Expert Systems with Applications 39 (2012) 11036–11047

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Texture extraction: An evaluation of ridgelet, wavelet and co-occurrence based methods applied to mammograms

Rodrigo Pereira Ramos [a,1], Marcelo Zanchetta do Nascimento [b,*], Danilo Cesar Pereira [b,2]

[a] Colegiado de Engenharia Elétrica, Universidade Federal do Vale do São Francisco, Avenida Antônio Carlos Magalhães, 510, 48902300 Juazeiro, BA, Brazil
[b] Centro de Matemática, Computação e Cognição, Universidade Federal do ABC, Rua Santa Adélia, 166, 09210170 Santo André, SP, Brazil

ARTICLE INFO

Keywords:
Mammography
CADx
Texture extraction
Co-occurrence
Wavelet and ridgelet

ABSTRACT

Image processing algorithms can be used in computer-aided diagnosis systems to extract features directly from digitized mammograms. Typically, two classes of features are extracted from mammograms with these algorithms, namely morphological and non-morphological features. Image texture analysis is an important technique that represents gray level properties of images used to describe non-morphological features. This technique has shown to be a promising technique in analyzing mammographic lesions caused by masses. In this paper, we evaluate texture classification using features derived from co-occurrence matrices, wavelet and ridgelet transforms of mammographic images. In particular, we propose a false positive reduction in computer-aided detection of masses. The data set consisted of 120 cranio-caudal mammograms, half containing a mass, rated as abnormal images, and half with no lesions. The following texture descriptors were then calculated to analyze the regions of interest (ROIs) texture patterns: entropy, energy, sum average, sum variance, and cluster tendency. To select the best set of features for each method, we applied a genetic algorithm (GA). In the ROIs classification stage, we used the Random Forest algorithm, a data mining technique that separates the data into non-overlapping segments. Experimental results showed that the best classification rates were obtained with the wavelet-based feature extraction using GA for selection of the most relevant features, giving an AUC = 0.90.

1. Introduction

Breast cancer is the most frequently diagnosed cancer in women all over the world. It is the second most common and leading cause of cancer death among women. Approximately one out of ten women could develop breast cancer during her lifetime. The highest incidence rate occurs in women over 35 years of age. According to published statistics, breast cancer has become a major health problem in both developed and developing countries over the past 50 years, and its incidence has increased in recent years (Tang, Rangayyan, Xu, El, & Yang, 2009). In Brazil there were about 49,240 new cases of breast cancer diagnoses and 11,860 deaths from this disease among women in 2010 (INCA, 2010). For this reason, studies have shown that early detection is the key to improve breast cancer prognosis (Elter & Horsch, 2009a, 2009b; Oliver et al., 2010).

Screen/film mammography has proved to be an effective aid for radiologists in the early detection of occult breast cancers in asymptomatic women and in the reduction of mortality rates. Although it is seen as the most reliable method for early detection of breast cancer, its interpretation is very difficult. Sensibility of screening mammography is affected by image quality and depends on the radiologist's level of expertise. As a result, some studies show that approximately 10–30% of malignant breast cancers visible on mammograms go undetected by radiologists during routine mammographic screening (Elter & Horsch, 2009a, 2009b). Also, it has been observed that only 15–34% of women who undergo a biopsy based on the results of a mammographic examination actually have malignant lesions. To improve the accuracy of mammography, double reading of the same screening mammogram has proven to increase the sensitivity rate (Gupta, Chyn, & Markey, 2006). However, the double reading performed by two radiologists increases the operating cost.

Computer-aided detection and computer-aided diagnosis (CAD) are procedures applied in mammography to assist radiologists in the interpretation of suspicious lesions such as calcification, mass and architectural distortion. A CAD system can act as a second reader, prompting the radiologist to review areas in a mammogram to be suspicious by specialized computer algorithms. Gromet

* Corresponding author. Tel./fax: +55 11 4996 0077.
E-mail addresses: rodrigo.ramos@univasf.edu.br, godoga@gmail.com (R.P. Ramos), marcelo.zanchetta@gmail.com, marcelo.nascimento@ufabc.edu.br (M.Z. Nascimento), danilo.pereira@ufabc.edu.br (D.C. Pereira).
[1] Tel./fax: +55 87 3614 1935.
[2] Tel./fax: +55 11 4996 0077.

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.eswa.2012.03.020

(2008) showed that both double reading and CAD are effective methods to increase the sensitivity of screening mammography for experienced mammogram readers. In his study, the use of a second reader increased sensitivity rates in 6.6%, growing from 81.4% to 88.0%, while with a single CAD reading sensitivity rates were enhanced to 90.4%. Therefore, the CAD system appears to be an effective alternative that provides similar, and potentially greater, benefits.

CAD systems can be broadly categorized into two types – computer-aided detection (CADe) and computer-aided diagnosis (CADx). CADe schemes are systems that automatically detect suspicious lesions in mammograms, being used as a localization task. For a survey of the state of the art of CADe systems, the reader is referred to a recent review article by Oliver et al. (2010). CADx systems extend the computer analysis to yield as output the characterization of a region or the estimated probability of lesion malignancy. This system is focused on the classification task. Image processing algorithms are used in CADx systems to extract features directly from digitized mammograms. Typically, two classes of features are extracted from mammograms with these algorithms, namely morphological and non-morphological features. Morphological features are intended to describe information related to the morphology of a lesion, such as lesion size and shape. Image texture analysis is an important class that represents gray level properties of images used to describe non-morphological features.

Texture analysis methods have been used in the analysis and interpretation of medical images. Texture is characterized by a set of local statistical properties of pixel intensities. In mammographic image processing, these features have been used to distinguish density patterns that indicate different levels of risk to developed malignant lesions. Texture analysis has shown to be a promising technique in analyzing mammographic lesions caused by masses (Gupta & Markey, 2005; Lim & Er, 2004; Mavroforakis, Georgiou, Dimitropoulos, Cavouras, & Theodoridis, 2006; Paquerault, Petrick, Chan, Sahiner, & Helvie, 2002). Textural information are important to outline the performance of CADx system, being required for the classification that distinguishes masses from normal tissues (Qian, Sun, Song, & Clark, 2001).

Digital image features can be extracted directly from the spatial pixel relations or from some different domain. Traditionally, texture features have been calculated using statistical properties such as first-order features based on gray-level intensity histograms and second-order features based on co-occurrence matrices (Mudigonda, Rangayyan, & Desautels, 2001; Sahiner, Chan, Petrick, Helvie, & Hadjiiski, 2001). Texture analysis methods based on co-occurrence matrices are often used for their ability to capture the spatial dependence of gray-level values within an image. There are several types of texture features that have been studied for CADx, in particular the texture analysis developed by Haralick, Shanmugam, and Dinstein (1973) are very commonly used.

Multiresolution analysis has proved to be useful in mammographic image processing, image enhancement, mass detection, and feature extraction. The common task is to decompose the original image into sub-bands that preserve high and low frequency information. Several studies have investigated the use of wavelet transform as a multiresolution analysis tool for texture analysis and classification (Liu, Babbs, & Delp, 2001; Mousa, Munib, & Moussa, 2005; Rashed, Ismail, & Zaki, 2007; Starck, Elad, & Donoho, 2004). When applied to images, wavelet transform produces a low resolution image and several detail images. The low resolution image is obtained by iteratively blurring the image; the detail images contain the information lost during this operation. The success of wavelets is mainly due to the good performance for piecewise smooth functions in one dimension. The wavelet is good at catching point singularities, but two dimensional piecewise smooth signals resembling images have one-dimensional singularities. Recently, the ridgelet transform has emerged as a new multiresolution analysis tool which deals effectively with line singularity in two dimensions (Do & Vetterli, 2003). The idea is to map a line singularity into a point singularity using the Radon transform (Deans, 1983). The ridgelet transform allows for representing singularity along lines in a more efficient way, in terms of compactness of the representation for a given reconstruction accuracy. This transform has been used in texture classification of CT medical (Semler, Dettori, & Kerr, 2006) and mammographic (Selvi & Malmathanraj, 2006) images.

