2014 22nd International Conference on Pattern Recognition

Improvement of Benign and Malignant Probability Detection based on Non Sub-sample Contourlet Transform and Super Resolution

Fatemeh Pak
Department of Computer and Information Technology Engineering
Qazvin Branch, Islamic Azad University
Qazvin, Iran
fatemehpak@qiau.ac.ir

Hamidreza Rashidy Kanan
Department of Electrical, Biomedical and Mechatronic Engineering
Qazvin Branch, Islamic Azad University
Qazvin, Iran
h.rashidykanan@qiau.ac.ir

Afsaneh Alikhassi
Department of Radiology, Faculty of medicine
Tehran University of Medical Sciences
Tehran, Iran
afsanehalikhassi@yahoo.co.uk

***Abstract*—Mammography is a standard method for early diagnosis of breast cancer. In this paper, a method has been provided for improving quality of mammographic images to help radiologists so that probability of benign or malign breast lesions can be detected faster and more accurate and false positive rate (FPR) can be reduced. The presented algorithm includes 3 main parts of preprocessing, feature extraction and classification. In the preprocessing stage, a region of interest (ROI) is determined and quality of images is improved by non sub-sample contourlet transform (NSCT) and super resolution (SR) algorithm altogether. In feature extraction stage, some features of the image components are extracted and skewness of each feature is calculated. Finally, support vector machine (SVM) is used to classify and determine probability of benign and malign disease. The obtained results on MIAS database indicate efficiency of the proposed algorithm.**

Keywords—Breast cancer, Mammography, Non sub-sample contourlet transform, Super resolution, Support vector machine, BI-RADS, MIAS database

I. INTRODUCTION

Diagnosis of breast cancer is known as one of the important subjects in medical science. Two cases of the most prevalent symptoms of breast cancer are masses and calcifications. Since some calcifications are very small and density difference between the healthy tissue and masses may be very low, diagnosis process will be difficult. Therefore, considering importance of proper diagnosis, computer aided detection (CAD) techniques were presented in recent years to help physicians and also reduce false positive rate (FPR) to perform diagnosis action faster, more easily and more accurately. Dominguez and Nandi [1] improved mammographic images based on statistical criteria. Then, they segmented suspicious regions using multi-level thresholding and extracted feature from any region. Finally, they utilized a ranking system for determining suspicious regions. Meenalosini and Janet [2] segmented the suspicious regions using region growing method after removing noise from the images and extracted some features from tissues using spatial gray level dependence. Then, they used support vector machines for classifying the extracted features. Oliver et al. [3] segmented region of interest using a level set algorithm based on region information. Then, they used Zernike moments to characterization of each segmented mass for modeling its shape. Finally, they applied Gentleboost algorithm to diagnosis of mass as benign or malignant.

Due to correspondence of the lesions on high frequency components [4], wavelet transform is one of the known methods in this research field which is able to determine breast lesions [5-7]. Wavelet transform has limitation in capturing directional information in images such as smooth contours and the directional edges, despite its expanded application among image processing techniques. Contourlet transform (CT) which is developed form of discrete wavelet transform is applied for solving this problem. Contourlet transform which has been presented by Do and Verrerli [8] has other properties such as directionality and anisotropy in addition to properties of wavelet transform. Moayedi et al. used contourlet transform for extracting feature from the determined masses [9] but, considering that contourlet transform lacks shift-invariance due to downsampling and upsampling, non sub-sample contourlet transform (NSCT) was presented by Cunha et al. [10] for compensating for this limitation.

In this paper, NSCT transform was applied for improving quality of images after determining region of interest (ROIs); then, super resolution (SR) algorithm was used to increase resolution of the images and a high-pass filter was used to highlight the desired regions. In the next stage, objects of the image are obtained and 7 features are extracted from them and skewness of each feature has been calculated. Finally, the obtained feature matrix is given to support vector machines (SVM) algorithm for determining breast imaging reporting and data system (BI-RADS).

