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

[1]A .Krishna Veni, [2]U.Sathish Kumar

[1]M .Tech Student, [2]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, coifelts 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_z =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. Fathima 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 $\varphi_{j,k}$ 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:

$$W_\varphi(j_0,k)=\frac{1}{\sqrt{M}}\sum_{x=0}^{M-1} f(x)\varphi_{j_0,k}(x) \qquad (1)$$

$$W_\psi(j,k)=\frac{1}{\sqrt{M}}\sum_{x=0}^{M-1} f(x)\psi_{j,k}(x) \qquad (2)$$

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).

<table>
<tr><td>LL</td><td>HL</td><td rowspan="2">HL</td></tr>
<tr><td>LH</td><td>HH</td></tr>
<tr><td colspan="2">LH</td><td>HH</td></tr>
</table>

Fig 1: Two level decomposition.

III. Proposed Method

Dataset:

The dataset for this work is a subset taken from the Mammographic image analysis society(MIAS) is an organization of UK research groups which have produced a digital mammography database . The data are in PGM – (Portable Gray Map) format for storage which consists of 332“normal”,”benign” and “malign” cases were selected[11]. In this study only circumscribed mass, ill-defined mass, speculated mass, architectural distortion and asymmetry are considered. In MIAS associated patient information and image information is given as below.

. There are four major groups for classifying breast density:

- Fatty (F) (106 images).
- Fatty-glandular (G) (104 images).
- Dense-glandular (D) (112 images).

The abnormalities are also described with their kind:

- CALC Calcification.
- CIRC Well-defined/circumscribed masses.
- SPIC Spiculated masses.
- MISC Other, ill-defined masses.
- ARCH Architectural distortion.

In this paper a subset of the database, i.e. 80 images which consists of 40 benign and 40 malignant are taken. These images are pre-processed using gabor filter, set of images filtered with the different sizes and orientations , to remove noise and cropped to obtain Region of Interests (ROI’s) of size 256×256. The selection of ROI’s removes the unwanted pixel information and background. Fig 1 shows original images of benign and malignant breast cancer. Fig 3 shows ROI’s extracted from different cases taken from the MIAS dataset.

Fig.1a. Original image of Benign breast Cancer 001

Fig.1b.Original image of Malignant breastCancer041

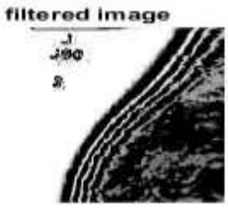


Fig.2a and 2bEnhanced image of benign and malignant breast cancer images

Fig. 3a and 3b ROI of benign and malignant breast cancer images

Feature extraction:

Image feature extraction is very important in image processing technique to detect and isolate desired portions of an image.The term feature can be stated as an “interesting’’ part of an image. Subsequent to image preprocessing, the features are extracted either at the tissue or cellular 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, coifelts 3, Daubechies 4,symlets3 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 featuers.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 predicated. 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

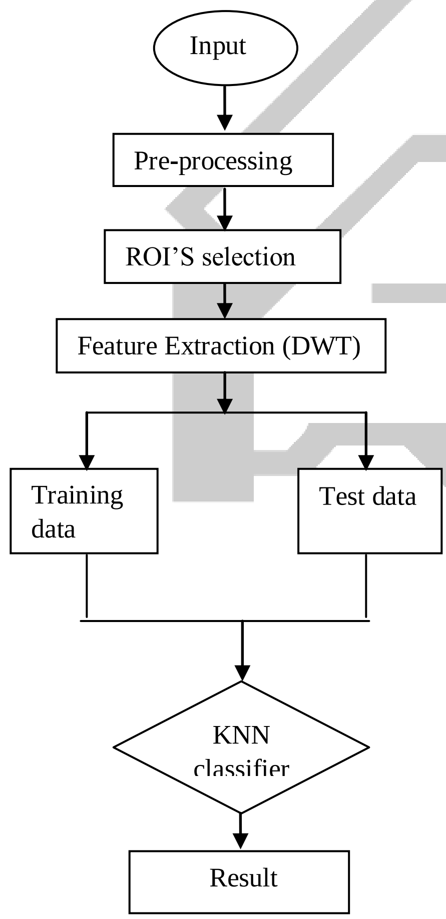


Fig: 4 Block diagram of proposed method.

IV.EXPERIMENTAL RESULTS

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, coifelts3, symlets3 when compared to 'Daubechies4' in terms of classification accuracy using KNN classifier.

Transform	Sensitivity	Specificity	Accuracy(%)
Daubechies4	0.90	0.90	90.00%
Biorthogonal 3.5	0.82	0.85	83.75%
Symlets3	0.87	0.87	88.75%
Coifelts3	0.87	0.85	86.25%

Table 1: Accuracy using coefficients of DWT functions

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

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

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 sensitivity (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.

