An Approach to Classify Digital Mammogram Using GLCM Features

Pravin M. Palkar
Department of Electronics Engineering
Priyadarshini Bhagwati
College of Engineering, Nagpur, India
E-mail: p_palkar@yahoo.com

Dr Pankaj Agrawal
Department of Electronics Engineering
G.H. Raisoni Academy of
Engineering & Technology, Nagpur, India
E-mail: pankaj.agrawal@raisoni.net

Abstract

Breast cancer is the one of the leading cause of death in women. Early and accurate detection of breast cancer can reduce the death rate effectively. Presently, mammography is the most reliable and effective technique in detection of suspicious masses. In digital mammogram, masses are early signs of breast cancer. The computer aided diagnosis (CAD) tool helps the radiologist in detecting the suspicious mass effectively in digital mammogram. But due to the diversity in shape, size, ambiguous margins and poor contrast between the mass and surrounding bright structure, detection and classification of suspicious mass is challenging. This paper presents an approach to classify digitized mammogram from the detected suspicious mass using GLCM features. The GLCM features are extracted after preprocessing, segmentation and region of interest ROI (suspicious mass) detection of the digital mammogram. Preprocessing is preliminary stage used in mammogram image enhancement. It detects and removes the unwanted labels and signs present on the mammogram and extract the breast region. Once the breast region has been found, pectoral muscle and side black strips are detected and eliminated. Then segmentation process is carried out to detect the suspicious mass in digital mammogram. GLCM features are extracted from the detected mass. Analysis of these GLCM features can provide clues and can be used to classify the digital mammogram as benign or malignant.

Keywords: *Breast cancer, Mammogram, GLCM Parameters, Preprocessing, Segmentation, ROI.*

INTRODUCTION

Cancer is a disease that causes cells in the body to change and grow out of control. It is uncontrolled multiplication of a group of cells in a particular location of the body [1]. Mostly the cancer cells is a tumor and the cancer are named after the part of the body where the tumor originates [2]. Breast cancer begins in breast tissue. In the recent years the incidence of breast cancer has increased significantly and deaths occurred due to it is a recognized world health problem. It is one of the leading causes of fatality, with approximately 1 out of 12 women affected by the disease during their lifetime. In India, breast cancer is the second most common cancer in females after lung cancers and the death rate of one in eight women has been reported due to breast cancer [3].Cancer is the result of fast and uncontrolled development of normal cells. In the case of breast cancer, these grown cells are located in the breast tissue. Both the number of people getting breast cancer and the number of people killed by breast cancer are rising faster in the developing world than in developed nations. To reduce the death rate caused by breast cancer, effective method is to detect and treat it early so that it can be treated successfully. Presently, many imaging techniques to detect the breast cancer are available [4] and mammography is one of the most effective and efficient system to detect suspicious masses [5] recommended for breast cancer screening. However, mammography is not a perfect method, since the detection of suspicious mass is repetitive and fatigue task to radiologist. For every thousand cases analyzed by him, only 3 or 4 are cancerous and thus an abnormality may be overlooked. Many Computer-Aided Detection (CAD) systems have been developed to support and aid radiologist in detecting the suspicious masses that may indicate the presence of breast cancer. Such systems act only as a second reader and the final decision is made by the radiologist. Studies shown that such systems have improved the radiologist's breast cancer detection accuracy[6]. Computer-Aided Detection (CAD) systems use the digitized mammography image. The computer software then searches for the suspicious mass that may indicate the presence of breast cancer. In preprocessing stage of CAD system, these suspicious areas are highlighted and alert the radiologist for further analysis. The CAD tools works in three stages. In first stage, mammogram is preprocessed and the breast region is detected. It is the input to the second stage, where suspicious mass is detected and finds the ROI. The statistical features are then extracted from the ROI, which specifies some quantifiable property and significant characteristics of an detected mass. Third stage is classification stage based on the statistical features obtained in the previous stage. Any suspicious mass is classified further as benign or malignant. Benign is not cancerous. Benign tumors may grow larger but do not spread to other parts of the body. Malignant is cancerous. Malignant tumors can invade and destroy nearby tissue and spread to other parts of the body.

