ISSN:0975-9646

Nidhi Singh et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 8 (6) , 2017,619-622

Recent Trends on various techniques for Feature Selection and Classification in Mammogram mass images.

Nidhi Singh[1], Dr. S. Veenadhari[2]

[1]Department of Computer Science & Engineering, Lakshmi Narain Institute of Technology, Bhopal, M.P, India
[2]Department of Computer Science, Aisect University, Bhopal, M.P, India

Abstract—**In order to ease radiologist's assessment for identification or classification of mammogram images, various techniques are proposed by researchers for the past two decades. Early mammography based breast cancer detection method depends on the production of excellent images qualities and expert analysis. In order to prove the affected area's abnormality as benign/malignant, the tissue must be removed from the affected portion and tested by using breast biopsy techniques. Diagnosis using mammograms is intended at classifying the detected cancerous regions as benign or malignant. Comprehensive literature survey has been done in connection with this research work and salient features of some of the relevant ones are outlined here. The review of literature given in this paper is centered upon various techniques for mammogram classification.**

Index Terms—**classification, Feature Extraction, mammogram.**

I. INTRODUCTION

Cancer is one of the disease problems which is made of various related diseases. In each and every cancer types, body's cells begin to divide involuntarily and spread in to surrounding tissue. Likewise, breast cancer arises from breast cell tissues. Breast is made by billions of microscopic cells. These cells start to multiply uncontrollably causing breast cancer. Breast cancer can be view as two types first one is ductal carcinoma and another is lobular carcinoma. A ductal carcinoma is the most common cancer that begins in the milk duct and lobular carcinoma is the cancer that begins in the lobules [1]. At the earlier times, the diagnosis and treatment proves to be ruinous without efficient techniques. At each stage, the death rate and the vigorosity of this disease is increasing. to reduce the death rate and minimize vigorosity of disease, there is a need of early breast cancer detection techniques. So automated computer aided detection is unavoidable. There are no causes of breast cancer, we can point it as only risk factors. It may be genetic or environmental. Genetic factors include family history, personal health history, menstrual and reproductive history, dense breast tissue, certain genome changes, age, gender etc. The environmental factors include obesity, poor diet, alcohol consumption, radiation, lack of physical activity etc [2]. The initial symptom of a breast cancer is the formation of a lump. This is due to tiny deposits of calcium called micro calcifications and tumors called circumscribed masses. These tumors are generally Benign and malignant. The benign tumors are generally non aggressive and non cancerous. They will not spread to other body parts [3]. For early detection of breast cancer there are various different imaging techniques are possible. These include MRI, X ray imaging, ultrasound imaging, digital mammography, screening etc. The digital mammography is widely used nowadays due to its advantages over others. X ray imaging is usually used for finding the signs of the cancer. While using a mammogram generally investigate the problem. Mammogram uses X rays to create images of the breast. Earlier there were film mammography's exists in which the images were stored on films, but now a day digital mammograms are widely used because they are captured and stored directly on digital computers as well as all the corners and nooks are visible for easy detection. In breast ultrasound the images are created using sound waves. But this is not used nowadays as it is done with a handheld device, it will generate false positives and false negatives when the person who operates it is not well experienced or skilled and thus the quality of the image will vary. The detailed detection of the images is done using feature extraction and texture extraction techniques. This itself opens another research area as a wide variety of techniques have been used for segmentation, feature extraction, enhancement. Done mainly by wavelet techniques, clustering, using GLCM matrix etc which are described clearly in the related works. So the ultimate aim of this survey is to provide different enhancement, detection and classification techniques for early breast cancer detection.

II. CLASSIFICATION TECHNIQUES

3.2.1 PSOWNN

PSO has been used for optimization problems. The optimal features of the images are selected and then classify the images based on wavelet neural network.

3.2.2 SVM Classifier

It is a machine learning method which uses a hyper plane that maximizes the margin in the training data to classify binary classes. Support vectors are the training data along the hyper plane. The distance between the support vectors and the class boundary is the margin. The decision planes that define decision boundaries are the basic idea of SVM.

3.2.3 K Means Clustering

The basic idea of k means clustering is that the n dataset is partitioned in to k clusters, in which each extracted image belongs to the cluster with nearest mean i.e., mean is taken as the criteria. Likewise all the images are classified.

3.2.4 SRAN Classifier

In the SRAN system, the training sample record arrives one by one and the network adapts its parameters on the basis of the difference in knowledge between the network and the current sample record. Uses basic concept of RBFN.

