Mammogram Classification using Two Dimensional Discrete Wavelet Transform and Gray-Level Co-occurrence Matrix for Detection of Breast Cancer

Shradhananda Beura, Banshidhar Majhi and Ratnakar Dash

Pattern Recognition Research Laboratory, Department of Computer Science and Engineering
National Institute of Technology, Rourkela, India

Abstract

In this paper, we propose a mammogram classification scheme to classify the breast tissues as normal, benign or malignant. Feature matrix is generated using GLCM to all the detailed coefficients from $2D$-DWT of the region of interest (ROI) of a mammogram. To derive the relevant features from the feature matrix, we take the help of *t-test* and *F-test* separately. The relevant features are used in a BPNN classifier for classification. Two standard databases MIAS and DDSM are used for the validation of the proposed scheme. It is observed that *t-test* based relevant features outperforms to that of *F-test* with respect to accuracy. In addition to the suggested scheme, the competent schemes are also simulated for comparative analysis. It is observed that the proposed scheme has a better say with respect to accuracy and area under curve (AUC) of receiver operating characteristic (ROC). The accuracy measures are computed with respect to normal vs. abnormal and benign vs. malignant. For MIAS database these accuracy measures are 98.0% and 94.2% respectively, whereas for DDSM database they are 98.8% and 97.4%.

Keywords:

Mammogram, ROI, GLCM, DWT, null hypothesis, confusion matrix

1. Introduction

Breast cancer is still the most common cancer throughout the world and a frequent cause of cancer death among women. According to Globocan project, it has been estimated for the year 2012 that, 1.67 million new cancer cases were diagnosed worldwide, which is 25% of all types of cancers. In India, the breast cancer is considered as the most common cancer and in the year 2012, $144,937$ women were newly detected with this cancer and $70,218$ patients died among them. So, it can be noticed that, one patient is dying out of every two newly diagnosed women [1]. It has been studied that, the recovery of the breast cancer as well as survival rate can be improved by the early detection through periodic screening. Regarding this context, mammography is the most effective and reliable method for an accurate detection of breast cancer in recent years [2]. Mammograms are x-ray images of breasts. Reading of mammograms is a very important task for radiologists as they suggest patients for biopsy. However, human interpretation of mammograms varies as it depends on training and experience. This leads to different judgments by different radiologists. Mammogram interpretation is a repetitive task which requires maximum attention for avoidance of misinterpretation. Therefore, computer-aided diagnosis (CAD) system is currently a very popular and efficient method which analyzes the digital mammograms with the use of image processing. CAD system helps radiologists in accurate interpretation of mammograms for detection

of suspicious lesions and classification. It has been observed that 60 to 90% of the biopsies of cancers predicted by radiologists found benign later [3]. So, it is very important to develop a CAD system, which can distinguish normal-abnormal as well as benign-malignant mammograms. The main objective of CAD system is to increase diagnosis accuracy and enhancing the mammogram interpretation. Thus, CAD system can reduce the variability in judgments among radiologists by providing an accurate diagnosis of digital mammograms. Regarding this responsibility, one important step is to find out a set of significant features from the mammographic images that can distinguish the normal mammograms from abnormal as well as the benign lesions from malignant ones. Different techniques and methods have been studied for this purpose.

For mammogram feature extraction and classification, several researches have been carried out over the year. One of the effective methods is the multiresolution analysis in which; the original mammographic image is decomposed into several sub-images that preserve informations about both high and low frequencies. Wavelet transform is one of the important methods for the texture analysis of the image. Many researchers worked on multiresolution analysis of mammograms based on wavelets by using different types of feature spaces. Dhawan et al. used wavelet decomposition and gray level image structure features for classification of mammograms and obtained an area under curve (AUC) of 0.81 in a receiver operating characteristic (ROC) curve [4]. Wei et al. achieved AUC of 0.96 through ROC analysis in the classification of abnormal-normal mammographic tissue classification by using multiresolution texture features [5]. In their method, wavelet transform has been used to decompose the mammographic region of interest (ROI) to collect different detail coefficients and consequently, texture features were extracted from these coefficients. Liu et al. used a set of statistical features based on wavelets and found 84.2% accuracy rate by using binary tree as classifier in mammogram classification [6]. Rashed et al. obtained 87.06% of classification accuracy by using different types of Daubechies wavelets in the classification of mammograms [7]. Prathibha et al. used multiscale wavelet transformation for extraction of texture features from the mammographic images. They obtained the classification performance as AUC of 0.95 in ROC to classify normal and abnormal mammograms by using the nearest neighbor classifier [8]. Buciu et al. achieved AUC values as 0.79 and 0.78 for classification of normal-abnormal and benign-malignant mammogram classes respectively [9]. They have used Gabor wavelets with principal component analysis for reduction in dimension of directional features with the help of support vector machine as classifier. Görgel et al. used wavelet based support vector machine (SVM) in their proposed method for mammographic mass classification and achieved an accuracy of 84.8% [10]. In another proposed method, Görgel et al. found 96.0% and 93.59% classification accuracy rates for normal-abnormal and benign-malignant mammogram classification using spherical wavelet transform (SWT) for extraction of features and SVM as the classifier [11]. In their proposed method, a local seed region growing algorithm has been used to detect ROIs of mammograms.

Texture of a mammographic image is the quantitative statistical measurements of pixel intensities in a region. The textural information of mammographic images is very important for distinguishing the abnormal pattern from the normal. A popular method for texture analysis can be taken into notice is the gray-level co-occurrence matrix (GLCM) which estimates the second order-statistical properties of images [12] [13] [14]. Chan et al. achieved performance index value of $AUC = 0.89$ by using texture morphology features based on GLCM in the classification of mammograms and [15]. In their proposed method, a feature selection technique based on genetic algorithm (GA) has been used to select

effective features from multidimensional feature spaces. Mutaz et al. developed a method in which the textural features were extracted from ROI using GLCM [16]. Utilizing these features, they discriminated the benign and malignant mammograms with the help of neural network and achieved an accuracy of 91.67% sensitivity. Jona et al. used GLCM to extract the features from the mammographic images [17]. They optimized the feature set by using a hybrid particle swarm optimization and genetic algorithm, and obtained 94% of classification accuracy by using SVM to classify the normal and abnormal mammograms.

The literature survey reveals about the existing classification schemes for digital mammogram images. However, most of them are not able to provide a good accuracy. It has been seen that the dimension of extracted feature space is so high due to large varieties of normal and abnormal tissues present in the breast. The use of high dimensional feature space may degrade the performance of the classification scheme. From a large feature space, only some of the features are effective and significant for the mammogram classification. Therefore, in addition to feature extraction, feature selection is also the key step in mammogram classification, which selects only the significant features from available feature space. So there is a need to develop some new feature extraction and selection algorithms to increase the accuracy of classification rate. In this paper, we have proposed an effective feature extraction algorithm using two dimensional discrete wavelet transform (2D-DWT) based multiresolution analysis along with gray-level co-occurrence matrix (GLCM) to compute texture features for mammographic images. A feature selection algorithm has been applied using two statistical feature selection methods such as *two-sample t* and *F-test* to select significant features from extracted features. Utilizing these significant features, a back propagation neural network (BPNN) has been used as classifier to predict the mammogram, whether it is a normal or abnormal. In addition, the severity with respect to malignant or benign is also estimated in abnormal cases. The rest of this paper is organized as follows: Section 2 deals with the proposed scheme, where extraction and selection of features and classification is discussed in detail. Section 3 describes the results obtained on standard databases. Section 4 gives the concluding remarks.

2. Proposed method

The proposed method of mammographic image classification includes two major algorithms such as feature extraction and selection. The feature extraction algorithm concentrates on the texture point in the mammographic image utilizing 2D-DWT and GLCM in succession on region of interest (ROI) to find out the feature descriptors of each detail coefficient of 2-level DWT. In the feature selection algorithm; effective and significant features are selected and provided to the neural network for the classification of mammograms as normal, benign or malignant. The overall block diagram of the scheme is shown in Figure 1.

2.1. Extraction of Region of Interest (ROI)

It may be noted that mammographic image is composed of different types of noises, artifacts in their background. The object area also contains the pectoral muscles. All these areas are unwanted portions for the texture analysis due to which; the full mammographic image is unsuitable for feature extraction and subsequent classification. Therefore, a cropping operation has been applied on mammogram images to extract the regions of interests (ROIs) which contain the abnormalities, excluding the unwanted portions of the image. This process is performed by referring the center
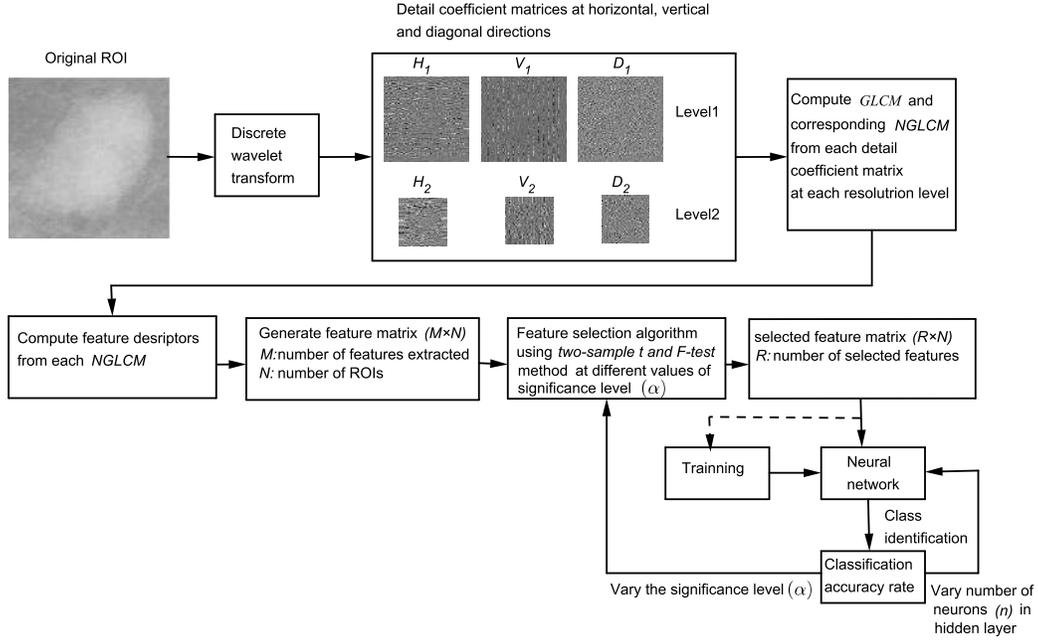
Figure 1: Block diagram of the proposed scheme for classification of mammograms using back propagation neural network (BPNN).

