Microcalcification and Macrocalcification Detection in Mammograms Based on GLCM and ODCM Texture Features Using SVM Classifier

Lincy Golda Careline S[1], Dr. J. S. Leena Jasmine[2]

[1]M.E, Applied Electronics, Velammal Engineering College, Chennai, Tamil Nadu, India
[2]Assistant Professor, Velammal Engineering College, Chennai, Tamil Nadu, India

Abstract— Breast cancer is a common cancer in women and the second leading cause of cancer deaths worldwide. Photographing the changes in internal breast structure due to formation of masses and microcalcification for detection of Breast Cancer is known as Mammogram, which are low dose x-ray images. These images play a very significant role in early detection of breast cancer. Usually in pattern recognition texture analysis is used for classification based on content of image or in image segmentation based on variation of intensities of gray scale levels or colours. Similarly texture analysis can also be used to identify masses and microcalcification in mammograms. However Grey Level Co-occurrence Matrices (GLCM) technique introduced by Haralick was initially used in study of remote sensing images. Radiologists f i n d i t d i f f i c u l t to identify the mass in a mammogram, since the masses are surrounded by pectoral muscle and blood vessels. In breast cancer screening, radiologists usually miss approximately 10% - 30% of tumors because of the ambiguous margins of tumors resulting from long-time diagnosis. Computer-aided detection system is developed to aid radiologists in detecting ma mammographic masses which indicate the presence of breast cancer. In this paper the input image is pre-processed initially that includes noise removal, pectoral muscle removal, thresholding, contrast enhancement and suspicious mass is detected and the features are extracted based on the mass detected. A feature extraction method based on grey level co- occurrence matrix and optical density features called GLCM -OD features is used to describe local texture characteristics and the discrete photometric distribution of each ROI. Finally, a support vector machine is used to classify abnormal regions by selecting the individual performance of each f e a t u r e . The results prove that the proposed system achieves an e x c e l l e n t detection performance using SVM classifier.

Keywords—Computer-aided detection system, grey level co-occurance matric, optical density co-occurance matrix, feature extraction.

I. INTRODUCTION

Breast cancer is a common form of cancer disease among women with nearly 1.7 million new cases diagnosed in 2015 and the second cause of cancer deaths worldwide [1]. Early detection of breast cancer is a key factor for successful cancer treatment. This represents about 1 2 % o f all new cancer cases and 25% of all cancers in women [1]. For every 2 women newly diagnosed with breast cancer in India one women is dying of breast cancer [1], [2].In women breast cancer is most common and deadliest forms of cancer found worldwide. Ten years back in India cervical cancer was having highest mortality rate whereas breast cancer was second highest but within ten years the statistics have changed, now breast cancer tops the list of cancer related mortality rate. The Indian breast cancer scenario is more worrisome and disturbing as compared to western countries or even with neighbours like china. If we look at the statistics of all Indian cities, breast cancer alone accounts to 25% to 31% of all cancers in women [1]. It has been observed that there is a significant age shift, and the average age of developing breast cancer in India has shifted from 50 - 70 years to 30 - 50 years; and sadly cancers in the young tend to be more aggressive. For just the year 2012, GLOBOCAN (WHO), had estimated 70218 women deaths in India due to breast cancer, more than any other country in the world, china was second with 47984 deaths and US third with 43909 deaths [2]. The vast difference in numbers of mortalities is alarming, and not just worth noticing but acting on it. It is projected For the years 2015, there will be an estimated 1,55,000 new cases of breast cancer and about 76000 women in India are expected to die of the disease. Breast cancer is related to hormones and the factors that modify the risk of this cancer

