An Optimized Feature Selection Method For Breast Cancer Diagnosis in Digital Mammogram using Multiresolution Representation

Mohamed Meselhy Eltoukhy[1,2,*] *and Ibrahima Faye*[1]

[1] Centre for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS, Malaysia
[2] Computer Science Department, Faculty of Computers and Informatics, Suez Canal University, Ismailia 41522, Egypt

Abstract: This paper introduces a method for feature extraction from multiresolution representations (wavelet,curvelet) for classification of digital mammograms. The proposed method selects the features according to its capability to distinguish between different classes. The method starts with both performing wavelet and curvelet transform over mammogram images. The resulting coefficients of each image are used to construct a matrix. Each row in the matrix corresponds to an image.The most significant features, in terms of capabilities of differentiating classes,are selected. The method uses threshold values to select the columns that will maximize the difference between the different classes'representatives. The proposed method is applied to the mammographic image analysis society (MIAS) dataset. The results calculated using 2x5-folds cross validation show that the proposed method is able to find an appropriate feature set that lead to significant improvement in classification accuracy.The obtained results were satisfactory and the performances of both wavelet and curvelet are presented and compared.

Keywords: Breast cancer; Wavelet transform; Curvelet transform; Feature selection; Digital mammogram.

1 Introduction

Breast cancer is the most common cancer in women worldwide. It is also the principle cause of death from cancer among women globally. The World Health Organization (WHO) has suggested that two components of early detection have been shown to improve cancer mortality. The first component is Education: to help people recognize early signs of cancer and seek prompt medical attention for symptoms. The second is the screening programs: to identify early cancer or pre-cancer before signs are recognizable, including mammography for breast cancer [1]. Digital mammography is one of the most suitable methods for early detection of breast cancer. However, the visual clues are subtle and vary in appearance, making diagnosis a challenging exercise even for specialist [2]. The false positive detection causes unnecessary biopsy. It has been estimated that only 20-30% of breast biopsy cases are proved to be cancerous [3,4]. On the other side, an actual tumor remains undetected in a false negative detection. Studies have

shown that 10-30% of the visible cancers are undetected [5]. Thus, there is a significant necessity for developing systems for automatic classification of suspicious areas in mammograms for aiding radiologists to avoid unnecessary biopsies and improve the efficiency of screening programs [6]. Computer aided detection (CAD) systems, which use computer technologies to detect abnormalities such as microcalcification, mass, architecture distortion and asymmetry, can play a key role in early detection of breast cancer and help to reduce the mortality rate among women with breast cancer. CAD in the field of digital mammography are divided into two main categories: computer aided detection (CADe) methods that are able to pinpoint the suspicion regions in mammograms and computer aided diagnosis (CADx) methods which are able to make a decision whether the suspicion regions consist of abnormal or healthy tissue and distinguish between benign and malignant tumors [7]. CAD systems for detecting masses or microcalcifications in mammograms have already been used and proven to be potential powerful tools [8]. The radiologists are attracted

---

* Corresponding author e-mail: mohamed.meselhy@petronas.com.my

by the effectiveness of clinical applications of CAD systems. CAD systems still need to be improved to meet the requirements of screening programs. Feature selection is a main point that should be taken under consideration when implementing a CADx system for recognizing breast tissue.Selecting the most significant features that have the capability to describe and maximize the differences between different tissues in an ample way. Feature selection is an important factor that directly affects the classification result.Most systems extract features to detect abnormalities and classify them as benign or malignant. The classification of malignant and benign is still a challenging problem for researchers [9]. There are various feature extraction methods that serve to condense input data and to reduce redundancies by highlighting important characteristics of the image. The features of digital images can be extracted directly from the spatial data or from a different space after using a transform such as Fourier transform, wavelet transform or curvelet transform [10,11,12]. Multiresolution analysis provides a very sparse and efficient representation for images. In recent years, several schemes for mammogram analysis using wavelet were introduced. Liu et al.[13] proved that the use of multiresolution analysis of mammograms improves the effectiveness of any diagnosis system based on wavelet coefficients. In their mammogram analysis study, they used a set of statistical features with binary tree classifier in their diagnosis system to detect the spiculated mass. The achieved successful classification rate was 84.2%. Mousa et al.[3] proposed a system based on wavelet analysis. They used anadaptive neuro-fuzzy inference system (ANFIS) for building the classifier to distinguish normal from abnormal and to determine whether the type of abnormality is mass or microcalcification. The maximum classification rate obtained was 85.4%.Rashed et al. [14] studied the multiresolution analysis of digital mammogram using wavelet transform to extract a fractional amount of the biggest coefficients. They used daubechies -4,-8,-16 wavelet functions with four levels of decomposition. Euclidian distance was used toclassify between microcalcification clusters, spiculated mass, circumscribed mass, ill-defined mass and normal tissue. The maximum classification rate achieved was 87.06%. Ferreira and Borges [15] proposed a system to classify mammogram images by transforming the images into wavelet bases and then using a set of the biggest coefficients from first level of decomposition as the feature vector toward separating microcalcification clusters, spiculated mass, circumscribed mass and normal classes of images. The maximum classification rate achieved was 94.85Moayedi et al. [16] presented a study of contourlet based mammography mass classification using support vector machine (SVM). In their study, a set of statistical properties of contourlet coefficients from 4 decomposition levels, co-occurrence matrix features and geometrical features are used as feature vectorfor the region of interest (ROI). Genetic algorithm was used for

