EFFECTIVE TEXTURE FEATURE MODEL FOR CLASSIFICATION OF MAMMOGRAM IMAGES

K. Rajendra Prasad[1] and M. Suleman Basha[2]
[1]Department of Computer Science and Engineering, Institute of Aeronautical Engineering, Dundigal, Hyderabad, India
[2]Department of Information Technology, RGM College of Engineering and Technology, Nandyal, Andhra Pradesh, India
E-Mail: krprgm@gmail.com

ABSTRACT

Breast cancer detection is an emerging need in mammography and it helps for radiologist for examining the stages of breast cancer detection. Mammogram classification is attempted in this paper using well-known support vector classification method. Mammogram classification follows three key steps, which are feature image enhancement, texture feature extraction, and classification. This paper presents the experimental results of mammogram classification for demonstrating the efficiency of SVM with underlying mechanisms of texture methods and it suggest the best combination of SVM and texture method to radiologist for better medical diagnosis of breast cancer detection.

Keywords: mammogram classification, SVM, feature extraction, texture methods.

1. INTRODUCTION

Mammography diagnosis is an important for radiologists to detect the stages of breast cancer detection. Problem of mammogram classification is simplified as key steps include enhancement of mammogram images, extraction of texture properties, and classification of mammogram images. Contrast stretching [1], histogram equalization [2], wavelet transformation [3], mean filter [4] and median filter [5] are techniques of image enhancement that are used for enhancing of mammogram images in respect to clarity and smoothing of noise levels. The enhanced mammogram image shows the clarity of masses during classification, which may be useful for obtaining more accurate mammogram classification results. Further, enhanced mammogram images are segment using region of interest (ROI) [4] method for choosing interested part of mammogram image and it removes unwanted region, thus size of mammogram image is reduced to 256 x 256 pixels which helps the faster classification results. Later, texture features are extracted from the resulting ROI mammogram images using various texture methods such as local binary patterns (LBP) [5], Histogram of gradients (HoG) [6], Gabor [7] and Gray-level co-occurrence matrix (GLCM) [8].

It is required to analyze texture features for each texture method and the best texture method with underlying mechanism of support vector machine (SVM) [9]. This paper is focused on developed mechanisms of LPB-SVM, HoG-SVM, Gabor-SVM and GLCM-SVM for addressing the best texture method for mammogram classification.

Breast cancer is the most common cancer among women. Thus efficient methods are required for early detection of breast cancer. Images can be classified into two categories based on physical properties such as shape, size. For identifying the cancer or non-cancer, texture data plays an important role and texture feature extraction can be done by different methods. Detection for this distortions or even milk ducts in a mammogram image can be done by classifying them into normal or abnormal.

Features extraction of enhanced mammogram images are shows the significant improvement when compared with normal mammogram images because noise and blur things are smoothed out in enhanced mammogram images. It is noted that only texture methods are applied only on either improved or enhanced mammogram images, in which accurate texture properties are retrieved that helps for detecting of any masses or tissues are found in mammogram images and this kind of accurate diagnosis is exactly needed for in mammography test for radiologist for finding the possibility for occurrence of breast cancer detection.

Contributions of the paper are summarized as follows:

a) Analyzed the enhanced mammogram images
b) Texture methods are applied on enhanced mammogram images
c) Extract the texture features of enhanced mammogram images using GLCM, Gabor, LBP and HoG methods
d) Study and analyze the texture features of mammogram images experimentally.
e) Develop classifier mechanisms of LBP-SVM, GLCM-SVM, Gabor-SVM and HoG-SVM for effective mammogram classification.

Organization of the paper is as follows: Section 2 presents the related work, Section 3 discusses the proposed method, Section 4 shows the experimental results and discussion, and Section 5 defines the conclusion and future scope of the paper.