In this paper, we present a comparative study for the use of co-occurrence matrices, wavelet and ridgelet transforms in texture analysis of mammographic images. The proposed model can be seen as a false positive reduction (FPR) algorithm for CADx systems. A FPR algorithm aims to validate whether a given sub-image containing a suspicious region contains a real lesion or it is only a region depicting a normal parenchyma (Lladó, Oliver, Freixenet, Martí, & Martí, 2009; Masotti, Lanconelli, & Campanini, 2009). A simplified FPR/CADx system is proposed intended to classify mammographic images as normal (containing no lesions) or abnormal (with a mass). Fig. 1 presents a schematic diagram of the system proposed in this work. Firstly, the mammograms were pre-processed to excerpt regions of interest. Feature extraction was then performed on each cropped sub-image. Texture features obtained from co-occurrence matrix were calculated at four pixel distances ($d = 1, 3, 6, 9$) (Bovis & Singh, 2000) with the orientation being quantized into four different directions ($\theta = 0°, 45°, 90°$, and $135°$). For multiresolution analysis tools, these features were computed with different decomposition levels based on Daubechies wavelet (Rashed et al., 2007), both for the wavelet and ridgelet transforms. We applied a genetic algorithm for feature selection based on the Random Forest algorithm, a data mining technique that separates the data into non-overlapping segments. At the final stage, we also used the Random Forest algorithm to classify the mammogram images as normal or abnormal. Receiver operating characteristic (ROC) curve analysis is the standard methodology for the evaluation of diagnostic systems in clinical medicine (Metz, 1978; Metz, Herman, & Shen, 1998). The overall performances of the methods were evaluated by means of the area under the ROC curve.

The paper is structured as follows: Sections 2.1 and 2.2 describe in more detail the multiresolution analysis and co-occurrence matrices for feature extraction on mammograms. In Section 2.3,

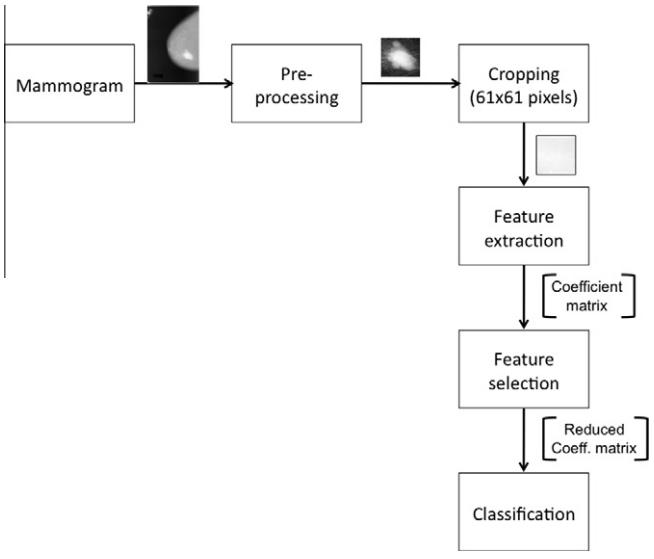

Fig. 1. Diagram of the system proposed in this work.

we present a subset of mammography cases selected from a publicly available database, the Digital Database for Screening Mammography. Next, in Sections 2.4 and 2.5, we describe the texture features selected and explain the genetic algorithm based feature selection method for classifying the images as normal or abnormal. Finally, Section 3 discusses the results of applying the proposed method for mass classification.

2. Materials and methods

2.1. Multiresolution analysis

Image analysis in multiple scales allows image resolution to be changed for processing as little data as possible by selecting relevant details for a given visual task (Mallat, 1996). The basic idea of multiresolution analysis is to represent an image on several sub-images, from coarse to fine resolution, and analyze them in the frequency domain. Broadly speaking, multiresolution allows for the zooming in or out on an image, when this is necessary.

When processing texture extraction algorithms, it is frequently necessary to measure texture features on neighborhoods of different size. Once multiresolution analysis evaluate image properties over domains of varying sizes, its zooming capacity makes the process not to be affected by the size of the pixel neighborhood.

In this article, it is employed two recently used multiresolution transforms for texture extraction: the wavelet transform and the ridgelet transform. The common task of both approaches is to decompose the original image into sub-bands that preserve high and low frequency information. This process is made recursively by high-pass and low-pass filtering the image. The individual characteristics of each transform are presented in the following.

Wavelet transform: A wavelet is a function $\psi(x) \in \mathbf{L^2}(\Re)$ (the set of square integrable real functions) with a zero average, i.e.:

$$\int_{-\infty}^{\infty} \psi(x)dx = 0 \tag{1}$$

normalized to have unit energy and centered in the neighborhood of $x = 0$. Given a function $f(x) \in \Re$, its wavelet transform $Wf(s,u) \in \Re$ is obtained from the inner product of $f(x)$ and a wavelet family, i.e.:

$$Wf(s,u) = \int_{-\infty}^{\infty} f(x)\psi_{s,u}(x)dx \tag{2}$$

where

$$\psi_{s,u}(x) = s^{-1/2}\psi\left(\frac{x-u}{s}\right) \tag{3}$$

are the wavelet family obtained by scaling $\psi(x)$ by s and translating it by u. The scale s is referred to as the transform resolution level.

To allow fast numerical implementations, it is common to impose that the scale and translation parameters vary only along discrete values, the most used being the dyadic wavelet decomposition, which is achieved when $s = 2^j$ and $u = k\,2^j$, for integers j and k. Therefore, one can construct wavelet and scaling families:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t-2^j k}{2^j}\right) \tag{4a}$$

$$\phi_{j,k}(t) = \frac{1}{\sqrt{2^j}}\phi\left(\frac{t-2^j k}{2^j}\right) \tag{4b}$$

that are orthonormal bases of some subspaces of $\mathbf{L^2}(\Re)$ related to the resolution 2^j. Many families of wavelets have been developed, such as Haar, Daubechies, Coiflet, cubic splines, among others (Mallat, 1999).

For discrete signals, the discrete wavelet transform (DWT) is obtained by the discretization of time as well as the translation and scale parameters. Mallat (1999) has proved that the dyadic DWT of a signal is equivalent to its decomposition through high-pass ($g[n]$) and low-pass ($h[n]$) filter banks, composed of as many banks as is the desired resolution. When dealing with images, which are 2-dimensional signals, the DWT can be computed with separable wavelet functions, what means that a 2-dimensional filter can be decomposed as the product of two one-dimensional filters (Mallat, 1999). In this case, a 1D-DWT is computed first in the image rows, and another 1D-DWT is applied to the columns of the two images generated. The 2D-DWT is illustrated in Fig. 2, in which an original image at the jth level (cA_j coefficients) is wavelet transformed, yielding four sub-band images at the $(j+1)$th level. There are three detail images, cD^h_{j+1}, cD^v_{j+1}, and cD^d_{j+1}, that represents horizontal, vertical and diagonal directions, respectively, and one approximation image cA_{j+1}, which is the original image at a coarse resolution.

For each desired resolution, the approximation coefficients from the upper level are applied to the 2D-DWT bank structure of Fig. 2, and others 4 sub-images are generated. An example of the application of a 2D-DWT with 2 resolution levels to a mammographic image is showed on Fig. 3.

Ridgelet Transform: It is well known that wavelets deal successfully with piecewise smooth functions in one dimension, that has only point singularities (Candès & Donoho, 1999), which are points of bad behavior. However, in spite of being widely used in image analysis in higher dimensions, other kinds of singularities are present, and wavelets are not as effective as it would be expected.

To overcome this problem, Candès and Donoho developed the ridgelet transform that deals effectively with 2-dimensional singularities, such as line singularities (Candès & Donoho, 1999). The idea is to map a line singularity into a point singularity using the Radon transform, and then to apply the wavelet transform into the Radon domain.

Given an integrable bivariate function $f(x_1, x_2) \in \Re^2$, its ridgelet transform is defined as:

$$R_f(s,u,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x_1,x_2)\psi_{s,u,\theta}(x_1,x_2)dx_1dx_2 \tag{5}$$

where the 2-dimensional ridgelets $\psi_{s,u,\theta}(x_1,x_2)$ are obtained from an one-dimensional wavelet mother $\psi(x)$ through:

$$\psi_{s,u,\theta}(x_1,x_2) = s^{-1/2}\psi\left(\frac{(x_1\cos\theta + x_2\sin\theta - u)}{s}\right) \tag{6}$$

This function is constant along "ridges" $(x_1\cos\theta + x_2\sin\theta) = \text{constant}$ and is oriented at an angle θ (Do & Vetterli, 2003).

To apply the ridgelet transform to digital images, its discrete version is required, which needs the discrete Radon transform. Do and Vetterli (2003) proposed a procedure to compute the discrete version of the ridgelet transform that is invertible, orthogonal and achieves perfect reconstruction: the finite ridgelet transform (FRIT). The FRIT is based on the finite Radon transform, which is defined as summations of image pixels over a certain set of lines, and can be computed in two steps (Dettori & Semler, 2007): (i) com-

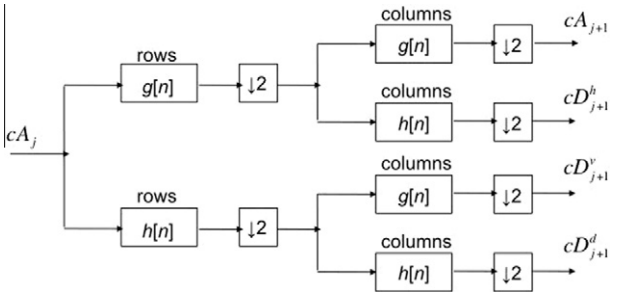

Fig. 2. One stage of a 2D-DWT filter bank.

