www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 3, Issue 11 November, 2014 Page No. 9097-9011

Approaches For Automated Detection And Classification Of Masses In Mammograms

C. Rekha[1], G. Gayathri[2]

[1]Assistant Professor, Department of Computer Science,
Alagappa University, R.D.Govt Arts College, Sivagangai

[2]M.Phil scholar, Alagappa University,
R.D.Govt Arts College, Sivagangai
gaya3aug@rediffmail.com

Abstract: ***Breast cancer is one of the most common cancer among women around the world. Several techniques are available for detection of breast cancer. Mammography is one of the most effective tools for early detection. The goal of this research is to increase the diagnostic accuracy of image processing and machine learning techniques for optimum classification between normal and abnormalities in digital mammograms. GLCM texture feature extractions are known to be the most common and powerful techniques for texture analysis. This paper presents an evaluation and comparison of the performance of two different classification methods used to classify the normal and abnormal patterns. The experimental result suggest that Artificial Neural Network is outperformed the other method.***

Keywords: ANN, GLCM, KNN, PSO.

1. Introduction

Breast cancer is a public health problem. The two types of breast positions are left and right. Breast cancer categorized into two such as normal and abnormal. Normal pattern typically have smooth surface. Abnormal pattern presents rough and complex surface. Abnormal cases are divided into two types[1]. There are mass and calcification. Masses are identified by their shape and margin characteristics. Calcification are small calcium deposit and appear as group of bright spot in mammogram. The recent advancement in medical field and more precisely the involvement of information technology in the medical field. It introduces a new diagnosis mechanism called Medical Image Processing. It is not only limited to the cancer disease, instead it has helps to greatly in the diagnosis of several kinds of diseases and it is evident through statistics. It has become easier to detect cancer from an infected breast and diagnose the breast cancer. The early detection can help in proper diagnosis and treatment resulting in minimizing the risk of most unwanted outcome of this death. Several techniques are available such as mammography, ultrasonography, MRI, where mammography is the most common tool. Mammography helps to detect characteristic of breast cancer lesions[2].

CAD system intends to provide the assistance to mammography detection, reducing misdiagnosis, and consequently allowing better treatment and prognosis. Several types of features are extracted from mammogram includes region based, shape based, texture based, position based and color based features. In this paper GLCM texture features are used to extract the features from mammogram. The effective feature selection method used to select the relevant feature and finally feed into the input of different classifier and then classify the breast cancer easily.

This paper is organized as follows: Section 2 summarizes the existing research. Section 3 is describes the proposed method. Section 4 is describes performance measures. Section 5 is describes the experiment results. Finally, in Section 6 is describes conclusion and future work.

2. Related Work

The numbers of researches are conducted in the area of breast cancer detection and classification. Nor Ashidi Mat Isa [3],The image enhancement technique is employed to improve the contrast of the tissues. Then constraint region growing based on local statistical texture analysis is applied to detect and segment out the mass from the mammograms. Sulochana Wadhwani [4], used ANN soft computing method for detect the cancer and easily differentiate the benign and malignant. Belal K. Elfarra [5], to enhance and introduce a new method for feature extraction and selection in order to build a CADx model to discriminate between cancers, benign, and healthy parenchyma. Nithya [6] compare the three types of texture feature extraction method. The results are proving that GLCM features based neural network is giving higher classification rate of 98%. A.Alofe et al, [7], use simple and effective feature selection method used to selelct the essential feature. The obtained classification accuracy was 100%. Mohd. Khuzi [8], used Otsu’s method with GLCM features for classification of masses obtained good. Aswini Kumar Mohanty [9], GLCM and GLRLM features are used to distinguishing malignant and benign mass with an accuracy rate is 104%.

3. Proposed Method

The proposed system consists of six major steps. The overview of proposed methodology is depicted in Fig.1.

