www.ioirp.com
International Journal of Innovative Research in Technology, Science & Engineering (IJIRTSE)
ISSN: 2395-5619, Volume – 1, Issue – 3. May 2015

Feature Extraction Based Breast Cancer Detection Using Mammogram on MRI Images

Anu.P[1], Roja.P[2], Seeja.V.M[3], Vigneshwar.N.D[4]
[1,2,3,4] Student, Department of ECE(ME-VLSI Design), Bannari Amman Institute of Technology, Erode, India.

***Abstract*— Early detection of breast cancer can improve survival rates to a great extent. Inter-observer and intra-observer errors occur frequently in analysis of medical images, given the high variability between interpretations of different radiologists. To offset this variability and to standardize the diagnostic procedures, efforts are being made to develop automated techniques for diagnosis and grading of breast cancer images. This review aims at providing an overview about recent advances and developments in the field of Computer Aided Diagnosis(CAD) of breast cancer using mammograms, specifically focusing on the mathematical aspects of the same, aiming to act as a mathematical primer for intermediates and experts in the field.**

Keywords*— Breast cancer, classifiers, Computer Aided Diagnosis(CAD), digital mammography, feature extraction techniques.

I. INTRODUCTION

CANCER refers to the uncontrolled multiplication of a group of cells in a particular location of the body. Breast cancer is any form of malignant tumor which develops from breast cells. Breast cancers are traditionally known to be one of the major causes of death among women. Mortality rates due to breast cancer have been reducing due to better diagnostic facilities and effective treatments. One of the leading methods for diagnosing breast cancer is screening mammography. This method involves X-ray imaging of the breast. Following the results of screening mammography, a follow up study is made for patients according to the level of suspicion of the abnormality. This stage is referred to as diagnostic mammography.

Early detection of breast cancer through screening and diagnostic mammography increases breast cancer treatment options and survival rates. Unfortunately, due to the human factor involved in the screening process, detection of suspicious abnormalities is prone to high degree of error. Studies have shown that radiologists have an error rate between 10%-30% for detection of cancer in screening studies.

A study by found that double reading of screening mammograms provided greater sensitivity than single reading without increasing recall rates. But, manpower is a major drawback with this approach. A major reason for these errors is due to the fact that radiologists depend on visual inspection. During manual screening of a large number of mammograms, radiologists may get easily missing out vital clues while studying the scans. As yet, there is no definitive literature which focuses on an elaborate discussion on the feature extraction, feature selection and classification methodologies used in breast cancer detection. The current study aims at filling this gap by documenting developments in that aspect.

II. COMPUTER-AIDED DIAGNOSIS SYSTEM PIPELINE

The pipeline used in a CAD system for breast cancer detection is consists of preprocessing, breast region segmentation, feature extraction and classification.

Preprocessing of mammograms is done to improve the contrast of mammograms which will be helpful in further stages of the detection pipeline. This step also includes denoising of the images. Segmenting the breast region from pectoral muscle and surrounding regions is carried out in order to make it easier to extract the suspicious tissues from breast segments. Feature extraction and classification steps are pattern recognition systems with not much of a difference.

The pipeline diagram for the Computer Aided Diagnosis system will be,

International Journal of Innovative Research in Technology, Science & Engineering (IJIRTSE)
ISSN: 2395-5619, Volume – 1, Issue – 3. May 2015

