Early stage Detection & Classification of Micro Calcification Clusters in the Digital Mammograms Using Digital Image Processing & SVM classifier

Sagar S. Tawani1, Dr. Ajay A. Gurjar2
1, 2Department of Electronics & Telecommunication, Sipna C.O.E.T., Amravati.

ABSTRACT:
One of the most widely renowned causes for higher death rates among women is Breast cancer. Early detection and care are of extreme importance for lowering the death rate due to breast cancer. Recent developments in digital mammography imaging systems intend to provide a better diagnosis of breast abnormalities. By early detecting the presence of micro-calcification in the mammograms we can diagnose the breast cancer. The proposed research paper is split into three stages to provide a system for classification between cancerous and noncancerous cases. The principle stage is the Segmentation Process, which applies threshold filters to separate the abnormal objects (foreground) from the breast tissue (background). The further stage in this research work is extraction of various features. This stage makes use of the segmented ROI images to extract characteristic features that may help in identifying abnormalities in mammograms. The various features like Histogram, Gray Level Co-occurrence matrix (GLCM) & Color Dominant are extracted from the distinct mammographic images. The final & concluding stage is classification, in which machine learning is used to distinguish the cancerous & noncancerous images. The proposed method have significantly high Sensitivity, Specificity and Accuracy for detecting microcalcifications for the database of 40 images which are taken from mammography unit of Rotary midtown Club Amravati.

Keywords: Breast cancer, Microcalcifications, Mammography, Histogram, GLCM, Color Dominant

INTRODUCTION:

India is encountering an uncommon ascent in the quantity of breast cancer growth cases over all segments of society, as are additionally different nations. Breast cancer is the most frequent cancer in women everywhere in India and has statistics of 25% to 31% of the entire cancers in women in cities of India. We can analyze that an age swing, that is the normal age of developing breast cancer has changed from 50-70 years to 30-50 years; and cancers in the youthful ages are seen to be more aggressive. Late reports note the nation that is home to 17 percent of the total populace has started to encounter higher paces of breast cancer. At present, India reports around 100,000 new cases every year. As per an investigation by International Agency for Research on Cancer (IARC), there will be around 250,000 new instances of breast cancer growth in India by 2020. Microcalcifications which are available in gatherings or groups have huge significance in clinical sciences. Microcalcification is a little affidavit of calcium which is collected in the tissues of breast, which is showed up as a bright spot enclosed inside a non-uniform bunches of microcalcifications these are the essential discoveries of quality of Breast Cancer in early stages and are ordinarily found in half of the mammograms which are analyzed by the radiologist. It might conceivable now and then radiologist may pass up a great opportunity to spot them. The best substitute to help radiologist as a subsequent reader to peruse different mammograms is a CAD System so as to avoid errors in the detection. The primary target of this paper is the characterization and recognizable proof of the different groups of microcalcification in the mammographic images to analyze cancerous part. In the proposed paper we have present another approach for recognizing the cancerous images, so radiologist will improve their interpretation results, as it is very difficult to find out abnormalities in dense breast tissue. The proposed method shows a new approach which includes image preprocessing, Segmentation, Feature extraction and classification of cancerous & non cancerous mammographic images. [7][8][11]

SYSTEM OVERVIEW

The Proposed System is implemented in MATLAB environment which has user friendly interface. The main stages of proposed system includes Image Pre-processing, Image Segmentation, Feature Extraction and finally classification using SVM classifier. In the proposed system the database from mammography unit of Rotary Midtown Club Amravati is used for experimental results. Figure 1 show the proposed system which is being implemented in order to detect & classify any abnormalities in the breast image.

1. Image Database

The proposed system is tested and implemented on the various mammographic images which are taken through the mammography unit of Rotary Midtown club, Amravati; it has a large database from the various regions of Vidarbha and Maharashtra. The database images are of 3540 x 4740 pixels, which is in an 8-bit format. In order to avoid the processing time for such a high quality images firstly we have convert these images into grayscale PNG images. We have tested our system on 40 Mammograms out of which 20 belongs to cancerous and rest of 20 are non cancerous.