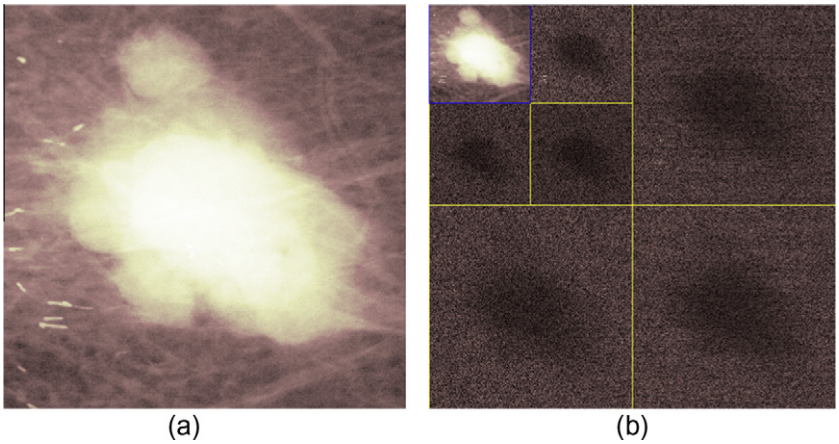

Fig. 3. Wavelet decomposition on 2 resolution levels of a mammographic image: (a) original ROI; (b) decomposed ROI.

pute the discrete Radon transform, through a 2-dimensional fast Fourier transform (FFT) followed by an one-dimensional inverse Fourier transform for each radial direction of the Radon projections; (ii) apply an one-dimensional wavelet transform over r resolutions for each radial direction of the Radon transform. Fig. 4 clarifies the procedure.

In this FRIT scheme, multiresolution comes from the application of 1D-DWT. Therefore, differently from 2D-DWT, FRIT application on a given image generates one approximation coefficient matrix and as many detail coefficient matrices as are the resolution levels. There is a total of $P+1$ directions (columns) for each of these matrices, where P is a prime number representing the size of a square image of dimensions $P \times P$. The primality on P is necessary since the FRIT is based on the finite Radon transform, which is obtained from mathematical operations on a finite field with modulo P (Candès & Donoho, 1999; Do & Vetterli, 2003). An example of the application of the ridgelet transform to a mammographic image can be seen on Fig. 5.

2.2. Co-occurrence matrices

Another way to acquire image characteristics is by the use of a set of information extracted from the gray level co-occurrence matrix, also known as spatial gray level dependence (SGLD) matrix (Gonzalez & Woods, 2002).

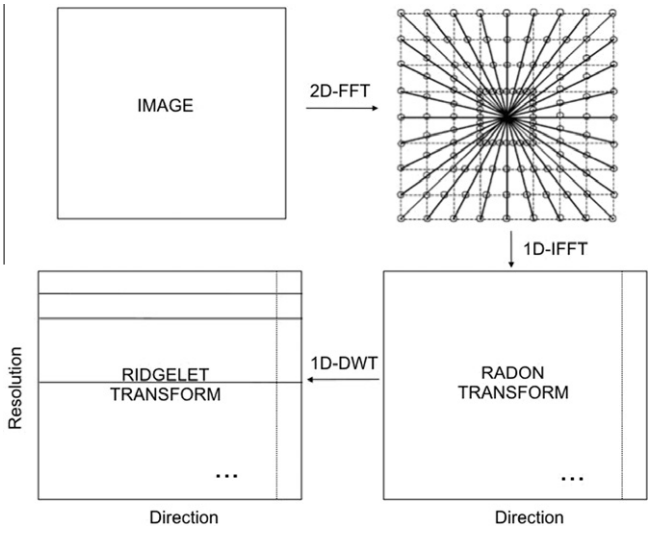

Fig. 4. Ridgelet transform computation.

To describe texture attributes using this technique, a set of characteristics must be computed. These characteristics are obtained from the co-occurrence matrix, which as its name suggests is constructed from the image by estimating the pairwise statistics of pixel intensities. A co-occurrence matrix is an $M \times M$ matrix whose rows and columns are indexed by the image gray levels $i = 1,2,\ldots,M$, where $M = 2^m$ for an m-bit image. Each normalized matrix entry $p(i,j)$ represents the statistical frequency for which a given pair of gray level pixels i and j is found in the image separated by a distance d at an θ direction, or, stated in another way, pixels separated from each other by a distance Δx and Δy in x and y directions, respectively. Once one matrix must be calculated for each pair (d,θ), it is usually computationally necessary to restrict these parameters to a limited number of values. In their paper, Bovis and Singh (2000) considered four pixel distances ($d = 1,3,6,9$) and discretized angles $\theta = 0°$, $45°$, $90°$, and $135°$. These values have also been considered in the present work.

2.3. Data set

The database used in this work encompasses mammographic screen/film digitized images taken from the Digital Database for Screening Mammography (DDSM) (Heath et al., 1998). The DDSM project is a joint effort of researchers from the Massachusetts General Hospital (D. Kopans, R Moore), the University of South Florida (K. Bowyer), and the Sandia National Laboratories – EUA (P. Kegelmeyer). The DDSM database has been widely used as a benchmark for numerous articles on the mammographic area, for being free of charge and having a vast and diverse quantity of cases. It is constituted of mammographic images and its corresponding technical and clinical information, including exam dates, age of patients, digitalization equipment (as well as resolution, number of rows, pixels per row and bits per pixel of the acquired images), lesion types (according to BIRADS®), and existent pathologies.

The data set consisted of 120 mammographic images in craniocaudal view taken from the DDSM set, half containing a mass, rated as abnormal images, and half with no lesions, rated as normal images in the classification process. The normal images have been used for estimating the false-positive ratio, and positive examples were extracted from malignant cases. We selected digitized images with a Lumisys laser film scanner at 50 μm and a Howtek scanner at 43.5 μm pixel size. Each image has a resolution of $M = 2^{12} = 4096$ gray level tones.

The location and size of a mass, when it exists, were taken from the code-chain of the ".ics" file available at the DDSM project and were used to automatically extract square sub-images called regions of interest (ROIs) from the original image. In mass CADx

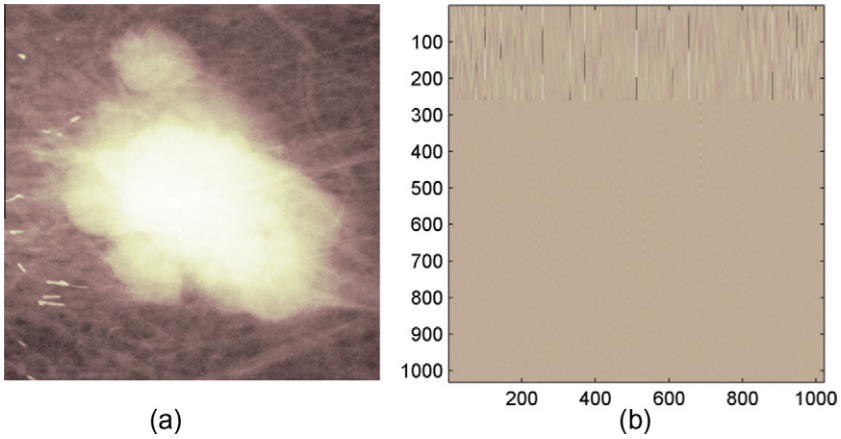

Fig. 5. Ridgelet decomposition on 2 resolution levels of a mammographic image: (a) original ROI; (b) transformed ROI.

approaches, texture analysis is not always performed on the full ROI but often only on special regions within the ROI. Texture analysis is often restricted to the mass region, excluding the background tissue region or on bands of pixels close to the mass margin (Elter & Horsch, 2009a, 2009b). Therefore, the images used in the experiments were cuttings of size 61×61 pixels done in the sub-image, whose centers correspond to the centers of the presented abnormalities. An example of the cropping process that eliminates image label and background is given in Fig. 6. The images were cropped to the referred dimensions because of the size requirements of a prime number for the FRIT and for the need to generate images where masses are fully contained on. To obtain sub-images with no mass, we have followed the same procedure, except that the location was randomly taken from a healthy part of the mammogram. With this approach, a total of 120 sub-images were acquired, and each of these cropped images was used for texture feature extraction and subsequent classification.