feature selection based on neural network pattern classification. They concluded that the contourlet features offer an improvement of the classification process. Eltoukhy et al.[17] presented a study of mammogram classification based on curvelet transform. A fractional amount of the biggest coefficients from each decomposition level is used as feature vector. They found that using the biggest coefficients of curvelet gave interesting results. In this paper a multiresolution based feature extraction method is proposed and tested using mammographic image analysis society (MIAS) dataset. The method starts with calculating class representatives after curvelet or wavelet transform. Then, it maintains the features that maximize the distance between the obtained class representatives.The proposed systemis tested on two classification problem stages. The first stage is to differentiate between normal and abnormal mammograms. The second stage is to distinguish between benign and malignant tumors.

2 Preliminaries

2.1 Wavelet transform

2.1.1 Multiresolution and one dimensional wavelet representation

Z and R denote the set of integers and real numbers respectively. The original signal$f(x)$ is measurable and has finite energy: $f(x) \in L^2(R)$. The multiresolution approximation of $f(x) \in L^2(R)$ at a resolution 2^j is defined as the orthogonal projection of a signal on the vectorspace V_{2^j} of $L^2(R)$. The V_{2^j} can be interpreted as the set of all possible approximations at the resolution 2^j of functions in $L^2(R)$, which is detailed in [18]. The approximation $A_{2^{j+1}}f(x)$ at resolution 2^{j+1} contains more information than the approximation $A_j f(x)$ at resolution 2^j. The details signal of $f(x)$ at resolution 2^j is denoted by $D_{2^j}f(x)$. The details can be defined as the difference between $A_{2^{j+1}}f(x)$ and$A_j f(x)$. $D_{2^j}f(x)$ is equivalent to the orthogonal projection of $f(x)$ on the complement Q_{2^j} of vector space V_{2^j} in $V_{2^{j+1}}$. According to the theory of multiresolution signal decomposition [18], there exists a unique scaling function $\varphi(x) \in L^2(R)$ and a unique corresponding wavelet function $\psi(x) \in L^2(R)$, where $\varphi_{2^j}(x) = 2^j \varphi(2^j x)$ and $\psi_{2^j}(x) = 2^j \psi(2^j x)$, such that $\{2^{-j/2}\varphi_{2^j}(x - 2^{-j}k)\}_{k\in z}$ and $\{2^{-j/2}\psi_{2^j}(x - 2^{-j}k)\}_{k\in z}$ are orthogonal bases of V_{2^j} and Q_{2^j} respectively. The approximation and detail signals of the original signal $f(x)$ at resolution 2^j are completely characterized by the sequence of $f(x)$ inner products of with φ_{2^j} and ψ_{2^j} as follows:

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

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

Let H be a low-pass filter and G be a high-pass filter, where the impulse response of the filter H is $h(k) = \langle \varphi_{-1}(x), \varphi(x-k) \rangle$, and the impulse response of the filter G is $g(x) = \langle \psi_{-1}(x), \psi(x-k) \rangle$. Define $\widetilde{H}$ with impulse response $\widetilde{h}(k) = h(-k)$ to be the mirror filter of H, and $\widetilde{G}$ with impulse response $\widetilde{g}(k) = g(-k)$ to be the mirror filter of G. The multi-resolution representation of $f(x)$ at any resolution 2^j can be implemented by a pyramidal algorithm as shown in Figure 1:

$$A_{2^{j-1}}f(x) = \sum_{-\infty}^{\infty} \widetilde{h}(2x-k)A_{2^j}f(k), \text{where } j = 0, -1, -2, \ldots \tag{3}$$

$$D_{2^{j-1}}g(x) = \sum_{-\infty}^{\infty} \widetilde{h}(2x-k)A_{2^j}f(k), \text{where } j = 0, -1, -2, \ldots \tag{4}$$