N=80	Predicted BENIGN	Predicted MALIGNANT
Actual BENIGN	36	4
Actual MALIGNANT	4	36

Table1: Confusion matrix of ROI images of db4 wavelet function

N=80	Predicted BENIGN	Predicted MALIGNANT
Actual BENIGN	36	4
Actual MALIGNANT	5	35

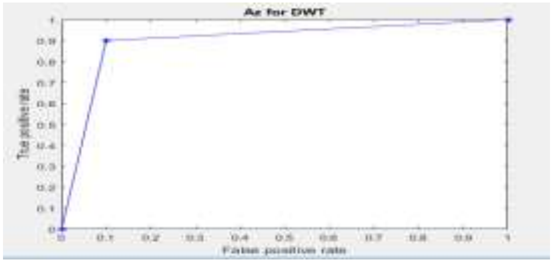
Table2: confusion matrix of ROI images of symlet3 wavelet function

N=80	Predicted BENIGN	Predicted MALIGNANT
Actual BENIGN	34	6
Actual MALIGNANT	7	33

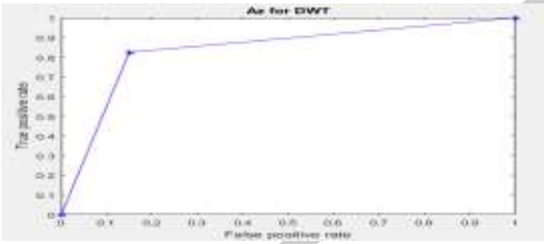
Table3: Confusion matrix of ROI images of coif3 wavelet function

N=80	Predicted BENIGN	Predicted MALIGNANT
Actual BENIGN	34	6
Actual MALIGNANT	5	35

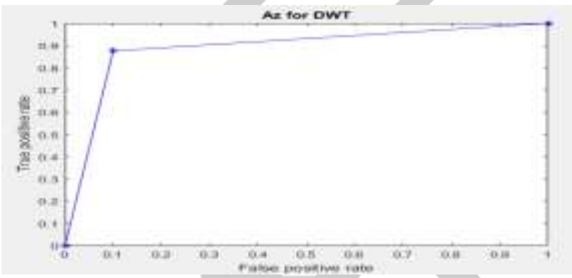
Table4: Confusion matrix of ROI images of bior3.5 wavelet function



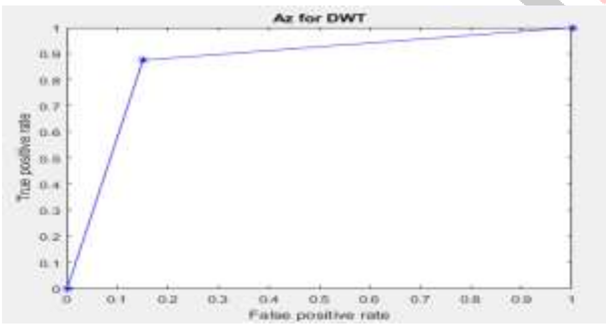
(a)Roc curve of Daubechies4 wavelet function



(b)Roc curve of Biorthogonal3.5 wavelet function



(c) Symlets3



(d) Coifelts3

Fig .5.ROC curve for classification of Malignant Vs Benign Breast Cancer using Different Discrete Wavelet Transform functions (a) Daubechies4(b)Biortogonal3.5(c)Symmlet3(d)coifelts3

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

[1] http://www.nationalbreastcancer.org/breast-cancer-facts.

[2] S. Dudea, C. Botar-Jid, D. Dumitriu, D. Vasilescu, S. Manole and M. Lenghel, "Differentiating benign from malignant superficial lymph nodes with sonoelastography," Medical Ultrasonography, vol. 15, no. 2, pp. 132-139, 2013

[3] Stamatia Detounis, "Computer-Aided Detection and Second Reading Utility and Implementation in a High-Volume Breast Clinic", Applied Radiology, pp: 8–15, 2004.

[4] NehaTripathi and S. P. Panda, "A Review on Textural Features Based Computer Aided Diagnostic System for Mammogram Mass Classification Using GLCM & RBFNN," International Journal of Engineering Trends and Technology (IJETT), vol. 17, no. 9, pp. 462-464, 2014

[5] R. Aarthi, N. K. K. Divya and S. Kavitha, "Application of Feature Extraction and clustering in mammogram classification using Support Vector Machine," in International Conference on Advanced Computing (ICoAC), Chennai, 2011.

[6]A. Oliver., J. Freixenet and A. Bosch, "Automatic Classification of Breast Tissue," Pattern Recognition and Image Analysis, vol. 3523, pp. 431-438, 2005.

[7] M. M. Fathima, D. Manimegalai and S. Thaiyalnayaki, "Automatic detection of tumor subtype in mammograms based On GLCM and DWT features using SVM," in International Conference on Information Communication and Embedded Systems (ICICES), Chennai, 2013

[8] N. V. Jog and S. R. Mahadik, "Implementation of Segmentation and Classification Techniques for Mammogram Images," International Journal of Innovative Research in Science, vol. 4, no. 2, pp. 422-426, 2015

[9]R. Campanini, D. Dongiovanni, E. Iampieri, N. Lanconelli, M. Masotti, G. Palermo, A. Riccardi, and M. Roffilli, "A novel featureless approach to mass detection in digital mammograms based on support vector machines," Phys. Med. Biol., vol. 49, no. 6, pp. 961–975, 2004.

[10] B. Sahiner, H. P. Chan, N. Petrick, D. Wei, M. A. Helvie, D. D. Adler, and M. M. Goodsitt, "Classification of mass and normal breast tissue: A convolution neural network classifier with spatial domain and texture images," IEEE Trans. Med. Imag., vol. no. 15, pp. 598–610,1996.

[11] http://peipa.essex.ac.uk/ ipa/pix/mias/.

[12] T. Cover and P. Hart, "Nearest neighbor pattern classification," in IEEE Transactions on Information Theory, Menlo Park, 1967.