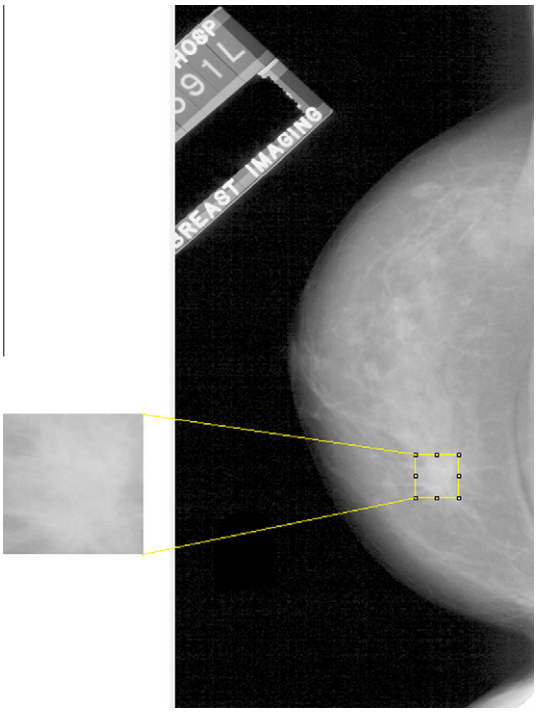

Fig. 6. Mammographic image and regions of interest.

2.4. Texture feature extraction

Once the mammographic images were pre-processed as described in the last section, the three approaches considered earlier were applied and textural feature vectors were extracted. The following texture descriptors were then calculated to analyze the ROIs texture patterns: entropy, energy, sum average, sum variance and cluster tendency. To define the five used descriptors, consider a given matrix $\mathcal{A}$ of size $M \times N$, whose entries are $p(i,j)$, and consider the following notation:

$$p_{x+y}(k) = \sum_{\substack{i=1 \\ i+j=k}}^{M} \sum_{j=1}^{N} p(i,j), \quad k = 2,3,\ldots,M+N \tag{7}$$

$$\mu = \sum_{i=1}^{M} \sum_{j=1}^{N} |p(i,j)| \tag{8}$$

The five descriptors are presented in what follows:

$$\text{Entropy} : f_1 = \sum_{i=1}^{M} \sum_{j=1}^{N} p(i,j) \log(p(i,j)) \tag{9}$$

$$\text{Energy} : f_2 = \sum_{i=1}^{M} \sum_{j=1}^{N} [p(i,j)]^2 \tag{10}$$

$$\text{Sum} - \text{average} : f_3 = \sum_{k=1}^{M+N} k p_{x+y}(k) \tag{11}$$

$$\text{Sum} - \text{variance} : f_4 = \sum_{k=1}^{M+N} (k - f_3)^2 p_{x+y}(k) \tag{12}$$

$$\text{Cluster tendency} : f_5 = \sum_{i=1}^{M} \sum_{j=1}^{N} (i + j - 2\mu)^2 p(i,j) \tag{13}$$

It should be noticed that matrix $\mathcal{A}$ can be any of the sub-images obtained from either the SGLD matrix or a given transform. In the first case, $p(i,j)$ is one entry of the SGLD matrix, and, since it is a square matrix, $M = N$. For wavelet and ridgelet transforms, $p(i,j)$ is taken from the absolute value of a given transform coefficient, since positive and negative contrast levels have the same meaning for mammographic images (Dettori & Semler, 2007). Feature extraction based on the five texture descriptors was implemented for each of the three considered approaches as follows:

- *Co-occurrence matrix*: an SGLD matrix for each ROI was computed for four directions (0°, 45°, 90°, and 135°) and the five descriptors were computed for each matrix, giving a total of

20 descriptors per ROI. These 20 descriptors are extracted from the SGLD matrices constructed at four pixel distances ($d = 1,3,6,9$) giving four separate feature vectors. Once there is 120 ROIs, a feature matrix was generated, with each row representing a feature vector of a given ROI, and this matrix was used by the classifier. For comparison of computational efforts, we have used a full SGLD matrix, with 12 bits of pixel resolution, and a subsampled SGLD matrix, with 8 bits of pixel resolution. This matrix is obtained reducing the image resolution to 8 bits and computing its SGLD matrix.

- *Wavelet transform*: for each ROI, the 2D-DWT was applied using the Db3 wavelet mother and two resolution levels, yielding six detail coefficient matrices and one approximation coefficient matrix. Since most of the information has been removed by performing low-pass filtering iteratively, the approximation matrix (cA_{j+m}) is not considered for further processing (Selvan & Ramarishnan, 2007). From each detail subband, the five descriptors were computed and used in the classification process, giving a total of 30 features for each ROI and a 120×30 feature matrix for all images was used in the classification process. As in the co-occurrence method, it was also considered a full image (12-bits resolution) and a subsampled image (8-bits resolution) as the input image, for comparison purposes.
- *Ridgelet transform*: as for wavelet transform, the FRIT was computed for each ROI, two resolutions and the Db3 wavelet mother was used for the 1D-DWT application, resulting in two detail coefficient matrices, each with 62 columns representing directions (once again it was not considered the approximation coefficients). Once the directions are relevant information returned by the FRIT, all 62 columns were considered as separated descriptors, and the five descriptors were computed for each of those columns of each detail matrix, used for subsequent classification, yielding a total of 620 descriptors for each ROI. A general 120×620 texture feature matrix was used for subsequent classification. The full and subsampled images were also taken into account.

2.5. Feature selection and classification

Feature selection and classification were performed with the help of Matlab environment and WEKA (*Waikato Environment for Knowledge Analysis*) computational software (Vibha et al., 2006). The feature selection was accomplished by using a genetic algorithm (GA) developed in Matlab together with the Random Forest classifier from WEKA. Once the main features were selected, they were passed to the Random Forest classifier for posterior classification.

Feature selection: For the feature selection procedure, it was connected both Matlab and WEKA environments with the aim of integrating the GA and the classifier algorithms (Mazurowski, Habas, Zurada, & Tourassi, 2008; Sahiner et al., 1996). This was done to determine which textural features weight more intensively in the classification process of malignant mammographic images.

Initially, a random population was generated having chromosome sizes equal to the set of features extracted from each image. The genes of these chromosomes were encoded as a binary string, where bit 0 (zero) corresponded to the absence of a given feature and bit 1 (one) to the presence of the same feature. Following, the features selected by the GA were passed to the WEKA Random Forest classifier algorithm to train it and test it. At the end of the execution, information about the classifier performance for the used data set were passed back to the GA, which used the classification area under the ROC curve as the chromosome fitness value.

In order to select the best individuals, a tournament selection procedure was employed. Four individuals of the current population were randomly chosen to compete with each other, and only that with the best fitness value was selected to belong to a mating pool with the same size of the population.

To combine the chromosomes, traditionally two operators are used: one-point and two-point crossover. These points are randomly generated and used to swap all the genes of the parent chromosomes to create two offsprings. In this paper, we performed a two-point crossover with different values of crossover probability. For each method of texture analysis, we applied different values of crossover probability to determine the best value of classification area under the ROC curve. The performance of the crossover probability are demonstrated in the results section. A random number can be generated within [0,1] associated with each pair of strings selected in the mating pool. If the random value is less than that the best value obtained by the texture analysis method under consideration, the crossover is performed. Otherwise, no crossover is implemented.

Mutation operator allows swapping the value of a certain gene depending on a random number generated within [0,1]. In this experiment, we used the mutation operator with different values of mutation probability applied to each analyzed method. In this case, different mutation probabilities were evaluated to determine the best value for classification. In the results section, we present the performance of the different mutation probabilities for the analyzed techniques. If the random number is less than the best obtained value for each method analyzed, a gene on chromosome at a random position is chosen and its value is altered (i.e. one changes 0 to 1 or 1 to 0). This process prevents the premature convergence and enables the GA to explore others research spaces.

The described process was iteratively repeated until 20 generations were reached, that is a number for which there is a stabilization of the best fitness value. After that, the chromosome with the highest fitness value (higher area under the ROC curve) was captured to provide the most relevant textural features to the classifier.

Classification: In the ROIs classification stage, we used a data mining technique that separates the data into non-overlapping segments. To this end, the Random Forest algorithm (Breiman, 2001) implemented on WEKA was utilized.

The Random Forest is defined as a classifier constructed from a collection of classification trees. It is a concept of regression trees induced by bootstrap samples of a training data set, with random features selected in the induction tree process. In this method, each tree is constructed in the following way:

Step 1: data is withdrawn from a training set through a random sampling process with bootstrap, where 2/3 of the data are used for growing the tree.
Step 2: A random number of features is selected from the training set and the one with the largest number of informations is used to split the node.
Step 3: The growing task continues until no node can be created for lack of information.
Step 4: The error rate is estimated using the 1/3 of the data left, by predicting their classes.