2. RELATED WORK

For identifying texture, we use some texture methods and computer aided diagnosis (CAD) tool is developed in [10] and it explains the classification of mammogram into normal and cancer pattern in order to support the radiologist for accurate diagnosis on the micro calcifications or masses type of ducts on mammogram

ARPN Journal of Engineering and Applied Sciences

images there are many similar works proposed on studies of [11] for classification of normal and abnormal patterns on mammograms. Since from past years CAD tool which was a cost effective method emerging as an important area for diagnosis CAD system initially extracts the texture features from the mammogram images and classify them in to benign or malignant.

Classification of normal and abnormal images from digital mammograms is required for mammography test. For this classification, extraction of texture features is needed and it is performed by Grey level co-occurrence matrix (GLCM) [11]. This GLCM method is initially defined by Harrick [12] which can be defined as a matrix of frequencies between two pixels however distribution among them would be depend upon the neighborhood distance between pixels. Statistical features used in this GLCM method have been used successfully for texture segmentation of an image. This method is accurate in breast cancer detection but one disadvantage was that here by using GLCM matrix small elements cannot be extracted from the details of a given image particularly region of interest (ROI) [13].

Histogram of oriented gradients (HoG) [14] were used to calculate gradient direction along x and y axis with in an image where as in cs-local binary patterns (LBP) [15] method both gradient and texture information can be compared. In HoG method initially gradient angle for each pixel with in a image was computed here higher magnitude values are considered as a part of edge direction whereas here lower values were discarded. Gabor wavelets [16] will be applied on a image in feature extraction stage .Gabor wavelet can be simply defined as a product of Gaussian kernel with sinusoid.

Enhanced mammogram images would be presents the better results compared to actual mammogram images and enhancement techniques are done generally to highlight certain features of interest in a mammogram image. Some of enhancement methods are contrast stretching [17], morphological operations [18], wavelet transformation [19], mean filter [20], median filter [21], and steerable filter [22]. For verification of quality of enhanced images, two key parameters were used, such as contrast to noise ratios (CNR) [23] which shows the quality improvement of images and peak signal to noise ratio (PSNR) [24].Higher PSNR value indicates that greater quality of mammogram image. Among the enhancement methods we can observe that highest PSNR value can be seen in steerable filter method images, in which images are decomposed into required level by using bank of steerable filter. In case of CNR parameter it is observed that it was highest in steerable filter which indicates the quality improvement of images and least in mean filter because of inefficiency of single pixel representatives. For the classification accuracy of mammogram images, we have used accuracy parameter [25], specificity [26], and sensitivity [27], precision [28]. Here also we used these parameters for discussing the proposed schemas for mammogram classification.

3. PROPOSED METHOD

The support vector machine (SVM) [18] is used for mammogram classification for finding severity of breast cancer. It is performed mammogram classification using extracted texture features and it is addressed that which texture feature gives the best classification results that helps for better diagnosis of breast cancer detection in early stages and saves the life of patient. Therefore, texture based SVM mechanisms are presented in this section; these are LBP-SVM, HoG-SVM, GLCM-SVM, Gabor-SVM.

In LBP, the features are transformed into an array and the LBP operator is applied on each pixel of the image that gives LBP coded image. The operator is depending on the neighborhood values, in which it takes eight neighbor pixels and center pixel intensity value is considered as a threshold. If the center pixel value is greater than neighbor pixel, then LBP operates the corresponding neighbor pixel is 1, otherwise corresponding neighbor pixel is 0. All these eight digits outputs 8 binary numbers and it can be shows the 8 digit number as feature vector and it converted into a decimal number. The example of LBP operator is shown in below:

121	140	80
223	135	90
77	87	66

LBP →

0	1	0
1	135	0
0	0	0

In this LBP processing, 135 is the intensity value of the centered pixel and it is compared with remaining eight-neighboring pixels in which the neighbor pixels are replace with binary number whose are greater than centered pixel value and other pixel values are replaced with zero. Local binary patterns (LBP) plays important role in texture classification i.e. whenever there is a need to assign new texture to texture classes. In this method initially bitmap image data gets converted in to a standard grey level image. Next LBP code can be analyzed by considering the pixel at a position and by comparing the grey level values with other pixel positions.

