Breast Cancer Detection in Digital Mammograms

Kanchan Lata Kashyap*, Manish Kumar Bajpai, Pritee Khanna
Computer Science & Engineering, Indian Institute of Information Technology, Design & Manufacturing
Jabalpur, India-482005
*Corresponding Author kanchan.k@iiitdmj.ac.in

Abstract— **This paper discusses an approach for automatic detection of abnormalities in the mammograms. Image processing techniques have been applied to accurately segment the suspicious region-of-interest (ROI) prior to abnormality detection. Unsharp masking has been applied for enhancement of the mammogram. Noise removal has been done by using median filtering. Discrete wavelet transform has been applied on filtered image to get the accurate result prior to segmentation. Suspicious ROI has been segmented using the fuzzy-C-means with thresholding technique. Tamura features, shape based features and moment invariants are extracted from the segmented ROI to detect the abnormalities in the mammograms. Proposed algorithm has been validated on the Mini-MIAS data set.**

Keywords—Unsharp masking, Median filtering, fuzzy-c-means, discrete wavelet transform

I. INTRODUCTION

Mammography has been considered as the most important technique to investigate the breast cancer. It can be used to detect the disease in the early stage when recovery is possible. Aim of Computer Aided Diagnosis system (CAD) is to read the mammograms, locate the suspicious regions of abnormalities and analyze its characteristics. Performance and reliability of CAD depends on accurate segmentation of the lesions and appropriate feature selection. Global segmentation of the lesions is the most challenging task due to the artifacts and healthy tissues present in the mammograms [1]. Various algorithms have been developed in literature for early detection of the breast cancer in mammograms. Kegelmeyer *et al.* presented laws textural feature and binary decision tree classifier to classify between lesions and normal tissues [2]. Campanini *et al.* presented an SVM classifier for mass detection in digital mammograms [3]. Sahiner et al. used four gray-level difference statistics (GLDS) texture features and convolution neural network for mass detection [4]. Bellotti *et al.* proposed an automated CAD system for mass detection based on gray level co-occurrences matrices (GLCM) features and back-propagation neural network classifier [5]. Mudigonda et al. and Dominguez et al. proposed Density slicing technique for mass segmentation [6, 7]. Sharma and Khanna developed a CAD system to detect the breast cancer using Zernike moments [8]. Kashyap et al. proposed an efficient segmentation algorithm for breast cancer detection in mammograms [9]. Present work proposes a shape based approach for breast cancer detection using mammograms. Organization of the paper is given as follows: Flow chart and the proposed algorithm is described in Section 2. Section 3 includes the validation of algorithm and experimental results followed by conclusions in Section 4.

II. PROPOSED ALGORITHM

Flow chart of the proposed algorithm is given in Fig 1 which includes breast region segmentation, enhancement, filtering, segmentation, feature extraction and classification steps. Different types of noises such as tape artifacts, labels or scanning artifacts are present in the mammogram images. Such noises must be suppressed and excluded from further processing. Breast region is segmented after removal of noises which contains the useful information. Original mammogram images are binarized by global thresholding technique. Morphological opening operation is performed to remove small regions [10]. Connected component labeling algorithm is applied to identify the largest connected component as a breast region. Breast region is superimposed on the original image to get the final segmented unlabeled image I_P. Contrast of Mammogram images are enhanced by using unsharp masking technique to get enhanced image I_{PE} [10]. Original unlabeled image is inverted to get inverted images $I_p^{'}$. Inverted image is also enhanced by using unsharp masking technique and it results enhanced inverted image $I_{PE}^{'}$. Enhanced inverted image is subtracted from Enhanced image I_{PE} and which results in subtracted image $I_{I.}$ Subtracted image is filtered by applying median filter to remove the overall noise of image and it results filtered image I_{1F}. Discrete wavelet transform (DWT) is applied to decompose the filtered image into approximated image and detailed image [10]. Wavelet transformation provides both the frequency as well as temporal information. Wavelet transform is a tool which has been used to locate the low frequency area and high frequency area in image I_{1F}. It also highlights the structural, geometrical and directional features of mammographic image. Detailed image is reconstructed and added to filtered image to sharpen the edges of mass. Sharpen image is used further to segment the suspicious mass region. Segmentation is performed to partition the images into homogeneous regions and extract the region-of-interest (ROI). Fuzzy C-means (FCM) clustering with thresholding algorithm is applied on sharpened image [11]. Connected component algorithm is applied to extract the segmented suspicious mass region. A set of tamura features, shape based and moment invariant features are extracted from each segmented suspicious regions and feature vector is created. Support vector machine (SVM) is used to classify the segmented region as an abnormal region or normal region [12].

