CLASSIFICATION OF BENIGN AND MALIGNANT BREAST TUMORS IN DIGITAL MAMMOGRAMS USING DIFFERENT WAVELET TRANSFORMS

A . Krishna Veni, Sathish Kumar

M .Tech Student, Assistant Professor
Department CSE
ANU College of Engineering & Technology
Acharya Nagarjuna University, Guntur

Abstract: Breast cancer is the leading cause of most deaths in the world. This paper deals with classification of Breast cancer, i.e. benign or malignant based on coefficients extracted from multiresolution analysis based on five different wavelet functions Biorthogonal 3.5, coiflets 3, Daubechies 4, Symlets 3. In this paper 80 Region of Interest (ROI)s from the Mammographic image analysis society(MIAS). The coefficients which extract the texture information from the ROI's of Mammogram are given as an input to the KNN classifier. The performance of the system is evaluated using Receiver Operating Characteristic curve (ROC). Experimental results show that the area under the curve (AUC) is $A_\text{ROC}=0.90$.

Keywords: Breast Cancer, KNN, MIAS

I. INTRODUCTION

Breast cancer is a disease that is commonly found in the women. It is a tumor that grows in women breast cells by growing in the surrounding tissues of the body. Therefore mammogram play important role in early diagnosing of breast cancer. Mammography is X-ray imaging technique for diagnosis breast tumor. Abnormal cell multiplication and growing into a tumor is called breast cancer. In breast cancer, cancer cells form in the tissues of the breast of the women[1]. There are two types of breast cancers based on the growing cell characteristics i.e., Benign and Malignant. A tumor which is easy to remove and can be stopped spreading in other parts of the body is known as benign tumor. Malignant Tumor is the cancerous cells which are dangerous and grows aggressively and spread into the other parts of the body is known as malignant[2].

Cancer is significant public health problem in the world today. Most of the people die because of breast cancer than any other types of cancers such as lung, colon and prostate cancers. According to the IARC(International Agency for Research on Cancer) of the WHO (World Health organization).

Recent statistics of the World health organization indicate around 8.2 million deaths caused by each year because of this type of cancer 2012 and 27 million of new cases of this disease are expected before 2030.According to the American cancer society 215 990 new cases of breast carcinoma has been detected in the United States alone in 2004.It is one of the major reasons of deaths occurring due to cancer in women[3]. It is the fifth most common cause of death from cancer in women. The alternative way to reduce the number of death caused by breast cancer is by early detection [4].If the tumor is diagnosed as benign or malignant in the early stage, the patient can be given treatment to increase of life span.

Different types of images are used to detection and diagnose the cancer, such as diagnostic mammograms (Xray), Magnetic Resonance imaging(MRI). Ultrasound (Sonography),and Thermography have better clarity, low noise and less distortion.

Research work has been done on detecting the breast cancer tumor in early stage. Aathi et al.[5] used statistical features and these features are given as inputs to Support Vector Machine achieved an accuracy of 86.11%. Oliver et al.[6] proposed the nearest neighbor and decision tree classification used to classify the breast cancer abnormalities. Fatima et al.[7] presented the first order and gradient features combined with GLCM,DWT using SVM classifier. Jog and Mahadik[8] represented a grey level difference method (GLDM) Gabor feature extraction methods along with SVM and KNN classifiers. Campanini et al. presented an SVM classifier for mass detection in digital mammograms [9]. Sahiner et al. used four gray-level difference statistics (GLDS) texture features and convolution neural network for mass detection [10].

II. METHODOLOGY

Discrete wavelet transform:

To represent an image into several sub images and analyzes them in the frequency domain multi resolution analysis is used. While processing, texture extraction it is necessary to measure texture coefficients of neighborhood of different sizes using wavelet multiresolution transform which preserves the original image into sub bands that preserve high and low information.DWT of a signal $x(t)$ can be stated as a shifted version of scaling function $\phi_{j,k}$ and shifted version is called as the mother wavelet function $\psi_{j,k}$.Once mother wavelet is selected the wavelet transform can be used to decompose a signal according to the scale.