of the abnormal area as the center of ROI and taking the approximate radius (in pixels) of a circle enclosing the abnormal area as shown in Figure 2. For the extraction of normal ROI, the same cropping procedure is performed on normal mammographic images with random selection of location. Thus, in this phase, the ROIs extracted are free from the background information and noises. Figures 3 and 4 show some extracted ROIs containing different classes of abnormality.
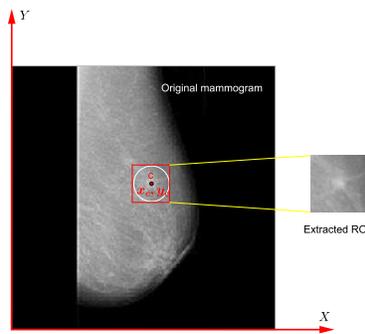
Figure 2: Cropping of ROI from mammographic image referring the center of the abnormal area.

2.2. Multiresolution Analysis using two dimensional Discrete Wavelet Transform (2D-DWT)

In the multiresolution technique, the underlaying texture of mammographic ROIs are analyzed by zooming in and out process. The discrete wavelet transform decomposes the mammographic ROI into a number of sub-images in different

(a) Fatty tissues   (b) Fatty-glandular tissues   (c) Dense-glandular tissues
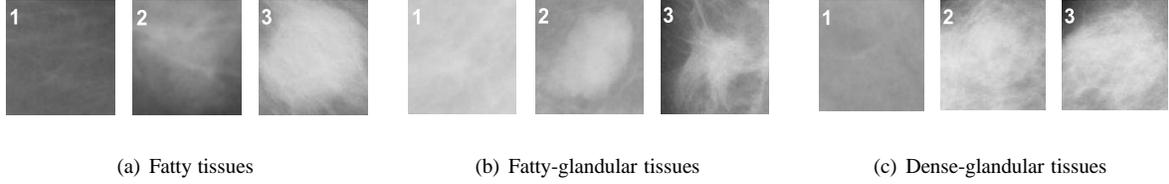
Figure 3: Mammographic ROIs of MIAS database [18]. The sub-figures indicate different types tissues present in mammograms. The levels 1, 2 and 3 of ROIs represents normal, benign and malignant classes respectively.

(a) Almost entirely fatty tissues   (b) Scattered fibro-glandular tissues

(c) Hetereogeneously dense tissues   (d) Extremely dense tissues
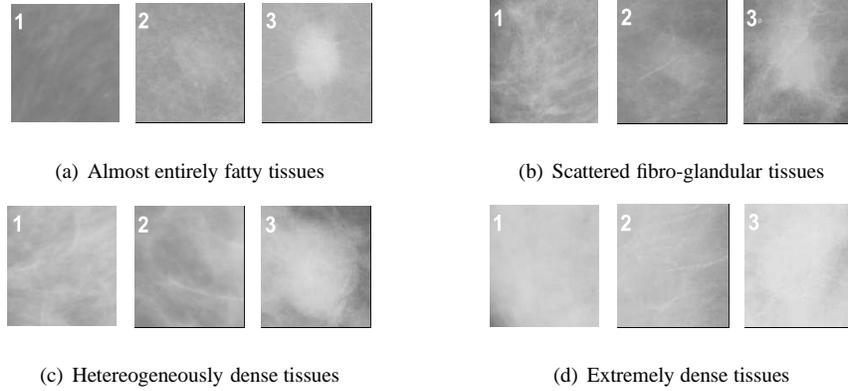
Figure 4: Mammographic ROIs of DDSM database from IRMA project [19]. The sub-figures indicate different types tissues present in mammograms. The levels 1, 2 and 3 of ROIs represents normal, benign and malignant classes respectively.

resolution levels preserving the high and low frequency information. This property leads the wavelet to extract better texture information from the mammographic ROIs. Given a continuous, square integrable function $f(x)$, its wavelet transform is calculated as the inner product of f and a real valued wavelet function ($\psi(x)$) [20] given by,

$$W\left[f\left(s,\tau\right)\right] = \langle f, \psi^k_{s,\tau} \rangle = \int\limits_{-\infty}^{\infty} f\left(x\right) \psi^k_{s,\tau}\left(x\right) dx \tag{1}$$

where $\psi^k_{s,\tau}(x) = \frac{1}{\sqrt{s}}\psi^k\left(\frac{x-\tau}{s}\right)$ is a wavelet family, $s \in Z$, τ and $k \in \{h, v, d\}$ are scale (resolution level), translation and orientation parameters respectively. The orientation parameters h, v and d represents to horizontal, vertical and diagonal directions respectively. Now the dyadic wavelet decomposition is achieved when $s = 2^j$ and $\tau = 2^j.n$, $\quad j, n \in Z$. Using the wavelet function $\psi(x)$ and scaling function $\varphi(x)$, the wavelet and scaling families are constructed as,

$$\psi^k_{j,n}(x) = \frac{1}{\sqrt{2^j}}\psi^k\left(\frac{x-2^j.n}{2^j}\right) \text{ and } \varphi^k_{j,n}(x) = \frac{1}{\sqrt{2^j}}\varphi\left(\frac{x-2^j.n}{2^j}\right) \tag{2}$$

These are orthonormal basis of sub-spaces and related to resolution 2^j. The wavelet atoms are defined by scaling and translating three mother atoms like ψ^h, ψ^v and ψ^d. These oriented mother atoms are computed as the tensor product of one dimensional $\psi(x)$ and $\varphi(x)$ given by,

$$\begin{aligned}&\varphi(x) = \varphi(x_1)\varphi(x_2), \psi^h(x) = \psi(x_1)\varphi(x_2), \\ &\psi^v(x) = \varphi(x_1)\psi(x_2) \text{ and } \psi^d(x) = \psi(x_1)\psi(x_2)\end{aligned} \tag{3}$$

5

A two dimensional discrete wavelet transform is implemented using the combination of digital filter banks and down-samplers. The digital filter banks consist of high-pass (g) and low-pass (h) filters. In the configuration of DWT structure, the number of banks is set as per the desired resolution [21]. As the image is a $2D$ signal, separable wavelet functions compute the discrete wavelet transform (DWT). The rows and columns of the image are separately undergone through the $1D$ wavelet transform to establish the $2D$-DWT. As shown in Figure 5, the original image $A_{2^{j+1}}f$ at resolution 2^{j+1} is decomposed into four sub-band images in the frequency domain. Among them, three sub-band images, $D_{2^j}^h f$, $D_{2^j}^v f$, $D_{2^j}^d f$ are the detail images at resolution 2^j in horizontal, vertical, and diagonal directions respectively. The Fourth one is the approximation image, $A_{2^j}f$ found at coarse resolution. So the whole image $A_{2^{j+1}}f$ is represented as,

$$A_{2^{j+1}}f = D_{2^j}^h f + D_{2^j}^v f + D_{2^j}^d f + A_{2^j}f \tag{4}$$

The decomposed sub-images are the representation of $2D$ orthogonal wavelet. Thus, the output of a wavelet decomposition of an image results into four orthogonal sub-band components like Low-Low (LL), Low-High (LH), High-Low (HL) and High-High (HH), that correspond to sub-images $D_{2^j}^h f$, $D_{2^j}^v f$, $D_{2^j}^d f$ and $A_{2^j}f$ respectively as shown in Figure 5.
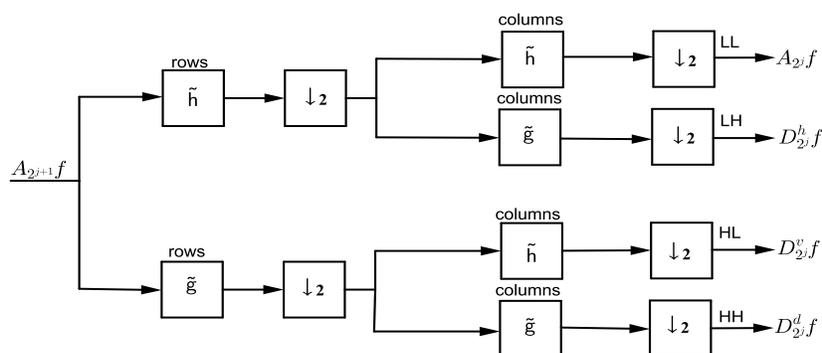
Figure 5: Wavelet decomposition using analysis filter banks.

