

Classification of Normal and Abnormal Patterns in Digital Mammograms for Diagnosis of Breast Cancer

R. Nithya
Assistant Professor, School of
Computing
SASTRA University
Tamil Nadu, India

B. Santhi
Professor, School of
Computing
SASTRA University
Tamil Nadu, India

ABSTRACT

In the last few years, computerized tool play important role in detection of breast cancer. This paper proposes a method for breast cancer diagnosis in digital mammograms using GLCM (Grey Level Co-occurrence Matrix) features. In this paper CAD (Computer Aided Diagnosis) system developed using GLCM feature and neural network. Mammography is an efficient tool for early detection of breast cancer. Computerized methods have recently show great tool in providing radiologists with second opinion about breast cancer diagnosis. Five GLCM features for mammogram images are extracted. Mammogram image is classified into normal image and cancer image. The effectiveness of this paper is examined on DDSM (Digital Database for Screening Mammography) database using classification accuracy, sensitivity and specificity. The overall accuracy can be improved by most relevant GLCM features, which is selected by feature selection algorithm.

Keywords

Mammograms, GLCM, Neural Network.

1. INTRODUCTION

Breast cancer is most common cancer among women. Mammographic images are X-ray images of breast region [1]. The commonly used diagnostic technique including biopsy, mammography, thermography and ultrasound image. Among these techniques mammography is best approach for early detection. In early stage visual clues are subtle and varied in appearance, it makes diagnosis difficult. The abnormalities are hiding by breast tissue structure. Breast cancer detection and classification of mammogram images is the standard clinical practice for the diagnosis of breast cancer. Different kind of soft computing techniques existing in the field of breast cancer detection such as neural network and fuzzy logic. Mammography is the efficient tool available for the detection of breast cancer before physical symptoms appear. The earlier the cancer detection is challenging and difficult task. The biopsy is a standard approach for cancer detection manually under a microscope. But biopsy is difficult and time consuming task. The features can be directly extracted from the images. Once an abnormality is detected on a mammogram using CAD system,

the radiologist's recommends biopsy. However, the detection of abnormality is a difficult task and unnecessary biopsy is performed. The CAD technique used to reduce unnecessary biopsy. The goal of diagnosis is distinguish between normal and cancer image. A good CAD system identified by low false negative and false positive rate. A mammogram abnormality consists of two types: mass and calcification. Calcification often associated with non-cancerous pattern. A mass is suspicious indicator of breast cancer.

Feature extraction is the first step in breast cancer detection. Texture feature is important for image classification. Various techniques have been used for computing texture features [2]. In this paper co-occurrence matrix are used. GLCM is a powerful tool for image feature extraction. Gray level pixel distribution described by statistics like probability of two pixels having particular gray level at particular spatial relationships. This spatial information is provided as two dimensional gray level matrices.

The remainder of this paper is organized as follows. Section 2 describes the existing techniques followed by the proposed research methodology in section 3. The experiments results are presented in section 4. Section 5 discusses the obtained results by the proposed technique. The conclusion and future direction are stated in the final section.

2. RELATED WORK

Various kinds of techniques have been proposed to improve the accuracy of breast cancer classification. The research in the areas of CAD systems for breast cancer detection has been conducted in last few decades. Brijesh Verma, Peter McLeod and Alan Klevansky [1] used novel soft cluster neural network. The highest classification accuracy obtained by this approach was 93% on mammograms from the DDSM. Ali Keles, Ayturk Keles and Ugur Yavuz [3] proposed neuro-fuzzy rules and sensitivity obtained was 76% on UCI machine learning repository. M.Vasantha, Dr.V.Subbiahbarathi and R.Dhamodharan [4] proposed hybrid algorithm for mammogram classification using GLCM features. They obtained accuracy was 87.5% on mammograms from the Mammograms Image Analysis Society (MIAS). Mohammed J.Islam, Majid Ahamadi and Maher A.Sid-Ahmed [5] used artificial neural network

technique and they obtained sensitivity was 90.1% on mammograms from the MIAS. T.S.Subashini, V.Ramalingam and S.Palanivel [6] used texture features to assessment of breast tissue density in digital mammograms and they obtained accuracy was 95.44% on mammograms from MIAS. Pasquale Delogu, Maria Evelina Fantacci, Parnian Kasae and Alessandra Retico [7] used texture features and neural classifier and obtained sensitivity of 80-85%.