2.1.2 Two-dimensional wavelet representation

The wavelet model can be extended to twodimensional signals by separable multiresolution approximation of $L^2(R^2)$ with scaling function $\varphi(x,y) = \varphi(x)\varphi(y)$. There are three associated wavelet functions $\psi^1(x,y) = \varphi(x)\psi(y)$, $\psi^2(x,y) = \psi(x)\varphi(y)$ and $\psi^3(x,y) = \psi(x)\psi(y)$, where $\psi(x)$ is the one-dimensional wavelet function associated to $\varphi(x)$. With this formulation, the wavelet decomposition of a two dimensional signal can be computed with a separable extension of the one-dimensional decomposition algorithm as shown in Fig. 1:
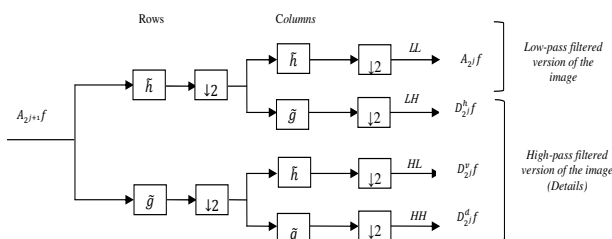

Fig. 1: A wavelet decomposition of an image

Fig. 2 illustrates the multilevel decomposition of the image $A_{2^{j+1}}$ into $A_{2^j}f$, $D_{2^j}^h f$, $D_{2^j}^v$ and $D_{2^j}^d f$ in the frequency domain. The images $A_{2^j}f$, $D_{2^j}^h f$, $D_{2^j}^v$ and $D_{2^j}^d f$ correspond respectively to the lowest frequencies, the vertical high frequencies (horizontal edges), the horizontal high frequencies (vertical edges) and the high frequencies in both directions (diagonal), i.e. $A_{2^{j+1}} = A_{2^j}f + D_{2^j}^h f + D_{2^j}^v f + D_{2^j}^d f$. This set of images is called an orthogonal wavelet representation in two dimensions [18]. The image A_{2^j} is the coarse

approximation at the resolution 2^j, and the images $D_{2^j}^h f$, $D_{2^j}^v f$ and $D_{2^j}^d f$ give the detail signals for different orientations and resolutions. If the original image has N pixels, then each of the images $D_{2^j}^h f$, $D_{2^j}^v f$ and $D_{2^j}^d f$ will have 2^{-j} pixels ($j > 0$), so that the total number of pixels in this new representation is N. This process starts with
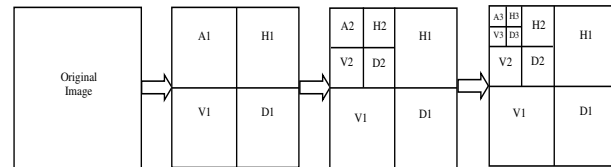

Fig. 2: Three levels multiresolution decomposition wavelet.

decomposing the image into orthogonal sub-bands with low-low (LL), low-high (LH), high-low (HL), and high-high (HH) components that correspond respectively to approximation, horizontal, vertical and diagonal. The LL sub-band is further decomposed into another four sub-bands. The low-low-low-low (LLLL) component represents the image approximation at this level and the process can be continued.

2.2 Curvelet transform

The discrete curvelet transform was proposed by Candes and Donoho [19], from the idea of representing a curve as superposition of functions of various length and width obeying the curvelet scaling law $width \approx length^2$. Fig. 3 presents the curvelet analysis method. The second generation of curvelet transform is presented in[20]. The work is done throughout in two dimensions, i.e. R^2 with x as spatial variable, ω as frequency domain variable, r and θ as polar coordinates in the frequency domain. A pair of windows $W(r)$ and $V(r)$ are defined as the radial window and angular window respectively. These are smooth, nonnegative and real-valued, with W taking positive real arguments and is supported on $r \in (1/2, 2)$ and V taking real arguments and is supported on $t \in [-1, 1]$. These windows will always obey the admissibility conditions:

$$\sum_{j=-\infty}^{\infty} W^2(2^j r) = 1, \quad r \in (3/4, 3/2) \tag{5}$$

$$\sum_{j=-\infty}^{\infty} V^2(t-1) = 1, \quad t \in (-1/2, 1/2) \tag{6}$$

For each $j \geq j_0$, a frequency window U_j is defined in the Fourier domain by:

$$U_j(r, \theta) = 2^{-3/4j} W(2^{-j}r) V\left(\frac{2^{\lfloor j/2 \rfloor}\theta}{2\pi}\right) \tag{7}$$
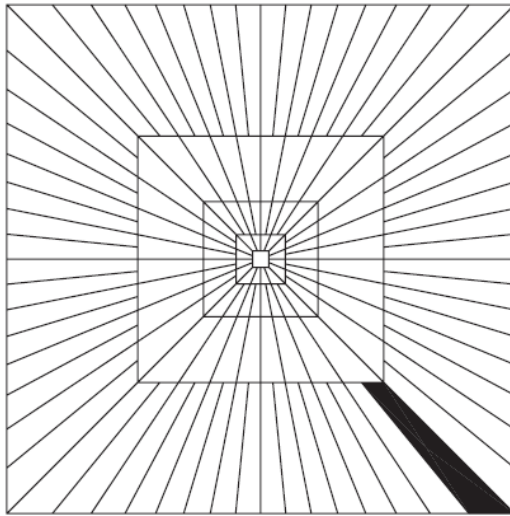
Fig. 3: Curvelet basic digital tiling in two dimensions. The shaded region represents one such typical wedge.

Where $\lfloor j/2 \rfloor$ is the integer part of $j/2$. Thus the support of U_j is a polar wedge defined by the support of W and V and is applied with scale dependent window widths in radial and angular directions. The symmetriezed version of (7), namely, $U_j(r,\theta) + U_j(r,\theta+\pi)$ is used to obtain real valued curvelet. The waveform $\varphi_j(x)$ is defined by means of its Fourier transform $\hat{\varphi}(\omega) = U_j(\omega)$. Let $U_j = (\omega_1, \omega_2)$ be the window defined in the polar coordinate system by (7). φ_j is the mother curvelet in the sense that all curvelet at scale 2^{-j} are obtained by rotations and translations of φ_j. A sequence of translation parameters $k = (k_1, k_2) \in Z^2$ and Rotation angles $\theta_l = 2\pi 2^{-\lfloor j/2 \rfloor} l$ are introduced, with $l = 0, 1, 2, ...$ such that $0 \leq \theta \leq 2\pi$ (the spacing between consecutive angles is scale-dependent). The curvelet functions are functions of $x = (x_1, x_2)$ defined at scale 2^{-j} , orientation angle θ_l and position $x_k^{(j,l)} = R_{\theta_l}^{-1}(k_1 2^{-j}, k_2 2^{-j/2})$ by:

$$\varphi_{j,l,k}(x) = \varphi_j(R_{\theta_l}(x - x_k^{(j,l)})) \tag{8}$$

Where R_θ is the rotation by θ radian and R_θ^{-1} is its inverse,

$$R_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

$$, R_\theta^{-1} = R_\theta^t = R_{-\theta}$$

A curvelet coefficient is the inner product of an element $f \in L^2(R^2)$ and a curvelet $\varphi_{j,l,k}$,

$$c(j,l,k) := \int_{R^2} f(x)\overline{\varphi_{j,l,k}}dx \tag{9}$$

Curvelet transform obeys an anisotropy scaling relation, $length \approx 2^{-j/2}, width = 2^{-j}$, such that $length \approx width^2$.

Fast digital curvelet transform can be implemented via two methods, using unequispaced FFTs or using wrapping.

3 Proposed method

The proposed method is based on maximizing the difference between two different classes representatives in a 2 classes classification problem (e.g. normal versus abnormal or benign versus malignant). The class representative is for example the mean of vectors (of wavelet or curvelet coefficients) representing images in the class. The vector of absolute differences between the two class representatives is then calculated. A dynamic threshold value is applied to the vector of absolute differences to remove the features (the columns here) that have values lower that the threshold.. The method requires an initial dataset, named building set. The building set contains images from the two classes considered.The building set is used to construct the vector of absolute differences. Let the number of images in the building set be K. All K images are decomposed using wavelet or curvelet transform. The obtained coefficients are used to construct a matrix $K \times N$ (K rows, N columns), where N is the number of coefficients obtained for each image. Typically is a large number. For each class, a mean feature vector is obtained by calculating the means of the coefficients of all images in that class. From the previous step, a second matrix $P \times N$ is constructed where P is the number of classes, i.e. each row corresponds to the mean of the classes' feature vectors. For two classes $P = 2$. Let the data of the two classes A and Bin the feature space be given as $\{x_{A_i}, i = 1, \cdots, n_A\}$ and $\{x_{B_j}, j = 1, \cdots, n_B\}$, where each element is a vector of coefficients. The mean of both classes A and B are defined as follows:

$$\mu_A = \frac{1}{n_A} \sum_{i=1}^{n_A} x_{A_i}$$

$$\mu_B = \frac{1}{n_B} \sum_{i=1}^{n_B} x_{B_i}$$

Therefore, the vector of absolute differences is calculated as follows:

$$D(A,B) = |\mu_A - \mu_B|$$

Each entry of $D(A,B)$indicates how the corresponding feature separates the classes. A feature will be kept, if the obtained value is greater than a threshold value, otherwise, it is suppressed. The most significant features are kept according to the applied threshold value. The obtained features are then used for classification using Support Vector Machine (SVM) classifier. In the classification stage, the dataset was divided into two groups, training group 70% and testing group 30%. The training group was used to build SVM classifier; the testing group was used to calculate the performance of the

classifier. To optimize the number of features with the maximum classification accuracy rate, the threshold value is changed and the classification is performed again using the new features set. This process is repeated until reaching a maximum performance with a minimum number of coefficients.