II. THE PROPOSED APPROACH

The presented technique includes 3 main stages including: 1) preprocessing, 2) feature extraction and 3) classification and BI-RADS detection. Fig. 1 shows structure of the presented method. Details of each section of the proposed algorithm will be presented.

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.164

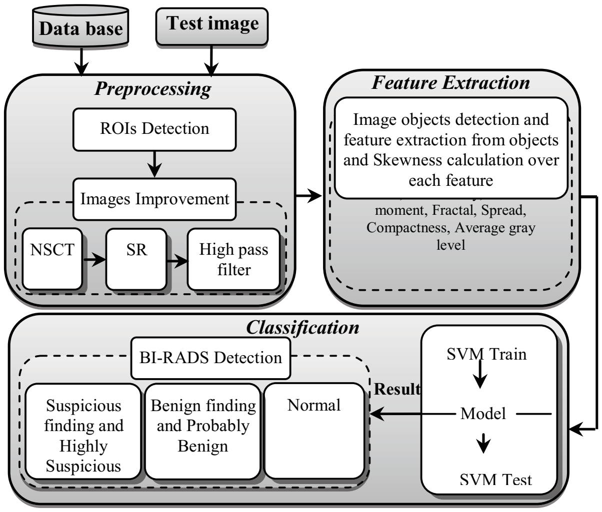


Fig. 1. The proposed system architecture

A. Preprocessing

To reach desirable results and distinguish between benign and malign lesions, preprocessing action is performed in the proposed method in two stages of determining region of interest and then improving quality of mammographic images.

a. Region Of Interest (ROIs) Detection

The first stage of the proposed method is to remove additional margins, which makes images smaller and finally reduces computational burden (Fig. 2 (b)). On the other hand, some features of masses and calcifications are very similar to regions of tissue such as pectoral muscle or similar to some artifacts such as label that include information of the patient in corner of the images and these regions should be removed for reducing false positive rate to obtain ROIs in mammographic images. For this purpose, all images have been aligned to the left side to place pectoral muscle for all images on the left side and label on the right side (Fig. 2 (b)). To remove pectoral muscle and label, thresholding methods and Erosion morphological operator have been utilized (Fig. 2 (c)).

b. Mammographic Images Improvement

It is necessary to promote quality of mammographic images for highlighting masses and calcifications in tissue of breast because it causes the edges and regions of the image to be extracted and studied effectively. Since edges of the image are superposed with high–frequency components, a method should be applied in frequency domain to achieve them. For this purpose, NSCT transform has been applied on the images and then SR algorithm has been utilized for increasing resolution of the image. At the end, a high-pass filter has been used for sharpening and highlighting desired regions.

- **NSCT:** In contourlet transform, the laplacian pyramid (LP) is first used to capture point discontinuities, and then followed by a directional filter bank (DFB) to link point discontinuities into linear structures [8]. The overall result is an image expansion using basic elements like contour segments, and thus called contourlet transform, which is implemented by a pyramidal directional filter bank (PDFB) [8]. The LP decomposition at each level generates a downsampled lowpass version of the original image, and the difference between the original image and the prediction results in a bandpass image. Due to downsampling and upsampling presented in both LP and DFB, contourlet transform is not shift-invariant. So, to achieve the shift-invariance property, NSCT was proposed by Cunha et al. [10]. The NSCT is built upon nonsubsampled pyramids (NSLP) and nonsubsampled directional filter bank (NSDFB) that is fully shift-invariant, multiscale and multidirection image decomposition.

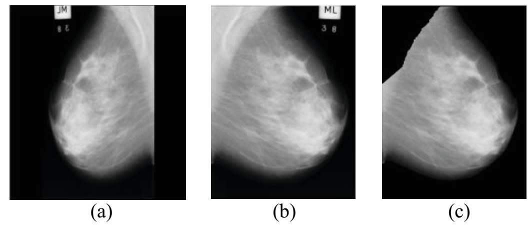


Fig. 2. ROIs detection. (a) original image. (b) additional margin removal and alignment image. (c) label and pectoral removal.