when diagnosed premenopausal and when diagnosed postmenopausal are not the same [1]. The countries with the top 20 highest incidence of breast cancer in 2015 are Belgium had the highest rate of breast cancer, followed by Denmark and France. More cases of breast cancer are diagnosed in less developed countries .The high incidence of breast cancer was in Northern America and Oceania; and the low incidence of breast cancer in Asia and Africa [1]. The aim of developing computer aided diagnosis system is to aid radiologists to improve breast cancerscreening and diagnosis [2], [3]. These systems act as a second opinion for radiologist in better reading and understanding of mammography images. Mammography is currently the most effective imaging modality used by radiologists for the screening of breast cancer [3]. In this paper we explore an automated technique for mammogram mass detection. The proposed method removes noises, separates background region from the breast profile region and removes the pectoral muscle for accentuating the breast profile region. A Computer Aided Diagnosis system is implemented under the MATLAB environment for classifying malignant masses in digital mammograms using Support Vector Machines (SVM) [16], [17]. The proposed method successfully achieves an accuracy of 95% which is considered as a good result when compared with similar works in the same research field. Jiři Grim, Petr Somol, Michal Haindl and Jan Daneš [2] proposed a new approach to diagnostic evaluation of screening mammograms based on local statistical texture models. The local evaluation tool has the form of a multivariate probability density of gray levels in a suitably chosen search window. The density function in the form of Gaussian mixture is estimated from data obtained by scanning of the mammogram with the search window Al Mutaz M. Abdalla, Safaai Dress, Nazar Zaki [14] proposed modern textural features analysis of breast tissues on mammograms by Detecting masses in digital mammogram based on second order statistics. The extraction of the textural features of the segmented region of interest (ROI) is done by using gray level co-occurrence matrices(GLCM) which is extracted from four spatial orientation. This is distinct from existing approach that tend to concentrate on the morphology of individual microcalcifications and global(statistical) cluster features. A set of microcalcification graphs are generated to represent the topological structure of microcalcification clusters at different scales [21].

II.   METHODS & METHODOLOGY

A. *Image Preprocessing*

The proposed CAD system for mammographic mass detection comprises four major stages: preprocessing, detection of suspicious mass region, feature extraction, and classification. Fig. 1 shows an overview of the proposed mass detection scheme, and the following subsection present each component in detail.
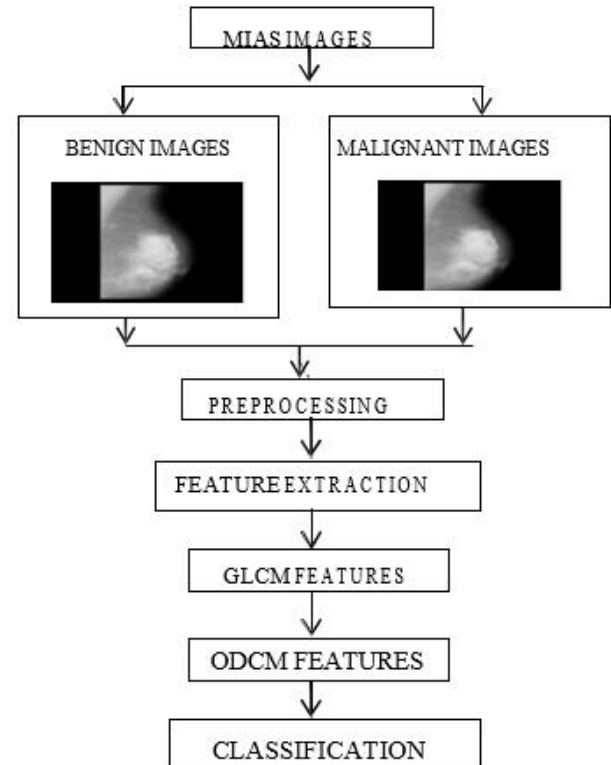

