Efficient Technique for Texture Features Based CAD System for Mammogram Images using GLCM and RBFNN (Breast Cancer Detection)

Ritu Tomar[1], Dr. Indu Kashyap[2]

[1]*Manav Rachna International University, Faridabad*
[1]*+91-8130986866*
[1]*ritutomar91@gmail.com*
[2]*Manav Rachna International University, Faridabad*
[2]*+91-9899798033*
[2]*indu.fet@mriu.edu.in*

Abstract: *Breast cancer is the most common malignancy of women and cause many deaths. The early detection is the best solution in order to avoid mastectomy, reduce the chances to return, and decrease the rate of death. This is done with the help of CAD system. CAD system has been developed to help radiologist for detecting cancer in women breast. CAD system examines mammographic images of a breast. The aim of this paper is to check the accuracy of GLCM features and improve system's accuracy using different features of GLCM which are being used less number of times. This is done by using GLCM (Grey level co-occurrence matrix) for extracting features and Neural Network for training and testing. In the future it can be used in a hierarchically CAD system that classify the images firstly according density, after indicates if there is a lesion and finally what kind of lesion is. The developed techniques have been implemented in MATLAB ® and tested on real case studies. The developed CAD system is second pair of eyes for the radiologist.*

1. INTRODUCTION

For years, Cancer is the most common malevolence of human and cause many deaths. In this, breast cancer is the most common reason of death in women. In most of the cases woman don't know about its presence, only in the later stages it gets noticed by its symptoms and this comes to be the main reason why there is number of deaths. So, the early detection plays an important role to save women life. But, the earlier cancer detection is challenging and difficult task. Most of the doctors suggest biopsy for early detection of cancer. Biopsy is detection of cancer manually under microscope. This is very difficult, time consuming and costly task. With the help of CAD, diagnosis with mammograms becomes more easy and accurate [1]. In breast cancer's diagnosis it provides easy tool for its detection at an early age [6]. Various study show that with the help of developed CAD System, detection of masses in breast cancer as benign and malignant is very much improved for the diagnostic purpose by radiologist [7]. CAD systems can

provide radiologist a "second pair of eyes," whose consistency and repeatability is very good.

With the objective of getting more accuracy in benign and malignant tissues we provide this work to analyze the performance of GLCM (textural feature extraction) with RBFNN classifier.

Various techniques have been used for computing textural features. GLCM is proved to be a good statistical method for extracting textural features from various images [3].

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. Radial Basis Function Neural Network (RBFNN) is a type of Artificial Intelligence (AI) system is used for tissue classification [2]. This work is proposed for analyzing the accuracy of GLCM features by passing these features for the classification of breast cancer from mammographic images as either benign or malignant.

The remainder of this paper is as follows. Section 2 describes the requirement analysis. Where section 3 describes proposed research with methodology. Section 4 describes the terms for analyzing performance using various terminologies. Section 5 discusses the obtained results by the proposed technique, conclusion and future scope.

2. REQUIREMENT ANALYSIS

Mammogram image is used for feature extraction and classification. Mammograms are low dose X-ray image of the breast to check the abnormal changes for the detection of breast cancer.

The database which we are using here for the proposed CAD system uses digitized screening mammograms, which is taken from the Mammography Image Analysis Society

(MIAS), mini mammography database. MIAS database is specially given for research purpose. This one is created by reducing the original MIAS database mammograms to 200 micron pixel edge and clipped so that every image becomes 1024 pixels * 1024 pixels. Basically, mammogram images consist of two views one is CranioCaudal view (CC) and second is Medio-Lateral Oblique, the database which we are using here consists of Medio-Lateral Oblique (MLO) views of the mammograms with ground truth of each abnormality. Only the architectural distorted mammograms are considered to make the feature extraction and to prove the efficiency of the algorithm [9].

3. METHODOLOGY AND PROPOSED WORK

We follow three steps for the proposed CAD system these are: Pre-processing for removal of noise and identifying Region of Interest (ROI), feature extraction using GLCM and classification using RBFNN classifier.
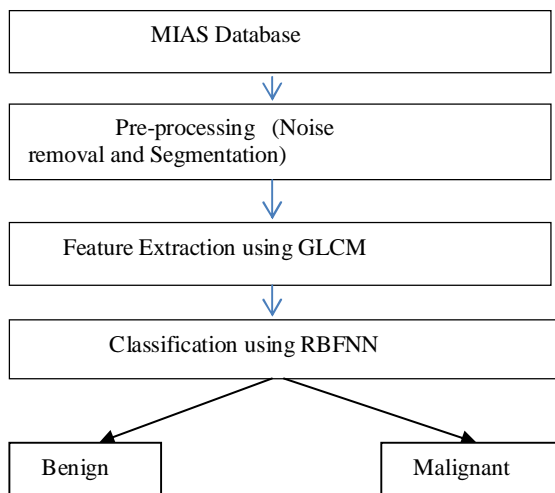

