Computers in Biology and Medicine 43 (2013) 765–774

Contents lists available at SciVerse ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Mammographical mass detection and classification using Local Seed Region Growing–Spherical Wavelet Transform (LSRG–SWT) hybrid scheme

Pelin Görgel [a,*], Ahmet Sertbas [a], Osman N. Ucan [b]

[a] Department of Computer Engineering, Faculty of Engineering, Istanbul University (IU), Istanbul, Turkey
[b] Department of Electrical Engineering, Faculty of Engineering, Istanbul Aydin University (IAU), Istanbul, Turkey

ARTICLE INFO

Article history:
Received 13 October 2010
Accepted 19 March 2013

Keywords:
Breast cancer
Image enhancement
Tumor classification
Spherical Wavelet Transform (SWT)
Homomorphic filtering
Local Seed Region Growing (LSRG)
Support Vector Machines (SVM)

ABSTRACT

The purpose of this study is to implement accurate methods of detection and classification of benign and malignant breast masses in mammograms. Our new proposed method, which can be used as a diagnostic tool, is denoted Local Seed Region Growing–Spherical Wavelet Transform (LSRG–SWT), and consists of four steps. The first step is homomorphic filtering for enhancement, and the second is detection of the region of interests (ROIs) using a Local Seed Region Growing (LSRG) algorithm, which we developed. The third step incoporates Spherical Wavelet Transform (SWT) and feature extraction. Finally the fourth step is classification, which consists of two sequential components: the 1st classification distinguishes the ROIs as either mass or non-mass and the 2nd classification distinguishes the masses as either benign or malignant using a Support Vector Machine (SVM). The mammograms used in this study were acquired from the hospital of Istanbul University (I.U.) in Turkey and the Mammographic Image Analysis Society (MIAS). The results demonstrate that the proposed scheme LSRG–SWT achieves 96% and 93.59% accuracy in mass/non-mass classification (1st component) and benign/malignant classification (2nd component) respectively when using the I.U. database with k-fold cross validation. The system achieves 94% and 91.67% accuracy in mass/non-mass classification and benign/malignant classification respectively when using the I.U. database as a training set and the MIAS database as a test set with external validation.

1. Introduction

Among various cancers, breast cancer places at the top in women, both in the developed and the developing countries. There is a parallel increase in the incidence of this disease with life expectancy and urbanization [1,2]. Previously, the most effective way to be able to survive breast cancer is detecting it in an early phase. The significance of mammography is to reduce deaths from breast cancer by early detection of masses. Although this technology has been developing, it remains difficult in some cases to interpret a dense mammogram, including some suspicious region of interest (ROIs). Whether the radiologist is not experienced enough or the contrast is inadequate, unnecessary biopsy tests are performed against the possibility of breast cancer. As biopsy tests are expensive and invasive, computer aided methods, which help to detect true positive masses (TPs) and eliminate false positives (FPs), have to be developed. Such methods have recently achieved adequate performance in assisting radiologists to make a malignant/benign decision by providing a "second eye" for breast cancer diagnosis.

As wavelets present an efficient decomposition in signals and images, several wavelet-based studies have been developed related to mammographical mass detection and classification in recent years [1–4]. However there are less studies about spherical wavelet and curvelet transforms because these methods are new in the literature. Karahaliou et al. [3] investigate clusters of microcalcifications with their texture properties. Three level multi-resolution decomposition is implemented using Laws' exture energy measures, first order statistics and cooccurrence matrices features to extract ROIs from the surrounding tissue. Their system, which uses a probabilistic neural network, produces 86% accuracy rate in classifying the masses as normal, benign or malignant. In the study of Angelini et al. [4] the system classifies the ROIs as either mass or non-mass. A pixel-based, Discrete Wavelet Transform-based (DWT) and Overcomplete Wavelet Transform-based (OWT) image representations are applied to an SVM system subsequently. The best results are obtained by DWT and OWT representations. Hwang et al. [5] extract mammographic image texture features using a Haar wavelet transform.

* Corresponding author. Tel.: +90 212 473 7070.
E-mail addresses: paras@istanbul.edu.tr (P. Görgel), asertbas@istanbul.edu.tr (A. Sertbas), uosman@aydin.edu.tr (O.N. Ucan).

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.03.008

They use neural networks, statistical discriminant analysis and SVM for classification and their system achieves 88% accuracy.

Curvelets represent the discontinuities through edges or curves in objects or images efficiently. Some studies performing curvelet transforms in image processing are as follows. Ali et al. [6] implement a curvelet transform approach to computed tomography (CT) images. Their system achieves satisfying results for the fusion of magnetic resonance. In the study of Binh and Thanh [7], a curvelet transform-based method is developed for object detection in speckled images. The constructed segmentation method presents a sparse expansion for typical smooth-contoured images.

In recent years Buciu and Gacsadi [8] present a study, in which the mammograms are filtered with Gabor wavelets, and directional features at different orientation and frequencies are extracted. Principal component analysis (PCA) is implemented to reduce the high dimension of filtered and unfiltered data and an SVM is used to classify the data. They achieve 97.56% sensitivity and 78.26% specificity. Tahmasbi et al. [9] present a study aiming to reduce the false negative rate by using Zernike moments as shape descriptors and margin characteristics. Two groups of the moments are extracted from the pre-processed mammograms. The moments that are the most effective ones are chosen and a backpropagation multilayer perceptron is used for classification, which performs at a 92.8% accuracy rate.

This paper presents a computer-aided diagnosis system including mammographic image enhancement, segmentation and diagnosis stages via filtering, mass detection and classification. SWT, which fits the geometric structure of spherical breast masses, helps to optimize a multiresolution transform prior to feature extraction. This study uses two different databases to indicate the superiority of SWT over DWT and the last scale coefficients over all coefficients. The new proposed method in this paper is based on a four-stage algorithm: enhancement with homomorphic filtering; segmentation with Local Seed Region Growing (LSRG); feature extraction with Spherical Wavelet Transform (SWT) and finally classification the ROIs and masses with SVM. The proposed system that we have called LSRG–SWT can be helpful to extract specific characteristics from raw data and provide true interpretation to radiologists.

The remainder of this paper is organized as follows. Section 2 gives a brief introduction to homomorphic filtering, Wavelet Transform, LSRG–SWT and SVM methods. Section 3 discusses the experimental work while Section 4 presents the results and Section 5 includes the conclusion.

2. Methodology

In this study the diagnosis task begins with contrast enhancement as seen in Fig. 1. First, we enhance the images by using homomorphic filtering and in this way local contrast is improved. Next, the suspicious regions such as masses are extracted using the proposed LSRG algorithm proposed by adding some local rules and descriptions to a Seed Region Growing algorithm. The detected ROIs are not always true positive masses, some of them are non-mass breast tissue and relatively brighter than the surrounding tissue. To prevent the increment of false positives and improve true positive detection, a Spherical Wavelet Transform is implemented prior to feature extraction. Each detected ROI is passed from a five-level SWT as the optimum results are achieved with a two-level DWT having six coefficients (approximation$_2$ (a_2), horizantal$_2$ (h_2), vertical$_2$ (v_2), diagonal$_2$ (d_2), approximation$_1$ (a_1) and the mean of (h_1, v_1, d_1)). To generate six coefficients ($w_1, w_2, w_3, w_4, w_5, c_5$) in SWT as well, the decomposition should contain five levels. Moreover, according to some previous studies in the literature [10] a five-level SWT performs better performance.