The objective of this paper is to present an effective approach to classify the digitized mammogram from the detected suspicious mass using GLCM features. All the stages are implemented using MATLAB. Preprocessing is preliminary stage used in mammogram image enhancement.[7]. It detects and removes the unwanted labels and signs present on the mammogram and extract the breast region. Once the breast region has been found, pectoral muscle and side black strips are detected and eliminated. Then segmentation process is carried out to detect the suspicious mass in digital mammogram [9].Then GLCM features are extracted, which can play very important role in classifying tumor or suspicious masses of mammograms as benign or malignant.

PREPROCESSING OF DIGITAL MAMMOGRAM

Digital mammograms are difficult to interpret directly, hence a preprocessing is needed in order to improve the quality of the mammogram and make the segmentation results more effective. The main objective of preprocessing is to improve the quality and make the mammogram ready for segmentation and feature extraction [6].As shown in Fig 1, noise observed in mammogram are high intensity rectangular label, low intensity label, tape artifacts etc [8].Preprocessing of mammogram is needed to detect and remove these noises.

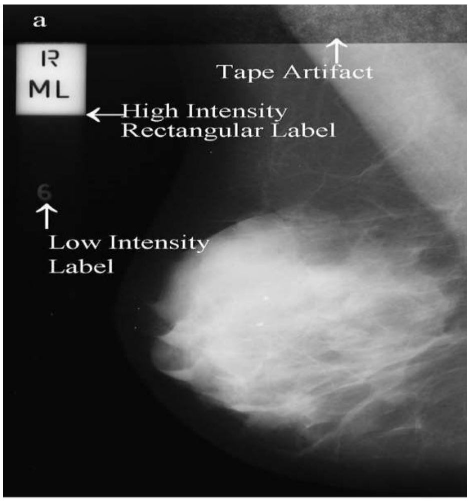

Fig. 1. *Mammogram with various noises*

In our proposed method, the mammogram samples are used from MIAS's database images [9].It includes both right and left breast images. To make processing simple and easier, we convert whole breast images to the left breast images .We considered only left sided mammogram and if we encounter any right sided mammogram, flip them in horizontal direction. In preprocessing we detect and remove the unwanted signs, labels and noises of the mammogram. First median filtering is performed to reduce and eliminate noise present in the mammogram image. Median filter preserves all useful information and also reduces the noise [10].It is nonlinear filter , efficient in removing salt and pepper noise and keeping the sharpness of the images [11].Median filtered mammogram is compared with threshold to get processed digital mammogram image [12]. Then regions with fewer pixels are removed by using the *bewareaopen* function in MATLAB. This process morphologically opens a binary image and removes all objects in the binary image, except the largest region (breast region). This operation uses an 8-connected neighbourhood. Next, a morphological operation reduce distortion and remove isolated pixels (individual 1's surrounded by 0's) is applied to the binary images using the *bwmorph* function in MATLAB with parameter 'clean'. These operations check all pixels and set a pixel to 1 if five or more pixel in its 3 by 3 neighbourhood are 1's, otherwise sets the pixel to 0. Hence got the binary image after removing all redundant regions and used it as a mask. Multiplied it with original gray scale image matrix to get gray scale image free from labels and noises as shown in Fig 2. After preprocessing, we observe that the mammogram is without any noise, signs and labels.