Fig. 1 An overview of proposed method

3.1. Pre-Processing

In pre-processing step we improve the quality of the image by removing noise. The pre- processing is important to eliminate some elements that are not connecting to the region of interest. So, the second step of pre-processing is removing the background area, removing the pectoral muscle and rib portion from the breast area because we are using mammograms of MLO view. And the last is segmentation to extract ROIs containing all masses and locate the suspicious mass from the ROI. Segmentation of the suspicious regions on a mammographic image is designed to have a very high sensitivity and a large number of false positives are acceptable. This step proves to be a critical success factor in the CAD system and helps in suppressing distortions and neglecting those parts which are not part of breast and this ensures a better

classification accuracy of CAD algorithm [11].

3.2. Feature Extraction

Feature Extraction is most important step, uses for extracting value of different-different features and plays a key role in pattern classification. It is a method of capturing visual content of images for indexing & retrieval. Image features can be either general features, such as extraction of color, texture and shape or domain specific features. Mammogram images consist of heterogeneous information that shows different types of tissues, blood vessels, glandular ducts, breast edges and mammogram machine characteristics. So to have better approach for detecting normal and abnormal tissues, we have to choose such a system which gives better result with more accuracy. Here we are using GLCM method for extracting the textural features.

3.2.1. GLCM (Grey Level Co-Occurrence Matrix)
It is widely used in much texture analysis application. The GLCM is table representation of how frequently various combinations of pixel brightness values (grey levels) occur in an image. Depending on the number of pixels or dots in each combination, Feature extraction based on GLCM is the second-order statistics that can be used to analyzing image as a texture. According to the number of intensity points (pixels) in each combination, statistics are classified into first-order, second-order and higher-order statistics. Harlick has defined 14 features [12], in most of the CAD system only 5 or 7 features have been used. Where using 5 features system gives 96% accuracy of 50 test cases and using 7 features system gives 95.50% accuracy of 112 test cases [9, 10]. In the proposed work, we use 8 features namely Contrast, Correlation, Auto-correlation, Sum Entropy, Variance, Information Measure of Correlation, Inverse Difference Momentum, Difference Entropy to measure the accuracy and other performance factors.

3.3. Classification

In proposed CAD technique we use artificial neural network for the classification of breast cancer in benign and malignant. Due to machine vision perception there is always a problem in detection of masses in digital mammography. This problem is solved to a great extent in this CAD, using Artificial Neural Network [5]. Neural network classification consists of two processes: Training and Testing. Neural network is the best tool in pattern classification application. The classifier is trained and then tested on mammogram image. The classification accuracy depends on training because it generates output based on its past experience of training when an unknown input is being passed.

3.3.1. RBFNN (Radial Basis Function Neural Network)
In order to find the potential micro calcification pixels based on the above mentioned features, a proper classification method must be used [4]. In our study, the classifier chosen is Radial

basis function neural network [8]. Neural network contains three layers: input layer, hidden layer and output layer [13,14]. The designing of neural network consist of input layer, hidden layer, output layer and activation function. Where input layer is the number of features extracted in feature extraction step, hidden layer depends on input and hardware configuration of the system and output layer declare whether it is benign or malignant.

4. PERFORMANCE ANALYSIS

Performance of CAD system is analyzed in terms of Accuracy, Sensitivity and Specificity. Where accuracy measures the standard of the binary classification. It takes into account of two classes these are true and false positives and negatives. Accuracy is generally estimated with balanced measure. Sensitivity/Precision deals with only positive cases and specificity deals with only negative cases.

Table 1: Measure Formula

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

A confusion matrix is the tabulation representation of actual and predicted classifications done by a neural network classifier [15]. Performance of such systems is commonly evaluated using the data in the matrix. The following table shows the confusion matrix for a two class classifier.

Table 2: Confusion Matrix

Actual	Predicted	
	Normal(Negative)	Cancer(Positive)
Normal	TN	FN
Cancer	FP	TP

The true positive rate (TP) is the amount of positive cases that were correctly identified. The false positive rate (FP) is the amount of negatives cases means normal cases that were incorrectly classified as positive/cancer. The true negative rate (TN) is defined as the amount of negatives cases that were classified correctly and the last is false negative rate (FN) which is the amount of positives cases means cancer that were incorrectly classified as negative/normal.

