



A quality enhanced preprocessing method for mammogram ROI extraction

T. R. Thamizhvani^{1*}, Bincy Babu¹, A. Josephin Arockia Dhivya¹, R. J. Hemalatha¹, Josline Elsa Joseph¹, A. Keerthana¹

¹ Department of Biomedical Engineering, Vels Institute of Science, Technology and Advanced Studies, Pallavaram, Chennai-600117, India

*Corresponding author E-mail: thamizhvani.se@velsuniv.ac.in

Abstract

Early detection of breast cancer is necessary because it is considered as one of the most common reason of cancer death among women. Nowadays, the basic screening test for detection of breast cancer is Mammography which consists of various artifacts. These artifacts leads to wrong results in detection of breast cancer. Therefore, Computer Aided Diagnosis (CAD) system mainly focus in removal of artifacts and mammogram quality enhancement. By this procedure, exact Region of Interest (ROI) can be obtained. This is a challenging procedure because detection of pectoral muscle and cancer region is difficult. Here a comparative study of different preprocessing and enhancement techniques are done by testing proposed system on mammogram mini-MIAS database. Result obtained shows that suggested system is efficient for CAD system.

Keywords: Mammogram; Computer Aided Diagnosis; Region of Interest.

1. Introduction

Breast cancer is considered as second major reason of cancer death, after lung cancer. It is developed from breast tissue. The various factors connected with breast cancer are age variation, obesity, history of breast feeding, menses asymmetry in breasts and menses [1]. If physician is suspecting something in patient, he will suggest for mammogram test. Mammography screening is performed for early detection of cancer and proper diagnosis can reduce cancer mortality rate. Breast calcifications are calcium depositions in breast tissues. Macro calcifications are non cancerous where micro calcifications are indication of breast cancer. Sometimes sign of mass or malignancy can be felt by the patient. From the guidelines of American Medical Association (AMA) and American College of Radiology (ACR), recommend mammography screening yearly for women at age 40. The studies verified that early detection can improve probability of treatment and thereby increase survival rate [2]. Mammography can provide detailed information of calcification area such as structure, pathology and shape [3]. The nature of normal and abnormal masses is almost similar, so the detection procedure is complex. Micro calcification regions are abnormal, it can be divided into two categories: malignant and benign [4]. Shape and margins are the main characteristics to distinguish between malignant and benign masses. Low contrast mammogram analysis is complicated due to the difficulty in identifying malignant and benign masses. Therefore, noise removal and contrast enhancement are the fundamental parts of mammogram analysis [5]. During acquisition time, if less amount of photons are falling on breast tissue, then the chance for low contrast mammogram is high. These type of mammograms probably have low Signal to Noise Ratio (SNR) [6].

Different mammogram pre-processing methods have been applied in breast cancer detection [7]. The existing mammogram pre-processing include Median, Frost, and Mean, Wiener filtering, Contrast Limited Adaptive Histogram Equalization (CLAHE) and pectoral muscle removal [8]. In Bhateja V et al [9] works, median filter is used to remove salt and pepper noise without distorting significant information. U Bick et al [10] discussed CLAHE technique for contrast enhancement of mammogram. Naveed et al [11] proposed non-local mean and frost filters for noise removal where mean value of both filters replaces the noisy pixel value. Qian et al [12] introduced Central Weighted Median Filter (CWMF) where more mammogram details were preserved. Compared with other filters CWMF suppress less amount of noise. Li et al [13] propose fractal modeling, which explores mammogram's similar properties. A Computer Aided Detection (CAD) system using genetic algorithm and median filtering is introduced by Thangavel et al [14]. Based on Kovalerchuk et al [15] works, Fuzzy logic concept has effective role in early detection of breast cancer. In Kobatake et al [16] method result shows more true positives and tumor regions are properly isolated. Ferrari et al [17] introduced region based enhancement, in which each pixel in mammogram is considered as seed in order to grow particular region. Based on the pixel value of seed point, transformations are applied and thereby contrast enhancement can be done. Bhangale et al [18] proposed unsharp masking technique to amplify high frequency while eliminates low frequency datas. Rogova et al [19] apply quantum noise theory in mammogram analysis. This method assumes square root model that represents exact approximation of noise in mammogram. In Chang et al [20] proposed system find application in dense and complex mammograms. Here multi scale features and wavelet transform are used for preprocessing.