Gabor filter is another feature extraction method and it is used three key steps before extraction of texture features from mammogram images. These steps are as follows: segment the mammogram image into sub-windows, apply Gabor filter bank on each of sub-window, important moment calculations such as mean, standard deviation, and skewness are computed from Gabor filter bank. Gabor filter describe the tool that defines the texture properties of the image in terms of spatial frequencies and their orientations. These orientations and frequencies are described by a 2D Gabor filter and it is modulated by a sinusoidal wave and the corresponding equations are shown in following Equations (1) and (2).

ARPN Journal of Engineering and Applied Sciences

$$g(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}e\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)\right]e(2\pi jW\overline{x}) \quad (1)$$

$$frequency(i)=f\max=0.2/(\sqrt{2})^{i-1}$$
$$\overline{x}=x\cos\theta+y\sin\theta \quad (2)$$

Here, x and y are denotes the spatial variables, 'σ_x' and 'σ_y' are used for scaling purpose and it takes the neighborhood summation and following Figure-1 shows the different number of Gabor filters with the orientations and frequencies of filter bank and these values are computed by following Eqn. (3) and (4)

$$orientation(i)=\frac{(i-1)\pi}{m} \quad (3)$$

$$frequency(i)=f\max=0.2/(\sqrt{2})^{i-1} \quad (4)$$

In Eqn. (3), 'm' refers to total number of orientations. This Gabor filter is applied on each window of region of interest (ROI) for extraction of Gabor based texture features in which texture features are formed with three statistical measures i.e. mean, standard deviation and skewness.

GLCM is another texture method and it is used for collecting the information about pixels. It tabulates the frequencies of pixels with specific brightness of an mammogram image. The information of pixels is constructed in matrix format for the different directions with angles of 0°,45°,90° and 135°. Multiple directions of the angle may outputs the reliable texture information, thus, these four directions of may be used for extraction of textual features of mammogram image. The co-occurrence matrix computation applied to a mammogram is shown in Figure-1.

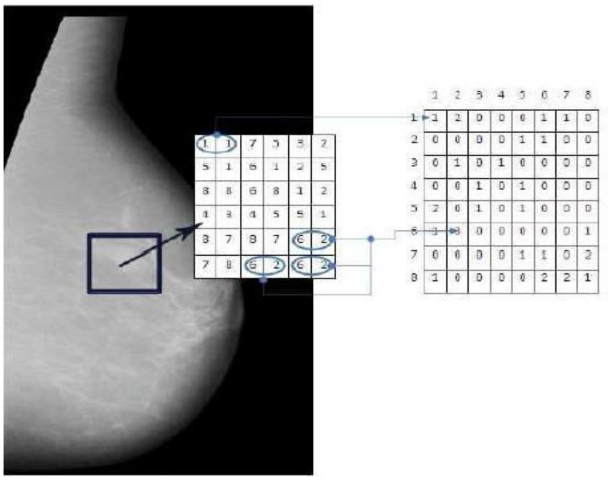

Figure-1. Example of a co-occurrence matrix with eight grey levels.

C. J. Rose and C. J. Taylor [14] proposed another texture method, namely Histogram of Gradients (HoG) and it has also another robust texture feature method, in which the mammogram image is divided into blocks or small cells and compute the weighted histogram on each of the block or cell. HoG normalize the frequencies in the interval of [0 1]. HoG descriptors are derived from the combination of those histograms of blocks.