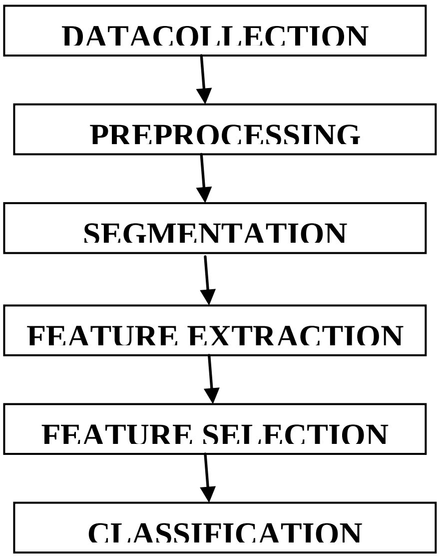

Figure 1: Proposed methodology

3.1 Data Collection

Digital mammograms are used as the standard inputs into the proposed method. The data used in this research is obtained from MIAS database. Images are digitized at 200 micron pixel edge, with a size of 1024 × 1024 pixels. Each pixel in the grayscale mammogram image represents the pixel intensity in the range of [0, 255] (8-bit).

3.2 Preprocessing

The pre-processing techniques are necessary, to remove the noise and to enhance the quality of the image. The mammogram images are preprocessed by using median filter and adaptive histogram equalization. The median filtering process removes the unwanted pixels(noises) in the mammogram images which helps in improving the performance of the process. Histogram equalization process helps to equalize the intensity of the image which will be helpful to identify the objects present in the image. The Fig. 2 shows that the after preprocessed image.

3.3 Segmentation

Segmentation is very difficult in the image processing[10]. Several segmentation techniques and algorithms have been developed for segmentation. The breast masses are segmented from the preprocessed images using Seeded Region Growing algorithm. The segmentation algorithm segments the breast masses from the image based on the clustered result of group of pixels in the image. For SRG, seed or a set of seeds can be quick, fairly robust and parameter free except for its dependency on the order of pixel processing.

In order to implement SRG a seed needs to be placed inside the pectoral muscle of the grayscale image. The following four steps are applied in the SRG process. 1) The region is iteratively grown by comparing all unallocated neighboring pixels to the region. 2) The difference between the pixel of interests intensity value and the region's mean is used as a measure of similarity. 3) The pixel with the smallest difference measure is allocated to the respective region. 4) The process stops when the intensity difference between the region mean and the new pixel become larger than the threshold value. The results of the segmented pectoral muscle obtained after the SRG process, is a binary image. Features cannot be directly computed from the segmented mammogram images. ROI sample is selected from segmented images after the segmentation process. The ROI of size 128 x 128 pixels is extracted with mass centered in the window. So, features need to be computed only from ROI sample, while excluding all other unimportant parts of the breast tissue.

3.4 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 the Gray Level co-occurrence matrix(GLCM) [11]. A gray level co-occurrence matrix (GLCM) contains information about the positions of pixels having similar gray level values. A co-occurrence matrix is a two-dimensional array, in which both the rows and the columns represent a set of possible image values. GLCM features are calculated in four directions which are 0^{0},45^{0},90^{0},145^{0} and four distances(1,2,3,4). GLCM calculates the probability of a pixel with the gray level of i occurring in a specific spatial relationship to a pixel with the value of j. The number of gray levels in the image determines the size of the GLCM. We have to extract the 12 different statistical features like 1.Contrast, 2.Correlation, 3.Cluster prominence, 4.Cluster shade, 5.Dissimilarity, 6.Energy, 7.Entropy, 8.Homogeneity, 9.Maximum probability, 10.Sum of squares, 11.Auto correlation, 12. Inverse different Moment. After features are extracted a feature selection is needed to extract an optimal subset of features for classification.

3.5 Feature Selection

Feature selection is important step in breast cancer classification. All features are extracted and not all are taken as input to the classifier. The main goal of feature selection is to reduce the dimensionality by eliminating irrelevant features and selecting the best discriminative features[12]. The advantage of feature selection is to improve the classification performance, to limit the number of input features, to achieve optimum accuracy. From the extracted features best features are selected using PSO algorithm. Then, a fitness function is evaluated to guide the search and select an appropriate classification system such that the number of incorrectly classified patterns is minimized. The basic idea of this algorithm is to identify features that are dissimilar between normal and abnormal pattern. Using this method, top five GLCM features are selected. The following procedure can be used to implementing the PSO algorithm:

1. Initialize the swarm by assigning a random position in the problem hyperspace to each particle.
2. Evaluate the fitness function for each particle.
3. For each and every individual particle, and then compare the particle's fitness value with its P_{best}. If current value is better than the P_{best}value, and then set this value as the Pbest and the current particle's position,value of xi as Pi.
4. Identify the particle that has the best fitness value. The fitness function value is identified as g_{best} and its position value as Pg.
5. Update the velocities and positions of all the particles using (1) and (2).
6. Repeat steps 2–5 until a stopping criterion is met t (e.g., maximum number of iterations or a sufficiently good fit ness value).