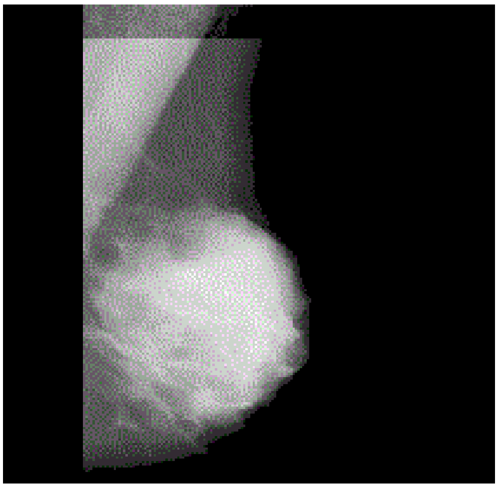

Fig. 2. *Pre-Processed Mammogram with pectoral muscle and background*

Then we have removed the redundant parts of the mammogram. The redundant parts are background parts, not covering the breast region other then right and left part of the mammogram. The background parts of mammogram are dark and their gray levels are near zero and do not have any difference with each other. Fig 2 shows this reality. After this, we detect and removed the pectoral muscle. As shown in Fig 2, the pectoral muscle is slightly brighter compared to the rest of the breast region. It must be removed so that detection of suspicious mass can be done efficiently. In left sided mammogram, it lies on the left top corner of the mammogram. We have to detect the position of pectoral muscle before removing it. The procedure for detecting the pectoral muscle is as shown in Fig 3 ; first get the straight line cut AB, we traced the mammogram from left side and found first non-zero column between the left background of the mammogram image and starting of actual part of breast region in the mammogram. Then determine the middle point C at the top margin of the mammogram and draw a straight CD from the point C to the lower left corner of the mammogram image. The line CD cross the line AB at point E. Observe that the right angled triangle CAE is our region of interest to detect the pectoral muscle. This

CAE region is cropped from the original mammogram image [13].To get the straight line cut FG, we traced the mammogram image from right side and found first non-zero column.

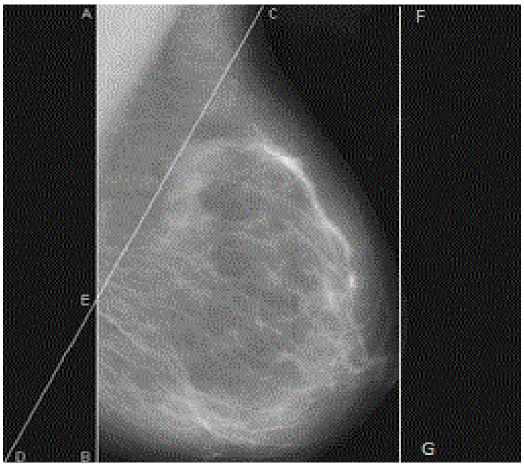

Fig 3. *Mammogram with pectoral muscle and background*

The above procedure successfully and efficiently removed the pectoral muscle, left hand and right hand black stripes from the mammogram image as shown in in Fig 4, with success ratio of 80 percent on different pairs of mammogram of different shapes and size [14].

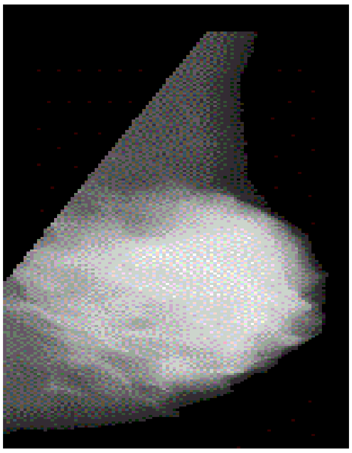

Fig 4. *Mammogram without pectoral muscle and background*

DETECTION OF MASS IN MAMMOGRAM

The information in the digital mammogram is present in the form of different intensities and textural variations. Our approach is to isolate the spatially interconnected structures in the image to form regions concentrated around prominent intensities. Hence after getting breast region as shown in Fig 4, we found prominent pixel to get threshold value for getting the spicules. By performing morphological operation, we filter the false spicules. After getting the intended spicules, we identified seed point seed point and grown that dynamically to get mass region as shown in Fig 5. We observed that suspicious mass is easily detected and size of ROI is dynamically defined with respect to the size of mass region. The GLCM features are extracted from this ROI of digital mammogram, to detect and classify the breast cancer [18] as benign or malignant.

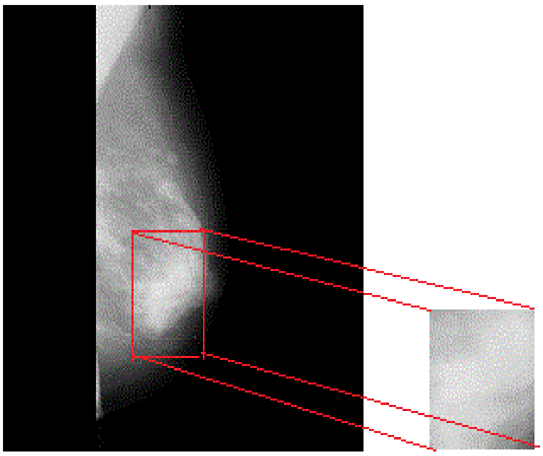

Fig 5. *Automatic Detection of Mass Region in Mammograms*

GLCM FEATURES EXTRACTION

The feature specifies some quantifiable property of an image and is computed such that it quantifies some significant characteristics of the image. Transforming the input data of the image into the set of features is called features extraction.The motive of feature extraction process is to represent the mammogram in the form to facilitate decision making process. The mammogram tissues are characterized by random and no homogeneous structures. Hence statistical features are preferred for the analysis of digital mammogram [15].