4 Implementation

In the present study, the MIAS data set is used to test the proposed technique. These images were previously investigated and labeled by an expert radiologist based on a technical experience and biopsy. The data set is composed of 322 mammogram of right and left breast from 161 patients, where 51 were diagnosed as malignant, 64 as benign and 207 as normal. Table 1 presents the data set distribution between different classes, benign and malignant.

Table 1: The distribution of (MIAS) dataset

Class	Benign	Malignant	Total
Microcalcification	12	13	25
Circumscribed masses	19	4	23
Ill-defined masses	7	7	14
Spiculated masses	11	8	19
Architectural distortion	9	10	19
Asymmetry lesion	6	9	15
Normal tissue	-	-	207
Total	64	51	322

The original mammograms are 1024×1024 pixels, and almost 50% of the image is comprised of the background with lot of noise. A manual cropping operation is therefore applied to the images to cut off the unwanted regions of the images,resulting to Region of Interest (ROIs)of size 128×128 pixels. The centers of abnormalities area (given by experts) are selected to be the centers of the ROIs. In that way, no abnormality is missed with the background. The data set was divided into two equal sets. The first set is used to construct the feature vectors, while the second set is used to test the proposed method. Wavelet and curvelet transforms are used to represent mammogram images as a pre-process for mammogram classification. In this work, four decomposition levels are used for both wavelet and curvelet. The four levels are obtained from the formula , where is the size of the image, in this work [21]. In this study, the proposed method is applied to solve two mammogram classification problem stages:
1. The classification of normal versus abnormal class.
2. The classification of risk level of cancerous cells based on tumor nature (i.e. benign versus malignant).

The proposed method can be summarized as shown in Fig. 4 as follows:
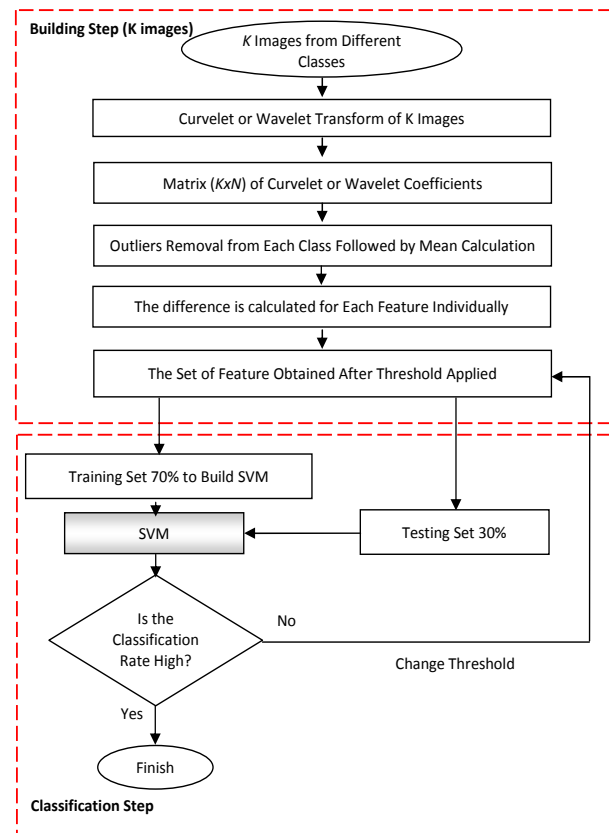