Researchers proposed and applied several computer aided methods for mammogram processing using different filters. In this proposed

system mini- Mammographic Image Analysis Society (MIAS) database is used. In 1994, J Suckling et al [21] structured this database for breast cancer analysis study purposes. MIAS mammogram films are taken from UK National Breast Screening Program (NBSP) conducted in UK and mammograms are digitized to 50 micron pixel with 8 bit word representation for each pixel. The database include 120 malignant and 202 normal mammograms, all having size 1024×1024 . If abnormalities are present, radiologist markings are there for reference. The important parts generally present in mammogram are tumor region, rectangular label, background, breast tissues, pectoral muscle, low intensity and large intensity region. Pectoral muscles are muscles that lie under breast and it connect shoulder and upper arm bones with chest. Pectoral muscles and rectangular label and background details are irrelevant in breast cancer detection so it coming under noise. Removals of these sections have an important role in mammogram preprocessing. Existing methods have several disadvantages such as some methods easily remove noise without considering significant data, difficulty in handling low contrast images and unnecessary enhancement background structures. The proposed system can eliminate the existing problems and thereby providing exact ROI for micro calcification detection.

2. Methodology

The proposed CAD system contains three sections. Section 1 is elimination of artifacts present in background such as label, low intensity region and background. Section 2 involve pectoral muscle removal and finally Section 3 include filtering and contrast enhancement by wiener filter and CLAHE respectively. The proposed concept is shown in Fig.1.

2.1. Mammogram label and background

Removal

A mammogram lab label is placed on mammogram when patient receives the breast examination result. Threshold method is used for label segmentation. Here binary images can be created from gray-scale images. This one of the earlier and easy method used for segmentation. Different type of threshold methods are histogram shape based, clustering, entropy based, object attribute method and spatial methods. Histogram shape methods analyze valleys, peaks and curvatures. Clustering methods samples clusters into two sections, foreground and background. Entropy of background and foreground are used in Entropy based methods. Object attribute measures similarity of binary image and original image. Mammogram pixel correlations are used in spatial methods. The important parameters considered in threshold are image intensity, I_{ij} and fixed threshold constant, T . The threshold concept is expressed in Eq.1.

$$I_{ij} < T \quad \{ \text{black pixels} \} \quad I_{ij} > T \quad \{ \text{white pixels} \} \quad (1)$$

If mammogram intensity is less than the prescribed cutoff limit T (constant), then threshold method restore each mammogram pixel value with a black pixel. If mammogram intensity is greater than T , it is replaced with white pixel.

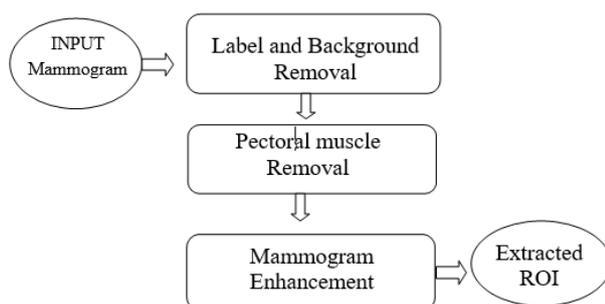


Fig. 1: Block Diagram of Mammogram Preprocessing System.

Binary technique removed background artefact and largest blob (breast part) is extracted by the organization of all connected blobs in mammogram. The original breast sketch can be obtained by the multiplication of input breast image with extracted blob. In mammogram image multiplication of two input images, multiplication of each consequent pixels in both input image takes place. Result image contain product of above multiplication in correspondent pixels. Scaling is image-processing tool where image multiplication is done by a constant. The following equation Eq.2. shows multiplication of two images where $P_1(i,j)$ and $P_2(i,j)$ are input image pixel values and $Q(i,j)$ is resultant multiplied product.

$$Q(i,j) = P_1(i,j) * P_2(i,j) \quad (2)$$

2.2. Removal of pectoral muscle

In mammography, three dimensional breast part is projected into a two dimensional mammogram. The pectoral region which positioned on anterior part of chest, include muscles such as subclavius, serratus, pectorails minor and pectoralis major. The breast part which cover chest muscle called pectoralis major. Pectoralis major is attached to ribs cartilage, breastbone and collarbone. But pectoralis minor lies under pectoralis major, attached to ribs and triangular shaped. Any of these chest muscles commonly known as pectoral muscle. Mammogram is breast X Ray, intensity of pixels in mammogram is the result from superposition of different breast structures through which the X Rays exposed. One of the main problem faced in mammogram analysis is superposition of pectoral muscle.