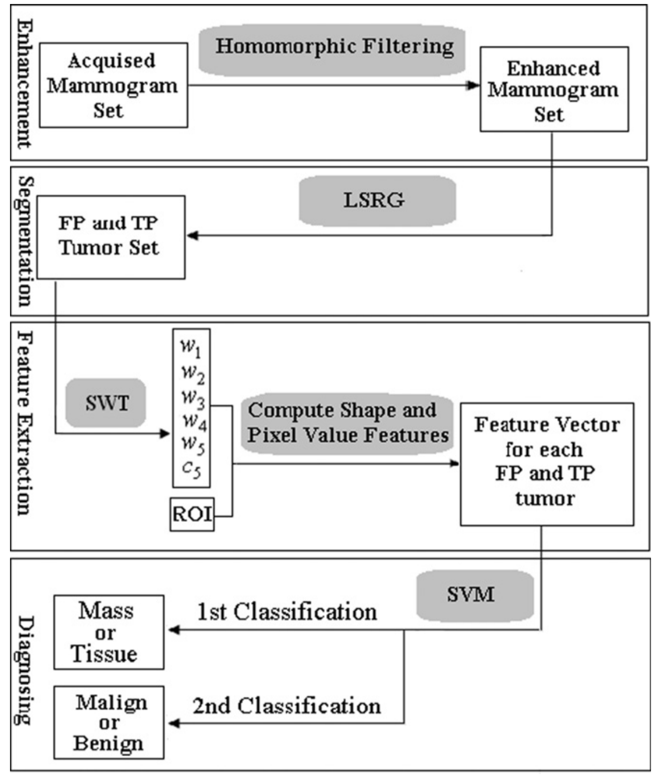

Fig. 1. The flow chart of LSRG–SWT method.

Each ROI is represented both with its own and SWT coefficients' shape and gray level-based feature matrices. 1st classification determines whether the ROI is mass or non-mass and the 2nd classification determines whether the mass is benign or malignant, which provides the breast cancer diagnosis. The software is developed with MATLAB Version 7.6 and the feature matrices are given to the SVM using WEKA 3.7.1.

2.1. Enhancement using homomorphic filtering

For correct segmentation and diagnosis mammogram contrast enhancement is implemented using homomorphic filtering,, which provides a good deal of control over the components of illumination and reflectance. This control requires the specification of a filter function $H(u,v)$ that affects the low and high frequency components of Fourier transform differently. An image $f(x,y)$ can be expressed as the product of illumination $i(x,y)$ and reflectance $r(x,y)$ components [11]:

$$f(x,y) = i(x,y)r(x,y) \tag{1}$$

and we define:

$$z(x,y) = \ln f(x,y) = \ln i(x,y) + \ln r(x,y), \tag{2}$$

$$Z(u,v) = F_i(u,v) + F_r(u,v) \tag{3}$$

where $Z(u,v)$, $F_i(u,v)$ and $F_r(u,v)$ demonstrate the Fourier transforms of $z(x,y)$, ln $i(x,y)$ and ln $r(x,y)$ respectively. If it is processed by means of a $H(u,v)$ filter function, $S(u,v)$ is yielded:

$$S(u,v) = H(u,v)Z(u,v) = H(u,v)F_i(u,v) + H(u,v)Fr(u,v) \tag{4}$$

so $s(x,y)$ is the inverse Fourier transform of $S(u,v)$ and can be expressed in the form:

$$s(x,y) = i'(x,y) + r'(x,y) \tag{5}$$

finally, the desired enhanced image is obtained as seen in Eq. (6).

$$g(x,y) = e^{s(x,y)} = e^{i'(x,y)} \times e^{r'(x,y)} = i_E(x,y)r_E(x,y) \tag{6}$$

in this filtering, the purpose is to separate the illumination and reflectance components in the form shown in Eq. (3). The homomorphic filter function $H(u,v)$ can then operate on these components separately as indicated in Eq. (4). $H(u,v)$ can be shown as:

$$H(u,v) = (\gamma_H - \gamma_L)[1 - e^{-c(D^2(u,v)/D_0^2)}] + \gamma_L \tag{7}$$

where D_0 is a specified distance from the origin of the transform and $D(u,v)$ is the distance from point (u,v) to the center of the frequency rectangle. Constant c controls the sharpness of the filter function slope as it is transmitted between the previously defined values 0.5 and 2 of the parameters γ_L (low) and γ_H (high) respectively. A brief of homomorphic filtering process and an enhanced mammogram are given in Figs. 2 and 3 respectively. Fig. 3(a) represents a dense mammogram, where it is hard to distinguish the marked masses from the surrounding tissue, causing false negative diagnosis. After the homomorphic filtering application, the suspicious regions having high attenuation properties and low local contrast gains more detectability.

2.2. LSRG segmentation method

Segmentation is an essential process in any image analysis study where an image is taken as input and some detailed description of the scene or object is used for output. It basically divides the spatial domain pixels into meaningful non-overlapping, constituent regions that are homogeneous with respect to some characteristics. Basic segmentation technique divides the image I into n non-overlapping regions represented by R_i $(i = 1, 2, 3, \ldots, n)$ satisfying the properties below.

a) $\cup_{i=1}^{n} R_i = I$
b) $R_i \cap R_j = \phi$
c) $H(R_i) = \text{TRUE}$
d) $H(R_i \cup R_j) = \text{FALSE}$ if R_i and R_j are adjacent

$H(R)$ represents the homogeneity criterion based on feature values that are established for the segmentation purpose over the region R. Property (a) ensures that every pixel in the image belongs to one of the non-overlapping sub-regions. The second property (b) guarantees that one pixel belongs to only one region in an image. The third property (c) ensures that the region satisfies the homogeneity criterion defined by the user. Finally the fourth property (d) ensures that the maximality of each region is satisfied.

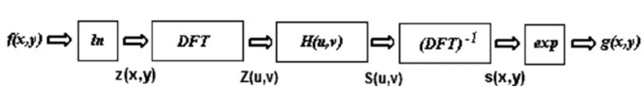

Fig. 2. Homomorphic filtering.

In this study we propose a new growing algorithm called Local Seed Region Growing (LSRG). It depends on the traditional similarity-based Seed Region Growing (SRG) segmentation algorithm that partitions an image directly into regions via some similarity measurements, without any search for boundaries or thresholds. The advantages, which differ LSRG from seed region growing, are the determination of similarity criterion and seed selection that are carried out according to both global and local conditions (neighbourhood of size 3×3). The steps of the improved LSRG algorithm, which divides the image I into n R_i regions for $i = 1,\ldots,n$ are listed below.

(1) Apply Seed Criterion (SeCr) to all pixels in image I and find the seeds belonging to the regions demonstrated as $R_i(s)$ where s are the coordinates of the seed.

$$\text{SeCr} = (I(x,y) - M_I) \geq \max(\text{Th}_I, S_I) \text{ and } (I(x,y) - M_{N(x,y)} \geq \max(\text{Th}_{N(x,y)}, S_{N(x,y)})$$

$I(x,y)$ is the current pixel specified as a seed while M_I, Th_I and S_I represent the mean, threshold and standard deviation of the entire image I respectively. $M_{N(x,y)}$, $\text{Th}_{N(x,y)}$ and $S_{N(x,y)}$ are the mean, threshold and standard deviation of the neighbourhood $(N(x,y))$ of this pixel respectively.