3.6 Classification

Classification is the final step in mammogram abnormality detection. The method has the objective to classify each image of one of two classes the cancer and normal. Classifications have the assignment to an unknown pattern of a predefined class, according to the pattern presented in the form of a feature vector. There are many classification techniques exist. After feature extractions of the mammogram images and classify the cancer and normal it helps to predict the texture feature it plays an important role in classification. Two different classifiers are used to classify the mammogram such as Artificial Neural Network(ANN) and K-Nearest Neighbor (KNN). The extracted features are considered as input to the classifier. The input features are normalized between 1 and 3. The desired output was specified as 1 for non-cancerous, 2 for cancerous in benign state and 3 for cancerous in malignant state.

Artificial neural networks are generally presented as systems of interconnected " neurons " which can compute values from inputs. The algorithm uses a Feed-Forward Neural Network. The representation of neural network with 'n' inputs, 'm' hidden units and one output unit. The extracted features are considered as input to the neural classifier. A neural network is a set of connected input/output units in which each connection has a weight associated with it. The neural network trained by adjusting the weights so as to be able to predict the correct class.

The K-Nearest Neighbours algorithm (k-NN) is a non-parametric method used for classification. It computes the distance from the unlabeled data to every training data point and selects the best k neighbors with the shortest instance. Suppose, given some data instance which belongs to one of the two categories or a class, and the goal is to determine which class the new data belongs to, is the problem of classification. There is no requirement for training process which actually makes this classifiers implementation as simple. If k = 1, then the object is simply assigned to the class of that single nearest neighbour. The accuracy, sensitivity, specificity of the classification is depends on the efficiency of the training.

4. Performance Measure

Three performance measure terms Accuracy (AC), Sensitivity (SE) and Specificity (SP) are used to evaluate the performance of the classifier. The respective formula are given in Table 1. Sensitivity is a proportion of positive cases that are well detected by the test and the specificity is a proportion of negative cases that are well detected by the test. Classification accuracy is depends on the number of samples correctly classified.

Table 1: Formula for Measures

Measures	**Formula**
Sensitivity	SE=TP/ (TP+FN)
Specificity	SP= TN / (TN+FP)
Accuracy	AC=(TP+TN)/ (TP+FP+TN+FN)

where, TP is the number of true positives; FP is the number of false positives; TN is the number of true negatives; FN is the number of false negatives. Confusion matrix is shown in Table 2. The highest value of both sensitivity and specificity shows better performance of the system.

Table 2: Confusion Matrix

Actual	Predicted	
	Positive	Negative
Positive	TP	FP
Negative	FN	TN

TP- predicts abnormal as abnormal.
FP- predicts abnormal as normal.
TN- predicts normal as normal.
FN- predicts normal as abnormal.

5. Experiment Result

For implementation among the MIAS datasets of 322 mammogram images, fatty, dense and glandular tissue images of Normal and abnormal severity of benign or malignant 1024X1024 pixel images are considered. These images are normalized to 256X256 pixel of Region of Interest and stored as trained and testing set respectively. Pre-processing and segmentation is applied to remove of noise and artifact and improve the efficiency of the images. In this analysis first is extract the features from mammograms. GLCM features are extracted from the mammograms and then selected features are input fed to the classifier. The proposed algorithm is run using Matlab software. Table 3 shows confusion matrix for this experiment. The performance measures are done by computed output. The evaluation results are presented in Table 4.

The result in Table 4 shows that different classifier methods used to discriminate normal and abnormal cases with different accuracy. In Artificial Neural Network classifier method, result shows that accuracy is 97%, specificity is 93%, sensitivity is 100%. It reveals that better classification rate in accuracy. The Fig. 3 shows that the relative performance measures. The Artificial Neural Network is outperformed the other method.