2. Image Preporcessing

This step involves the cropping of breast image, filling of empty rows and columns, deleting background in order to remove noise or artifacts which are present in the mammograms. Distinct morphological operations & median filtering is done in order to enhance the visibility of an image. By performing the pre-processing step we can simply avoid number of false positive results which can be misinterpret by radiologist due to the presence of background noise. [4]

3. Image Segmentation

Segmentation plays a vital role in the image processing techniques. Without proper segmentation one cannot detect the exact presence of microcalcifications for the denser breast images, where we have to make accurate differentiation between abnormalities and the breast muscles. Basically Image segmentation is nothing but the partition of an image into some constituent components to get suspicious region. In proposed system, a novel wavelet based segmentation process is done which includes thresholding. Wavelets based image segmentation provides a novel approach which works on each and every pixel of the image and produced the number of slices in all three directions, so that segmentation is accurately done on every mammographic image. By selecting proper threshold the slices which are generated shows how the central pixel is varied in accordance with neighborhood pixel in all three directions. The basic steps for global threshold segmentation are 1) Select a proper threshold value for the image. 2) In order to convert the enhanced image into binary apply threshold to the image. 3) If at any pixel value is over the threshold value it is considered as foreground or abnormal otherwise background or breast tissue. After Segmentation various regions which are segmented,

4. Feature Extraction

After Segmentation the various patches (ROI) of mammographic images are identified and several features like Histogram, Color Dominant and Gray Level Co-occurrence Matrix (GLCM) are extracted. In GLCM four second order parameters Contrast, Correlation, Energy and Homogeneity are computed. Energy measure the smoothness of image. Contrast measures the local level variations which show the lower values for low contrast image and high values for high contrast image. Homogeneity measures how closely the distribution of elements in the GLCM occurs, it is normally occur in the range of 0 to 1. Entropy measures the randomness. After calculating all these values finally we form a data vector which is then given as an input to SVM classifier which classifies the cancerous and non cancerous images accurately.

5. Classification

After getting the dataset from the feature extraction step the Support Vector Machine (SVM) is trained with these complete data vector and the classification of various benign and malignant lesions is performed. Various SVM classifiers are applied out of which fitcsvm function has highest accuracy. The SVM with RBF kernel is employed in proposed method. The various advantages of using RBF kernel over others are: Non-linearly separable classification problems can be solved, their stability is good, and it has less parameter to calculate. Gaussian RBF kernel of tuning parameter σ:

$$k(xi, xj) = \exp\frac{(xi - xj)}{2\sigma^2}$$

5. Output of Proposed System

The output of a proposed system is represented through the Graphical User Interface (GUI) which is shown in Figure 3 and the novel method for segmentation is shown in Figure 2.