(2) Start the growing process for $i = 1$ to n for all seeds.

a) Constitute the set labeled as 'last' which demonstrates the coordinates of the pixels joined last to the related region and compute the mean (M_1) of the seed and these pixels.

$$M_1 = \frac{R_i(s) + R(\text{last})}{\text{last} + 1}$$

b) Determine all neighbours $(N(\text{last}))$ of the pixels in the set 'last' and compute the mean (M_2) of each jth pixel in $N_j(\text{last})$ and $M(\text{last})$ (the mean of the set 'last') respectively. Find the appropriate neighbours to join to ith region using the Similarity Criterion (SiCr).

$$M_2 = \frac{M_{last} + N_j(\text{last})}{2}$$

$$\text{SiCr} = |M_2 - N_j(\text{last})| \leq S_{N(j)}$$

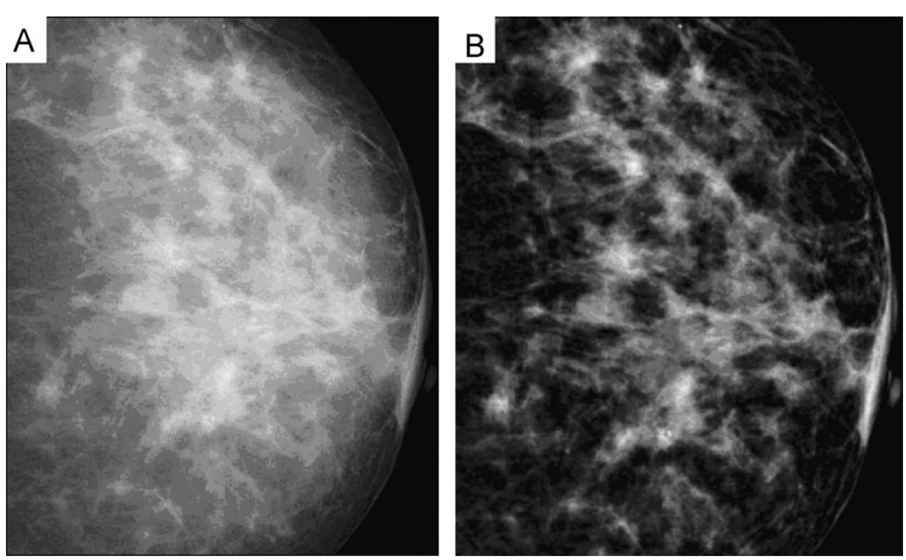

Fig. 3. (A) Dense mammogram (raw data) and (B) enhanced mammogram by using homomorphic filtering.

where $S_N(j)$ is the standard deviation of the neighbours of jth pixel.

c) If SiCr is TRUE, grow the ith region by joining the jth pixel to it.

$$R_i' = R_i \cup N_j(\text{last})$$

d) Go to step (2) until all seeds are grown.

(3) In LSRG algorithm the tested thresholds are listed in Table 1.

Table 1
The thresholds and definitions used in LSRG algorithm.

Thresholds	Definition
Thr1	Gray level values of 100, 128 and 200 that represent different gray color tones. The values close to 255, correspond to brighter gray color tones
Thr2	(max. gray level of the mammogram+mean of the mammogram)/2
Thr3	The mean of the pixels that are larger than the mammogram mean
Thr4	The mean of the pixels that are larger than 100
Thr5	The mean of the pixels that are larger than 128
Thr6	(max. gray level of the mammogram+Thr3)/2
Thr7	(max. gray level of the mammogram+Thr4)/2
Thr8	(max. gray level of the mammogram+Thr5)/2
Thr9	(mean of the mammogram+standart deviation of the mammogram)/2

In this study, Thr6 is preferred as it gives the maximum performance in LSRG algorithm. According to the obtained results, 191 non-masses and 78 masses (all of the true positives) are detected with Thr6. The result which can also considered to be a satisfying detection is obtained with Thr7 (182 non-masses and 71 masses). If the threshold value gets larger, the number of non-masses (FPs) decreases but also the number of detected masses (TPs) might decrease. Fig. 4a and c represents two enhanced mammograms while Fig. 4b and d represents the suspicious regions detected by LSRG algorithm that could be masses.

2.3. Wavelet transform

The signal is decomposed into various scales at different levels of resolution after wavelet transform and by dilating the mother wavelet multiresolution analysis is provided. A one dimensional signal $f(x) \in L^2(R)$ at 2^j resolution is orthogonal to the signal belonging to V_{2^j} subspace [12,13]. $W_{2^j}^A f(x), W_{2^j}^{D^h} f(x), W_{2^j}^{D^v} f(x)$ and $W_{2^j}^{D^d} f(x)$ represent $f(x)$ signal's approximation, horizontal detail, vertical detail, and diagonal detail respectively. The approximation $W_{2^{j+1}}^A f(x)$ at resolution 2^{j+1} carries more information than $W_{2^j}^A f(x)$ at resolution 2^j. $\phi(x)$ And $\psi(x)$ demonstrate the scaling and wavelet functions respectively which satisfy $\varphi_{2^j} = 2^j \varphi(2^j x)$ and $\psi_{2^j} = 2^j \psi(2^j x)$. O_{2^j} has an orthogonal base of $\{2^{-j/2} \psi_{2^j}(x - 2^{-j}k)_{k \in Z}$ and V_{2^j} has an orthogonal base of $\{2^{-j/2} \psi_{2^j}(x - 2^{-j}k)_{k \in Z}$.