Fig. 4: The proposed method for mammogram classification using difference vector

5 Results and discussion

This section presents and evaluates the results of the experiments carried out according to the two stages of mammogram classification problem. The first stage is the classification of normal versus abnormal classes. The second stage is to differentiate between benign and malignant tumors. In this work, the average of classification accuracy rate is calculated based on 2x5-fold cross validation using SVM. Firstly, we apply the wavelet to the dataset, followed by the proposed method for feature extraction. Then SVM is built and the classification accuracy rate is calculated. Fig. 5 and Fig. 6 illustrate the performance of the SVM classifier to distinguish between normal and abnormal or between benign and malignant corresponding to the number of extracted feature with different threshold values. It shows that the maximum accuracy obtained is 91.67% in case of normal and abnormal, while it is 92.16% in case of benign and malignant. On the other side, the curvelet is applied to the dataset and the maximum classification accuracy rate obtained is 94.41% in case of normal and abnormal, while it is 96.07% in case of benign and malignant as shown in Fig. 7 and Fig. 8. To validate the
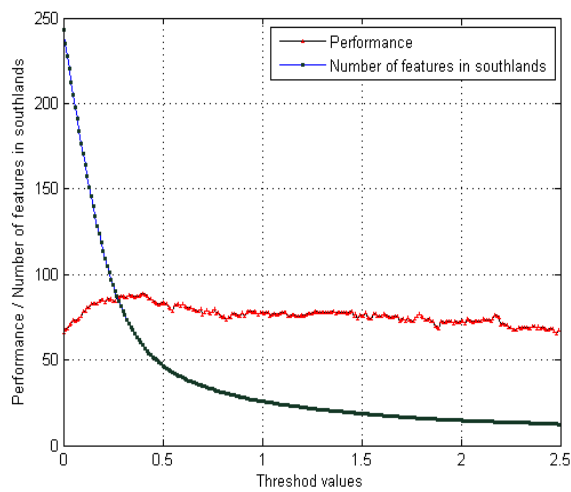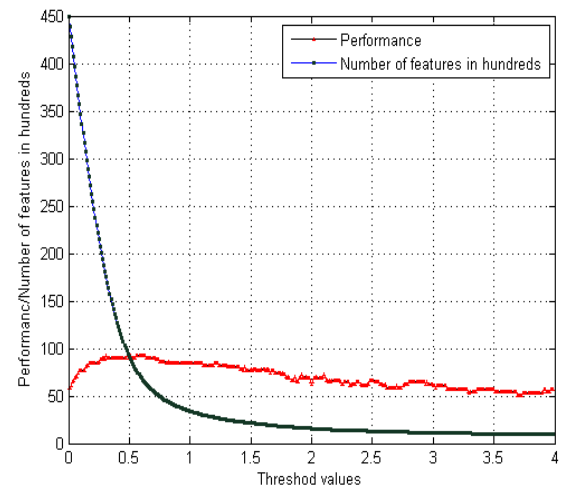
Fig. 5: The classification accuracy rates for normal and abnormal classes corresponding to the number of features with different thresholds using wavelet coefficients.
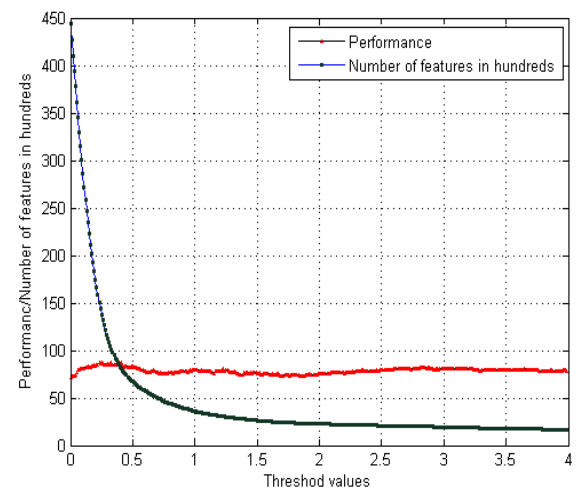
Fig. 7: The classification accuracy rates for normal and abnormal classes corresponding to the number of features with different thresholds using curvelet coefficients.
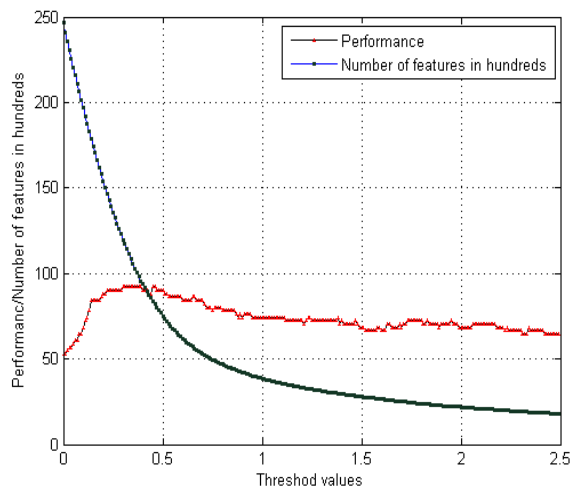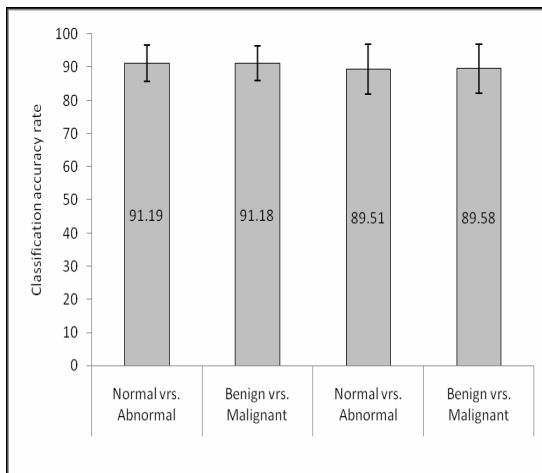