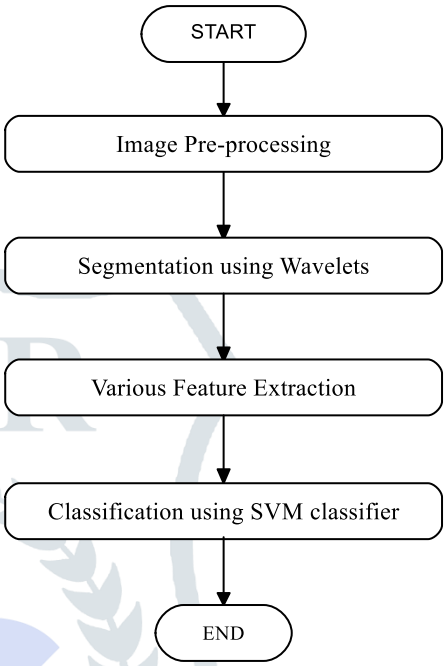

Fig.1. Overview of Proposed System

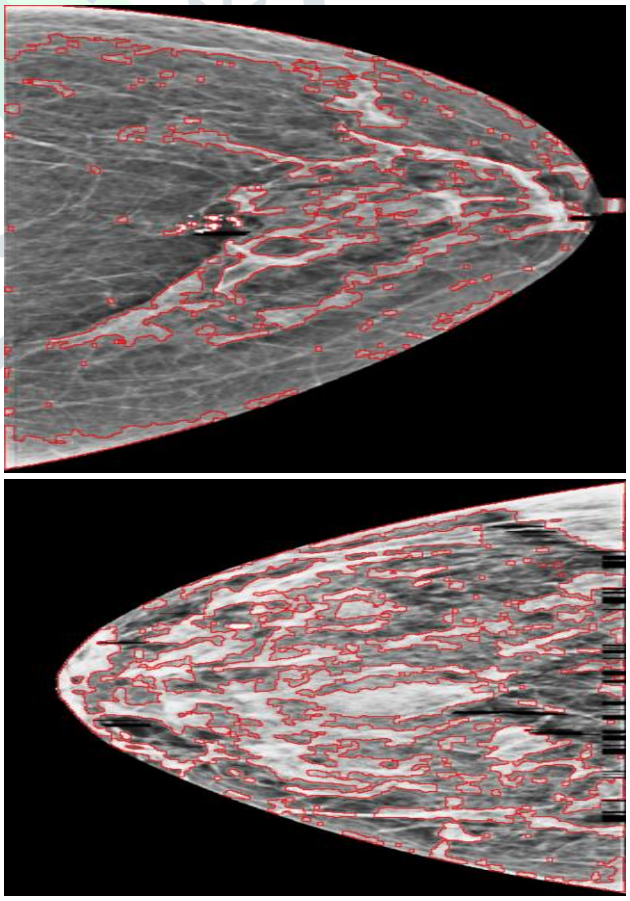

Fig.2. Segmentation of various Mammographic images using proposed method

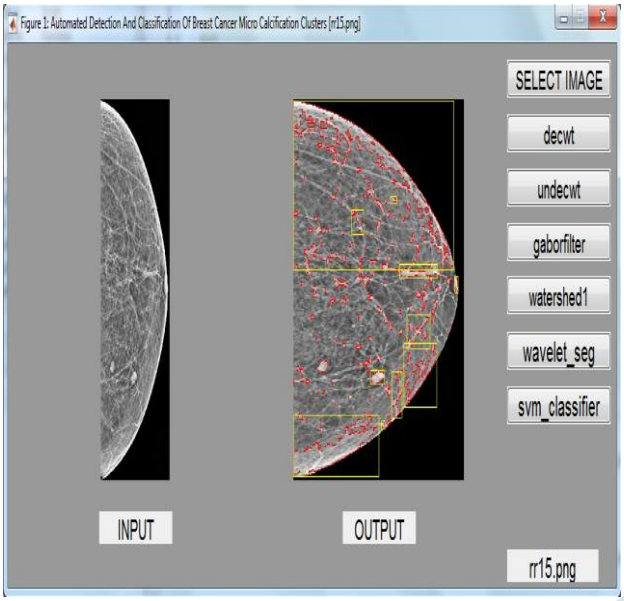

Fig.3. Final Output showing suspicious regions by yellow patches

FEATURE EXTRACTION

Feature extraction is a technique for catching visual substance of mammograms for retrieving and indexing. It is utilized to mean a snippet of data which is significant for comprehending the computational errand identified with a specific application.

1. Color Histogram:

Histogram is basically a illustration of the colors distribution in an image. It is an approximation of the probability allocation of a continuous variable. The density of data can be plot by the Histogram. The color histogram consists of two axes where x axis represents the intensity values ie 0 – 255 in 8 bit image while y axis represents the number of pixels having same intensity levels in an image. If we taken 'n' be the total number of observations and 'k' is the total number of bins, the histogram mi will meets the following conditions:

$$n = \sum_{i=1}^{k} \text{mi}$$

2. Color Dominant:

All in all, colour is one of the most predominant and recognizable low-level visual highlights in portraying image. The R, G, B colour model describes colors with three components depicting the intensities of red, green and blue lights. This model is closely related to the fact that the human eye contains three kinds of receptors that have different sensitivities to specific wavelength ranges. As a result, almost any colour perception can be faked with a combination of three suitably selected wavelengths at suitable intensities.

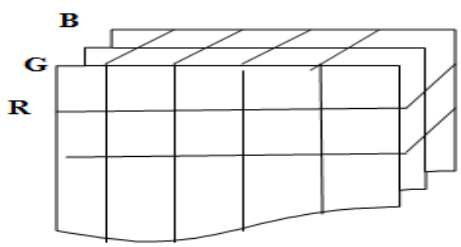

Fig.4. Coarse Division of RGB in 8 partitions

3. Gray Level Co-occurrence Matrix (GLCM):

The GLCM is an organization of how regularly various blends of pixel brightness grey levels present in a image. It comprises of two stages for extraction of features. First is to compute the co-occurrence matrix and second step is to calculate texture Features based on the co-occurrence matrix. Components in the matrix are processed by the equation:

$$P(i,j|d,\theta) = \frac{P(i,j|d,\theta)}{\sum_i \sum_j P(i,j|d,\theta)}$$

GLCM represents the texture feature according to the correlation of the pixels with its neighborhood pixels. In this paper, Four second order statistics will be taken into considerations which are Energy, Contrast, Homogeneity and Entropy in order to define the texture feature.

$$\text{Energy } E = \sum_x \sum_y P(x, y)\, 2$$

$$\text{Contrast } I = \sum_x \sum_y (x-y)\, 2\, P(x, y)$$

$$\text{Correlation} = \sum \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} i.j\, Pi\, j - \mu x \mu y}{\sigma x \sigma y}$$

$$\text{Homogeneity } H = \sum_{i,j} (p(i, j) / 1 + \xi i\, j \xi)$$

Features of texture are computed when distance = 1 and direction =0°, 45°, 90°, 135°. Four texture features are calculated in each direction.