- **SR algorithm:** Goal of SR is to achieve an image with high resolution based on one or a set of images with low resolution. Since SR technique is a way of increasing resolution without altering the existing imaging hardware, it can be suitable for medical images [11]. Goal of SR algorithm is to improve resolution of the images which firstly include high-frequency components and secondly aliasing and degradation have occurred in the images with high-frequency components. Therefore, SR can increase sampling rate by utilizing information of high-frequency components and also reducing aliasing effects caused by application of NSCT [11]. In this research, SR technique based on learning algorithms has been used among the SR techniques [12] because it is suitable for single image problems and is able to predict high-frequency components.
- **Final Image Improvement:** In the proposed method, after determining ROIs, dimensions of the image have decreased by half of the main dimensions due to largeness of mammographic images and high computational rate. Then, three-level NSCT decomposition has been applied on the images and after applying this transform, 2 subbands in the first level, 4 subbands in the second level, 8 subbands in the third level and 1 low-pass subband which has low-frequency components are obtained. Among the obtained subbands, only subbands with high-frequency components are used. To achieve edges of image, "prewitt" edge detector has been used on each subbands and to obtain stronger edges of the image, standard deviation of each subbands are calculated and applied as a threshold in "prewitt" edge detector. After finding the edges in each subbands, a weight is given to some regions of subbands which include edges to highlight edges. Then, the image is reconstructed after making changes in its high-frequency subbands. Results of this section are found in Fig. 3 (c). In the proposed method, SR algorithm has been used based on fuzzy learning algorithm [13] after reconstruction of the image to improve quality of the image again and increase its resolution (Fig. 3 (d)). Finally, a high-pass filter is used for sharpening and highlighting the desired regions. The final improved image is shown in Fig. 3 (e).

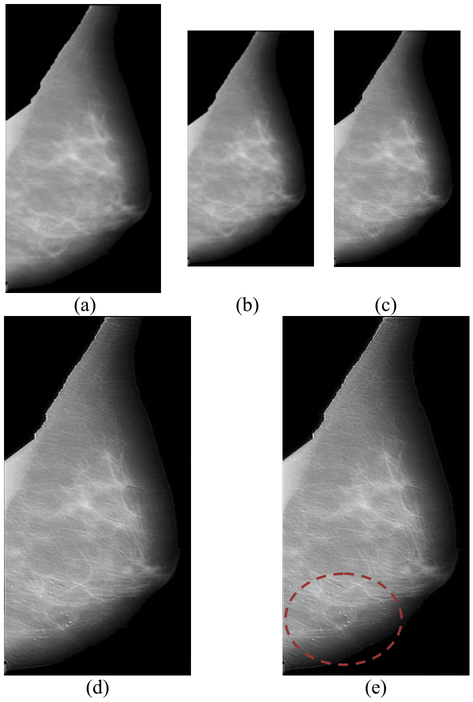


Fig. 3. Image improvement steps. (a) the original image after ROIs detection. (b) Resized Image. (c) improved image by applying NSCT. (d) increase resolution of (c) by applying fuzzy SR algorithm. (e) final improvement Image by applying high pass filter to (d).

B. Feature Extraction

In feature extraction stage, it is necessary to extract the information from mammographic images so that the system can distinguish between normal and abnormal tissues correctly. Since masses and calcifications have higher density than other tissues of the breast and are brighter than them, a threshold value is considered in the proposed method to produce binary images including masses and calcifications. Then, the regions available in the images are labeled and objects of the images are obtained. At the end, features are extracted from the objects of the images. Due to high importance of lesions' shape in determination of benign or malign probability, the features which study shape of lesions are mostly used in this paper. These features include area [14, 15], eccentricity [14], central moment [14], fractal [15], spread [14], compactness [14, 15] and average gray level.