The texture features of four methods, namely, LBP, GLCM, HoG and Gabor are used in training phase and testing phase of SVM classification for building the models of different cases of mammogram images (i.e normal and abnormal).The learning of SVM for mammogram image derives the objective function of training samples such as mean square error and it minimize a generalized error. The SVM is defined by the function and it shown in Equation (5) and (6)

$$f(x)=w^T\phi(x)+b \quad (5)$$

$$f(x)\geq 0\,for y_i=+1 and f(x)<0\,for y_i=-1 \quad (6)$$

Here w refers to a cost function to find the closest point from sour point and this closet point is maximal, usually known as 'optimal separating hyperplane', and decision support vectors 'y' are determined by minimizing the square of w. The texture features are varying for each texture method and it is observed in SVM classification method through finding the accuracy of respective classifier model and it is discussed in detailed in following experimental section.

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

Manual analysis use the computer aided diagnosis (CAD) for predicting breast cancer from mammogram images so automated analysis play an important role. To extract the visual data from a given mammogram image different set of feature extraction methods will be used in an image processing. Initially pre-processing will be performed on a mammogram image which is a major step to improve the image quality later image segmentation which transforms blur image into clear one and then feature extraction, classification methods were performed.

In this work four feature extraction methods such as local binary patterns (LBP), histogram of oriented gradients (HOG), Gabor, Grey level co-occurrence matrix (GLCM) are used for testing them on SVM classifiers for breast cancer detection. Support vector machines (SVM) is an optimal choice for learning of mammogram data as it is a supervised learning method which works effectively on mammogram classification and SVM completely relies on finding best separable hyper plane which separates data in to its two classes.

SVM classifier being able to predict multiclass problems from the obtained training examples which can be able to classify objects based on nearest class labels. In training phase we map the support vectors into a high

ARPN Journal of Engineering and Applied Sciences

multidimensional space with an each vector having a class label of either cancer or non-cancer. In classification phase, based on the test point classification will be done by assigning the most frequent one among training samples. For the four feature extraction methods it is required to find accuracy percentage by calculating the correct predictions by total number of predictions from the confusion matrix [28]. Performance measures such as accuracy, precision, sensitivity and specificity were used. Another important measure Receiver operating characteristics (ROC) curve can be plotted by taking the fraction of predicted true positives rates by false rates or chances of being abnormal. The ROC curve used for better diagnosis for comparing performances of four methods. The final experimental result demonstrates that SVM would be most suitable and accurate mammogram classifier method.

5.1 Dataset collection

Initially we take a set of mammographic images from the MIAS database [15] which belongs into two categories like normal and abnormal. Here initially breast images will be transformed in to digital format at a fixed resolution of 1024*1024 pixels (grey level).The proposed work focus on the effective accuracy method for classification into normal and abnormal one. Ina mammographic image not every portion is necessary for detection hence ROI extraction is needed. By knowing about the position of abnormalities per each mammogram image ROI features were extracted.

Accuracy: The classification accuracy of these each feature extraction method depends upon the no of correctly classified mammogram data in to normal or abnormal and it shown in Equation (7).

$$\text{Accuracy}=\frac{(TP+TN)}{(P+N)} \tag{7}$$

Sensitivity: It can be defined as fraction of positive value.We can measure the classification range between 1 and 0 which indicates worst and best classification accuracy and it shown in Eqn. (8). Other measures specificity and precision are shown in Eqn. (9) and Eqn. (10)

$$\text{Sensitivity}=\frac{TP}{(FP+TN)} \tag{8}$$

$$\text{Specificity}=\frac{TN}{(FN+TN)} \tag{9}$$

$$\text{Precision}=\frac{TP}{(TP+FP)} \tag{10}$$

Whereas here ‘TP’ makes count of sample images being cancer, ‘FP’ makes count of sample images which are incorrectly classified of being cancer while they are normal, ‘TN’ makes count of sample images of being non cancer, ‘FN’ counts sample images which are incorrectly classified of being non cancer while they are being cancer.