2.3. Gray-Level Co-occurrence Matrix (GLCM)

The gray-level co-occurrence matrix (*GLCM*) is used to extract the texture in an image by doing the transition of gray level between two pixels. The *GLCM* gives a joint distribution of gray level pairs of neighboring pixels within an image [13]. The co occurrence matrix of the ROI is useful in classification of types of breast tissues by extracting descriptors from the matrix. For the computation of *GLCM*, first a spatial relationship is established between two pixels, one is the reference pixel, and the other is a neighbor pixel. This process forms the *GLCM* containing different combination of pixel gray values in an image. Let $q(i, j)$ is the element of *GLCM* of a given image f of size $M \times N$ containing the number of gray levels G ranging from 0 to $G - 1$. Then q can be defined as the matrix element and given by,

$$q(i, j) = \sum_{x=1}^{M} \sum_{y=1}^{N} \begin{cases} 1, & \text{if } f(x, y) = i \text{ and } f(x + \Delta x, y + \Delta y) = j \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

where (x, y) and $(x + \Delta x, y + \Delta y)$ are the locations of reference pixel and its neighboring pixel respectively. Each element of GLCM, $q(i, j | \Delta x, \Delta y)$ represents the relative frequency with which two pixels in a given neighborhood separated by a distance $(\Delta x, \Delta y)$ having gray level values i and j respectively [14]. It can be represented as $q(i, j | D, \theta)$, where the parameter D is the distance of separation between two neighboring resolution cells with two pixels having intensities i and j in the image. The other parameter θ represents the direction of neighboring pixel with respect to pixel of reference. The directionality used in GLCM is shown in Figure 6. The parameter D is also called as set distance as it specifies the distance of all neighboring resolution pairs contain in a set. For the texture calculation; the GLCM must be symmetrical, and each entry of the GLCM should be a probability value. For this purpose, a normalization process is followed. Each element of the normalized gray-level co-occurrence matrix (NGLCM) is defined as,

$$p(i, j) = q(i, j) / \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} q(i, j) \quad (6)$$

where n represents the size of NGLCM.
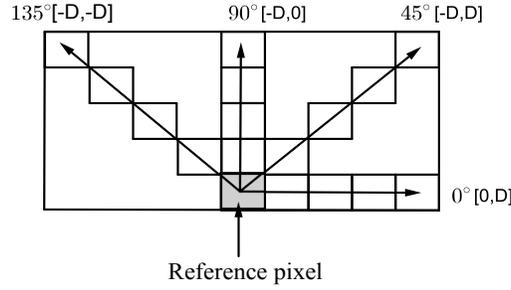
Figure 6: Directionality used in the gray-level co-occurrence matrix.

The size of GLCM is same as the number of gray levels of input image. The GLCM is highly dependent on the parameters D and θ. Several matrices can be obtained with small changes in the parameter D and θ. For the digital mammograms, the distance parameter D is limited to integral multiples of the pixel size, and the value of a direction parameter θ can be $0°, 45°, 90°$ and $135°$. Figure 7 describes the process of computation of GLCM of a given test image intensity matrix. Here, the number of gray level is four and the offset values are taken as $[0, 1]$, $[-1, 0]$, $[-1, 1]$, and $[-1, -1]$. The offset values represent set distance $D = 1$ in four possible neighboring pixel directions $\theta = 0°, 90°, 45°$ and $135°$ with respect to reference pixel. It can be seen that the occurrence of resolution cells pair $(0, 2)$ in the intensity matrix of input image is 4 in the horizontal direction $(\theta = 0°)$ due to the symmetric property. Therefore, the element in the $(0, 2)$ position of the horizontal GLCM is 4 as shown in Figure 7(b). In the same manner other three GLCMs are computed. Figure 8 shows the normalized gray-level co-occurrence matrices (NGLCM) where each cell in the matrices contains probability value. Each element of NGLCM is computed by dividing 24 in case of horizontal and vertical directions, and 18 in case of left diagonal and right diagonal directions to each element of corresponding symmetrical GLCM.

7

	0	1	2	3
0	0	2	1	2
1	2	3	2	0
2	0	1	3	2
3	2	0	2	1

(a)

(b)

	0	1	2	3
0	0	1	4	0
1	1	0	3	1
2	4	3	0	3
3	0	1	3	0

(c)

	0	1	2	3
0	0	1	5	0
1	1	0	2	1
2	5	2	0	3
3	0	1	3	0

(d)

	0	1	2	3
0	0	0	0	3
1	0	0	2	1
2	0	2	6	0
3	3	1	0	0

(e)

	0	1	2	3
0	2	1	0	1
1	1	0	2	1
2	0	2	4	0
3	1	1	0	2
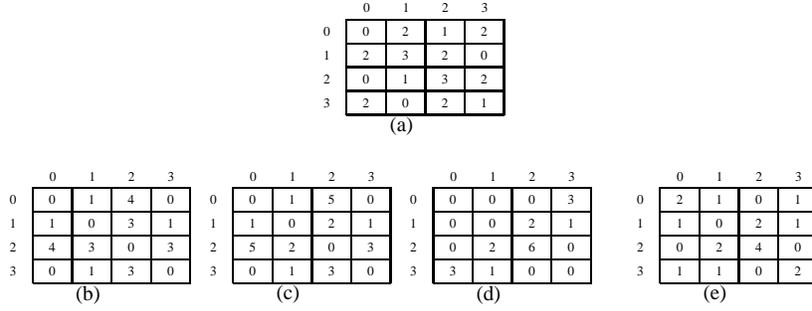
Figure 7: Computation of co-occurrence matrices. (a) Intensity values of input image with 4 gray levels. Different co-occurrence matrices (*GLCM*) for set distance $D = 1$ at four different directions such as (b) horizontal ($\theta = 0°$), (c) vertical ($\theta = 90°$), (d) right diagonal ($\theta = 45°$), (e) left diagonal ($\theta = 135°$).

	0	1	2	3
0	0	0.0417	0.1667	0
1	0.0417	0	0.1250	0.0417
2	0.1667	0.1250	0	0.1250
3	0	0.0417	0.1250	0

(a)

	0	1	2	3
0	0	0.0417	0.2083	0
1	0.0417	0	0.0833	0.0417
2	0.2083	0.0833	0	0.1250
3	0	0.0417	0.1250	0

(b)

	0	1	2	3
0	0	0	0	0.1667
1	0	0	0.1111	0.0556
2	0	0.1111	0.3333	0
3	0.1667	0.0556	0	0

(c)

	0	1	2	3
0	0.1111	0.0556	0	0.0556
1	0.0556	0	0.1111	0.0556
2	0	0.1111	0.2222	0
3	0.0556	0.0556	0	0.1111

(d)
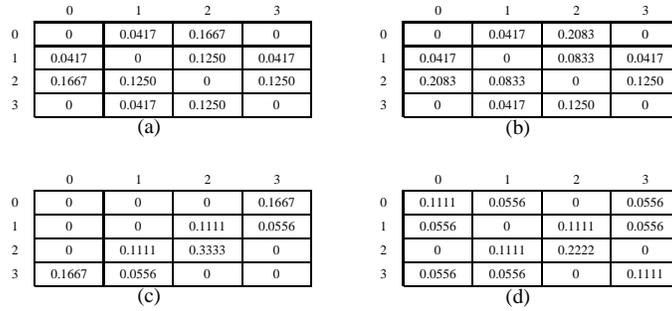
Figure 8: Normalized co-occurrence matrices (*NGLCM*) of corresponding co-occurrence matrices (*GLCM*) in Figure 7 at directions (a) $\theta = 0°$, (b) $\theta = 90°$, (c) $\theta = 45°$ and (d) $\theta = 135°$.

2.4. Feature Extraction using 2D-DWT and GLCM

In the discrete wavelet decomposition, the output detail images give the detail coefficients of the original image. It is found that, the approximation sub-image carries little energy due to which it is not taken in the texture analysis of mammographic ROI. But the wavelet detail coefficients provide the texture descriptors of the mammographic ROI. Using 2-D DWT, the three detail coefficient matrices at each resolution level are obtained, which represent horizontal, vertical, and diagonal sub-structures of the ROI as shown in Figure 9. Then the gray-level co-occurrence matrices are calculated at each resolution level by taking the absolute value of each coefficient in the corresponding matrices. For analysis of texture patterns of each ROI, the following five texture descriptors such as energy, correlation, entropy, sum variance, and sum average are calculated [13]. The texture features of ROIs are computed using *GLCM*. Now $p(i, j)$ is the (i, j)th entry of normalized *GLCM*. Let $p_x(i)$ is the ith entry in the marginal probability matrix by summing the rows of $p(i, j)$, defined as, $p_x(i) = \sum_{j=1}^{G} p(i, j)$, where G is the number of distinct gray levels in the quantized ROI. Similarly, $p_y(j) = \sum_{i=1}^{G} p(i, j)$ and $p_{x+y}(k) = \sum_{\substack{i=1 \\ i+j=k}}^{G} \sum_{j=1}^{G} p(i, j)$, $k = 2, 3, ..., 2G$. The expressions for different texture feature descriptors (*FD*) are given in Table 1. The feature matrix of each ROI generated using the 2*D*-DWT and *GLCM* method is described

in Algorithm 1.
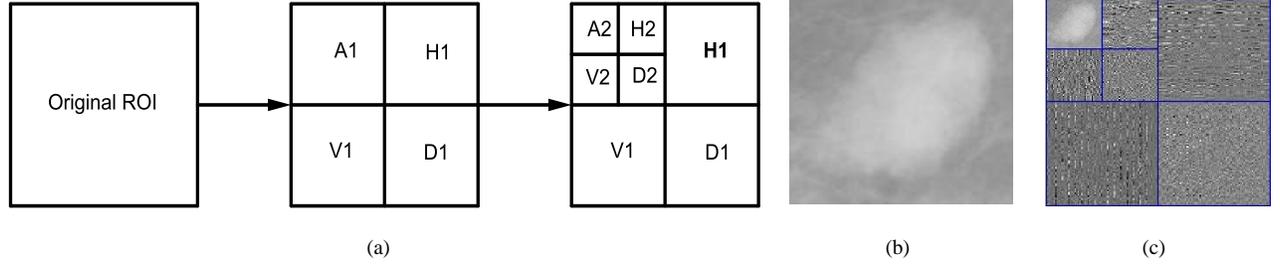
Figure 9: $2D$ discrete wavelet transform of mammographic ROI, (a) wavelet decomposition at two resolution level, (b) original ROI (mdb015) and (c) transformed ROI.

