

# Mastography Classification using 2 D and GLCM for Detection of Breast Cancer

Mahesh S. Kedare

PG Student, Department of Computer Science & Engineering  
PES Collage of Engineering  
Aurangabad.

Dr. V. B. Kamble

Department of Computer Science & Engineering  
PES Collage of Engineering  
Aurangabad.

**Abstract-** Breast cancer is found to be the most common form of cancer found in women which is the leading cause for cancer death worldwide. Detection of abnormality at the earliest increases the chances of successful treatment and can reduce the mortality rate. MRI is a widely used medical imaging technique. Noise in MRI negatively affects image processing and analysis works. The main objective of pre-processing stage is to improve the quality of image by removing the irrelevant noises and unwanted portions in the image so as to convert the image into some other representation that is more meaningful, thus making it easier to interpret the details in an image. In this proposed work various filtering algorithms are discussed and compared and an automated scheme for Magnetic Resonance Imaging (MRI) breast segmentation is proposed. It is found that there are several types of abnormalities in breast. Among those, signs of breast cancer are normally associated with asymmetry between images of left and right breasts. Other type of abnormalities related to breast tumours is presence of micro-calcifications in the breast, presence of masses in the breast and Architectural Distortion (AD). Architectural Distortion refers to, disruption of the normal arrangement of the tissue strands of the breast resulting in a radiating or haphazard pattern without an associated visible centre. Micro-calcifications (MC) are tiny deposits that range from 50 to several hundred microns in diameter, which usually appear in clusters. Masses are signs of breast cancer. Masses with speculated margins have a high likelihood of malignancy. Architectural distortion (AD) is the third most common mammographic finding of breast cancer. Literature informs that about 81% of spiculated mass and 48-60% of AD is malignant and it is estimated that 12-45% of cancers not found in mammographic screening are AD. The detection sensitivity of the current computer systems for v speculated mass and AD is not as effective as micro-calcification detection algorithms and thus there is a pressing need for improvements in their detection.

**Keywords :** Mammographic images, Malignant & Benign, Multiresolution Analysis, GLCM.

## INTRODUCTION

Breast cancer is the most common disease throughout the world and it has been said that out of eight women one is affected with the breast cancer. According to Globocan project it has been estimated that 1.62 Million new cancer case found and in India 1, 44,937 women affected with cancer and 70,000 women died so we can say that out of two one women is dying due cancer[1]. It can be recovered by using the early detection, screening etc [2]. Mammography is the most useful method for detection of

breast cancer, mammograms are nothing but the x-ray images of the breast [3].

Human interpretation of the breast cancer is done by using the training and experience but this method can give miss results so that a diagnosis is developed which is known as the Computer Aided Diagnosis (CAD). CAD is the most popular method which analyzes the image with the use of image processing. Biopsies given by the radiologists got failed 60 to 90 %, so this is the main reason behind the development of the CAD [4].

Computer Aided Diagnosis gives the radiologist clear picture of the images so he can accurately classify the tissue as benign, malignant or normal rather we can say normal or abnormal. CAD system is purely based on the feature selection and feature extraction technique and so many methods have been developed in order to classify the mammogram or to extract the mammograms.

This research paper will discuss the method developed by using DWT & GLCM matrix.

## LITERATURE SURVEY

This section reveals about the previous work which has been done related for improving the feature selection and extraction for CAD.

First method is the hybrid of the particle swarm optimization and genetic algorithm to optimize the feature set called as Genetic Swarm Optimization. Particle Swarm Optimization (PSO) is technique which can be trapped in local minima so to overcome difficulties genetic algorithm is used. First step is the to segment the image by using the expectation maximization algorithm then 78 gray level co occurrence matrix is generated and reduced by PSO , GA , and GSO technique. The feature set is released and then SVM classifier is applied which will show breast tissue as normal or abnormal. The performance of GA, PSO & GSO based SVM is compared by Receiver Operating Characteristic Curve (ROC). EM method characteristics breast tissues as clusters and compute maximum like hood are generated by mixture of Gaussian. EM algorithm estimate missing values of clusters. Feature extraction is to reduce original image into set of features. These features are selected by PSO, in this method each tissue is called as particle and a vector is selected for the fitness of the solution. It then characterized as gbest and pbest and five best values are given to GA whether features are called as chromosome and these are called as population. GSO optimization technique is uses for feature set optimization

and the particles are optimized based on their parents and lastly a statistics are made and advantage of this SVM is it classifies small training samples in high dimensional space [5].

Second method is combination of artificial neural network and second order statistics. As per the developers of this method second order statistics not been studied in depth and they also used neural network for the selection of feature. The first step in this to select images from the DDSM database which contains some normal, abnormal and some malignant images abnormal area is extracted by the considering the region of interest and their texture descriptors as well as statistical features which contains energy, homogeneity, and correlation of gray level values. The GLCM is created by means of certain parameters as well as feature frequency of certain pairs. For the feature selection they used ANN which will used for training and if the error occurs it find the weight of the all the network it is based on the back propagation algorithm with least mean square algorithm. [6].