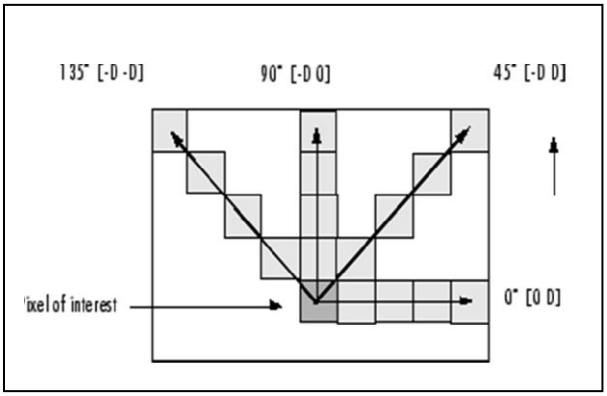

Fig.5. Directional Analysis of GLCM

Contrast		Correlation		Energy		Homogeneity	
Cancerous	Non-Cancerous	Cancerous	Non-Cancerous	Cancerous	Non-Cancerous	Cancerous	Non-Cancerous
0.2151	0.1501	0.9775	0.9832	0.1325	0.2577	0.9185	0.9403
0.4729	0.2956	0.9506	0.9669	0.1151	0.2321	0.8704	0.9088
0.7208	0.4312	0.9247	0.9517	0.1067	0.2189	0.8397	0.8886
0.9462	0.5574	0.9012	0.9375	0.1016	0.2098	0.8179	0.8722
0.2963	0.2398	0.9690	0.9731	0.1234	0.2427	0.8917	0.9214
0.6188	0.4787	0.9352	0.9463	0.1081	0.2181	0.8467	0.8849
0.9090	0.6993	0.9048	0.9214	0.1010	0.2066	0.8168	0.8617
1.1654	0.8857	0.8780	0.9003	0.0965	0.2002	0.7966	0.8454
0.1858	0.1817	0.9806	0.9797	0.1336	0.2530	0.9325	0.9353
0.3799	0.3687	0.9604	0.9587	0.1160	0.2255	0.8767	0.8996
0.5719	0.5420	0.9403	0.9392	0.1083	0.2122	0.8482	0.8772
0.7592	0.6890	0.9207	0.9227	0.1035	0.2044	0.8280	0.8611
0.2856	0.2414	0.9702	0.9729	0.1242	0.2412	0.8990	0.9200
0.5906	0.4754	0.9382	0.9465	0.1089	0.2170	0.8489	0.8840
0.8654	0.6925	0.9094	0.9219	0.1022	0.2057	0.8201	0.8609
1.1113	0.8744	0.8837	0.9012	0.0977	0.1994	0.7996	0.8447

Table 1. GLCM Features Extracted Values

The Final values which are derived from the second order statistical parameters are shown in following Table 1. The values are derived for both Cancerous & Noncancerous images. These complete dataset along with histogram & color dominant values are combined in a vector form and given as a input to SVM classifier which classify it accurately. [4] [12] [13] [14]

RESULTS & DISCUSSION

The proposed system is tested on 40 images which are taken from the database of Rotary club Midtown Amravati. A complete database of over 200 images is available out of which 20 cancerous & 20 noncancerous images are taken into considerations. Primarily in preprocessing stage, mammographic images were preprocessed to remove noise and artifacts & also image cropping has been taken place. Again images are resized to reduce the processing & execution time. After preprocessing the image, automatic segmentation takes place which accurately segment the images in order to detect the microcalcifications. Now in the feature extraction process, Histogram, color dominant and second order statistical features are obtained by GLCM. The final values obtained from these statistical parameters are shown in Table I. Finally, the complete dataset are combine in a vector will be given as input to the SVM classifier which classify the cancerous & non-cancerous images with the very high accuracy of 99.67%. The performance of the classifier is evaluated with the help of a confusion matrix which summarizes the number of samples predicted correctly or incorrectly by the classifier.