Fig. 6: The classification accuracy rates for benign and malignant classes corresponding to the number of features with different thresholds using wavelet coefficients

Fig. 8: The classification accuracy rates for benign and malignant classes corresponding to the number of features with different thresholds using curvelet coefficients.

results, 2×5-folds cross validation method was applied at the optimized threshold point. The classification accuracy rates are calculated using the obtained coefficients from the optimized threshold values. The average of the classification rates of 2×5 folds obtained for classification of normal and abnormal or between benign and malignant while using either wavelet or curvelet are presented in Fig. 9. The error bar represents

the standard deviation between different folds. A comparison study with previous work has been achieved, the proposed method compared to the work presented in [22]. To test whether the difference between the classifications rates of the proposed method and the method presented in [22] is statistically significant, a paired t-test is performed on the results of the 2×5-fold cross-validation. Let μ_1 and μ_2 be respectively the mean accuracies of the proposed method and the method in ref [22]. The null hypothesis is that the difference between

Fig. 9: Classification accuracy rates obtained through 2×5-folds cross validation.

6 Conclusion

A feature extraction method for finding the most significant coefficients was proposed and implemented to classify a set of mammogram images. This work focuses on combining the advantages of multiresolution representations and SVM. The method is based on identifying the most significant feature among the multiresolution representation, by investigating the statistical characteristics of the row data (wavelet or curvelet coefficients). The proposed method tested over two classification problems of mammogram. The reported accuracy is 91.19% for the first stage to distinguish between the normal and abnormal tissue. For the second problem, the accuracy is 91.18% for differentiating between benign or malignant tumors. We believe that improvement in the classification overall accuracy rates is attributed to the selection of the most significant features. The obtained results support the claim that curvelet are able to present the curvatures better then wavelet.

Acknowledgment

the means of the two methods is zero $H_0 : \mu_1 - \mu_2 = 0$, and the alternative hypothesis is that the difference is positive ($H_a : \mu_1 - \mu_2 > 0$). For a trial i of the cross validation, let P_1^i and P_2^i be respectively the results obtained by using the proposed method and the method in [23]. The test statistic is computed as follows:

$$ t = \frac{\overline{P}\sqrt{n}}{\sqrt{\frac{\sum_i^n (P^i - \overline{P})^2}{n-1}}} $$

where $\overline{P} = \frac{1}{n}\sum_1^n P^i$, $P^i = P_1^i - P_2^i$ and n is the number of runs, $n = 10$ in this work. The P-value is obtained from a t-distribution table at the degree of freedom (n-1), and is compared to the critical value 0.05 (i.e. 5% significance level). If the P-value is less than 0.05, i.e. the null hypothesis is rejected at 0.05 significance level. The results at 5% significance level are summarized in Table 2. The null hypothesis is accepted at 0.05 significance level for all classification problems. It means that the classification rates obtained using the proposed method are similar to those obtained by using the method presented in [22], and there is no statistically significant between the proposed method and the feature ranking method presented in [22].

Table 2. The results of *t-test* at significance level $\alpha = 5\%$ to compare the proposed method with feature ranking method.

Function	Method	T-value	P-value	Null Hypothesis (H_0)
Abnormal vs. Normal	Curvelet	1.9355	0.0849	accepted
	Wavelet	1.9355	0.0747	accepted
Benign vs. Malignant	Curvelet	1.3066	0.2206	accepted
	Curvelet	0.4057	0.6945	accepted

References

[1] http://www.worldwidebreastcancer.com/learn/breast-cancer-statistics-worldwide/ (Visited June 2013).

[2] B. Verma, P. Zhang, A novel neural-genetic algorithm to find the most significant combination of features in digital mammograms, Applied soft computing, **7**, 612-625 (2007).

[3] R. Mousa, Q. Munib, A. Moussa, "Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural". Expert systems with applications, **28**, 713-723 (2005).

[4] H.S. Zadeh, F.R. Rad, S.P. Nejad, "Comparison of multiwavelet, wavelet, Haralick, and shape features for microcalcification classification in mammograms", Pattern recognition, **37**, 1973-1986 (2004).