3. METHODOLOGY

An overview of proposed method is presented in Figure 1.

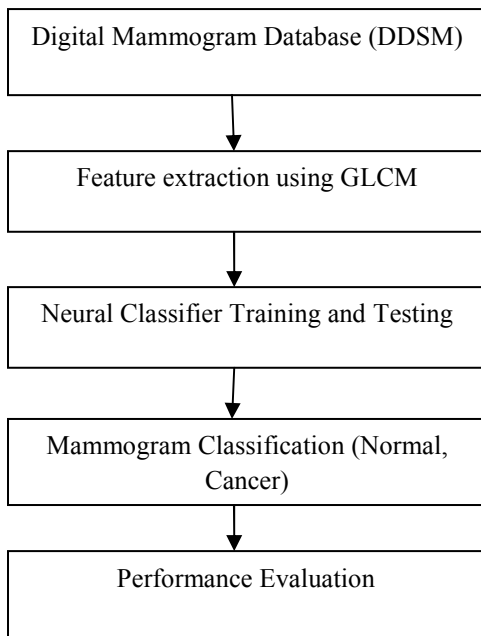


Fig 1: An Overview of Proposed Method

3.1 Data Collection

Data for experiment in the proposed method taken from DDSM. DDSM is a resource for mammographic image. The total 250 mammograms have been used for training and testing. DDSM contains more than 2500 mammograms available at <http://marathon.csee.usf.edu/Mammography/DDSM>.

3.2 Feature Extraction

Feature extraction is very important part of pattern classification. To identify texture in image, modeling texture as a two dimensional array gray level variation. This array is called Gray Level co-occurrence matrix. GLCM features are calculated in four directions which are $0^\circ, 45^\circ, 90^\circ, 145^\circ$ and four distances (1,2,3,4). Five statistical measures such as correlation, energy, entropy, homogeneity and sum of square variance are computed based on GLCM [8-13]. Table 1 provides explanation and equation for five features. The size of GLCM is determined by number of gray level in an image. For each of the formula: G is the number of gray level used. The matrix element P (i, j |Δx, Δy) is the relative frequency with two pixels separated by pixel

distance (Δx, Δy), occur within a given neighborhood, one with intensity i and other with intensity j . Table 2 shows GLCM features for normal and cancer class. μ_i, μ_j are mean and σ_i, σ_j are standard deviation of P (i, j), where

$$\mu_i = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j | \Delta x, \Delta y)$$

$$\mu_j = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} j \times P(i, j | \Delta x, \Delta y)$$

$$\sigma_i = \sum_{i=0}^{G-1} (i - \mu_i) \sum_{j=0}^{G-1} P(i, j)$$

$$\sigma_j = \sum_{i=0}^{G-1} (i - \mu_j) \sum_{j=0}^{G-1} P(i, j)$$

3.3 Classification

Neural classification consists of two processes: Training and Testing. Neural network is the best tool in pattern classification application. The classifier is trained and tested on mammogram image. The classification accuracy depends on training. Neural network contains three layers: input layer, hidden layer and output layer [14, 15]. The designing of neural network consist a number of input, hidden, output units and activation function. The first layer has 5 nodes and second layer has two nodes. One node is needed for output layer. The actual output is compared with desired output by error rate. For the neural network model error (E) is calculated using equation 1

$$E = d - a \text{ ----- (1)}$$

where d is the desired output, a is the actual output. The output of the network is determined by activation function such as sigmoid. Neural network are trained by experience, when fed an unknown input into neural network, it can generalize from past experience and produce a result. Five GLCM features fed to neural input layer. The output layer produce either 1(normal) or 0(cancer). Figure 2 shows the neural network architecture.