Pectoral muscles are generally seen as rectangular areas at borders of mammogram. Pectoral muscle can be eliminated by region growing technique which is simple segmentation method based on pixel values. Selection of seed points is key factor in this extraction method. Region growing is done by comparing seed point's neighboring pixels and take decision whether it could be added or not. The selection of seed points can be pixels from a grid or from a particular gray scale. It can be decided manually or automatically. Depending on region membership factor, regions are grown from selected seed points. This membership factor may be color, intensity or texture. We can use four connected or eight connected neighborhood to develop region growing from seed points. The process keep continue watching adjacent pixels. The pixels are classified into seed points if these they have the equal intensity with respect to seed points. This method is an iterative phase where main aim is to classify the likeness of mammogram into regions.

2.3. Mammogram enhancement

Image enhancement is important in image processing to improve perception of image data. It provides more accurate and better image results for other image processing applications. Quality is key factor for human analysis and interpretation; these image enhancement techniques can produce more quality enhanced images depending on applications. In this proposed system mammogram enhancement include noise removal and contrast enhancement.

2.3.1. Filtering

Filtering techniques are applied in image processing in order to modify image and to highlight remarkable features. These techniques include edge enhancement, sharpening and smoothing. High frequencies are suppressed in smoothing and low frequencies are suppressed in edge detection. One of the important noise filters are wiener filters which is based on Fourier iteration and have the advantage that it takes less computational time. The wiener filtering gives an optimal balance between noise smoothing and inverse filtering. It is capable of both blur inversion and additive noise removal. Thereby it reduces overall mean square error. Wiener filtering can be expressed as linear assessment of input image. Fourier domain of wiener is expressed asin Eq.3.

$$F(f_1, f_2) = \frac{G^*(f_1, f_2)S_{xx}(f_1, f_2)}{|G(f_1, f_2)|^2 P_{xx}(f_1, f_2) + P_{nn}(f_1, f_2)} \quad (3)$$

Where $P_{xx}(f_1, f_2)$ power spectra of original image are, $P_{nn}(f_1, f_2)$ are power spectra of additive noise and $G(f_1, f_2)$ is blurring filter. From this equation, it is clearly visible that wiener filter is not only combination of low pass filter but also high pass filtering.

2.3.2. Contrast enhancement

In image processing contrast is an important factor in determining quality. Contrast help to distinguish adjacent objects. In other words, it can be computed by difference in brightness and color of nearby objects. Human eyes are more sensitive to contrast than compared to luminance. Histogram equalization, Adaptive Histogram Equalization (AHE) are commonly used in mammogram processing. The chances for noise enhancement are more in case of AHE. Thereby AHE leads the appearance of artifacts on enhanced regions. For reducing these noises and artifacts, Contrast Limited Adaptive Histogram Equalization (CLAHE) is introduced which is a modification of AHE. Cumulative Distributive Function (CDF) is calculated for intensity levels. The contrast enhancement amount for particular intensity is directly proportional to slope of calculated CDF. For a particular bin slope of CDF is computed by taking the histogram height for that bin. In CLAHE, contrast can be limited by controlling the slope of CDF which can limit by adjusting histogram height. The difference between CLAHE and AHE is that CLAHE contain extra contrast limiting operation by clipping the histogram. The algorithm for CLAHE include following steps. First step is calculation of grid size, 32 pixels square is the minimum size. Default size is choosing as grid size if window size is not indicated. Next step is to identify the grid points and calculating histogram for each grid, the process start from top left corner. Based up on the specified clipping level, histogram is clipped and new CDF is calculated from that histogram. Then all these steps are repeating for all pixels in input image. Final step is to interpolate the obtained values to develop the mapping.

Histogram can be expressed as function in Eq.4. Grey values in image are $m_0, m_1, m_2 \dots m_{L-1}$, P is total number of pixels and P_n is number of pixels having m_k grey value.

$$H(m_k) = \frac{P_n}{P} \quad (4)$$

3. Results and discussion

The proposed system is tested on MIAS database [22]. Different mammogram images such as Glandular, Dense and Fatty types are taken for analysis. Various filters and contrast enhancement techniques are applied to compare the performance of proposed work..

