Study of Mammogram Micro calcification to Aid tumour detection using Artificial Neural Network Based Classifier

B.Vijayalakshmi[1], R. Bhanumathi[2], G.R. Suresh[3]

PG Student, Dept. of CSE, Apollo Priyadarshanam Institute of Technology, Tamilnadu, India [1]

Assistant professor, Dept. of CSE, Apollo Priyadarshanam Institute of Technology, Tamilnadu, India [2]

Professor, Dept. of ECE, Easwari Engineering College, Tamilnadu, India [3]

ABSTRACT: In mammograms, a cluster of micro calcifications shows an early sign of breast cancer. An accurate and efficient classification scheme is needed for detection of tumour in mammograms. A Computer-Aided-Detection (CAD) system for automatic detection of tumour of micro calcification clusters on mammograms is discussed in this paper. The approach is very helpful for radiologist to diagnose the tumour and performs faster than typical screening programs. The images are collected from the Mammographic Image Analysis Society (MIAS) databases for implementation. The feature is extracted by Local Binary Pattern and classified using ANN. The proposed scheme provides high accuracy in classification of micro calcifications**.**

KEYWORDS: Artificial Neural Network, GLCM, Local Binary Pattern, Micro calcifications, Texture.

I.INTRODUCTION

For years cancer has been one of the biggest threats in human life, deaths caused by cancer are expected to increase in the future with an estimated 12 million people dying from cancer in 2030. Treatment of breast cancer at an early stage can significantly improve the survival rate of patients. Mammography is currently the most sensitive method for detecting early breast cancer. Retrospective studies have shown that radiologists can miss the detection of a significant proportion of abnormalities in addition to having high rates of false positives. The estimated sensitivity of radiologists in breast cancer screening is only about 75%. In order to improve the accuracy of interpretation, a variety of Computer-Assisted Detection (CAD) techniques have been proposed. In real sense, the Malignancy [1] or Benign, its type and from it detection of stage of cancer as invasive and non-invasive is a very fuzzy kind of decision making. Benign tumors are "well-differentiated," that the tumor cells differ only slightly in appearance and behavior from their tissue of origin. Malignant is used to describe a cancer that generally grows rapidly and is capable of spreading throughout the body. Detection of breast cancer is conducted by means of mammography and ultrasonography (USG) imaging. Microcalcifications (MCCs) clusters are one of the important radiographic indications related to breast cancer because they are present in 30%–50% of cancers found mammographic ally MCCs are tiny bits of calcium that may show up in clusters or in patterns (like circles) and are associated with extra cell activity in breast tissue. Scattered micro-calcifications are usually a sign of benign breast tissue. MCCs appear as small bright arbitrarily shaped regions on the large variety of breast texture background and characterize early breast cancer are detectable in mammograms shown in Fig1. For MCCs, the interpretations of their presence are very difficult because of its morphological features. The dense tissues especially in younger women may easily be misinterpreted as MCCs due to film emulsion error, digitization artefacts or anatomical structures such as fibrous strands, breast borders or hypertrophied lobules that almost similar to MCCs. Other factors that contribute to the difficulty of MCCs detection are due to their fuzzy nature, low contrast and low distinguish ability from their surroundings.

However, microcalcification detection from mammograms may be troublesome. To overcome this problem, CAD is developed to improve the diagnostic accuracy and the consistency of the radiologists' image interpretation. Issam El-
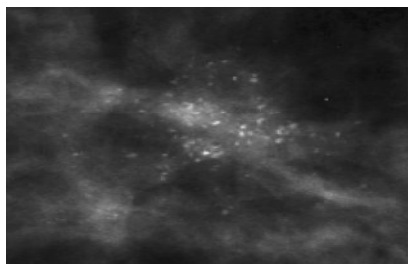
Fig1: Mammograms with Microcalcifications

Naqa et al [2] proposed that Support Vector Machine Classifier used to automatically detect the presence of MCCs in a mammogram using two class pattern classifications to locate the position of the mammogram. Tomasz Arodza et al [3] used CAD system technique, in which filter the original image with microcalcification contrast shape, enhance by wavelet-based sharpening algorithm and visual analysis. The evaluated system significantly improves the detection of microcalcifications in small field digital mammography. Alain Tiedeu et al [4] proposed that clustered microcalcifications are detected on mammograms based on texture analysis. Ryohei Nakayama et al [5] proposed a Computer – Aided Diagnosis scheme using a Filter bank technique with Hessian Matrix to classifying NC (nodular component) and NLC (nodular and linear component). Alolfe et al [6] proposed Computer aided diagnostic system based on wavelet analysis. Computer-Aided Diagnosis system that can be very helpful in diagnosing microcalcifications' patterns in digitized mammograms earlier and faster than typical screening programs.