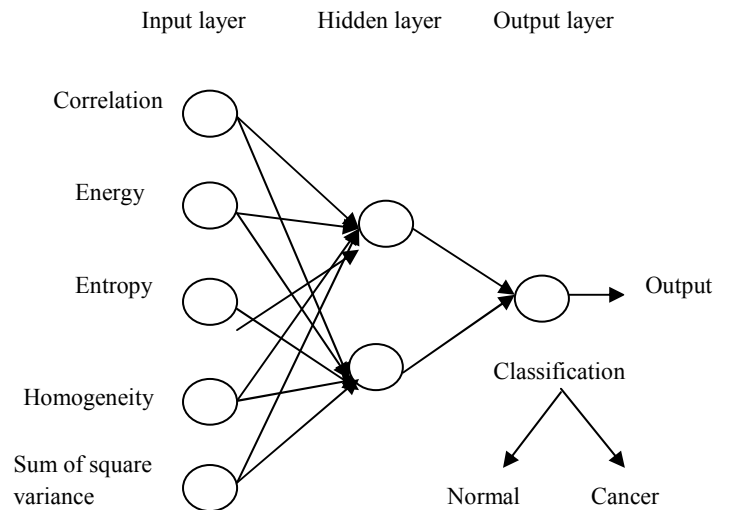


Fig 2: Neural Network Architecture

Step 6: Use test patterns and calculate the accuracy.

Algorithm

- Step 1: Extract features from mammograms
- Step 2: Create input and target for normal class
- Step 3: Create input and target for cancer class
- Step 4: Initialize weights at small random values
- Step 5: calculate output

4. PERFORMANCE MEASURE

The performance of the prediction was evaluated in terms of sensitivity, specificity, accuracy, the respective formula are given in Table 3. Accuracy measures the quality of the binary classification (two-class). It takes into account true and false positives and negatives. Accuracy is generally regarded with balanced measure whereas sensitivity deals with only positive cases and specificity deals with only negative cases. TP is number of true positives, FP is number of false positives, TN is number of true negatives and FN is number of false negatives.

Table 1.Explanation and Formula for GLCM Features

Features	Explanation	Formula
Entropy	Entropy measures the statistical randomness.	$-\sum_{i,j=0}^{G-1} P(i, j) \log (P(i, j))$
Energy	Energy is also known as uniformity of ASM (angular second moment) which is the sum of squared elements from the GLCM.	$\sum_{i,j=0}^{G-1} P(i, j)^2$
Homogeneity	Homogeneity is to measure the distribution of elements in the GLCM with respect to the Diagonal.	$\sum_{i,j=0}^{G-1} \frac{P(i, j)}{1+ i-j }$
Correlation	Correlation measures the joint probability occurrence of the specified pixel pairs.	$\sum_{i,j=0}^{G-1} P(i, j) (i - \mu_i)(j - \mu_j) / \sigma_i \sigma_j$
Sum of square variance	This feature puts relatively high weights on the elements that differ from the average value of P (i, j).	$\sum_{i,j=0}^{G-1} P(i, j) (i - \mu)^2$

Table 2.GLCM Features Value for Normal and Cancer Class

Image id	Image class	Correlation	Energy	Entropy	Homogeneity	Sum of square variance
Mam1	Normal	0.258	0.47	1.4270	0.8880	8.6527
Mam2	Normal	0.638	0.213	1.9880	0.8320	7.1783
Mam3	Normal	0.223	0.194	2.2070	0.7700	8.5741
Mam4	Normal	0.493	0.274	1.7510	0.8780	7.0682
Mam5	Normal	0.423	0.375	1.6550	0.8700	8.648
Mam6	Cancer	0.916	0.424	1.3270	0.9610	3.743
Mam7	Cancer	0.912	0.231	1.7640	0.9200	7.8946
Mam8	Cancer	0.963	0.309	1.5340	0.9580	5.9444
Mam9	Cancer	0.957	0.295	1.5680	0.9460	7.5551
Mam10	Cancer	0.894	0.285	1.6620	0.9270	5.8116

A confusion matrix provides information about actual and predicted cases produced by classification system [3].The performance of the system is examined by demonstrating correct and incorrect patterns. They are defined as confusion matrix in Table 4. The higher value of both sensitivity and specificity shows better performance of the system.

