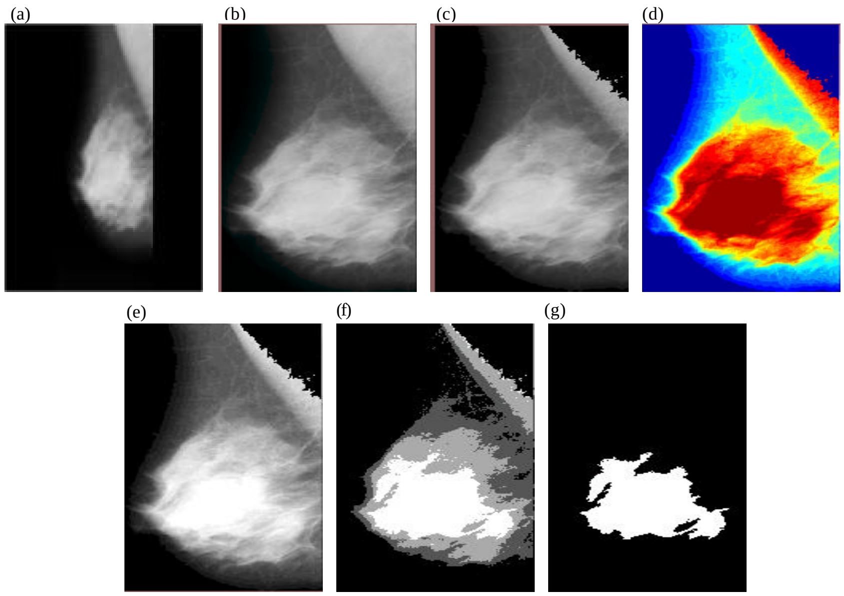

Fig. 7: a) Original image mdb001; b) Remove label and background; c) Remove pectoral muscle; d) Convert to density slicing; e) Convert to gray level with threshold 20; f) Cluster the image by adaptive k-mean and g) Convert to black and white with threshold 0.79

Gray-level threshold: a level threshold value (gray-level) is selected in this method to facilitate entirely pixel values below this threshold to be drawn to zero and an upper threshold value is as well selected with the intention of all pixel values higher than this threshold to be mapped to 255. These can be utilized to generate a binary mask for an image. Regularly, these masks can be employed to segment an image into dual classes (binary) with the purpose of applying supplementary processing to each class autonomously (Gonzalez and Woods, 2002) (Fig. 6).

Adaptive k-mean: The main disadvantage of the k-means process is the clusters number that will divide up the data set, defined by the user as an input factor. The texture, intensity and shape in mammogram images database are varying from one image to another. Thus, considering a constant regions number and the identical regions for entirely database images is inappropriate. Hence, two methodologies have proposed for with dynamism and routinely determined regions of interests for individual mammogram. So, we started using k-means clustering to divide a mammogram to different density tissue region, automatically selected the seed point to determinate threshold values of each region (Elmoufidi *et al*., 2015).

The k-means process is straightforward unsupervised learning procedures that resolve the recognized clustering problem. The procedure has uncomplicated procedural steps to categorize a particular data set throughout a certain k clusters number. The attitude of the k-means procedure is as follows (Fig. 7).

Describe K cluster centers, either arbitrarily or using particular heuristics. Every pixel allocating to the adjacent cluster is based on the smallest Euclidean distance amid the point and the centers of K cluster. Re-calculate the centers of cluster. Iterate step 2 and 3 by a loop. This criteria will break the loop if the center does not move. For a specified set of N observation {S1, 2, ... , SN}, k-means process segments the remark into k cluster {C1, 2, ... , Ck}. Their cluster center is {μ1, 2, ... , μk} (k<N) in an attempt to diminish the input cluster sum of squares as in Eq. 1 stands for the mean of ith in during each iteration that can be calculated by:

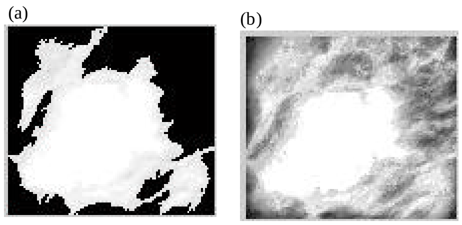

Fig. 8: a) ROI segment and b) Enhancement by histogram equalization

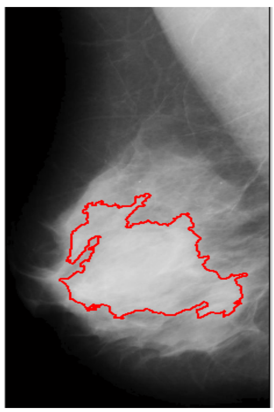

Fig. 9: Automatic edge map detection of ROI

$$m_i = \frac{\sum_J^{C_i} X_j^i}{n_j}$$

The planned method steps of this for resultant Fig. 7-9 are as follows:

- Read mammogram image file
- Convert to grayscale image
- Use the median filter to remove noise from image
- Remove label and background
- Remove pectoral muscle
- Enhancement using density slicing
- Apply k-means algorithm
- Detect the boundary of injury area

CONCLUSION

The suggested method shows that it is easy to regulate the number of regions, the number of pixels in a split area and the edge map and we can extract the divided areas with features. On the other hand, the boundaries of fragmented areas are accurate. This is very useful, for example, to highlight small tumors very close to dense glandular tissue in the medical image and calculate the number of pixels (area measure the effect of the drug before and after taking the medicine. In addition, it is useful to extract an area of the image separately to diagnose and analyze images. On the other hand, if we first use assembly techniques such as k-means to include gray images, then using our method, we can provide very accurate edge maps, area extraction and image segmentation results. Finally, this methodology was tested on many medical images. The results of experiments on the computer show that this method can quickly, extract areas of interest in gray biomedical images and show that the proposed method is effective and possible.