Table 3: Confusion matrix for Testing

Actual	Predicted		
	Benign	Malignant	Normal
Benign	7(TP)	3	1(FP)
Malignant	2	13	5
Normal	0(FN)	3	28(TN)

Table 4: Evaluation Result

Measures	ANN	KNN
Sensitivity	100%	100%
Specificity	93%	91%
Accuracy	97%	95%
F - Measure	97.5%	93.76%
Precision	97%	95.6%
Recall	98%	92.54%

6. Conclusion

The digital mammogram images are taken from MIAS database. The proposed algorithm is automated since the process segments the breast masses automatically. This paper investigated a classification of mammogram using GLCM texture feature. The simple and effective feature selection methods used to select the relevant feature and then fed into the classifier, finally classified as normal, benign and malignant. The experimental result shows that when compared to other methods, Artificial Neural Network Classifier shows better accuracy rate is 97%. In future, other texture based techniques can be evaluated in this research and to perform a comparative study.

References

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

[2] Michael A.Yachoub, A.S.Mohamed and Yasser M.Kadah,, ”A CAD system for the detection of malignant patterns in digitized mammogram films,” CARIO International Biomedical Engineering Conference, 2006.

[3] Nor Ashidi Mat Isa, Ting Shyue Siong, “Automatic Segmentation and Detection of Mass in Digital Mammograms,” Recent Researches in Communications, Signals and Information Technology, pp.143-146.H.H. Crokell, “Specialization and International Competitiveness,” in Managing the Multinational Subsidiary, H. Etemad and L. S, Sulude (eds.), Croom-Helm, London, 1986. (book chapter style)

[4] Sulochana Wadhwani, A.K.Wadhwani, Monika Saraswat, “Classification of Breast Cancer Detection Using Artificial Neural Networks,” Current Research in Engineering, Science and Technology (CREST) Journals, Vol.1, Issue 3, pp.85-91, 2013.

[5] Belal K. Elfarra and Ibrahim S. I. Abuhaiba, “Mammogram Computer Aided Diagnosis,” International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 5, No. 4, pp.1-30, 2012.

[6] Nithya .R, and Shanthi. B, “Comparative Study On Feature Extraction Method For Breast Cancer Classification,” Journal Of Theoretical And Applied Information Technology, Vol. 33 No.2, pp.220-226, 2011.

[7] Mohamed A.Alolfe, “Feature selection in computer aided diagnostic system for microcalcification detection in digital mammograms,” 26th National Radio Science Conference, 2009.

[8] A Mohd. Khuzi, R Besar,WMD Wan Zaki, NN Ahmad, “Identification Of Masses In Digital Mammogram Using Gray Level Co-Occurrence Matrices,” Biomedical Imaging and Intervention Journal, 5(3):e17, pp.1-13, 2009.

[9] Aswini Kumar Mohanty, Swapnasikta Beberta, and Saroj Kumar Lenka, “Classifying Benign and Malignant Mass using GLCM and GLRLM based Texture Features from Mammogram,” International Journal of Engineering Research and Applications, Vol. 1, Issue 3, pp.687-693.

[10] Indra Kanta Maitra, Sanjay Nag And Prof. Samir K. Bandyopadhyay, “Automated Digital Mammogram Segmentation For Detection Of Abnormal Masses Using Binary Homogeneity Enhancement Algorithm,” Indian Journal Of Computer Science And Engineering, Vol. 2 No. 3, pp.416-427, 2011.

[11] Haralick,R.M., Shanmugam,K., Dinstein,I., “Textural features for image classification,” IEEE Trans Sys ManCyb, pp.610—21, 1973.