Statistical textures feature is a measure of the arrangement of intensities in a region. The texture features can be computed from the gray tones themselves and easier to compute. This approach is less intuitive but is computationally efficient and will be best for classification of textures [16].The extraction methods of GLCM features play very important role in detecting suspicious mass because of the nature of mammograms. GLCM texture features have been proven to be useful in differentiating masses of breast region as benign and malignant [17].

The GLCM is a tabulation representation of different combinations of pixel brightness values (gray levels) occur in an image. To calculate the second order texture features, GLCM is used in our proposed method. First order statistical texture measures are statistics calculated from the original image values like variance, mean, etc. These measures do not consider pixel neighbour relationships. The second order statistical measures unlike first order, consider the relationship between groups of two neighbouring pixels in the original image. Third and higher order statistical textures considering the relationships among three or more pixels are theoretically possible but not commonly implemented due to calculation time and interpretation difficulty. In our proposed method, we have extracted second order statistical features of GLCM. These features provide information about the texture of the mammogram. We have considered Contrast, Energy, Homogeneity and Correlation features of GLCM matrices. Contrast is referred as sum of square variance and is measure of local level variations and takes high gray level values for the mass of mammogram of high contrast. Energy is referred as uniformity and takes high gray level values for the mass of mammogram which has very similar pixels. Homogeneity is referred as inverse different moments and is measures of closeness of the distribution of elements in GLCM to the diagonal of GLCM and is high for uniform gray levels. Correlation measures the joint probability occurrence of the specified pixel pairs.

EXPERIMENT RESULTS

Experimental results of preprocessing stage shows that it effectively flip the right sided mammogram if occur, the median filtering efficiently removes unwanted objects and texts, good success ratio for detection and suppression of pectoral muscle and elimination of side stripes of mammogram. The mammogram preprocessing results obtained over samples of mammogram of MIAS database have shown good results. The segmentation results also show that the seed point detection and region growing method to determine the suspicious mass in digital mammograms is promising and the results show that the proposed algorithm achieved 90% true result. After extraction of features, the graphical representation of GLCM features with respect to different offset are very interesting and are sited in figure 6 and 7 respectively.