Texture features have been widely used in mammogram classification. The texture features have the ability to distinguish between abnormal and normal cases. It can be characterized as the space distribution of gray levels in a neighborhood [1, 2]. Extracted texture features provide information about textural characteristics of the image. Different classifiers are used for medical imaging application including artificial intelligence, wavelet etc. Texture features has been measured from the extracted LBP image and used as parameter to enhance the classification result with ANN. The performance analysis of ANN is compared with K-Nearest neighbor and SVM classifier which has been discussed in this paper.

II. METHODOLOGY

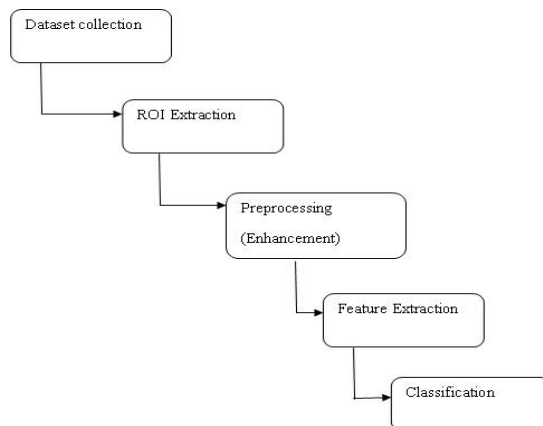The framework of proposed approach is shown in Fig 2.

Fig 2: Framework of proposed approach

The steps involved are as follows:
1. Images are collected from MIAS database.
2. ROI of 256X256 pixels of images are extracted from 1024X1024 pixel images. Preprocessing is used to remove the noise and Enhancement is performed to improve the adjustment of image

3. Features are extracted using Local Binary Pattern.
4. Texture features such intensity based features (mean, standard deviation) and GLCM based features (Energy, Entropy, Correlation and Homogeneity) are measured. Classification is performed with Artificial Neural Network based classifier.

A. Dataset Collection

It is difficult to access real medical images for experimentation due to privacy issue. Data is collected from the Mammographic Image Analysis Society (MIAS) consists of 322 images, which belong to three categories: normal, benign and malign, which are considered abnormal. All images are digitized at a resolution of 1024×1024 pixels and eight-bit accuracy (gray level). The existing data in the collection consists of the location of the abnormality (like the center of a circle surrounding the tumor), its radius, breast position (left or right), type of breast tissues (fatty, fatty-glandular and dense) and tumor type if exists (benign or malign). The proposed approach focus only on the microcalcification of benign and malignant images for abnormal and normal mammogram images.

B. ROI Extraction

Since all the portion of mammogram images are not necessary for detection of microclassification of clusters, ROI extraction of certain portion is needed. Using the locations of any abnormalities supplied by the MIAS for each mammogram, the ROI of size 256×256 pixels is extracted by entering the coordinates of X, Y and radius in pixel. Then the images are divided into two sets the training set and the testing set respectively. The extracted portion of ROI of mammography with benign and malignant is shown in Fig 3.

(a)   (b)

Fig 3: ROI extracted images (a) Benign (b) Malignant

C. Preprocessing and Enhancement

Mammograms are medical images that are difficult to interpret, thus a pre-processing phase is needed in order to improve the image quality and make the segmentation results more accurate. The first step involves the removal of noise and unwanted parts in the background of the mammograms. Then, Image enhancement operations can be used to improve the appearance of images, to eliminate noise or error, or to accentuate certain features in an image. The result of the process for detecting the suspicious lesions contains MCCs, and some noise. To reduce the amount of signatures which are not MCCs, perform a sequence of morphological operations on the filtered image.

The original image is smoothed and subtracted from an enhanced contrast image. The contrast enhancement technique increases the contrast by mapping the values of the input intensity image to new values such that, 1% of the data is saturated at low and high intensities of the input data. The Gaussian filters are used to smooth the original image, to attenuate as much as possible the MCC signals while preserving at best the mammary tissue signals. The adaptive technique is used to isolate MCCs from their immediate surroundings rather than from a different region that may have a different density level. The kernel size of 23 pixels is used, so that MCCs that are larger than 1.2mm in size are eliminated by the background correction process. By subtracting pixel wise the smoothed-image from the contrast-enhanced image, the background (mammary tissue) is strongly attenuated. The difference-image $D(i,j)$ obtained here above is binarized using a local adaptive thresholding algorithm. For a rectangular window centered on a pixel of coordinates (i,j) of width w_{bin} and height h_{bin}, the pixel *(i, j)* in the resulting image is $BI(i,j)$ is computed as:

$$BI(i,j) = 1 \; if \; \; if \; BI^{pix}(i,j) > \bar{m}_{i,j}^{bin} + k\sigma_{i,j}^{bin} \quad BI(i,j) = 0 \; otherwise \tag{1}$$

where k is a pre-selected integer and the mean grey-level value $m_{i,j}^{bin}$ and the standard deviation $\sigma_{i,j}^{bin}$ are computed. Then it performs morphological top-hat filtering on the grayscale or binary input image. The transformed binarized image from ROI extracted image is shown in Fig 4.
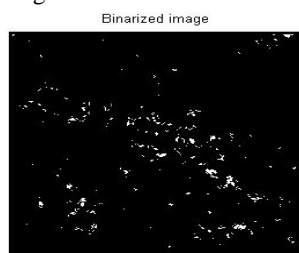

Fig 4: Transformed Binarized Image from ROI Extracted Image

D. Feature Extraction

This is a way of dimensionality reduction in medical image processing. When the input data to an algorithm is too large to be processed and suspected to be redundant. It has to be transformed into a reduced representation set of features. The features are used to extract the relevant information from the input data to perform the desired task.

1) Local Binary Pattern

Local Binary Pattern operator is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel with the value of the center pixel and considers the result as a binary number. The original LBP method is a complementary measure for local image contrast. LBP extracts texture feature in spatial domain. The LBP value is determined as:

$$LBP = \sum_{i=1}^{8} E_i . 2^{i-1} \tag{2}$$

where

$$E_i = \begin{cases} 1 \, \text{if} \; V_i \geq V_0 \\ 0 \; \text{if} \; V_i < V_0 \end{cases} \tag{3}$$

The preprocessed and enhanced mammography images are extracted using Local Binary Pattern and its image and histogram is shown in fig 5 (a) and (b) respectively.
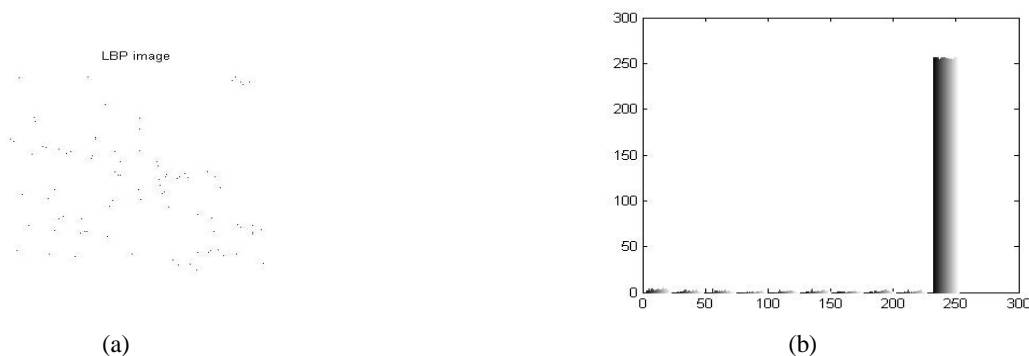

(a)   (b)

Fig 5 a) LBP image of Micro calcification of Mammogram and b) Histogram of LBP image

2) Texture Features:-

Intensity Based Features and GLCM based features are measured to analyze the performance. Intensity based features are first order statistics depends only on individual pixel values. The intensity and its variation inside the mammograms can be measured by features like mean and standard deviation using 40 samples of mammograms. The

Gray Level Co-occurrence Matrix (GLCM) texture measurement is a method to analyze image texture [5, 6]. It is a robust method that has been developed for calculating first and second order texture features from image. The texture features are calculated as follow:

1)Mean Value: The mean gives the average intensity value of an image. Mammographic images that contain micro calcifications have a higher mean than those of normal images. It is calculated as follows:

$$\mu = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} P(i,j) \tag{4}$$

where '*i,j*' indicates rows and columns of the image and *P (i, j)* is the cell denoted by row and column of the image.

