

EARLY DETECTION OF BREAST CANCER USING HYBRID CLASSIFICATION TECHNIQUE

C.Charanya¹

¹Student, Computer Science, Paavai Engineering College, Tamil Nadu, India

Abstract

In this paper, we are about to see the comparison of an existing system with the proposed system by achieving high accuracy with less time delay. Breast cancer which is a deadly disease can be seen numerous in number among women now-a-days. Mammography is considered to be one of the main tools for the early detection of breast cancer. Medical practitioners use various methods to identify the presence of micro calcifications. The medical image processing techniques are used for many medical problems and the same can be applied for the problem of cancer detection. In general, the mammogram images are used to identify the presence of cancer. But the accuracy of identifying the presence of cancer requires many constraints and cannot be easily said that there exist breast cancer. Radiologist uses the mammogram images which are not much accurate and clear. The mammogram images have only black and white pixels with various gray scale values. Even though, the presence of micro calcification has been identified, identifying the location where the exact cancer cells are present is highly a tedious job. The image processing technique for micro calcification detection has different stages namely preprocessing, segmentation and micro calcification detection. In general the mammogram images would have the presence of noise which is removed at the stage of preprocessing. In the segmentation stage, the gray scale pixels are grouped according to the value of gray scale. Based on the gray scale values the method identifies the pixels which are affected by the cancer and the shape of mass can be used for the determination to check if it is malignant or benign in the detection stage. Finally comes the classification stage and then the results.

Keywords: Mammography, Micro calcification, Malignant, Benign

1. INTRODUCTION

The rate of survival increases if breast cancer is detected at the early stage. Screening mammography, radiographic imaging of the breast, is considered to be the most effective tool for the detection of breast cancer in advance. These are performed on an asymptomatic woman for detecting early.

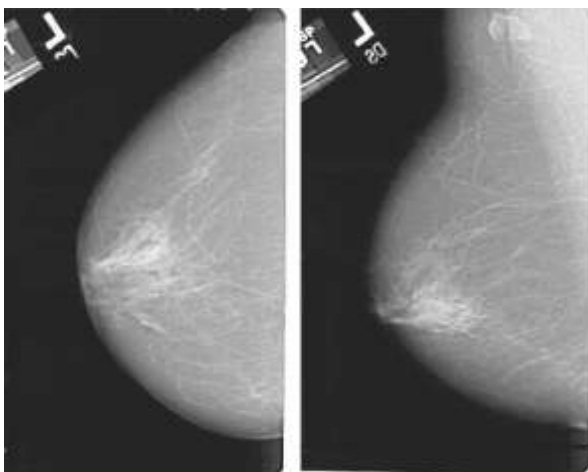


Fig-1 (a) MLO view

(b) CC view

There are two views of each breast recorded namely craniocaudal (CC) view, that is taken top-to-bottom, and a mediolateral oblique (MLO) view, (fig 1) that is a side view taken at an angle.

Radiologists visually search mammograms for specific abnormalities. Some of the important signs of breast cancer that radiologists look for are clusters of micro calcifications, masses, and architectural distortions. A mass is defined as a space-occupying lesion seen in at least two different projections. Masses are described by their shape and margin characteristics. Calcifications are tiny deposits of calcium, which appear as small bright spots on the mammogram. They are characterized by their type and distribution properties.

2. DETECTION OF CALCIFICATIONS

Calcifications can be seen as a small calcium deposits that are white in color, formed in the breast. They are present in various sizes and shapes. The characteristics that are considered to be the most important for calcifications are their size, shape or morphology, number, and distribution. It is to be noted that they are generally very small. Variation in size takes place between 0.1 mm to 1 mm and the average diameter as 0.3 mm. due to overlapping, small calcifications may be missed.