Fig.1: Block diagram of proposed mammographic mass detection scheme

The preprocessing method consists of the following steps for noise reduction, removal pectoral muscle, separation of the image from the background and contrast enhancement.
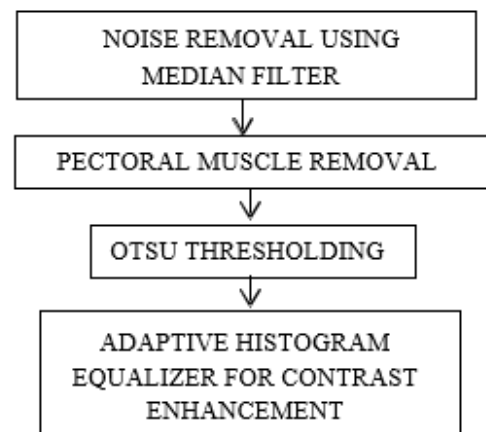

Fig.2: Proposed preprocessing system

1. Noise reduction

The noise reduction in mammogram is done by using median filter to remove the salt and pepper noise present in the input mammogram image.
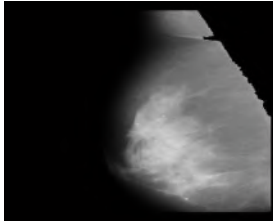

Fig.1.1: Represents the noise removal using median filter

2. Pectoral muscle removal

The pectoral muscle are brightest pixels present in the top corners of the mammogram image. Because the pectoral muscle represents a brighter region, it may affect the detection results [14]. Hence, the maximum connected component finds the position of the pectoral region and adaptive thresholding is done inorder to remove the pectoral muscle from the input mammogram image. Consequently, the breast region is obtained by removing the pectoral muscle from the foreground.
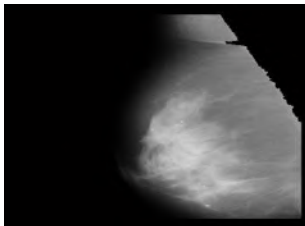

Fig.1.2: Represents the pectoral muscle removal from the noise reduction image

3. Thresholding

The thresholding is done to separate the input image from the background. The Otsu thresholding method is applied to the digital mammogram to find the foreground of concern, which contains a breast region in most mediolateral oblique (MLO) views of mammograms [9]. The otsu thresholding is applied to the mammogram input image to separate the breast region from the background.

Fig.1.3: Represents the thresholding performed using otsu thresholding method.

4. Contrast enhancement

The contrast enhancement is done to increase the brightness of the image. The adaptive histogram equalization technique is applied to increase the contrast of the mammogram image. The preprocessing method reduces noises, removes the pectoral muscle, separates the breast image from the background and increases the contrast of the mammogram input image.
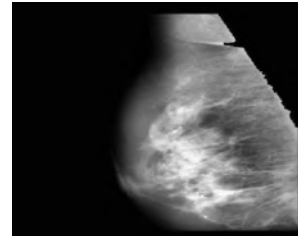

Fig.1.4: Represents the contrast enhancement using adaptive histogram equalization technique.

B. Feature Extraction Module

After ROI detection of suspicious masses [4], [5] some features are extracted to express the characteristics of the suspicious mass region. The intensity distribution of masses is an important characteristic for mass detection.

1. Grey-level co-occurrence matrix

Therefore, some pattern recognition methods use a gray level co-occurrence matrix (GLCM) to extract characteristics [9], [14]. Statistical distributions of observed combinations of intensities at specified positions relative to each other in an image are used to obtain statistical textural features. These features can be classified into first-order, second-order, and higher order according to the number of intensity points (connected pixels) considered in each combination. The Grey level Co-ocurrence Matrix (GLCM) technique is used to compute second-order statistical textural features. A GLCM is a matrix where the number of rows and columns is equal to the number of gray levels, in the image. The matrix element P(i, j / Δx, Δy) is the relative frequency with which two pixels, separated by a pixel distance (Δx, Δy), occur within a given neighborhood, one with intensity i and the other with intensity j. One may also say that the matrix element P(i, j 1 d,8) contains the second order statistical probability values for changes between gray levels i and j at a particular displacement distance d and at a particular angle (8). Figure 1 below illustrates the geometrical relationships of GLCM measurements made for four distances d (d = max {1 Δ x I, 1 Δ y I}) and angles of 8 = 0, rr/4, rr/2 and 3 rr /4 radians under the assumption of angular symmetry. The idea behind GLCM