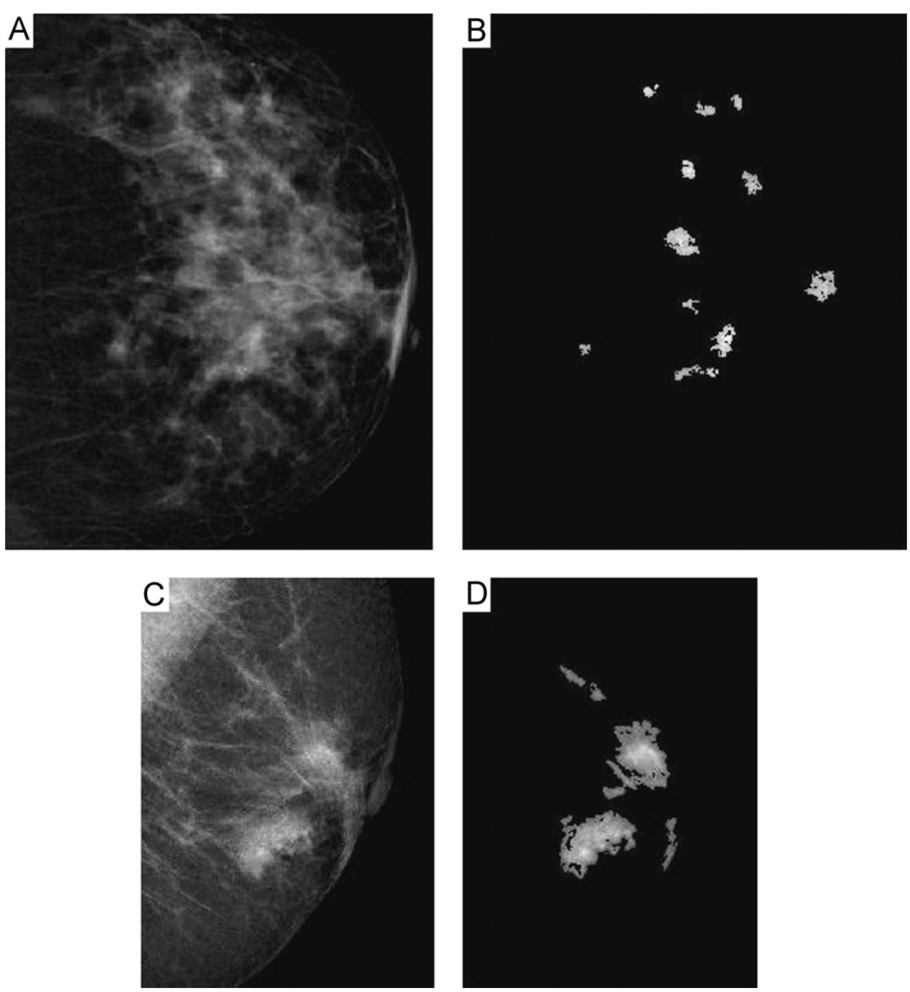

Fig. 4. (A and C) Enhanced mammograms. (B and D) ROIs detected by LSRG method.

The original signal $f(x)$ at resolution 2^j has approximation and detail components that are characterized as follows:

$$\{W^A_{2^j}f(k)\}_{k\in Z} = \{\langle f(o), \varphi_{2^j}(0-2^{-j}k)\rangle\}_{k\in Z} \quad (8)$$

$$\{W^D_{2^j}f(k)\}_{k\in Z} = \{\langle f(o), \psi_{2^j}(0-2^{-j}k)\rangle\}_{k\in Z} \quad (9)$$

where h is a low-pass and g is a high-pass filter satisfying $h(k) = \langle \phi_{-1}(x), \phi(x-k)\rangle$ and $g(k) = \langle \psi_{-1}(x), \psi(x-k)\rangle$. $f(x)$ At resolution 2^j can also be demonstrated by the mirror filters $\hat{h}(k) = h(-k)$ and $\hat{g}(k) = g(-k)$ for $j = 0, -1, -2, \ldots$.

$$W^A_{2^{j-1}}f(x) = \sum_{k=-\infty}^{\infty} \hat{h}(2x-k)W^A_{2^j}f(k) \quad (10)$$

$$W^D_{2^{j-1}}f(x) = \sum_{k=-\infty}^{\infty} \hat{g}(2x-k)W^A_{2^j}f(k) \quad (11)$$

2.4. Spherical Wavelet Transform

Wavelets are no longer optimal in the analysis of data containing anisotropic features. This has started the development of different multiscale decompositions such as the ridgelet, spherical and curvelet transforms [14–16]. Firstly Starck et al. [16] have demonstrated that spherical transforms can be useful for detection and discrimination of non Gaussianity in astronomical images. The full-sky data is mapped to a sphere to implement a curvelet transform on the sphere. The goal of this paper is to implement a medical image processing study based on Spherical Wavelet Transform using the advantage of SWT complying well with spherical shapes. Multi-resolution analysis (MRA) of $L_2(S^2)$, where S^2 is the unit disc, is $S^2 = \{(x,y)\in R^2 : x^2 + y^2 \le 1\}$. Accordingly the associated Legendre functions are;

$$P_n^{(m)}(t) = (1-t^2)^{m/2}\frac{1}{2^n n!}\frac{d^{n+m}}{dt^{n+m}}(t^2-1)^n \quad \text{for } n\ge m \quad (12)$$

$$\langle P_n^{(m)}, P_p^{(m)}\rangle = \frac{2(n+m)!}{2n+1(n-m)!}\delta np \quad \text{for } n\ge m,\ \ p\ge m \quad (13)$$

the local surface coordinates are introduced in S^2 as follows:

$$x = \begin{bmatrix} \sin\theta & \cos\phi \\ \sin\theta & \sin\phi \\ \cos\theta & \end{bmatrix} \in S^2 \quad \text{where } \theta\in[0,\pi], \phi\in[-\pi,\pi] \quad (14)$$

in these local coordinates the scalar product is expressed as,

$$\langle f, g\rangle = \int_0^{\pi}\int_{-\pi}^{\pi} f(\theta,\phi)g(\theta,\phi)\sin\theta d\phi d\theta \quad (15)$$

spherical harmonics represent the angular portion of a set of Laplace's equation solutions. Laplace's spherical harmonics set, which has the equation below, forms an orthogonal system in spherical coordinates [13].

$$Y_l^m(\theta,\phi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}}P_l^m(\cos\theta)e^{im\phi} \quad (16)$$

in $Y_l^m(\theta,\varphi)$ spherical harmonics $\theta$and φ represent spherical polar angles while l and m indicate the level and the order respectively. P_l^mdenotes the Legendre polynomials and equals $1, x, (1/2)(3x^2-1), (x/2)(5x^2-3), (1/8)(35x^4-30x^2+3)$ for $l = 0,1,2,3,4$ respectively. As the reconstruction of an image from its wavelet coefficients $I = \{w_1, \ldots, w_j, c_j\}$ is straightforward and Eq. (17) can be written as follows:

$$c_0(\theta,\phi) = c_j(\theta,\phi) + \sum_{j=1}^{j} w_j(\theta,\phi) \quad (17)$$

also we can write the equation of $c_0(\theta,\phi) = \varphi_{lc}(\theta,\phi)\times f(\theta,\phi)$. In this study we use Shannon scaling function (φ_{lc}), which is demonstrated by Eq. (19) [17]. Eq. (18) is the basic form of scaling functions with l_c cut-off frequency and $\hat{\varphi}_{lc}(l,0)$ spherical harmonic coefficients.

$$\varphi_{lc}(\theta,\phi) = \varphi_{lc}(\theta) = \sum_{l=0}^{l=lc}\hat{\varphi}_{lc}(l,0)Y_{l,0}(\theta,\phi) \quad (18)$$

$$\varphi_j(x,y) = \sum_{n=0}^{\min[2^j, M-1]}(|x||y|)^{-n-1}\frac{2n+1}{4\pi}P_n(|x\|y|) \quad (19)$$

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. Wavelet coefficients are the difference between two consecutive resolutions, $w_{j+1}(\theta,\varphi) = c_j(\theta,\varphi) - c_{j+1}(\theta,\varphi)$, which corresponds the following specific choice for ψ_{lc}:

$$\hat{\psi}_{\frac{lc}{2^j}}(l,m) = \hat{\varphi}_{\frac{lc}{2^{j-1}}}(l,m) - \hat{\varphi}_{\frac{lc}{2^j}}(l,m) \quad (20)$$

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

$$\hat{h}_j(l,m) = \sqrt{\frac{4\pi}{2l+1}}h_j(l,m) = \begin{cases} \dfrac{\hat{\varphi}_{\frac{lc}{2^{j+1}}}(l,m)}{\hat{\varphi}_{\frac{lc}{2^j}}(l,m)} & \text{if } l < \frac{lc}{2^{j+1}} \\ 0 & \text{otherwise} \end{cases} \quad (21)$$

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

$$\hat{g}_j(l,m) = \sqrt{\frac{4\pi}{2l+1}}g_j(l,m) = \begin{cases} 1 & \text{if } l \ge \frac{lc}{2^{j+1}} \\ \dfrac{\hat{\psi}_{\frac{lc}{2^{j+1}}}(l,m)}{\hat{\varphi}_{\frac{lc}{2^j}}(l,m)} & \text{if } l < \frac{lc}{2^{j+1}} \end{cases} \quad (22)$$

as seen in the flow chart of SWT algorithm (Fig. 5), the aim is to obtain all coefficients $(w_1, w_2, w_3, w_4, w_5, c_5)$ of the transform including the wavelet and scaling coefficients.