3.1. Pre-processing results

Testing and analysis of proposed system done in MIAS database. The mammogram contain non-significant components such as label and pectoral muscle. Fig1. Shows sampled mammogram and steps involved in removal of label. Binary mammogram is obtained by threshold technique. Two blobs are visible in binary image, the large one is breast part and small one is label. The connected components are systematized and largest blob is extracted. By multiplication process label removed mammogram can be obtained. Fig.2. shows extracted pectoral muscle, pectoral muscle removed mammogram, finally filtered and contrast enhanced output Reduction of pectoral muscle is done by region growing method. Image quality is enhanced by wiener filtering and CLAHE.

3.3.1. Performance evaluation

Performance of system is compared with different filters and enhancement techniques. Performance evaluation metrics are used to analyze the working.

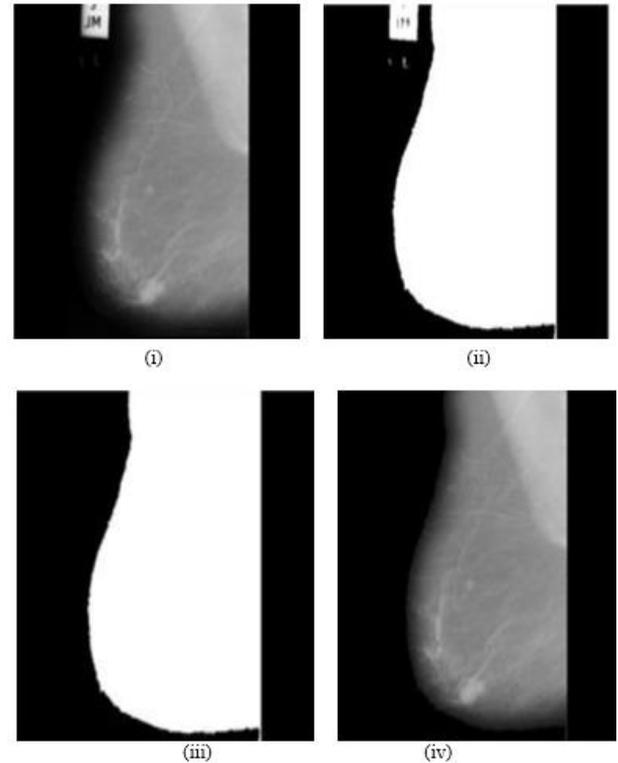


Fig. 2: (I) Sample Mammogram Image (II) Binary Mammogram (III) Extracted Breast Part (IV) Label Removed Mammogram.

3.1.1. Filtering

The performance of proposed system can be analyzed by two metrics are The Peak Signal to Noise Ratio(PSNR) and Root Mean Square Error(RMSE). The PSNR is a performance evaluation factor used to measure quality between an original image and processed output. Higher PSNR value indicates better quality. For MSE analysis both original and processed images are considering and calculating the cumulative squared error. Lower value of MSE denotes less error. Calculation of Mean Square is primary step in computation of PSNR. Eq.5 shows expression for MSE where m and n represents number of rows and columns in original image. M_1 is original image and M_2 is result image of proposed system. Eq.6 shows PSNR expression in which maximum fluctuation in original image is R

$$MSE = \frac{\sum_{P,Q} [M_1(m,n) - M_2(m,n)]^2}{P*Q} \quad (5)$$

$$PSNR = 10 \log_{10} \left(\frac{R^2}{MSE} \right) \quad (6)$$

Median and Standard Deviation (SD) filters are compared with wiener filter by using metrics PSNR as shown in Fig.3. The median filter is a nonlinear noise removing filter. It is commonly used because of its capability to preserve the

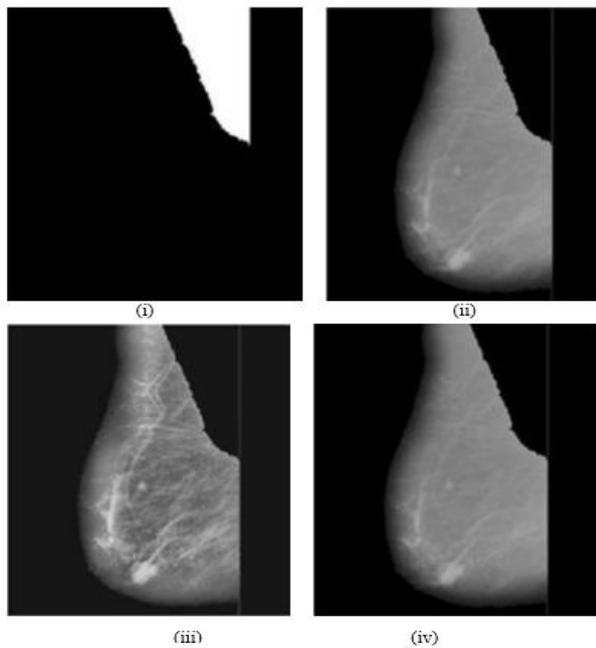


Fig. 3: (I) Extracted Pectoral Muscle (II) Pectoral Muscle Removed Mammogram (III) Filtered Output (IV) Contrast Enhanced Mammogram.