Third method is detection of masses using classification using local seed region growing and spherical wavelet transform hybrid scheme. The purposed method consists of four steps , first one is the homographic filtering for the enhancement which will gives you the classification as the mass or non mass , next one is to find out the region of interest by means of local seed algorithm where each tissue is called as seed and by default some static parameters and extract the region of interest. The third step is to apply wavelet transform for the feature extraction which will give the multiple zoomed images of the abnormal regions and next one is feature selection which will consist two components first one mass or non mass classification and the other one is to distinguish as normal or malignant by using the support vector machine. ). The proposed scheme LSRG-SWT scheme achieved 96 % and 93.59 % accuracy in first component which is also known as mass or non mass classification and begin/malignant classification when it used k cross validation. The system achieves 94% and 91.67% accuracy in mass / non mass classification and begins/malignant classification respectively when they used I.U database for training set and MIAS database as test set with external validation [7].

#### PROPOSED SYSTEM

The proposed system approach consists of two main algorithms such as feature extraction and feature selection. The feature extraction algorithm concentrates on the texture point in the mammographic image utilizing 2D-DWT and GLCM in succession on region of interest (ROI) to find out the feature descriptors of each detail coefficient of 2-level DWT. In the feature selection algorithm, effective and significant features are selected and provided to the neural network for the classification of mammograms as normal, benign or malignant. There are five key steps in the proposed system approach:

1. Preprocessing
2. Feature Extraction
3. Feature Selection

## 1 Preprocessing

This section reveals image preprocessing, discrete wavelet transform & gray level cooccurrence matrix of the selected image.

### a) Region Of Interest(ROI) :

Image is composed of different types of noises, artifacts in their background. The object area contains pectoral muscle. Due to this image is not suitable for feature extraction therefore a cropping operation is performed for the removal of noises, this operation is performed by means of considering the center of area .The result of this operation gives you the image which is noise free.

### b) Multiresolution Analysis using 2 D – DWT :

In the multi resolution technique, the under laying texture of mammographic ROIs is analyzed by zooming in and out process. The discrete wavelet transform decomposes them a mammographic ROI into a number of sub-images in different resolution levels preserving the high and low frequency information. This property leads the wavelet to extract better texture information from the mammographic ROIs. Given a continuous, square integrable function  $f(x)$  its wavelet transform is calculated as the inner product of  $f$  and areal valued wavelet function  $(\psi x \delta \mathbb{P})$  and it is given by:

$$W[f(s, \partial)] = (f, \psi k_t^s) = \int_{-\infty}^{\infty} f, \psi k_t^s(x)$$

### c) Gray Level Cooccurrence Matrix (GLCM) :

The gray-level co-occurrence matrix (GLCM) is used to extract the texture in an image by doing the transition of gray level between two pixels. The GLCM gives a joint distribution of gray level pairs of neighboring pixels with in an image. The co occurrence matrix of the ROI issue in classification of types of breast tissues by extracting descriptors from the matrix.

There are two types of relationship been developed between two pixels one is reference pixel and other one neighbor pixel. Let  $q(I,J)$  be element and it is given by

$$\sum_{x=1}^M \cdot \sum_{Y=1}^N \cdot \begin{cases} 1 \\ 0 \text{ otherwise} \end{cases}$$

## 2 Feature Extraction

The outputted image contains little energy due to this energy the texture analysis of the image cannot be done .So for these three coefficient matrix found out i.e. horizontal, vertical and diagonal . For analysis of texture

patterns of each ROI, the following five texture descriptors such as energy, correlation, entropy, sum variance, and sum average are calculated. There is one algorithm provided for this which is explained above which will give you the correct GLCM matrix.

### 3. Feature Selection

This is the final stage where the image can be considered as the benign, normal or malignant. One major problem lies with the large number of features that is very difficult to determine which feature or combination of features achieves better classification accuracy rate. So it is very essential to find out feature which can identify the mammogram easily. To select feature T and F test are performed where T test provides great accuracy. In MIAS database 86.00% for three stages and for same parameters 88 % in DDSM database

This approach for feature selection faces some difficulties when it provided with large amount of data. So here we used first rank representation (FRR) algorithm, because it will remove repeated indices rather we can say that features in feature matrix.