Table 3.Formula for Measures

Measures	Formula
Sensitivity	$TP/(TP+FN)$
Specificity	$TN/(TN+FP)$
Accuracy	$(TP+TN)/(TP+FP+TN+FN)$

Table 4.Confusion Matrix

Actual	Predicted	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

TP-predicts cancer as cancer

FP-predicts cancer as normal

TN-predicts normal as normal

FN- predicts normal as cancer.

5. EXPERIMENT RESULT

This section details the results of automatic classification on mammograms using GLCM features and neural network. In order to evaluate this work, conduct experiments on the DDSM database. The proposed method trained with 200 mammograms (100 normal, 100 cancer) and tested with 50 mammograms (25 normal, 25 cancer).In this analysis first procedure is extract features from mammograms. The proposed neural network has 5 units in input layer, 2 units in hidden layer and 1 unit in the output layer. Five GLCM features are extracted from the 200 mammograms and these input fed to neural input layer. At the end of the training process updated weight values are stored. The trained network tested for 50nmammograms.In training process outputs are calculated using saved weight value. Table 5

and Figure 3 shows confusion matrix for this experiment. The performance measures are done by computed output. Table 6 and Figure 4 show the computed sensitivity, specificity and accuracy for testing data of proposed method. The obtained classification accuracy of this method was 96% whereas sensitivity and specificity were 100% and 93%.The overall classification accuracy for normal is 100% and for cancer is 92%.

Table 5.Confusion Matrix for Testing

Actual	Predicted	
	Cancer (Positive)	Normal (Negative)
Cancer(Positive)	23(TP)	2(FP)
Normal(Negative)	0(FN)	25(TN)

Table 6.Performance Measure

Test case	Accuracy	Sensitivity	Specificity
50	96%	100%	93%

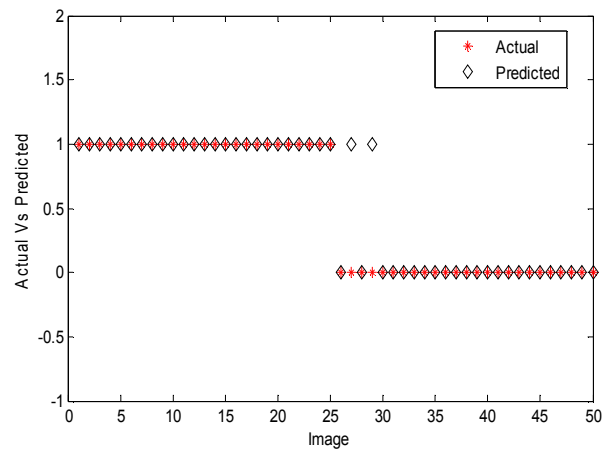


Fig 3: Confusion Matrix for Testing Result

6. CONCLUSION

The method employed in this paper has given better performance. The CAD system is developed for the classification of mammogram into normal and cancer pattern with the aim of supporting radiologists in visual diagnosis .This paper has investigated a classification of mammogram images using GLCM features. The maximum accuracy rate for normal and cancer classification is 96%. For future work, GLCM features combined with statistical moment features to improve the results in classification of mammogram images. Using

proper feature selection method accuracy may be improved efficiently.