978-1-4799-8633-0/15/$31.00 ©2015 IEEE

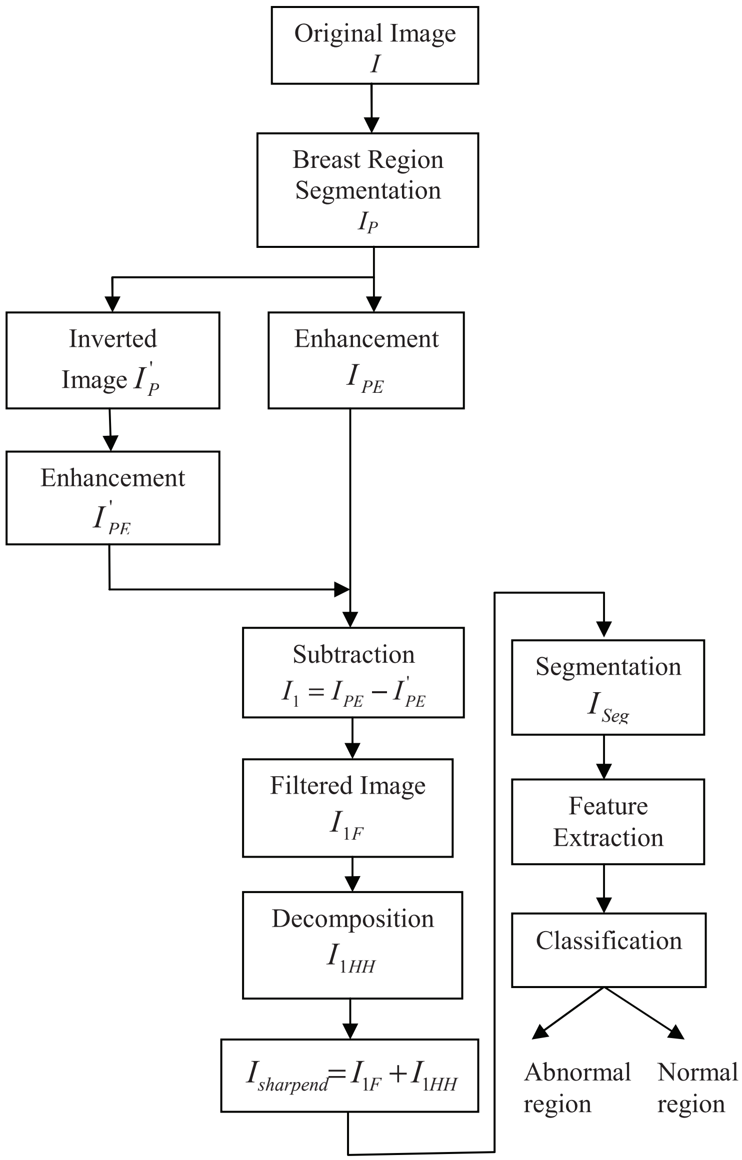


Fig. 1. Flow chart of proposed algorithm

Algorithm of the proposed methodology is given as follows:

Algorithm1 (I, Class)

Input : I

- ***I : input Image***
- ***I_P : preprocessed image***
- ***I_{PE} : enhanced preprocessed image***
- ***I'_P : inverted preprocessed image***
- ***I'_{PE} : enhanced Inverted preprocessed image***
- ***I_1 : subtracted Image***
- ***I_{1F} : filtered Image***
- ***I_{1FA} : approximated decomposed filtered image***
- ***I_{1FD} : detailed decomposed filtered image***
- ***I_{seg} : segmented image***
- ***FV : feature vector***
- *Preprocessing ( )* : preprocessing of image
- *Unsharp ( )* : standard unsharp masking mthod
- *Invert ( )*: standard inversion method of image
- *Median ( )* : standard median filter
- *Wavelet ( )*: standard wavelet method for image decomposition
- *FCMthreshold ( )* : fuzzy c means clustering with thresholding
- *Feature_vector ( )*: function to extract features
- *SVM ( )*: standard function for classification