2.5. SVM

Support Vector Machine (SVM), introduced by Vapnik in 1995 [18], is a method to estimate the data classification function [19]. The basic idea of an SVM is to construct a hyperplane as the decision surface in such a way that the margin of separation between positive and negative examples is maximized [20]. A classification task usually involves separating data into training and test sets. Each instance in the training set contains one target value and several attributes. The goal of the SVM is to produce a model (based on the training data) that can predict the target values of test data even the attributes are given only. An SVM uses a kernel function, in which the nonlinear mapping is implicitly embedded. In Cover's theorem, a function can be considered as a kernel provided that it satisfies Mercer's conditions [21]. The following relation should be maximized to optimize the SVM classifier boundary in a given training set of instance-label pairs $(x_i, y_i),\ i = 1, \ldots, l$ where $x_i \in R^n$ and $y \in \{1, -1\}^l$:

$$L(c) = \sum_{i=1}^{l} c_i - \frac{1}{2}\sum_{i,j=1}^{l} y_i y_j c_i c_j K(x_i, x_j), \quad 0 \le c_i \le P \quad (23)$$

while

$$\sum_{i=1}^{l} y_i c_i = 0, \quad w = \sum_{i=1}^{N} c_i y_i x_i, \quad c_i[y_i(w^T x_i + b) - 1 + \xi_i] = 0 \quad (24)$$

where P is a user-specified positive parameter to control the tradeoff between SVM complexity and the number of non-separable points. l Shows the number of samples and $K(x_i, x_j)$ is the SVM kernel. Here a solution to $c=(c_1, c_2, \ldots, c_l)$ is obtained, where c_i is a Lagrange coefficient. The slack variables ξ_i are used to relax the constraints of the canonical hyperplane equation. In a typical SVM the kernel function plays an important role in mapping the input vector implicitly into a high-dimensional feature space, in which better separability can be achieved.

3. Experimental work

3.1. Image dataset

In the present study two different databases are used in order to validate our methodology. 60 Images have been acquired from 30 patients from the Radiology Department of the Faculty of Medicine Hospital of Istanbul University, Turkey. There are 78 masses in these 60 images, among which 35 are malignant and 43 are benign. No masses are found in 6 mammograms. Mammograms have also been taken from the free MIAS, which comprises the second database, and contains 25 malignant and 35 benign masses. As known, the final diagnosis of a breast mass is made by biopsy tests in medical centers. Therefore the masses (abnormal tissue) have been marked and classified as benign or malignant (Fig. 6) by expert radiologists from the Radiology Department according to the biopsy results. In the hospital GIOTTO IMAGE SDL/W, which is a modern mammography system for diagnostic and screening examinations, is used. It utilizes the latest technologies: the 2nd generation amorphous selenium (A-Se) digital detector of 24×30 cm and a special tungsten anode x-ray tube for patient dose reduction with direct energy conversion (direct conversion of the x-photons into electric charges). The mammogram set has been selected from various patients at different ages to make the images invariant to contrast.

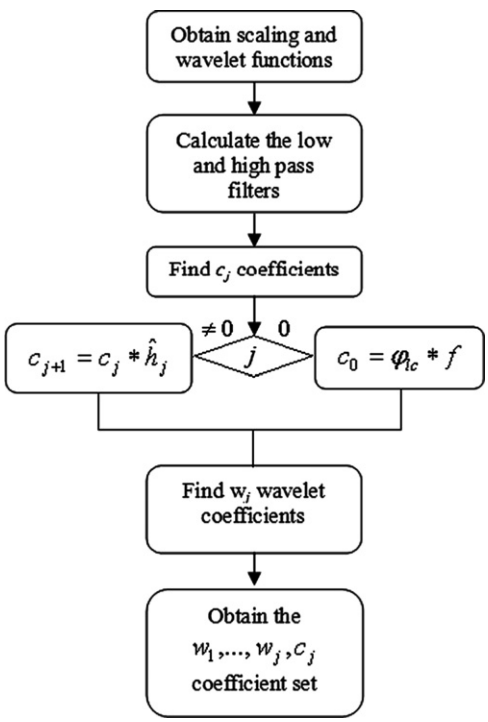

Fig. 5. Flow chart of the SWT method.

3.2. Feature extraction

With appropriate feature extraction, relevant information of input data can be used to perform the desired task instead of using full size input [21]. In this study we extract some features related to mass size, geometrical shape and boundary contour from SWT coefficients and raw ROIs. The preferred features related with size are as follows: Area is the actual scalar number of pixels in the region [22]; Centroid is the center of the region; BoundingBox is the smallest rectangle containing the region; Filled Area is the number of pixels in filled region and Equiv Diameter ($\sqrt{4 \times \text{Area}/\pi}$) is the diameter of a circle with the same area as the region. The features related with geometrical shape are as follows. Euler Number is the number of objects in the region minus the number of holes in those objects and Extrema is the extremal points in the region. The rows of the matrix contain the *x*- and *y*-coordinates of the points. Convex Hull is the smallest convex polygon that can contain the region. Solidity is the proportion of the pixels in the region that are also in the convex hull. The features related with boundary are as follows. Major Axis Length is the length (in pixels) of the major axis of the ellipse that has the same second-moment as the region while Minor Axis Length is the length (in pixels) of the minor axis of the ellipse that has the same second-moment as the region. Eccentricity is the eccentricity of the ellipse that has the same second-moment as the region and it is the ratio of the distance between the foci of the ellipse and its major axis length. Orientation means the angle between the x-axis and the major axis of the ellipse