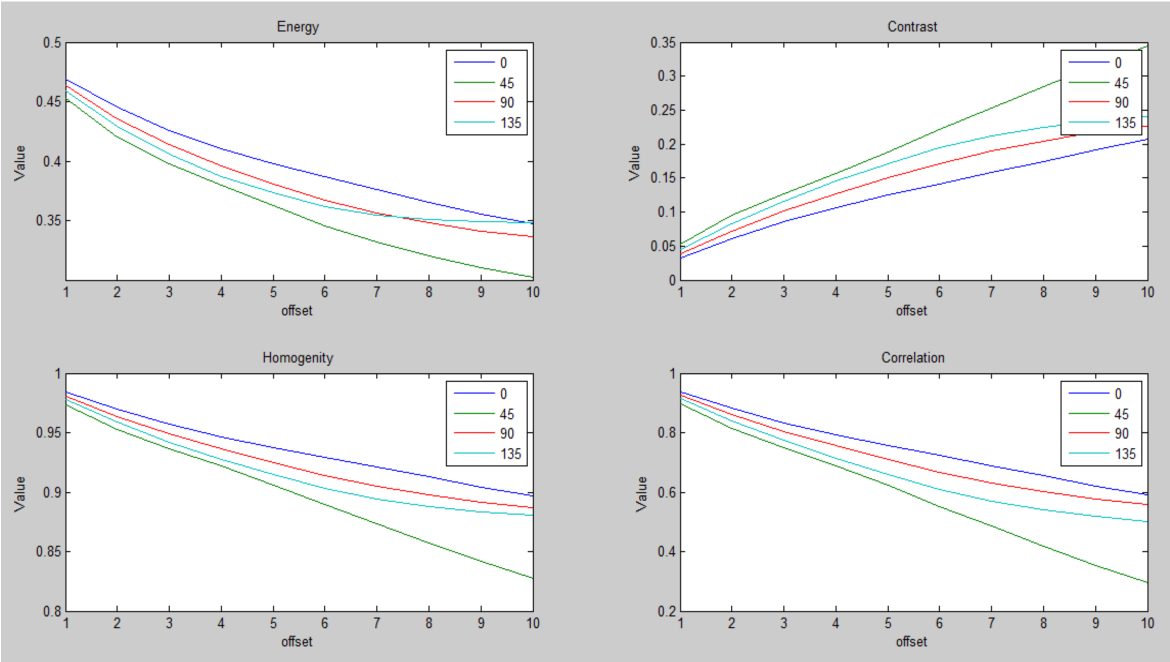

Fig 6. *Co-relations of GLCM features for Benign Mass*

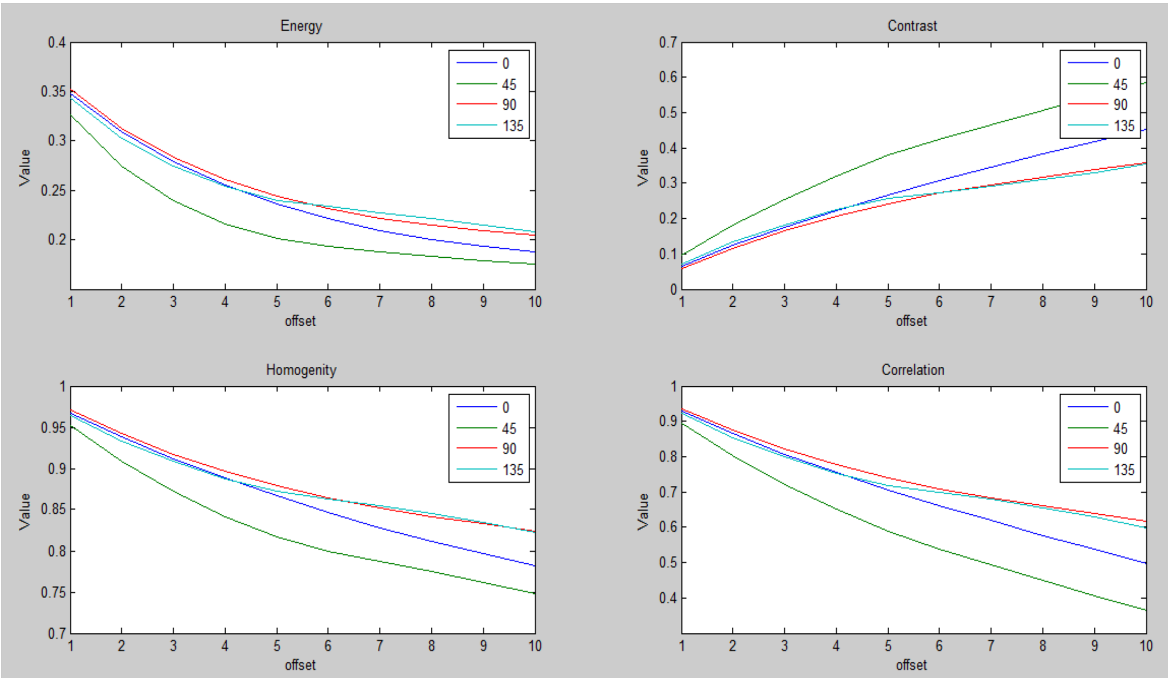

Fig 7. *Co-relations of GLCM features for Malignant Mass*

The graphical representation shows that the co-relations between GLCM features for benign and malignant mass are highly different as per our expectation and easily recognizable. We observed that intersection between the curves are less in benign compared to the curves of malignant. Multiple intersections are observed in malignant case. It shows effectiveness of the proposed method and the observations can be used in discrimination of the masses as benign and malignant.

CONCLUSION

The presented approach can be successfully used to detect the suspicious mass of a digital mammogram and to characterize the mass region based on co-relation of GLCM features. Investigation of the preliminary results obtained reveals that co-relation of GLCM features distinguish between benign and malignant. Thereafter, these could be used to define criteria of decision that will permit to distinguish a mammogram as benign or malignant. This presented approach can acts as a supporting tool to radiologists in detecting masses/tumors in mammograms.