Discrete transform coefficients can be defined as:
The coefficients can be obtained by expanding \( f(x) \) into numbers. By iterating the single scale filter bank, multi scale filter banks can be generated by approximation of to the input of another filter. In the one dimensional case \( f(x, y) \) is used as input, which decompose an image into four sub bands. Where as in two level decomposition it further decompose into sub bands with low-low (LL), low-high (LH), high-low (HL) and high-high (HH).

\[
W_{\phi}(j_0, k) = \frac{1}{\sqrt{M}} \sum_{x=0}^{M-1} f(x)\phi_{j_0, k}(x)
\]

\[
W_{\phi}(j, k) = \frac{1}{\sqrt{M}} \sum_{x=0}^{M-1} f(x)\phi_{j, k}(x)
\]

The properties of individual cells can be extracted based on tissue level. Quantification and distribution of the cells across the tissue are based on tissue level feature extraction. The properties of individual cells can be extracted based on cellular level feature extraction. There are many feature extraction techniques. In this paper used discrete wavelet functions like biorthogonal 3.5, coiflets 3, Daubechies 4, symlets 3 is applied for extracting coefficients to obtain the texture information of the ROI’s of the breast cancer. Extracted coefficients are sorted in ascending order. These coefficients are used to represent the corresponding breast cancer, i.e. each breast cancer image consists of these coefficients which are passed to classification step.

**Classification:**

There are different classifiers to classify the type of cancer based on features. The K-Nearest Neighbour (K-NN) is the modest algorithm from among the entire machine learning algorithm[12]. In this KNN classifier is used. The data mining research community identified the K nearest...
neighbor (KNN) as one of the top ten classifications. Based on the classes of its K nearest neighbors it predicts the class of new object. It is a non-parametric method used for classification. In this Euclidean distance between the feature vector of the test image and feature vector of the training images is calculated. In training set two inputs, namely benign and malignant are considered. Accuracy of the KNN algorithm depends on noise and unwanted features. So effort must be placed on selecting features given to the classifier.

Coefficients calculated from different breast cancer images using different wavelet functions are given as input to the KNN classifier. Statistical feature vector of a test image is given and Euclidean distance is measured between the test image set and training image set. Therefore the distances are sorted in ascending order. Here the image with shortest distance is treated as a reference image. Based on this image the test sample can be predicted. Cross validation is used to evaluate the classification accuracy for different transforms. The coefficients extracted are divided into 80% of data as training data and 20% as testing data. Overall performance is evaluated and area under the curve (AUC) is calculated.

The performance measure of the proposed method using different Dwt functions can also be supported by metrics like sensitivity (True positive rate) given by Eq.3 and specificity (True negative rate) given by Eq.4 using confusion matrix obtained. The values of specificity and sensitivity calculated from the confusion matrix of different DWT functions are given in Table: 1,2,3,4.

<table>
<thead>
<tr>
<th>Transform</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Accuracy(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Daubechies4</td>
<td>0.90</td>
<td>0.90</td>
<td>90.00%</td>
</tr>
<tr>
<td>Biorthogonal 3.5</td>
<td>0.82</td>
<td>0.85</td>
<td>83.75%</td>
</tr>
<tr>
<td>Symlets3</td>
<td>0.87</td>
<td>0.87</td>
<td>88.75%</td>
</tr>
<tr>
<td>Coiflets3</td>
<td>0.87</td>
<td>0.85</td>
<td>86.25%</td>
</tr>
</tbody>
</table>

Table 1: Accuracy using coefficients of DWT functions

Sensitivity = \( \frac{TP}{TP + FN} \) (3)

Specificity = \( \frac{TN}{TN + FP} \) (4)

The main objective behind the study is to classify whether the given ROI’s of the breast cancer are benign or malignant. Comparison is made between different DWT functions for extracting coefficients. Results are presented in Figure 5 for different ROI’s using different DWT functions in extracting coefficients. The value for Az is low for Biorthogonal3.5, coiflets3, symlets3 when compared to ‘Daubechies4’ in terms of classification accuracy using KNN classifier.
<table>
<thead>
<tr>
<th></th>
<th>Predicted BENIGN</th>
<th>Predicted MALIGNANT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual BENIGN</td>
<td>34</td>
<td>6</td>
</tr>
<tr>
<td>Actual MALIGNANT</td>
<td>5</td>
<td>35</td>
</tr>
</tbody>
</table>

Table 4: Confusion matrix of ROI images of bior3.5 wavelet function

![Fig. 5. ROC curve for classification of Malignant Vs Benign Breast Cancer using Different Discrete Wavelet Transform functions](image)

(a) Daubechies4 wavelet function
(b) Biorthogonal3.5 wavelet function
(c) Symlets3
(d) Coiflets3

V. CONCLUSION

In this paper classification is done using DWT and KNN by considering coefficients. The different DWT filters are considered. In this paper, accuracy depends on number of coefficients used for classification. In this paper, Experimental work has been done on a subset of breast cancer images obtained from MIAS database. The classification accuracy for classifying malignant and benign using Discrete wavelet function ‘Daubechies4’ and KNN classifier is 90%.

REFERENCES