INTERNATIONAL JOURNAL OF ELECTRONICS, MECHANICAL and MECHATRONICS ENGINEERING
Vol.2 Num.1 pp.(79-85)

A Comparative Study of Breast Mass Classification based on Spherical Wavelet Transform using ANN and KNN Classifiers

Pelin GÖRGEL[1], Ahmet SERTBAS[1], Osman Nuri UCAN[2]

[1]*Department of Computer Engineering, Faculty of Engineering*
Istanbul University (IU), Istanbul, TURKEY
[2]*Department of Electrical Engineering, Faculty of Engineering*
Istanbul Aydin University (IAU), Istanbul, TURKEY
e-mail: paras@istanbul.edu.tr, asertbas@ istanbul.edu.tr, uosman@aydin.edu.tr

Abstract: *Breast cancer may be missed by the radiologists at the early ages because of the mammography artifacts. Computer aided diagnosis can decrease the mortality rate by providing a second eye. The artifacts exist due to the noise and the inappropriate contrast in mammograms. Therefore a study that classifies the cropped region of interests (ROI's) as benign or malign and provides a second eye to the radiologists is proposed. The study consists of two steps: First step is the application of Spherical Wavelet Transform (SWT) to the original ROI matrix prior to feature extraction. Second step is to extract some predetermined pixel and shape features both from wavelet and scaling coefficients. Finally, for classification the prepared feature matrix is given to Artificial Neural Networks (ANN) and K-Nearest Neighbour (KNN) systems which are widely used in image processing. The algorithms are tested on 60 abnormal digitized mammogram ROIs acquised from The Mammographic Image Analysis Society (MIAS) which is a free mammogram database.*

Key Words: *Breast cancer, tumor classification, Spherical Wavelet Transform (SWT), ANN, K-Nearest Neighbour.*

1. Introduction

Breast cancer is one of the leading causes of cancer deaths among women in the world. Early detection of breast cancer is essential in reducing life fatalities. Mammography is widely recognized as the most reliable technique for early detection of breast cancers. Although this technology has been developing rapidly, the interpretation of a mammogram is often difficult and depends on the expertise and experience of the radiologist. Many times the radiologist recommends unnecessary biopsies depending on the probability of breast cancer because of the suspicious masses. Since the breast biopsy is an expensive and invasive procedure, methods to increase mammographic and accuracy without missing true positive (cancer) masses have to be developed. These methods are generally a combination of the computer aided systems and the radiologist's interpretation and have recently achieved a great performance in assisting radiologists in the malignant or benign decision, by providing them a second eye about diagnosing breast cancer disease. The computerized methods are applied by using digital mammography technology in which the mammograms are able to be visualized on computers.

The wavelet technology provides an efficient analysis for digital mammography. Hence, several studies for detection and classifying problem of mammographical masses were introduced in recent years. Unfortunately there are not many studies on the spherical wavelet and curvelet transforms as much as discrete wavelet transform because these are recently developed technologies. Karahaliou and Boniatis (2008) investigated texture properties of microcalcification clusters on mammograms.

Gray-level texture and wavelet coefficient texture features at three decomposition levels were extracted from surrounding tissue regions of interest. Specifically, gray-level firstorder statistics, gray-level cooccurrence matrices features, and Laws' texture energy measures are extracted from original image regions. They achieved 86% accuracy at classifying the masses normal, benign or malign. Hwang and Choi (2005) described breast tissue image analyses using texture features from Haar wavelet transformed images to classify breast lesion. The approach for creating a classifier is composed of two steps: feature extraction and classification. They extracted texture features from wavelet transformed images with 10 times magnification. In the classification step, they created three classifiers from each image of extracted features using statistical discriminant analysis, neural networks (back-propagation algorithm) and SVM. The system produced a satisfying classification performance of 91%. Curvelet efficiently represents discontinuities along edges or curves in images or objects. Some studies using curvelet transform in image processing have been carried out. Ali, Eldokany and Saad (2008) presented a curvelet approach for the fusion of magnetic resonance and computed tomography images. They found that curvelet transform achieved good results in their fusion. Bind and Tahan (2007) presented a method for object detection of speckle image based on curvelet transform. They constructed a segmentation method that provides a sparse expansion for typical images having smooth contours.