In this work training examples can be taken from the set of texture image classes. By concatenating all texture classes in to single texture class and representing them as a single set of points. The final important step is using the SVM classifier in order to recognize the texture class by accuracy, results obtained by using SVM would be compared with those obtained using a nearest classifier. SVM generally operates on a numeric attributes. Hence here detected cases of normal or abnormal can be analyzed from the confusion matrix mentioned below Table 1.

Table-1. Format of confusion matrix.

	Normal	Abnormal
Normal	TP	FN
Abnormal	FP	TN

From the extracted features SVM is trained with N normal and cancer images. SVM classifier here undergoes testing and training that gathers the features extracted.

5.2 Accuracy with histogram of oriented gradients

This is a dense model which extracts feature data with in every position of an image. In this algorithm initially gradients will be computed for divided blocks of an mammogram image. In this extraction method SVM classifier is trained with a sample 75 normal and cancer training examples here 25 images will be tested among the 25 normal images and SVM classified 22 images were identified as normal and 24 as cancer mentioned in Table-2.

Table-2. HoG-SVM results.

SVM	Training	Testing	detected
Normal	75	25	22
Cancer	75	25	24

5.3 Accuracy with Grey Level Co-occurrence Matrix (GLCM)

In this method texture features will be calculated by five statistical measures such as correlation, energy, homogeneity, sum of square variance, entropy. GLCM size can be estimated by considering a number of grey levels in an image. In this method elements in grey level matrix indicates the frequency of a pixel distance in an image with its neighbourhood pixel. The classifications of GLCM-SVM are presented in Table-3.

Correlation: It measures the probability occurrences of an image pixel point.

Energy: It can be defined as a sum of squared pixel pair elements within grey level matrix.

Homogeneity: It can be defined as finding the distribution of elements in a grey level matrix with respect to diagonals.

Sum of square variance: this method assigns high weight on the elements that differ from the average value of p (i, j).

Entropy: measures the randomness of a pixel element.

Table-3. GLCM-SVM classification results.

Texture features	Normal training	Normal testing	Detected classification ratio (%)	Abnormal training	Abnormal testing	Detected classification ratio (%)
Correlation	75	25	89.1	75	25	88.5
Energy	75	25	78.1	75	25	77.1
Homogeneity	75	25	81.3	75	25	80.4
Sum of square variance	75	25	87.2	75	25	86.2
entropy	75	25	89.5	75	25	89.3

5.4 Accuracy with Gabor filters

In this method feature extraction can be done by Gabor cancer detection algorithm. After Gabor feature extraction at a final stage probability can be considered of being nearest cluster that belongs to cancer by specifying factor such as threshold. In Gabor wavelets case fuzzy set works best than other classifiers such as SVM in the stage of classification can be compared by accuracy factor w .In the case of Gabor wavelets for reducing here image dimensionality principle component analysis were used important thing about this wavelet was that here it even reduces the frequency domain of an image pixel.

The comparative results of LBP-SVM, HoG-SVM, GLCM-SVM, and Gabor-SVM are presented in following Table-5. It is noted that, GLCM texture method support more for SVM mammogram, classification than other method. From Figure-2, it is clear that GLCM+SVM perform as best than other hybrid mechanisms.