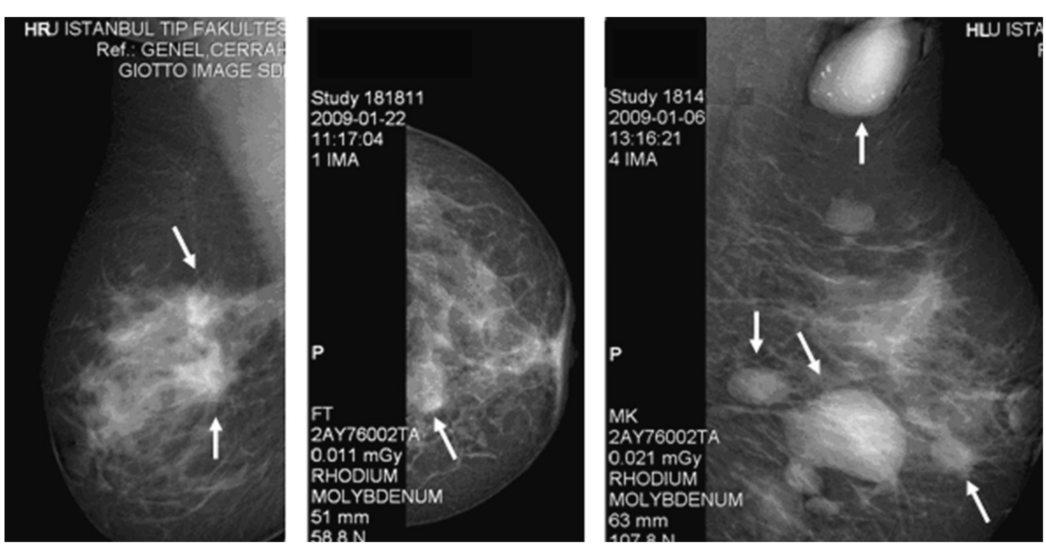

Fig. 6. Mammograms from the I.U. database with annotated masses: the first and second mammograms have malignant masses and the third one has benign masses.

that has the same second-moment as the region. Extent represents the proportion of the pixels in the bounding box that are also in the region. We also define and extract two further features: boundary based Mean Center-Border Distance representing the similarity between the ROI and a typical circle; and shape based Symmetry. All features mentioned above are calculated to provide feature matrices for each ROI. Those matrices are used as input vectors to the supervised learning system SVM.

3.3. Classification of the detected ROIs

A classification system results as false positive (FP) if the system labels a negative point as positive, false negative (FN) if the system labels a positive point as negative, true positive (TP) and true negative (TN) if the system correctly predicts the label respectively [23]. In this study, diagnosis of the breast ROIs consists of two classifications. The 1st classification helps to determine whether the ROI is a mass (TP) or non-mass (FP). This classification aims to reduce the non-mass number which can cause incorrect diagnosis. The 2nd classification, which is more significant, distinguishes the masses as benign (FP) or malignant (TP). Masses existing in breast tissue might have different shape, margin, orientation, lesion boundary, echogenic pattern and vascularity. Malignant masses leading to breast cancer disperse into the normal breast tissue, have irregular boundary and sharp corners like stars, while benign masses, which do not prevent the survival, have smooth, distinct and regular margin (Fig. 7). Radiologists sometimes make false negative diagnosis (which may cause mortality) by missing the masses due to noise and contrast inadequacy or they make false positive diagnosis (which may cause redundant biopsies) by assuming the non-masses were masses due to density and shape similarity.

In this study the 1st and 2nd classification steps are carried out using an SVM classifier. The feature extraction process is applied to both the raw ROIs and their SWT coefficients $(w_1, w_2, w_3, w_4, w_5, c_5)$ to obtain the comprehensive feature matrices. K-fold cross validation, in which whole data are randomly divided into k mutually exclusive and approximately equal sized subsets, is used for the I.U. database to separate the test and training data. The classification algorithm is trained and tested k times [24,25]. Different k values listed in Table 5 are used in the trials conducted to reach optimum accuracy. Furthermore to produce more objective results via external validation, the MIAS database is used as the test set and I.U. database is used as the training set.

3.4. Performance metrics

Receiver operating characteristic (ROC) analysis, which is applied extensively to diagnostic systems in clinical medicine, is based on statistical decision theory and developed in the context of electronic signal detection. ROC curve is a plot of the classifier's true positive diagnosis rate versus its false positive diagnosis rate [23]. In this study we use ROC curves to compare the performance of the coefficient sets and also calculate some well-known image processing performance metrics, which are as follows.

The sensitivity is defined as the ratio between the number of true positive predictions and the number of regions in the test set. It is defined as follows:

$$\text{Sensitivity} = \frac{\text{TP}}{(\text{TP}+\text{FN})} \times 100\% \qquad (25)$$

the specificity is defined as the ratio between the number of false positive predictions and the number of regions in the test set. It is

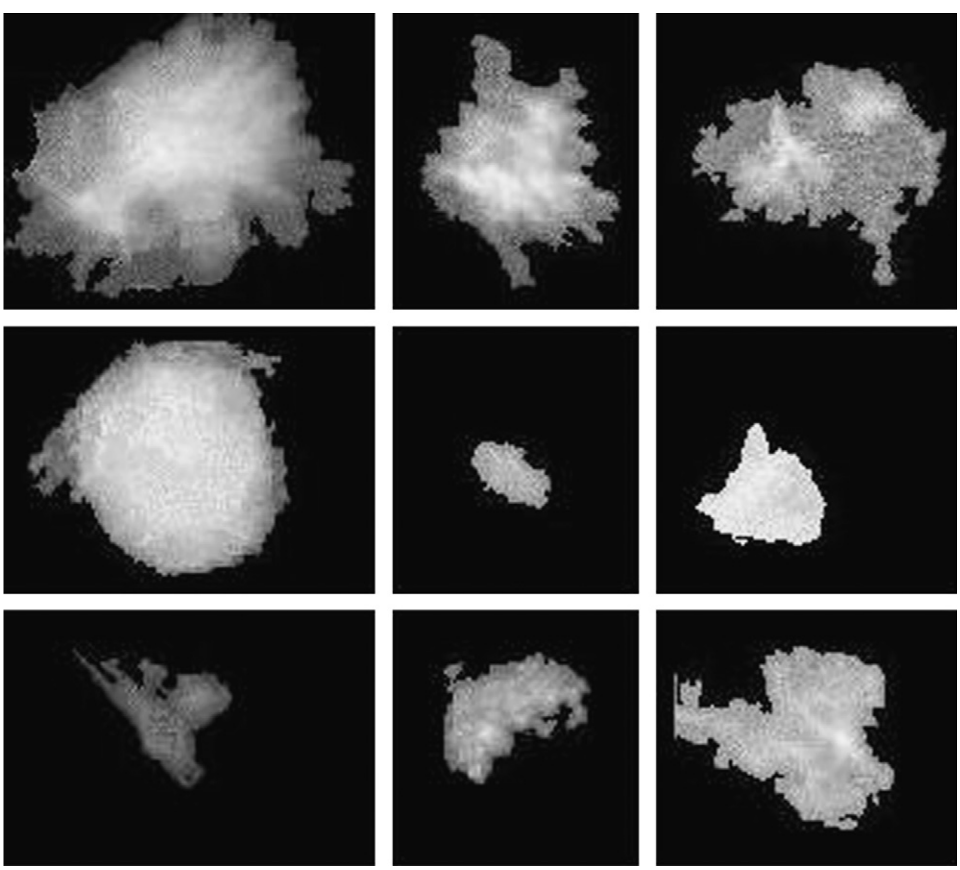

Fig. 7. Some ROIs detected by LSRG algorithm, the top line: malignant masses, the middle line: benign masses and the bottom line: non-masses.