It is commonly used because of its capability to preserve the edges. Median filter replaces each pixel with average value of nearby pixels. The SD filter calculates local standard deviation of input image. Response of five MIAS mammograms are taken for analysis as shown in Fig.3. From this study it can be observed that wiener filter have high PSNR value than median and SD filter. Higher value of PSNR denotes better performance. Thereby the result state that SD filter shows very less performance for MIAS mammogram images and wiener is best among other filtering methods.

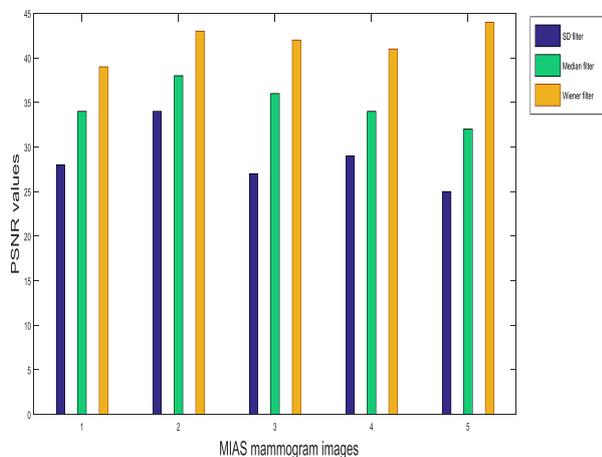


Fig. 4: PSNR Value Comparison for Different Filters.

3.1.2. Contrast enhancement comparison

The final step in preprocessing is contrast enhancement. Histogram Equalization (HE), imadjust and CLAHE are compared for evaluation. Fig.4. shows histogram for these contrast enhancement techniques. The mammogram intensities are adjusted in Histogram equalization. Adaptive Histogram Equalization(AHE) differs from HE that AHE computes many histograms to improve the contrast of filtered mammogram. It is helpful in dealing with local contrast and enhancement of edges in each tile of input image. Over amplification is one of the main problem faced by AHE. A modified version of AHE is CLAHE that prevent over amplification by limiting the histogram. From the comparison of all these methods, it can be observed that CLAHE technique is efficient in handling MIAS mammogram images.

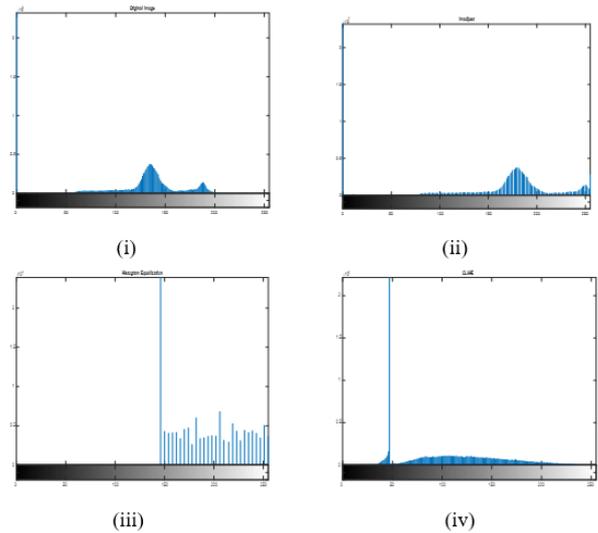


Fig. 5: Histogram (I) Original Mammogram (II) Imadjust (III) Histogram Equalization (IV) CLAHE.