Table 1: Computation of feature descriptors for mammographic ROIs.

Feature Descriptor	Name	Computation
FD_1	Energy	$\sum\limits_{i=1}^{G} \sum\limits_{j=1}^{G} \{p(i,j)\}^2$
FD_2	Correlation	$\dfrac{\sum\limits_{i=1}^{G} \sum\limits_{j=1}^{G} (i,j)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
FD_3	Entropy	$-\sum\limits_{i=1}^{G} \sum\limits_{j=1}^{G} p(i,j) \log(p(i,j))$
FD_4	Sum variance	$\sum\limits_{i=2}^{2G} (i - sum\,entropy)^2 p_{x+y}(i)$
FD_5	Sum average	$\sum\limits_{i=2}^{2G} i p_{x+y}(i)$

where, μ_x, μ_y, σ_x and σ_y are the means and standard deviations of p_x and p_y, and

$$sum\,entropy = -\sum_{i=2}^{2G} p_{x+y}(i) \log\{p_{x+y}(i)\}.$$

In Algorithm 1, a $2D$-DWT is applied on N mammographic ROIs to produce different detail coefficient matrices (DM) at r different directions such as horizontal, vertical and diagonal directions for l resolution levels. A co-occurrence matrix ($GLCM$) and its corresponding normalized co-occurrence matrix ($NGLCM$) are calculated from each DM in four directions ($p = 4$) i.e at $0°, 45°, 90°$ and $135°$ at a set distance D. Then all the feature descriptors (FD) mentioned in Table 1 are computed from each $NGLCM$ and combined to form a feature descriptor matrix (FDM). Thus, a feature matrix is generated by concatenating all the FDMs from all $NGLCM$s for whole N ROIs.

Algorithm 1 Feature matrix generation

Require: N: Total number of images in dataset

j and l: resolution level and total number of resolution levels respectively

DM: Detail coefficient matrix for various resolution level

r: Total number of directions in which DM is to be computed, here $r = 3$

$GLCM$: Co-occurrence matrix

$NGLCM$: Normalized co-occurrence matrix

θ and p: Direction and number of directions for computing $NGLCM$ respectively

D: Set distance for computing $GLCM$

FD and $FDM[1 : S, 1 : N]$: Feature descriptor and feature descriptor matrix

s: Number of feature descriptors

M: Total number of features

Ensure: $feature\_matrix[1 : M, 1 : N]$

Function wavedec() does wavelet transform of ROI. Function detcoef() extracts three detail and one approximation coefficient matrix from transformed ROI at lower resolution levels using DWT technique. And function graycomatrix() computes $GLCM$ from DM

1: Initialize the required values to l, r, p and s

2: $M \leftarrow l * r * p * s$

3: **for** $i \leftarrow 1$ to N **do**

4:      Read ROI_i

5:      $D \leftarrow 1$

6:      **for** $j \leftarrow 1$ to l **do**

7:          $TROI_i \leftarrow$ wavedec(ROI_i)   {$TROI_i$ is the wavelet transform of ROI_i}

8:          **for** $d \leftarrow 1$ to r **do**

9:              $DM_{jd} \leftarrow$ detcoef($TROI_i$)

10:             **for** $k \leftarrow 1$ to p **do**

11:                 $GLCM_{jd\theta_k} \leftarrow$ graycomatrix(DM_{jd}, θ_k, D)

12:                 $NGLCM_{jd\theta_k} \leftarrow GLCM_{jd\theta_k} /$ sum(elements of $GLCM_{jd\theta_k}$)

13:                 **for** $q \leftarrow 1$ to s **do**

14:                     Compute FD_s from $NGLCM_{jd\theta_k}$ and append to $FDM_{jd\theta_k}$

15:                 **end for**

16:             **end for**

17:         **end for**

18:         $D \leftarrow D + 1$

19:     **end for**

20: **end for**

21: $feature\_matrix \leftarrow$ concatenate (FDMs)

2.5. Feature Selection and Classification

The features extracted from the textures of ROIs are expressed as mathematical descriptions. This helps the classifier to distinguish the breast tissues as normal, benign or malignant. However, one major problem lies with the large number of features that is very difficult to determine which feature or combination of features achieves better classification accuracy rate [3]. Therefore, it is important to select a suitable and optimized set of features from a high dimensional feature matrix that has the ability to distinguish between different types of mammograms. In this scheme, two statistical methods such as *two-sample t* and *F-test* have been used comparatively to select the most significant features from the fetaure matrix. *Two-sample t* and *F-tests* are performed on two classes, and a test decision is returned for the null hypothesis that the data in two vectors v_1 and v_2 come from normal distributions with equal means. The test determines whether the data from two vectors v_1 and v_2 are related or not. In the proposed feature selection algorithm, a null hypothesis value, $h = 1$ indicates that the null hypothesis is incorrect and rejected. An incorrect null hypothesis implies that, data from two vectors v_1 and v_2 are different and independent. In the *two-sample t* and *F-test* method, the t and F values are computed as,

$$t = \frac{\left|\mu_{v_1} - \mu_{v_2}\right|}{\sqrt{\frac{(\sigma_{v_1})^2}{N_{v_1}} + \frac{(\sigma_{v_2})^2}{N_{v_2}}}} \tag{7}$$

$$F = \frac{S_{v_1}^2}{S_{v_2}^2} \tag{8}$$

Where, N_{v_1} and N_{v_2} are the numbers of ROIs in two classes. Here, μ_{v_1} and μ_{v_2} are means, σ_{v_1} and σ_{v_2} are standard deviations, and S_{v_1} and S_{v_2} are the variances of two classes. A higher t and F value indicates more significant differences between the means of the two vectors. For a certain threshold t and F value, corresponding p_1 and p_2 values define probabilities of obtaining a t and F value more than the threshold. A significance level, α defines the lower threshold for the p_1 and p_2 values. The value of α is in the range 0 and 1. As the α value decreases from one to zero, the selection of the number of features reduces. The selection of significant features has been described in Algorithm 2.

In this paper, the artificial neural network has been used as a classifier. Artificial neural network is a powerful parallel dynamic system consisting of multiple simple and interconnected processing units, that performs tasks like the biological brains. A neural network can perform the necessary transformation operation automatically by means of neuron's state response to their input information. These networks are trained with a set of samples known as the training set. The network is trained by learning the values of its internal parameter from the training set so that, an input leads to a specific output. A feed-forward back-propagation multilayer neural network (BPNN) is one of the most common and effective network structures used for classification in the feature space. A set of R selected significant features are supplied to BPNN for the classification purpose. Different measures of classification are obtained with the different number of significant features by varying the α values and changing the number of neurons (n) in the hidden layer of neural network as shown in Figure 1. The performance of the classifier is evaluated with the help of a confusion matrix [22]. The confusion matrix is a table that shows the output and actual class classification accomplished by the classifier. The confusion matrix for two class-problem and corresponding measures of performance are represented in Figure 10.

Algorithm 2 Feature Selection

Require: $feature\_matrix[1:M, 1:N]$, $target\_class[1:N]$

   α: Significance level

Ensure: $selected\_feature_1[1:R, 1:N]$ and $selected\_feature_2[1:R, 1:N]$

   R: Total number of selected features

   Functions ttest() and vartest() compute the null hypothesis values of two vectors at different values of significance

   level, by *two-sample t* and *F-test* respectively

1: Create two empty vectors v_1 and v_2

2: Initialize α, $0 < \alpha < 1$

3: **for** $i \leftarrow 1$ to M **do**

4:    Clear contents of vector v_1 and vector v_2

5:    **for** $j \leftarrow 1$ to N **do**

6:       **if** $target\_class[j] = 1$ **then**

7:          Append $feature\_matrix[i, j]$ to v_1

8:       **else**

9:          Append $feature\_matrix[i, j]$ to v_2

10:      **end if**

11:   **end for**

12:   $h_1[i] \leftarrow$ ttest(v_1, v_2, α)

13:   $h_2[i] \leftarrow$ vartest(v_1, v_2, α)

14:   **for** $l \leftarrow 1$ to 2 **do**

15:      **if** $h_k[i] = 1$ **then**

16:         Append $feature\_matrix[i, 1:N]$ to $selected\_feature_k$

17:      **end if**

18:   **end for**

19: **end for**

Confusion matrix   Performance measures

		Positive	Negative	
Output class	Positive	True Positive (TP)	False Positive (FP)	Positive Predictive Value =TP/(TP+FP)
	Negative	False Negative (FN)	True Negative (TN)	Negative Predictive Value = TN/(TN+FN)
Performance measures		True Positive Rate (Sensitivity) = TP/(TP+FN)	True Negative Rate (Specificity) = TN/(TN+FP)	Accuracy = (TP+TN)/ Total number of samples

Positive   Negative

Actual class

Figure 10: Confusion matrix for two-class problem with different performance measures.

Among the different measures mentioned in Figure 10, specificity and sensitivity are two important parameters for performance evaluation. Sensitivity determines the percentage of true positive rate while specificity determines the percentage of true negative rate. Positive predictive value and negative predictive value calculate the percentage of positive and negative characteristics in case of positive and negative tests respectively. For an ideal performance, both specificity and sensitivity should be high. The evaluation of a classifier performance can also be accomplished by means of receiver operating characteristics (ROC) curves [3]. It is a two dimensional plot of true positive rate (sensitivity) versus false positive rate (1-specificity) in vertical and horizontal axes respectively. The area under the ROC curve referred by an index AUC is an important factor for evaluating the classifier performance. The value of AUC is 1.0 is a perfect performance of the classifier.