is to describe textures by a matrix of pair gray level appearing probabilities. The fourteen texture features are defined as follows: Entropy, Energy, Local homogeneous, Contrast, Intensity, Correlation, Inverse difference moment, Sum average, Sum of squares variance, Sum entropy, Difference entropy, Inertia, Cluster Shade, and Cluster Prominence [9]. GLCM texture considers the relation between two pixels at a time, called the reference and the neighbour pixel. The neighbour pixel is chosen to at the right of each reference pixel. This can also be expressed as a (1,0) relation: 1 pixel in the x direction, 0 pixels in the y direction. Each pixel within the window is a reference pixel , starting in the upper left corner it proceeds to the lower right. Pixels along the right edge have no right hand neighbour, so they are not used in count.

2. *Optical density co-occurance matrix*

The background information is considered in the discrete texture feature category because it transforms an intensity into an optical density value. The optical density transformation for each pixel (i,j) of an object region is defined as

$$OD_{ij} = \log (I_{ij} / I_O)$$

where I_{ij} is the intensity value of pixel, and I_o is the average background intensity. The background is an ROI excluding the pixels belonging to the object region. This study proposes two complex feature extraction methods to achieve a complete description of quantitative characteristics. The first feature extraction module adopts GLCM features and optical density features. This type of complex texture feature extraction method extracts the information of local intensity relation and discrete photometric distribution. The proposed scheme computes four co-occurrence matrices with one pixel distance in four directions: left diagonal, right diagonal, vertical, and horizontal [9]. Another complex feature extraction method is also constructed that is similar to the proposed complex module, but translates the gray level co-occurrence matrix into the optical density co-occurrence matrix (ODCM) to characterize the photometric textures [14]. The optical density co-occurance matrix is a co-occurrence matrix of the optical density image. An optical density image can be obtained by converting the intensity of the gray level image into optical density image and linearly mapping all the optical density values one by one to an image that has an 8-bit depth information [9]. The minimum optical density value was mapped to 0, and the maximum optical density value was mapped to 255. After transformed the gray level image into the

optical density image, the differences between gray level values are enlarged, enhancing the simulated mass region. Since the background represents the surrounding normal tissues in an ROI with appropriate thresholding, an optical density image can serve as a graph that represents the degree of malignant tissue based on the intensity (the lighter area represents greater possibility of malignant tissue) [14]. Finally, the two proposed methods combining texture features and optical density features use seventy-six statistics to achieve a complete description of characteristics.

C. *Classification*

Several classification methods are developed in recent years. One classification technique that is widely used for the diagnosis of breast tumors is the Support Vector Machine (SVM) [16]. SVM is one of the shining peaks in many learning algorithms which is inspired by statistical learning theory and has appeared in the machine learning community in the last few decades [16],[17].Consequently, the proposed classification module consists of reduced features which were selected by support vector machine classifier after the performance comparison of classifiers. The single stage SVM classifier identifies the image from the testing dataset to be benign or malignant by comparing the image from the trained datasets. Thus the SVM classifier proves to be significant compared to other classifiers by providing an accuracy of 95% with sensitivity of 0.9 which is considered as better performance than other experimental methods.