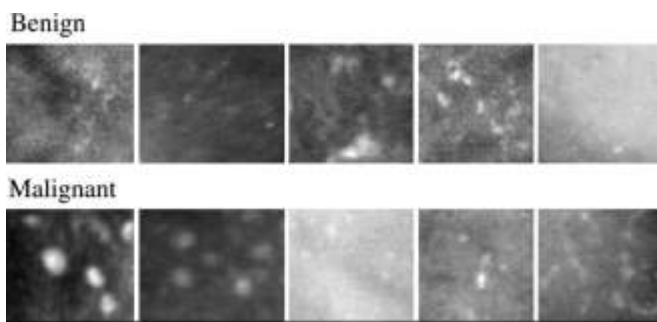


Fig-2 (a) Benign(small) (b) Malignant(big)

Sensitivity = $\frac{\text{Number of True Positive Classification}}{\text{Number of Malignant Lesion}}$

Specificity = $\frac{\text{Number of True Negative Classification}}{\text{Number of Benign Lesion}}$

Till date even expert radiologists finds it very difficult to characterize MCs(micro calcification) as benign or malignant according to their appearance in mammograms. This in turn increases the number of unnecessary breast biopsies. The difficulty in classifying benign and malignant MCs only with their appearance alone can be clearly seen from their visual similarity, as illustrated in the set of example images shown in Fig. 2.

3. METHODS OF EXISTING SYSTEM

3.1 Grey Level Co-occurrence Matrix (GLCM)

Grey level Dependency Matrix is known as GLCM in short. GLCM can be defined as “A two dimensional histogram of gray levels for a pair of pixels, which are separated by a fixed spatial relationship.” Computation of an image in GLCM is done using a displacement vector d , defined by its radius δ and orientation θ . Let us consider a 4x4 image which is represented with four grey-tone values 0 to 3 (figure 3-a). A generalized GLCM for that image is shown in figure (3-a) in which $\#(i,j)$ stands for the number of times grey tones occurs. the first-order statistics, second-order statistics or higher-order statistics are measured based on the number of pixels or dots.

Feature extraction based on Grey-level co-occurrence matrix (GLCM) comes under second-order statistics that is used to analyze image as a texture. The figure below represents the formation of the GLCM with 4 levels at the distance $d = 1$ and the direction of 0° .

Grey tone	0	1	2	3
0	$\#(0,0)$	$\#(0,1)$	$\#(0,2)$	$\#(0,3)$
1	$\#(1,0)$	$\#(1,1)$	$\#(1,2)$	$\#(1,3)$
2	$\#(2,0)$	$\#(2,1)$	$\#(2,2)$	$\#(2,3)$
3	$\#(3,0)$	$\#(3,1)$	$\#(3,2)$	$\#(3,3)$

General form of GLCM

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

Fig-3(a) Test image for GLCM

4	2	1	0
2	4	0	0
1	0	6	1
0	0	1	2

GLCM for $\delta=1$ and $\theta=0$

6	0	2	0
0	4	2	0
2	2	2	2
0	0	2	0

GLCM for $\delta=1$ and $\theta=90$

4	1	0	0
1	2	2	0
0	2	4	1
0	0	1	0

GLCM for $\delta=1$ and $\theta=45$

2	1	3	0
1	2	1	0
3	1	0	2
0	0	2	0

GLCM for $\delta=1$ and $\theta=135$

Fig-3(b) The four GLCM for angles equal to 0° , 45° , 90° and 135°

3.2 Expectation-Maximization Logistic-Regression

Algorithm (EMLR)

A decision is taken for an image as benign/malignant, by a two step procedure. Separate classifiers are applied on the CC view and MLO views to determine if the cluster is malignant or not. The image-level decision that is modeled as a hidden random variable can neither be seen in the train image nor in the test image. by integrating the decisions of both views, the final biopsy-level decision is determined as the second step. In this study we can learn about the forecasting of biopsy result in each view and combing of two different views into an optimal decision. logistic regression is used to model each view-level binary decision. In the E-step of EM algorithm, estimation of hidden random variables are being done. In the M-step of EM algorithm, updating of values of the model parameters are being done.

3.2.1 EM algorithm

Input: Pairs of feature vectors $(x_{1,cc}, x_{1,mlo}), \dots, (x_{n,cc}, x_{n,mlo})$ with corresponding binary labels z_1, \dots, z_n .

Output: model parameters w_{cc} , w_{mlo} and θ .