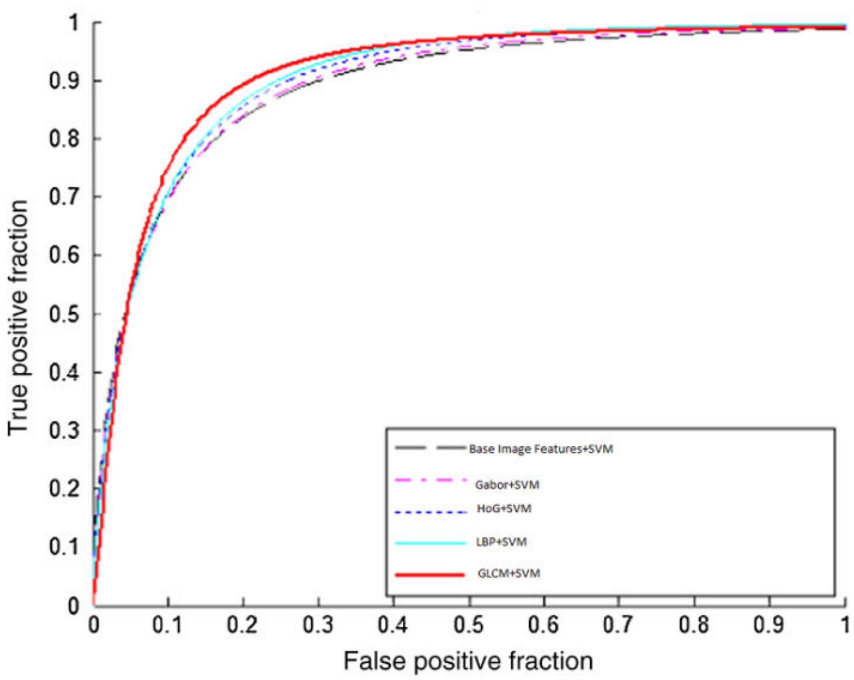

Figure-2. ROC curve for texture features.

Table-4. Comparative results for mammogram classification.

	Accuracy	Sensitivity	Specificity	Precision
LBP_SVM	89.2%	0.88	0.89	0.90
HoG_SVM	84%	0.83	0.87	0.88
GLCM_SVM	92%	0.94	0.93	0.95
GABOR_SVM	85%	0.86	0.76	0.85

CONCLUSIONS

Effective mammogram classification is an immediate requirement for radiologist to find the breast cancer detection at earlier stages for women. Four texture methods are applied on mammogram images for deriving of texture features and the technique of SVM classification is used for classification of normal and abnormal mammogram images. From the study of experimental results, this paper suggests the hybrid model of GLCM-SVM to radiologist for accurate mammogram classification.

REFERENCES

[1] Shen-Chuan Tai, Zih-Siou Chen, and Wei-Ting Tsai. 2014. An Automatic Mass Detection System in Mammograms Based on Complex Texture Features. IEEE Journal of Biomedical and Health Informatics. 18(2).

[2] P. Janani, J. Premaladha and K. S. Ravichandran. 2015. Image Enhancement Techniques: A Study. Indian Journal of Science and Technology. 8(22).

[3] M. Bitenc, D. S. Kieffer, K. Khoshelham. 2015. Evaluation of Wavelet Denoising Methods for Small-Scale Joint Roughness Estimation Using Terrestrial Laser Scanning. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. II-3/W5.

[4] F. Godtliebsen, E. Spjotvoll, and J. S. Marron. 1997. A nonlinear Gaussian filter applied to images with discontinuities. J. Nonparametr. Statist. 8(1): 21-43.

[5] S. M. Astley. 2004. Computer-based detection and prompting of mammographic abnormalities. Brit. J. Radiol. 77(suppl 2): S194- S200.

[6] K. Wongsritong, K. Kittayaruasiriwat, F. Cheevasuvit, K. Dejhan and A. Somboonkaew. 1998. Contrast enhancement using multipeak histogram equalization with brightness preserving.in Proc. IEEE Asia-Pacific Conf. Circuits and Systems. pp. 455-458.

[7] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller,M. P. Braeuning, and S. M. Pizer. 1998. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. J. Digit. Imag. 11(4): 193-200.

[8] N. Petrick, H.-P. Chan, D. Wei, B. Sahiner, M. A. Helvie and D. D. Adler. 1996. Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification. Med. Phys. 23(10): 1685-1696.

[9] Egidio Ieno Junior; Luis Manuel Garces; Alejandro Jose Cabrera; Tales Cleber Pimenta. 2016. Performance Analysis of Algorithms over FPGA for Removing Salt and Pepper Noise. IEEE Latin America Transactions. 14(5): 2120-2127.