3.2.5 Probabilistic Neural Network

PNN is a form of Probabilistic density function; training is very easy and fast for PNN. The computational load in the training phase is transferred to the evaluation phase in PNN which makes it different from other classifiers.

III. Literature Review

In 2000, S. Baeg and N. Kehtarnavaz [4] proposed tow novel texture image classifier to classify the abnormalities in the mammogram. The first texture feature provides a measure of smoothness/denseness and is obtained by applying a morphological operator to maxima/minima image points. The second texture feature reflects a measure of architectural distortion and is derived from image gradients. For classifying the masses a three layer BPN is used as the classifier. 150 images are taken for the classification which results with ROC of 0.91.

In (R J Nandi et al. [5], 2006), five feature selection methods, including three statistical measures such as Student's t-test, Kolmogorov-Smirnov Test, and Kullback-Leibler Divergence are explained in order to refine the pool of features available. Both the training and test accuracies obtained are high: above 99.5% for training and typically above 98% for test experiments. A leave one out experiment shows 97.3% success in the classification of benign masses and 95.0% success in the classification of malignant tumors. A shape feature known as fractional concavity is found to be the most important among those tested, since it is automatically selected by the genetic programming classifier in almost every experiment.

In 2009 Pelin Gorgel et al. [6] proposed DWT and SVM based mammogram mass classification technique that classifies the masses as benign or malignant. They have applied their method in two stages: First, feature extraction by computing the wavelet coefficients and Second, classification using the classifier trained on the extracted features. 66 images are taken for the research. Experimental test performed on mammogram showed that 84.8% classification accuracy is achieved by using the SVM with RBF kernel. Also to show the classification performance of SVM confusion matrix, accuracy, sensitivity and specificity analysis with different kernel are used.

In order to detect the masses, K-means and co-occurrence matrix is described in by Leonardo de Oliveira Martins et al. [7] in 2009 and SVM classifier is used. For classifying, images are separated into two groups based on shape and texture descriptor i.e. masses and non masses. Eccentricity, circularity and convexity are used as Shape descriptor. For extracting the descriptors 4*4 window size is used to allow the calculation of co-occurrence. 85% of accuracy is achieved by this method.

In 2010 Mohammed J. Islam et al. [8] proposed a method to classify the mass in mammogram images.ANN is used to classify the masses, which performs benign/malignant classification on ROI that contains mass in image. Texture is characterized by the textural features mean, SD, entropy, skewness, kurtosis and uniformity. This method is used to improve the efficiency and effectiveness. For classifying the masses seven features of ANN is proposed. 90.91% and 83.87 % is achieved rate of both sensitivity and specificity respectively.

In 2010 Fatemeh Saki and Amir Tahmasbi [9], has proposed a novel opposition based classifier which classify breast masses into benign and malignant categories. Their aim was to improving the convergence rate of MLP learning rules as well as increasing the mass diagnostic performance A multilayer perceptron network with a novel learning rule, called Opposite Weighted Back Propagation (OWBP) is utilized as the classifier. The features include circularity, Zernike moments, contrast, average gray level, round, nodular and stellate derivatives and exact shape of the contour of the masses. It evaluated the classifier has been trained with both traditional back propagation and OWBP learning rules.

In 2011, Huanping Zhao et al. [10] proposed a novel multi view information fusion algorithm based on multi agent method to improve the accuracy of classification of masses. 128 ROIs from DDSM database composed by 64 pairs of Craniocaudal view and mediolateral oblique view are chosen for the experiments. This demonstrated that the proposed algorithm improved the accuracy and reduced the false positive rates.

Comparative analysis of ANN and K Nearest Neighbor (KNN) classifiers based on Spherical wavelet transform (SWT) is proposed by Pelin Gorgel et al. [11] in 2013. Shape and pixel value features both from the coefficients obtained by the SWT algorithm and the original ROI image are extracted. These features are area, centroid, bounding box, major-minor axis length, eccentricity, orientation, filled area, Euler number, extrema, convex hull, equiv diameter, solidity, extent and finally the mean center-border distance. The calculated numeric features mentioned below provide feature matrix which is used as the input vector of the supervised learning system ANN and KNN. From MIAS database 60 abnormal images are taken for testing.