The EM algorithm:

E-step: For each mammographic pair data $(x_{t,cc}, x_{t,mlo})$ calculate

$$c_{ij} = \frac{\theta_{ij}(z_t)p(y_{t,cc} \equiv i | x_{t,cc}; w_{cc})p(y_{t,mlo} \equiv j | x_{t,mlo}; w_{mlo})}{P(z_t | x_{t,cc}, x_{t,mlo}; w_{cc}, w_{mlo}, \theta)}$$

$$i=0,1 \quad j=0,1 \quad t=1, \dots, n$$

M-step:

- Update the view fusion parameter θ .

$$\theta_{i,j} = \frac{\sum_{\{t|z_t \equiv i\}} c_{ij}}{\sum_t c_{ij}}, \quad i=0,1, \quad j=0,1.$$

- To find w_{cc} , apply gradient ascent.

$$L(w_{cc}) \equiv \sum_t \sum_{i=0,1} (c_{i0} + c_{i1}) \log p(y_{t,cc} \equiv i | x_{t,cc}; w_{cc})$$

$$w_{cc} \leftarrow w_{cc} + \lambda \sum_t (c_{t10} + c_{t11} - \sigma(w_{cc} x_{t,cc})) x_{t,cc}$$

- To find w_{mlo} apply gradient ascent.

$$L(w_{mlo}) \equiv \sum_t \sum_{j=0,1} (c_{t0j} + c_{t1j}) \log p(y_{t,mlo} \equiv j | x_{t,mlo}; w_{mlo})$$

$$w_{mlo} \leftarrow w_{mlo} + \lambda \sum_t (c_{t01} + c_{t11} - \sigma(w_{mlo} x_{t,mlo})) x_{t,mlo}$$

Table 1: the expectation-maximization (em) algorithm

4. PROPOSED SYSTEM

4.1 Steps

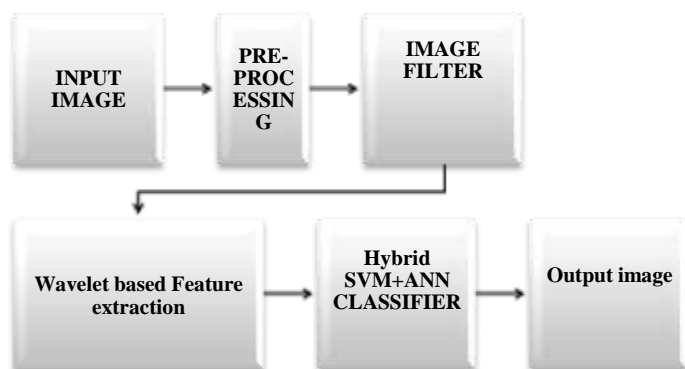


Fig-4: Steps in proposed system

The mammogram images have only black and white pixels with various gray scale values. Even though, the presence of micro calcification has been identified, identifying the location where the exact cancer cells are present is highly a tedious job. The image processing technique for micro calcification detection has different stages namely preprocessing, segmentation and micro calcification detection. In general the mammogram images would have the presence of noise which is removed at the stage of preprocessing. In the segmentation stage, the gray scale pixels are grouped according to the value of gray scale. Based on the gray scale values the method identifies the pixels which are affected by the cancer and the shape of mass can be used to determine whether it is malignant or

benign in the detection stage. Finally comes the classification stage and then the results.

4.2 Preprocessing

The preprocessing is used to eliminate any interfere in the image to identify the tumor. It is significantly increasing the reliability, robustness of the image. It is used to remove the noise and enhance the image using Gaussian filter. To reduce the work area only to the relevant region that exactly contains the breast. It acquisition the image from the database as an input image. Gaussian filter Gaussian filter is used to remove the noise from the image and its impulse response is Gaussian function. So the noise added and removed from the input image. It is a class of linear smoothening filters with the weight chosen according to the shape of Gaussian function. Gaussian kernel is mainly used for smoothening purpose. It is used to removing noise from the normal distribution. The filter window is symmetric is center in the time domain in the non-casual filters. It makes Gaussian filter as unrealized. This is generally no consequences for application where the bandwidth is much larger than the signal. Gaussian functions have five properties that make them useful. These properties of Gaussian filter smoothening filters are effective low-pass filters from the view of both the spatial and frequency domains are efficient.