REFERENCES

Ahmad, A.M., G.M. Khan, S.A. Mahmud and J.F. Miller, 2012. Breast cancer detection using cartesian Genetic programming evolved artificial neural networks. Proceedings of the 14th Annual International Conference on Genetic and Evolutionary Computation (GECCO'12), July 07-11, 2012, ACM, Philadelphia, Pennsylvania, USA., ISBN:978-1-4503-1177-9, pp: 1031-1038.

Chakraborty, J., A. Midya and R. Rabidas, 2018. Computer-aided detection and diagnosis of mammographic masses using multi-resolution analysis of oriented tissue patterns. Exp. Syst. Appl., 99: 168-179.

Chakraborty, J., S. Mukhopadhyay, V. Singla, N. Khandelwal and R.M. Rangayyan, 2012. Detection of masses in mammograms using region growing controlled by multilevel thresholding. Proceedings of the 2012 25th IEEE International Symposium on Computer-Based Medical Systems (CBMS), June 20-22, 2012, IEEE, Rome, Italy, ISBN:978-1-4673-2049-8, pp: 1-6.

Cordeiro, F.R., W.P. Santos and A.G. Silva-Filho, 2016. A semi-supervised fuzzy GrowCut algorithm to segment and classify regions of interest of mammographic images. Exp. Syst. Appl., 65: 116-126.

Dominguez, A.R. and A.K. Nandi, 2008. Detection of masses in mammograms via. statistically based enhancement, multilevel-thresholding segmentation and region selection. Comput. Med. Imag. Grap., 32: 304-315.

Elmoufidi, A., K. El Fahssi, S. Jai-Andaloussi and A. Sekkaki, 2015. Automatically density based breast segmentation for mammograms by using dynamic k-means algorithm and seed based region growing. Proceedings of the 2015 IEEE International Conference on Instrumentation and Measurement Technology (I2MTC), May 11-14, 2015, IEEE, Pisa, Italy, ISBN:978-1-4799-6114-6, pp: 533-538.

Eltonsy, N.H., G.D. Tourassi and A.S. Elmaghraby, 2007. A concentric morphology model for the detection of masses in mammography. IEEE. Trans. Med. Imag., 26: 880-889.

Gao, X., Y. Wang, X. Li and D. Tao, 2010. On combining morphological component analysis and concentric morphology model for mammographic mass detection. IEEE. Trans. Inf. Technol. Biomed., 14: 266-273.

Gonzalez, R.C. and R.E. Woods, 2002. Digital Image Processing. 2nd Edn., Prentice Hall, Upper Saddle River, New Jersey, USA., ISBN:9780130946508, Pages: 793.

Henriksen, E.L., J.F. Carlsen, I.M. Vejborg, M.B. Nielsen and C.A. Lauridsen, 2016. The efficacy of using Computer-Aided Detection (CAD) for detection of breast cancer in mammography screening: A systematic review. Acta Radiol., 60: 13-18.

Islam, M.M., H. Iqbal, M.R. Haque and M.K. Hasan, 2017. Prediction of breast cancer using support vector machine and k-Nearest neighbors. Proceedings of the 2017 IEEE International Conference on Region 10 Humanitarian Technology (R10-HTC), December 21-23, 2017, IEEE, Dhaka, Bangladesh, ISBN:978-1-5386-2176-9, pp: 226-229.

Karssemeijer, N. and G.M. Te Brake, 1996. Detection of stellate distortions in mammograms. IEEE. Trans. Med. Imag., 15: 611-619.

Kooi, T., G. Litjens, B. Van Ginneken, A. Gubern-Merida and C.I. Sanchez *et al*., 2017. Large scale deep learning for computer aided detection of mammographic lesions. Med. Image Anal., 35: 303-312.

Miranda, E., M. Aryuni and E. Irwansyah, 2016. A survey of medical image classification techniques. Proceedings of the 2016 International Conference on Information Management and Technology (ICIMTech), November 16-18, 2016, IEEE, Bandung, Indonesia, ISBN:978-1-5090-3353-9, pp: 56-61.

Mudigonda, N.R., R.M. Rangayyan and J.E.L. Desautels, 2001. Detection of breast masses in mammograms by density slicing and texture flow-field analysis. IEEE Trans. Med. Imaging, 20: 1215-1227.

Thakran, S., S. Chatterjee, M. Singhal, R.K. Gupta and A. Singh, 2018. Automatic outer and inner breast tissue segmentation using multi-parametric MRI images of breast tumor patients. PloS One, 13: 1-21.

Varela, C., P.G. Tahoces, A.J. Mendez, M. Souto and J.J. Vidal, 2007. Computerized detection of breast masses in digitized mammograms. Comput. Bio. Med., 37: 214-226.