Output : Class

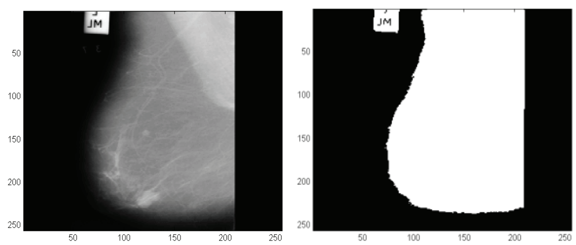
begin

I_P = *Preprocessing (I)*
I_{PE} =*Unsharp*(I_P)
I'_P =*Invert*(I_P)
I'_{PE}=*Unsharp*(I'_P)
$I_1 = I_{PE} - I'_{PE}$
I_{1F}= *Median*(I_1)
$[I_{1FA}, I_{1FD}]$ = *Wavelet*(I_{1F})
$I_{Sharpened} = I_{1F} + I_{1FD}$
I_{Seg} = *FCMthreshold*($I_{Sharpened}$)
FV = *Feature_vector*(I_{seg})
Class= SVM(FV)

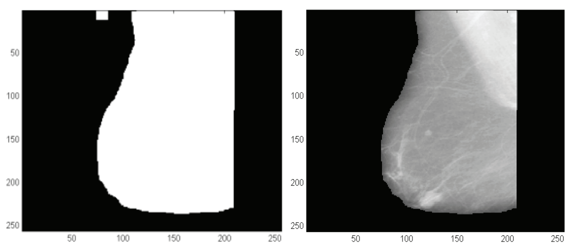
end

III. EXPERIMENTAL RESULTS AND DISCUSSION

Proposed algorithm is validated on publicly available Mammographic Image Analysis Society (MIAS) data set containing 322 mammogram images available in this data set. All sample images are scaled down to 256×256 pixels using nearest neighbor interpolation method. Removal of tape artifacts and labels are done in first stage of proposed algorithm. Fig. 2 shows the intermediate results of breast region segmentation. Fig 2(a), (b), (c) and (d) show the original mammogram, binarized image after performing thresholding operation, image after performing morphological opening operation and unlabeled image, respectively. It is clear from Fig. 2 that background information is successfully removed from the mammographic image.



(a) Original image (b) Binarized image



(c) Morphological opened image (d) Unlabeled image

Fig. 2. Intermediate results of breast area identification

Fig. 3 shows the enhanced unlabeled image. Inverted unlabeled mammogram and enhanced inverted mammogram images are shown in Fig. 4. Subtracted image is shown in fig 5. It is clear that suspicious mass region is enhanced in the subtracted images. Filtered image and segmented suspicious mass regions are shown in Fig. 6. The sample segmented mammogram images after applying the FCM algorithm with thresholding technique are shown in Fig. 7. Connected component labeling algorithm is applied to extract the segmented suspicious mass. Fractal dimension using box counting method, three tamura based features i.e., coarseness, contrast and directionality along with ten region based features i.e., area, perimeter,

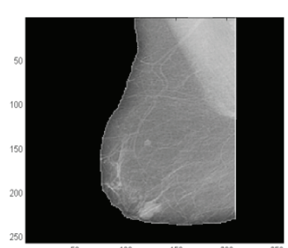
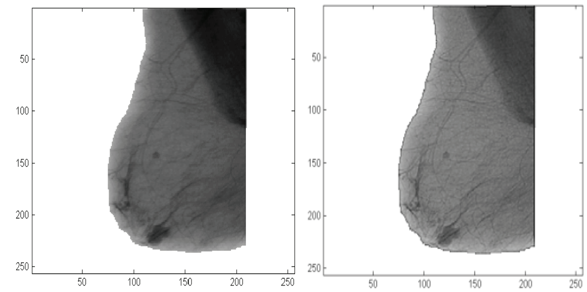
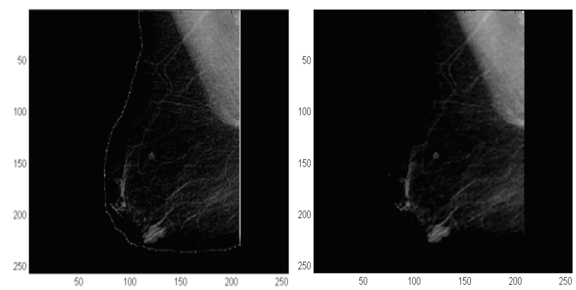