4.3 Feature Extraction

The features are extracted from the mammograms using wavelet based feature extraction. It is used to extract the textural patterns of the mammograms to identify the suspicious region of masses. The wavelet is used to analyze different frequencies of an image using different scales. This approach is enabling of both local and global features are present in the image. The feature extraction is used to measure the properties from the enhanced image are Area, Origin, major axis length, minor axis length, eccentricity, orientation, filled area, extreme, solidity, equiv diameter. The area is the scalar value; it computes the actual number of pixels in the region. Then the origin is the vector and it computes the centre of the mass region.

4.4 Classification

Classification is used to classify the tumor as normal, benign, malignant as it helps to predict the feature using support vector machine SVM and ANN. The extracted feature image will be an input to the classification system. Classification process is defined by two phases namely training and testing. Known data is given as an input and the classifier is being trained in the training phase and unknown data is given as an input and the classifier is being trained in the testing phase. Classifications have the assignment to an unknown pattern of a predefined class, according to the pattern presented in the form of a feature vector.

4.4.1 Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a supervised learning model with an associated learning algorithm that

analyzes data and recognize pattern. In other words, given labeled training data the algorithm outputs an optimal hyper plane which categorizes new examples. It searches for the separation of hyper plane into positive and negative with maximum margin..SVM classifier is designed to solve a binary classification problem by finding the class boundary of the hyper plane maximizing the margin in the given training data. Kernel functions are used in SVM to solve the nonlinear boundary problems. The optimal hyper plane is for linear and are extended to patterns that are non linear. The transformation of original data into new space is separable in kernel function.

4.4.2 Artificial Neural Network (ANN)

Artificial Neural Networks are the computational system whose concept is derived from biological neural networks.

An ANN consists of a collection of processing elements that are highly interconnected. ANN transforms a set of inputs to a set of desired outputs. Transformation results are determined by the characteristics of the elements and the weights associated with the interconnections among them. Neural networks have the advantages of adaptive learning, self-organization, and real time operation. The neural network based system should be efficient and generalized enough to diagnose new unseen cases of breast cancer which was not present in training data. Therefore the proposed work also focuses on improving the generalization property of the designed network. This technique has provided a better accuracy. ANN tools have shown to be valuable in reducing the workloads on clinicians by detecting artifact and providing decision support.

Table-2 Performance of the various classifiers

S.No	Image Classifiers	Total No. of Images	TP	FN	TN	FP	Sensitivity (%)	Specificity (%)	Accuracy (%)
	LOGISTIC	248	205	23	14	6	97.16	62.16	65.9
	SVM	248	200	29	13	6	97.09	69.05	69.1
	EM-LR	248	197	24	18	9	95.63	57.14	70.6
	MULTIVIEW	248	210	20	14	4	98.13	58.82	72