This paper aims mammographical diagnose via classification of tumors. The computer-aided diagnosis system described in this paper is based on feature extraction using Spherical Wavelet Transform (SWT) and ANN classification. First, several features related to size, shape and pixel values of the lesions are computed with a new technique based Spherical Wavelet Transform (SWT). Finally, to achieve the diagnosis, two different classification methods are used as classifiers, one of them is ANN and the other is K-Nearest Neighbour algorithm. The proposed system can be helpful to extract specific characteristics from a data and provide true interpretation to the radiologists.

The remaining of this paper is organized as it follows. Section 2 gives a brief introduction to wavelet and spherical wavelet transforms and Section 3 discusses the feature extraction work. Section 4 presents the comparative accuracy results and discussions, while Section 5 contains the conclusion of the work.

2. Methodology

2.1 Discrete Wavelet Transform (DWT)

The wavelet transform decomposes the signal into different scales with different levels of resolution that provide multiresolution analysis (MRA) by dilating a single prototype function, the mother wavelet. The multiresolution approximation of one-dimensional signal $f(x) \in L^2(R)$ at a resolution 2^j is defined as the orthogonal projection of a signal on subspace V_{2^j} of $L^2(R)$ (Eltoukhy, Faye and Samir 2010). $W^A_{2^j} f(x), W^{D^v}_{2^j} f(x), W^{D^h}_{2^j} f(x)$ and $W^{D^d}_{2^j} f(x)$ are the approximation, vertical detail, horizantal detail and diagonal detail of $f(x)$ respectively. $W^A_{2^{j+1}} f(x)$ approximation at resolution 2^{j+1} contains more information than the approximation $W^A_{2^j} f(x)$ at resolution 2^j. The details are equivalent to the orthogonal projection of $f(x)$ on the complement O_{2^j} of vector space V_j in V_{j+1}. The scaling function and corresponding wavelet function are denoted by $\varphi(x)$ and $\psi(x)$ respectively that satisfy $\varphi_{2^j} = 2^j \varphi(2^j x)$ and $\psi_{2^j} = 2^j \psi(2^j x)$. $\{2^{-j/2} \varphi_{2^j}(x - 2^{-j}k)\}_{k \in Z}$ and $\{2^{-j/2} \psi_{2^j}(x - 2^{-j}k)\}_{k \in Z}$ are orthogonal bases of O_{2^j} and V_{2^j} respectively. The approximation and detail signals of the original signal $f(x)$ at resolution 2^j are completely characterized by the sequence of inner products of $f(x)$ with φ_{2^j} and ψ_{2^j} as follows:

$$\{W^A_{2^j} f(k)\}_{k \in Z} = \{\langle f(o), \varphi_{2^j}(0 - 2^{-j}k)\rangle\}_{k \in Z} \quad (1)$$

$$\{W^D_{2^j} f(k)\}_{k \in Z} = \{\langle f(o), \psi_{2^j}(0 - 2^{-j}k)\rangle\}_{k \in Z} \quad (2)$$

Let *h* be a low-pass filter and *g* be a high-pass filter that $h(k) = \langle \varphi_{-1}(x), \varphi(x-k)\rangle$ and $g(k) = \langle \psi_{-1}(x), \psi(x-k)\rangle$. The mirror filters are defined as $\hat{h}(k) = h(-k)$ and $\hat{g}(k) = g(-k)$. The multiresolution representation of $f(x)$ at any resolution 2^j can

also be implemented by the mirror filters for $j = 0,-1,-2,\ldots$.

$$W^{A}_{2^{j-1}} f(x) = \sum_{k=-\infty}^{\infty} \hat{h}(2x-k) W^{A}_{2^{j}} f(k) \tag{3}$$

$$W^{D}_{2^{j-1}} f(x) = \sum_{k=-\infty}^{\infty} \hat{g}(2x-k) W^{A}_{2^{j}} f(k) \tag{4}$$

2.2 Spherical Wavelet Transform (SWT)

In analyzing data that contains anisotropic features, wavelets are no longer optimal and this has motivated the development of new multiscale decompositions such as the ridgelet and the spherical or curvelet transforms (Abrial 2007, Donoho 2000 and Starck 2003). Firstly, In Starck's (2003) study , it has been shown that the spherical transform could be useful for the detection and the discrimination of non Gaussianity in astronomical images. In this area, further insight will come from the analysis of full-sky data mapped to the sphere thus requiring the development of a curvelet transform on the sphere. The goal of this paper is to extend existing studies with wavelet transform based biomedical image processing by implementing the new technology spherical wavelet transform which works well on spherical shapes.