Fig. 3. Enhanced unlabeled image



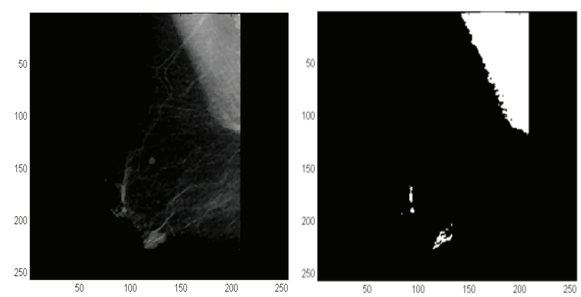
(a) (b)

Fig. 4. (a) Inverted unlabeled image (b) Enhnaced inverted image



(a) (b)

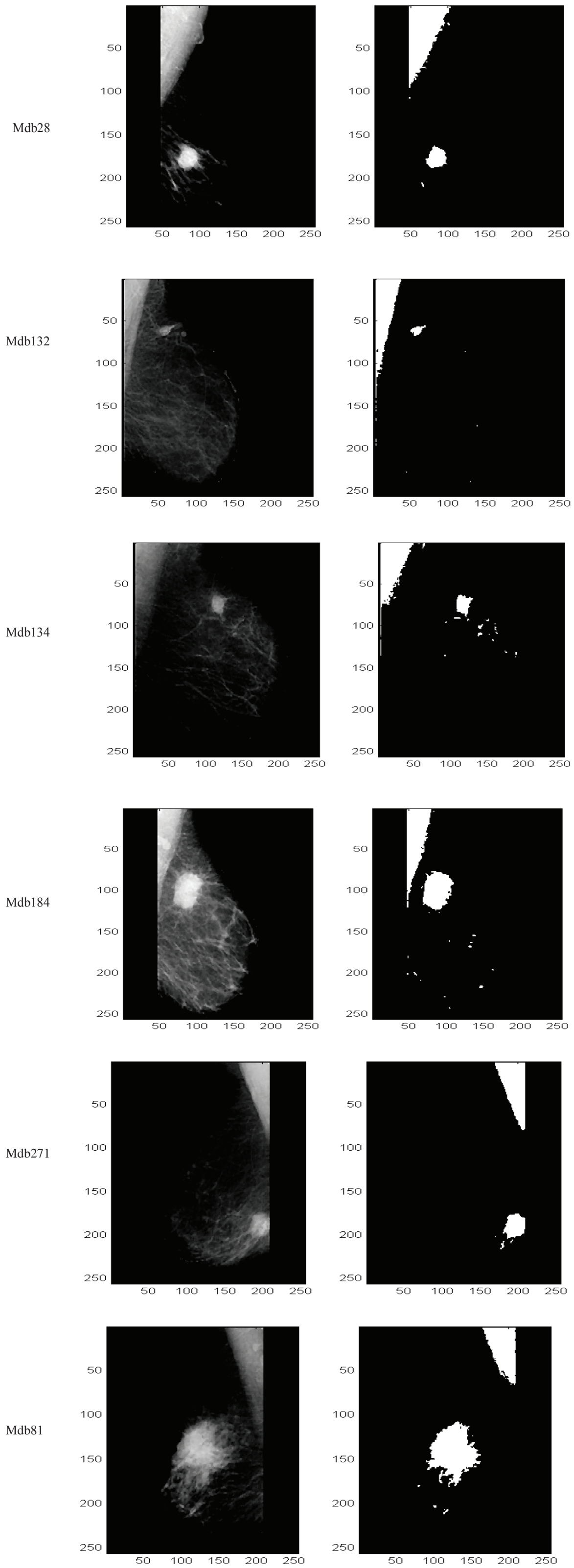
Fig. 5. (a) Subtracted image (b) Filtered image



(a) Sharpened image (b) Segmented image

Fig. 6. Results of suspicious region segmentation

major axis length, minor axis length, eccentricity, orientation, convex area, equi- diameter, solidity, extent are extracted from segmented region. Seven moment invariant are also extracted from the segmented suspicious mass region [10, 13, 14, 15]. Fisher score technique has been used to select the prominent features from the all extracted features [12]. Three region based features eccentricity, solidity and extent are selected from nine features and the first three features are selected from seven moment based features. Support vector machine with radial basis function (RBF) kernel is used to classify. Performance of the classification is measured in terms of sensitivity, specificity, accuracy and ROC curve [16]. Sensitivity measures the proportion of actual positives which are correctly identified as cancerous. Specificity measures the proportion of actual negative which are correctly identified as noncancerous.