defined as follows:

$$\text{Specificity} = \frac{\text{TN}}{(\text{TN}+\text{FP})} \times 100\% \tag{26}$$

the overall accuracy is the ratio between the total number of correctly classified regions and the test set size (total number of regions). It is defined as follows:

$$\text{Accuracy} = \left(\frac{N_R}{N}\right) \times 100\% \tag{27}$$

where N_R is the number of correctly classified regions during the test run and N is the total number of test set. False positive fraction (FPF) gives the numbers of FPs per case (mammogram) while true positive fraction (TPF) gives the true positive detection rate according to Eq. (28). Sensitivity, specificity and accuracy define the performance of the 1st and 2nd classifications. On the other hand FPF and TPF define the performance of the proposed LSRG algorithm.

$$\text{FPF} = \frac{\text{FP}}{\text{Total case number}}$$
$$\text{TPF} = \frac{\text{TP}}{\text{TP}+\text{FN}} \tag{28}$$

4. Results

To evaluate the entire LSRG–SWT system performance, the detection rate (TPF) of LSRG mass detection algorithm is firstly measured as 1 (78/(78+0)) according to Eq. (28). That is, LSRG is able to detect all benign and malignant masses in the mammograms. However the total number of detected masses is obtained as 269 containing 191 non-masses (FPs) and 78 masses (TPs) for the I.U. database. Since there are 60 cases (mammograms) in image data set, the FPF is 3.2 according to Eq. (28). The 1st classification, which distinguishes the detected 269 ROIs as either mass or non-mass, is implemented to reduce the FPF value as it causes false positive diagnosis. This classification achieves 96% accuracy and the number of non-masses (FPs) is reduced to 3 (Table 3) and FPF is decreased to 0.1. The confusion matrix and some well-known performance metrics related to the 1st classification are listed in Tables 2 and 3. On the other hand LSRG algorithm produces 94% accuracy for the MIAS database in mass/non-mass classification. Kappa statistics in Table 3 is typically an assessment, for which two or more raters examining the same data specify the degree of agreement in assigning data to categories. For medical statistics, the raters are radiologists that analyze an x-ray and computers that analyze the same x-ray for diagnosis [26,27].

The purpose of 2nd classification is construction of a system that could help radiologists for an accurate diagnosis by distinguishing the masses as either malignant (TP) or benign (FP). For the 2nd classification we began with the I.U. database and have made several trials to measure the change of the performance depending on using only ROI's own features and using its various SWT coefficients' features. The performance is unfortunately 75% when using ROI's own features without SWT. The accuracy increases to 91.03% with additional features of six SWT coefficients. The feature matrices, which include all coefficients, are of size 17×7 due to 17 features for each of the 6 coefficient sets and one for the ROI's own matrix. In the trials, the 4th and 5th level coefficients' (last scale coefficients) features, which are more meaningful, produce higher classification accuracy of 93.59%. Table 4 represents the confusion matrix of optimum performance.

When only wavelet coefficient $(w_1, w_2, w_3, w_4, w_5)$ features are added to ROI's own feature matrix, the accuracy is 84.62%. The performance is 87.18% when approximation coefficient (c_5) features are added. On the other hand the results obtained by using Discrete Wavelet Transform (DWT), which achieves its highest accuracy of 83.21%, are listed in Table 5 with contributed coefficient set features. To calculate the performance of this entire breast cancer diagnosis system with the I.U. database, 2nd classification results are multiplied with the 1st classification result (96%) one by one to represent the more realistic classification results (Table 5).

As seen in Table 5, optimum LSRG–SWT entire system performance represents 97% sensitivity, 91% specificity and 90% classification accuracy according to Eqs. (25)–(27) with the optimal parameters. Fig. 8 points out the performance analysis of the coefficient sets, which give the highest two performances with SWT and DWT methods and ROI's own feature matrix, with ROC curves.

To implement external validation, the MIAS database is used as the test set and the I.U. database is used as the training set. In the trials, the last scale coefficient features which are more meaningful, produce higher classification accuracy of 91.67% with SWT. On the other hand Discrete Wavelet Transform (DWT), which achieves its best accuracy of 80%, is also applied to the same training and test set to present the comparison between SWT and DWT methods. Further results are listed in Table 6 with contributed coefficient set features. To calculate the performance of the entire breast cancer diagnosis system, 2nd classification results are multiplied with the 1st classification result (94%) one by one to represent the more realistic results using the MIAS database (Table 6).

5. Conclusion

Breast cancer is among the most prevalent cancer types in the world if it can be diagnosed early [25]. Classification systems used

Table 2
Confusion matrix obtained with 1st classification using the I.U. database.

Mass	Non-mass
69 (TP)	9 (FN)
3 (FP)	188 (TN)

Table 4
Confusion matrix of 2nd classification using the I.U. database.

Malignant	Benign
34 (TP)	1 (FN)
4 (FP)	39 (TN)

Table 3
Performance metrics of 1st classification using the I.U. database.

TP rate	FP rate	Precision	Recall	*F*-measure	ROC area	Kappa statistics	Mean absolute error	Root mean squared error	Relative absolute error (%)
0.955	0.086	0.955	0.955	0.955	0.971	0.889	0.062	0.205	14.93

Table 5
The comparative test results of the proposed LSRG–SWT method for the I.U. database: Lsc, Ac, Appc, Wvc, Roi and Appc&MWvc represent Last Scale Coefficients, All Coefficients, Approximation Coefficients, Wavelet Coefficients, ROI's own matrix and Approximation Coefficients with the mean of the Wavelet Coefficients respectively. These sets are fed into either SWT or DWT blocks.

Method	Coefficient set	*k* Value of cross validation	Sensitivity (%)	Specificity (%)	Accuracy of 2nd classification (%)	Accuracy of the entire system (%)
SWT	Lsc	8	97	91	93.59	90
SWT	Ac	9	92	90	91.03	87
SWT	Appc	4	89	86	87.18	84
SWT	Wvc	8	86	83	84.62	81
–	Roi	9	68	81	74.96	72
DWT	Appc&MWvc	7	78	88	83.21	80
DWT	Ac	5	77	81	79.59	76
DWT	Appc	10	83	79	80.77	78
DWT	Wvc	6	77	79	78.21	75

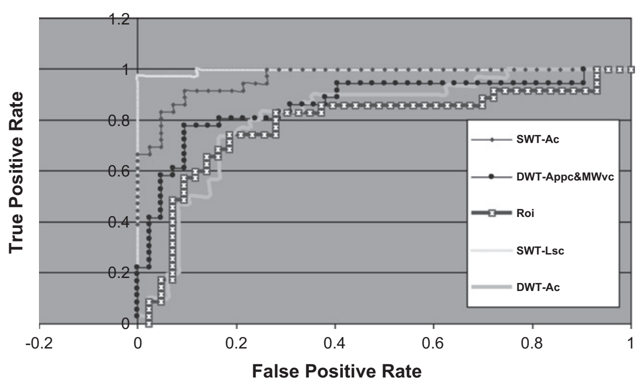

Fig. 8. Performance analysis for the I.U database.

Table 6
The comparative test results of the proposed LSRG–SWT method using the I.U. database for the training set and the MIAS database for the test set: Lsc, Ac, Appc, Wvc, Roi and Appc&MWvc represent Last Scale Coefficients, All Coefficients, Approximation Coefficients, Wavelet Coefficients, ROI's own matrix and Approximation Coefficients with the mean of the Wavelet Coefficients respectively. These sets are fed into either SWT or DWT blocks..