3. Experimental Results and Analysis

To validate the proposed feature extraction and mammogram classification scheme, simulations have been carried out in the MATLAB environment. For the analysis of the proposed method, mammographic images are taken from two databases such as Mammographic Image Analysis Society (MIAS) database [18] and Image Retrieval and Medical Applications (IRMA) project [19]. The MIAS database is built by Suckling et al. and openly available for scientific research. The mammographic image database in IRMA project is made by Deserno et al., who collected images from several other databases including Digital Database for Screening Mammography (DDSM). Both MIAS and IRMA databases provide appropriate informations based on types of background tissues, and the class of abnormalities present in the mammograms. The class of abnormality consists of normal-abnormal class, and again based upon the severity of abnormality; the abnormal class is divided into two sub-classes such as benign and malignant. The MIAS database contains 322 images, which are categorized into three according to tissue types like fatty, fatty-glandular and

dense-glandular. Out of 322 images, 207 images are normal, 115 images are abnormal; and again among abnormal images the number of benign and malignant types are 64 and 51 respectively. We considered all the images for our experiment from this database. In IRMA project, the database is divided into 12 and 20 class problems. In 12 class problem, the mammograms are categorized according to tissue density, and each category is divided into three classes; normal, benign and malignant. In 20 class problem, the mammograms are of two categories of different types of lesions. The 12 class database consists of mammograms of four tissue types; almost entirely fatty, scattered fibro glandular, heterogeneously dense and extremely dense. This database consists of 2796 images out of which 2576 images are from DDSM database. We have taken a total of 550 DDSM images from 12 class problem, out of which 300 images are normal and 250 images are abnormal. The abnormal class consists of 129 benign images and 121 malignant images. Each mammographic ROI has been taken of size 128×128 used in the feature extraction phase to find several types of features.

3.1. Results for feature extraction

In this paper, the symmetric biorthogonal 4.4 wavelet has been used to compute DWT of images. It has been observed that at $l = 2$, the 2D-DWT gives the suitable results on feature extraction. At each resolution level (j) the DWT results three detail coefficient matrices and thus a total of six detail coefficient matrices (DM) such as H_1, V_1, D_1 at $j = 1$, and H_2, V_2, D_2 at $j = 2$ are obtained in three different directions. Furthermore, four $GLCM$ and corresponding $NGLCM$ are computed from each detail coefficient matrix (DM) at each resolution level. The resolution level (j) of wavelet transform act as the distance parameter (D) for $GLCM$ computation. The value of D has been taken 1 and 2 for resolution level $j = 1$ and $j = 2$ respectively. From each $NGLCM$, a total of five feature descriptors ($s = 5$) such as energy, correlation, entropy, sum variance and sum average are extracted and consequently, form a feature descriptor matrix. Thus, for $l = 2$, $r = 3$, $p = 4$ and $s = 5$, a total 120 ($M = l \times r \times p \times s$) features are extracted from N number of ROIs. This M number of features are kept in rows with corresponding N number of ROIs in columns to generate a feature matrix, which is to be used in feature selection algorithm. Tables 2, 3, 4, and 5 show the values of different texture feature descriptors for different types of ROIs at each resolution level (j).

3.2. Results for mammogram type classification using most significant features and neural network

During the experiment, different number of significant features (R) have been selected through *two-sample t* and *F-test* methods. Figure 11 shows the variation of the number of selected features with respect the various values of α for MIAS and DDSM databases. It has been observed that, the reduced number of selected features (R) is obtained at lower values of significance level (α) using both statistical methods. It is also observed that for same value of α, the dimension reduction is more in DDSM images as compared to MIAS images. The selected features were used in the classifier to find the optimal classification accuracy rate. After getting several sets of significant features, we conducted the classification experiments on both MIAS and DDSM dataset using a three-layer BPNN. In the experiment, 70% of the total dataset have been used for training. From the remaining dataset, 15% data were used for validation and rest 15% were used for testing purposes. The different number of images are used in the three phases of the classifier which is shown in Table 6.

Table 2: Different values of various feature descriptors at $\theta = 0°$ with set distance $D = 1$ for $j = 1$ and $D = 2$ for $j = 2$.

Type of ROI	Detail coefficien	Feature Descriptors									
		MIAS					DDSM				
		FD_1	FD_2	FD_3	FD_4	FD_5	FD_1	FD_2	FD_3	FD_4	FD_5
Normal	H_1	0.2779	487.5895	2.0317	68.9951	9.7276	0.3516	280.7523	1.8023	46.6090	8.0955
	H_2	0.0970	692.6964	2.8604	41.6123	8.1096	0.1532	844.4912	2.2438	91.2830	11.0730
	V_1	0.1877	466.5697	1.9693	61.5468	9.0858	0.1731	459.2810	2.1313	55.5010	8.7921
	V_2	0.0996	767.3355	2.8207	51.9270	8.8618	0.1101	653.7500	2.6209	51.4534	8.5875
	D_1	0.2179	444.1351	1.8202	65.0145	9.2864	0.1450	549.6110	2.2607	57.1411	8.9572
	D_2	0.2331	347.4291	2.2023	43.2433	7.9591	0.2361	325.8900	2.0760	47.0090	8.0912
Benign	H_1	0.2281	419.9608	1.9828	52.9509	8.6760	0.9296	219.4603	0.2129	98.9842	9.9898
	H_2	0.0968	874.2193	2.7809	57.3901	9.2558	0.2355	294.7308	1.6366	34.1920	6.9669
	V_1	0.3602	239.8866	1.6763	48.3035	8.0619	0.9139	213.1101	0.2764	95.1445	9.9783
	V_2	0.0884	741.4346	2.8748	47.4457	8.5482	0.5363	184.5213	1.1469	47.6322	7.7436
	D_1	0.2263	432.9716	1.8142	65.5525	9.3108	0.7223	148.4800	0.5889	54.3605	7.8384
	D_2	0.1219	668.9592	2.5554	56.7644	9.0768	0.9639	130.3821	0.1285	61.9740	7.9811
Malignant	H_1	0.2096	669.7592	1.9406	88.8764	10.8185	0.9341	217.4412	0.1970	97.1415	9.9778
	H_2	0.1253	663.9185	2.4817	56.4908	9.0556	0.3882	246.0833	1.3712	42.7977	7.5013
	V_1	0.4441	216.5229	1.4351	51.0913	8.0865	0.9127	214.5800	0.2823	95.3071	9.9841
	V_2	0.2024	498.4734	2.0593	65.2262	9.4488	0.2998	239.7766	1.5814	40.7870	7.4602
	D_1	0.3583	258.8663	1.5051	53.2372	8.3248	0.3350	284.9154	1.2925	58.8471	8.5895
	D_2	0.1205	645.2020	2.6058	53.9703	8.8808	0.7726	143.5648	0.5099	55.7352	7.8725

H_1, V_1, D_1: horizontal, vertical and diagonal detail coefficient matrices at $j = 1$.
H_2, V_2, D_2: horizontal, vertical and diagonal detail coefficient matrices at $j = 2$.
FD_1, FD_2, FD_3, FD_4 and FD_5 are feature descriptors defined in Table 1.

Table 3: Different values of various feature descriptors at $\theta = 90°$ with set distance $D = 1$ for $j = 1$ and $D = 2$ for $j = 2$.

Type of ROI	Detail coefficien	Feature Descriptors									
		MIAS					DDSM				
		FD_1	FD_2	FD_3	FD_4	FD_5	FD_1	FD_2	FD_3	FD_4	FD_5
Normal	H_1	0.2326	479.8930	2.1199	71.5657	9.7237	0.2907	276.4500	1.8987	47.2782	8.0948
	H_2	0.0903	677.5727	2.8725	42.0656	8.1228	0.1745	819.9701	2.1207	98.9453	11.0805
	V_1	0.1969	468.5895	1.9439	60.1313	9.0849	0.1951	461.4608	2.0561	54.6023	8.7895
	V_2	0.1015	762.7643	2.7947	51.3033	8.8450	0.1123	675.1321	2.6809	48.5760	8.5925
	D_1	0.2177	442.5444	1.8138	65.2780	9.2845	0.1451	548.7622	2.2601	57.3110	8.9543
	D_2	0.2250	346.4965	2.2173	43.3453	7.9539	0.2515	324.4001	2.0657	47.2960	8.0953
Benign	H_1	0.2011	413.7083	2.0464	55.6596	8.6787	0.9139	213.2400	0.2881	95.4910	9.9834
	H_2	0.0897	864.0826	2.8057	58.5130	9.2778	0.2197	270.7105	1.7934	34.3020	6.9534
	V_1	0.4352	241.9732	1.5906	48.3573	8.0619	0.9259	219.8802	0.2182	96.9100	9.9857
	V_2	0.1005	746.1557	2.8445	47.4000	8.5461	0.5974	194.7628	0.9838	50.0120	7.7476
	D_1	0.2291	433.3259	1.8093	65.8895	9.3106	0.7195	148.5254	0.5898	54.3786	7.8380
	D_2	0.1213	674.2864	2.5647	57.0645	9.0855	0.9639	130.3410	0.1285	61.9856	7.9811
Malignant	H_1	0.1930	664.6917	1.9807	92.7172	10.8194	0.9127	212.2101	0.2813	95.1182	9.9746
	H_2	0.1234	637.6616	2.4486	56.7353	9.0446	0.3230	231.8769	1.5543	41.3890	7.4696
	V_1	0.5052	216.8881	1.3503	50.3845	8.0858	0.9182	222.9300	0.2372	96.7241	9.9957
	V_2	0.2093	506.7676	2.0243	64.7834	9.4423	0.3256	254.5963	1.4196	42.2070	7.4696
	D_1	0.3559	258.9083	1.5028	53.5048	8.3264	0.3324	285.0214	1.2995	58.7192	8.5893
	D_2	0.1279	634.4545	2.5620	53.8055	8.8830	0.7804	142.9786	0.4945	55.9823	7.8772

Table 4: Different values of various feature descriptors at $\theta = 45°$ with set distance $D = 1$ for $j = 1$ and $D = 2$ for $j = 2$.