The decision algorithm of the Random Forest method works as follows:

1. Select the number of growing trees, represented by the parameter N, and an integer m not greater than the number of features passed to the classifier.
2. For i from 1 to N train the forest.
3. Randomly select a bootstrap sample of the data. The data that are not selected in this step are named as out-of-bag.
4. Grow a random tree, where at each node the best set is chosen among m variables randomly selected.

5. Use the tree to predict the out-of-bag data.
6. At the end, use the results obtained from the out-of-bag data to indicate the class.

In this work, we used the Random Forest algorithms with two classes to classify the objects from the data set as normal or abnormal images. We used each feature matrix obtained from the feature selection approaches described earlier as the input data to the classifier. We employed 10-time 10-folded cross validation method to evaluate the proposed classifier accuracy as well as its generalization capability. In this experiment, the feature vectors are divided into 10 disjoint groups containing 90% of the data for training and the remaining 10% for validation. This process is repeated 10 times (Moayedi, Azimifar, Boostani, & Katebi, 2010). In the application of the 10-fold cross-validation, the groups of data used in the GA feature selection procedure are different from the groups used in the training and testing procedure by randomly defining them. Since we use the entire data set for feature selection, the use of the selected features in Random Forest classification is optimistically biased.

Performance evaluation: To give support to the performance evaluation of each proposed method for feature extraction, we computed sensitivity (true positive rate) and specificity (true negative rate) for each of the confusion matrices (Kohavi & Provost, 1998). A confusion matrix shows the predicted and actual classifications accomplished by a classifier. It has dimensions $L \times L$, where L is the number of classes evaluated by the classifier. In our case, $L = 2$ and the confusion matrix can be stated as shown in Table 1.

The parameters employed for performance evaluation can be summarized in Table 2, which shows the most commonly used formulas. Specificity measures the percentage of positive instances that were predicted as positives, while sensitivity measures the percentage of negative instances that were predicted as negatives.

Performance evaluation was accomplished by means of the *receiver operating characteristic* (ROC) curves. An ROC curve is a two-dimensional graph of test sensitivity, plotted on the y-axis, versus its false positive rate (or 1-specificity), plotted on the x-axis. An ROC graph depicts relative trade-offs between benefits (TP) and costs (FP) (Fawcett, 2006), and is produced by varying the decision threshold of the classifier. Although it sometimes can lose some subtle information on classification results, the area under the ROC curve (AUC) is a good accurate measure of a given classifier performance. A test classification with an AUC of 1.0 is a perfect test and an AUC of 0.0 results in a perfectly inaccurate test.

For validation of the results, it was performed a statistical F-test to evaluate the statistical differences between the features obtained from each proposed method. The F-test is an inference on the equality of the variances of two populations, say, σ_1^2 and σ_2^2, within a given confidence interval α (DeCoursey, 2003). Two data sets are evaluated by testing the hypotheses:

$$H_0 : \sigma_1^2 \neq \sigma_2^2 \quad (14)$$

$$H_1 : \sigma_1^2 = \sigma_2^2 \quad (15)$$

The null hypothesis, H_0, indicates that the data from the two populations are independent random samples with different means and unknown and different variances. The result of the test is given by a binary variable F. The value $F = 1$ indicates the rejection of the null hypothesis at the confidence interval α. Contrarily, a value of $F = 0$ indicates a failure to reject the null hypothesis at the α significance level. In this work, the tests were performed by comparing the methods in a pairwise manner, that is, the feature vectors (rows) of each matrix were pairwise compared, considering a confidence interval $\alpha = 5\%$. In all tested cases, we obtained a value $F = 0$, i.e., a failure to reject H_0, indicating that the data are statistically independent and distinct.

Table 1
Confusion matrix for $L = 2$ classes.

Actual class	Predicted class	
	Positive	Negative
Positive	True positive (TP)	False negative (FN)
Negative	False positive (FP)	True negative (TN)

Table 2
Definition of performance evaluation parameters.

Parameter	Definition
Sensitivity	TP/(TP + FN)
Specificity	TN/(TN + FP)

3. Results and discussion

We have applied the combined feature selection/classification method described earlier to each feature extraction scheme showed in Section 2.4, namely the co-occurrence matrix based method, the wavelet transform based method and the ridgelet transform based method. For comparison purposes, we have divided the results obtained with and without the use of the GA for feature selection.

In Figs. 7 and 8, we have plotted the ROC curves with its corresponding AUC without considering feature selection for SGLD matrices with different pixel distances ($d = 1, 3, 6, 9$), for images of 8 and 12 bits of resolution, respectively.

The results presented in Figs. 7 and 8 show that the variation of the distance parameter influences the classifier performance on images of 8 bits and 12 bits. There is a little gain when increasing image resolution, with the AUC going from 0.73 to 0.74 with no feature selection with pixel distance $d = 1$. However, we can observe that the increase in pixel distances makes it possible to obtain more relevant results for the classifier performance. In the case of subsampled images we can observe that there was a gain for the AUC and the best results were obtained for $d = 6$ and $d = 9$ (AUC = 0.76). For full images (12-bits), the change of distances was able to provide an increase of the AUC to 0.82 (at a distance

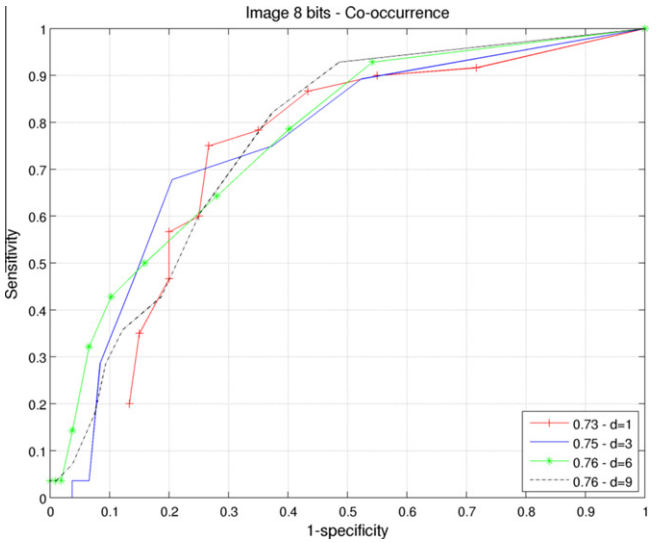

Fig. 7. ROC curves for SGLD matrices constructed at four pixel distances ($d = 1, 3, 6, 9$), with no feature selection algorithm, for subsampled images (8 bits of resolution).

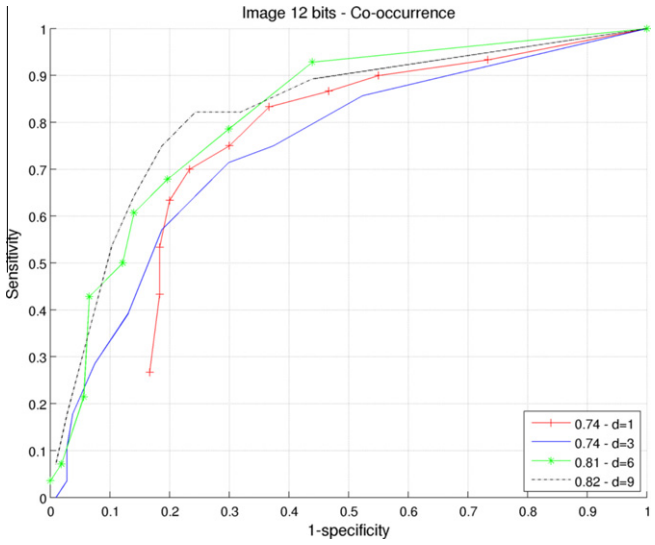

Fig. 8. ROC curves for SGLD matrices constructed at four pixel distances ($d = 1,3,6,9$), with no feature selection algorithm, for full images with 12 bits of resolution.

$d = 9$). Therefore, when comparing with the other two extraction methods, for now on we will use only the SGLD matrices obtained with distance $d = 9$.

In Figs. 9-11, we show the influence of crossover and mutation probabilities for the co-occurrence, wavelet and ridgelet based texture extraction methods, respectively, on the classification accuracy. Table 3 shows the best values of the mutation and crossover probability and the number of selected attributes for each method of texture analysis.

The GA procedure was able to reduce the number of attributes for the three methods. It can be seen that the large *reduction occurred* with attributes obtained from the co-occurrence matrix. We observed that the best reduction rates were obtained with the wavelet-based feature. Also, we have split the results into images with full resolution (12 bits) and subsampled images (8 bits). This has been done for computational complexity evaluation of the [i]three methods, once the more bits of resolution, the more computational complexity is present.