MRA of $L_2(S^2)$ where S^2 is the unit disc : $S^2 = \{(x,y) \in R^2 : x^2 + y^2 \le 1\}$. Accordingly the associated Legendre functions are;

$$P_n^{(m)}(t) = (1-t^2)^{m/2} \frac{1}{2^n n!} \frac{d^{n+m}}{dt^{n+m}} (t^2-1)^n, \quad n \ge m \tag{5}$$

$$\langle P_n^{(m)}, P_p^{(m)} \rangle = \frac{2(n+m)!}{2n+1(n-m)!} \delta np, \quad n, p \ge m \tag{6}$$

The spherical harmonics are the angular portion of a set of solutions to Laplace's equation. Represented in a system of spherical coordinates, Laplace's spherical harmonics set which has the equation below forms an orthogonal system.

$$Y_l^m(\theta,\phi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(\cos\theta) e^{im\phi} \tag{7}$$

It can be shown that the spherical harmonics, almost always written as $Y_l^m(\theta,\phi)$, form an orthogonal and complete set functions of the spherical polar angles, θ and ϕ, with ℓ and m indicating degree and order of the function. This implies that the harmonics can be used to describe a function of θ and ϕ in the form of a linear expansion; the expansion coefficients may be used as linear regression parameters, which means that they may be chosen such that the original and expanded function "resemble" each other as closely as possible. The more spherical symmetry the original function possesses, the shorter the expansion and the fewer fit parameters have to be determined. P_l^m denotes the Legendre polynomials and it equals $1, x, \frac{1}{2}(3x^2-1), \frac{x}{2}(5x^2-3), \frac{1}{8}(35x^4-30x^2+3)$ for $l = 0,1,2,3,4$ respectively. Because of the reconstruction of an image from its wavelet coefficients $I = \{w_1,\ldots,w_j,c_j\}$ is straightforward, the equation below can be written.

$$c_0(\theta,\phi) = c_j(\theta,\phi) + \sum_{j=1}^{j} w_j(\theta,\phi) \tag{8}$$

Also we can write $c_0(\theta,\phi) = \varphi_{lc}(\theta,\phi) * f(\theta,\phi)$. It is one of the most important parts of the wavelet transform to determine the scaling function φ_{lc}. In this study we use the Shannon scaling function which is demonstrated by the Eq. 10 (Wade 1995). Eq. 9 is the basic form of scaling functions with l_c cut-off frequency and $\hat{\varphi}_{lc}(l,0)$ spherical harmonic coefficients.

$$\varphi_{lc}(\theta,\phi) = \varphi_{lc}(\theta) = \sum_{l=0}^{l=lc} \hat{\varphi}_{lc}(l,0) Y_{l,0}(\theta,\phi) \tag{9}$$

$$\varphi_j(x,y) = \sum_{n=0}^{\min\left[2^j, M-1\right]} (|x||y|)^{-n-1} \frac{2n+1}{4\pi} P_n(|x||y|) \tag{10}$$

The other scaling functions with increasing scales are obtained respectively by using $\varphi_{j+1} = \varphi_0(2^{-(j+1)}x)$ until the desired scale is reached. Just as with the à trous algorithm, the wavelet coefficients can be defined as the difference between two consecutive resolutions, $w_{j+1}(\theta,\phi) = c_j(\theta,\phi) - c_{j+1}(\theta,\phi)$ which in fact corresponds to making the following specific choice for the wavelet function ψ_{lc} :

$$\hat{\psi}_{\frac{lc}{2^j}}(l,m) = \hat{\varphi}_{\frac{lc}{2^{j-1}}}(l,m) - \hat{\varphi}_{\frac{lc}{2^j}}(l,m) \tag{11}$$

The above multi-resolution sequence can also be obtained recursively by a low pass filter h_j for each scale j by

$$\hat{h}_j(l,m) = \sqrt{\frac{4\pi}{2l+1}} h_j(l,m) = \begin{cases} \dfrac{\hat{\varphi}_{\frac{lc}{2^{j+1}}}(l,m)}{\hat{\varphi}_{\frac{lc}{2^{j}}}(l,m)} & \text{if } l < \dfrac{lc}{2^{j+1}} \\ 0 & otherwise \end{cases} \quad (12)$$

It is then easily pointed that c_{j+1} derives from c_j by convolution with $\hat{h}_j$: $c_{j+1} = c_j * \hat{h}_j$. In the same way a high pass filter can be derived with ψ_{lc} wavelet function on each scale j and $w_{j+1} = c_j * g_j$.

