Automatic Detection of Tumor Subtype in Mammograms Based On GLCM and DWT Features Using SVM

M.Mohamed Fathima, PG Student
Department of Information Technology
National Engineering College
Kovilpatti, Tamilnadu

Dr.D.Manimegalai, M.E., Ph.D.,
Department of Information Technology
National Engineering College
Kovilpatti, Tamilnadu

Ms.S.Thaiyalnayaki, M.E.,
Department of Information Technology
National Engineering College
Kovilpatti, Tamilnadu

Abstract-Mammography images are employed in diagnosing breast cancers, since they are most effective, low cost and one of the highly sensitive techniques such that they can detect even small lesions. The proposed work increases the accuracy of classification and reduces the percentage of false positives. The images from the data set are initially preprocessed and contrast enhanced which makes the image most effective for further analysis. Then Region Of Interest (ROI) is determined from morphological top hat filtered image by means of thresholding segmentation. Various features like first order textural features, Gray Level Co-occurrence Matrix (GLCM) features, Discrete Wavelet Transform (DWT) features, run length features and higher order gradient features are derived for the particular ROI. Support Vector Machine (SVM) classifier is trained with the above mentioned features using MATLAB bioinformatics tool box. Thus the classified results are obtained for the query image based on the trained SVM structure. The mammography data set has been taken from the Mammographic Image Analysis Society (MIAS) in which there are 322 images available along with ground tooth information.

Index terms: Mammograms, Segmentation, SVM Classifier, GLCM & Run length features

I. INTRODUCTION

With the tremendous growth of medical field, the reason of the cancer is unknown. Therefore mammograms play a vital role in early diagnosis of breast cancer and gives promising results in controlling it. The nature of tissues in mammography images derives the main complexity in identifying the type of cancer. Usually, the X-ray component of a mammogram is required for breast cancer screening purposes. A lesion will usually appear brighter than the surrounding tissue on a mammogram. This is due to the fact that the area denser than fat will stop more x-ray photons.

Most of the researchers have chosen digital mammograms for their processing, since by using such images good contrast is achieved over dense breast tissue. Also the image acquisition is very fast and the patient is exposed to radiation only for small amount of time. Normally it is very tedious for the physicians and radiologists to analyze between a malignant and benign mass. Recent studies show that still there is about 10-25% of misclassification and false diagnosis. Hence it is inevitable to go for some automatic CAD approaches which do not involve manual intelligence. This study involves some novel classification approaches that resulted in good accuracy rates in classifying benign and malignant tumors. The results obtained were analyzed for its efficiency using some of the performance metrics like accuracy, sensitivity, specificity, precision and fitness value in the case of neural networks etc.

II. RELATED WORK

Abundant studies have been made on troubles in diagnosing breast cancer, based on digital mammograms. Region of interest (ROI) is selected in [2] by using constrained region growing segmentation methods by Alfonso et.al and using thresholding in [4] by Casico. In this case the number of ROIs increases as the density of breast tissue increases. Hadhoud (2005) has enhanced the image by mathematical morphology methods involving Top hat algorithm [3] using some structural elements like low pass Gaussian filters (LPGF) through which even very fine details were obtained. Morphological features such as Density, Shape, Margin, Abnormality assessment rank etc were extracted by Brijesh in [14] & [15] from the area extraction stage using Region Based Segmentation.

The Back Propagation Neural Network (BPNN) classifier used in [5] achieved classification rate of 83.30% for benign cases and 77.80% for malignant cases. The hidden layers and the weights associated with them can be varied to get optimum results. In [10] textural analysis based statistical descriptors such as averages, standard deviations and higher order statistics of intensity values were extracted as features. By selecting such features the percentage of classification rate was 79.31%.

III. DATASET COLLECTION

The mammography case samples required for the study has been taken from the Mammographic Image Analysis Society (MIAS). The Mammographic Image Analysis Society (MIAS) is an organization of UK research groups interested in the understanding of mammograms and has generated a database of digital mammograms. It contains 322 mammogram images of size 1024 × 1024 pixels with ground truth information about the abnormalities, i.e., type of cancer, severity of the diagnosis (Benign or Malignant), center coordinates of location of the abnormality and radius of the circle enclosing the abnormality.