5. Experimental Result and Conclusion

The developed algorithm helps the radiologist in decision making whether the mammograms is normal or abnormal. This section details the result of classification on mammograms using GLCM features and RBFNN. In order to analyze this work, experiment is done on MIAS database. This proposed method is trained with 112 mammograms (56 benign, 56 malignant) and then tested. Started with pre-processing then pass the pre-processed images for extracting 8 Harlick's features. These extracted features are passed to neural network as an input known as 8 units in input layer, hidden layer (111 units, based on input layer) and 1 unit in output layer, which provides output either benign or malignant.

Table 3: Confusion Matrix for Tested database

Actual	Predicted	
	Normal	Cancer
Normal	56(TN)	3(FN)
Cancer	0(FP)	53(TP)

Table 4: Performance measure

Test Cases	Accuracy	Specificity	Sensitivity
112	97.3%	100%	94%

Table 3 shows the confusion matrix for tested database and table 4 shows the performance measure. During training system has identified all the benign images correctly but cannot identify all malignant images correctly, because of which accuracy goes down. The classification accuracy obtained by using 8 features was 97.3% whereas sensitivity and specificity were 94% and 100%.

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 work has investigated a classification of mammogram images using GLCM features. The maximum accuracy rate for normal and cancer classification is 97. 3% by using 8 extracted features. For future work, we can use more than one feature extraction technique like, extract feature using GLCM and then pass it to another feature extraction technique. So, we can get new feature. Using proper feature selection method accuracy may be improved efficiently.

6. REFERENCES

[1] 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.

[2] Atam P. Dhawan, YateenChitre, Christine Bonasso and Kevin Wheeler"; Radial-Basis-Function Based Classification of Mammographic Microcalcifications Using Texture Features; Engineering in Medicine and Biology Society, 1995., IEEE 17th Annual Conference (Volume:1) 535 - 536 vol.1

[3] 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.

[4] I Claristoyintzni, E. Derimatas ,G. Kokkinakis; Neural Classification Of Abnormal Tissue In Digital Mammography Using Statistical Features Of The Texture; Electronics, Circuits and Systems, 1999. Proceedings of ICECS '99. The 6th IEEE International Conference on (Volume:1)

[5] J. Jiang , P. Trundle, J. Ren ; Medical image analysis with artificial neural networks; Computerized Medical Imaging and Graphics;Volume 34, Issue 8, December 2010, Pages 617–631

[6] Jinshan Tang, Rangaraj. M. Rangayyan, Issam El Naqa, Yongyi Yang. "Computer-aided detection and diagnosis of breast cancer with mammography: Recent Advancements," *IEEE Trans. Biomed. Eng*, vol 13.No 2,March 2009.

[7] Khalid bashir, Anuj Sharma." *Review Paper on Classification onMammography",International* Journal of Engineering Trends and Technology (IJETT) ,Volume 14 Number 4 – Aug 2014.

[8] Meng Joo Er,.Shiqian Wu, Juwei Lu and Hock Lye Toh, Face Recognition with Radial Basis Function(RBF) Neural Networks," IEEE Trans. on Neural Networks, vol. 13, no.3, pp: 697 –710, May 2002.

[9] Neha Tripathi, Supriya P. Panda, "A Review on Textural Features Based Computer Aided Diagnosis System for Mammogram Mass Classification Using GLCM and RBFNN", IJETT, Volume 17 November 9- November 14.

[10] Nithya R. , Santhi B., "Classification of Normal and Abnormal Patterns in

Digital Mammograms for Diagnosis of Breast Cancer",IJCA, 0975 – 8887, Volume 28– No.6, August 2011.

[11] NorliaMdYusof, NorAshidi Mat Isa and HarsaAmylia Mat Sakim. "Computer-Aided Detection and Diagnosis for Microcalcifications in Mammogram: A Review,'' IJCSNS International Journal of Computer Science and Network Security, VOL.7 No.6, June 2007

[12] Robert M. Harlick,K.Shannugam,Its'hakDinsten, "Textural features for image classification"IEEE,Transaction on Systems Man &Cybernetics,Vol.SMC-3,No-6,Nov-1973,pp 610-621

[13] 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.

[14] 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.

[15] Sheshadri HS, Kandaswamy A, "*Detection of breast cancer by mammogram Image segmentation*"J Cancer Res Ther - December 2005 - Volume 1 – Issue, 4,232-234