$$\hat{g}_j(l,m) = \sqrt{\frac{4\pi}{2l+1}} g_j(l,m) = \begin{cases} 1 & \text{if } l \geq \dfrac{lc}{2^{j+1}} \\ \dfrac{\hat{\psi}_{\frac{lc}{2^{j+1}}}(l,m)}{\hat{\varphi}_{\frac{lc}{2^{j}}}(l,m)} & \text{if } l < \dfrac{lc}{2^{j+1}} \end{cases} \quad (13)$$

The purpose is to obtain all coefficients $(w_1, w_2, w_3, w_4, w_5, c_5)$ of the transform including wavelet and scaling coefficients respectively. These coefficients are used in the classification of the diagnosis system.

2.3 Artificial Neural Network

In this study a supervising three layer back propagation neural network is used for pattern classifying (Fig. 1).

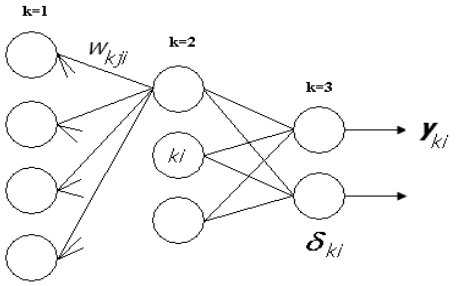

Figure 1. ANN model

The outputs of the neurons and the error of the output neurons are calculated using Eq. 14 and 15 respectively:

$$y_{(k+1)i} = \sum_{j=1}^{N_k} y_{kj} w_{(k+1)ji} \quad (14)$$

$$\delta_{ki} = y_{ki}(1 - y_{ki})(\text{target}_i - y_{ki}) \quad (15)$$

The Eq. 16, 17 and 18 updates the weights of the output layer, calculates the error of the hidden layer neurons and updates the weights of the hidden layer respectively.

$$w_{kji}^{+} = w_{kji} + \eta \delta_{ki} y_{(k-1)j} \quad (16)$$

$$\delta_{ki} = y_{ki}(1 - y_{ki}) \sum_{j=1}^{N_{(k+1)}} \delta_{(k+1)j} w_{(k+1)ij} \quad (17)$$

$$w_{kji}^{+} = w_{kji} + \eta \delta_{ki} y_{(k-1)j} \quad (18)$$

2.4 K-Nearest Neighbour

K-Nearest Neighbor (KNN) classification is one of the most fundamental and simple classification methods. When there is little or no prior knowledge about the distribution of the data, KNN should be one of the first choices for a classification study. KNN is a nonparametric method in that no parameters are estimated. Instead, the proximity of neighboring input (a) observations in the training data set and their corresponding output values (y) are used to predict the class of the objects in the validation data set. Firstly two input variable case is considered since it is easy to represent in two-dimensional space. Euclidean distance between two input vectors a_1 and a_2 is used in KNN (Panigrahi and Pandi 2009).

$$a'_1 = (a_{1,x}, a_{1,y}) \quad (19)$$

$$a'_2 = (a_{2,x}, a_{2,y}) \quad (20)$$

The distance between these two vectors is computed as the length of the difference vector $a_1 - a_2$, denoted by

$$d(a_1, a_2) = |a_1 - a_2| = \sqrt{(a_{1,x} - a_{2,x})^2 + (a_{1,y} - a_{2,y})^2} \quad (21)$$

More generally the distance between two p-dimensional vectors $u' = (u_1, u_2, ..., u_p)$ and $v' = (v_1, v_2, ..., v_p)$ is calculated as

$$d(u,v) = |u - v| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + ... + (u_p - v_p)^2} \quad (22)$$

The minimum distance between the vectors gives the closest neighbour so it is predicted that it belongs to the same class with the test object.

3. Feature Extraction

The images from the MIAS data set are previously investigated and labeled by the experts based on technical experience and biopsy (Fig. 2,3). The data set is composed of 60 cropped regions of interest (ROI's) which 25 are diagnosed as malign, 35 as benign.