Type of ROI	Detail coefficien	Feature Descriptors									
		MIAS					DDSM				
		FD_1	FD_2	FD_3	FD_4	FD_5	FD_1	FD_2	FD_3	FD_4	FD_5
Normal	H_1	0.2264	480.5639	2.1465	70.6623	9.7249	0.2847	276.8162	1.9086	46.7880	8.0924
	H_2	0.0879	681.0941	2.8792	42.3220	8.1304	0.1485	827.6910	2.2421	94.6770	11.0822
	V_1	0.1864	467.7653	1.9784	60.9509	9.0844	0.1721	459.9882	2.1365	55.0801	8.7910
	V_2	0.0926	766.4563	2.8347	51.7463	8.8526	0.0978	671.1874	2.7181	49.0870	8.5902
	D_1	0.2174	443.6977	1.8201	64.9795	9.2851	0.1449	550.2200	2.2622	56.8473	8.9557
	D_2	0.2095	348.9067	2.2437	43.2632	7.9568	0.2114	329.7618	2.1021	44.7035	8.0979
Benign	H_1	0.1974	415.2558	2.0749	54.6695	8.6772	0.9139	213.2245	0.2897	95.4060	9.9831
	H_2	0.0906	858.3919	2.8054	58.5568	9.2685	0.2050	271.6973	1.8368	33.2631	6.9557
	V_1	0.3516	241.0157	1.6931	47.6765	8.0624	0.9139	213.1512	0.2776	95.0765	9.9784
	V_2	0.0898	748.6347	2.8727	47.1760	8.5424	0.5371	184.2814	1.1599	47.4170	7.7424
	D_1	0.2249	434.0044	1.8205	65.2272	9.3105	0.7210	148.7421	0.5919	54.2596	7.8378
	D_2	0.1199	668.5166	2.5641	56.9308	9.0802	0.9657	130.4118	0.1256	62.1110	7.9824
Malignant	H_1	0.1883	666.8222	2.0078	91.3533	10.8204	0.9123	212.2347	0.2843	95.0021	9.9747
	H_2	0.1233	643.4498	2.4670	56.9562	9.0486	0.3243	231.1132	1.5688	41.1360	7.4945
	V_1	0.4392	216.4788	1.4470	50.4909	8.0867	0.9123	214.6300	0.2834	95.1750	9.9846
	V_2	0.2056	494.2931	2.0431	65.5455	9.4444	0.3078	239.7369	1.5993	40.3093	7.4633
	D_1	0.3515	259.9105	1.5157	52.6011	8.3250	0.3352	285.1724	1.3013	58.1961	8.5885
	D_2	0.1197	643.0346	2.6010	53.7630	8.8819	0.7873	143.3512	0.5003	55.8630	7.8764

Table 5: Different values of various feature descriptors at $\theta = 135°$ with set distance $D = 1$ for $j = 1$ and $D = 2$ for $j = 2$.

Type of ROI	Detail coefficien	Feature Descriptors									
		MIAS					DDSM				
		FD_1	FD_2	FD_3	FD_4	FD_5	FD_1	FD_2	FD_3	FD_4	FD_5
Normal	H_1	0.2269	480.7278	2.1466	70.6295	9.7249	0.2881	276.6600	1.9064	47.0070	8.0924
	H_2	0.0900	684.9990	2.8778	42.1644	8.1296	0.1502	828.8966	2.2446	94.3992	11.0817
	V_1	0.1861	467.8358	1.9786	60.8743	9.0849	0.1733	459.8334	2.1368	55.1521	8.7912
	V_2	0.0954	760.9155	2.8241	51.8240	8.8542	0.0978	671.1200	2.7155	49.0154	8.5895
	D_1	0.2164	443.6506	1.8209	65.1938	9.2854	0.1448	549.9213	2.2617	56.9170	8.9548
	D_2	0.2122	347.3934	2.2333	43.1518	7.9576	0.2160	329.5301	2.1044	44.8295	8.0979
Benign	H_1	0.1970	414.9199	2.0684	54.9266	8.6774	0.9139	213.2773	0.2873	95.4790	9.9833
	H_2	0.0910	859.0862	2.8037	58.3728	9.2670	0.2068	271.7451	1.8327	33.4358	6.9550
	V_1	0.3523	240.9230	1.6918	47.7658	8.0624	0.9139	2132.0892	0.2766	95.1265	9.9782
	V_2	0.0901	743.2662	2.8825	47.8501	8.5424	0.5364	184.2615	1.1616	47.362	7.7424
	D_1	0.2239	433.9719	1.8191	65.1063	9.3110	0.7229	148.6612	0.5919	54.2346	7.8348
	D_2	0.1183	672.4989	2.5709	56.7228	9.0802	0.9671	130.4906	0.1244	62.0631	7.9820
Malignant	H_1	0.1880	666.9464	2.0085	91.2963	10.8207	0.9123	212.2300	0.2835	95.0051	9.9747
	H_2	0.1227	647.4172	2.4699	56.9847	9.0494	0.3285	232.2644	1.5734	40.9761	7.4945
	V_1	0.4382	216.5055	1.4477	50.4490	8.0867	0.9123	214.5783	0.2840	95.1964	9.9838
	V_2	0.2016	492.9544	2.0538	65.8033	9.4437	0.3088	239.6123	1.6006	40.2593	7.4645
	D_1	0.3474	259.8169	1.5155	52.7306	8.3250	0.3358	285.1300	1.3011	58.2996	8.5887
	D_2	0.1165	642.1973	2.6072	53.6483	8.8819	0.7916	143.4742	0.4983	55.8124	7.8766

(a) Normal-abnormal class   (b) Benign-malignant class
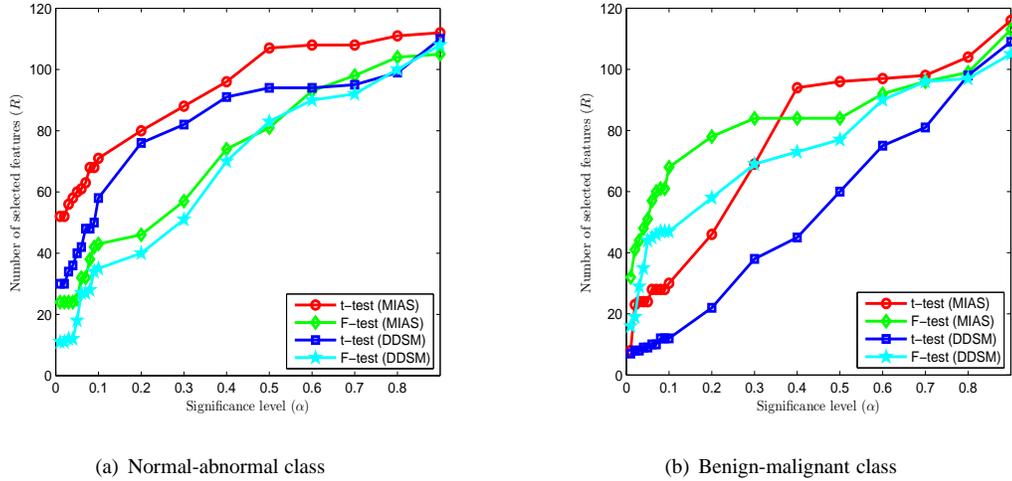
Figure 11: Feature selection by *two-sample t* and *F-test* method. The reduced number of selected features (R) is obtained at lower values of significance level (α).

Table 6: Various Numbers of mammographic images used in different phases of BPNN classifier.

Mammogram database used	Mammogram image class	Total number of images	Number of mammographic images		
			Training (70%)	Testing (15%)	Validation (15%)
MIAS	Normal–abnormal	322	226	48	48
	Benign–malignant	115	81	17	17
DDSM	Normal–malignant	550	384	83	83
	Benign–malignant	250	174	38	38

As mentioned in the proposed model (Figure 1), the magnitude of significance level (α) for feature selection and number of neurons in the hidden layer (n) of the BPNN influence the performance of the classifier. It is very difficult to find the best significant feature set through which the classifier achieves optimal performance. Therefore, several feature sets obtained at various values of significance level (α) are used in the classifier to find the optimum results. In fact, for the same value of α, the classifier achieves different performance results at the different number of hidden layer neurons (n). In our experiments, the values of n have been chosen as 5, 10, 15 and 20 to investigate the best performance. It has been found that, at $n = 15$ with respect to different α, the classifier attains its best performance. Different performance measures, including sensitivity, specificity and test classification accuracy using two feature

selection methods are presented in Table 7. It is observed that, the higher classification accuracy rates are obtained with *two-sample t-test* feature selection method for both the databases. These values are as 98.0% (normal-abnormal), 94.2% (benign-malignant) for MIAS datbase, and 98.8% (normal-abnormal), 97.4% (benign-malignant) for DDSM database.

Table 7: Different values of performance measures of the classifier using two feature selection methods with $n = 15$.