3.1.3. Breast tissue type comparison

Breast tissues generally comprises of milk ducts, milk glands and breast tissues. Breast tissues have different nature, they may be dense tissue or non dense tissue which is called fatty tissue. According to doctor’s opinion, if person having dense breast tissue it is difficult to do screening process and also it shows cancer risk. Dense tissue breast have more appearance of tissues on mammogram. Fatty tissues seem as transparent and dark where dense tissues are white areas on mammogram. The lobules present in breast which is responsible in milk production are called glandular tissues. Different wiener filter masks are applied to mammogram. Fig. shows PSNR values for different type of mammograms. Dense tissues have high PSNR among other two types. 1x1 to 8x8 masks are chosen for comparison. The proposed system shows better performance in detection of micro calcification

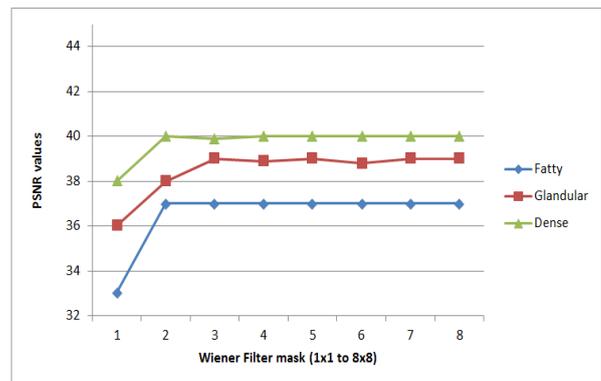


Fig. 6: PSNR Value Comparisons for Different Wiener Filter Masks.

Root mean square error value is calculated to analysis the performance of this work. Different breast tissue types are compared by varying wiener filter masks and RMSE is calculated as shown in Fig.6. Fatty breast tissues shows less RMSE value and dense tissue have maximum. 1x1 wiener filter mask have entirely different response compared to all other masks.

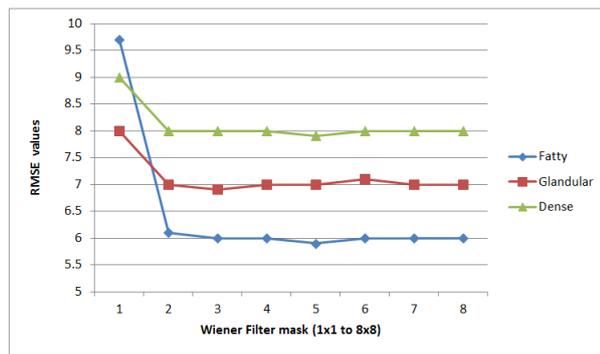


Fig. 6: RMSE Value Comparisons for Different Wiener Filter Masks.

4. Conclusion

Breast cancer is prominent reason of cancer death in women. The appropriate screening will help to identify the presence of abnormality. Mammogram comprises different non-significant components which results in false detection and create difficulty in decision-making. The proposed system is applied in MIA's database. The suggested approach removes background artefacts, labels and pectoral muscle. In addition to that, it enhances the quality of mammogram by filtering and contrast enhancement. From the analysis and comparison, wiener filter mask (3, 3) and CLAHE contrast limit 0.2 are influencing features in mammogram enhancement. Region growing technique eliminates pectoral region. The extracted ROI is effective and appropriate for any CAD system, which is using for early breast cancer detection

5. Future work

Future work of this paper is to segment the calcification cluster, to extract the texture features and classify micro calcification into benign or malignant.