Table.1:Performance Analysis of Svm Classifier

SVM CLASSIFIER		
PARAMETERS	MICROCALCIFICATION IMAGES	MASS IMAGES
ACCURACY	95.0	93.5
SPECIFICITY	9	10
SENSITIVITY	10	9

III.   EXPERIMENTAL RESULTS

A. *Datasets*

The mammogram datasets were taken from the Mammogram image analysis society (MIAS)

database for both benign and malignant breast cancer images [20]. A set of benign and malignant images are trained to determine the breast tumors and undergo the following training process such as preprocessing, feature extraction. The testing process consists of a set of benign and malignant breast cancer images. Images are digitized to 50 micron pixel edge representing each pixel with an 8- bit word. The database contains 322 digitized images. The database has been reduced to a 200 micron pixel edge and padded/clipped so that all the images are 1024x1024.MIAS database is used in for this work, this is freely available and has been widely used for mammogram classification. This database is made up of medio-lateral oblique views of both right and left breast of women.

Table.1: Table showing the distribution of Cases in MIAS

Class	Begnin	Malignant	Total
Microcalcification	12	13	25
Circumscribed masses	19	4	23
III-defined masses	7	7	14
Spiculated masses	11	8	19
Architectural distortion	9	10	19
Asymmetry lesion	6	9	15
Normal tissue	-	-	207
Total	64	51	322

We randomly choose 20 samples each from normal, benign and malignant cases for our experimental purpose

B. Performance and Analysis of the Proposed System

(c)   (d)

(e)   (f)

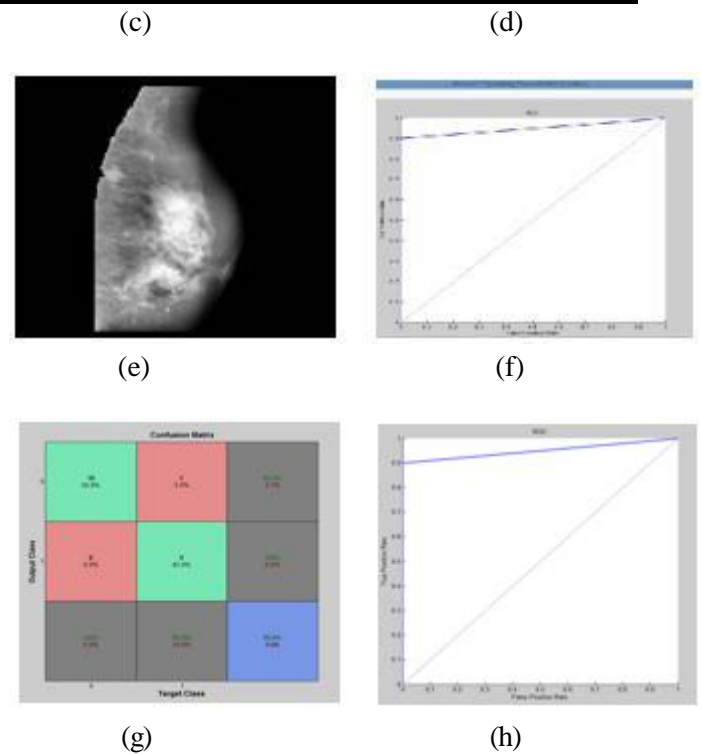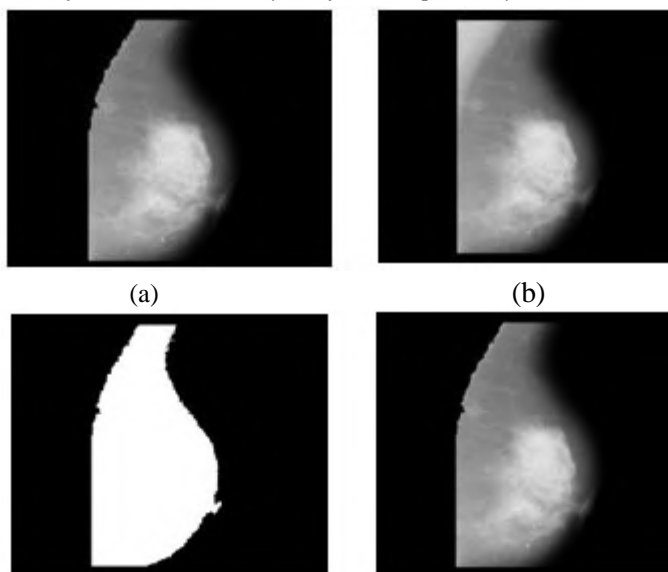(g)   (h)

Fig.4: (a) Original mammogram, (b) pectoral muscle removal image, (c) thresholding image, (d) noise removal image, (e) contrast enhancement image, (f) receiver operating characteristics curve of mass image, (g) confusion matrix, (h) receiver operating characterestics of microcalcification images.