Method	Coefficient set	Sensitivity (%)	Specificity (%)	Accuracy of 2nd classification (%)	Accuracy of the entire system (%)
SWT	Lsc	96	89	91.67	86.17
SWT	Ac	96	86	90.00	84.60
SWT	Appc	92	83	86.67	81.47
SWT	Wvc	88	77	81.67	76.77
–	Roi	76	74	75.00	70.50
DWT	Appc&MWvc	84	77	80.00	75.20
DWT	Ac	80	74	76.67	72.07
DWT	Appc	80	77	78.33	73.63
DWT	Wvc	76	74	75.00	70.50

in medical decision provide medical data to be examined in shorter time and more detailed. The research presented in this article aims to decrease the mortality rate related to breast cancer by reducing the number of malignant masses that radiologists would not notice using the current imaging technologies. It is also desirable to decrease the number of requested biopsy tests due to false positive detection. In this work, we develop a hybrid scheme consisting of homomorphic filtering, LSRG and SWT methods and denote it as LSRG–SWT system that segments the ROIs, detects the masses and classifies them. The satisfying LSRG detection results are 96% and 94% in the I.U. and the MIAS databases respectively. Spherical Wavelet Transform is applied to the ROIs, along with shape, boundary and gray level-based feature extraction. This multi-resolution decomposition study is efficient for solving the real-world problems related to spherical shapes like breast masses as spherical harmonics and equations associated with sphere are used. As SWT fits the geometric structure of the spherical breast masses, it provides optimum multiresolution and produces malignant/benign classification accuracy of 93.59% with the I.U. database using k-fold cross validation. On the other hand the accuracy reduces to 91.67% with external validation when the MIAS database is used for testing and I.U. database is used for training. This study also indicates the superiority of the last scale coefficients (4th and 5th level coefficients – w_4,w_5,w_5) over all coefficients. Furthermore DWT is applied to the masses to present the superiority of SWT method over DWT. Consequently the satisfying performance demonstrates that this study is valuable to improve early diagnosis and reduce the number of unnecessary biopsies.

Conflict of interest statement

None declared.

Acknowledgment

We thank Prof. Dr. Siddiqi Abul Hasan for his scientific information on Spherical Wavelet Transform approach.

References

[1] N.R. Pal, B. Bhowmick, S.K. Patel, S. Pal, J Das, A multistage neural network aided system for detection of microcalcifications in digitized mammograms, Neurocomputing 71 (2008) 2625–2634.
[2] P. Görgel, A. Sertbas, O.N. Ucan, A wavelet based mammographic image denoising and enhancement with homomorphic filtering, J. Med. Syst. 34 (6) (2010) 993–1002.
[3] N.A. Karahaliou, I.S. Boniatis, et al., Breast cancer diagnosis: analyzing texture of tissue surrounding microcalcifications, IEEE Trans. Inf. Technol. Biomed. 2 (2008) 731–738.
[4] E. Angelini, R. Campanini, et al., Testing the performance of image representations for mass classification in digital mammograms, Int. J. Mod. Phys. C 17 (2006) 113–131.
[5] H. Hwang, H.Choi, et al.,Classification of breast tissue images based on wavelet transform using discriminant analysis, in: Proceedings of 7th International Workshop on Enterprise networking and Computing in Healthcare Industry, Gimhae, South Korea, 2005, pp. 345-349.
[6] F.E. Ali, I.M. Eldokany, A.A. Saad, F.E. Abdelsamie, Curvelet fusion of MR and CT images, Prog. Electromagn. Res. 3 (2008) 215–224.
[7] N.T. Binh, N.C. Thanh, Object detection of speckle image base on curvelet transform, ARPN J. Eng. Appl. Sci. 2 (2007) 14–16.
[8] I. Buciu, A. Gacsadi, Directional features for automatic tumor classificiation of mammogram images, Biomed. Signal Process. Control 6 (2011) 370–378.
[9] A. Tahmasbi, F. Saki, S.B. Shokouhi, Classificaiton of benign and malignant masses based on Zernike moments, Comput. Biol. Med. 41 (2011) 726–735.
[10] P. Yu, X. Han, F. Ségonne, Cortical surface shape analysis based on spherical wavelet transformation, in: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop, 2006.
[11] R. Gonzales, R. Woods, Digital Image Processing, Prentice Hall, USA, 2002, pp. 191–193, 793 pp (Chapter 4).

[12] M.M. Eltoukhy, I. Faye, B.B. Samir, A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram, Comput. Biol. Med. 40 (2010) 384–391.
[13] P. Görgel, A. Sertbas, O.N. Ucan, A comparative study of breast mass classification based on spherical wavelet transform using ANN and KNN classifiers, Int. J. Electron. Mech. Mechatronics 2 (2011) 79–85.
[14] P. Abrial, Y. Moudden, et al., Morphological component analysis and inpainting on the sphere: application in physics and astrophysics, J. Fourier Anal. Appl. (JFAA), Special İssue on "Analysis on the Sphere" 13 (2007) 729–748.
[15] D. Donoho, M. Duncan, 2000, in: H. Szu, M. Vetterli, W. Campbell, J. Buss (Eds.), Proceedings of the Aerosense 2000, Wavelet Applications VII, vol. 4056, SPIE, p. 12.
[16] J. Starck, E. Candès, D.L. Donoho, Astronomical image representation by the curvelet transform, Astron. Astrophys. 398 (2003) 785–800.
[17] W.R. Wade, A Walsh, System for polar coordinates, Comput. Math. Appl. 30 (1995) 221–227.
[18] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.
[19] C.J.C. Burges, A Tutorial on Support Vector Machines for Pattern Recognition, Kluwer Academic Publishers, Dordrecht, 1998.
[20] G.B. Junior, A.C. Paiva, et al., Classification of breast tissues using Moran's index and Geary's coefficient as texture signatures and SVM, Comput. Biol. Med. 39 (2009) 1063–1072.
[21] J. Yan, B. Zhang, N. Liu, et al., Effective and efficient dimensionality reduction for large-scale and streaming data preprocessing, IEEE Trans. Knowl. Data Eng. 18 (2006) 320–333.
[22] MATLAB R2008a, Product Help, 2008.
[23] M. Karnana, K. Thangavel, Automatic detection of the breast border and nipple position on digital mammograms using genetic algorithm for asymmetry approach to detection of microcalcifications, Comput. Methods Programs Biomed. 87 (2007) 12–20.
[24] D. Delen, G. Walker, A. Kadam, Predicting breast cancer survivability: a comparison of three data mining methods, Artif. Intell. Med. 3 (2005) 113–127.
[25] S. Sahan, K. Polat, H. Kodaz, et al., A new hybrid method based on fuzzy-artificial immune system and *k*-nn algorithm for breast cancer diagnosis, Comput. Biol. Med. 37 (2007) 415–423.
[26] J. Cohen, A coefficient of agreement for nominal scales, Educ. Psychol. Meas. 20 (1960) 37–46.
[27] R.J. Cook, Kappa and its dependence on marginal ratesin: P. Armitage, T. Colton (Eds.), The Encyclopedia of Biostatistics, Wiley, New York, 1998, pp. 2166–2168.