References

- [1] R Siegel, Naishadham D, Jemal A. Cancer Statistics, 2013. CA: Cancer J Clinicians 2013; 63: 11-30.
- [2] JS Mandelblatt, Cronin KA, Bailey S, Berry DA, de Koning HJ, Draisma G, Huang H, Lee SJ, Munsell M, Plevritis SK. Effects of mammography screening under different screening schedules, model estimates of potential benefits and harms. *Annal Internal Medicine* 2009; 151:738-747.
- [3] Sundaram KM, Sasikala D, Rani PA. A Study on Preprocessing a Mammogram Image Using Adaptive Median Filter. *Int J Innovative Res Sci Eng Technol* 2014;3: 10333-10337.
- [4] Acharya RU, Ng EYK, Chang YH, Yang J, Kaw GJL. Computer-Based Identification of Breast Cancer Using Digitized Mammograms. *J Medical Systems* 2008; 32:499-507. Bland, P.H. and C.R. Meyer, Robust three-dimensional object definition in CT and MRI. *Med. Phys.*, 1996, 23(1), p. 99-107
- [5] Mustra M, Grgic M. Robust automatic breast and pectoral muscle segmentation from scanned mammograms. *Signal Processing* 2013; 93: 2817-2827.
- [6] Mencattini A, Salmeri M. Noise estimation in mammographic images for adaptive denoising. *EFOMP European Conference on Medical Physics* 2007.
- [7] Liu CC, Tsai CY, Liu J, Yu CY, Yu SS. A pectoral muscle segmentation algorithm for digital mammograms using Otsu thresholding and multiple regression analysis. *Computer and Mathematics with Applications* 2012;64(5), pp.100-1107.
- [8] Ferrari RJ, Rangayyan RM, Desautels J E, Borges RA, Frere AF. Automatic identification of the pectoral muscle in mammograms, *IEEE Transactions on Medical Imaging* 2004;23(2), pp.232-245.
- [9] Bhateja V, Verma A, Rastogi K, Malhotra C, Satapathy SC. Performance Improvement of Decision Median Filter for Suppression of Salt and Pepper Noise. *Advances in Signal Processing and Intelligent Recognition Systems* 2014, Springer: 287-297.
- [10] U. Bick, M.L. Giger, R.A. Schmidt, R.M. Nishikawa, D.E. Wolverton, and K. Doi, "Automated Segmentation of Digitized Mammograms", *Academic Radiology*, vol. 2, no. 2, pp. 1-9, 1995.
- [11] Naveed N, Hussain A, Arfan JM, Choi TS. Quantum and impulse noise filtering from breast mammogram images. *Comput Methods Programs Biomed* 2012; 108: 1062-1069.
- [12] Qian, W., Clarke, L.P., Kallergi, M. and Clark, R.A. "Tree-Structured Nonlinear Filters in Digital Mammography," *IEEE Transactions on Medical Imaging*, Vol. 12, No. 1, pp. 25-36, 1994.
- [13] Li, H., Liu, K.J.R. and Lo, S.C.B. "Fractal modeling and segmentation for the enhancement of microcalcifications in digital mammograms," *IEEE Transactions on Medical Imaging*, Vol.16, No. 6, pp. 785-798, 1997.
- [14] K Thangavel, M Karnan, R Sivakumar, AK Mohideen, "Automatic detection of microcalcification in mammograms-a review", *International Journal on Graphics, Vision and Image Processing*, 5 (5), 31-61 2005
- [15] Kovalerchuk, B., Traintaphyllou E.J.F. Ruiz and J. Clayton, J. "Fuzzy Logic in Computer-Aided Breast Cancer Diagnosis: Analysis of Lobulation," *Artificial Intelligence in Medicine*, Vol. 11, pp. 75-85, 1997.
- [16] Kobatake, H., Murakami, M., Takeo, H. and Nawano, S. "Computerized Detection of Malignant Tumors on Digital Mammograms," *IEEE Transactions on Information Technology in Biomedicine*, Vol. 18, No. 5, pp. 369-378, 1999.
- [17] Ferrari, R.J., De Carvalho, F., Marques, P.M.A. and Frere, A.F. "Computerized classification of breast lesions: shape and texture analysis using an artificial neural network," 7th international conference on Image Processing and its applications, pp. 517-521, 1999.
- [18] Bhangale, T., Desai, U.B. and Sharma, U. "An unsupervised scheme for detection of microcalcifications on mammograms," *IEEE International Conference on Image Processing*, pp. 184-187, 2000.
- [19] Rogova, G.L., Stomper, P.C. and Ke, C. "Microcalcification texture analysis in a hybrid system for computer aided mammography," *SPIE*, Vol. 3661, pp. 1426-1433, 1999.
- [20] Chang, R.F., Wu, W.J., Tseng, C.C., Chen, D.R. and Moon, W.K. "3-D Snake for US in Margin Evaluation for Malignant Breast Tumor Excision Using Mamotome," *IEEE Transactions on Information Technology in Biomedicine*, Vol. 7, no. 3, pp. 197-201, 2003.
- [21] J. Suckling, J. Parker, D.R. Dance, S. Astley, I. Hutt, C.R.M. Boggis, I. Ricketts, E. Stamatakis, N. Cernaez, S.L. Kok, P. Taylor, D. Betal, J. Avage, The mammographic image analysis society digital mammogram database, in: *Proceedings of the 2nd International Workshop on Digital Mammography*, York, England, 10-12 July 1994, Elsevier Science, Amsterdam, 1994, pp. 375-378.
- [22] <http://peipa.essex.ac.uk/info/mias.html>.