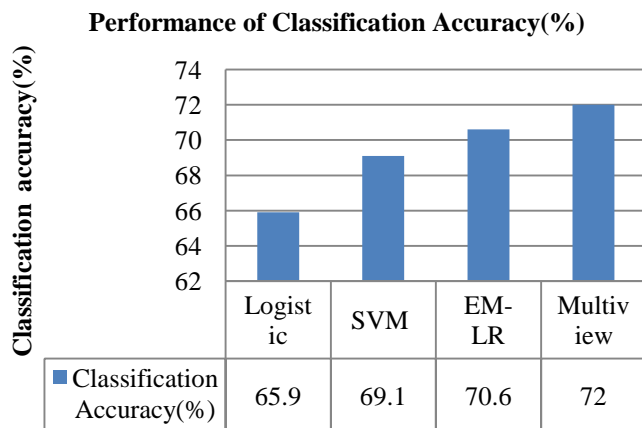


Chart-1 Performance Comparison Graph

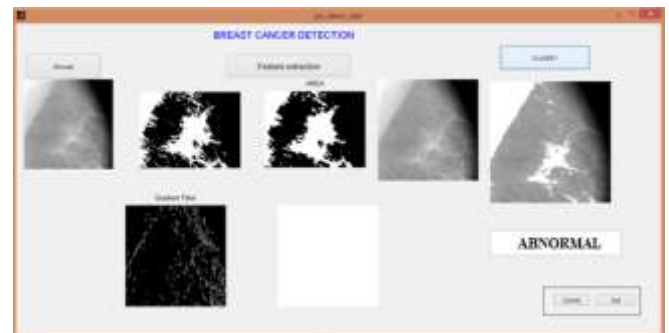


Fig 5: Screen Shot of Implementation of Detection

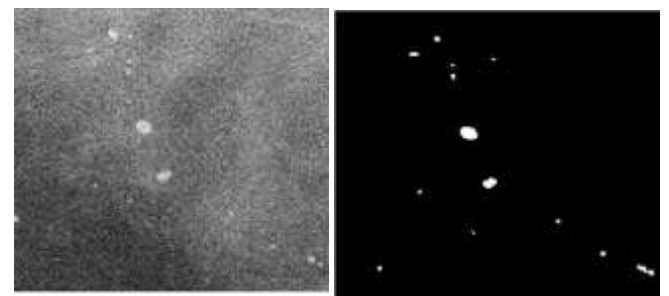


Fig 6: (a) Input Image (b) Output Image

4.5 Output Results

The proposed method has been implemented and evaluated for their Normalization Accuracy using various data sets. The methods have been implemented using Matlab and evaluated using various data sets. For evaluation the DDSM data set has been used. The DDSM data set contains 126 classes of mammograms. Among them 90 percent is taken as training set and 10 percent taken for testing.

Table -3 Details of evaluation parameters

Parameter	Value
Data Set	DDSM
No of mammograms	126
Tool Used	Matlab 2013 a

5. CONCLUSION

The earlier method uses only the threshold values of the pixels of mammogram images and based on the value they perform clustering. This reduces the accuracy of clustering and detection of micro calcification this is due to computes the similarity based on the gray scale values. This also results in higher time complexity.

The proposed system focuses on the depth of the white mass values i.e. the effective classification method which uses the white mass values and their similarity by considering each region of the mammogram image using 3D volumetric analysis. This improves the accuracy and performance of micro calcification detection is increased with less time taken for processing the image.

REFERENCES

- [1] HD. Cheng, X. Cai, X. Chen, L. Hu and X. Lou, "Computer aided detection and classification of microcalcifications in mammograms: a survey", Pattern Recognition, Vol. 36, No. 12, 2003, pp. 2967–91.
- [2] M. Elter and A. Horsch, "CADx of mammographic masses and clustered microcalcifications: a review", Medical Physics, Vol. 36, No. 6, 2009, pp. 2052–68.
- [3] HP. Chan, B. Sahiner, KL. Lam, N. Petrick, MA. Helvie, MM. Goodsitt, et. al., "Computerized analysis of a mammographic microcalcifications in morphological and texture feature spaces", Medical Physics, Vol. 25, No. 10, 1998, pp. 2007– 19.
- [4] Jayalakshmi T. and Santhakumaran A. 2011. Statistical normalization and back propagation for classification. International Journal of Computer Theory and Engineering, 3, 89-93.
- [5] I. Leichter, R. Lederman, S. Buchbinder, P. Bamberger, B. Novak and S. Fields, "Optimizing parameters for computer aided diagnosis of microcalcifications at mammography", Acad. Radiol., Vol.7, No. 6, 2000, pp. 406–12.
- [6] Manikandan, G., Sairam, N., Sharmili S. and Venkatakrishnan S. 2013. Achieving privacy in data mining using normalization, Indian Journal of Science and Technology, 6, 4268-4272.

BIOGRAPHY



C.Charanya, She was born in Attur, Salem (D.t), TamilNadu, India. She completed her B.TECH. Information Technology from Sona College of Technology, Salem, affiliated to ANNA University Chennai in 2012. She is currently doing the Post-graduation in

Computer Science and Engineering from Paavai Engineering College, Namakkal, affiliated to ANNA University Chennai Tamil Nadu.