- **Area (A):** A suitable descriptor for expressing size of the objects available in the image.
- **Eccentricity (E):** The range of values for this feature is [0-1]. The value which is closer to 0 will give circular objects and the value which is closer to 1 will give linear objects.
- **Central moment (M):** It obtains some information about roughness of the shapes. The central moment of order k of a distribution is defined as Equation (1):

$$m_k = E(x-\mu)^k \tag{1}$$

where, $E(X)$ is the expected value of X. The range of values for this feature is also [0-1]. If this value is larger and close to 1, irregularity in the shapes will occur more.

- **Fractal (F):** Micro calcifications are small light local anomaly points which represent sharp local changes in contrast of the image from the fractal standpoint and rare events in global sense [16]. A higher value of fractal corresponds to irregularity contour and thus to a higher probability of malignancy.
- **Spread (S):** It measures how unevenly objects are distributed about their centroid and it is based on the central moments of the boundary pixels. Spread is defined as Equation (2):

$$S = \mu_{o,2} + \mu_{2,0} \tag{2}$$

where, the moments *f(x,y)* translated by an amount *(p, q)*, are defined as Equation (3):

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y) \tag{3}$$

The range of values for this feature is also obtained between 0 and 1. Again, a lower value represents a circular object while a large value defines a linear and non-uniform object.

- **Compactness (C):** Compactness is a dimensionless quantity which provides a simple measure for complex counters. Compactness is independent of translation, rotation and scale that defined as Equation (4):

$$C = \frac{P^2}{4\pi A} \tag{4}$$

where, P is perimeter and A is area of the objects. A larger value of this feature describes an irregular and elongated object while a smaller value is representative of a more symmetric and regular object.

- **Average gray level (AG)**: Since masses and calcifications are brighter than the background, therefore, average values of pixels of the image objects can be applied as a suitable feature.

Considering the introduced features and their definition, skewness (SK) of each feature is calculated as Equation (5):

$$SK = \frac{E(x-\mu)^3}{\sigma^3} \tag{5}$$

where, μ is the mean of x, σ is the standard deviation of x and $E(t)$ represents the expected value of the quantity t. Skewness is a measure of the asymmetry of the data which basis one can find roughness or smoothness, regularity or irregularity, symmetry or asymmetry, circularity or linearity and brightness of the objects of the image and decide if they are malign or benign. Therefore, feature vector will be obtained for each image as Equation (6):

$$F = [SK(A) \quad SK(E) \quad SK(M) \quad SK(F) \quad SK(S) \quad SK(C) \quad SK(AG)] \tag{6}$$

C. Classification and BI-RADS Detection

In this paper, SVM [17] are used for classification and the related BI-RADS have been determined. American college of radiology (ACR) [18] classifies mammographic images using BI-RADS in 6 general categories considering breast lesions.

- BI-RADS 0: Evaluation is not complete
- BI-RADS I: Normal
- BI-RADS II: Benign finding
- BI-RADS III: Probably Benign

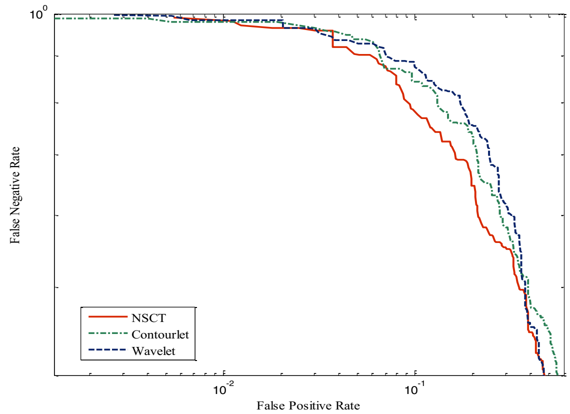


Fig. 4. DET curves for comparing of different time-frequency transforms

TABLE I. COMPARISON OF MAMMOGRAPHIC IMAGE IMPROVEMENT BY USING DIFFERENT TIME-FREQUENCY TRANSFORMS

methods	Mean accuracy	Mean AUC	Mean F-measure	FPR	Max accuracy
NSCT	87.26%	0.8567	80.90%	0.0955	96.29%
Contourlet	85.87%	0.8411	78.81%	0.1059	92.59%
Wavelet	82.17%	0.7994	73.26%	0.1337	96.29%