S. Deepa [12] et.al discussed the different ranges of Contourlet Coefficient Cooccurance Matrix features in the analysis of mammogram images and classification. The ROI is enhanced using histogram equalization and are reduced using contourlet transform. The cooccurance matrices are produced for various directions. Feature selection is done by Sequential Floating forward Selection Algorithm (SFFS). Classification is done through Probabilistic Neural network (PNN). The demerit is that the system is not automated. MIAS database is used providing 92.5% accuracy.

J. Dheeba et.al. [13] proposed a Particle Swarm Optimized Wavelet Neural Network (PSOWNN) to classify and detect breast cancer. WNN posses both wavelet and neural network properties, here the ROI detection is done through Global thresholding technique.WNN along with PSO proves to be best for classification as it decreased false negatives and false positives. From the mammographic images the laws texture energy measures are extracted with an abnormality detection algorithm. The textural features are extracted through a windowing operation and by convolution kernels applied to ROI.216 mammographic images obtained from mammographic screening centers have been used. Advantage is that this method increased convergence of back propagation algorithm error and the disturbances in learning are avoided.

K. Subashini et. al. [14] proposed breast cancer detection through ultrasound images. Wavelet domain techniques i.e., DWT translates the images in to wavelet coefficients to remove noise. Here the noise representing coefficients are suppressed and image features are enhanced. Segmentation is done with active contour model and the texture features are extracted using auto covariance coefficients. The back propagation neural network are used for classification which found better in performance than linear classifiers.

S. Julian Savari et.al [15] proposed a paper that uses histogram equalization to increase the image quality. The intensity features are extracted and computed to calculate volumetric values. The classification is done by K means clustering algorithm. Noise is removed by Gabor filter. The dataset used are MIAS and DDSM database. Provide 99% classification accuracy.

Neetha Jog et.al [16] used GLDM and Gabor features as extraction methods with SVM and KNN classifiers. Here MIAS database is used. Gabor feature along with wavelets are also used. Basic idea of SVM was to find an hyper plane to separate D dimensional data in to two classes finely. KNN classify objects based on nearest training samples in the feature space. Result showed that SVM+GLDM provide 95.83% accuracy than KNN+GLDM and SVM+Gabor filter has 71.83% accuracy than KNN+Gabor Filter.

IV. The quality assessment for Image segmentation

The performance of the classification algorithm is evaluated with the help of confusion matrix, ROC curve with AUC score, and other parameters like F-measure and Matthews's correlation coefficient (MCC).

A confusion matrix is a table that provides information about the predicted and actual class classification performed by the classifier. The confusion matrix for two classes (benign and malignant) and performance measures are given in Tables 1 and 2, respectively.

Table 1

Actual Class	**Predicted class**	
	Positive	***Negative***
Positive	TP (true positive)	FN (false negative)
Negative	FP (false positive)	TN (true negative)

Table 2

Measure	**Definition**
TPR or recall	TP/(TP + FN)
FPR	FP/(FP + TN)
precision	TP/(TP + FP)
ACC	(TP + TN/K)

The TPR (true positive rate) and FPR (false positive rate) are two important measures for performance evaluation. The TPR calculates malignant ROIs correctly classified out of the total number of malignant ROIs. The FPR parameter calculates benign ROIs incorrectly classified out of the total number of benign ROIs. The F-measure and MCC play an important role in the quality evaluation of binary classification. The F-measure is computed as the harmonic mean of 'precision' and 'recall' and given by

$$F - measure = \frac{2 \times recall \times precision}{recall + precision}$$

The MCC is a correlation coefficient between the observed and predicted classification and given by

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{(TP + FN)(TN + FP)(TP + FP)(TN + FN)}$$

The F-measure ranges from 0 to + 1 and MCC ranges from −1 to +1. Larger values of both F-measure and MCC indicate higher classification quality. The evaluation of a classifier performance can also be achieved by means of receiver operating characteristics (ROC) curve. ROC curve is represented in a 2D graph which plots TPR against FPR. The ROC curve has an important index value known as the area under the curve (AUC), which determines the classifier's performance. The AUC of value 1.0 is an ideal performance of the classifier.

V. Conclusion

It is evident from the literature survey that the classification or detection mass in digital mammograms is mainly based on texture and statistical based features. Though there are many approaches, multiresolution and multi directional analysis has recently been proposed as a new method for feature extraction and image representation other than wavelet analysis

This survey paper concludes that there are several techniques that deal with feature extraction and classification of diagnosing images that gave different accuracies. Also found that the mammographic images are giving better accuracy than ultrasound images, MRI images etc and most of the works used MIAS database which contain 322 mammographic images. But in future ultrasound images can be used as it is devoid of radiations. Of these a reliable and effective method has to be found out. The advancement of curvelets are a promising approach as it has got many advantages than wavelets. The use of SVM also increases the accuracy than any other classifiers.