2)Standard Deviation: It refers to the dispersion of values in a mammographic image around the mean.

$$SD = \sqrt{(mean)^2} \tag{5}$$

3) Energy: Energy represents the orderliness of a mammographic image. Energy is calculated as follows:

$$E = \sum_{i,j=0}^{n-1} P(i,j)^2 \tag{6}$$

4) Contrast: Contrast is a measure of the extent to which an object is distinguishable from its background. It represents the local variations present in an image, and calculates the intensity contrast between a pixel and its neighbour. It is calculated from the image as per the following equation.

$$C = \sum_{i,j=0}^{n-1} (i-j)^2 P(i,j) \tag{7}$$

where n denotes the number of pixels in the image and *P (i, j)* is the cell denoted by row and column of the image.

5) Coorelation: It measures how correlated a pixel is to its neighbor over the whole image. Correlation is 1 or -1 for a perfectly positively or negatively correlated image. So that it can be used to match the data to the predicted value.

$$Corr = \sum_{i,j=0}^{n-1} \frac{(i \times j)P(i,j) - \mu_i \mu_j}{\sigma_i \sigma_j} \tag{8}$$

6) Homogeneity: It measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal. These measure the values or pixel which is closer to the microcalcification of mammograms.

$$H = \sum_{i,j=0}^{n-1} \frac{P(i,j)}{1+|i-j|} \tag{9}$$

The list of features used for experimental work for images of mammograms is as shown in Table 1.

Texture Features	Normal		Abnormal	
	Training set	Test set	Trainingset	Test set
mean	252.581955	253.48	254.1333	253.604
standard deviation	18.63845532	14.85518	11.2435	14.19793
Contrast	0.2792	0.1696	0.1081	0.1549
Correlation	-0.0029	-0.0017	-0.0011	-0.0016
Energy	0.9887	0.9931	0.9956	0.9937
Homogeneity	0.995	0.997	0.9981	0.9972

Table 1: Texture feature measures for normal and abnormal of mammograms

E. Classification

ANN Classification is the process of learning to separate samples into different classes by finding common features between samples of known classes. It consist of input layer, hidden layer and output layer A set of samples may be taken from biopsies of two different tumor types, and their gene expression levels measured. Because making predictions on unknown samples is often used as a means of testing the ANN classifier. The two-layer feed-forward network is used. The training sets of images are compared with testing set of images and classify the tumor detected in those images. The classifier is used to classify the tumor as either normal or abnormal by detection of microcalcifications clusters of mammography images.

.

III. RESULT AND DISCUSSION

For implementation among the MIAS datasets of 322 images, fatty, dense and glandular tissue images of Normal (40) and Microcalcification of abnormal severity of benign or malignant (40) of 1024X1024 pixel images are considered. These images are normalized to 256X256 pixel of Region of Interest and stored as trained and testing set respectively. Preprocessing and enhancement is applied to remove of noise and artefact and improve the efficiency of the images. The local binary pattern is applied and Gray Level Cooccurrence Matrix (GLCM) features and Intensity based features

such as mean and standard deviations are measured. The classification is performed using ANN, in which network is created, trained for some samples and tested with remaining samples. The results are interpreted in confusion matrix is shown in Table 2. The confusion matrix describes the actual and predicted classes of the proposed method. It contains the information of samples related with normal and cancerous portion of mammographic images.

Output Class	Target Class	
	Normal	*cancerous*
Normal	40(TP)	0(FP)
cancerous	3(FN)	37(TN)

Table 2: Confusion Matrix for classification of breast cancer

A number of different measures are commonly used to evaluate the performance of the proposed method. These measures including Accuracy, sensitivity, specificity and precision are calculated from confusion matrix using the equation 10. It returns a value from -1(inverse prediction) to +1(perfect prediction).

$$\text{Accuracy} = (TP+TN)/(TP+TN+FP+FN)$$
$$\text{Sensitivity} = TP/(FP+TN)$$
$$\text{Specificity} = TN/(FN+TN)$$
$$\text{Precision} = TP/(TP+FP)$$

(10)

The performance measure of proposed approach and comparison with other classifiers are formulated in Table 3 and their performance measure analysis is depicted in Fig 6. It shows that accuracy is 96.25%, specificity is 92.50%, sensitivity is 97.5% and precision is 93.02% for ANN classifier. It reveals that better classification rate in accuracy and precision, when compared to k-Nearest Neighbor (k-NN) and Support Vector Machine (SVM) classifier.