IV. PROPOSED METHODOLOGY

The proposed method studies the presence of multiple features for discriminating breast tissues according to severity of cancer i.e. benign and malignant. Here we use morphological top hat filtering algorithm for improving the contrastness in the mammogram. ROI has been determined using region-based segmentation technique based on thresholding. The proposed method derives several features from the ROI which are essential for classification. For classification assignment SVM has been used which greatly reduces the misclassification rates. The proposed methodology has five major steps as shown in fig1

A. Preprocessing and Segmentation

It is expected for each mammogram to be preprocessed in order to reduce the error probabilities and increase the processing speed. In the preprocessing stage the operations performed are contrast enhancement, resizing such as cropping and scaling etc. First and foremost the relation ship given below has been used for the increasing the image contrast

$$f_{cont}(x,y)= \frac{f^2(x,y)*255}{\max\{f^2(x,y)\}}$$

Enhancement has the ability to determine the most unnoticed or hardly seen features of a mammogram. Mammogram is enhanced using Morphological top hat filtering algorithm with the help of disk shaped structuring element of size 12. The filtered image is processed with local thresholding algorithm for the detection of so called Region of Interest. The ROI obtained by means of the above mentioned procedure is clearly shown in fig 2 from its initial steps.

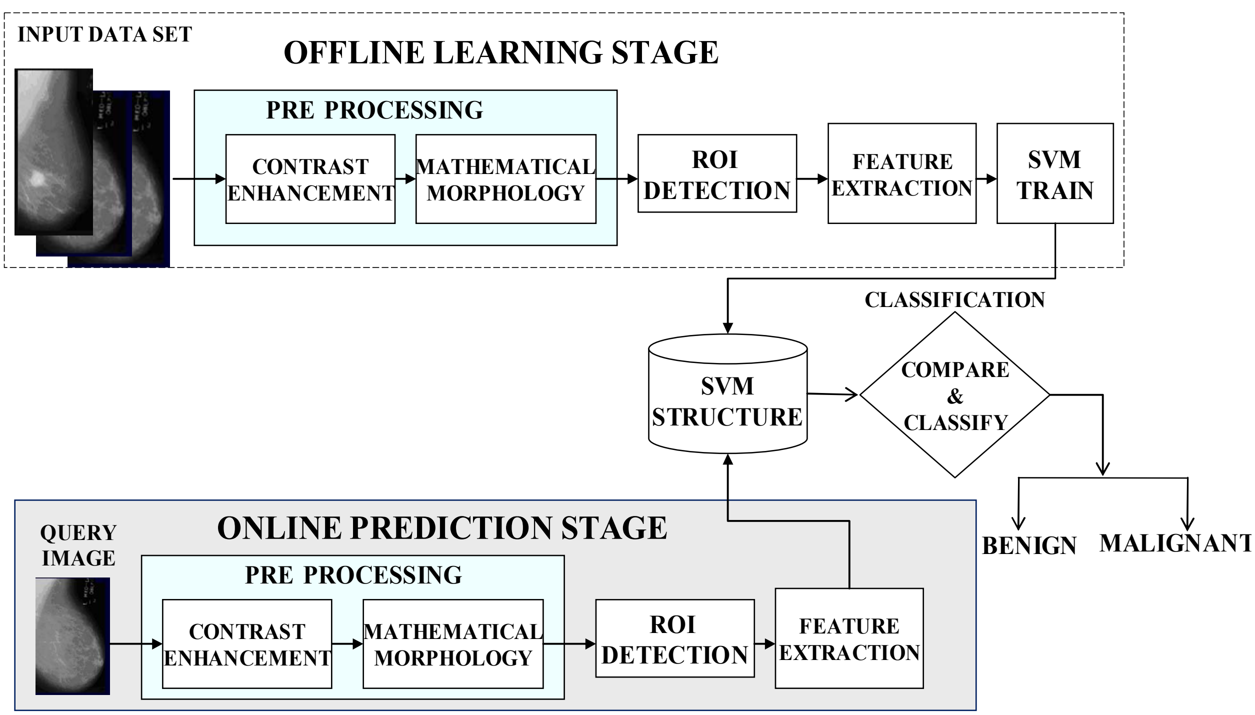
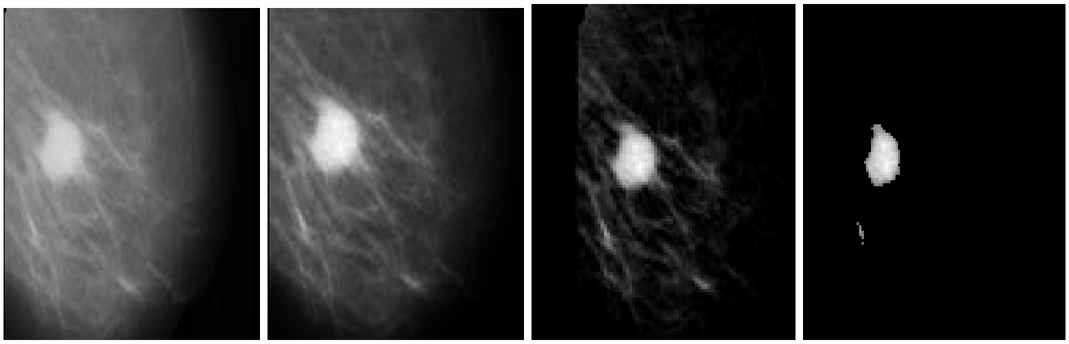