In Figs. 12 and 13, we have plotted the ROC curves with its corresponding AUC without considering feature selection, for images of 8 and 12 bits of resolution, respectively. Figs. 14 and 15 show the ROC curves and its AUC when using GA for feature selection, for 8 and 12-bit resolution images, respectively.

When comparing the results for full and subsampled images (8 and 12 bits of resolution), one can observe that the co-occurrence method is sensitive to changes on image resolution when applied matrices with pixels distance $d = 9$. When using GA, the AUC becomes higher when increasing resolution, going from 0.83 to 0.89. This sensitivity to resolution is due to the spatial dependency on pixel values of the SGLD matrix, which is lost with resampling. Therefore, the reduction of image resolution generated negative consequences on the classifier performance when we use the SGLD matrix to subsampled images, showing that is not interesting to reduce computational efforts when dealing with textural features. The use of GA for feature selection causes significant improvement on the classifier performance (7% and 6% gains for 8 and 12 bits, respectively).

The same behaviors are observed for the other two methods, in which spatial pixel dependency is not the main focus. For wavelet transform, there are slightly differences from the random classifier for subsampled images with no feature selection, with AUC = 0.51. Also, a great increase in performance occurs for full resolution images, with the AUC augmenting to 0.84. The same tendency is perceived when considering GA for feature selection, in which the AUC goes to 0.62 and 0.90 for 8 and 12 bits, respectively. Comparing performance improvements with and without the use of GA for feature selection, it can be noticed a greater efficiency of GA for subsampled images than for full images, with performance gains of 11% and 6%, respectively. That indicates that subsampling images gives rise to the advent of some redundancy on data textural features when wavelet transformed. However, even presenting a higher performance improvement when GA is applied, the AUC is still small for wavelet-based feature extraction on subsampled images compared to full images, either using GA or not.

The ridgelet-based procedure has little performance improvements when comparing subsampled and full images, with AUC

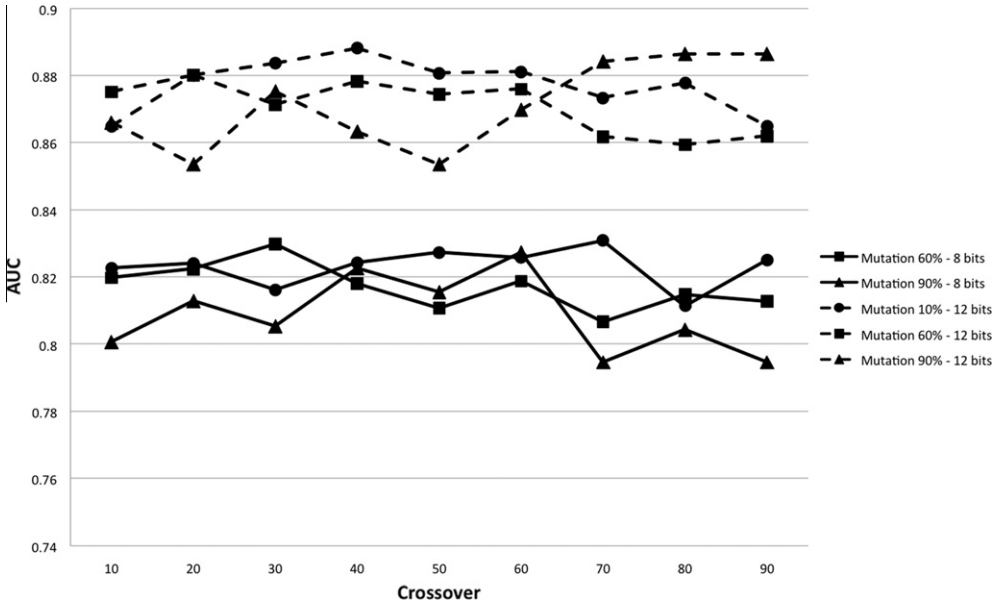

Fig. 9. Evolution of the influence of crossover and mutation probabilities on AUC for co-occurrence.

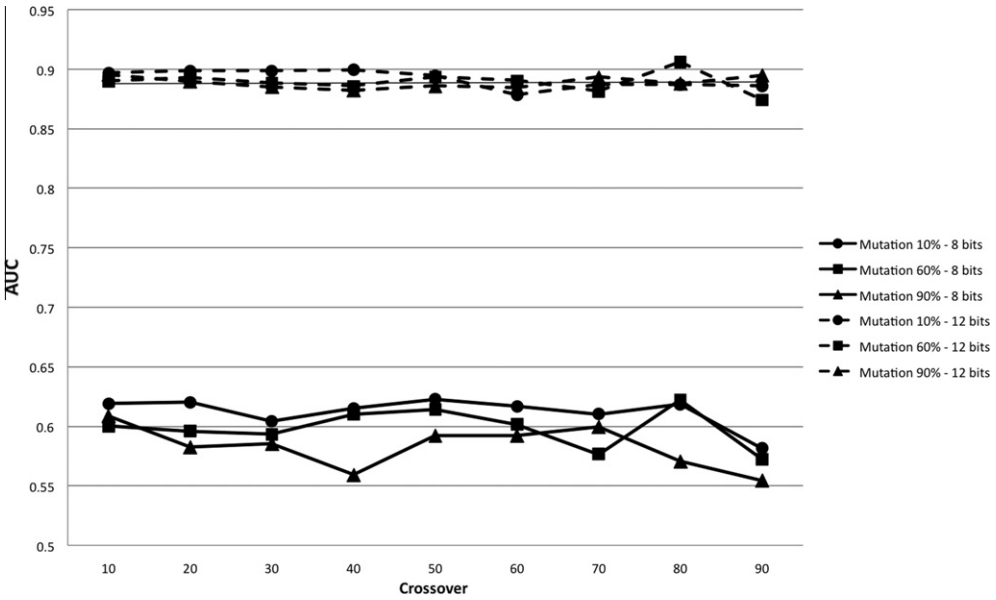

Fig. 10. Evolution of the influence of crossover and mutation probabilities on AUC for wavelet.

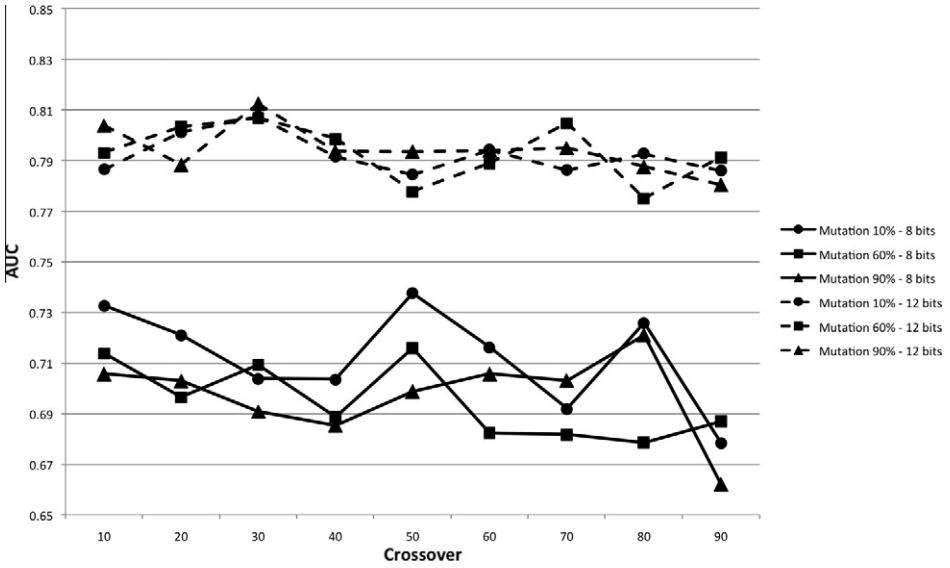

Fig. 11. Evolution of the influence of crossover and mutation probabilities on AUC for ridgelet.

Table 3
Parameters of GA used in each technique of texture analysis.

Technique	Image (resolution)	Cross-over probability	Mutation probability	Number of features
Co-occurrence	8-bits	0.7	0.1	10
	12-bits	0.4	0.1	11
Wavelet	8-bits	0.5	0.1	14
	12-bits	0.8	0.6	13
Ridgelet	8-bits	0.5	0.1	300
	12-bits	0.9	0.3	313

ranging from 0.63 to 0.66 with no features selection, and from 0.73 to 0.81 using GA. An interesting behavior occurs with the ridgelet-based approach when we observe the results with and without GA feature selection, namely the higher rate of performance increase compared to the other two methods. For subsampled images, the gain obtained with no feature selection is 10%, while the increase for full images is 15%. Since the ridgelet transform is by its nature redundant, it was already expected that the feature vectors originated by its application would lead to very redundant data, and GA deals effectively with this redundancy minimizing the number of relevant descriptors.