IV.   DISCUSSION AND CONCLUSION

In the proposed method an automatic CAD system for mammographic mass detection that uses complex texture features for classifying the suspected mass region preprocesses the input mammogram image to acquire the breast region and suppress the effects of noise using median filter. Then the two feature extraction methods using grey-level co-occurance matrix and optical density co-occurance matrix which then combines the GLCM features and optical density features to describe both the grey level characteristics of local textures and photometric discrete textures. The optical density image enhances the difference of grey level based on the normal tissue intensity to strengthen the description of the suspicious area shape for feature extraction in CAD system..

The proposed method proves that the classification using support vector machine achieves satisfactory detection with sensitivity of 0.9, accuracy of 95% for both two feature extraction methods. The ODCM-optical density features that can increase the mass

detection rate of the CAD system for the dense breast is proposed in this study to reduce the burden of radiologists and conserve resources. The future work of the proposed method comprises of an automatic CAD system for microcalcification identification in mammogram detection adapting various texture features using suitable classifier to increase the accuracy, sensitivity and reduce false positive rates and improve overall performance of the system.

REFERENCES

[1] American cancer society, "Breast cancer facts &figures 2015-2016",Atlanta: American cancer society, Inc.2015.

[2] J. Grim, P. Somol, M. Haindl, and J. Danesˇ, "Computer-aided evaluation of screening mammograms based on local texture models," Trans. Img. Proc., vol.18, pp. 765–773, Apr. 2009.

[3] Mohamed E. Elmanna, Yasser M. Kadah,"Implementation of Practical Computer Aided Diagnosis System for Classification of Masses in Digital Mammograms", International Conference on Computing, Control, Networking, Electronics and Embedded Systems Engineering, 2015

[4] Jawad Nagi, Sameem Abdul Kareem, Farrukh Nagi, Syed Khaleel Ahmed,"Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms", 2010 IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES 2010), Kuala Lumpur, Malaysia, 30th November - 2nd December 2010.

[5] Indra Kanta Maitra, Sanjay Nag and Prof. Samir K. Bandyopadhyay, "Automated digital mammogram segmentation for detection of abnormal masses using binary homogeneity enhancement algorithm", Indian Journal of Computer Science and Engineering (IJCSE) Vol. 2 No. 3 Jun-Jul 2011.

[6] N. Eltonsy, G. Tourassi, and A. Elmaghraby, "A concentric morphology model for the detection of masses in mammography," Medical Imaging, IEEE Transactions on, vol. 26, pp. 880 –889, June 2007.

[7] Naga R. Mudigonda, Rangaraj M. Rangayyan and J.E Leo Desautels, "Detection of Breast Masses in Mammograms by Density Slicing and Texture Flow- Field Analysis", IEEE Transactions On Medical Imaging, Vol. 20, No. 12, December 2001.

[8] S. Xu and C. Pei, "Hierarchical matching for automatic detection of masses in mammograms,"

in Electrical and Control Engineering (ICECE), 2011 International Conference on, pp. 4523–4526,September2011.

[9] Shen-Chuan Tai, Zih-Siou Chen, and Wei-Ting Tsai, "An Automatic Mass Detection System in Mammograms based on Complex Texture Features",IEEE Journal Of Biomedical And Health Informatics,Vol.18.No .2,March 2014.