Fig1 Schematic flow diagram of the proposed system for mass classification on mammograms



(a) **(b)** **(c)** **(d)**

Fig 2 (a) mdb028 from data set, (b) Contrast Enhanced Image, (c) Morphological top hat filtered image, (d) ROI obtained from Segmentation

B. Feature Extraction

Transformation of input data i.e. the Region of Interest into the set of features necessary for classification has been derived. The features extracted must be carefully chosen, because it is expected to perform the desired classification task using this reduced representation instead of the complete ROI. Performance with ample amount of information generally requires a large amount of memory and computation power or a classification algorithm which over fits the training data. The features extracted are first order textural features, higher order gradient features, discrete wavelet transform features using daubechies, Coiflets and haar filters. The First order and gradient features for different types of classes such as normal, benign and malignant are compared in the table1. The run length features extracted are displayed with their corresponding values in table2. The gray level features energy, contrast, correlation and homogeneity which are obtained using gray level co-occurance matrix are given in table3.

TABLE 1 FIRST ORDER AND GRADIENT FEATURES

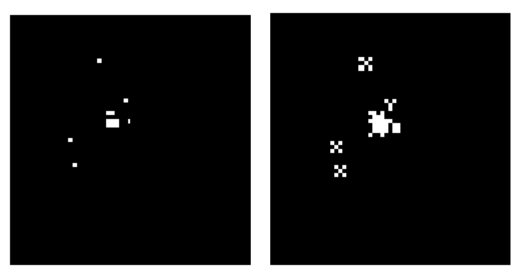
Features	Class		
	Normal	Benign	Malignant
Mean	948.28	106.94	794.92
Std Dev	5134.14	1684.25	5784.85
Skew ness	0.97	0.14	1.67
Kurtosis	4.08	1.42	1.19
Abs Std	1967.40	1062.68	1572.49
Slope	27228.6	20466.2	28090.8

TABLE 2 RUN LENGTH FEATURES

Run Length Features	Values
SRE1	0.43
LRE1	5527.69
GLN1	85.29
RP1	0.02
RLN1	88.72
LGRE1	0.47
HGRE1	81.59
Max run	129

TABLE 3 GLCM FEATURES

GLCM Features	Values
Energy	0.9662
Correlation	0.879
Contrast	0.178
Homogeneity	0.9668



(a) (b)

Fig 3 (a) Wavelet feature using Daubechies filter
(b) Wavelet feature using Coiflet filter

C. Classification

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates among a set of features having different class memberships. SVM is superior machine learning methodology with good accuracy in classification of high dimensional mammography datasets. Classification task in SVM involves with training and testing the image which consist of some specified features. Each case in the training set contains one 'target value' (class labels) and several 'attributes' (extracted features).First and foremost the statistical learning support vector machine is trained to create an SVM structure. Normal cases are selected for label 1; benign cases are selected for label 2 and malignant cases for label 3. In the testing phase the mass in the mammogram is classified by comparing the features with SVM structure. Hence SVM predicts the target value of data instances in the testing set more effectively. The response of SVM for a multi class solution i.e. benign, malignant, normal is shown in fig 3

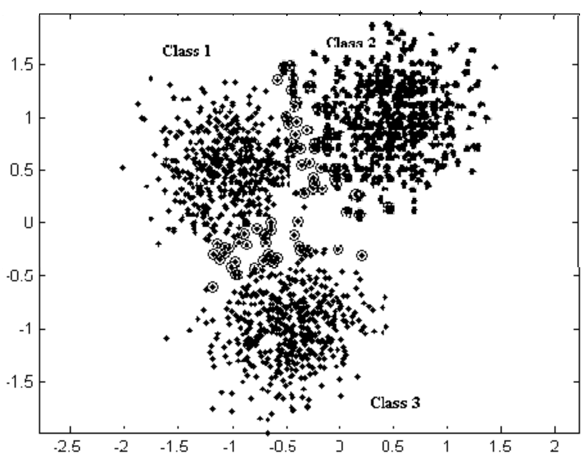


Fig 3 Multi class SVM Classification

VI. PERFORMANCE CRITERIA

In Mammography image classification techniques, the performance metrics to be determined are Accuracy, Sensitivity and Specificity. The sensitivity of a test is the proportion of people with have the disease who test positive for it. The specificity of a test is defined as the proportion of Victims without the disease who will test negative for it.