Fig. 2. Mammogram samples with marked benign tumors

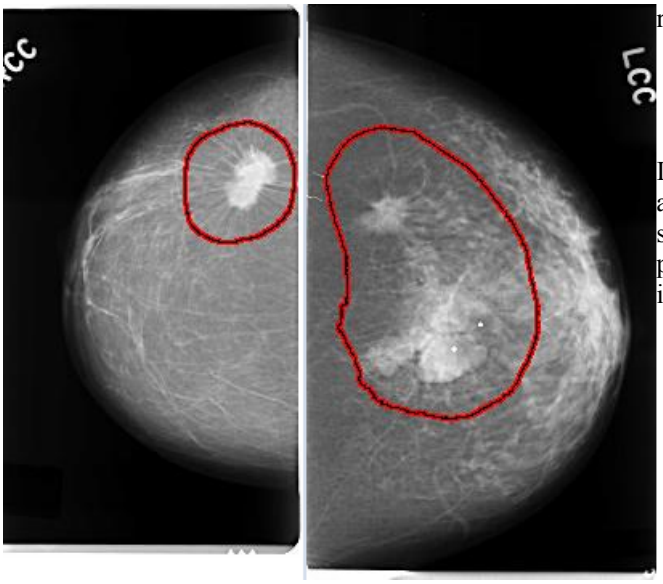

Figure 3. Mammogram samples with marked malign tumors

The tumors existing in breast tissue can differentiate in shape, margin, orientation, lesion boundary, echogenic pattern, posterior acoustic features, effect on surrounding parenchyma, effect on surrounding parenchyma, and vascularity. Malign masses leading to breast cancer disperse into the normal breast tissue, have irregular boundary and sharp corners like stars, and the benign masses which do not cause loss in human lives have smooth, distinct and regular margin.

In this study we extract some shape and pixel value features both from the coefficients obtained by the SWT algorithm and the original ROI image. These features are a*rea, centroid*, *bounding box, major-minor axis length* (the length of the major and minor axis of the ellipse that has the same second-moments as the region), e*ccentricity* (belonging to the ellipse that has the same second-moments as the region and the ratio of the distance between the foci of the ellipse and its major axis length), o*rientation* (the angle between the x-axis and the major axis of the ellipse that has the same second-moments as the region), f*illed area, euler number* (the number of objects in the region minus the number of holes in those objects), *extrema* (8-by-2 matrix; the extremal points in the region), c*onvex hull* (the smallest convex polygon that can contain the region), *equiv diameter* (the diameter of a circle with the same area as the region), s*olidity* (the proportion of the pixels in the convex hull that are also in the region), e*xtent* (the proportion of the pixels in the bounding box that are also in the region) and finally the m*ean center-border distance* (similarity between a circle and the ROI). The calculated numeric features mentioned below provide feature matrix which is used as the input vector of the supervised learning system ANN and KNN. Each column and row of the matrix demonstrates a feature and an other ROI respectively.

4.Results

In this work, the classification step is carried out after SWT, using an ANN classifier which gives satisfying results in image processing. The proposed breast cancer diagnosing system is shown in Fig.4.

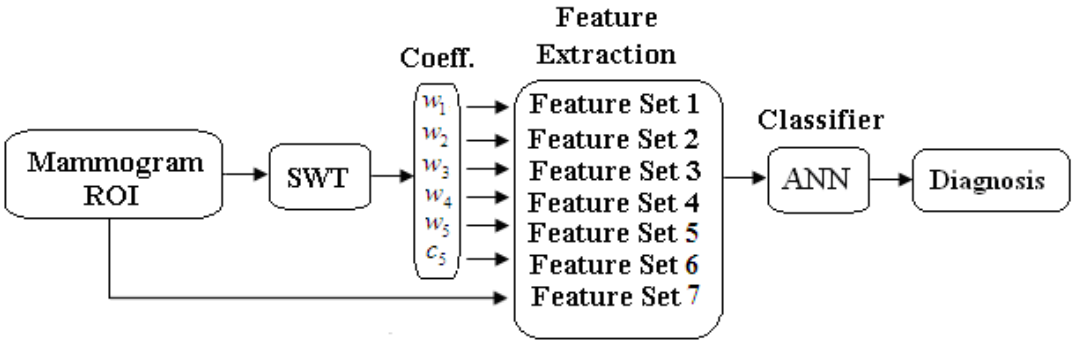

Figure 4. The diagnosis system

The effectiveness of the technique is determined by extracting features of the cropped ROI of mammograms' 5 scale SWT coefficients. The features mentioned in Section 3 which are extracted from the original ROI image and the related coefficients are submitted as inputs to the ANN system. Since the number of the features is 17, feature matrix (input matrix) has a size of 17x7; one for ROI matrix and six for SWT coefficient matrices $w_1, w_2, w_3, w_4, w_5, c_5$. In addition to this work DWT is also applied before classification and KNN algorithm is also used for the classification. The purpose is to present the superiority of SWT over DWT (Discrete Wavelet Transform) and ANN over KNN.