A surprising result appears when we compare the three proposed methods. Because of its large number of feature descriptors, ridgelet would be a natural candidate to present the best

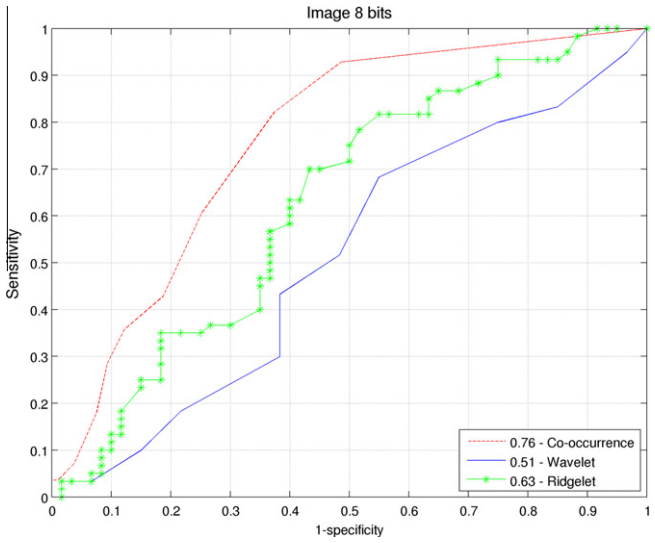

Fig. 12. ROC curves for SGLD, wavelet and ridgelet based methods, with no feature selection algorithm, for subsampled images (8 bits of resolution).

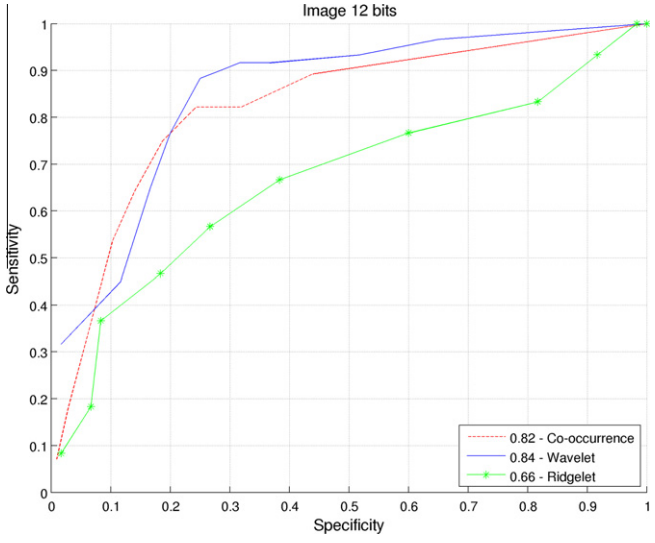

Fig. 13. ROC curves for SGLD, wavelet and ridgelet based methods, with no feature selection algorithm, for full images (12 bits of resolution).

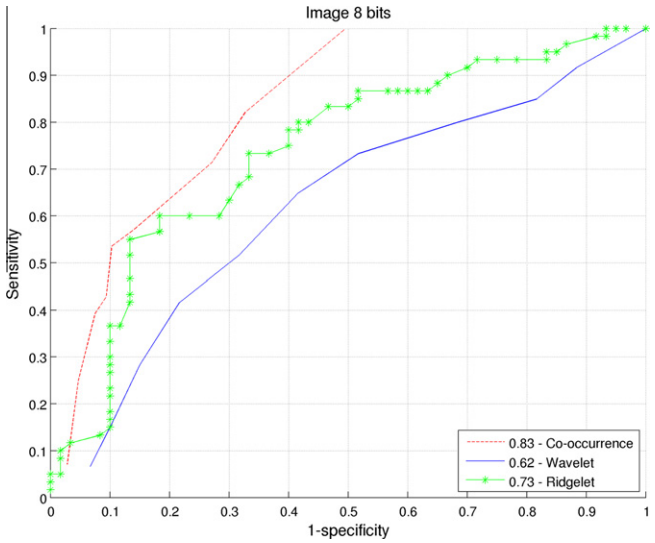

Fig. 14. ROC curves for SGLD, wavelet and ridgelet based methods, using GA for feature selection, for subsampled images (8 bits of resolution).

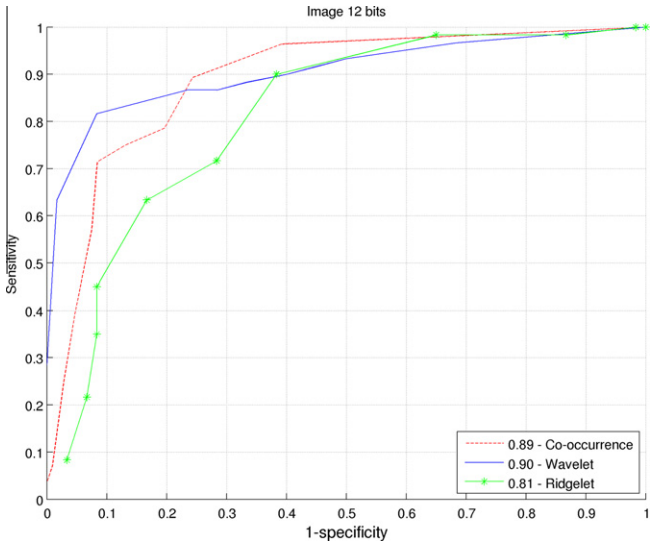

Fig. 15. ROC curves for SGLD, wavelet and ridgelet based methods, using GA for feature selection, for full images (12 bits of resolution).

performance results. However, it was never the case; either the SGLD-based or the wavelet-based methods performed better than ridgelet. Comparing the best cases of all three methods, it can be observed that wavelet performed better, with an AUC = 0.90, followed by SLGD, with AUC = 0.89, and ridgelet being the worst, with an AUC = 0.81.

When comparing the three methods, one has to make a clear distinction on computational complexity for each one. At the feature extraction step, care must be taken on the number of gray levels used and pixel distance to obtain the SGLD matrix, since its results is dependent on those parameters and affects the obtainment of textural descriptors. In the other two methods, computational complexity to acquire textural descriptors depends on the image size, on the decomposition levels for wavelet, and on the decomposition levels and radial directions for ridgelet. Both at the GA feature selection step, when used, and at the classification step, computational complexity is concerned about the number of descriptors, which is more significant for the ridgelet-based approach, for its higher number of descriptors.

Considering only the best cases, which were all obtained using GA for feature selection, the times taken by each method to generate the results are shown in Table 4. These results were acquired using a PC with a 1.5 GHz Intel Core2 Duo processor and 2 GB of RAM. It can be seen that the best computational time is achieved by the SGLD method, valid for subsampled mammograms. A close look at that table shows, however, that the overall time for the SGLD procedure is nearly the same as the wavelet-based for full images, which performed better in terms of AUC, making the latter the best choice. As it was expected for its large features vector, the ridgelet-based method presented the worst computational complexity, remarkably at the GA feature selection step, which took more than 13 times longer than wavelet.

For the sake of completeness of this article, we qualitative compare the techniques used in the present work with other important approaches utilized in the literature. In (Sahiner et al., 1996), the authors proposed a similar method based on GA and SGLD matrix and obtained an AUC of 0.90. Qian et al. (2001) used an approach based on wavelet transform and neural network to acquire an

Table 4
Computational complexity in terms of time to get the results for all images set.

Method	Feature extraction time (seconds)	GA feature selection time (seconds)	Classification time (seconds)
Wavelet	112.85	1320.04	3.20
Ridgelet	120.17	18030.42	12.10
SGLD	106.14	1080.30	1.00

AUC of 0.86 and 0.93, with a two-channel wavelet transform and an adaptive four-channel wavelet transform, respectively. It is important to note that our scheme has no adaptive iteration. Lladó et al. (2009) presented a work based on local binary patterns for textural extraction and a support vector machine for classification and obtained AUC results ranging from 0.91 to 0.94, where it was considered different ROI sizes.

4. Conclusion

Multiresolution techniques are very useful for image processing due to its similarity to the human visual system. Both systems process less image data by selecting relevant details to perform visual recognition tasks. In this work, we proposed a FPR algorithm for a CADx system. We used two multiresolution techniques, namely wavelet and ridgelet transforms, to extract textural features from mammogram images for breast cancer classification. It was also considered a genetic algorithm processing for feature selection purposes. We compared both methods with the well know co-occurrence based method. Five textural descriptors were extracted from each image and used for classification using the Random Forest algorithm. Using GA procedures, it was possible to obtain a new reduced set of features with improved classification ratios. Performance evaluation was computed in relation to the area under the ROC curve.