- True positive: Victims correctly diagnosed as sick (TP)
- False positive: Fit people incorrectly diagnosed as sick (FP)
- True negative: Fit people correctly diagnosed as healthy (TN)
- False negative: Victims incorrectly diagnosed as healthy. (FN)

Accuracy = (TP+TN)/(TP+TN+FP+FN)

TABLE 4. CONFUSION MATRIX

Classified Group	Actual Group		FPV	TPV
	Abnormal	Normal		
Abnormal	28(TP)	3(FN)	0.96	--
Normal	2(FP)	67(TN)	--	0.93

The classified results are compared using the confusion matrix for all possible cases in the mammograms. The True Positive Rate (TPV) or specificity in this case is 0.97 and the False Positive Rate obtained is 0.96. Another Performance metric, Precision=TP/(TP+FP) has been found as 93.3% and Sensitivity=TP/(TP+FN) has been found to be 90.92%.

V. CONCLUSION

Mammography is one of the finest methods in breast cancer recognition, but in a few cases radiologists face complexity in directing the tumors. The proposed methodology is completely automatic as it does not oblige human intervention and increases the accuracy in detection.. The segmentation algorithm applied here yields better results than region based techniques as well as global thresholding. There is no necessity for selecting the seed point to start segmentation process. The purpose of using SVM is to obtain the acceptable results in a fast and easy manner. The statistical risk of misclassification is minimized by maximizing the margin between the support vectors and the hyper plane. Accuracy has been obtained upto 95%. The main advantage of the proposed method is that the number of false positives has been reduced up to 1 for every 100 images. In future enhancements can be made by extracting shape features and by using clustering algorithms for segmenting the ROI.

VI. REFERENCES

[1] Ta. Cheng Chen, Tung. Chou Hsu, "A GAs Based Approach for Mining Breast Cancer Pattern" Elsevier, Expert Systems with Applications, 30, 674-681, 2006

[2] Alfonso Rojas Dominguez, Asoke K. Nandi, "Toward Brest Cancer Diagnosis Based on Automated Segmentation of Masses in Mammograms" Elsevier, Pattern Recognition, 42, 1138-1148, 2009

[3] M. Hhadhoud, M. Amin, W. Dabbour, "Detection of Breast Cancer Tumor Algorithm Using Mathematical Morphology and Wavelet Analysis" ICGST International Journal on Graphics, Vision and Image Processing(GVIP). pp. 75-80. 2006

[4] D. Cascio, F. Fauci, R. Magro, G. Raso, "Mammogram Segmentation by Contour Searching and Mass Lesions Classification With Neural Network" IEEE TRANSACTIONS ON NUCLEAR SCIENCE, VOL. 53, NO. 5, OCTOBER 2006

[5] Brijesh Verma, John Zakos, "A Computer-Aided Diagnosis System for Digital Mammograms Based on Fuzzy-Neural and Feature Extraction Techniques", IEEE Transactions on Information Technology in Biomedicine, Vol.5, No. 1, march 2001

[6] Osmar R. Za¨ıane, Maria-Luiza Antonie, Alexandru Coman, "Mammography Classification by an Association Rule-based Classifier" MDM/KDD2002: International Workshop on Multimedia Data Mining (with ACM SIGKDD 2002)

[7] Sumeet Dua , Harpreet Singh, H.W. Thompson, "Associative classification of mammograms using weighted rules", Expert Systems with Applications 36 (2009) 9250–9259

[8] Manrique, D., Rios, J., Rodriguez-Paton, A, "Evolutionary System for Automatically Constructing and Adapting Radial Basis Function Networks" Neurocomputing,69, 2268-2283,2006.

[9] B.N. Prathibha, Dr. V. Sadasivam, "An Analysis on Breast Tissue characterization in combined transform domain using Nearest Neighbor Classifiers" International Conference on Computer, Communication and Electrical Technology – ICCCET 2011, 18th & 19th March, 2011

[10] Zhang H, Foo SW. "Computer Aided Detection of Breast Masses from Digitized mammograms" IEICE. Transaction INF. &SYST. E89-D (6):1955-1961. 2006.

[11] Anna N. Karahaliou, Ioannis S. Boniatis, Spyros G. Skiadopoulos, "Breast Cancer Diagnosis: Analyzing Texture of Tissue Surrounding Microcalcifications"IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, VOL. 12, NO. 6, NOVEMBER 2008

12] Maurice Samulski and Nico Karssemeijer, "Optimizing Case-Based Detection Performance in a Multiview CAD System for Mammography" IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 30, NO. 4, APRIL 2011

[13] Brijesh Verma, PeterMcLeod, AlanKlevansky, "Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer", Expert Systems with Applications 37 (2010) 3344–3351

[14] Brijesh Verma, PeterMcLeod, AlanKlevansky, "A novel soft cluster neural network for the classification of suspicious areas in digital mammograms" Pattern Recognition 42 (2009) 1845 -- 1852

[15] Heang-Ping Chan, Jun Wei, Berkman Sahiner, Elizabeth A. Rafferty, Tao Wu, "Computer-aided Detection System for Breast Masses on Digital Tomosynthesis Mammograms" Radiology 2005; 237:1075–1080