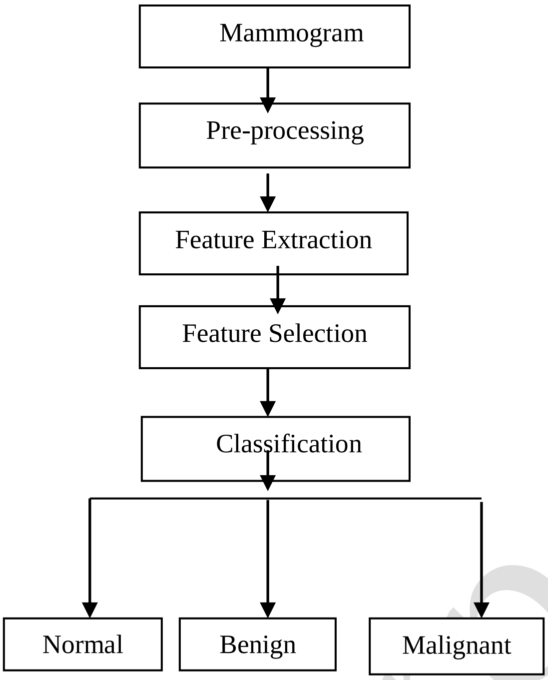

Figure 1. Computer-Aided Diagnosis system pipeline

III. PREPROCESSING OF MAMMOGRAMS

De noising and enhancement of mammograms are very important for both manual inspection stage and the computer-aided diagnosis second reading stage. Mammograms do not provide a very good contrast between normal glandular and malignant tissues. Because of using X-rays, the attenuation level of the contrast of an image does not give that much of efficiency. Image contrast can be obtained through the linear coefficients. This is mostly difficult for younger women with denser breast tissues.

For the enhancement process of the mammograms, the contrast level of an image has to be finding. Noise equalization is the solution to obtain the contrast level at all image intensities. The noise makes the detection of small and weak structures more difficult.

The enhancement from better noise estimation will be done by the following formulae. Let consider the image location as (x,y).

c(x,y) = f(x,y) – (x,y) (1)

where c(x,y) is the estimated local contrast, f(x,y) is the image gray level at (x,y) and (x,y) is the median gray level within the neighborhood of (x,y).

In the preprocessing, input image has to be involving in the filtering process. For this low pass filter and high filter can be used. The equation(1) can be equated to a high-pass spatial filter. The local contrast provides a measure of the high-frequency image noise.

Image enhancement is generally based wavelet transforms such as discrete wavelet transform and continuous wavelet transform.

Similarly the conventional contrast enhancement technique can be written by the formulae as follows,

θ(t+1):

= 0 (2)

The equation(2) will be assumed to be as differentiability. In that Q(θ;θ(t)) will be maximized. To maximize an image Expectation Maximization algorithm will be used. By using this algorithm probability density function(pdf) can be calculated to handle the missing data.

IV. DATABASE IMAGE PROCESSING

In the preprocessing the input image will be enhanced by using the wavelet transform. Similarly the database images also have to be segment for the comparison process. For the segmentation process the transforms will be used. Here the using transforms will be listed as,

Daubachies transform
Symlet transform
Bi orthogonal transform3.3
Bi orthogonal transform 3.5
Bi orthogonal transform 3.7

The segmentation is generally based on the pixels. Initially the input image will be involved in daubachies transform. There the segmentation process will be done. Some of the clusters may not be found by this transform.

Then the image will be involved in the symlet transform. There also the segmentation will be done. There some of the cluster pixels can be found.

Then the image will be involved in the bi orthogonal transforms. The three versions of this transform will be done combinable. There the remaining clusters will be found.

The Sum Average(SA) is found from the pixel in consideration and the size of the gray scale will be,

SA = (3)

V. FEATURE SELECTION

Class separability measures are used to estimate the best feature sets for a classification problem. It is necessary to carefully choose features since the choice of feature is critical in the final accuracy. The simplest method for this feature selection is done through dimensionality reduction using Principle Component Analysis(PCA).

The PCA will be done using eigen value decomposition on covariance matrices of the observed regions and finds their principle components. Once the principle components are found, the first *n* significant components can be extracted to reduce the dimensionality of the feature vector. Though PCA is a widely used method for dimensionality reduction.