Mammogram Database	Selection method	Significance level (α)	Performance measures (%)							
			normal-abnormal				benign-malignant			
			R	Sensitivity	Specificity	Accuracy	R	Sensitivity	Specificity	Accuracy
MIAS	*Two-sample t-test*	0.9	112	77.8	93.3	87.5	116	87.5	88.9	88.2
		0.5	107	88.2	93.3	91.7	96	100	90.0	94.2
		0.2	80	100	97.0	98.0	46	77.8	87.5	82.4
		0.05	60	85.7	97.1	93.8	24	75.0	77.8	76.5
		0.01	52	100	77.4	85.4	08	66.7	54.5	58.9
	Two-sample F-test	0.9	105	69.2	100	91.7	113	87.5	77.8	82.4
		0.5	81	75.0	93.8	87.5	84	100	66.7	82.4
		0.2	46	88.2	100	95.8	78	88.9	87.5	88.2
		0.05	25	82.6	96.0	89.6	51	87.5	66.7	76.5
		0.01	24	64.7	93.5	83.3	32	62.5	77.8	70.6
DDSM	*Two-sample t-test*	0.9	110	93.5	86.4	90.3	109	93.3	86.9	89.4
		0.5	94	91.4	83.3	87.9	60	94.1	90.4	92.1
		0.2	76	100	97.9	98.8	22	100	94.7	97.4
		0.05	40	100	95.1	97.6	09	92.3	96.0	94.7
		0.01	30	97.8	89.4	93.9	07	93.3	91.3	92.1
	Two-sample F-test	0.9	108	93.6	84.6	89.1	105	86.6	86.9	86.8
		0.5	83	93.3	89.4	91.5	77	93.3	86.9	89.4
		0.2	40	100	95.1	97.6	58	93.8	90.9	92.1
		0.05	18	95.6	92.1	93.9	44	86.7	91.7	89.4
		0.01	11	95.4	84.6	90.3	16	92.3	88.0	89.5

We have also evaluated the performance of two feature selection methods by comparing the obtained *AUC* values of ROC curves at different magnitudes of significance level (α) with respect to the different number of hidden layer neurons (n). A heat-map has been used to demonstrate the comparison as shown in Figure 12. It is clearly observed that, the best values of *AUC* have been accomplished with the significance level (α) of 0.2 for classification of MIAS and DDSM datasets. One tenuous deviation in AUC is observed at $\alpha = 0.5$ for benign-malignant classification in MIAS data. This is due to some irregular tissue pattern in mammograms.

During our experiment, we compared the performances achieved by the BPNN classifier along with statistical
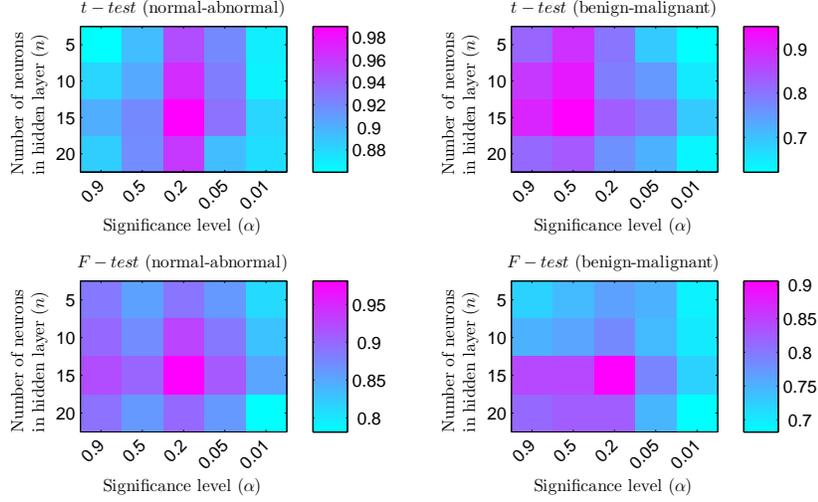
(a) MIAS Database
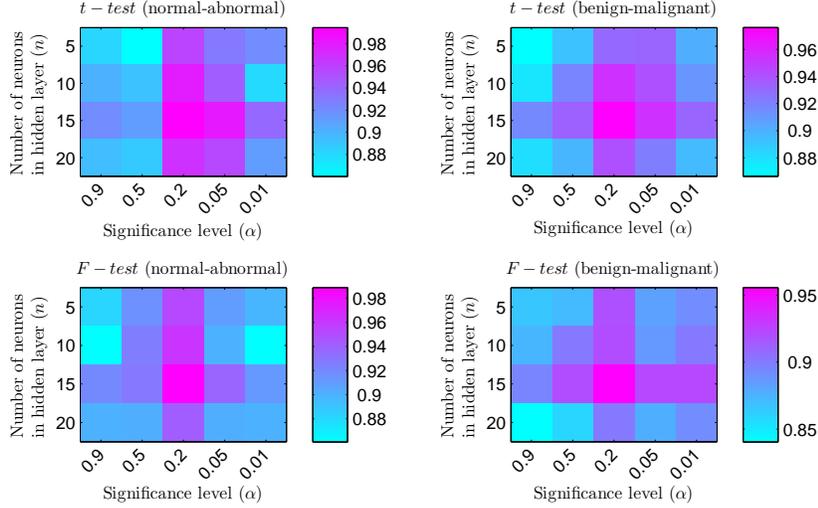
(b) DDSM Database

Figure 12: Heat-maps of AUC measurements using selective feature sets by varying significance levels (α) in two-sample t and F-test methods with respect to different numbers of hidden layer neurons (n) in BPNN classifier.

two-sample t and *F-test* method with random forest method [23]. ROC curves obtained using the proposed scheme, and the random forest method are shown in Figure 13. It has been inferred that the proposed scheme outperforms the random forest method with respect to AUC measurements. Table 8 presents the comparison of the test accuracies and AUC measurements for BPNN classifier and random forest technique. The maximum AUC values obtained by the BPNN and *t-test* method are 0.9899 and 0.9504 in MIAS, and 0.9945 and 0.9761 in DDSM database for both normal-abnormal and benign-malignant pattern classification. It is clearly observed that the *two-sample t-test* has quite higher-performance values in comparison to other methods mentioned for both databases.
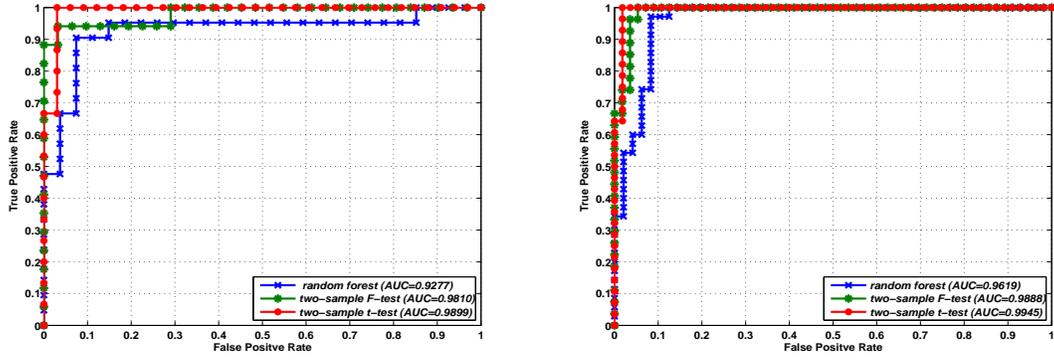
(a) ROC for prediction of abnormal tissues using MIAS database (b) ROC for prediction of abnormal tissues using DDSM database
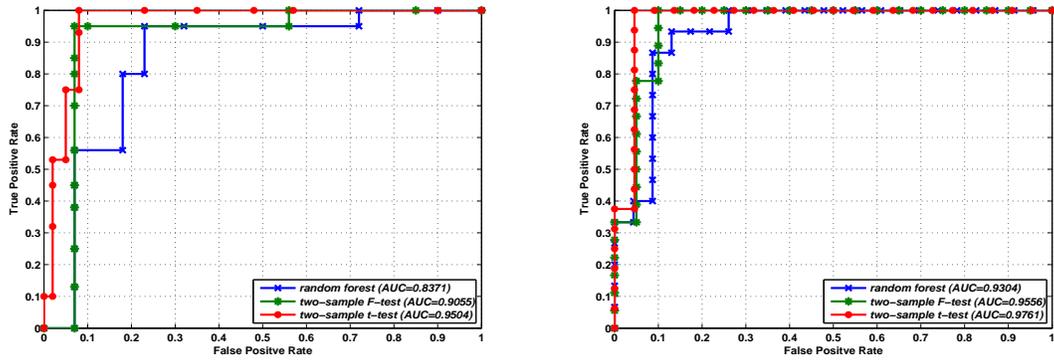
(c) ROC for prediction of malignant tissues using MIAS database (d) ROC for prediction of malignant tissues using DDSM database

Figure 13: Comparison of ROC curves of mammogram classification using proposed scheme and random forest method utilizing optimal selected feature set.

Table 8: Comparison of optimal test accuracy rates and *AUC* measurements between proposed and random forest methods.