Classifiers	Performance Measures			
	Specificity	Precision	Accuracy	Sensitivity
SVM	90%	90.70%	93.75%	97.50%
k-NN	90%	90.91%	95%	100%
ANN	92.50%	93.02%	96.25%	100%

Table3: Performance analysis of various classifiers

Fig 6: Performance Measure Analysis of classifiers

IV.CONCLUSION

Medical Imaging refers to view the human body in order to diagnose, monitor or treat medical conditions. Breast cancer has become a public health problem among women around the world. The features are extracted using LBP which provides statistical features. The texture features such as GLCM features and mean and standard deviations are measured and the classification of mammograms using ANN classifier is presented. The proposed experimental results show that when compared to several other methods 96.25% microcalcification detection in mammograms. The high accuracy of classification of MCCs obtained with the proposed ANN classifier can help radiologists in making an accurate diagnostic decision, which can reduce unnecessary biopsies. It will be shown that proposed system perform better detection than the existing system in-terms of sensitivity and classification rate.

REFERENCES

[1] W. liyang, Y. Yongyi, N. Robert, and J. Yulei, "A study on Several machine-Learning Methods for Classification of Malignant and Benign Clustered microcalcifications," *IEEE transactions on medical imaging*, vol. 24(3), March 2005.

[2] Issam El-Naqa, and Yongyi Yang, Miles N. Wernick, "Support Vector Machine Learning For Detection Of Microcalcifications In Mammograms" , IEEE International Symposium on Bio-Medical Imaging, 2002.

[3] Tomasz Arodza ,Marcin Kurdziel , "Detection of clustered micro calcifications in small field digital mammography ", Elsevier Journal of Computer Methods and Programs in Biomedicine, 81, 2005,pp.56-65.

[4] Alain Tiedeu, Christian Daul, "Texture-based analysis of clustered microcalcifications detected on mammograms", Elsevier Journal of Digital Signal Processing, Vol.22(1), January 2012,pp.124-132.

[5] Ryohei Nakayama, Yoshikazu Uchiyama, "Computer-Aided Diagnosis Scheme Using a Filter Bank for Detection of Microcalcification Clusters in Mammograms", IEEE Transactions On Biomedical Engineering, Vol. 53(2), February 2006, pp.273-283.

[6] M. A. Alolfe, A. M. Youssef, "Computer-aided Diagnostic System Based On Wavelet Analysis For Microcalcification Detection In Digital Mammograms", IEEE Bio-Medical Engineering Conference, December, 2008.

[7] Samir Kumar Bandyopadhyay,"Pre-processing of Mammogram Images", International Journal of Engineering Science and Technology, Vol. 2(11), 2010, pp.6753-6758.

[8] V. Saravanan, S.Pitchumani Angayarkanni, "A Novel Approach for Cancer Detection in MRI Mammogram Using Decision Tree Induction and BPN", International Journal of Image Processing (IJIP), Vol 4(6), pp.661-668.

[9] Vishnukumar K. Patel, Prof. Syed Uvaid, Prof. A. C. Suthar, " Mammogram of Breast Cancer detectionBased using Image Enhancement Algorithm", International Journal of Emerging Technology and Advanced Engineering, Vol.2(8), August 2012, pp. 143-147.

[10] Dheeba. J, Wiselin Jiji. G, "Detection of Microcalcification Clusters in Mammograms using Neural Network", International Journal of Advanced Science and Technology Vol.19, June, 2010, pp.13-22.

[11] Fatima Eddaoudi, Fakhita Regragui, "Microcalcifications Detection in Mammographic Images Using Texture Coding", Applied Mathematical Sciences, Vol. 5(8), 2011, pp. 381 – 393.

[12] Jawad Nagi, Sameem Abdul Kareem, Farrukh Nagi, Syed Khaleel Ahmed, "Automated Breast Profile Segmentation for ROI Detection Using Digital Mammograms", IEEE EMBS Conference on Biomedical Engineering & Sciences (IECBES 2010), Kuala Lumpur, Malaysia, 30th November - 2nd December 2010, pp. 87-92.

[13] A. K. Jain, "Fundamentals of Digital image Processing", Englewood Cliffs, Prentice Hall, 1989.

[14] Rafael C Gonzalez, Richard E Woods, Steven L. Eddins, Digital Image Processing using MATLAB,2nd ed, Pearson Education, Singapore, 2005.

[15] M.Vasantha, V. Subbiah Bharathi, "Medical Image feature extraction selection and classification", International Journal of engineering science and technology vol.2 (6), 2010,pp. 2071-2076.

[16]R.Nithya,B.Santhi,"comparative study of feature extraction, method for breast cancer classification", Journal of theoretical and applied Information Technology, vol.33(2), 2011, pp 220-224.

[17] Brijesh Verma, Peter McLeod, Alan   Klevansky, "Classification of benign and malign patterns in digital mammograms for the diagnosis of breast cancer", Expert System with Applications, 37,2010, pp.3344-
3351.