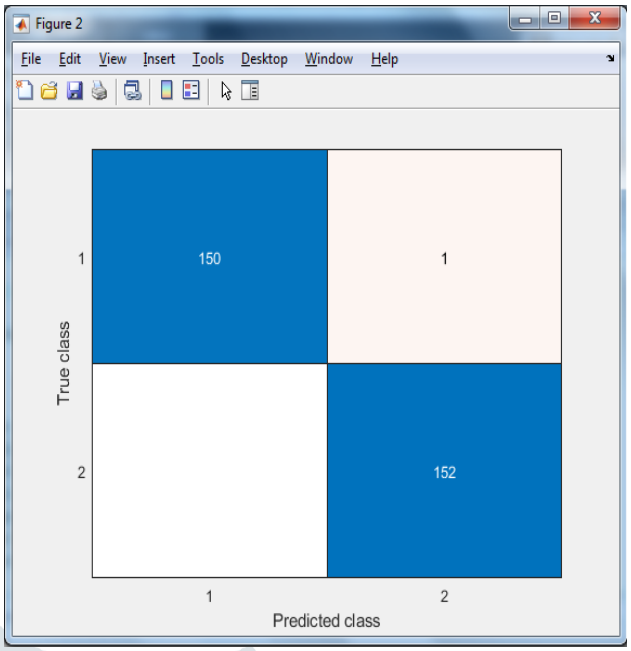

Fig.6. Confusion Matrix of Proposed System

CONCLUSION

From the distinct outcome present in the paper, we can conclude that the proposed method for differentiating cancerous and noncancerous clusters reduces the interpretation errors which are usually found in differentiation; additionally it is extremely proficient in correct detection in mammograms by enhancing resultant image visibility. Again by extracting the various features like Histogram, Color Dominant & Second order GLCM statistics along with SVM classifier gives accuracy of 99.67% is achieved on tested dataset of 40 images. Hence the proposed system improves the overall execution of the structure towards early detection and classification of microcalcifications in the mammographic images.

REFERENCES

[1] Heinlein P., Drexl J., Schneider W., Integrated waveletsfor enhancement of microcalcifications in digital mammography, IEEE Transactions on Medical Imaging, Vol. 22, No.3, pp. 402-413, 2003.

[2] A microcalcification detection system for digital mammography using the contourlet transform, Manzano-Lizcano J.A, Sanchez-Avila C and Moyano L (2004), Advances in Computational & Experimental Engineering & Science, pp.611-616.

[3] Wavelet based microcalcifications detection in digitized mammograms, Bouyahia S, Mbainaibeye J and Ellouze N (2009), ICGST-GVIP Journal, Vol.8, No.5, pp.23-31.

[4] Early Stage Breast Cancer Detection System using GLCM feature extraction and K-Nearest Neighbor (k-NN) on Mammography image, Htay & Maung, 18th International symposium on communications & information technologies, Thailand.

[5] Alarcon Aquino, Starostenko, Detection of microcalcifications in digital mammograms using dual tree complex wavelet transform, Vol. 17, No. 1, Engineering Intelligent System, March 2009.

[6] A Computer Aided diagnosis system for microcalcification cluster detection in digital mammogram, Balakumaran T and ILA Vennila

(2011),International Journal of Computer Applications, Vol.34, No.1, pp.39-45.

[7] American cancer Society, 2014a, Breast Cancer Facts & Figures.

[8] Statistics of Breast cancer in india: Global comparison cancerindia.org, ICMR, NICPR.

[9] Automatic Segmentation of Brain Tumors from MR images Using Undecimated wavelet transform and Gabor wavelets, ICECS 2010, G.Mirajkar, Barbadekar.

[10] Leborgne , R.A. (1953) The breast in Roentgen Diagnosis.

[11] Breast cancer in India - Global Comparison.

[12] Detection of Normal mammograms based on mammographic tissue density using GLCM features, Mona Elshinawy, Abdel Hameed, A journal of Biomedical Engineering, April 2011.

[13] Color and texture image retrieval using chromaticity histograms and wavelet frame, IEEE Transactions on Multimedia, 676–686, 2004, Liapis S, TziritasG.

[14] A histogram-based approach for object-based query-by-shape-and-color in image and video databases, Image and Vision Computing, Ediz Saykol, Ugur Gudukbay, Ozgur Ulusoy (2005) 1170–1180.

[15] MCs detection approach using Bagging and Boosting based twin support vector machine, Xinsheng Zhang, Xinbo Gao and Minghu Wang (2009), IEEE International Conference on Systems, Man and Cybernetics (SMC 2009), pp.5000-5505.

[16] J. Dheeba, N. Singh, S. Selvi, "Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach", Journal of biomedical informatics, Volume 49, June 2014, Pages 45-52.

[17] Wavelet transforms for detecting microcalcifications in mammograms IEEE Transactions Volume: 15, Issue: 2,April 1996 by R.N. Strickland, Hee II Hahn.

[18] Jinshan Tang, Rangaraj M. Rangayyan, Jun Xu, Issam El Naqa, Yongyi Yang,Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances IEEE Transactions On Information Technology In Biomedicine, Vol. 13, No. 2, March 2009.