The feature selection will be done by using the most efficient method called as *bottom-up* or a *top-bottom* approach. To select the single feature the correlation between the features has to be calculated. If is the *k*th feature of the *n*th pattern, with *n*=1,2,3,4,…,*N* and *k*=1,2,3,4,…,*m*, the cross correlation between any two of them is given by,

= (4)

This will be used to select the second best feature. This is only used most widely.

VI. CLASSIFICATION

After the patterns in a data have been extracted to form feature representations, classifiers can be developed using several approaches depending on the features available and the pattern classification problem. The classification is not a simple task. In previous CAD techniques the threshold method has been used. For the best result the nowadays neural network will be used. The mostly used classifiers are listed as below,

(i)Probability Neural Network
(ii)*K*-Nearest Neighbor
(iii)Support Vector Machine

www.ioirp.com

International Journal of Innovative Research in Technology, Science & Engineering (IJIRTSE)
ISSN: 2395-5619, Volume – 1, Issue – 3. May 2015

The main way to describe these classifiers is that, consider training pattern …. , with each of these being assigned to one of the two classes or , a weight vector w and a threshold is determined such that,

x + x (5)

where,

T is the threshold vector
w is the weight vector
is the threshold value

Comparing with this threshold technique *k*-means clustering is extremely used in mammographic analysis. This will provide a considerably good classification accuracy level. An improved normal *k*-means technique of clustering will be done with adaptive region growing.

VII. COMBINATION OF CLASSIFIERS

Even though the classification is done individually by using the different types of classifiers the combination of classifiers will give the accurate result. Yet another reason to combine classifiers is the fact that designing a classifier specific to the data set might not result in a good generalization of the classifier. This combining process involved in several schemes such as parallel, serial or cascading, hierarchal or trees combinations.

The average output by the combination of classifiers will be shown as the following formulae,

= (6)

where,

= (7)

This is the required equation for the classification process

VIII. RESULTS

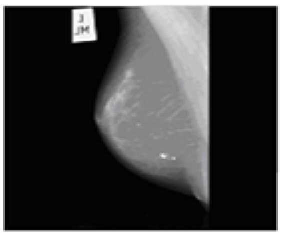

Figure 2.Input Image

Figure 3.Output Image

In the result, the classification have done by using the classifiers. In an output image the clusters have been pointed by the white pixels. The black pixels are used to describe the normal tissues. Thus by the result have been derived.

IX. CONCLUSION

Though the idea of using computer-aided detection is gaining popularity, it should not be missed that CAD techniques can serve only as a double-reading aid and cannot replace human readers. The main goal of CAD must be to increase diagnostic accuracy with advanced mathematical and computational techniques.

Though several advances in this respect have been made in the past 30 years, with mathematical advances and recent improvements in computing techniques and improved speed of computation. The previous techniques are very hard to implment in real-time setting, but now by using neural networks it can be easily implmented.

References

[1] Kathikeyen Ganesan, U.Rajendra Acharya, Chau Kuang Chua, 2013, “Computer-aided breast cancer detection using mammograms: a review,IEEE Review in Biomedical Engineering, vol. 6.

[2] M.F.Akay, “Support vector machines combined with feature selection for breast cancer diagnosis”, *Expert syst. With Applicant*., vol. 36., pp.3240-3247, March 2013.

[3] K.J. McLoughlin, P.J.Bones and N.Karssemeijer, “Noise equalization for detection of microcalcification clusters in direct mammogram images”,*IEEE Trans. Med. Imag.*, vol. 23, pp.313-320, March 2004.

[4] E. P. Simoncelli and E. H. Adelson, “Noise removal via Bayesian wavelet coring”, in *Proc, Int. Conf. Image Processing*, vol. 1, pp. 379-382, Sep. 1996.

[5] C.J. Vyborny, M. L. Giger and R.M. Nishikawa, “Computer aided detection and diagnosis of breast cancer”, *Radiologic Clinics N. Amer.*, vol. 38, pp. 725-740, July 2000.