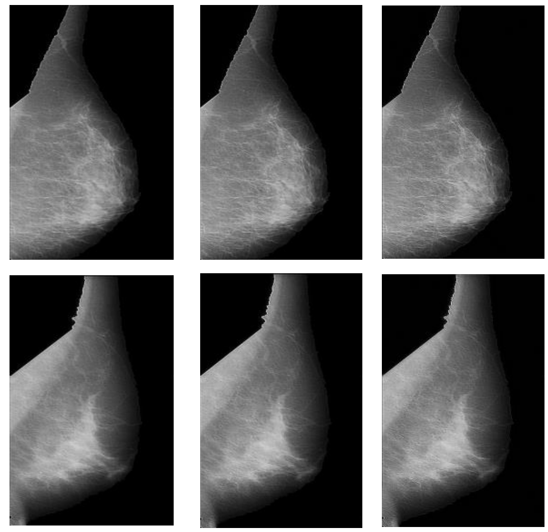
- BI-RADS IV: Suspicious finding
- BI-RADS V: Highly Suspicious

In the proposed system with available information of existing database, the related BI-RADS has been determined in three classes: normal (BI-RADS I), benign (BI-RADS II,III) and malign (BI-RADS IV, V).

III. Experimental Results

In this paper, a mammographic image analysis society (MIAS) database [19] has been used to evaluate efficiency of the proposed system. This database includes 322 images with dimensions of 1024×1024 of which 208 images are normal and 80 images include masses and calcifications. 47 out of these 80 abnormal images have benign lesion and 33 images have malign lesion. It is necessary to note that the set including 288 normal and abnormal images have been used to test the proposed method.

In the proposed method, a three-level NSCT decomposition has been applied after determining ROIs in stage of improving quality of mammographic images and in each level, 2, 4 and 8 subbands are obtained respectively. In image decomposition with NSCT, “maxflat” and “dmaxflat7” are used as NSLP and NSDFB filter. After improving the images and extracting feature of image objects, the formed feature matrix is given to SVM for determining BI-RADS. Simulation results show that when Gaussian kernel with degree 5 is used and value of parameter C is considered equal to 10000, the best result is obtained. To evaluate the proposed method, 16-fold cross validation has been used. In addition, to show efficiency of the provided system, some measures such as sensitivity, specificity, accuracy, area under curve (AUC), F-measure and FPR have been also applied.



(a) (b) (c)

Fig. 5. Comparison of different SR methods. (a) original image after applying NSCT (b) applying SR algorithm based on bicubic interpolation (c) applying SR algorithm based on fuzzy learning.

In process of improving quality of mammographic images, NSCT was used as said before due to need for edges of the image and efficiency of this methods was studied with other methods in frequency domain (contourlet and wavelet). The detection error trade-off (DET) curves in Fig. 4 and also Table I indicate efficiency of the provided system. The highest mean accuracy which is equal to 87.26% relates to NSCT method with FPR of 0.095.

In the second stage of improving quality of mammographic images, SR algorithm was applied based on fuzzy algorithm on the images. To show efficiency of the system, the proposed method has been compared with bicubic interpolation method, which is shown in Fig. 5 Mean accuracy is equal to 82.40% when bicubic is used as resolution increase and FPR will be 0.131 which has lower accuracy and higher false positive rate than the used SR technique.