[10] Xiaofei He, Deng Cai, Shuicheng Yan, Hong-Jiang Zhang. 2005. Neighborhood preserving embedding. Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference.

[11] H. D. Cheng, X. J. Shi, R. Min, L. M. Hu, X. P. Cai, and H. N. Du. 2006. Approaches for automated detection and classification of masses in mammograms. Pattern Recognit. 39: 646-668.

[12] R. M. Haralick, K. Shanmugan, and I. Dinstein. 1973. Textural features for image classification. IEEE Trans. Syst., Man, Cybern. SMC-3(6): 610-622.

[13] G. N. Lee, T. Hara, and H. Fujita. 2006. Classifying masses as benign or malignant based on co-occurrence matrix textures: A comparison study of different grey-level quantizations.in Proc. 8th Int. Workshop on Digital Mammography, S. M. Astley, Ed. et al., Berlin-Heidelberg, Germany. Vol. 4046, LNCS, pp. 332-339.

[14] C. J. Rose and C. J. Taylor. 2003. A statistical model of texture for medical image synthesis and analysis. Med. Image Understand. Anal. pp. 1-4.

[15] http://www.mammoimage.org/databases.

[16] Jalalian A., Mashohor S. B., Mahmud H. R., Saripan M. I. B., Ramli A. R. B. & Karasfi B. 2013. Computer-aided detection/diagnosis of breast cancer in mammography and ultrasound: a review. Clinical imaging. 37(3): 420-426.

[17] Ramirez-Villegas J. F. & Ramirez-Moreno D. F. 2012. Wavelet packet energy, Tsallis entropy and statistical parameterization for support vector-based and neural-based classification of mammographic regions. Neurocomputing. 77(1): 82-100.

[18] Sharma S. & Khanna P. 2015. Computer-Aided Diagnosis of Malignant Mammograms using Zernike Moments and SVM. Journal of digital imaging. 28(1): 77-90.

[19] N. F. Boyd *et al*. 2010. Breast tissue composition and susceptibility to breast cancer. J. Nat. Cancer Inst. 102: 1224-1237.

[20] L. Häberle *et al*. 2012. Characterizing mammographic images by using generic texture features. Breast Cancer Res. 14(2): R59.

[21] R. J. Ferrari, R. M. Rangayyan, R. A. Borges, and A. F. Frère. 2004. Segmentation of the fibro-glandular disc in mammograms using Gaussian mixture modelling. Med. Biol. Eng. Comput. 42: 378-387.

[22] A. P. Dhawan *et al*. 1996. Analysis of mammographic microcalcifications using gray-level image structure features. IEEE Trans. Med. Imag. 15(3): 246-259.

[23] H. Soltanian-Zadeh*et al.* 2004. Comparison of multiwavelet, wavelet, haralick, and shape features for microcalcification classification in mammograms. Pattern Recog. 37(10): 1973-1986.

[24] E. A. Sickles. 1986. Breast calcifications: Mammographic evaluation. Radiology. 160(2): 289-293.

[25] M. Elter and A. Horsch. 2009. CADx of mammographic masses and clustered microcalcifications: A review. Med. Phys. 36(6): 2052-2068.

[26] M. Lanyi. 1985. Morphologic analysis of microcalcifications. in Early Breast Cancer, J. Zander and J. Baltzer, Eds. Berlin, Germany: Springer. pp. 113-135.

[27] M. Suleman Basha, B Rama Subbaiah, K Rajendra Prasad. 2014. An Approach for Speech Recognition Technique is published in Articles of i. manager's journal on Pattern Recognition. 1(2).

[28] M. Suleman Basha, B Rama Subbaiah. 2014. A Review of Biometric Identification in Signal Processing is published in the Journal of International Journal of Engineering Research. Vol. 3, ISSN: 2319-6890,pp. 22-25.