(a) Sharpened image (b) Segmented image

Fig. 7. Results of suspicious region segmentation

Mathematical expression of sensitivity, specificity and accuracy are given as follows:

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

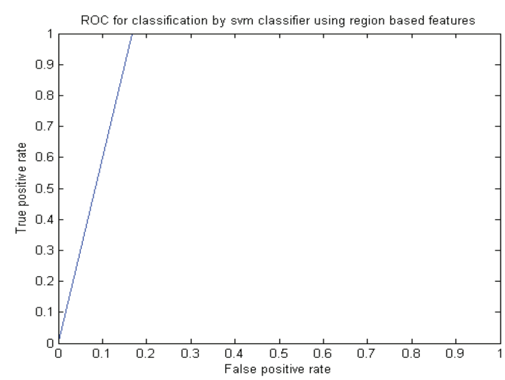
$$Specificity = \frac{TN}{FP + TN} \tag{2}$$

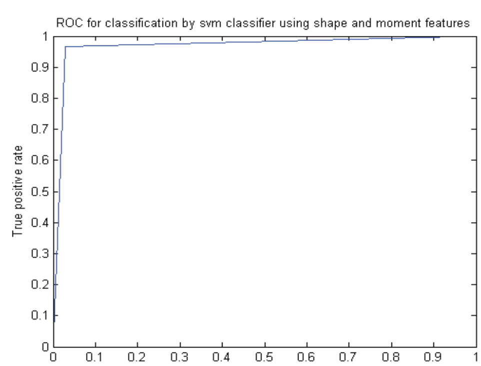
$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{3}$$

here TP, TN, FP and FN represents the true positive, true negative, false positive and false negative, respectively. ROC curve is a plot between the true positive rate (TPR) and false positive rate (FPR). Performance measurement of proposed algorithm is given in Table 1.

TABLE I
PERFORMANCE MEASURE OF PROPOSED ALGORITHM

Features	Sensitivity	Specificity	Accuracy
Region based features	100%	83.34%	91%
Momentinvariant +fractal dimension	97.14%	96.67%	96.92%
Tamura features	85.71%	70%	78.46%





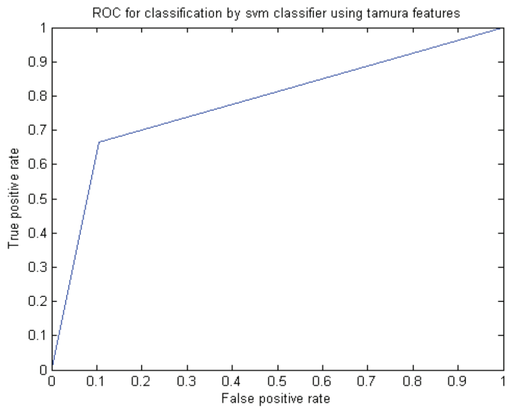


Figure-8 ROC curve for abnormality classification

100%, 97.14%, 85.17% sensitivity and 83.34%, 96.67%, 70% specificity has been achieved using region based features, combination of moments with fractal dimension and tamura based features, respectively as shown in the table 1. Area under curve A_z of 0.91, 0.96, 0.78 has been achieved as shown in the fig 8. Performance of the proposed algorithm with other existing methods in this field is summarized in Table-2.

TABLE II
COMPARISIONS WITH EXISTING METHODS

Methods	**Features**	**Classifier**	**No. of images**	**Accuracy**
Proposed Algorithm	Region based,	SVM	322	91%
	Moment invariant +fractal dimension,			96.92%
	Tamura features			78.46%
Brzakovic et al[17]	Shape based features	Decision tree	25	85%
Cao et al.[18]	Intensity, shape and texture	SVM	60	90.7% Sensitivity
Pasquale et al.[19]	Intensity, shape, texture	Feed Forward Neural Network	226	80% Sensitivity

IV. CONCLUSIONS

Proposed method presents an algorithm to enhance, segment and classify abnormalities present in the mammogram. The following conclusions can be derived from the work:

- Subtraction of enhanced preprocessed and enhanced inverted preprocessed image improves the detection of suspicious region in mammograms.
- Edges of the suspicious region are sharpened by adding detailed coefficients of the wavelet decomposition with filtered image.
- Accuracy of the segmentation is improved by using FCM algorithm.
- Moment based features has given better result than the region based and tamura based features.