Experimental results showed that the best classification rates were obtained with the wavelet-based feature extraction using GA for selection of the most relevant features, giving an AUC = 0.90 with a negligible computational complexity loss compared to the best time consumption method based on the SGLD matrix. Surprisingly, once for its rather superior number of feature descriptors, the ridgelet-based approach showed smaller best AUC than the wavelet-based algorithm. For all methods, it was observed that GA helps to improve classification performance, and the SGLD-based and ridgelet-based methods are less sensitive to full images than the other methods. Furthermore, the results obtained in the present work were similar to those obtained in recent works on the literature.

The procedures developed in this article are, by no means, completely generalizable. For example, the use of different classifiers may lead to different results. As well, the inclusion of different texture features, other mammogram data sets, and different selection algorithms may have other influences on the FPR algorithm performance. These are possibilities to be explored in future works.

Acknowledgments

The authors would like to thank the financial support of FAPESP.

References

Bovis, K., & Singh, S. (2000) Detection of masses in mammograms using texture features. In *Proceedings of the 15th international conference on pattern recognition,* 2, pp. 267–270.
Breiman, L. (2001). Random forests. *Machine Learning, 45*, 5–32.
Candès, E. J., & Donoho, D. L. (1999). Ridgelets: A key to higher-dimensional intermittency? *Philosophical Transactions of the Royal Society of London Series A-Mathematical Physical and Engineering Sciences*, 2245–2509.
Deans, S. R. (1983). *The Radon transform and some of its applications*. New York: John Wiley and Sons.
DeCoursey, W. J. (2003). Statistics and probability for engineering applications. Newnes.
Dettori, L., & Semler, L. (2007). A comparison of wavelet, ridgelet and curvelet-based texture classification algorithms in computed tomography. *Computers in Biology and Medicine, 37*, 486–498.
Do, M. N., & Vetterli, M. (2003). The finite ridgelet transform for image representation. *IEEE Transactions on Image Processing, 12*, 16–28.
Elter, M., & Horsch, A. (2009a). CADx of mammographic masses and clustered microcalcifications: A review. *Med. Phys., 36*, 2052–2068.
Elter, M., & Horsch, A. (2009b). CADx of mammographic masses and clustered microcalcifications: A review. *Medical Physics, 36*, 2052–2068.
Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters, 27*, 861–874.
Gonzalez, R., & Woods, R. (2002). Digital Image Processing, Addison-Wesley, 2nd Ed.
Gromet, M. (2008). Comparison of computer-aided detection to double reading of screening mammograms: Review of 231,221 mammograms. *American Journal of Roentgenology, 190*, 854–859.
Gupta, S., Chyn, P. F., & Markey, M. K. (2006). Breast cancer CADx based on BI-RADS descriptors from two mammographic views. *Medical Physics, 33*, 1810–1817.
Gupta, S., & Markey, M. K. (2005). Correspondence in texture features between two mammographic views. *Medical Physics, 32*, 1598–1606.
Haralick, R. M., Shanmugam, K., & Dinstein (1973). Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, 10*, 610–621.
Heath, M., Bowyer, K., Kopans, D., Kegelmeyer, P., Moore, R., Chang, K., et al. (1998). Current status of the digital database for screening mammography. *Digital Mammography*, 457–460.
Instituto Nacional do Câncer (INCA), (2010). Estimativa da Incidência e Mortalidade por Câncer no Brasil. Available at http://www2.inca.gov.br/wps/wcm/connect/tiposdecancer/site/home/mama/cancer_mama.
Kohavi, R., & Provost, F. (1998). Glossary of terms. *Machine Learning, 30*, 271–274.
Lim, W. K., & Er, M. J. (2004). Classification of mammographic masses using generalized dynamic fuzzy neural networks. *Medical Physics, 31*, 1288–1295.
Liu, S., Babbs, C. F., & Delp, E. J. (2001). Multiresolution detection of spiculated lesions in digital mammograms. *IEEE Transactions on Image Processing, 10*, 874–884.
Lladó, X., Oliver, A., Freixenet, J., Martí, R., & Martí, J. (2009). A textural approach for mass false positive reduction in mammography. *Computerized Medical Imaging and Graphics, 33*, 415–422.
Mallat, S. (1996). Wavelets for a Vision. In *Proceedings of the IEEE*, 84, pp. 604–614.
Mallat, S. (1999). *A wavelet tour of signal processing* (2nd ed.). Academic Press.
Masotti, M., Lanconelli, N., & Campanini, R. (2009). Computer-aided mass detection in mammography: False positive reduction via gray-scale invariant ranklet texture features. *Medical Physics, 36*, 311–316.
Mavroforakis, M. E., Georgiou, H. V., Dimitropoulos, N., Cavouras, D., & Theodoridis, S. (2006). Mammographic masses characterization based on localized texture and dataset fractal analysis using linear, neural and support vector machine classifiers. *Artificial Intelligence in Medicine, 37*, 145–162.
Mazurowski, M. A., Habas, P. A., Zurada, J. M., & Tourassi, G. D. (2008). Decision optimization of case-based computer-aided decision systems using genetic algorithm with application to mammography. *Physics in Medicine and Biology, 53*, 895–908.
Metz, C. E. (1978). Basic principles of roc analysis. *Seminars in Nuclear Medicine, 8*, 283–298.
Metz, C. E., Herman, B. A., & Shen, J. H. (1998). Maximum likelihood estimation of receiver operating characteristic (roc) curves from continuously-distributed data. *Statistics in Medicine, 17*, 1033–1053.
Moayedi, F., Azimifar, Z., Boostani, R., & Katebi, S. (2010). Contourlet-based mammography mass classification using the SVM family. *Computers in Biology and Medicine, 40*, 373–383.
Mousa, R., Munib, Q., & Moussa, A. (2005). Breast cancer diagnosis system based on wavelet analysis and fuzzy-neural. *Expert Systems with Applications, 28*, 713–723.
Mudigonda, N. R., Rangayyan, R. M., & Desautels, J. E. (2001). Detection of breast masses in mammograms by density slicing and texture flow-field analysis. *IEEE Transactions on Medical Imaging, 20*, 1215–1227.
Oliver, A., Freixenet, J., Martí, J., Pérez, E., Pont, J., Denton, E. R. E., et al. (2010). A review of automatic mass detection and segmentation in mammographic images. *Medical Image Analysis, 14*, 87–110.
Paquerault, S., Petrick, N., Chan, H. P., Sahiner, B., & Helvie, M. A. (2002). Improvement of computerized mass detection on mammograms: Fusion of two-view information. *Medical Physics, 29*, 238–247.
Qian, W., Sun, X., Song, D., & Clark, R. A. (2001). Digital mammography: Wavelet transform and Kalman-filtering neural network in mass segmentation and detection. *Academic Radiology, 8*, 1074–1082.
Rashed, E. A., Ismail, I. A., & Zaki, S. I. (2007). Multiresolution mammogram analysis in multilevel decomposition. *Pattern Recognition Letters, 28*, 286–292.
Sahiner, B., Chan, H. P., Petrick, N., Helvie, M. A., & Hadjiiski, L. M. (2001). Improvement of mammographic mass characterization using spiculation measures and morphological features. *Medical Physics, 28*, 1455–1465.
Sahiner, B., Chan, H. P., Wei, D., Petrick, N., Helvie, M. A., Adler, D. D., et al. (1996). Image feature selection by a genetic algorithm: Application to classification of mass and normal breast tissue. *Medical Physics, 23*, 1671–1684.
Selvan, S., & Ramarishnan, S. (2007). SVD-based modeling for image texture classification using wavelet transformation. *IEEE Transactions on Image Processing, 16*, 2688–2696.

Selvi, S. T., & Malmathanraj, R. (2006). Segmentation and SVM Classification of Mammograms. In *Proceedings of the IEEE International Conference on Industrial Technology,* pp. 905–910.

Semler, L., Dettori, L., & Kerr, W. (2006). Ridgelet-based texture classification in computed tomography. In *Proceedings in the eighth IASTED international conference on signal and image processing, Honolulu*.

Starck, J. L., Elad, M., & Donoho, D. L. (2004). Redundant multiscale transforms and their application for morphological component analysis. *Advances in Imaging and Electron Physics, 132*, 287–348.

Tang, J., Rangayyan, R. M., Xu, J., El, N., & Yang, Y. (2009). Computer-aided detection and diagnosis of breast cancer with mammography: Recent advances. *IEEE Transactions on Information Technology in Biomedicine, 13*, 236–251.

Vibha, L., Harshavardhan, G. M., Pranaw, K., Deepa Shenoy, P., Venugopal, K. R., & Lalit, M. P. (2006). Classification of mammograms using decision trees. *IDEAS*, 263–266.