In this work, ROC (Receiver Operating Characteristic) analysis based on statistical decision theory, has been applied to the diagnosis system. The ROC curve is a plot of the classifier's true positive detection rate versus its false positive rate. A classifying system results as *false positive* (FP) if the system labels a negative point as positive and *true positive* (TP) if the system correctly predicts the positive label. Fig. 5 and Table 1 demonstrate performance analysis using ROC curves and the test results respectively.

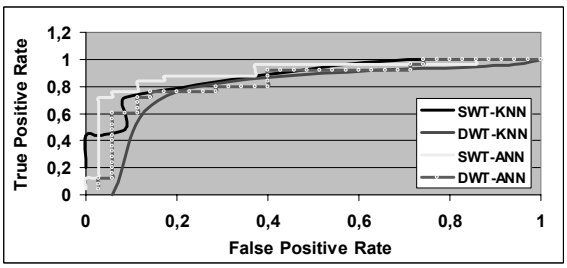

Fig. 5. Performance Plot of SWT-ANN, DWT-ANN, SWT-KNN and DWT-KNN Diagnosis Systems

Table 1. The test results of the Diagnosis Systems

Classification Method	Wavelet Method	Sensitivity (%)	Specificity (%)	Classification Accuracy (%)
ANN	SWT	89	84	87
ANN	DWT	89	72	82
KNN	SWT	91	72	83
KNN	DWT	80	76	78

5.Conclusion

The research presented in this article aims to decrease the mortality rate related to breast cancer by reducing the number of malign tumors which radiologists can not notice by means of computer aided image processing . On the other hand it is desired to decrease the number of the requested tumor biopsies which are assumed to be cancer but in fact indicate benign features. In this study breast tumors that have been correctly classified by the experts are tested in the software system constructed by the SWT, DWT, ANN and KNN techniques in MATLAB Version 7.6 (The Language of Technical Computing) environment. The satisfying experimental results demonstrate that the best performance is obtained by using SWT-ANN method with a 87% classification accuracy. The SWT-KNN system has 83% classification accuracy while DWT-ANN system has 82%. The poorest performance is demonstrated with DWT-KNN system which has only 78% classification accuracy. According to the results it is seen that this study is valuable to improve early diagnosis and reduce the number of unnecessary biopsies.

References

[1] Abrial, P & Moudden, Y 2007, "Morphological component analysis and inpainting on the sphere: application in physics and astrophysics" , Journal of Fourier Analysis and Applications (JFAA), special issue on ``Analysis on the Sphere", Vol. 13, pp. 729-748.

[2] Ali, F, Eldokany, I, Saad, A and Abdelsamie, F 2008, "Curvelet fusion of MR and CT images", Progress in Electromagnitics Research, pp. 215–224.

[3] Binh, N & Thanh, C 2007, "Object detection of speckle image base on curvelet transform", ARPN Journal of Engineering and Applied Sciences, Vol. 2, pp. 14-16.

[4] Donoho, D & Duncan, M 2000, in Proc. Aerosense 2000, Wavelet Applications VII, ed. H. Szu, M. Vetterli, W. Campbell, & J. Buss, SPIE, 4056, 12.

[5] Eltoukhy, M, Faye, I & Samir, B 2010, "A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram, Computers in Biology and Medicine, Vol. 40, pp. 384–391.

[6] Hwang, H & Choi, H 2005, "Classification of breast tissue images based on wavelet transform using discriminant analysis, neural network and SVM.", IEEE, pp. 345-349.

[7] Karahaliou, N, A & Boniatis, I, S 2008, "Breast cancer diagnosis: analyzing texture of tissue surrounding microcalcifications", IEEE Transactions on Information Technology in Biomedicine, Vol. 12, pp. 731-738.

[8] Panigrahi, B, K & Pandi, V, R 2009, "Optimal feature selection for classification of power quality disturbances using wavelet packet-based fuzzy k-nearest neighbour algorithm ", Generation, Transmission & Distribution,Vol. 3, pp. 296 – 306.

[9] Starck, J, Candès, E & Donoho, D, L 2003, "Astronomical image representation by the

[10] W.R. Wade, "A Walsh System for Polar Coordinates", Computers Math. Applic., Vol. 30, pp. 221-227, 1995.