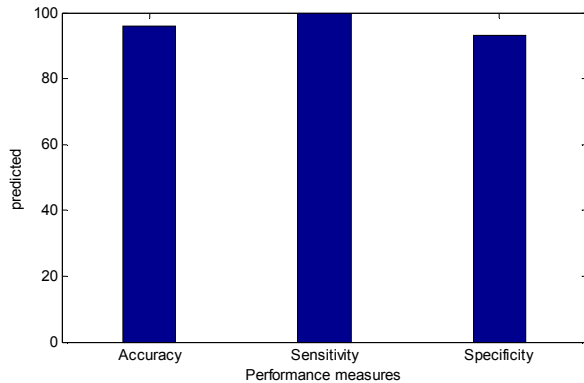


Fig 4: Performance Measure for Testing

7. REFERENCES

- [1] Brijesh Verma and Ping Zhang, "A novel neural-genetic algorithm to find the most significant combination of features in digital mammograms", *Applied Soft Computing*, no.7, pp.513-525, 2007.
- [2] H.S.Sheshadri and A.Kandaswamy, "Experimental investigation on breast tissue classification based on statistical feature extraction of mammograms", *Computerized Medical Imaging and Graphics*, no.31, pp.46-48, 2007.
- [3] Ali Keles, Ayturk Keles and Ugur Yavuz, "Expert system based on neuro-fuzzy rules for diagnosis breast cancer", *Experts Systems with Applications*, pp.5719-5726, 2011.
- [4] M.Vasanth. DR, V.Subbiahbarathi and R.Dhamodharan, "Medical Image Feature, Extraction, Selection and Classification", *International Journal of Engineering Science and Technology* Vol. 2(6), pp. 2071-2076, 2010.
- [5] Mohammed J.Islam, Majid Ahamadi and Maher A.Sid-Ahmed, "An efficient automatic mass classification method in digitized mammograms using artificial neural network", Vol.1, No.3, 2010.
- [6] T.S.Subashini, V.Ramalingam and S.Palanivel, "Automated assessment of breast tissue density in digital mammograms", *Comuter Vision and Image Understanding*", no.114, pp.33-44, 2010.
- [7] Pasquale Delogu, Maria Evelina Fantacci, Parnian Kasae and Alessandra Retico, "Characterization of mammographic masses using a gradient-based segmentation algorithm and a neural classifier". *Computers in and Biology Medicine*, no.37, pp.1479-1491, 2007.
- [8] S.Daniel Madan Raja, A.Shanmugam, "Artificial neural networks based war scene classification using invariant moments and GLCM features: A comparative study", *International Journal of Engineering Science and Technology*, Vol.3, No.2, 2011.
- [9] Andrea Baraldi Flavio Parmiggiani, "An investigation of textural characteristics associated with gray level co-occurrence matrix statistical parameters", *IEEE transactions on geosciences and remote sensing*, Vol.33, No.2, 1995.
- [10] Dr. H.B. Kekre, Sudeep D.Thepade, Tanuja K Sarode and Vashali Suryawanshi, "Image retrieval using texture features extracted from GLCM, LBG and KPE", *International journal of Computer Theory and Engineerin*, Vol.2, No.5, 2010.
- [11] Ramana Reddy, A.Suresh, M.Radhika Mani and V.Vijaya Kumar, "Classification of textures based on features extracted from preprocessing images on random windows", Vol.9, 2009.
- [12] Dipankar Hazra, "Texture recognition with combined GLCM, wavelet and rotated wavelet features", *International Journal of Computer and Electrical Engineering*: Vol.3, No.1, 2011.
- [13] Norila Md Yusof, Nor Ashidi Mat Isa and Harsa Amylia Mat Sakim, "Computer -aided detection and diagnosis for microcalcifications in mammogram", *International Journal of Computer Science and Network Security*", Vol.7, No.6, 2007.
- [14] S.Saheb Basha, Dr.K.Satya Prasad, "Automatic detection of breast cancer mass in mammograms using morphological operators and fuzzy c-means clustering", *Journal of Theoretical and Information Technology*.
- [15] Shekar Singh, Dr.P.R.Gupta, "Breast cancer detection and classification using neural network", *International Journal of Advanced Engineering Sciences and Technologies*, Vol No.6, Issue No.1, pp.004-009.