Accuracy of the proposed algorithm will be improved by extracting some other features of suspicious region. The Classification of the benign and malignant mass will be done in further research work.

REFERENCES

[1] Sharma, P. Khanna, “ROI Segmentation using Local Binary Image,” IEEE International conference on System, Computing and Engineering, pp 136-141, 2013.

[2] W. P. Jr Kegelmeyer, J. M. Pruneda, P. D. Bourland, A. Hillis, M. W. Riggs, and M. L. Nipper, “Computer-aided mammographic screening for speculated lesions,” Radiology, vol. 191, no. 2, pp. 331–337, 1994.

[3] R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi, and M. Roffilli, “A novel featureless approach to mass detection in digital mammograms based on support vector machines,” *Phys. Med. Biol.*, vol. 49, no. 6, pp. 961–975, 2004.

[4] B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, “Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images,” *IEEE Trans. Med. Imag.*, vol. no. 15, pp. 598–610, 1996.

[5] R. Bellotti, F. D. Carlo, S. Tangaro, G. Gargano, G. Maggipinto, M. Castellano, R. Massafra, D. Cascio, F. Fauci, R. Magro, G. Raso, A. Lauria, G. Forni, S. Bagnasco, P. Cerello, E. Zanon, S. C. Cheran, E. L. Torres, U. Bottigli, G. L. Masala, P. Oliva, A. Retico, M. E. Fantacci, R. Cataldo, I. D. Mitri, and G. D. Nunzio, “A completely automated CAD system for mass detection in a large mammographic database,” *Med. Phys.*, vol. no. 33, pp. 3066–3075, 2006.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[6] N. R. Mudigonda, R. M. Rangayyan and J. E. Desautels, “Detection of breast masses in mammograms by density slicing and texture flow-field analysis,” IEEE Trans. Med. Imaging. 20, pp 1215-122, 2001.

[7] A. R. Dominguez, A. K. Nandi, “Detection of masses in mammograms via statistically based enhancement multilevel-thresholding segmentation, and region selection,” Comput. Med. Imaging Graphics. 32, pp 304-315, 2008.

[8] S. Sharma and P. Khanna, “Computer-Aided Diagnosis of Malignant Mammograms using Zernike Moments and SVM,” Journal of Digital Imaging, vol. 28, pp 77-90, 2015.

[9] K. L. Kashyap, M. K. Bajpai, P. Khanna, “A Novel Approach for Segmentation of Mammographic image to detect breast cancer,” 7th International Symposium on Process Tomography Dresden, Germany, 2015 [Accepted].

[10] R. C. Gonzalez., "Digital image processing," 2nd Edition. India: Prentice Hall, 2003.

[11] J. C. Bezdek, "Pattern Recognition with Fuzzy Objective Function Algorithms," Plenum Press, New York, 1981.

[12] R. O. Duda, P. E. Hart, D. G. Stork, Pattern Classification, second ed., John Wiley and Sons, New York 2001.

[13] H. Tamura, S Mori , T. Yamavaki " Textural features corresponding to visual perception," IEEE Trans On Systems, Man and Cybernetics, vol 8, pp 460-472, 1978.

[14] M. K. Hu, "Visual pattern recognition by moment invariants," IRE Transactions on Information Theory, vol. 8, pp 179-187, 1962.

[15]] T. M. Nguyen and R. M. Rangayyan, ''Shape Analysis of Breast Masses in Mammograms via the Fractal Dimension," in Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference Shanghai, China, 2005.

[16] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic _ROC_ curve," Radiology **143**, pp. 29–36, 1982.

[17] D. Brzakovic,X. M. Lau, P. Brzakovic, ''An Approach to Automated Detection of Tumors in Mammograms,'' IEEE Trans. On Med. Imaging, vol.9 , pp. 233-241, 1990.

[18] A. Cao, Q. Song, X. Yang, "Robust information clustering incorporating spatial information for breast mass detection in digitized mammograms,'' Computer Vision and Image Understanding, 109, pp. 86-96, 2008.

[19] P. Delogu, M. E. Fantacci, P. Kasae, A. Retico, "Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier," Computers in Biology using a gradient-based segmentation algorithm and a neural classifier," Computers in Biology and Medicine 37, pp. 479 – 1491 2007.

[20] J. Suckling, " the mammographic image analysis society digital mammogram database," in 2nd international workshop on digital mammography, 1994.