In the final experiment, to show the performance of proposed classification, the result of SVM compares with different methods in Fig. 6 by using box-whisker plots. Horizontal axis shows accuracy of system and vertical axis shows the values are obtained for this measure. Box-plot can be showed average and median. On each box, the horizontal line denotes median, the circle denotes mean and the horizontal lines outside each box identify the upper and lower whiskers, and dot points denote the outliers. The dotted line in Fig. 6 shows the highest mean accuracy for SVM among the other classifiers.

In Table II, shows the comparison results of the proposed method with other methods which have been tested on MIAS database. The obtained results indicate high efficiency of the presented system.

IV. Conclusion

In this paper, a method for improving quality of the mammographic images to help radiologists was presented, so that probability of benign or malign breast lesions can be detected faster and more accurate and FPR can be reduced.

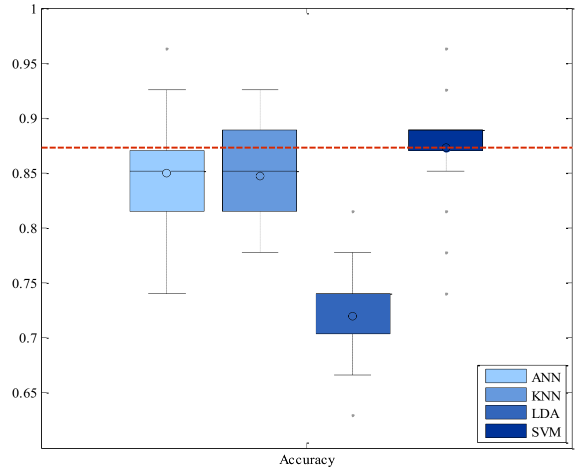


Fig. 6. Comparison of different classifier based on accuracy. ANN: artificial neural network, KNN: k-nearest neighbor, LDA: linear discriminant analysis and SVM: support vector machine.

TABLE II. COMPARISON WITH OTHER METHODS

Methods	Accuracy	Sensitivity	Specificity
Ferreira and Borges [5]	94.85%	-	-
Mohanalin et al. [20]	-	0.9375	-
Moayedi et al. [9]	82.1%	-	-
Meenalosini and Janet [2]	-	0.952	0.944
Proposed approach (mean accuracy)	87.26%	0.8090	0.9045
Proposed approach (max accuracy)	96.29%	0.9444	0.9722

The presented algorithm includes 3 main parts of preprocessing, feature extraction and classification. Advantages of this method include: firstly, NSCT has been used in the proposed method unlike the previous methods which have used wavelet transform for determining edges of images. NSCT is preferred over other methods in frequency domain. Secondly, SR algorithm has been used to predict high-frequency components better and remove distortions in the proposed method. Also, to obtain a system which can distinguish between normal and abnormal tissues, 7 features were extracted from the objects in mammographic images and each feature was analyzed by calculating skewness of each feature to decide if they are malign or benign.

Results of tests show that the proposed method can obtain mean and maximum accuracy of 87.26% and 96.29% respectively based on cross validation strategy on MIAS database. Also, FPR of the radiologists which is equal to 15% in mammographic images is reduced averagely to 9.55% and maximum 2.87%.

REFERENCES

[1] A. Dominguez and A. Nandi, “Detection of masses in mammograms via statistically based enhancement, multilevel-thresholding segmentation, and region selection”, journal of Computerized Medical Imaging and Graphics, Vol. 32, No. 4, pp. 304-315, 2008.

[2] S. Meenalosini and J. Janet, “Computer Aided Diagnosis of Malignancy in Mammograms”, Eropean Journal of Scientific Research, Vol. 72, No. 3, pp. 360-368, 2012.

[3] A. Oliver, A. Torrent, X. Liado and J. Marti, “Automatic diagnosis of masses by using level set segmentation and shape description”, International Conference on Pattern Recognition, pp. 2528-2531, 2010.

[4] J. Bozek, M. Mustra, K. Delac and M. Grgic, “A Survey of Image Processing Algorithms in Digital Mammography”, Advances in Multimedia Signal Processing and Communication, Vol. 231, pp. 631-657, 2009.