[5] H.D. Cheng, X. Cai, , X. Chen, L. Hu, X. Lou, "Computer aided detection and classification of mocrlcalcification in mammograms: a survey", Pattern recognition, **36**, 2967-2991 (2003).

[6] I. Christoyianni, A. Koutras, E. Dermatas, G. Kokkinakis, "Computer aided diagnosis of breast cancer in digitized mammograms", Computerized medical imaging and graphics, **26**, 309-319 (2002).

[7] J. Tang, R. Ranjayyan, I. El Naqa, Y. Yang, "Computer aided detection and diagnosis of breast cancer with mammogram: recent advances", IEEE Transaction information technology in biomedicine, **13**, 236-251 (2009).

[8] R. Ranjayyan, F. Ayers, J. Desautles, "A review of computer aided diagnosis of breast cancer: Toward the detection of subtle signs". Franklin Institute, **344**, 312-348 (2007).

2928

M. M. Eltoukhy, I. Faye: An Optimized Feature Selection Method For...

[9] B. Verma, P. McLeod, A. Klevansky, "Classification of benign and malignant patterns in digital mammograms for the diagnosis of breast cancer", Expert systems with applications, **37**, 3344-3351 (2010).

[10] M.M. Eltoukhy,I. Faye, B.S. Belhaouari, "Breast cancer diagnosis in digital mammogram using multiscale curvelet transform",Computerized medical imaging and graphics, **34**, 269-276 (2010).

[11] M.M. Eltoukhy, I. Faye, B.S. Belhaouari, A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram, computers in biology and medicine, **40**, 384-391 (2010).

[12] U. Sayed, M.A. Mofaddel, W.M. Abd-Elhafiez, M.M. Abdel-Gawad, "Image Object Extraction Based on Curvelet Transform", An International Journal of Applied Mathematics & Information Sciences, **7**, 133-138 (2013).

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

[14] E.A. Rashed, I.A. Ismail, S.I. Zaki., "Multiresolution mammogram analysis in multilevel decomposition". Pattern Recognition Letters, **28**, 286-292 (2007).

[15] C. B. R. Ferreira, D. L. Borges. "Analyses of mammogram classification using a wavelet transform decomposition". Pattern Recognition Letters, **24**, 973-982 (2003).

[16] F. Moayedi, Z. Azimifar, R. Boostani, S. Katebi, "Contourlet-based mammography mass classification using the SVM family", Computers in biology and medicine, **40**, 373-383 (2010).

[17] M. M. Eltoukhy, I. Faye, S. Brahim-Belhaouari, "Using curvelet transform to detect breast cancer in digital mammogram," proc. of The 5th International Colloquium on Signal Processing and its Application (CSPA) , Kuala Lumpur, (2009).

[18] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, **7**, 674-93 (1989).

[19] E. Candes, D. Donoho, Curvelets, multi-resolution representation, and scaling laws, Wavelet Applications in Signal and Image Processing VIII, A. Aldroubi, A. F. Laine, M. A. Unser eds., Proc. SPIE 4119, (2000).

[20] E. Candes, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms. Multiscale Model. Simul., **5**, 861-899 (2006).

[21] http://www.curvelet.org,CurveLab-2.1.2Toolbox, (2004).

[22] M.M. Eltoukhy, I. Faye, B.B. Samir, "A statistical based feature extraction method for breast cancer diagnosis in digital mammogram using multiresolution representation", Computers in Biology and Medicine, **42**, 123-128 (2012).

Mohamed Meselhy is an Assistant professor at Computer Science Department, Faculty of Computers and informatics, Suez Canal University, Egypt. He received his PhD in Electrical and Electronic Engineering, Universiti Teknologi PETRONAS, Malaysia. Currently, he is a Postdoctoral researcher in Center for Intelligent Signal and Imaging Research (CISIR), Universiti Teknologi PETRONAS, Malaysia. He is a member of the IEEE. He has published in several peer reviewed reputed journals and conference proceedings. He is the author of one book and a reviewer for several international journals. His main research interests are: Artificial Intelligence, medical image analysis and Multi resolution Representation.

Ibrahima Faye is an Associate Professor at Universiti Teknologi PETRONAS, Tronoh, Malaysia. He is attached to the Department of Fundamental and Applied Sciences and the Centre for Intelligent Signal and Imaging Research. He received a BSc, MSc and PhD in Mathematics from University of Toulouse and a MS in Engineering of Medical and Biotechnological data from EcoleCentrale Paris. His research interests include Engineering Mathematics, Image Processing and Pattern Recognition. He is a member of the IEEE, the IEEE Engineering in Medicine and Biology Society, and the French Mathematical Society.