Mammogram database used	Mammogram image class	Test accuracy rate (%)			*AUC* measurements		
		BPNN and *t-test* method	BPNN and *F-test* method	Random forest method	BPNN and *t-test* method	BPNN and *F-test* method	Random forest method
MIAS	Normal–abnormal	98.0	95.8	93.3	0.9899	0.9810	0.9277
	Benign–malignant	94.2	88.2	82.4	0.9504	0.9055	0.8371
DDSM	Normal–abnormal	98.8	97.6	92.8	0.9945	0.9888	0.9619
	Benign–malignant	97.4	92.1	89.5	0.9761	0.9556	0.9304

Further, a training error comparison for the proposed scheme and random forest method is shown in Figure 14 to evaluate the training convergence. The training error of the classifier is expressed as mean squared error (mse) values at multiple numbers of training iteration in BPNN classifier and number of growing trees in the random forest method. The mean squared error is the average squared difference between output classes generated by the classifier and existing actual classes. The training error curves of *two-sample t-test* method shows that it converges faster than other methods for both normal-abnormal and benign-malignant mammogram classes. Finally, a comparative analysis between the proposed scheme with other existing schemes has been made and shown in Table 9. It has been clearly observed that the proposed scheme performs better classification than other schemes with respect to different performance measures. However, the proposed scheme requires an additional feature selection phase using statistical tests unlike random forest and DWT methods, which are simple to implement.
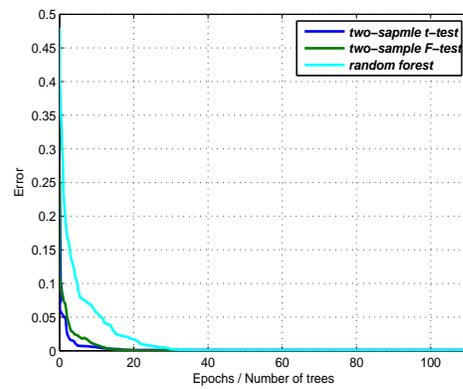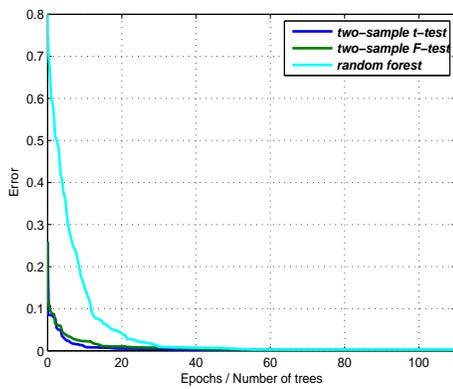
(a) For normal-abnormal ROI classification (MIAS database) (b) For normal-abnormal ROI classification (DDSM database)
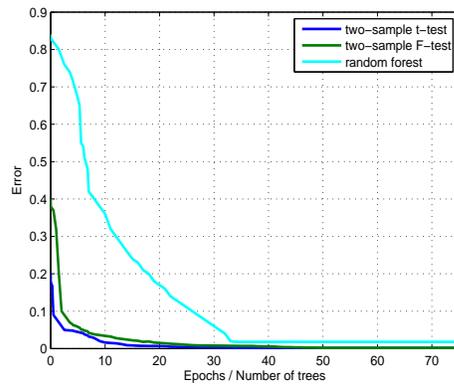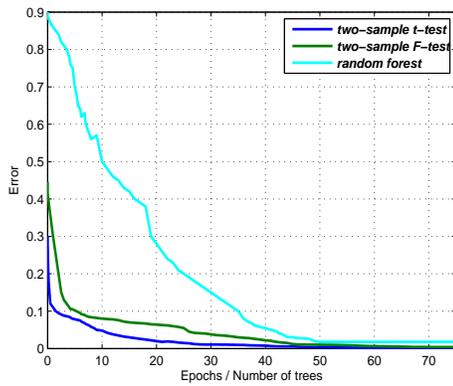
(c) For benign-malignant ROI classification (MIAS database) (d) For benign-malignant ROI classification (DDSM database)

Figure 14: Training error comparison by proposed neural network using *two-sample t* and *F-test* feature selection methods and random forest classification method.

Table 9: Performance comparison by different methods with the proposed scheme.

Method reference	Techniques	Classification performance measure
Prathibha et al. (2010) [8]	DWT, Nearest negihbor classifier	$AUC = 0.95$ (normal-abnormal)
Buciu et al. (2011) [9]	Gabor wavelets and PCA, SVM classifier	sensitivity=97.56%, Specificity=60.86% $AUC = 0.79$ (normal-abnormal) sensitivity=84.61%, Specificity=80.0% $AUC = 0.78$ (benign-malignant)
Mutaz et al. (2011) [16]	GLCM, ANN classifier	Sensitivity=91.6%, Specificity=84.17% (benign-malignant)
Jona et al. (2012) [17]	GLCM, SVM classifier	Accuracy=94.0% (normal-abnormal)
Görgel et al. (2012) [10]	DWT, SVM classifier	Accuracy rate=84.8% (benign-malignant)
Görgel et al. (2013) [11]	SWT, SVM classifier	Accuracy rate=96.0% (normal-abnormal) Accuracy rate=93.59% (benign-malignant)
Proposed scheme	Combination of DWT and GLCM, BPNN	sensitivity=100%, Specificity=97.0% Accuracy =98.0%, $AUC = 0.9899$ (normal-abnormal,MIAS) sensitivity=100%, Specificity=90.0% Accuracy =94.2%, $AUC = 0.9504$ (benign-malignant, MIAS) sensitivity=100%, Specificity=97.9% Accuracy =98.8%, $AUC = 0.9945$ (normal-abnormal, DDSM) sensitivity=100%, Specificity=94.7% Accuracy =97.4%, $AUC = 0.9761$ (benign-malignant, DDSM)

4. Conclusion

In this paper, we propose an efficient mammogram classification scheme to support the decision of radiologists. The scheme utilizes 2D-DWT and GLCM in succession to derive feature matrix form mammograms. To select the relevant features from the feature matrix, both *t-test* and *F-test* are applied. It is observed that, t-test based relevant features achieves higher classification accuracy with BPNN than that of *F-test*. To validate the efficacy of the suggested scheme, simulation has been carried out using MIAS and DDSM databases. Its competent schemes are also simulated in the similar platform. Comparative analysis with respect to accuracy and AUC of ROC reveals that the suggested scheme outperforms other schemes. An accuracy of 98.0% and 94.2% have been obtained for normal-abnormal and benign-malignant respectively in MIAS database. The similar parameters are 98.8% and 97.4% are achieved in DDSM database.

References

[1] Globocan project 2012, International Agency for Research on Cancer (iarc), World Health Organisation: cancer fact sheets, `http://globocan.iarc.fr`.

[2] L. Tabar, P. Dean, Mammography and breast cancer: the new era, International Journal of Gynecology & Obstetrics 82 (3) (2003) 319–326.

[3] H. Cheng, X. Shi, R. Min, L. Hu, X. Cai, H. Du, Approaches for automated detection and classification of masses in mammograms, Pattern recognition 39 (4) (2006) 646–668.

[4] A. Dhawan et al, Analysis of mammographic microcalcifications using grey-level image structure features, IEEE Transactions on Medical Imaging 15 (3) (1996) 246–259.

[5] D. Wei, H.-P. Chan, N. Petrick, B. Sahiner, M. A. Helvie, D. D. Adler, M. M. Goodsitt, False-positive reduction technique for detection of masses on digital mammograms: Global and local multiresolution texture analysis, Medical Physics 24 (6) (1997) 903–914.

[6] S. Liu, C. F. Babbs, E. J. Delp, Multiresolution detection of spiculated lesions in digital mammograms, Image Processing, IEEE Transactions on 10 (6) (2001) 874–884.

[7] E. A. Rashed, I. A. Ismail, S. I. Zaki, Multiresolution mammogram analysis in multilevel decomposition, Pattern Recognition Letters 28 (2) (2007) 286–292.

[8] B. Prathibha, V. Sadasivam, Breast tissue characterization using variants of nearest neighbour classifier in multi texture domain, IE (I) journal 91 (2010) 7–13.

[9] I. Buciu, A. Gacsadi, Directional features for automatic tumor classification of mammogram images, Biomedical Signal Processing and Control 6 (4) (2011) 370–378.

[10] P. Gorgel, A. SERTBAŞ, N. Kilic, O. OSMAN, et al., Mammographic mass classification using wavelet based support vector machine, IU-Journal of Electrical & Electronics Engineering 9 (1) (2012) 867–875.

[11] P. Görgel, A. Sertbas, O. N. Ucan, Mammographical mass detection and classification using local seed region growing–spherical wavelet transform (lsrg–swt) hybrid scheme, Computers in biology and medicine 43 (6) (2013) 765–774.

[12] A. S. Kurani, D.-H. Xu, J. Furst, D. S. Raicu, Co-occurrence matrices for volumetric data, in: 7th IASTED International Conference on Computer Graphics and Imaging, Kauai, USA, 2004, pp. 447–452.

[13] R. M. Haralick, K. Shanmugam, I. H. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man and Cybernetics, (6) (1973) 610–621.

[14] F. Albregtsen, et al., Statistical texture measures computed from gray level coocurrence matrices, Image Processing Laboratory, Department of Informatics, University of Oslo (1995) 1–14.

[15] H.-P. Chan, B. Sahiner, K. L. Lam, N. Petrick, M. A. Helvie, M. M. Goodsitt, D. D. Adler, Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces, Medical Physics 25 (1998) 2007–2019.

[16] M. A. Al Mutaz, S. Dress, N. Zaki, Detection of masses in digital mammogram using second order statistics and artificial neural network, International Journal of Computer Science & Information Technology (IJCSIT) 3 (3) (2011) 176–186.

[17] J. Jona, N. Nagaveni, A hybrid swarm optimization approach for feature set reduction in digital mammograms, WSEAS Transactions on Information Science and Applications 9 (2012) 340–349.

[18] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S.-L. Kok, et al., The mammographic image analysis society digital mammogram database.

[19] T. M. Deserno, M. Soiron, J. E. de Oliveira, Texture patterns extracted from digitizes mammograms of different BI-RADS classes, Image Retrieval in Medical Applications project, release: V1.0 , 2012, `http://ganymed.imib.rwth-aachen.de/irma/datasets_en.php`.

[20] C. Gonzalez, E. Woods, Digital Image Processing, Pearson, 2011.

[21] S. Mallat, A wavelet tour of signal processing, Access Online via Elsevier, 1999.

[22] R. Kohavi, F. Provost, Glossary of terms, Machine Learning 30 (2-3) (1998) 271–274.

[23] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.