[5] C. B. R. Ferreira and D. L. Borges, “Analysis of mammogram classification using a wavelet transform decomposition”, Pattern Recognition Letters, Vol. 24, No. 7, pp. 973-982, 2003.

[6] Y. Sung-Nien, L. Kuan-Yuei and H. Yu-Kun, “Detection of microcalcifications in digital mammograms using wavelet filter and Markov random field model”, Computerized Medical Imaging and Graphics, Vol. 30, No. 3, pp. 163-173, April 2006.

[7] V. Alarcon-Aquino, O. Starostenko, J. M. Ramirez-Cortes, R. Rosas-Romero, J. Rodriguez-Asomoza, O. J. Paz-Luna and K. Vazquez-Munoz, “Detection of Microcalcification in Digital Mammograms Using the Dual-Tree Complex Wavelet Transform”, Engineering Intelligent Systems, Vol. 17, No. 1, pp. 49-63, 2009.

[8] M. N. Do and M. Vetterli, "The contourlet transform: an efficient directional multiresolution image representation", IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, vol. 14, No. 12, pp. 2091-2106, 2005.

[9] F. Moayedi, Z. Azimifar, R. Boostani and S. Katebi, “Contourlet-based mammography mass classification using the SVM family”, Journal of Computers in Biology and Medicine, vol. 40, No. 4, pp. 373-383, 2010.

[10] A. L. Cunha, J. Zhou and M. N. Do, “The Nonsubsampled Contourlet Transform: Theory, Design, and Applications”, IEEE Transaction On Image Processing, Vol. 15, No. 10, pp. 3089-3101, 2006.

[11] H. Greenspan, “Super-Resolution in Medical Imaging”, Published by Oxford University Press on behalf of the British Computer Society, Vol. 52, No. 1, pp. 43-63, February 2008.

[12] J. Tian and K. K. Ma, “A survey on super-resolution imaging”, Signal, Image and Video Processing, Vol. 5, pp. 329-342, September 2011.

[13] P. Purkait and B. Chanda, “Fuzzy-Rule Based Approach forSingle Frame Super Resolution”, IEEE Inetrnational Conference on Fuzzy Systems, pp. 1-7, 2013.

[14] M. Tembey, Computer-aided diagnosis for mammographic microcalcification clusters, M.S.C.S. Thesis, University of South Florida, 2003.

[15] W. N. Street, W. H. Wolberg and O. L. Mangasarian, “Nuclear Feature Extraction For Breast Tumor Diagnosis”, International Symposium on Electronic Imaging: Science and Technology, Vol. 1905, pp. 861-870, 1993.

[16] B. C. Patel and G. Sinha, “Early Detection of Breast Cancer using Self Similar Fractal Method”, International Journal of Computer Applications, Vol. 10, No. 4, pp. 39-43, November 2010.

[17] C. J. C. Burges, "A tutorial on Support Vector Machines for pattern recognition", Data Mining and Knowledge Discovery, Vol. 2, No. 2, pp. 121-167, Jun 1998.

[18] G. Giannakopoulou, G. M. Spyrou, A. Antaraki, L. Andreadis, D. Koulocheri, F. Zagouri, A. Nonni, G. M. Filippakis, K. S. Nikita, P. A. Ligomenides and G. C. Zografos, “Downgrading BIRADS 3 to BIRADS 2 category using a computer-aided microcalcification analysis and risk assessment system for early breast cancer”, Journal of Computer in biology and Medicine, Vol. 40, pp. 853-859, 2010.

[19] MIAS database, http://peipa.essex.ac.uk/ipa/pix/mias/

[20] Mohanalin, Beenamol, P. K. Kalra and N. Kumar, “A novel automatic microcalcification detection technique using Tsallis entropy & a type II fuzzy index”, Journal of Computers and Mathematics with Applications, Vol. 60, No. 8, pp. 2426-2432, 2010.