[10] M. Sameti, R. Ward, J. Morgan-Parkes, and B. Palcic, "Image feature extraction in the last screening mammograms prior to detection of breast cancer," Selected Topics in Signal Processing, IEEE Journal of, vol. 3, pp. 46 –52,February 2009.

[11] Bhagwati Charan Patel , G. R. Sinha ,"Mammography Feature Analysis and Images",2014 International Conference on Electronic Systems, Signal Processing and Computing Technologies.

[12] Cheng, H.D., Shi, X.J., Min, R., Hu, L.M., Cai, X.P., Du, "H.N: Approaches for Automated Detection and Classification of Masses in Mammograms",Pattern Recognition 39(4), 646–668 (2006).

[13] I. Kitanovski, B. Jankulovski, I. Dimitrovski, and S. Loskovska, "Comparison of feature extraction algorithms for mammography images," in Image and Signal Processing (CISP), 2011 4th International Congress on, vol. 2, pp. 888– 892,2011.

[14] Al Mutaz M. Abdalla, Safaai Dress, Nazar Zaki,"Detection of Masses in Digital Mammogram Using Second Order Statistics and Artificial Neural Network", International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 3, June 2011.

[15] Snehal A. Mane, Dr. K. V. Kulhalli, "Mammogram Image Features Extraction and Classification for Breast Cancer Detection", International Research Journal of Engineering and Technology (IRJET),Volume 2,Issue 7 , October 2015.

[16] Nasseer M. Basheer, Mustafa H. Mohammed, "Classification of Breast Masses in Digital Mammograms Using Support Vector Machines"International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 10, October 2013.

[17] M. Hussain, S. Wajid, A. Elzaart, and M. Berbar, "A comparison of svm kernel functions for breast cancer detection," in Computer Graphics, Imaging and Visualization (CGIV), 2011 Eighth International Conference on,

pp. 145–150, 2011.

[18] Xiaoming Liu, Xin Xu, Jun Liu, Zhilin Feng," A new automatic method for mass detection in mammography with false positives reduction by supported vector machine",2011 4th International Conference on Biomedical Engineering and Informatics.

[19] X. Gao, Y. Wang, X. Li, and D. Tao, "On combining morphological component analysis and concentric morphology model for mammographic mass detection," Information Technology in Biomedicine, IEEE Transactions on, vol. 14, pp. 266 –273, March 2010.

[20] MIAS database, http:// pcipa.csscx.ac.uk/ipa/pix/mias/

[21] Zhili Chen, Harry Strange, Arnau Oliver, Erika R. E. Denton, Caroline Boggis, and Reyer Zwiggelaar,"Topological Modeling And Classification Of Mammographic Microcalcification Clusters", IEEE Transactions On Biomedical Engineering, Vol. 62, No. 4, April 2015.

[22] Akshay S. Bharadwaj and Mehmet Celenk, "Detection Of Microcalcification With Top-Hat Transform And The Gibbs Random Fields", 2015 IEEE.

[23] Luqman Mahmood Mina, Nor Ashidi Mat Isa,"Breast Abnormality Detection In Mammograms Using Artificial Neural Network", 2015 IEEE , International Conference On Computer, Communication And Control Technology (I4ct 2015).

[24] C.Abirami, R.Harikumar, S.R.Sannasi Chakravarthy, "Performance Analysis And Detection Of Micro Calcification In Digital Mammograms Using Wavelet Features", IEEE WISPNET 2016 CONFERENCE.

[25] Daniel Ruiz-Fernandez, Senior Member, Ieee, Juan José Galiana-Merino, And Manuela B. Pacheco Lloret, "Influence Of The Surrounded Tissue In The Detection Of Microcalcifications Using Wavelets", 2015 IEEE.

[26] Juan Wang, Yongyi Yang, and Robert M. Nishikawa, "Quantitative Study Of Image Features Of Clustered Microcalcifications Inforpresentation Mammograms", ICIP 2016.