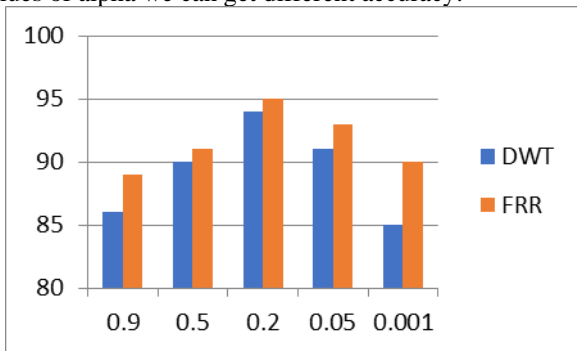
First rank representation gives 89 % accuracy in same set of database which we provide to discrete wavelet transform

#### PERFORMANCE ANALYSIS

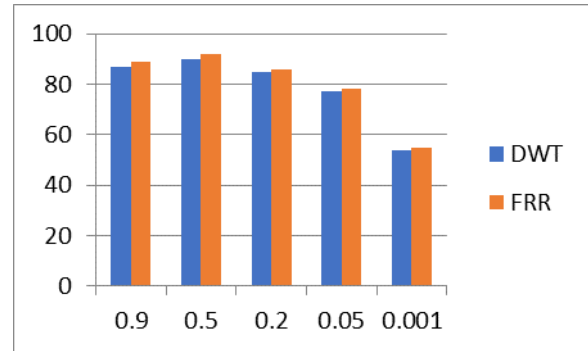
To validate all the methods like feature extraction & selection, algorithms are written in MATLAB environment. Mammographic images used for the validation of results which taken from MIAS (Mammographic Image Analysis Society) database.

MIAS database contains 322 images which are categorized as normal, abnormal, benign & malignant. Discrete wavelet transform & GLCM method gives 86 % accuracy in MIAS database so as first rank representation gives 89 % database.

The accuracy is based on the significance level (alpha), sensitivity & selected feature. By using different values of alpha we can get different accuracy.



The above graph shows accuracy of benign & malignant images



The above graph shows specificity of benign & malignant images

So from above we can say that Fixed Rank Representation gives better accuracy than the discrete wavelet transform

### CONCLUSION

In this paper, we proposed an efficient mammogram classification scheme to support the decision of radiologists. The scheme utilizes 2D-DWT and GLCM in succession to derive feature matrix from mammograms. To select the relevant features from the feature matrix, both t-test and F-test are applied. It is observed that t-test based relevant features achieve higher classification accuracy with BPNN than that of F-test. To validate the efficacy of the suggested scheme, simulation has been carried out using MIAS and DDSM databases. Its competent schemes are also simulated in the similar platform. Comparative analysis with respect to accuracy and AUC of ROC reveals that the suggested scheme outperforms other schemes. An accuracy of 86.0% and 88.2% has been obtained for normal-abnormal and benign-malignant respectively in MIAS database. The similar parameters are 88.8% and 89.4% achieved in DDSM database. Further, a training error comparison for the proposed scheme and random forest method is to evaluate the training convergence. The mean squared error is the average squared difference between output classes generated by the classifier and existing actual classes. The training error curves of two-sample t-test method show that it converges faster than other methods for both normal-abnormal and benign-malignant mammogram classes

### ACKNOWLEDGMENT

I express my gratitude towards my guide Prof. P. B. Bhalerao. Because of his guidance I have completed my work satisfactorily.

### REFERENCES

- [1] "Globocan Project 2012, International Agency for Research on Cancer (IARC), and World Health Organization: cancer fact sheets".
- [2] "L. Tabar, P. Dean, Mammography and breast cancer: the new era, Int. J. Gynecol. Obstet. 82 (3) (2003)".
- [3] "Mammogram classification using two dimensional discrete wavelet transforms and gray-level co-occurrence matrix for detection of breast cancer".
- [4] A. Dhawan et al., Analysis of mammographic micro calcifications using grey-level image structure features, IEEE Trans. Med. Imaging 15(3) (1996) 246-259..

- 
- [5] "J. Jona, N. Nagaveni, A hybrid swarm optimization approach for feature set reduction in digital mammograms, WSEAS Trans. Inf. Sci. Appl. 9 (2012)".
- [6] "M.A. Al Mutaz, S. Dress, N. Zaki, Detection of masses in digital mammogram using second order statistics and artificial neural network, Int. J. Comput. Sci. Inf. Technol. 3 (3) (2011)".
- [7] "P. Gorgel, A. Sertbas, O.N. Ucan, Mammographical mass detection and classification using local seed region growing-spherical wavelet transform (lsrg-swt) hybrid scheme,".
- [8] "Banshidhar Majhi, Ratnakar Dash Mammogram Classification using 2 dimensional discrete wavelet transform & gray level cooccurrence matrix for detection of breast cancer".
- [9] "T.M. Deserno, M. Soiron, J.E. de Oliveira, Texture patterns extracted from digitized mammograms of different BI-RADS classes, Image Retrieval in Medical Applications Project, release:".
- [10] "F. Albrechtsen, et al., Statistical texture measures computed from gray level cooccurrence matrices, Image Processing Laboratory, Department of Informatics, University of Oslo, vol. 20, 1995, pp. 1-14.".
- [11] "R.M. Haralick, K. Shanmugam, I.H. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. (6) (1973) 610-621.".
- [12] "A.S. Kurani, D.-H. Xu, J. Furst, D.S. Raicu, Co-occurrence matrices for volumetric data, in: 7th IASTED International Conference on Computer Graphics and Imaging, Kauai, USA, 2004, pp. 447-452..".