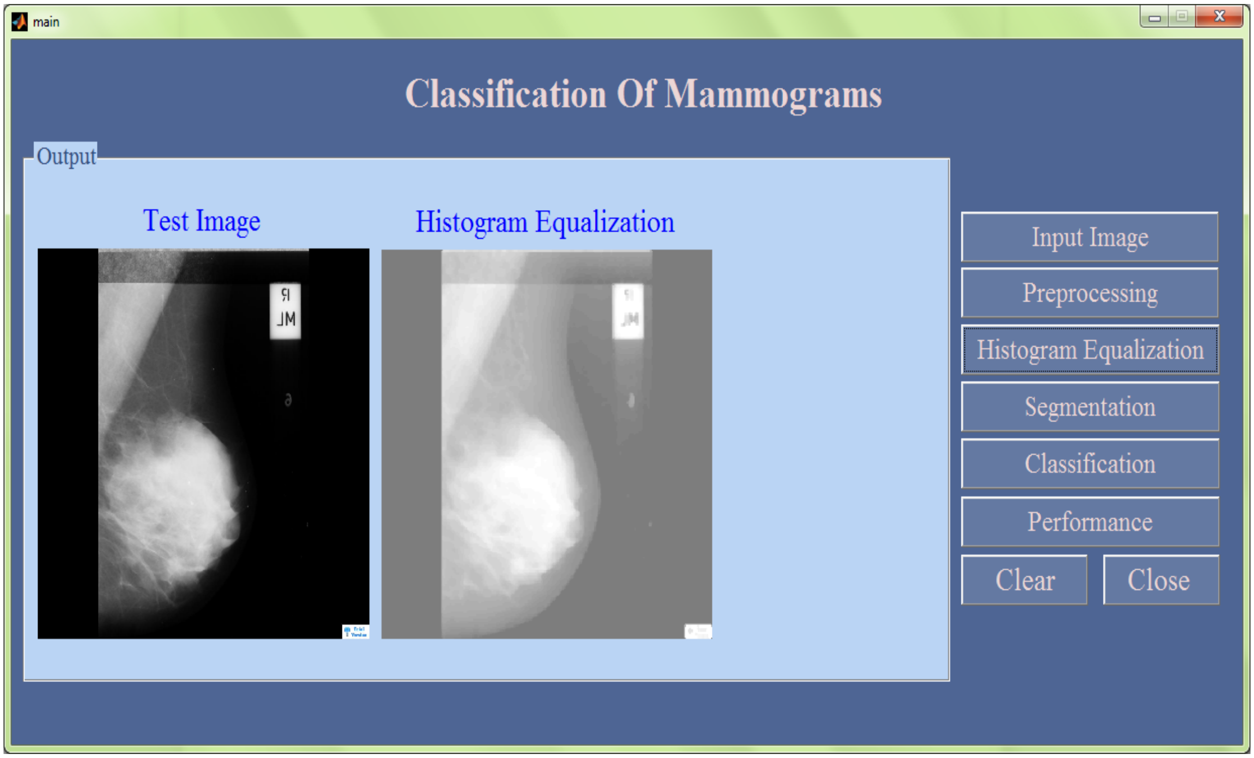

Figure 2: After Preprocessed image

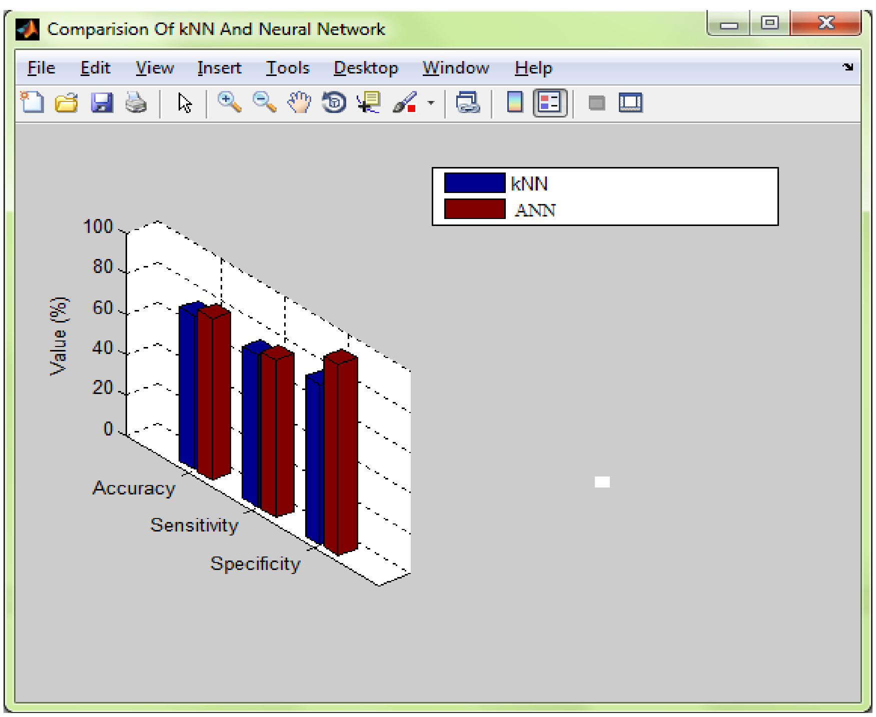

Figure 3: Relative Performance