REFERENCES

[1] Saurabh PrasadLori Mann BruceJohn E Ball," A Multi-classifier and Decision Fusion Framework for Robust Classification of Mammographic Masses" International Conference of the IEEE Engineering in Medicine and Biology , 2008.

[2] K.Subashini, K.Jeyanthi, 'Masses detection and classification in ultrasound images', IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS),Volume 9, Issue 3 Ver. II (May -Jun. 2014), PP 48-51 .

[3] Ojo J. A., Adepoju T. M., Omdiora E. O., Olabiyisi O. S. and Bello O. T, 'Pre-Processing Method for Extraction of Pectoral Muscle and Removal of Artefacts in Mammogram,' IOSR Journal of Computer Engineering (IOSR-JCE) e-Volume 16, Issue 3, Ver. V (May-Jun. 2014).

[4] R. J. Nandi,A. K. Nandi,R. M. Rangayyan,D. Scutt," Classifi- - cation of breast masses in mammograms using genetic programming and feature selection", Medical and Biological Engineering and Computing August 2006, Volume 44, Issue 8, pp 683–694.

[5] S. Baeg, N. Kehtarnavaz,"Texture based classification of mass abnormalities in mammograms", 13th IEEE Symposium on Computer-Based Medical Systems, 2000.

[6] GORGEL Pelin & SERTBA Ahmet & Kilic Niyazi & Ucan Osman , Mammographic mass classification using wavelet based support vector machine, JOURNAL OF ELECTRICAL & ELECTRONICS ENGINEERING ,Vol. 9. No. 1,2009.

[7] Leonardo de Oliveira Martins, E.C. da Silva, A.C. Silva, A. C. de Paiva," Classification of Breast Masses in Mammogram Images Using Ripley's K Function and Support Vector Machine", International Workshop on Machine Learning and Data Mining in Pattern Recognition ,MLDM 2009.

[8] Islam M.J., Ahmadi M., Sid-Ahmed M.A. (2010) Computer-Aided Detection and Classification of Masses in Digitized Mammograms Using Artificial Neural Network. Advances in Swarm Intelligence. ICSI 2010.

[9] F Saki, A Tahmasbi, SB Shokouhi ,"A novel opposition-based classifier for mass diagnosis in mammography images" 17th Iranian Conference of Biomedical Engineering (ICBME), 2010.

[10] Huanping Zhao,Weidong Xu,Lihua Li,Juan Zhang "Classification of Breast Masses Based on Multi-View Information Fusion Using Multi-Agent Method", 5th International Conference on Bioinformatics and Biomedical Engineering, (iCBBE) 2011 .

[11] *Pelin GÖRGEL, Ahmet SERTBAS, Ainura TURUSBEKOVA,"* Classification Of Breast Masses Using Anfis-Based Fuzzy Algorithms: A Comparative Study", IU-Journal of Electrical & Electronics Engineering,2013.

[12] S. Deepa, Dr.V.Subbiah Bharathi, 'Textural Feature Extraction and Classification of Mammogram Images using CCCM and PNN', IOSR Journal of Computer Engineering (IOSR-JCE) ,Volume 10, Issue 6 (May. - Jun. 2013) .

[13] J. Dheeba, N. Albert Singh, S. Tamil Selvi 'Computer-aided detection of breast cancer on mammograms: A swarm intelligence optimized wavelet neural network approach', Journal of Biomedical Informatics 49 (2014) 45–52, 2014 Elsevier Inc.

[14] K.Subashini, K.Jeyanthi, 'Masses detection and classification in ultrasound images', IOSR Journal of Pharmacy and Biological Sciences (IOSR-JPBS),Volume 9, Issue 3 Ver. II (May -Jun. 2014), PP 48-51

[15] S.Julian Savari Antony ,Dr.S.Ravi, 'A New Approach to Determine the Classification of Mammographic Image Using K-Means Clustering Algorithm', International Journal of Advancements in Research & Technology, Volume 4, Issue 2, February -2015 .

[16] Neeta Jog, Arvind Pandey, 'Implementation of Segmentation and Classification Techniques for Mammogram Images', IOSR Journal of Engineering (IOSRJEN), Vol. 05, Issue 02 (February. 2015) .