REFERENCES

1. Karthikeyan Ganesan, U. Rajendra Acharya, Chua Kuang Chua, Lim Choo Min, K. Thomas Abraham, and Kwan- Hoong Ng: Computer-Aided Breast Cancer Detection Using Mammograms: A Review .Methodological Review IEEE REVIEWS IN BIOMEDICAL ENGINEERING, VOL. 6, 2013
2. Indra K Maitra, Sanjay Nag,Prof. Samir K Banyopadhyay :Identification of Abnormal Masses in Digital Mammography Images , International Journal of Computer Graphics ,Vol. 2,No. 1 May, 2011.
3. Sara Dehghani , Mashallah Abbasi Dezfooli :A Method For Improve Preprocessing Images Mammography, International Journal of Information and Education Technology, Vol. 1, No. 1, April 2011.
4. U. Bick, F. Diekmann:Digital mammography, Springer, Berlin, 2009.
5. H. Gholam:Registration of vibro-acoustography images and x ray mammography,Proceeding IEEE, 2005.
6. D.Narain Ponraj, M.Evangelin Jenifer: A Survey on the Preprocessing Techniques of Mammogram for the Detection of Breast Cancer, Journal of Emerging Trends in Computing and Information Sciences, Vol. 2, No. 12, December 2011.
7. Sharanya Padmanabhan and Raj i Sundararajan :Enhanced Accuracy of Breast Cancer Detection in Digital Mammograms using Wavelet Analysis , fN 47907, USA 978-1-4673-2322-2112/$31.00, 2012 IEEE
8. Pradeep N. , Girisha H. , Sreepathi B. And Karibasappa K.,"Feature Extraction Of Mammograms, " International Journal Of Bioinformatics Research ,ISSN: 0975–3087 & E-ISSN: 0975–9115 , Volume 4, Issue 1, 2012, pp.-241-244
9. Samir Kumar Bandyopadhyay: Detection of Abnormal Masses in Mammogram Images, International Journal of Computer Science and Information Technologies, Vol. 1 (5) , 2010.
10. J. Sharma, J. K. Rai and R. P. Tewari: Identification of pre-processing technique for enhancement of mammogram images, 2014 International Conference on Medical Imaging, m-Health and Emerging Communication Systems (MedCom), Greater Noida, pp. 115-119, 2014.
11. R. Ramani, Dr. N.Suthanthira Vanitha: The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images, MECS (http://www.mecs-press.org/), 5, 47-54, April 2013.
12. J. Nagi, S. Abdul Kareem, F. Nagi and S. Khaleel Ahmed: Automated breast profile segmentation for ROI detection using digital mammograms," 2010 IEEE EMBS Conference on Biomedical Engineering and Sciences (IECBES), Kuala Lumpur, 2010, pp. 87-92.
13. Prof. Samir Kumar Bandyopadhyay: Pre-processing of Mammogram Images, International Journal of Engineering Science and Technology, Vol. 2(11), 2010.
14. Pravin Palkar, Dr Pankaj Agrawal: Preprocessing of Mammograms for Computer-Aided Diagnosis of Breast Cancer International Journal of Modern Computer Science

(IJMCS),Volume 4, Issue 5,, ISSN: 2320-7868 (Online), October, 2016
15. L. Sheng, “The analysis of digital mammograms : speculated tumor detection and normal mammogram characterisation”, PhD. Thesis,Purdue University, 1999
16. Pravin Palkar, Dr Pankaj Agrawal:A Technique to Extract Statistical Parameters of Digital Mammogram to Detect Breast Cancer, International Journal of Advanced Research in Science, Engineering and Technology Vol. 3, Issue 12 , December 2016
17. Pravin Palkar, Dr Pankaj Agrawal:An Approach to Analyse Digital Mammogram Using Statistical Parameters, International Journal for Research in Applied Science & Engineering Technology (IJRASET) Volume 5 Issue I, January 2017.
18. Pravin Palkar, Dr Pankaj Agrawal:ROI Detection in Digital Mammograms for Detection of Breast Cancer, International Journal of Engineering Science and Computing, February 2017 Volume 7 Issue No. 2.

